2019 19th International Symposium on Communications and Information Technologies (ISCIT)

MB-PBA: Leveraging Merkle Tree and Blockchain to Enhance User Profile-based Authentication in E-Learning Systems

Duy-Minh Nguyen, Quang-Huan Luu, Nguyen Huynh-Tuong and Hoang-Anh Pham
Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology, VNU-HCM
{1870042, 1770474, htnguyen, anhpham}@hcmut.edu.vn

***Abstract*—User authentication in online systems is always a big challenge in determining whether the logged-in user is the legitimate account owner. This challenge becomes critical in E-learning systems since it affects the assessment result of learners accurately. User profile-based authentication (PBA) has shown its advantages compared with other existing approaches. The PBA system will generate random questions and then evaluate the users' answers based on the collected information when the user registered into the system. However, this approach may reveal the users' data when the system is compromised and exploited. In this paper, we utilize the dominant features of the Merkle tree data structure and smart contracts on Blockchain to improve the PBA approach. The proposed method simplifies identity authentication, protects user confidential information, and enhances the security factor in terms of data integrity, user information transparency, and system availability. Our preliminary experiments have shown the potential and feasibility of the proposed authentication method.**

***Index Terms*—authentication, e-learning, privacy, merkle tree, blockchain**

I. Introduction

In recent years, online training becomes an attractive choice instead of the full-time on-campus training model due to the reasonable costs and flexible study time. Many leading universities such as the Massachusetts Institute of Technology, Harvard University, and the University of Pennsylvania have launched online programs for numerous majors up to the master level.

The online training is facilitated by a software system known as E-learning system to manage the teaching and learning process [1]. With only an Internet-connected computer, students no longer have to go to the campus for attending the lectures and meeting their instructors. Instead, they can remotely access learning materials (e.g., lecture notes and books), ask questions, do assignments, and take exams like the traditional training model. However, the online training model asks for efforts and truthfulness from its learners. Because learners and teachers are not meeting and interacting personally, it is easy for the students to cheat deliberately.

The E-learning systems must demonstrate its reliability and bring trust to users and society regarding its training quality and transparency (e.g., how to know if a student's performance in the system is indeed his or hers). The critical challenge of E-learning systems is the ability to authenticate the person who is accessing the course resources, performing learning activities, and especially taking online exams. From computer science's perspective, it is a matter of identifying and referencing a real-world person as a digital entity in digital systems, which leads to several security issues needed to be addressed thoroughly [2], [3].

Additionally, in the traditional on-site training model, academic records, including transcripts and examination results, have been stored and managed via paper-based documents that are currently digitalized and saved as a digital copy. However, digital data seems more likely to be manipulated and altered than physical data [4], which is same in E-learning systems. Therefore, it is a considerable challenge regarding data storage and management in E-Learning systems. It must be transparent, meanwhile assuring data security and privacy.

In this paper, we address two challenging questions leading to security issues in E-learning systems as follows:

1) How can an E-learning system handle identity misuse?
2) How can an E-learning system ensure data integrity of students' academic results?

Regarding identity misuse, the major challenge to have an appropriate and effective authentication mechanism to examine the users whenever they log in the system for ensuring their legitimate identity to mitigate exam cheating, involuntary or voluntary tampering, and impersonation. For example, the student can actively share the account information to his/her friends to help in taking the online tests, which is a typical situation for most E-learning platforms. In another scenario, the student account is hacked and manipulated in unexpected such as the students' profile can be leaked, and the academic results can be altered. This aspect concerns the data integrity of students' records.

Many research efforts have been studied over the past years to tackle the authentication in E-learning systems such as userid and password-based, biometric-based, user profile-based, and multifactor-based approaches. In this paper, we leverage the dominant features of the Merkle tree and Blockchain to improve the user profile-based authentication for mitigating identity misuse and enhancing data integrity of students' records.

The remainder of this paper is organized in the following manners. Section 2 briefly summarizes our review of the existing authentication approaches in terms of advantages and disadvantages. Section 3 describes our proposed authentication

978-1-7281-5009-3/19/$31.00 ©2019 IEEE

method in details. Several preliminary experiments to evaluate our proposed approach are presented and discussed in Section 4, and Section 5 provides the final concluding remarks and future works.

II. Related Work

Many research efforts have been studied to tackle the challenges in online system authentication. In this section, we briefly review several existing methods, including traditional ones (e.g., password/ID-based authentication) and modern ones (e.g., biometric-based, smartcard-based, user profile-based, and multifactor-based authentication). We summarize the advantages and disadvantages of these methods in Table I.

In [5], [6], the authors proposed various solutions of biometric-based approach. Additionally, a group of researchers in Pakistan [7] has recently introduced biometrics and IoT-combined solution. These biometrics-based models often require dedicated devices (e.g., fingerprint reader). Therefore, these approaches are not much feasible for E-learning systems where students have to equip biometrical devices with high costs in general.

With the rapid advancement of image processing systems in recent years, the image-based authentication methods have been investigated and delivered positive results such as solutions proposed by Andeep S. Toor et al. [8] and Pavel D.Gusev et al. [9]. However, accuracy in the image-based authentication is still open questions while this approach also faces the risks of user images being back-and-forth attacked.

As the growth of E-learning models, the research interests in mining education data have significantly attracted research community. The authentication challenge can be solved by applying user attribute and behavior. Several studies have been used to authenticate online users based on a set of challenging questions [10], [11]. In addition to the usual security questions regarding personal information, users will be asked by challenging questions related to their activities in the past. This method does not require any additional hardware capabilities like in the biometrics-based methods.

Another advanced approach so-called profile-based authentication (PBA) that utilizes multi-factor knowledge such as login identifier and password and challenging questions for the authentication process. For example, Fig. 1 describes a profile-based authentication framework, in which after being asked for a username and password to log in to the system, users will be asked to answer challenging questions for authentication. These questions are generated from the user's profile, so only users can know precisely this information. The most advantage of PBA method is not affected by external factors (e.g., camera and biometric reader) that interfere with the authentication process, which makes this authentication method consistent with the E-learning system. However, this method still has limitations as follows:

- The privacy of user's personal information is not guaranteed;
- The user profile database can be attacked and changed in order to pass the authentication process;
- It does not guarantee the data integrity of user records such as student academic results;
- It can not against replay/playback attacking.

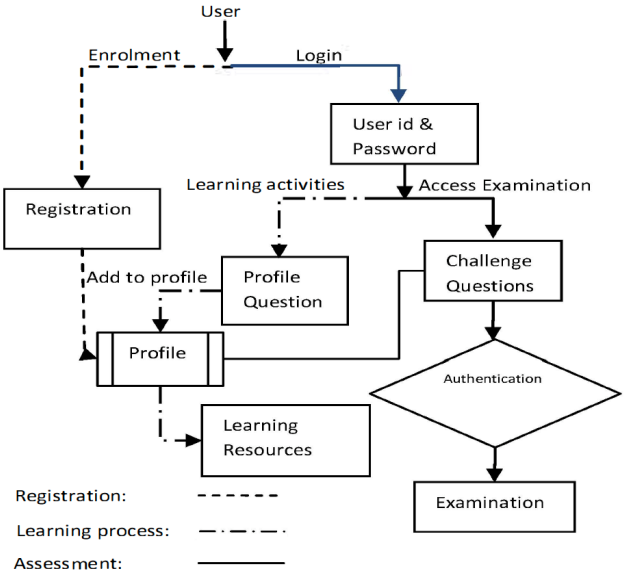

Fig. 1. Profile-Based Authentication Framework [12]

III. The Proposed Approach

Whenever users participate in the system at the beginning, users are required to answer several security questions so-called *security/secret question set* in order to complete the registration process. These provided answers are stored to build users' profile that will be used to generate challenging questions to authenticate users. However, different from the traditional PBA methods, our proposed method will process and reorganize the answers of the security question set in a secure way to protect user privacy by utilizing Merkle tree structure and Blockchain.

Additionally, users are suggested to provide extended information to improve system security. Besides, while users are using the system, their activities and records will be collected and processed to build user profiles. The challenging questions generated from user profiles are called the *profile question set*.

These security and profile question sets are then used to generate challenging questions to authenticate users whenever they log in to the system to access system resources for learning or taking tests.

Fig. 2 depicts the overall framework of the proposed approach, so-called MB-PBA, in which we summarize our contributions as follows:

- Proposing a secure method to store user profiles by Merkle tree structure and Blockchain.
- Proposing an efficient authentication process to maintain the accuracy and transparency at a high degree.

TABLE I
REVIEW OF POPULAR AUTHENTICATION TECHNOLOGIES

Authentication Model	Advantages	Disavantages	References
Password/ID based	- Simple and easy to use. - Low cost.	- Low entropy passwords are prone to dictionary attacks. - Have to regular renewal to keep the security.	[12], [13]
User profile based	- Don't require integration and/or additional hardware. - Low cost.	- The privacy of user data should be considered. - Does not guarantee user data integrity.	[12], [14]
Smart card based	- Multiservice and flexibility. - Easy to use. - Data integrity.	- Can be used only with the help of special devices called "smart card readers". - It gives liability issues if stolen or lost. - The accuracy of information is small.	[15], [16]
Biometrics-based	- Improved customer experience. - Cannot be forgotten or lost.	- Require integration and/or additional hardware. - Biometric features are not useable if compromised. - Environment and usage can affect measurements. - High cost.	[12], [17], [8]
Multifactor based	- More factors used to determine a person's identity, the greater the trust of authenticity.	- Inconvenient and logistically difficult. - High cost.	[17], [18]

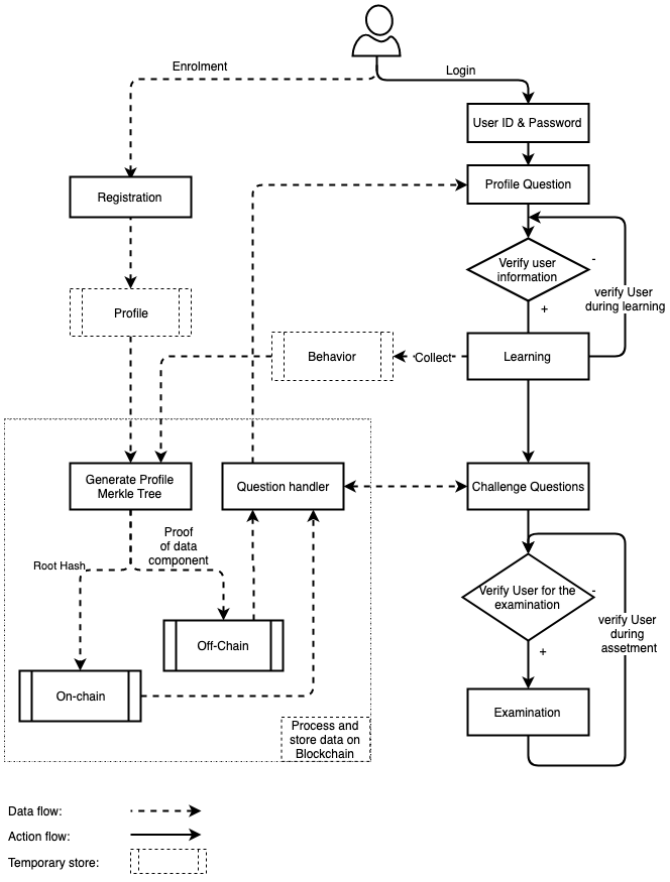

Fig. 2. The proposed MB-PBA Framework

A. Secure Method to Store User Profiles

User profiles are always updating and expanding by the time, so data structures storing user profiles plays a significant role in the system. With these strict requirements, the compact spare Merkle tree (CSMT) is an appropriate choice due to its outstanding features [19] as follows:

- *Support operations*: insert, delete, (non-) membership proof;
- *Data structure*: CSMT is nearly balanced. We can reach to any particular entry in $log_2(n)$ step if the entry exists in the tree.
- *Space*: is bounded by twice the number of keys in the tree. $\sum_{i=0}^{\lceil log_2(n) \rceil} 2^i = 2^{\lceil log_2(n) \rceil + 1} - 1 < 2n$
- *Max proof size*: if SHA256 is used, the max proof size would be $2 \times 256 \times log_2(n)$.

For a data component of user profiles that can be decomposed into a certain number of elements, the traditional Merkle Tree (MT) data structure will be applied due to its advantage of tree construction speed. The tree root value will become a component of the CSMT. Additionally, we utilize blockchain technology to store user authentication evidence for ensuring the privacy and integrity of personal user data, in addition to improving the system's defense against data alteration or falsification attacks. However, if entire data is stored in the blockchain, it becomes more difficult and expensive in synchronization [20]. Therefore, we combine off-chain and on-chain storages in the proposed MB-PBA framework.

Users academic results are also an essential part of the user profile, which means that the user profile will be continuously growing as well as diversifying challenging question banks. Besides, taking advantage of the immutable nature of the blockchain will ensure data integrity of students academic results.

In our proposed approach, we decompose user profile's raw data into sections to build the CSMT, in which the Merkle root value will be stored on the blockchain by smart contracts. Then, this root value will be used in the authentication process as the evidence to validate user answers for challenging questions. Meanwhile, each leaf node storing the corresponding audit path will be kept off-chain to save costs and speed up

processing. This audit path, which is a hash value itself, will be taken together with the hash value from the user answers to recalculate the Merkle root. If this new root is matching with the root value stored on-chain, it means the user answers are valid; otherwise, it may be identity fraud or misuse.

Our proposed method not only makes authentication process more effective because it avoids asking users for too much sensitive personal information, but also saves user data entirely confidential because the system store only the audit paths represented by hash values.

B. Efficient Authentication Process

In E-learning systems, before accessing the system resource (e.g., lessons, tests), the student will be asked to log-in by username and password. If there are abnormal signs in the process of using the system, the student will be asked challenging questions to validate his/her legitimate account ownership.

The system randomly generates challenging questions by extracting personal information (e.g., birth date, national ID number, student ID number, parents' full names, and academic results) and user activities in the past (e.g., the name of a course the user has previously enrolled in, and the previous email address associated with this account).

User authentication by challenging questions can protect against sniffing attacks where intruders can obtain users' login credentials by capturing network traffic (e.g., accounts and passwords). In the session of a replay attack, the attacker listens to previously exchanged packets and re-sends them to impersonate as the authorized user in the original session. With a good enough set of challenging questions, the system can resist replay attacks because each pair of challenging question and answer is unique at different moments. If an attacker monitors the exchange of login information in the past and tries to reuse this information, he will not succeed in gaining access.

However, after monitoring the authentication procedure for a long enough time, the attacker will know the user's personal information when answering challenge questions. In order to enhance security, the answers are encrypted or hashed before being sent to the server for avoiding user information exposed. With the blockchain-based approach, the answer will be hashed and combined with the corresponding audit path to recalculate the Merkle root value. This value will be then compared to the Merkle root saved on the blockchain to validate the answers of challenging questions. Thanks to this process, personal user data are secured, while the accuracy and transparency of the authentication process are preserved at a high degree since the nature of blockchain guarantees that on-chain data cannot be altered illegally to bypass the identity check.

Fig. 3 depicts our proposed authentication process, as described above. For answers to secret questions, this process also ensures the personal user information cannot be revealed during authentication, thereby withstanding the risks of network intrusion attacks. The exchanged data over the network is entirely encrypted to protect users' confidential information.

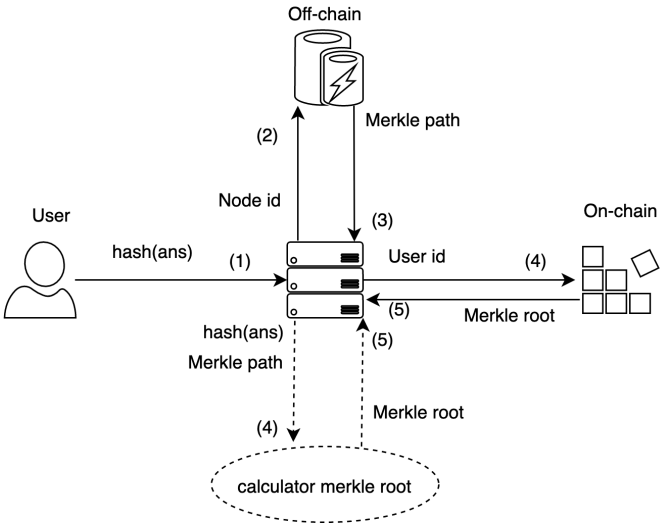

Fig. 3. Process of validating user's answer

IV. Experiment and Evaluation

A. Implementation and Experiment

We implement three following function modules as well as perform several preliminary experiments to evaluate the proposed BM-PBA framework.

1) Build Merkle tree (MT) and compact spare Merkle tree (CSMT): A user profile consists of three attributes that will be constructed under MT and CSMT structure, as shown in Fig. 4, in which data component with key 3 contains information about phone numbers, organized into an MT structure and the original value of this MT will be the component value of CSMT. Proof of each component of the data will be calculated to verify the data later. Proof of the 2^{th} data component is described as: $\{"left" : "52a99d...74402f", "key" : 1\}, \{"right" : "f66abc...dc6a84", "key" : 7\}$.

To assess the scalability of the authentication method, we conducted numerous experiments to measure and evaluate processing time for MT and CSMT construction as well as its storage space. The experimental results, shown in Table II and Fig. 5, indicate that the MT is perfectly appropriate for data components with a fixed number of elements (e.g., phone number, and date of birth) because of its fast processing speed. With the advantage of storage space, as well as the ability to insert new data, CMST represents the suitability for extending user profile structure.

2) Smart contract to manage proof of user profile: In the proposed MB-PBA framework, when a user completes answering security questions or profile question, these answers will be processed and then stored on the Blockchain by calling the corresponding function in smart contracts as an implemented shown in Fig. 6.

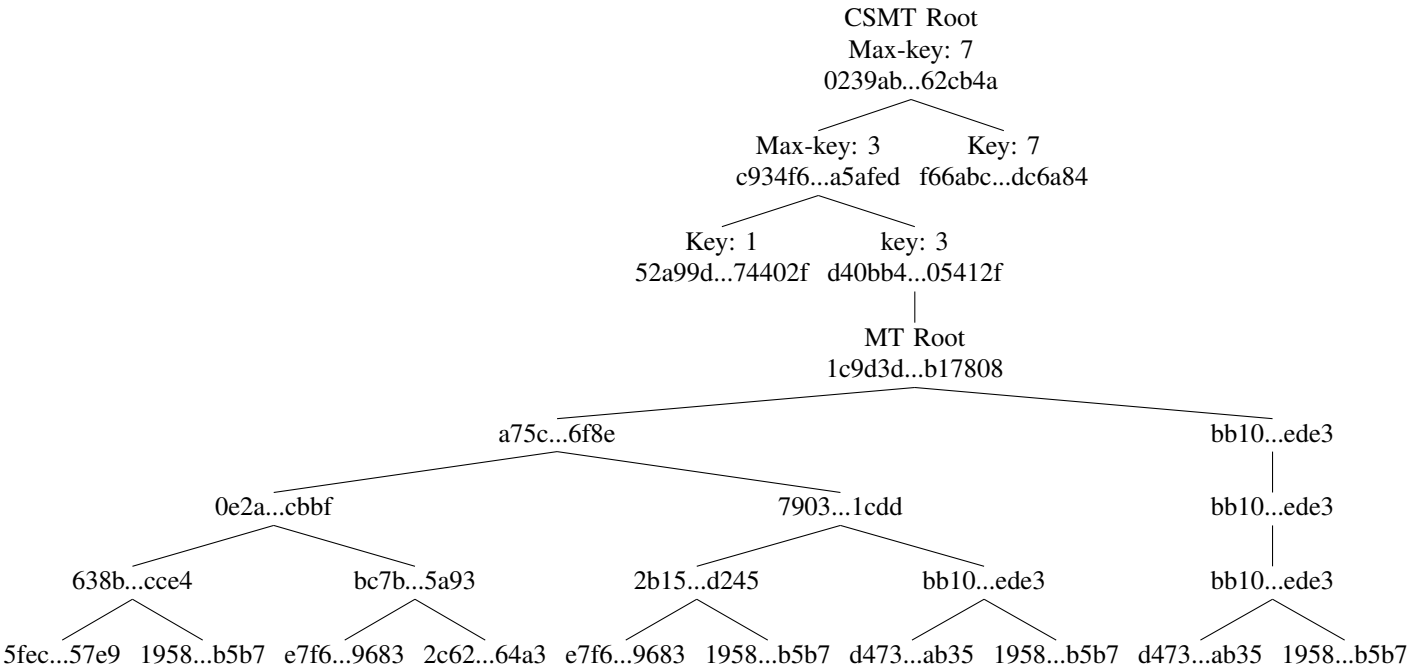

Fig. 4. User profile has three attributes in our data strucutre approach

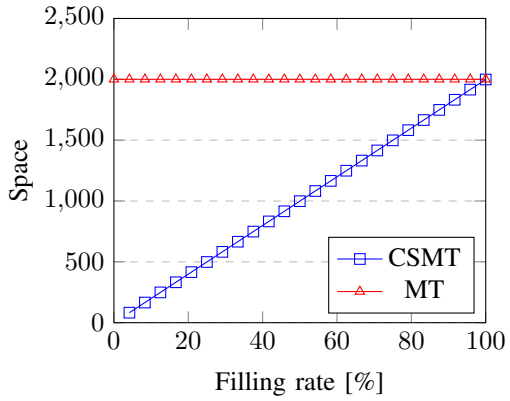

Fig. 5. Space of CSMT/MT has 1000 data compenents with filling rate

```
CONTRACT
AuthChainBK

FUNCTION
append_profile_hash(user_id: uint256, value: string)

INPUTS
1870042, 0239abaf58d97924438fd29550e7a8953157f4e8ac5c59344660bb4d7f62cb4a
```

Fig. 6. Call function of smart contract to store evidence of information

3) Challenging question to authentication: Whenever a user successfully logs in the system, the user will be redirected to the challenging questions page for authentication as Fig. 7. The challenging questions are randomly generated from the collected security questions set and profile questions set.

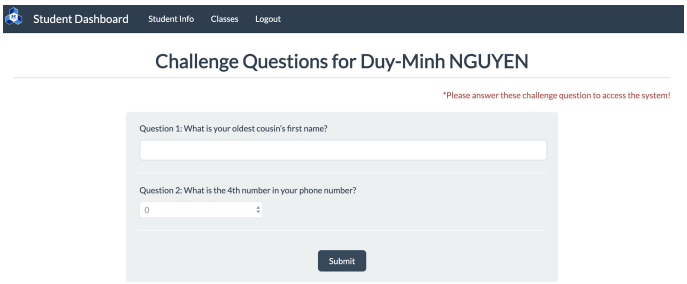

Fig. 7. Challenging questions page for authentication

B. Evaluation and Discussion

The computation time increases when the size of the users' profile becomes larger. Merkle tree construction has a linear relationship with the number of data component (i.e., leaf nodes). Meanwhile, the computation time for audit path (proof) generation and verification is logarithmically proportional to the number of data blocks because the number of elements is related to the tree height. The experimental results summarized in Table II have shown that the proposed MB-PBA shortens the authentication process and improves user experience.

Based on the theoretical background and empirical results of the proposed authentication method after being implemented and integrated into an E-learning management system, Table III summarizes some objective assessments of security features for an information system (CIA triad).

Unlike the PBA method, the user's personal information in MB-PBA is stored in the compact spare Merkle tree structure,

TABLE II
THE EXPERIMENTAL COMPUTATION TIME (SECONDS)

Number of data components	Merkle tree generation	Audit path (proof) generation	Verification answer
10	0,00003267	0,00001233	0,00001566
50	0,00009821	0,00001438	0,00001899
100	0,00018302	0,00001521	0,00002193
150	0,00027680	0,00001633	0,00002376
200	0,00037893	0,00001772	0,00002527
250	0,00042041	0,00002167	0,00002575
300	0,00055822	0,00002289	0,00002599

TABLE III
COMPARISON BETWEEN CONVENTIONAL PBA AND MB-PBA METHODS

Criteria comparison	PBA	MB-PBA
1. Confidentiality & Data privacy		
1.1 Is store user infomation as clear text?	Yes	No
1.2 Is tranfer user infomation as clear text?	Yes	No
1.3 Can against replay attack/playback attack?	No	Yes
1.4 Can against keylogger (password attack)?	Yes	Yes
1.5 Can against privacy leak?	No	Yes
2. Integrity		
2.1 Can resist attack to database ?	No	Yes
2.2 Can the system administrator modify the data?	No	Yes
3. Availability		
3.1 Is the data distributed distributed ?	No	Yes
3.2 Does academic results be accessed even if the system is down?	No	Yes

in which data is hashed to protect data privacy. Besides, MB-PBA ensures data integrity and transparency with on-chain storage. The proposed design can resist attacks on databases where the attacker cannot replace any part or whole CSMT.

Compared to the traditional centralized model, the Blockchain-based model provides availability in nature. Therefore, our proposed method can provide high available authentication service.

By requiring answers to challenging questions, including security and profile questions, it can prevent the disclosure of login information intentionally or unintentionally hacked because only authentic user can exactly know complete information for a long time. Consequently, it mitigates the identity misuse.

V. CONCLUSION

By utilizing Blockchain and Merkle tree structure, we proposed a secure method to store user profile and an efficient authentication process to improve the existing PBA framework. The proposed MB-PBA framework not only enhances the data security and integrity of user-profiles but also mitigates identity misuse that is a big challenge in E-learning systems. Our preliminary experiments and evaluation demonstrate the feasibility of the MB-PBA.

For future works, we will conduct large-scale experiments to evaluate the effectiveness and usability of the proposed framework. In addition, how to store and generate challenging questions based on user behavior efficiently is also a difficult task in profile-based authentication approaches.

ACKNOWLEDGEMENTS

This research was supported by Infinity Blockchain Labs (IBL), Vietnam Blockchain Corporation (VBC).

REFERENCES

[1] Ryann K Ellis. Learning Management Systems. *ASTD learning circuits*, pages 1–7, 2009.
[2] N.H.M. H M Alwi and Ip-Shing Fan. Information security management in e-learning. (July 2015):1–6, 2014.
[3] Christophe Kiennert, Pierre Olivier Rocher, Malinka Ivanova, Anna Rozeva, Mariana Durcheva, Joaquin Garcia-alfaro, Christophe Kiennert, Pierre Olivier Rocher, Malinka Ivanova, Anna Rozeva, and Mariana Durcheva. Security challenges in e-assessment and technical solutions To cite this version : HAL Id : hal-01699388. 2018.
[4] Jorge Miguel, Santi Caballé, and Fatos Xhafa. Security for e-Learning. *Intelligent Data Analysis for e-Learning*, pages 7–23, 2016.
[5] Kornelije Rabuzin, Miroslav Bača, and Mario Sajko. E-learning: Biometrics as a security factor. *Proceedings of the International Multi-Conference on Computing in the Global Information Technology, ICCGI'06*, 00(c), 2007.
[6] P. Punithavathi, S. Geetha, Marimuthu Karuppiah, SK Hafizul Islam, Mohammad Mehedi Hassan, and Kim Kwang Raymond Choo. A lightweight machine learning-based authentication framework for smart IoT devices. *Information Sciences*, 484:255–268, 2019.
[7] Muhammad Farhan, Sohail Jabbar, Muhammad Aslam, Mohammad Hammoudeh, Mudassar Ahmad, Shehzad Khalid, Murad Khan, and Kijun Han. IoT-based students interaction framework using attention-scoring assessment in eLearning. *Future Generation Computer Systems*, 79:909–919, 2018.
[8] Andeep S. Toor, Harry Wechsler, Michele Nappi, and Kim Kwang Raymond Choo. Visual Question Authentication Protocol (VQAP). *Computers and Security*, 76:285–294, 2018.
[9] Pavel D. Gusev and Georgii I. Borzunov. The analysis of modern methods for video authentication. *Procedia Computer Science*, 123:161–164, 2018.
[10] Abrar Ullah, Hannan Xiao, Mariana Lilley, and Trevor Barker. Using Challenge Questions for Student Authentication in Online Examination. *International Journal for Infonomics*, 5(3/4):631–639, 2016.
[11] Libor Juhaák, Jií Zounek, and Lucie Rohlíková. Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92:496–506, 2019.
[12] A Ullah, Hannan Xiao, and Mariana Lilley. Profile Based Student Authentication in Online Examination. *International Conference on Information Society (i-Society)*, pages 109–113, 2012.
[13] Huiping Jiang. Strong password authentication protocols. *ICDLE 2010 - 2010 4th International Conference on Distance Learning and Education, Proceedings*, pages 50–52, 2010.
[14] Abrar Ullah, Hannan Xiao, Trevor Barker, and Mariana Lilley. Graphical and text based challenge questions for secure and usable authentication in online examinations. *2014 9th International Conference for Internet Technology and Secured Transactions, ICITST 2014*, pages 302–308, 2014.
[15] Sahu Ajay Kumar, Kumar Ashish, and Gupta Tarun. Survey of Remote User Password Authentication Scheme Using Smart Cards. *International Journal of Advanced Research*, 3(4), 2015.
[16] Amin Abd Elwahab, Ayman M.Bahaa Eldin, Ayman M. Wahba, and Mohamed A. Sheirah. A security layer for smart card applications authentication. *Proceedings - The 2009 International Conference on Computer Engineering and Systems, ICCES'09*, pages 514–517, 2009.
[17] Shu Yun Lim, Pascal Tankam Fotsing, Abdullah Almasri, Omar Musa, Miss Laiha Mat Kiah, Tan Fong Ang, and Reza Ismail. Blockchain Technology the Identity Management and Authentication Service Disruptor: A Survey. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4-2):1735, 2018.
[18] Aleksandr Ometov and Sergey Bezzateev. Multi-factor authentication: A survey and challenges in V2X applications. *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, 2017-November:129–136, 2018.
[19] Faraz Haider. Compact Sparse Merkle Trees. pages 1–8, 2018.
[20] Giulia Fanti, Nina Holden, Yuval Peres, and Gireeja Ranade. Communication cost of consensus for nodes with limited memory. pages 1–62, 2019.