

Application of Blockchain and Fully Homomorphic Encryption in Online Questionnaires

Xinhong Yu *

Changwang School of Honors, Nanjing University of Information Science & Technology, Nanjing, China

* Corresponding author email: 201983190025@nuist.edu.cn

Abstract. An online questionnaire is a simple and efficient method of acquiring diverse data for surveys. However, a proportion of respondents have strong concerns about whether questionnaires in this form will leak their privacy, which causes their unwillingness to fill out the questionnaires. Blockchain is well known for decentralization, traceability, and transparency and is now widely used in all kinds of platforms. Fully homomorphic encryption is another popular technique that ensures the safety of privacy, it helps to compute and use encrypted data without decrypting it. To solve privacy-preservation problems and enhance online data sharing, this paper dedicates to demonstrating how the merits of Blockchain and Fully Homomorphic Encryption (FHE) can be utilized to realize a secure, reliable, and decentralized online-questionnaire platform that guarantees the authenticity of data and privacy security for clients. This paper also proves that the combination of blockchain and FHE can successfully meet the expectation of an ideal online questionnaire platform.

Keywords: Blockchain; Fully Homomorphic Encryption; Online Questionnaires.

1. Introduction

Questionnaires are often used when researchers need to acquire a large amount of data and do data analysis. Most questionnaires contain a few choice questions, yes or no questions which respondents can fill in minutes. Investigators became aware of privacy preservation problems a long time ago so anonymous questionnaires are now more welcoming to respondents. Paper questionnaires have unavoidable flaws. For instance, handing out paper questionnaires is tiring, and filling questionnaires with pens is tiring too, which means the investigation costs too much unnecessary energy and time for both investigators and respondents. Anonymous paper questionnaires also can't ensure privacy preservation when submitting questionnaires. In recent years, online questionnaires have replaced paper questionnaires for their convenience. Researchers can easily have access to much more data and analyze it more efficiently by copying data to the software immediately [1-3].

However, online questionnaires still face privacy leaking problems. Worries about whether the data can be kept safe always exist because data analysis processes are not open to the public, so respondents can't know how investigators utilize their data. To make the situation worse, online privacy leakage frequently happened and the scale of every leakage is huge, billions of sensitive records may be leaked through one attack. Consequently, online questionnaires barely win people's trust when they are not anonymous, let alone convince them to submit sensitive data. Paper questionnaires and online questionnaires have their distinct advantages in terms of privacy preservation, but both of them can't protect privacy security very well. To address these problems, this paper will introduce two popular technologies: blockchain and fully homomorphic encryption to achieve a new type of online-questionnaire platform, which provides secure data sharing efficiently. This platform is fully decentralized and able to finish simple polynomial computation with encrypted data so that respondents' privacy and sensitive data can be well preserved.

2. Overview of Blockchain and Fully Homomorphic Encryption Technologies

2.1 Blockchain

In this section, this paper will introduce the background of blockchains such as the mechanism of blockchain, and the history of blockchain. Characteristics of blockchain will be also introduced at the end of this section.

2.1.1 Development of Blockchain

The concept of blockchain was first proposed in Bitcoin White Paper, 2008, to address problems existing in traditional bank systems and invent a peer-to-peer electronic cash system based on distributed timestamp server. Since 2008, blockchain has experienced three main stages of its development [4] (Fig.1). Now applications of blockchain are far beyond currency and financial services. Blockchain has been applied as a core technique in NFT, metaverse...

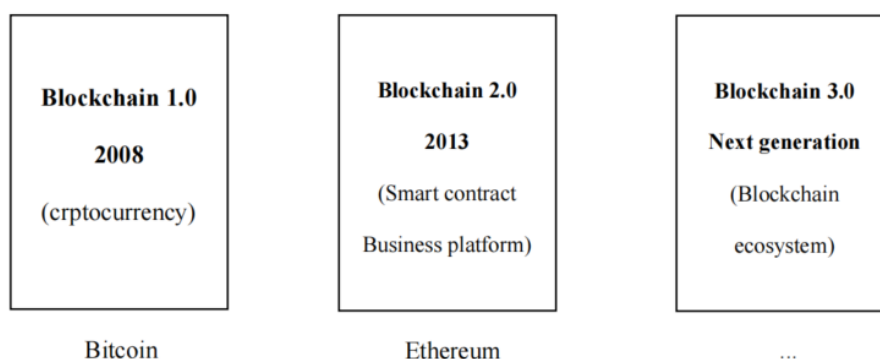


Fig 1. The development of blockchain

2.1.2 Categories of Blockchain

To adapt different applications in different situations, blockchain technologies can be divided into three main categories: public chain, private chain, and consortium chain. as the original blockchain technology, the public chain is completely decentralized and applied in cryptocurrency like bitcoin. Anyone who owns computing power can participate in this distributed blockchain system and contributes to verifying, writing, and supervising records on the chain. Public chain programs like bitcoin will give out corresponding bonuses to award those who help to maintain the justice of public chains according to how much their contribution is. the private chain is a centralized chain. One needs central authorities' permission to get involved in private chains. The authorities can decide if they want to open access for others to read data. Considering its centralization, clients of private chain technologies are mostly small organizations and governments that deal with very sensitive messages.

The consortium chain is partially decentralized and open to members of specialized groups and third parties with limited authority [5,6]. The generation of blocks is jointly decided by pre-selected nodes. Other nodes have limited access and do not function in the consensus process. Consortium chains let multiple organizations build systems to achieve common purposes.

2.1.3 Consensus Mechanism

As a decentralized system, Blockchain doesn't need a third party's supervision by adopting consensus mechanisms. There are four popular consensus mechanisms in blockchains: Proof-of-Work (PoW), Proof-of-Stake (PoS), Proof-of-Burn (PoBurn), and delegated Proof-of-Stake (DPoS). PoW and PoS are the two most widely used mechanisms, with Bitcoin and Ethereum being the two most famous applications. PoW is a mechanism that utilizes hash functions to prove the authenticity of data. Provers have to solve puzzles produced by the hash function with much computing power. Hash function puzzles are very hard to compute but easy to verify [7-9]. After solving the puzzle, the prover will broadcast it to other nodes then others can verify the data. In bitcoin, this hash function algorithm is SHA-256, which was published in 2001 by NASA and proved secure. Take the PoW

mechanism applied in bitcoin as an instance, *Prehash* denotes the hash value of the last blocks, T_i indicates a list of transactions. Provers have to compute a nonce to satisfy (1), which costs much time and computing power.

$$H(\text{Prehash} || T_1 || T_2 || \dots || \text{nonce}) = 000(k \text{ times})0 \text{ *****} \quad (1)$$

PoS mechanism uses ownership of cryptocurrency instead of lots of hardware power to prove the authenticity of data. According to PoS, owners who have more cryptocurrency are more likely to get chances to prove the credibility of data. Provers need to pay a certain amount of cryptocurrency before creating blocks. After the blocks are generated and validated, provers will gain some cryptocurrency as awards. If validation fails, provers will get penalties. PoS brings less pollution to the environment than PoW does and avoids competition for computing power.

2.1.4 Features of Blockchain

(1) Decentralization: Decentralization is the most significant feature of blockchain. Blockchain is made up of linked blocks, generation of blocks, information adding and information sharing don't need a third party's intervention, so a large amount of expense for management is saved. Since the whole system runs automatically and orderly, mistakes seldom happen unless serious bugs exist in consensus algorithms. The responsibility distribution of maintaining the systems is relatively fairer than traditional methods.

(2) Anonymity: Data on the blockchain is encrypted by asymmetric encryption algorithms. The digital signature is used to indicate identity and check whether the transaction has been identified when transmitting encrypted data. Anonymity is the premise of decentralization because it prevents the factors of real-world identities and authorities to influence the running of blockchain systems. In addition, anonymity preserves privacy to some extent. However, this feature may favor illegal activities like theft and fraud.

(3) Transparency and Traceability: As a decentralized system, records on blockchains are open to the public because blockchain needs the public's verification to run. Timestamp servers are used in blockchain so that all records on the chains are listed in time order, this enables people to trace old records. Meanwhile, data is irreversibly modified according to the principles of blockchain. Once a block and the data are confirmed, no modification can be done to it. As a result, records on blockchains are transparent, traceable, and unforgeable.

(4) Credibility: Credibility relies much on consensus mechanisms. Taking the PoW mechanism as an example, all verification on blockchains needs lots of computation. Attackers need to occupy at least 51% of all computing power to approve fraudulent information, which is almost impossible to happen because it costs too much. All mechanisms in use are already proven to be safe.

2.2 Fully Homomorphic Encryption

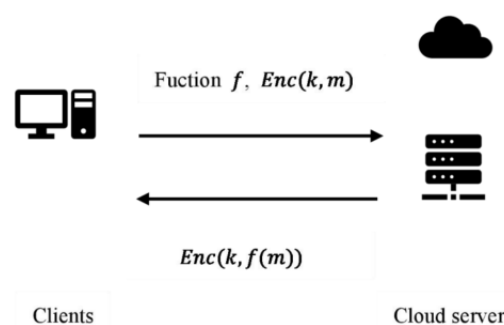


Fig 2. Ideas of FHE

Fully homomorphic encryption (FHE) is a technique proposed to operate on encrypted data without knowing the secret key and decrypting the data. An operation on encrypted data will achieve the same result as on decrypted data. With FHE, cloud servers can provide service to clients without knowing clients' data (Fig.2) and thus accomplish the purpose of privacy preservation. In this section, this paper will introduce some definitions of FHE.

2.2.1 Homomorphism

Homomorphism is a structure-preserving map between two algebraic structures of the same type.

$$\delta: A \rightarrow B \text{ where } \delta(x *_A y) = \delta(x) *_B \delta(y) \quad (2)$$

For example: $\delta(x) = e^x$, then:

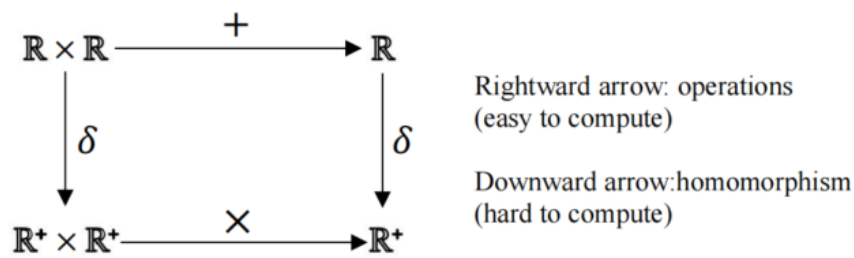


Fig 3. An example of homomorphism

In applications of homomorphic encryption, there are usually many complex operations combining together. In other words, HE is a combination of homomorphism and complex theory.

2.2.2 Procedures of Fully Homomorphism Encryption

FHE consists of four algorithms, three of which are similar to algorithms of public key encryption (PKE):

- (1) Key Generation: $Gen(\lambda, r) \rightarrow (pk, sk)$.
- (2) Encryption: $Enc(pk, m, r) \rightarrow c$.
- (3) Decryption: $Dec(sk, c) \rightarrow m$.
- (4) Evaluate: $Dec(Eva(pk, f, c_1, c_2, \dots, c_i)) \rightarrow f(m_1, m_2, \dots, m_i)$, $f \in F$ (a family of functions).

The security of FHE should be nearly the same as PKE. FHE also satisfies semantic security definition.

2.2.3 Applications of Fully Homomorphic Encryption

The first scheme was proposed by Craig Gentry in 2009 but was too slow to implement on any project. Since 2009, there have been four generations of FHE and it's getting faster about 8 times every year. FHE has been applied in neural networks [10], genome-wide association studies [11], private information retrieval (PIR) [12], and many other projects. Standardizations on FHE are now in progress to evaluate security and APIs for different schemes.

3. Findings

Online questionnaire platforms are now facing problems such as privacy preservation and data authenticity proof. To address these problems and make online-questionnaire platforms work more efficiently, this paper proposes a scheme that combines blockchain and FHE technologies to design a decentralized and reliable online-questionnaire platform.

According to the scheme, the design can be separated into 2 parts. The first part is to realize decentralization and proof of authenticity, using blockchain technology [13]. Detailed illustrations are as follows:

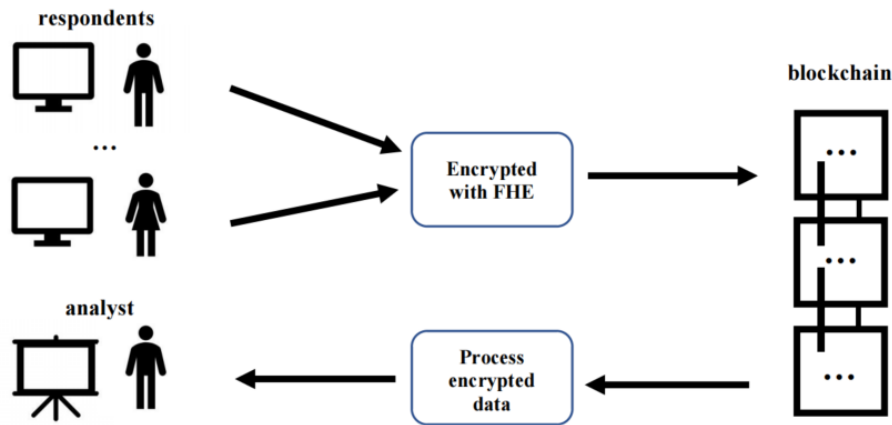


Fig 4. System model of an online-questionnaires platform

1) The platform issues exclusive cryptocurrency. Users can spend their cryptocurrency to release questionnaires, they earn cryptocurrency by answering questionnaires. The more frequently users participate in platform activities, the more cryptocurrency they will spend or earn.

2) Proof-of-Stake is used as the consensus mechanism of this platform.

3) The platform will issue or destroy cryptocurrency to dynamically adjust the total amount of the cryptocurrency according to the number of users and the number of platform questionnaires. The nature of the issued currency is close to that of Ethereum.

4) Validators of data in questionnaires will receive some cryptocurrency as awards.

5) Users who own more cryptocurrency are more likely to be validators. The detailed possibility is computed according to coin-age-based selection schemes to guarantee equality.

6) The platform uses a public chain and it is open to the public.

The second part is to achieve privacy preservation during the process of doing inquiries and allowing other users to analyze data. To do operations such as addition, subtraction, and multiplication on the encrypted sharing data, FHE is an ideal scheme other than zero-knowledge proof (ZKP) or multi-party computation (MPC). ZKP needs too much complex computation, and it doesn't fit in scenarios where more than two parties are involved. MPC can't guarantee most of the parties won't collude. Like ZPC, MPC also needs lots of computation. Partial homomorphic encryption is not considered because PHE schemes like ElGamal and RSA fail to do addition and subtraction simultaneously. This paper suggests a lattice-based FHE scheme as a solution to privacy-preservation problems [14,15]. In this scheme, noise is added to the messages to mask homomorphism. The noise can make it more secure but noise grows as the computation increases and may finally influence the accuracy of outcomes. To address the noise accumulation problems in this scheme, Craig Gentry proposed bootstrapping (re-encrypting) and squashing. It is worth noting that bootstrapping is the most time-consuming process in FHE, so optimizing bootstrapping process is the key to accelerating the whole scheme [16].

The advantages of blockchain and FHE-based online questionnaire platforms:

(1) Decentralization

By applying PoS consensus mechanism and public chain, this new type of online-questionnaire platform can run automatically without a third party's supervision. For a platform, decentralization helps save much expense for maintenance and management, consequently having less constriction on scaling up.

(2) Credibility

Addition and inquiries can be done to the data while deletion is not allowed. All data are listed in time order and traceable. According to PoS mechanism, approving fraudulent data will be fined, which is an efficient means to prevent malicious intent.

(3) Security

Encrypted data is of no use to hackers, a FHE scheme makes it feasible to do computation on encrypted data without decryption. In addition, standards such as NIST have been made to evaluate the security of FHE schemes against quantum computers. Among all FHE schemes, lattice-based FHE schemes are the most recognized. The security of ideal lattice-based FHE is close to that of PKE. In other words, lattice-based FHE schemes have been proven secure.

(4) Attractiveness

Cryptocurrency is adopted in this platform so it encourages more people to get involved to earn money. Because of its decentralization, the threshold of the platform is quite low, people without professional experiences can also participate and customize their questionnaires according to their interests.

The flaws of this platform:

(1) Profit-driven intent

Cryptocurrency also brings negative problems to this platform. For those who are eager to earn a profit, the adoption of cryptocurrency may attract dishonest people to use malicious plugins to submit numerous meaningless data and affect the accuracy of statistics.

(2) Data processing speed limitation

Although promising, now FHE is still far from perfect. The amount of data that an online-questionnaire platform needs to deal with is so huge that current FHE schemes couldn't process timely. Apart from the speed limit of FHE, verification of blocks also takes a lot of time. So, the speed limit is the most urgent problem to overcome.

4. Conclusion

In this paper, a novel online questionnaire platform scheme combining blockchain and FHE technologies are presented to guarantee good privacy preservation and data authenticity. Proof-of-Work mechanism is used to achieve decentralization at a relatively lower cost of computing power. Among many popular HE schemes; this paper applies a lattice-based FHE scheme for it is fully homomorphic and secure against quantum attacks. In the future, more work is required to speed up the data processing by reducing the time of bootstrapping in FHE schemes. Additionally, reducing the transaction fee of the cryptocurrency issued on this platform is worth more attention. Since the verification of blocks is time-consuming, accelerating the verification at the sacrifice of some security is a feasible strategy in some cases. The future research direction is to discuss how to achieve the balance between security and data processing speed.

References

- [1] Martyn Denscombe. The length of responses to open-ended questions: A comparison of online and paper questionnaires in terms of a mode effect. *Social Science Computer Review*, 2008, 26(3): 359-368.
- [2] Samúel Lefever, Michael Dal, Ásrún Matthíasdóttir. Online data collection in academic research: advantages and limitations. *British Journal of Educational Technology*, 2007, 38(4): 574-582.
- [3] Paul van Schaik, Su Luan Wong, Timothy Teo. Questionnaire layout and national culture in online psychometrics. *International journal of human-computer studies*, 2015, 73: 52-65.
- [4] Yang Lu. The blockchain: State-of-the-art and research challenges. *Journal of Industrial Information Integration*, 2019, 15: 80-90.
- [5] Kei Leo Brousmiche, Thomas Heno, Christian Poulain, Antoine Dalmieres, Elyes Ben Hamid. Digitizing, securing and sharing vehicles life-cycle over a consortium blockchain: Lessons learned. 2018 9th IFIP international conference on new technologies, mobility and security (NTMS). IEEE, 2018: 1-5.
- [6] Gu Jingjing, Sun Binglin, Du Xiaojian, Wang Jun, et al. Consortium blockchain-based malware detection in mobile devices. *IEEE Access*, 2018, 6: 12118-12128.

- [7] Lin Iuon-Chang, Liao Tzu-Chun. A survey of blockchain security issues and challenges. *Int. J. Netw. Secur.*, 2017, 19(5): 653-659.
- [8] Sercan Ozcan, Unalan Serhan. Blockchain as a general-purpose technology: Patentometric evidence of science, technologies, and actors. *IEEE Transactions on Engineering Management*, 2020.
- [9] Li Xiaoqi, Jiang Peng, ChenTing, Luo Xiapu, et al. A survey on the security of blockchain systems. *Future Generation Computer Systems*, 2020, 107: 841-853.
- [10] Bernardo Pulido-Gaytan, Andrei Tchernykh, Jorge M Cortés-Mendoza, et al. Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities. *Peer-to-Peer Networking and Applications*, 2021, 14(3): 1666-1691.
- [11] Alexander Wood, Kayvan Najarian, Delaram Kahrobaei. Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Computing Surveys (CSUR)*, 2020, 53(4): 1-35.
- [12] Song Weitao, Hu Bin, Zhao XiuFeng. One Improved RLWE-based FHE and Fast Private Information Retrieval. *KSII Transactions on Internet and Information Systems (TIIS)*, 2019, 13(12): 6260-6276.
- [13] Shen Meng, Tang Xiangyun, Zhu Liehuang, Du Xiaojiang, et al. Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities. *IEEE Internet of Things Journal*, 2019, 6(5): 7702-7712.
- [14] Craig Gentry. A fully homomorphic encryption scheme. Stanford university, 2009.
- [15] Pin-Chun Chen, Kuo Tzu-Hsiang, Wu Ja-Ling. A Study of the Applicability of Ideal Lattice-Based Fully Homomorphic Encryption Scheme to Ethereum Blockchain. *IEEE Systems Journal*, 2021, 15(2): 1528-1539.
- [16] Jung Hee Cheon, Hyeongmin Choe, Donghwan Lee, Yongha Son. Faster Linear Transformations in HELib, Revisited. *IEEE Access*, 2019, 7: 50595-50604.