International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

User's Truthfulness Identification in Data Market

Ms. Swati Phalke and Dr. Chaitanya Kulkarni

Department of Computer Engineering VPKBIET, Baramati Pune, Maharashtra, India.

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Many of the online information systems have arisen as an important business model to meet the needs of society for a person-specific data, whereas the service provider receives raw data from the data contributors and then delivers value-added data services to the data users i.e. data consumer. Nonetheless, data users are confronted with a pressing problem in the data exchanging framework, that is, how to check that the service provider has obtained and stored data in a fair manner? In fact, data users are usually unable to reveal to data buyers their private personal data and actual identities. For big security problem which obtained for privacy use blockchain technology which is distributed in networks. Blockcnain-based verifying method can potentially help people to find out whether the data they obtain is precisely exactly what they expect. The suggested method that identifies participants is Truthfulness. In this system consumer buy product that he / she should give the system analysis to verify first whether or not the contributors are an approved individual.

Keywords: *Data truthfulness, Privacy Preservation, Homomorphic Encryption, Identity Based Signature, Service Provider, Data Markets, Blockchain Technology.*

Introduction

Throughout the big data of society, has a voracious demand for revealing personal developed data in society. Taking into account the possible economic value of a personal data in the decision making and enhancing experience for users, several open and free platforms have surfaced to allow the interchange of personal digital information [3]. For example if any system which provide online service to user's they are creating privacy that works for all. It has a responsibility that comes with the development of affordable and accessible goods and the services for anyone. Instead of those services they can use to direct their goods, procedures, and employees in maintaining the information of their user's private, safe and comfortable data. There is, a major security issue in such market-based networks, i.e., in terms of information collection and processing, it is difficult to ensure the truthfulness, especially when data contributors' privacy needs to be maintained [9]. At first data contributors have to submit their data in a system but in case, when original data sets are combined with more than a small amount of fake or synthetic specimens to draw these metrics, the final election outcome will be badly affected.

Next service provider have been add some or more bogus data in the original data for his personal profit only that can be makes much more effect on the system. It will cause much more loss cause to the consumer. The guaranteeing veracity and preserving data contributors ' privacy are both critical for the deep-term growth and development of data markets. There are four notable difficulties to integrate integrity and security defense in a common sense information showcase.

The first task of the strategy is that ensuring the integrity of the collection of knowledge and preserving the protection seems to be opposing objectives. The second task is check is data analysis that makes it even more difficult to verify the truthfulness of data collection. The third task is to test the truthfulness and data protection of data acquisition guaranty of asymmetric information around data consumer and service provider. The last task is to gather raw information from a larger amount of low throughput data contributors [8]. That is, the integrity of the gathering of knowledge in our model. Only authorized staff obtain this personal data. However, acknowledgement in computer signature plans involves the training of crude personal data, and can launch the complex character of a supporter of information without too much of a stretch. The enterprise's motivation is that TPDM is really the main secure feature in online set of data for the Truthfulness and Privacy Preservation. TPDM is the first main secure mechanism in data markets. TPDM imposes honest collection and transmission of real data by the service provider. TPDM instantiate with the two ways

of real-world data facilities, which are profile matching and data distribution. These can be implement with concrete data markets, and extensively evaluate with their performances in given data sets.

In order to protect the privacy of users from malicious data shared by other people and to lease database energy, a Blockchain based verification approach is used in TPDM for securing system. By using the Blockchain to record the hash value and other required data sharing information from other people, it ensure that data users accessed from third party source is indeed the original data uploaded [11].

Literature Survey

Matthew Franklin, Dan Boneh, presented a paper on a fully functional encoding scheme (IBE) based on identity. The
system is based on two group bilinear maps, in which encrypted mechanism is implemented in this paper based on
identity. The system's main limitation is that private key revocation is not possible [1].

Privacy Preserving: - Wang, Wang, Wang, Ren, K., and W. Lou those individual provides the cloud data storage public audit service to ensure that consumers can use batch auditing to audit the outsourced data to an independent third-party auditor (TPA) [3].

Data Markets In The Cloud: -Mag dalena Balazinska, Bill Howe, and Dan Suciu addressed it explains some of the problems that businesses are facing and also discussed the related research issues that our group can help to solve. They also addressed the implications for the database research community of the emerging cloud-based information markets
[4].

Jan Camenisch, Susan Hohenberger, Michael stergaard Pedersen, put forward a paper on the first email batch verifier. They also proposed a new signature scheme with very short
signatures, which also makes batch authentication extremely effective for many applications [5].

CryptoNets: - Dowlin had been working on the technique i.e. CryptoNets neural networks, this neural network that is linked to the user scrambled data in the encrypted format, because of this data owner gets trust to send over the cloud their private encrypted information. By using GPUs and FPGAs to speed up the calculation, the efficiency and idleness can be substantially improved. Another path for further development is to discover increasingly successful encoding plans that take littler parameters into account, and therefore faster homomorphic calculation [6].

AccountTrade: -T. Jung, X.-Y. Li, W. Huang, J. Qian, L. Chen,J. Han, J. Hou, and C. Su say that, Accountable Protocols for Big Data Trade Against Fraudulent Consumers uses a system which provides transparency for fraudulent vendors who can re-sell data sets from others. This is a modern quantitative calculation of the quality of the dataset that can be computable efficiently [9].

Zhenzhe Zheng proposes VENUS, which for community detected data markets is the key gain driVEN knowledge acqUiSition method. VENUS provides two related programs
in particular: VENUS-PRO revenue-driven amplification and VENUS-PAY for deployment minimization. VENUSPAY
sells out of the agreed second-value as far as increments are concerned [10].

Trust in the durability of blockchain technology is that in a multitude of blockchain based applications and experiments. Scalability and consensus algorithms are areas of growing research to make blockchain more adaptable for larger-scale businesses [12].

System Model

The fig.1 shows the two layers- **1. Data Acquisition layer.**

2. Data Exchange layer.

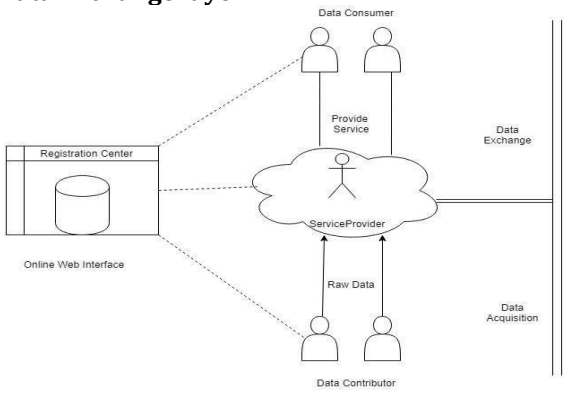

Fig 1: - System Model

In the data acquisition layer service provider sends a request for access a raw data with its authority, both the service provider and data contributor have to be a member of the data market system. In which one can store or upload his/her data for the further process i.e. data contributor, and second for the accessing data from data contributor with valid request to them, which is nothing but a dataset of personal identifications of users in the social area of the system's users. Service provider checks the authenticity of each record in the given dataset from data contributor.

In a data exchange layer the service provider is provide a dataset which is he/she taken by data contributor in a system to the data consumer with its valid request. A service provider may change dataset for his/her own profit in system for that purpose the data consumer check truthfulness and authentication of each data which is provided by service provider.

Proposed Methodology

In the proposed system with the help of the Blockchain Technology we've get idea to plan secure mechanism

i.e. TPDM for data markets in an economic system as well as for protecting privacy of each user and achieving truthfulness of data.

1. What is Blockchain?

Blockchain: - It is referred to as Distributed Ledger Technology (DLT), which makes the history of any digital benefit permanent and transparent through the usage of decentralization and cryptographic hashing. It is an especially gifted and innovative technology because it help us to reduce risk, stamps out fraud and takes transparency in a scalable way for numerous uses.

Truthfulness and Privacy Preservation (TPDM) is the first secure mechanism which gives the guarantee of honesty and security of data in proposed system. In this system data contributor have to submit their data in truthful way and which cannot accessed by others i.e. third party. The service provider have to collect authenticated data and to provide without adding bogus data for their profit only.

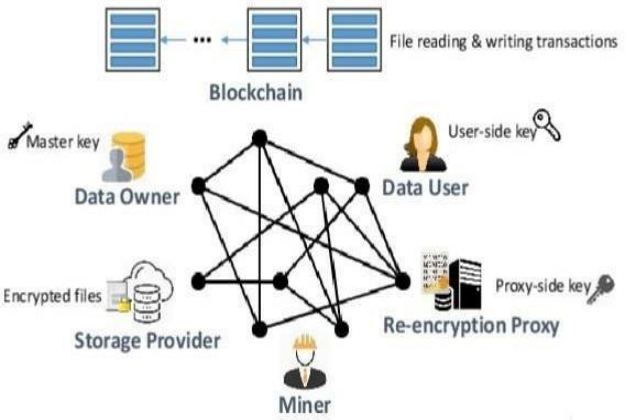

Fig 2: - Blockchain Architecture.

The above fig. 2 shows the architecture of the blockchain Technology. What is more, each the in person identifiable info and also the sensitive information of information contributors are well protected. Additionally for better security we use Blockchain Technology which is based on decentralized, peer- to-peer transactions. It offers a better protection and eliminates the need for other single controlling agency that maintains data base management powers.

Table 1: - Types of Blockchain

Based on Blockchain access	Based on Blockchain data access
Unauthorized: -Anyone with processing power may enter.	**Public: -**Anyone who accesses can change
Authorized: - Approved users.	**Private: -**Authorized users can change/modify.

In principle, four types of a blockchains can be identified as shown in Table 1 based on who can access the blockchain network and how the authorization to write to the blockchain network are allocated. It is noted, that the words public and unauthorized are used interchangeably, and so are the private and authorized.

2. Blockchain Protocols

A foolproof, purpose-appropriate consensus mechanism is essential for maintaining data dignity and uniformity among the network's participating nodes. Although a lot of work has been done on blockchain protocols, some main algorithms are described in brief here whose variants are being used and further adapted to suit different blockchain applications. Table 2 provides a quick comparison of the blockchain protocols.

Table 2: - Comparison of Blockchain Protocols

Protocols	Advantages	Disadvantages
Proof of Work	-Considered very stable -Miners receive rewards	- Quite slow at the moment. - Consumes lot of electricity.
Proof of Stake	-Minus chance of hardware centralization. -Possibly quicker than Proof-of- work protocol.	-Economic penalties for fraudulent attempts.
Practical Byzantine Fault Tolerance	-It tolerate 1/3rd of the nodes to be faulty or Adversarial. - Fast and efficient	-Revelries requisite approve to the precise contribution of groups. -Originates at the rate of obscurity

Architecture and Algorithms

1. Architecture Of TPDM

System architecture of our proposed method gives a valuable security analyze in the data authentication and integrity, the way of truthfully collection and processing of data, also both the confidentiality of data and way of preserving a identity with higher security.

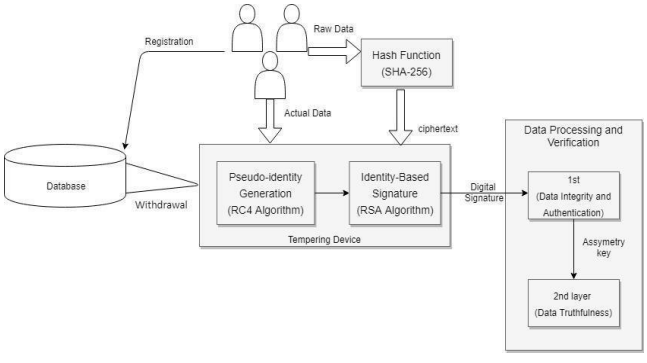

Fig 3:- Architecture of TPDM

The profile matching which can go through only the correctness and completeness of each user. If the at the time of authentication and processing of digital signature the record of user which is provide by service provider to consumer that will be match too existing one else consumer declare it as a fake user. The system architecture shows in a following figure. 3. The TPDM which integrates data truthfulness and privacy preservation in data markets.

Following are the steps in TPDM

1.1 Initialization:-

Registration center having multiplicative groups. In that registration centers have two keys i.e. secret key and public key. The registration center sets up parameters for a partially homomorphic cryptosystem: a secret key SK, a public key PK, an encryption scheme E (·), and a decryption scheme D (·).

1.2 Signing Key Generation:-

To achieve unidentified authentication in data markets, the tampering device is utilized to generate a pair of pseudo identity and secret key for each registered data contributor i.e. Identity-Based Signature.

1.3 Data Submission: -

We need to consider many criteria for safe submission of raw data, including confidentiality, authentication and honesty. We use partly homomorphic encryption to ensure data confidentiality. In addition, the encrypted raw data must be signed prior submission to ensure data protection and data integrity, and checked after receipt.

A. Pseudo-Identity Generation.

To generate of Pseudo-Identity the encryption can be done with symmetry key i.e. RC4. It having a capability of both encryption and decryption. Encrypted key used for data processing in system, the valid users only allow to access data in system.

B. Hash Function

Hash function which is here we used SHA-256 generates an almost-unique 256-bit signature for a text. A hash is not encoding, it cannot be decrypted back to the original text. It is a 'one-way' cryptographic function, and is a fixed size for any size of source text.

C. Identity-Based Signature

With help of RSA algorithm we get signature which is digital signature. In the system it makes its own hash which comparable to the existing hash of particular user. If those are equally match to each other then and then only further processing is done.

1.4 Data Processing and Verification

A. Data Integrity and Authentication

The data users are also able to verify the accuracy of the processing of data. The data users should know his / her text under the various circumstances.

B. Data Truthfulness.

The data consumer can validate the authority of applicant data sources comparable to the first-layer batch verification. The homomorphic properties also allow data consumers to verify the truthfulness of the processing of data. Provided that the data user knows the plaintext of each user in system, all the cross terms involved can be evaluated by means of a constant user multiplication.

1.5 Revocation and Tracing.

The two-layer batch verifications it will hold if and if only when all the signatures are valid, and when there is a single invalid signature then fail. In a signature batch may hold invalid one(s) which produced by unintentional data exploitation or possibly nasty activities launched by an attacker. If there is a single invalid signature, it reject the entire batch. Tracing and recalling illegal data items and their corresponding signatures are important in practice. If the verification of the second layer batch fails, the data user may allow the service provider to figure out which signature(s) are invalid. Likewise, if the verification of the first-layer batch fails the service provider will figure out the invalid one(s) by itself.

2. Algorithms

- Obtains the public key (n, e).
- Represents the plaintext message as a positive integer m with 1<m<n
- Computes the ciphertext c=m^e mod n. ➢ Sends the ciphertext c.

2.1 Encryption

Encryption is done with the help of RC4 algorithm. RC4 is a symmetric key algorithm for the stream cipher. For both encryption and decryption, the same algorithm is used as the data stream is simply X-OR with the key sequence generated. The main flow depends entirely on the plaintext used. It is used to produce pseudo-random bits subsequently and then generate a pseudo-random stream that is X-OR with the plain-text to give the cipher text.

RC4 is a key generation algorithm having a following steps –

1. Initialization (S, T, k, i, len).
2. Pseudo random generation (Stream Generation).
3. Encryption with X-OR.

2.2 Hash Key Generation:-

Hash processes transform arbitrary huge bit strings called messages into small, fixed-length bit strings called message digests, such that digests identify very high probability messages that generated them. *Input: -* Message (M)

Output: - New hash (Digital signature)

- Message is processed in 512 blocks sequentially.
- Message digest is 256 bits instead of 160 bits of SHA1's.
- 64 rounds instead of 80 rounds of compression.

• Algorithm structure as follows:- 1. Padding bits.
2.Appending length as 64 bit unsigned.
3.Buffer initiation.
4.Processing of message.
5.Output (Digital Signature).

2.3 Generation Of Digital Signature:- The RSA crypto system is the world's most popular used algorithm for public key cryptography. This can be used to encrypt a message and no need for separate transaction of a secret key. With both public key encryption and digital signatures, the RSA algorithm may be used. The reliability is based on the complexity of wide arithmetic factoring.

Key Generation

- Generate two prime numbers, p and q
- Let n=pq
- Let m=ø (n) = (p-1) (q-1)
- Choose a small number e, co-prime to m, with GCD

(ø (n), e)-1; 1<e<ø (n)

- Find d, such that de mod ø (n) =1
- Publish e and n as public key.
- Keep d and m as a secret key

Generate digital signature.

Result and Discussion

The primary focus of data preservation is to improve conventional data mining techniques that mask private information through data modification. The major issues is a way of changing information and discovering the result of information mining from the changed data. The first method is the data Disturbing principles to protect confidentiality. The method is to use cryptographic methods to construct models of data mining. Preserving privacy is chosen to use it when the intruder are unable to learn something extra from of the data, even if his context information is derived from other sources. The following fig. shows that the actual processing and verification in the TPDM with blockchain technology. In the following figure shows that the Data Processing and verification in the Blockchain Technology. The fig 4 shows the main file claim in which hash will be calculate for each user in system and stored only necessary information in the blockchain, and upload it into the cloud.

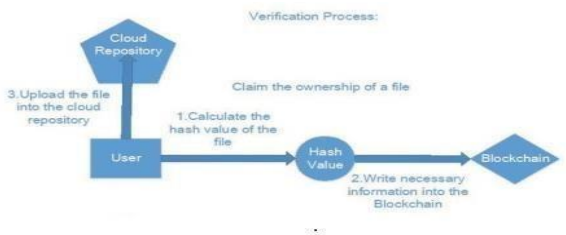

Fig 4: - The Ownership file Claim

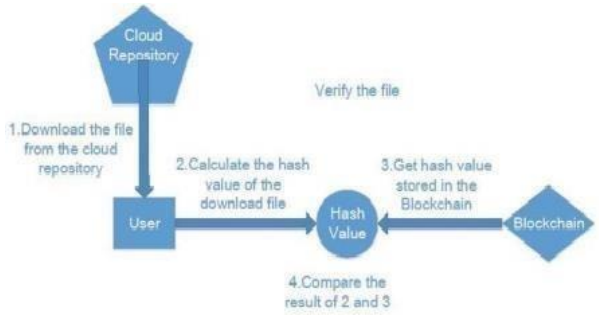

Fig 5: Verify The File

In fig 5 shows that the after claiming the file data consumer download the file from database and calculate its hash value. After that calculate the hash value of stored data in blockchain and at last compare it with first hash value. It will gives us a completeness and correctness of each users in the system.

Advantages

1. The system is highly secure mechanism for the data markets.
2. It is organized locally and instead signs manipulation authentication and identity-based signature in part.
3. It is the first mechanism in data trading for both encryption and truthfulness.
4. It verifies honesties of privacy preservation in knowledge market.

Applications

1. **Data market is the best application of TPDM**.
2. **All government applications**: - It will central to the end of several long standing businesses and professions.
3. **Private Agencies: -** With increased number of mobile applications seeking complete access to user data such as contacts, messages, photos and a variety of other personal data. .
4. **Banking System:** -In the baking time consuming process it provides necessary transparency and speed via smart contracts with requirements.
5. **Education:** - It will provide a verifiable, easily shareable and permanent record of such educational records and rewards.

Conclusion

In this proposed system to develop Truthfulness and Privacy Preservation (TPDM) for authenticity and truthfulness of data, which simultaneously guarantees truthfulness of data and protection of privacy. The data contributors should truthfully send their own information in TPDM, but they cannot impersonate others. The service provider is also required to collect and process data in a fair manner. Here set up of TPDM with two different data providers and tested their

performance with blockchain technology extensively on the dataset. Author should mention limitations of the proposed system. The author is also asked to mention the future research line.

Reference

[1]. Dan Boneh, Matthew Franklin, Fellow, "Identity-based encryption from the well pairing," in CRYPTO, 2001.

[2]. Allam Mousa and Ahmad Hamad,"Evaluation of the RC4 Algorithm for Data

[3]. Encryption", International Journal of Computer Science Applications Vol.3, No 2, June 2006.

[4]. C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," Proc. IEEE INFOCOM, 2010, pp. 1–9.

[5]. M. Balazinska, B. Howe, and D. Suciu, Senior Members, IEEE "Data markets in the cloud: An opportunity for the database community," Vol. 4, no. 12, pp. 1482–1485, 2011

[6]. J. Camenisch, S. Hohenberger, and M. Ø. Pedersen, "Batch verification of short signatures," Journal of Cryptology, vol. 25, no. 4, pp. 723–747, 2012.

[7]. R. Gilad-Bachrach, N Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy," in ICML, 2016.

[8]. Anna L. Buczak, and Erhan Guven,"A Survey of Data Mining and Machine

[9]. Learning Methods for Cyber Security Intrusion Detection", IEEE communications surveys tutorials, vol. 18, NO. 2, second quarter 2016.

[10]. ChaoyueNiu, Student Member, IEEE, Zhenzhe Zheng, "Achieving Data Truthfulness and Privacy Preservation in Data Markets," IEEE transactions on kno wledge and data engineering, vol. xx, no. xx, xxxx 2017.

[11]. T. Jung, X.-Y. Li, W. Huang, J. Qian, L. Chen,J. Han, J. Hou, and C. Su,

[12]. "AccountTrade: accountable protocols for big data trading against dishonest consumers," in INFOCOM, 2017.

[13]. Z. Zheng, Y. Peng, F.Wu, S. Tang, and G. Chen, "Trading data in the crowd: Profit driven data acquisition for mobile crowd sensing," IEEE Journalon Selected Areas in Co mmunications, vol. 35, no.2, pp. 486–501, 2017.

[14]. Yu Liu, Haopeng Chen, Fei Hu, "A Blockchain-based Verification for Sharing Data Securely", 978-1-5386-1978-0/17/$31.00©2017IEEE.

[15]. Ask,:https://en.wikipedia.org/wiki/Privacy preservation in Blockchain.

[16]. Ask,:https://en.wikipedia.org/wiki/truthfulness.

[17]. Quora, http://www.quora.com.

[18]. Acedemia, https://www.academia.edu/