iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3131

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Raw Data Truthfulness and Privacy Preservation in a Practical Data Market

M. Geetha[1], V. Harini[2], S. Sherin Benita[3], A. Supriya[4]
[1]*Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India*
[2, 3, 4]*Student, Department of CSE, Panimalar Institute of Technology, Chennai, India*

Abstract: ***In the era of block chain, ensuring data truthfulness and protecting the privacies of data contributors are both important to the long term healthy development of data markets. On one hand, the ultimate goal of the service provider in a data market is to maximize their profit. Yet, to reduce operation cost, a strategic service provider may provide data services based on the whole raw data set, or even return a fake result without processing the data from designated data sources. The content of raw data should not be disclosed to data consumers to guarantee data confidentiality, even if the real identities of the data contributors are hidden. TPDM is structured internally in associate degree Encrypt-then-Sign fashion, using partially homomorphic encryption and identity-based signature. It at the same time facilitates batch verification, data processing, and outcome verification, while maintaining identity preservation and data confidentiality. We additionally instantiate TPDM with a profile matching service and an information distribution service, and extensively evaluate their performances***
Keywords: ***Data markets, data truthfulness, data confidentiality.***

I. INTRODUCTION

The collection of digital data by governments, corporations, and individuals has created tremendous opportunities for knowledge-based decision making. Driven by mutual edges, or by regulations that require certain data to be published, there is a demand for the exchange and publication of data among various parties[1].Data publishing is equally ubiquitous in other domains.

A task is to develop ways and tools for business knowledge during a lot of hostile surrounding so the revealed knowledge remains much helpful whereas individual privacy is preserved. This undertaking is called privacy-preserving data publishing (PPDP). In the past few years, research communities have responded to this challenge and proposed many approaches [1]. While the analysis field continues to be quickly developing, it is a good time to discuss the assumptions and desirable properties for PPDP, clarify the differences and requirements that distinguish PPDP from different connected issues, and systematically summarize and evaluate different approaches to PPDP. This survey aims to achieve these goals[1]. Neither the service nor other users of the service should be able to learn the exact context information (e.g., duration, type of service request) of a user, unless the user decides to divulge such information. Users' context information should be protected against both outsiders and service providers they interact with. In this paper, we propose a user privacy preserving authentication and access control scheme at the application level to address the security and user privacy concerns in Pervasive Computing Environments (PCEs). In this paper, we assume both buyers and sellers interact with the broker via secure communication channels. The communication is encrypted and decrypted with pre-distributed keys to ensure that the dataset is not receptive to the general public. This also implies authentication is in place since the broker needs to use the correct entity's key for communication. We present Account Trade, a set of accountable protocols for big data trading via data brokers. It enables data brokers to achieve trading-related accountability against dishonest consumers by blaming them when misbehaviour is detected[3].

II. LITERATURE SURVEY

A Novel Privacy Preserving Authentication and Access Control Scheme for Pervasive Computing Environments [1] uses a scheme that integrates two cryptographic primitives, blind signature and hash chain, into a highly flexible and lightweight authentication and session key establishment protocol. But the number of desirable security properties needed by this system are numerous and hence becomes complex.

1) *Account Trade:* Accountable Protocols for Big Data Trading Against Dishonest Consumers[3]uses a scheme that achieves the accountability against dishonest sellers who may re-sell others' datasets. This is a novel rigorous measurement of the dataset uniqueness which can be efficiently computable. But there are always negligible chances for the attacker to break cryptographic tools, and the leveraged predictive models can hardly be perfect regarding the precision and recall.

2) *Privacy-Preserving Data Publishing:* A Survey of Recent Developments[6] shows a system involves, removing sensitive information or removing potential linking information that can associate an individual person to the sensitive information in a document. But the typical concerns include the degradation of data/service quality, loss of valuable information, increased costs, and increased complexity.

In the paper [7], the system solves cases where users have several prioritized location preferences on real mobile devices. But the system is not supported by the majority of current hardware and software platforms.

In the paper [8] , fine-grained personal profiles enable finer differentiation among the users having different levels of interest in the same attribute. It supports a spread of private-matching metrics at completely different privacy levels.But the protocols incur high communication costs which increase with impact on dimensions of a profile.

In the paper [9], the Public auditing service for cloud data storage ensures that users can resort to an independent third party auditor (TPA) to audit the outsourced data when needed by using batch auditing. But Batch auditing for multiple owners is tedious due to variation in their parameters.

III. EXISTING SYSTEM

In the existing system , the client will not be able to identify if the data service providers have manipulated the data from the contributor for their financial greed. Some providers will submit the data to consumers without processing it to reach their target. Privacy of the data transmitted from the contributor to the client is not ensured.

A. Blind Signature Algorithm

One of the key features of the signing algorithm discussed is that User knows exactly what they are signing[1]—they sign it because they are the author of the message. But the thought of authentication applies to a lot of things that merely proving one is that the author of a specific message.

1) *Disadvantage:* One problem with this scheme as written is that while the customer can figure out that user tried to defraud them, they can also forge transcripts to frame them.

Thus, the customer cannot prove to anyone else that user was guilty[1]. But assuming the user has a digital signature scheme of their own, it is easy to modify the above protocol to protect them.

B. Min Hash Algorithm

The MinHash signature[4] remains the same with high probability over the random choices made in the hash functions. Thus, with such signatures, personally relevant information is protected. While cumulative traffic patterns can still be effectively extracted. Our most vital contribution is that the MinHash hierarchy on the checkpoints, with which one can efficiently answer popular path queries. The MinHash hierarchy built in levels takes randomly sampled checkpoints in a recursive manner, where a checkpoint on level i appears in level i+1 with fixed probability. At every level, the checkpoints record whether they are 'neighbors' on this level – defined as whether there is a popular path connecting them without other checkpoints of the same level in between.

1) *Disadvantage*

a) Hash collisions are practically unavoidable. When hashing a random subset of a large set of possible keys.

b) Hash tables become quite inefficient when there are many collisions.

c) Hash table does not allow null values, like hash map

C. Interleaved Policy Evaluation

Interleaved comparison strategies[3]which compare rankers using click data, are a promising alternative to traditional information retrieval evaluation methods that require expensive explicit judgments. A major limitation of these methods is that they assume access to live data, meaning that new data must be collected for every pair of rankers compared[3].

1) *Disadvantage:* It cannot support policies that require other semantics, such as creating a log entry when a violation occurs**.**

IV. PROPOSED SYSTEM

In the Proposed System, TPDM consists of 5 phases: initialization, signing key generation, data submission, data processing and verifications, and tracing and revocation. In this system, user's raw data is encrypted using homomorphic algorithm.

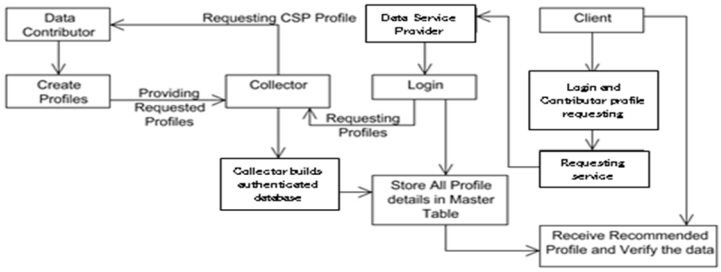
In the Modification, we include Block chain technology for effective data storage and analysis part. We collect the data of query from all the public and all the data are encrypted and stored and finally made block chain using ethereum tool.

First, we tendto think about that a malicious knowledge contributor or associate external aggressor might impersonate alternative legitimate knowledge contributors to submit probably fake data. Besides, some malicious attackers might deliberately modify data throughout submission. Hence, the service provider needs to confirm that raw data are indeed sent unaltered by registered data contributors, i.e., to guarantee data authentication and data integrity in the data acquisition layer. Second, the service provider in the data market might be greedy, and attempts to maximize her profit by launching the following two types of attacks:

Partial information collection: To bring down the expenditure on knowledge acquisition, the service provider may insert bogus data into the raw data set.

No/Partial information processing: To cut down the operation price, the service provider may try to return a fake result without processing the data from designated sources, or to provide data services based on a subset of the whole raw data set.

V. ARCHITECTURE DIAGRAM



The signature scheme is applied to the plaintext space,the data consumer needs to know the content of raw data for verification. However, if we employ a conventional public key encryption scheme to construct the ciphertext space, the service provider has to decrypt and then process the data. Even worse, such a construction is vulnerable to the no/partial data processing attack, because the data consumer, only knowing the ciphertexts, fails to verify the correctness and completeness of the data service. Thus, the greedy service supplier might cut back operation price, by returning a false result or manipulating the inputs of information process. Therefore, we turn to the partially homomorphic cryptosystem for encryption, whose properties facilitate both data processing and outcome verification on the ciphertexts.

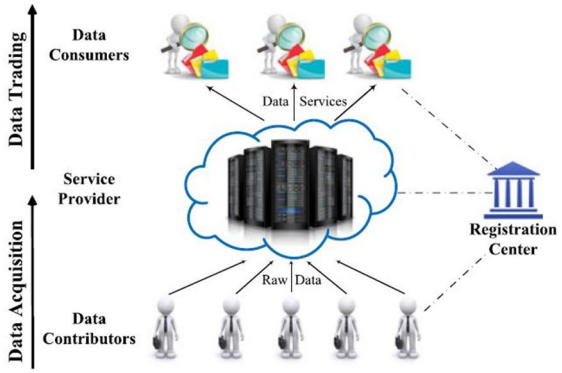
VI. MODULE DESCRIPTION

A. User Interface

In this module, Intermediate user will be updating all the data regarding the information aboutthe different merchants and all there products to be verified and data analysis process has to beperformed by the public users. The inputs are given by the general merchants to verify the public utility of the different users.

B. Server

The centralized server is deployed to monitor all the activities of the general public to monitor the reaction of the public towards consumption of the particular product. Main server is the important to monitor all the activities of the public which is stored in the server. All the questionnaire is processed in the main server for any data analysis based services.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

C. Review Questionnaire And Analysis

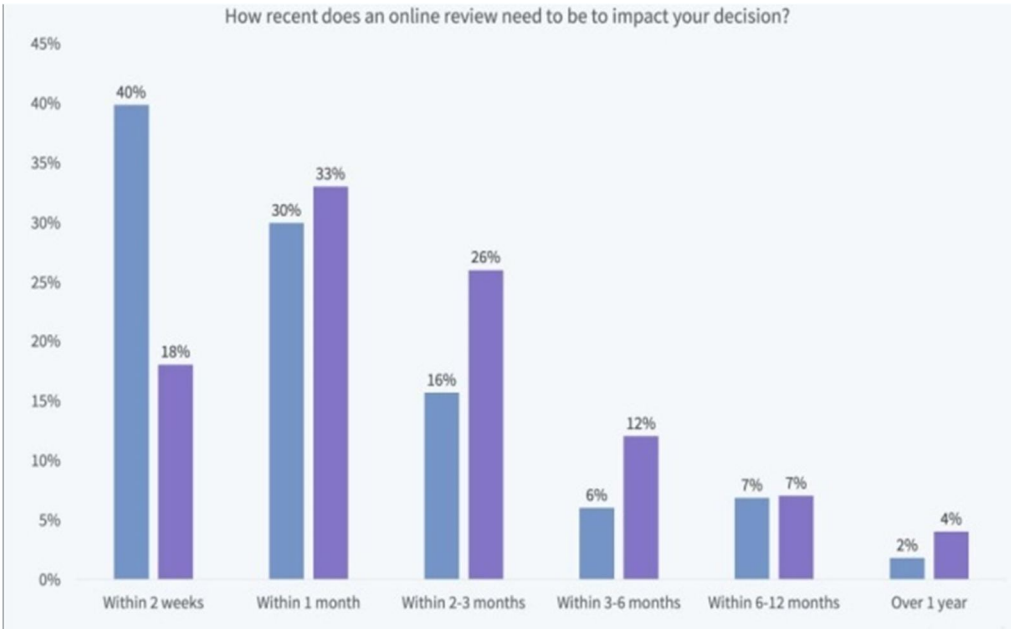
In this module questionnaire is proposed to all the consumers so that the final output is extractedfrom the questionnaire to find out the best product as per the common public opinion. Overalldata science and data analysis is applied in this module to find out the final result from the userpoint.

D. Block Chain Formation

In this module, Block chain is used to the secured data analysis so that the data stored in the mainserver for any data analysis is placed securely. Block chain is formed through Ethereum Software and based cryptographic algorithm.

VII. RESULTS AND DISCUSSION

The main focus ofprivacy preserving data publishing was to enhance traditional data mining techniques which mask the sensitive information by modifying the data. The major problems were a way to modify information andthe way to discover the information mining result from the changed data. The data Perturbing values for preservation of customer privacy is the first approach. The other approach is Cryptographic tools to build data mining models. Privacy preserving is preferred to be go out when the attacker is unable to know anything extra from the given data even though with the presence of his background knowledge obtained from other sources.



VIII. CONCLUSION

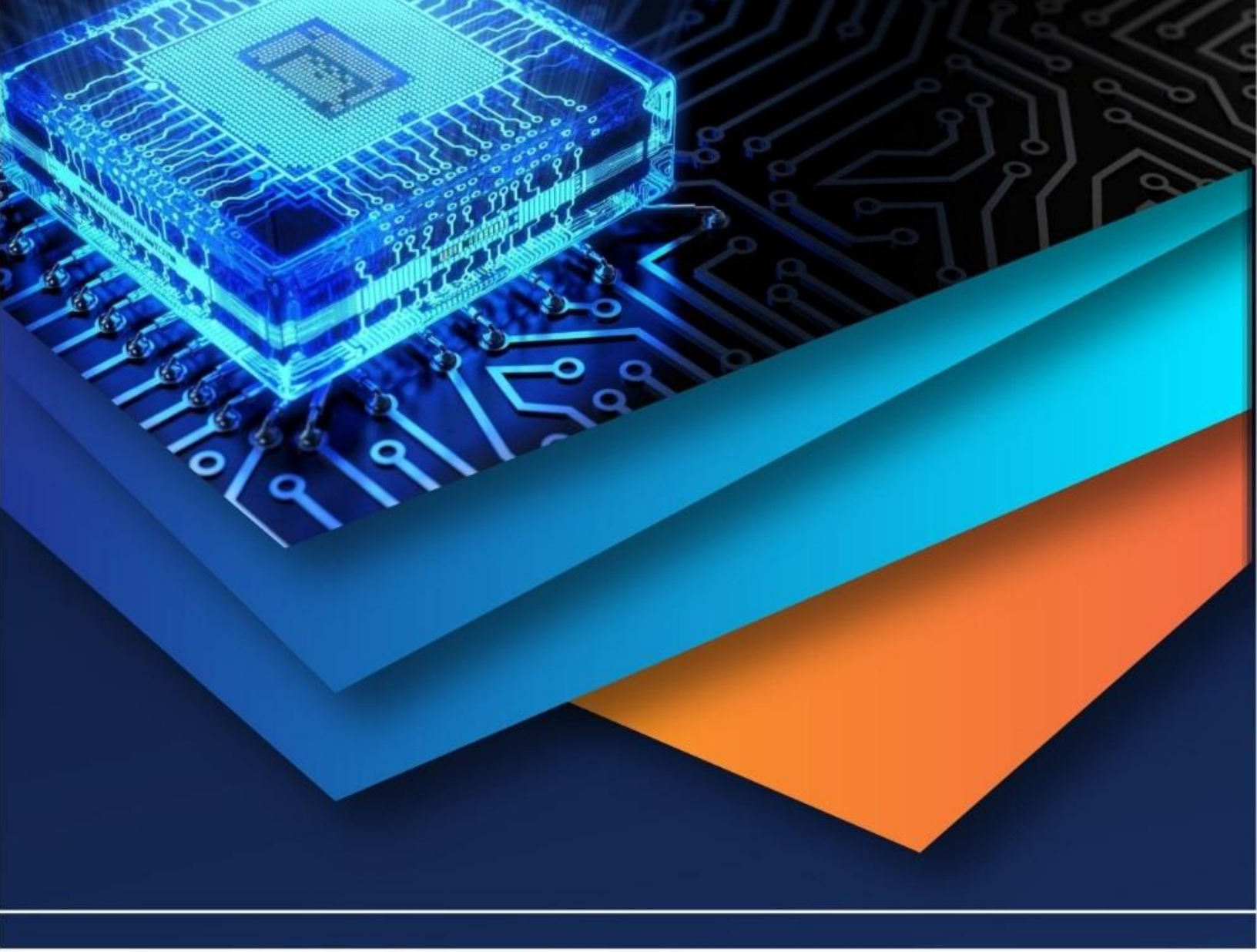
In this paper, we have proposed the first efficient secure scheme TPDM for data markets, which simultaneously guarantees data truthfulness and privacy preservation. In TPDM, the data contributors have to truthfully submit theirown data, but cannot impersonate others. Besides, the service provider is enforced to truthfully collect and process data. Furthermore, both the personally identifiable information and the sensitive raw data of data contributors are well protected. In addition, we have instantiated TPDM with two different data services, and extensively evaluated their performances on two real-world datasets. Evaluation results have demonstrated the scalability of TPDM in the context
of large user base, particularly from computation and communication overheads. At last, we have shown the feasibility of introducing the semi-honest registration centre with detailed theoretical analysis and substantial evaluations.

IX. FUTURE ENHANCEMENTS

As for further work in data markets, it would be interesting to consider diverse data services with more complex mathematic formulas. Under a specific data service, it is well-motivated to uncover some novel security problems, such as privacy preservation and verifiability.

REFERENCES

[1] K. Ren, W. Lou, K. Kim, and R. Deng, "A novel privacy preserving authenticationand access control scheme for pervasive computing environments, "IEEE Trans. Veh. Technol., vol. 55, no. 4, pp. 1373–1384, Jul. 2006.
[2] M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunityfor the database community,"Proc. VLDB Endowment, vol. 4, no. 12, pp. 1482–1485, 2011
[3] P. Upadhyaya, M. Balazinska, and D. Suciu, "Automatic enforcement of data usepolicies with datalawyer," in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2015, pp. 213–225.
[4] T. Jung, X.-Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su,"AccountTrade: Accountable protocols for big data trading against dishonestconsumers,"Proc. IEEE Conf. Comput. Commun., 2017, pp. 1–9.
[5] G. Ghinita, P. Kalnis, and Y. Tao, "Anonymous publication of sensitive transactionaldata,"IEEE Trans. Knowl. Data Eng., vol. 23, no. 2, pp. 161–174, Feb. 2011.
[6] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving datapublishing: A survey of recent developments,"ACM Comput. Surveys, vol. 42, no. 4, pp. 1–53, Jun. 2010.
[7] I. Bilogrevic, M. Jadliwala, V. Joneja, K. Kalkan, J. P. Hubaux, and I. Aad,"Privacy-preserving optimal meeting location determination on mobile devices,"IEEE Trans. Inf. Forensics Security, vol. 9, no. 7, pp. 1141–1156, Jul. 2014.
[8] R. Zhang, Y. Zhang, J. Sun, and G. Yan, "Fine-grained private matching forproximity-based mobile social networking,"Proc.IEEE INFOCOM, 2012, pp. 1969–1977.
[9] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing fordata storage security in cloud computing,"Proc. IEEE INFOCOM, 2010, pp. 1–9.
[10] 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 J. Sel. Areas Commun., vol. 35, no. 2, pp. 486–501, Feb. 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY