ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

Maintaining Identity Preservation and Data Confidentiality in Data Markets using Encrypt-Then-Sign Fashion

Rajnandini Kumawat [1], Prof. Harish Barapatre [2] , Prof. Ankit Sanghvi [3]

Dept. of Computer Engineering, Alamuri Ratnamala Institute of Engineering and Technology, Maharashtra, India [1,3]

Dept. of Computer Engineering, Yadavrao Tasgaonkar Institute of Engineering and Technology, Maharashtra, India [2]

ABSTRACT: Privacy and security are two important but seemingly contradictory objectives in a pervasive computing environment. On one hand, service providers want to authenticate legitimate users and make sure they are accessing their authorized services in a legal way. On the other hand, users want to maintain the necessary privacy without being tracked down for wherever they are and whatever they are doing. As a significantly business paradigm, many online information platforms have emerged to satisfy society's needs for person-specific data, where a service provider collects raw data from data contributors, and then offers value-added data services to data consumers.
In this paper, we propose TPDM, which efficiently integrates Truthfulness and Privacy preservation in Data Markets. TPDM is structured internally in an Encrypt-then
Sign fashion, using partially homomorphic encryption and identity-based signature. It simultaneously facilitates batch verification, data processing, and outcome verification, while maintaining identity preservation and data confidentiality and also instantiate TPDM with a profile matching service and a data distribution services.

KEYWORDS*:* Data, Truthfullness, privacy preservation,

I. INTRODUCTION

In the era of big data, society has developed an insatiable appetite for sharing personal data. Realizing the potential of personal data's economic value in decision making and user experience enhancement, several open information platforms have emerged to enable person-specific data to be exchanged on the Internet [1], [2], [3], [4], [5]. For example, Gnip, which is Twitter's enterprise API platform, collects social media data from Twitter users, mines deep insights into customized audiences, and provides data analysis solutions to more than 95% of the Fortune 500 [2]. However, there exists a critical security problem in these market-based platforms, i.e., it is difficult to guarantee the truthfulness in terms of data collection and data processing, especially when privacies of the data contributors are needed to be preserved. Let's examine the role of a pollster in the presidential election as follows. As a reliable source of intelligence, the Gallup Poll [6] uses impeccable data to assist presidential candidates in identifying and monitoring economic and behavioral indicators. In this scenario, simultaneously ensuring truthfulness and preserving privacy require the Gallup Poll to convince the presidential candidates that those indicators are derived from live interviews without leaking any interviewer's real identity (e.g., social security number) or the content of her interview. If raw data sets for drawing these indicators are mixed with even a small number of bogus or synthetic samples, it will exert bad influence on the final election result.

Ensuring 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 her profit. Therefore, in order to minimize the expenditure for data acquisition, an opportunistic way for the service provider is to mingle some bogus or synthetic data into the raw data sets. Yet, to reduce operation cost, a strategic service provider may provide data services based on a subset of the whole raw dataset, or even return a fake result without processing the data from designated data sources. However, if such speculative and illegal behaviors

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

cannot be identified and prohibited, it will cause heavy losses to the data consumers, and thus destabilize the data market. On the other hand, while unleashing the power of personal data, it is the bottom line of every business to respect the privacies of data contributors. The debacle, which follows AOL's public release of "anonymized" search records of its customers, highlights the potential risk to individuals in sharing personal data with private companies [7]. Besides, according to the survey report of 2016 TRUSTe /NCSA Consumer Privacy Infographic - US Edition [8], 89% say they avoid companies that do not protect their privacies. Therefore, 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.

II. REVIEW OF LITERATURE

In this paper, we have studied the varies paper and most of the author suggest the idea and suggest the varies techniques and algorithm. In the existing system the author suggest a two-layer system model for data markets. The model has a data acquisition layer and a data trading layer. There are four major kinds of entities, including data contributors, a service provider, data consumers, and a registration center. In the data acquisition layer, the service provider procures massive raw data from the data contributors, such as social network users, mobile smart devices, smart meters, and so on. In order to incentivize more data contributors to actively submit high-quality data, the service provider needs to reward those valid ones to compensate their data collection costs. For the sake of security, each registered data contributor is equipped with a tamper-proof device. The tamper-proof device can be implemented in the form of either specific hardware [6] or software [7]. It prevents any adversary from extracting the information stored in the device, including cryptographic keys, codes, and data. They consider that the service provider is cloud based, and has abundant computing resources, network bandwidths, and storage space. Besides, she tends to offer semantically rich and value-added data services to data consumers rather than directly revealing sensitive raw data, e.g., social network analyses, data distributions, personalized recommendations, and aggregate statistics.

The seminal paper [10] by Balazinska *et al.* discusses the implications of the emerging digital data markets, and lists the research opportunities in this direction. Li *et al.* [39] proposed a theory of pricing private data based on differential privacy. Upadhyaya *et al.* [11] developed a middleware system, called Data Lawyer, to formally specify data use policies, and to automatically enforce these pre-defined terms during data usage. Jung *et al.* [12] focused on the datasets resale issue at the dishonest data consumers. However, the original intention of above works is pricing data or monitoring data usage rather than integrating data truthfulness with privacy preservation in data markets, which is the consideration of our paper.

The registration center maintains an online database of registrations, and assigns each registered data contributor an identity and a password to activate the tamper-proof device. Besides, they maintains an official website, called certificated bulletin board [1], on which the legitimate system participants can publish essential information, e.g., white lists, blacklists, resubmit-lists, and reward-lists of data contributors. Yet, another duty of the registration center is to set up the parameters for a signature scheme and a cryptosystem. To avoid being a single point of failure or bottleneck, redundant registration centers, which have identical functionalities and databases, can be installed.

III. OBJECTIVE

Relevant objective of the proposed system are a s follows
1. To propose a efficient secure scheme for data markets, which simultaneously guarantees data truthfulness and privacy preservation.
2. To achieve the ultimate goal of the service provider in a data market is to maximize their profit.
3. To first secure mechanism for data markets achieving both data truthfulness and privacy preservation.
4. To Ensure truthfulness and protecting the privacies of data contributors.
5. To hide content of raw data from data consumers to guarantee data confidentiality, even if the real identities of the data contributors are hidden.
Ensuring truthfulness and protecting the privacies of data contributors are both important to the long term healthy development of data markets. Therefore, the content of raw data should not be disclosed to data consumers to guarantee

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

data confidentiality, even if the real identities of the data contributors are hidden. It is difficult to guarantee the truthfulness in terms of data collection and data processing, especially when privacies of the data contributors are needed to be preserved.
When we purchase a product at that time first we check user reviews according to that he/she decide the product is good or not, so we proposed this system to check the contributor is truthful or not limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you? In recent years, data market design has gained increasing interest, especially from the database community.

IV. PROBLEM DEFINITION

The motivation of the project is TPDM is the first secure mechanism for datamarkets achieving both data truthfulness and privacy preservation. To integrate truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truthfulness of data collection allows the data consumers to verify the validities of data contributors identities and the content of raw data, whereas privacy preservation tends to prevent them from learning these confidential contents. Specifically, the property of non-repudiation in classical digital signature schemes implies that the signature is unforgeable, and any third party is able to verify the authenticity of a data submitter using her public key and the corresponding digital certificate, i.e., the truthfulness of data collection in our model. However, the verification in digital signature schemes requires the knowledge of raw data, and can easily leak a data contributors real identity To the best of our knowledge, TPDM is the first secure mechanism for data markets achieving both data truthfulness and privacy preservation. TPDM is structured internally in a way of Encrypt then-Sign using partially homomorphic encryption and identity-based signature. It enforces the service provider to truthfully collect and to process real data. Besides, TPDM incorporates a two-layer batch verification scheme with an efficient outcome verification scheme, which can drastically reduce computation overhead

V. PROPOSED SYSTEM

To integrate truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truthfulness of data collection allows the data consumers to verify the validities of data contributors' identities and the content of raw data, whereas privacy preservation tends to prevent them from learning these confidential contents. Specifically, the property of non-repudiation inclassical digital signature schemes implies that the signature is unforgeable, and any third party is able to verify the authenticity of a data submitter using her public key and the corresponding digital certificate, i.e., the truthfulness of data collection in our model. However, the verification in digital signature schemes requires the knowledge of raw data, and can easily leak a data contributor's real identity [9]. Regarding a message authentication code (MAC), the data contributors and the data consumers need to agree on a shared secret key, which is unpractical in data markets.

Yet, another challenge comes from data processing, which makes verifying the truthfulness of data collection even harder. Nowadays, more and more data markets provide data services rather than directly offering raw data. The following three reasons account for such a trend: 1) For the data contributors, they have several privacy concerns [8]. Nevertheless, the service-based trading mode, which has hidden the sensitive raw data, alleviates their concerns; 2) For the service provider, semantically rich and insightful data services can bring in more profits [10]; 3) For the data consumers, data copyright infringement [11] and datasets resale [12] are serious. However, such a data trading mode differs from most of conventional data sharing scenarios, e.g., data publishing [13]. Besides, the result of data processing may no longer be semantically consistent with the raw data [14], which makes the data consumer hard to believe the truthfulness of data collection. In addition, the digital signatures on raw data become invalid for the data processing result, which discourages the data consumer from doing verification as mentioned above. Moreover,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

although data provenance [15] helps to determine the derivation history of a data processing result, it cannot guarantee the truthfulness of data collection.

The third challenge lies in how to guarantee the truthfulness of data processing, under the information asymmetry between the data consumer and the service provider due to data confidentiality. In particular, to ensure data confidentiality against the data consumer, the service provider can employ a conventional symmetric/asymmetric cryptosystem, and can let the data contributors encrypt their raw data. Unfortunately, a hidden problem arisen is that the data consumer fails to verify the correctness and completeness of a returned data service. Even worse, some greedy service providers may exploit this vulnerability to reduce operation cost during the execution of data processing, e.g., they might return an incomplete data service without processing the whole raw data set, or even return an outright fake result without processing the data from designated data sources.

Last but not least, the fourth design challenge is the efficiency requirement of data markets, especially for data acquisition, i.e., the service provider should be able to collect data from a large number of data contributors with low latency. Due to the timeliness of some kinds of person specific data, the service provider has to periodically collect fresh raw data to meet the diverse demands of high quality data services. For example, 25 billion data collection activities take place on Gnip every day [2]. Meanwhile, the service provider needs to verify data authentication and data integrity. One basic approach is to let each data contributor sign her raw data. However, classical digital signature schemes, which verify the received signatures one after another, may fail to satisfy the stringent time requirement of data markets. Furthermore, the maintenance of digital certificates under the traditional Public Key Infrastructure (PKI) also incurs significant communication overhead. Under such circumstances, verifying a large number of signatures sequentially certainly becomes the processing bottleneck at the service provider.

In the proposed system firstly the efficient secure scheme for data markets simultaneously guarantees data truthfulness and privacy preservation. In this system, user purchases product than he/she can send review to the system than system first check whether the contributors are authorized person or not. Under a specific data service, this system provides privacy preservation and verifiability.

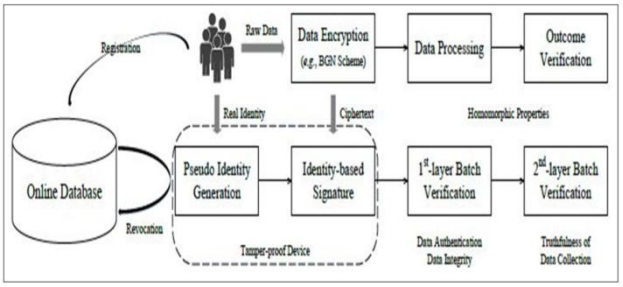

Figure N0-1 system Architecture

VI. DESIGN DETAIL

Following the guidelines given above, we now introduce
Data Truthfullness in detail. Data Truthfullness module consists of 5 phases: initialization, signing key generation, data submission, data processing and verifications, and tracing and revocation.

a) INITIALIZATION

In this module initialize the data contributor and registration centre which every unit has ready to play our role and initialize the algorithm which encrypts the uploaded data by data contributor.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

b) DATA ENCRYPTION

In this unit data contributor upload the data with private key verification and provide this encrypted data to service provider with valid key.

c) **DATA PROCESSING AND VERIFICATIONS**

In this phase, we consider two-layer batch verifications, *i.e.*, verifications conducted by both the service provider and the data consumer. Between the two-layer batch verification, we introduce data processing and signatures aggregation done by the service provider. At last, we present outcome verification conducted by the data consumer.

d)**DATA PROCESSING AND SIGNATURES AGGREGATION**

Instead of directly trading raw data for revenue, more and more service providers tend to trade value-added data services, *e.g.*, social network analysis, personalized recommendation, location-based service, and data distribution.

e) DATA DISTRIBUTOR

Data distributor gets the valid data with processed verification phase. Validators as per request of distributor provide 2 layer verification and submission process also provides the privacy preservation of specified file and finally provides the valid key to data distributor for the decryption purpose. Finally data distributor decrypts that data and achieves the data truthfulness.

f) **Truthfulness of Data Collection**

To guarantee the truthfulness of data collection, we need to combat the partial data collection attack defined. We note that it is just a special case, where the service provider is the attacker. Hence, it is infeasible for the service provider to forge valid signatures on behalf of any registered data contributor. Such an appealing property prevents the service provider from injecting spurious data undetectably, and enforces her to truthfully collect real data.

VII. SECURITY ANALYSIS

In this paper we have generate the varies keys for data contributor and data distributor. Our main aim to provide the security and privacy preservation which data have uploaded by data contributor. For the security purpose we consider the feasibility of the registration center from the perspectives of computation, communication, and storage overheads. We implement the identity based signature scheme with MNT159. In addition, for the profile matching service, the number of attributes is fixed at 10, and the number of valid data contributors' m is set to be 10000. Accordingly, the number of matched ones φ is 449 at $\delta = 12$. For the data distribution service, we fix the number of random variables β at 8, and set the number of valid data contributors to be 10000.

VIII. DEPTH-TRACING ALGORITHM

To evaluate the feasibility of `-DEPTH-TRACING algorithm when the batch verification fails, we generate a collection of 1024 valid signatures, and then randomly corrupt an α-fraction of the batch by replacing them with random elements from the cyclic group G1. We repeat this evaluation with various values of α ranging from 0 to 20%, and compare the verification latency per signature in batch verification with that in single signature verification. Here, the batch verification time includes the time cost spent in identifying invalid signatures. Batch verification is preferable to single signature verification when the ratio of invalid signatures is up to 16%. The worst case of batch verification happens when the invalid signatures are distributed uniformly. In case the invalid signatures are clustered together, the performance of batch verification should be better. The service provider can preset a practical tracing depth, and let those unidentifified data contributors do resubmissions.

ALGORITHM 1:- DEPTH-TRACING

1. Initialization: S = $\{\sigma_1,\cdots,\sigma_n\}$, head = 1, tail = n, limit = l
2. whitelist = ∅, blacklist = ∅, resubmitlist = ∅ 1

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

3. Function l-DEPTH-TRACING(S,head,tail,limit)
4. If |whitelist| + |blacklist| = n or limit = 0 then
5. return
6. else if CHECK-VALID (S,head,tail) = true then
7. ADD-TO-WHITELIST(head,tail)
8. else if head = tail then Single signature verification
9. ADD-TO-BLACKLIST(head,tail)
10. else Batch signatures verification from σ_{head} to σ_{tail}
11. mid = [head+tail]/2
12. l-DEPTH-TRACING(S,head,mid,limit−1)
13. l-DEPTH-TRACING(S,mid + 1,tail,limit−1)

ALGORITHM 2:-AES

1. Key Expansions-round keys are derived from the cipher key using Rijndael's key schedule. AES requires a separate 128-bit round key block for each round plus one more.
2. Initial Round

(a) Add Round Key - each byte of the state is combined with a block of the round key using bitwise xor.

3. Rounds

(a) Sub Bytes - a non-linear substitution step where each byte is replaced with another according to a lookup table.

(b) Shift Rows-a transposition step where the last three rows of the state are shifted cyclically a certain number of steps.

(c) Mix Columns - a mixing operation which operates on the columns of the state, combining the four bytes in each column.

(d) Add Round Key

4. Final Round (no Mix Columns)

(a) Sub Bytes

(b) Shift Rows

(c) Add Round Key

IX. RESULTS AND DISCUSSION

Finally we have received the varies results, they may depends on data contributor. How the data contributor contributes our data i.e. types of data and size of the data which they have uploaded for the service provider when signature verification and batch verification get the more time which happens in process of contributor. Final results analysis graph was shown in following figure no-2

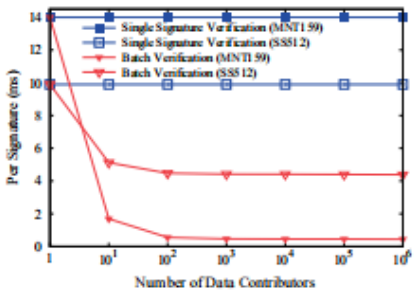

Figure N0-2 Analysis

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

X. CONCLUSION

The data contributors have to truthfully submit their own data, but cannot impersonate others. Besides, the service provider is enforced to truthfully collect and process data. In addition, In this system instantiated two different data services, and extensively evaluated their performances on two real-world datasets. The personally identifiable information and the sensitive raw data of data contributors are well protected.

REFERENCES

[1] T. Jung, X. Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su, Account Trade: accountable protocols for big data trading against dishonest consumers," in INFOCOM, 2017.
[2] "TRUSTe/NCSA Consumer Privacy Infographic – US Edition,"https://www:truste:com/resources/privacy research/ ncsa-consumer-privacy-index-us/. 2016
[3] P. Upadhyaya, M. Balazinska, and D. Suciu, "Automatic enforcement of data use policies with datalawyer," in SIGMOD, 2015
[4] R. Ikeda, A. D. Sarma, and J.Widom, "Logical provenance in dataoriented workflows?" in ICDE, 2013
[5] B. C. M. Fung, K.Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," ACM Computing Surveys, vol. 42, no. 4, pp. 1–53, Jun. 2010.
[6] T. W. Chim, S. Yiu, L. C. K. Hui, and V. O. K. Li, "SPECS: secure and privacy enhancing communications schemes for VANETs," Ad Hoc Networks, vol. 9, no. 2,pp. 189 – 203, 2011.
[7] G. Ghinita, P. Kalnis, and Y. Tao, "Anonymous publication of sensitive transactional data," IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 2, pp. 161–174, 2011
[8] M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunity for the database community," PVLDB, vol. 4, no. 12, pp. 1482–1485, 2011.
[9] M. Barbaro, T. Zeller, and S. Hansell, "A face is exposed for AOL searcher no. 4417749," New York Times, Aug. 2006.
[10] K. Ren,W. Lou, K. Kim, and R. Deng, "A novel privacy preserving authentication and access control scheme for pervasive computing environments," IEEE Transactions on Vehicular Technology, vol. 55, no. 4, pp. 1373–1384, 2006