RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

ENSURING DATA ACCURACY AND PRIVACY PRESERVATION IN DATA MARKETS

[1]Thambala Ramesh,[2]Sanga Ravikiran,[3]Jellapuram. Roja, [4]Sk Yakubi
[123]Assistant Professor,[4]Student
Department of CSE Engineering
Abdul Kalam Institute of Technological Sciences, Kothagudem,Telangana

ABSTRACT:

Many online information platforms have arisen as an important business paradigm in order to serve the demands of society for person-specific data. These platforms are characterized by the fact that a service provider receives raw data from data contributors and then provides value-added data services to data consumers. The data consumers, on the other hand, are confronted with a significant dilemma at the data trading layer, which is the question of how to check whether the service provider has obtained and processed data in an honest manner. Additionally, the individuals that provide data are often hesitant to divulge their sensitive personal information as well as their true identities to the individuals who consume the data. TPDM is a proposal that we provide in this work. It is designed to incorporate Truthfulness and Privacy protection in Data Markets in and efficient manner. Internally, TPDM is organized in an Encrypt-then-Sign way, with partly homomorphic encryption and identity-based signatures being used. In addition to preserving identification and ensuring the confidentiality of data, it concurrently makes batch verification, data processing, and result verification easier to do. We additionally instantiate TPDM with a profile matching service and a data distribution service, and we conduct an in-depth analysis of how well these services work on the Yahoo! Music ratings dataset and the 2009 RECS dataset, respectively. The findings of our study and assessment show that TPDM is capable of achieving a number of desired features while also incurring minimal computing and communication overheads when it comes to supporting large-scale data marketplaces or markets.

1. INTRODUCTION

As a result of the expectations that society has for data that is personal to individuals, several online information platforms have emerged as a significant alternative to traditional business models. A service provider gets raw data from data contributors and then delivers value-added data services to data consumers. This is the defining characteristic of these platforms, which are distinguished by another characteristic. On the other side, the data consumers are presented with a fundamental difficulty at the data trading layer, which is the question of how to verify whether the service provider has gotten and processed data in an honest way. This is a problem since the data consumers are the ones who are using the data. Further, the persons who offer the data are often reluctant to reveal their sensitive personal information as well as their genuine identities to the others who consume the data. This is because the data is being consumed by other individuals. One of the proposals that we provide in this paper is called TPDM. The purpose of this system is to ensure that the protection of privacy and truthfulness in data markets is carried out in an effective way. On the inside, TPDM is structured in an Encrypt-then-Sign manner, with identity-based signatures and partially homomorphic encryption being used. Batch verification, data processing, and result verification are all made simpler as a consequence of this, in addition to the fact that it protects the confidentiality of data and

maintains the identify of individuals. TPDM is further instantiated with a profile matching service and a data distribution service, and we perform an in-depth examination of how well these services operate on the Yahoo! Music ratings dataset and the 2009 RECS dataset, respectively. TPDM is a service that allows users to match profiles with other users. When it comes to providing support for large-scale data marketplaces or markets, the outcomes of our research and evaluation indicate that TPDM is capable of attaining a number of characteristics that are required while also incurring minimum processing and communication overheads.

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 the 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 data 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 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 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 set. Yet, to reduce operation cost, a cunning service provider may provide data services based on a subset of the whole raw data set, or even return a fake result without processing the data from designated sources. However, if such speculative and illegal behaviors cannot be identified and prohibited, it will cause heavy losses to 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/NCSAConsumerPrivacy Infographic- US Edition [8], 89% of consumers say they avoid companies that do not respect privacy. Therefore, the content of raw data should not be disclosed to the data consumers to guarantee data confidentiality, even if the real identities of the data contributors are hidden.

2. RELATED WORK

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 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 set. Yet, to reduce operation cost, a cunning service provider may provide data services based on a subset of the whole raw data set, or even return a fake result without processing the data from designated sources. However, if such speculative and illegal behaviors cannot be identified and

prohibited, it will cause heavy losses to 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/NCSAConsumerPrivacy Infographic- US Edition [8], 89% of consumers say they avoid companies that do not respect privacy. Therefore, the content of raw data should not be disclosed to the data consumers to guarantee data confidentiality, even if the real identities of the data contributors are hidden.

3. EXISTING SYSTEM:

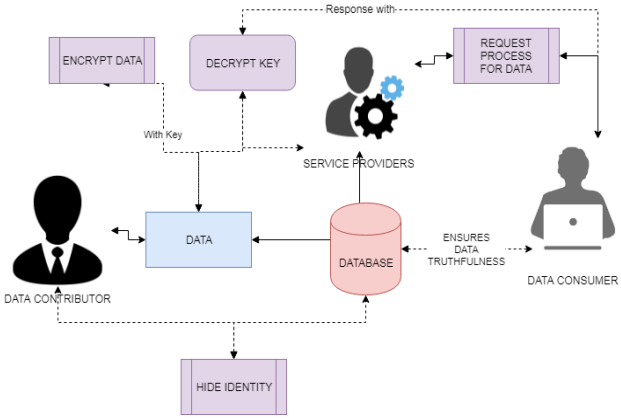
To integrate truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truthfulness 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 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 contributor's real identity. 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.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; 3) for the data consumers, data copyright infringement and datasets resale are serious. However, such a data trading mode differs from most of conventional data sharing scenarios, e.g., data publishing. Besides, the result of data processing may no longer be semantically consistent with the raw data, 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, although data provenance helps to determine the derivation history of a data processing result, it cannot guarantee the truthfulness of data collection.

4. PROPOSED SYSTEM:

In this Project, by jointly considering above four challenges, we propose TPDM, which achieves both Truthfulness and Privacy preservation in Data Markets. TPDM first exploits partially homomorphic encryption to construct a ciphertext space, which enables the service provider to launch data services and the data consumers to verify the correctness and completeness of data processing results, while maintaining data confidentiality. In contrast to classical digital signature schemes, which are operated over plaintexts, our new identity-based

signature scheme is conducted in the ciphertext space. Furthermore, each data contributor's signature is derived from her real identity, and is unforgeable against the service provider or other external attackers. This appealing property can convince data consumers that the service provider has truthfullycollected data. To reduce the latency caused by verifying a bulk of signatures, we propose a two-layer batch verification scheme, which is built on the bilinearity of admissible pairing. At last, TPDM realizes identity preservation and revocability by carefully adopting ElGamal encryption and introducing a semi-honest registration center. We summarize our key contributions as follows. 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 Encryptthen-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.

5. SYSTEM ARCHITECTURE:



6. MODULES

DATA CONTRIBUTOR

The user undergoes registration method the Registration centre can give the pseudo identity and provides them to the user. We have a tendency to assume that the registration centre sets up the system parameters at the start of information commerce. The verification conducted by each the service supplier and also the knowledge shopper. Between the two-layer batch verifications, we have a tendency to introduce processing and signatures aggregation done by the service supplier. At last, we have a tendency to gift outcome verifications conducted by the information shopper. The information contributors is in have to expose the service that provided by them, in terms of the entire package of the service. The package that comprises the small print like a product that give by the contributor and also the various price for the every product in commission. And a complete price of the service. The information contributor is ready to turn out any range (N numbers) of service and every are declared as separate package

DATA COLLECTOR

The collector surfs with the contributor services and choose the required package of services. And also the collector submits the resource request to the various CSP of service. If the CSP acknowledge the collector request of resource, currently the collector is prepared to access the resource details and to supply the various resource to requesting service supplier. Collector serves a intermediate between the broker and also the CSP.

DATA PROVIDER

The service provider will ready to choose the service that in would like from the service provided by the collector from the CSP. If the service supplier selects their desired package of service, then the service supplier ought to pay money for the various services. If the service

supplier is paid with the service the service provider will access the service from collector.

DATA CONSUMER

The buyer hunt for the service that they have from the assorted service suppliers. And if the buyer finds the required service they request the service to the service supplier and obtain use with the resource. And verify or cross check the resource that bought from the service supplier that whether or not service provider serves the proper resource in cheap price.

7. 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 their own 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, especially from computation and communication overheads. At last, we have shown the feasibility of introducing the semi-honest registration center with detailed theoretical analysis and substantial evaluations.

REFERENCES

1. "Microsoft Azure Marketplace," https://datamarket.azure.com/home/.
2. "Gnip," https://gnip.com/.
3. "DataSift," http://datasift.com/.
4. "Datacoup," https://datacoup.com/.
5. "Citizenme," https://www.citizenme.com/.
6. "Gallup Poll," http://www.gallup.com/.
7. M. Barbaro, T. Zeller, and S. Hansell, "A face is exposed for AOL searcher no. 4417749," New York Times, Aug. 2006.
8. "2016 TRUSTe/NCSA Consumer Privacy Infographic – US Edition," https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/.
9. 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.
10. 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.
11. P. Upadhyaya, M. Balazinska, and D. Suciu, "Automatic enforcement of data use policies with datalawyer," in SIGMOD, 2015.
12. 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 dishonest consumers," in INFOCOM, 2017.
13. 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.
14. 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.

15. R. Ikeda, A. D. Sarma, and J. Widom, "Logical provenance in dataoriented workflows?" in ICDE, 2013.
16. M. Raya and J. Hubaux, "Securing vehicular ad hoc networks," Journal of Computer Security, vol. 15, no. 1, pp. 39–68, 2007.
17. 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.
18. D. Boneh, E. Goh, and K. Nissim, "Evaluating 2-dnf formulas on ciphertexts," in TCC, 2005.