OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SURVEY ON ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Mr. Shanteshwar Bhagwan Khankare[1], Professor Varsha Dange[2]

PG Students Department of Computer Engineering, Dhole Patil College of Engineering, Wagholi, Near Eon IT park., Pune-412207, Maharashtra, India[1]

Assistant Professor, Department of Computer Engineering, Dhole Patil College of Engineering, Wagholi, Near Eon IT park., Pune-412207, Maharashtra, India[2]

Abstract: **As a significant 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. However, in the data trading layer, the data consumers face a pressing problem, i.e., how to verify whether the service provider has truthfully collected and processed data? Furthermore, the data contributors are usually unwilling to reveal their sensitive personal data and real identities to the data consumers. The proposed a system which finds the contributors are Truthfulness or not using SVM algorithm. In this system user purchase product than he/she can send review to the system than system first check whether the contributors are authorized person or not.**

Keywords: Data truthfulness, service provider

I INTRODUCTION

In this project 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. The motivation of the project is TPDM is the first secure mechanism for data markets achieving both data truthfulness and privacy preservation.

II LITERATURE SURVEY

In this Paper, Machine learning classification is used in numerous settings nowadays, such as medical or genomics predictions, spam detection, face recognition, and financial predictions. Due to privacy concerns, in some of these applications, it is important that the data and the classifier remain confidential. In this work, we construct three major classification protocols that satisfy this privacy constraint: hyper plane decision, Naive Bayes, and decision trees. We also enable these protocols to be combined with AdaBoost. At the basis of these constructions is a new library of building blocks for constructing classifiers securely; we demonstrate that this library can be used to construct other classifiers as well, such as a multiplexer and a face detection classifier. We implemented and evaluated our library and classifiers. Our protocols are efficient, taking milliseconds to a few seconds to perform a classification when running on real medical data sets.[1]

Author Presents, Proximity-based mobile social networking (PMSN) refers to the social interaction among physically proximate mobile users directly through the Bluetooth/WiFi interfaces on their smartphones or other mobile devices. It becomes increasingly popular due to the recently explosive growth of smartphone users. Profile matching means two users comparing their personal profiles and is often the first step towards effective PMSN. It, however, conflicts with users' growing privacy concerns about disclosing their personal profiles to complete strangers before deciding to interact with them. This paper tackles this open challenge by designing a suite of novel fine-grained private matching protocols. Our protocols enable two users to perform profile matching without disclosing any information about their profiles beyond the comparison result. In contrast to existing coarsegrained private matching schemes for PMSN, our protocols allow finer differentiation between PMSN users and can support a wide range of matching metrics at different privacy levels. The security and communication/computation overhead of our protocols are thoroughly analyzed and evaluated via detailed simulations.[2]

Cloud-computing is transforming many aspects of data management. Most recently, the cloud is seeing the emergence of digital markets for data and associated services. We observe that our community has a lot to offer in building successful cloud-based data markets. We outline some of the key challenges that such markets face and discuss the associated research problems that our community can help solve.[3]

As a significant business paradigm, data trading has attracted increasing attention. However, the study of data acquisition in data markets is still in its infancy. Mobile crowd sensing has been recognized as an efficient and scalable way to acquire large-scale data. Designing a practical data acquisition scheme for crowd-sensed data markets has to consider three major challenges: crowd-sensed data trading format determination, profit maximization with polynomial computational complexity, and payment minimization in strategic environments. In this paper, we jointly consider these design challenges, and propose VENUS, which is the first profit-driVEN data acqUiSition framework for crowd-sensed data markets. Specifically, VENUS consists of two complementary mechanisms: VENUS-PRO for profit maximization and VENUS-PAY for payment minimization. Given the expected payment for each of the data acquisition points, VENUS-PRO greedily selects the most "costefficient" data acquisition points to achieve a sub-optimal profit.[4]

III PROBLEM STATEMENT

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.

IV SYSTEM ARCHITECTURE

In the proposed system first efficient secure scheme for data markets, which simultaneously guarantees data truthfulness and privacy preservation. In this system user purchase 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.

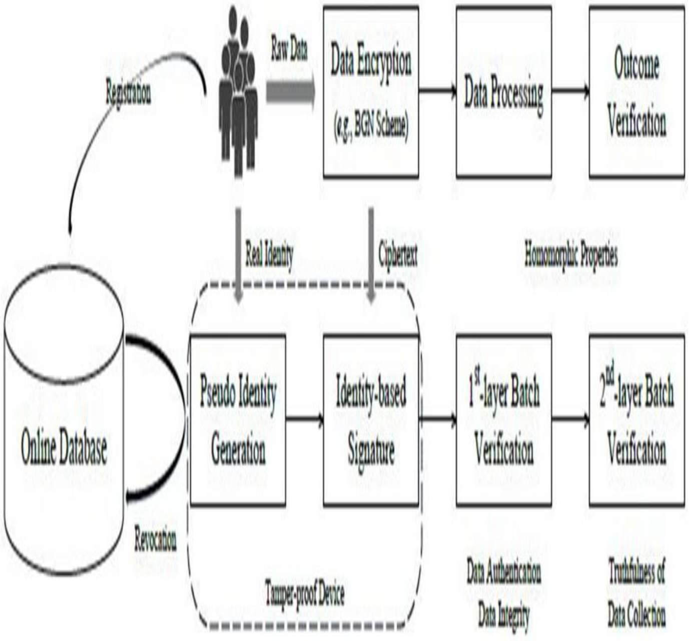


Figure 1 System architecture

V ALGORITHM

1.AES ALGORITHM

1. KeyExpansions

For each round AES requires a separate 128-bit round key block plus one more.

2. InitialRound

AddRoundKey—with a block of the round key, each byte of the state is combined using bitwise xor.

3. Rounds

SubBytes—in this step each byte is replaced with another byte.

ShiftRows— for a certain number of steps, the last three rows of the state are shifted cyclically.

MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.

AddRoundKey

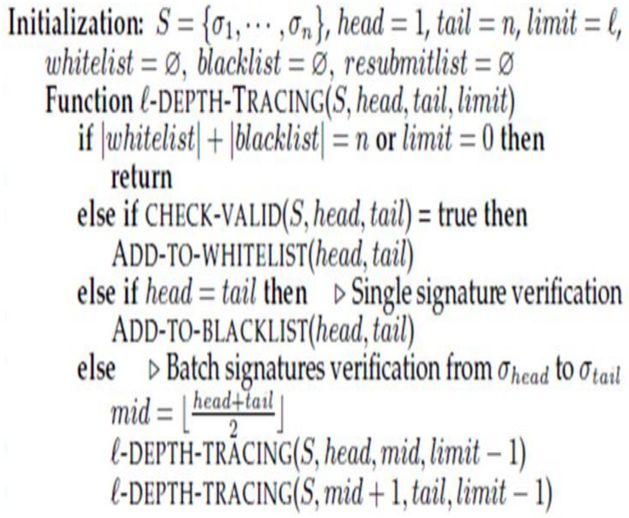
Final Round (no MixColumns)

SubBytes

ShiftRows

AddRoundKey.

2. L-DEPTH-TRACING



```
Initialization: $S = \{\sigma_1, \cdots, \sigma_n\}$, $head = 1$, $tail = n$, $limit = \ell$,
  $whitelist = \emptyset$, $blacklist = \emptyset$, $resubmitlist = \emptyset$
  Function $\ell$-DEPTH-TRACING($S, head, tail, limit$)
    if $|whitelist| + |blacklist| = n$ or $limit = 0$ then
      return
    else if CHECK-VALID($S, head, tail$) = true then
      ADD-TO-WHITELIST($head, tail$)
    else if $head = tail$ then    ▷ Single signature verification
      ADD-TO-BLACKLIST($head, tail$)
    else    ▷ Batch signatures verification from $\sigma_{head}$ to $\sigma_{tail}$
      $mid = \lfloor \frac{head+tail}{2} \rfloor$
      $\ell$-DEPTH-TRACING($S, head, mid, limit - 1$)
      $\ell$-DEPTH-TRACING($S, mid + 1, tail, limit - 1$)
```

VI CONCLUSION AND FUTURE WORK

In this project, 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, AccountTrade: 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.