Achieve Accuracy of Data and Confidentiality Preservation in Data Markets

Miss. Aishwarya Pratap Jadhav, Prof. M. A. Wakchaure,
Department of Computer Engineering Amrutvahini College of Engineering

Abstract- People are posting their data on different sites. Huge amount of data were collected daily by the different users. Research has been done on the accumulation, different kind of applications proceeded rapidly. The collection of digital data by governments, corporations, and individuals has created tremendous opportunities for knowledge based decision making. Whenever any user want to buy any online product, usually people will inform themselves by reading the different online reviews. Their might be chances that, those particular user is fake user and posting bogus data. Also security is also the main issue. furthermore, within the data trading layer, the data user facing a many trouble, i.e., how to confirm whether the data is truly collected and organized by the service provider or not? Also, the data subscribers are normally not ready to uncover their sensitive information and real personalities to the data user. In this paper, we represent the TPDM concept, which productively combine Truthfulness and Privacy Preservation in Data Markets. TPDM is based on encryption-decryption mechanism, and also uses a few degree fully- homomorphic encryption and personality primarily based mark. Also we moreover introduced TPDM with the profile coordinating management, it check whether user's Id is real or false Id .

Keywords- Data truthfulness, identity based signature, privacy preservation, data makets.

I. INTRODUCTION

Now a day large amount of data were present, many users share their private data. There are many open records systems were present. Due to which many users exchanged their data on the internet. For example, Facebook and twitter's API platforms were present, that collects personal social media data of many users. But, there is a security issues in those market platforms, i.e., it is hard to assure that the data which is collected is true data. Also privateness of the data subscribers are needed to be preserved.

Fig.1: A layer system model for data market

To combine truthfulness and privateness maintenance in a data marketplace, there are main four challenges. The most important first design challenge is that to verifying the truthfulness of data which is collected from different users and gives confidentiality for the privacy of that data. Data collected from user must be valid user. The raw data is contributed by the data provider. The next challenge comes up from data processing, which makes verifying the truthfulness of facts series even more difficult. these days, increasingly records markets provide information services rather than giving raw data directly to the user. In this paper TPDM were introduced, Truthfulness and Privacy Preservation in Data Markets. The Fully homomorphic encryption technique were used for providing the protected mechanism to the data by performing specific operations on that specific data i.e, addition and multiplication operation on plain text data and make it ciphertext. The third challenge is based on how to give assurance of the the accuracy of data processing, under the information asymmetry between the data users and the service provider due to data truthfulness. Mainly, to ensure data confidentiality against the data giver, the service provider can employ a conventional symmetric /asymmetric cryptosystem, and can let the data subscriber do encryption on their raw data to generate ciphertext. Unfortunately, a hidden problem arisen is that the data users fails to confirm the correctness and completeness of a again records carrier. And the last design challenge is the effectiveness need of data markets, the service provider need to be able to gather all given records from a big wide variety of data subscribers with low latency. Also, the service provider must be verify the data confidentiality.

II. LITARATURE SURVEY

In the paper [2], C. Wang, Q. Wang, K. Ren, and W. Lou this people gives 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.

Jan Camenisch, Susan Hohenberger , Michael stergaard Pedersen, proposed a paper on the first batch verifier for messages. Furthermore they also propose a new signature scheme with very short signatures, for which batch verification for many users is also highly efficient. Although the new signature scheme which they proposed has some

limitations, it is very efficient and still practical for some communication applications.[3]

Magdalena Balazinska, Bill Howe, and Dan Suciu, They discussed about It outline some of the some challenges that markets face and also discussed the associated research issues that our community can help solve. Also they told the implications of the emerging cloud-based data markets on the database research community. Our community has a great opportunity in making a significant impact on these data markets, while solving exciting data management research challenges.[4]

Dan Boneh, Matthew Franklin proposed a paper on fully functional identity-based encode scheme (IBE). The system is based on the bilinear maps between groups. In this paper identity-based encrypted mechanism is introduced. The privacy of the system is a natural analogue of the computational assumption. The main restriction of this system is Revocation for private key is not present. [5]

Seung Hyun Seo, Mohamed Nabeel, Xiaoyu Ding, proposed a paper on An Efficient Certificateless cyphertext were established for Secure Data Sharing in Public System storage clouds. The Safely share sensitive data and informative data in public system storage clouds. Additionally has downside that Network Connections Dependency furthermore Cost is more this calculation utilized is public key encryption algorithms.[6]

Ricardo Mendes and Joaqo P. Vilela these fellows proposed a paper and gives the study on the most relevant PPDM techniques from the literature and the measured used to evaluate such techniques and represents typical applications of PPDM methods in different similar fields. The pattern were introduced also known as Privacy-Preserving Data Mining (PPDM). In this survey, the reviews will be given on the data mining methods which are mostly relevantnt to PPDM of huge amount of information is provided.[7].

III. PROPOSED METHODOLOGY

In the proposed research work to design and implement a system which is first efficient secure scheme for the data markets, which simultaneously garantees the data accuracy and confidentiality preservation. The TPDM is structured internally in a way of Encode -then-Sign, using fully homomorphic encryption technique and identity based signature. Also the data which were used is taken by hotel data . There are two different data sets were used for identifying the real and false comments of different data users. The service provider must be Collect the true data and process that data.

A. System Architecture

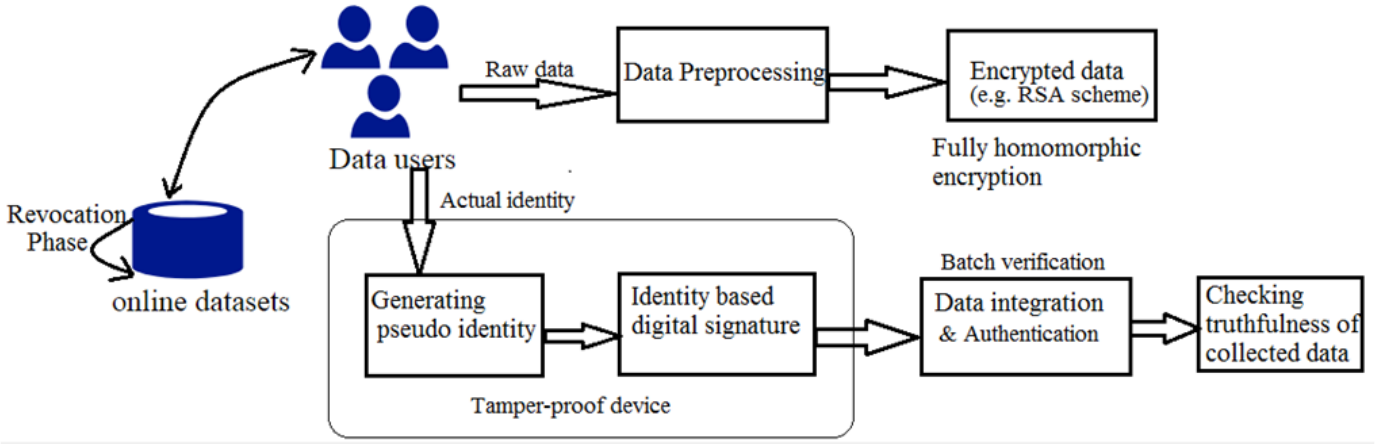

Fig.2: System Architecture

1. Key generation

To provides data confidentiality, we gives fully Homomorphic encryption . Which uses addition and multiplicative operation on data and encrypt that given data. Also it provides more security than partial homomorphic encryption .These are the simple operation performed on plain text to make it in encrypted format. So that any third party member can not easily accessed the informative data easily.

2. Data submission

Two-layer batch verification is considered in this phase which is done by both sides means at the service provider side and the data users side rerspectively. Batch verification is done for security purpose. Data pre-processing and signatures aggregation done by the service provider. And outcome verification conducted by the data users

- **Batch Verification:** After constructing the encrypted

data, we can allow each data subscriber digitally sign their encrypted raw data. But, digital certificates has incurs significant communication overhead. To handel these problem, we gives an identity-based signature scheme.

- **Space Construction:** The main problem is that how to enable the data user to verify the validnesses of user's signatures, while maintaining data confidentiality. Hence the public key encryption method is apply for the construction of the encrypted text, and then service provider has to decode/decrypt that data using decryption algorithm and then process the data. The fully homomorphic cryptosystem for is used for encryption and decryption.

3. Data preprocessing and verification:
   The fully homomorphic encryption properties used. Also the data users able to verify the accuracy of data processing. Under the different circumstances the data users should knows his/ her plaintext. Verification done for checking the truthfulness of data.

4. Revocation and Tracing:
   In any user misbehave the that particular user get revocate and donot have authority to access the data,modify the data.

IV.ALGORITHMS

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

B. Decryption :

1. Person A recovers m from c by exploitation his or herprivate key exponent, d, by the computation m cd (mod n).
2. Consider m, Person A will recover the first original message M by reversing the padding scheme.

This procedure works since c me (mod n), cd (me)d (mod n), cd mde (mod n).
By the symmetry property of mods we have that mde mde (mod n).
Since de = 1 + k(n), we can write
mde m1 + k(n) (mod n), mde m(mk)(n) (mod n), mde m (mod n).
B. ⊦- Depth tracing Algorithm [1] Initialization: S = {δ1,δ2,…..δn}, head = 1, tail = n,
limit = ⊦,
whitelist = Ø, blacklist = Ø, resubmitlist = Ø 1.Function ⊦-DEPTH-TRACING(S, head, tail,limit)
2. if │whitelist│ + │blacklist│ = n or limit = 0

I. RSA Algorithm
INPUT: Required modulus bit length, k OUTPUT: An RSA key pair ((N,e),d)
then then

3. return
4. else if CHECK-VALID(S, head, tail) = true
5. ADD-TO-WHITELIST(head, tail) 6.else if head = tail then //Single signature

1. Select a value of e from 3,5,17,257,655373,
2. repeat
3. p ← genprime(k/2)
4. until (pmode)≠1
5. repeat
6. q ← genprime(k - k/2)
7. until (qmode)≠1
8. N ← pq
9. φ (n) ← φ (p)* φ (q)↔(p-1)(q-1) // 'φ' Euler's totient function.
10. e ← 1<e< φ (n)
11. d ← ≡ e (mod φ(N))
12. return (N,e,d)

A. Encryption:
Sender does the following:
verification
7.ADD-TO-BLACKLIST(head, tail)
8. else // Batch signatures verification from δhead to δtail
9. mid = │ head+tail │
10. ⊦-DEPTH-TRACING(S,head, mid, limit -1)
11. ⊦-DEPTH-TRACING(S,mid + 1, tail, limit - 1)

V. SYSTEM REQUIREMENTS

A. Hardware Requirement

1. Main Processor : Any Processor above 1 GHz
2. Hard Disk : 20 GB.
3. Ram : 1GB.
4. Input device: Standard Keyboard and Mouse.
5. Output device : Monitor with normal visual Resolution.

B. Software Requirement

1. Platform : Java, My SQL 5.5 onwards
2. Software : Net Bean IDE 8.1
3. Operating System : Windows Family

VI. MATHEMATICAL MODEL

A. Mapping diagram

A function is a relationship that pairs each input with exactly one output. A function can be represented by ordered pairs or a mapping diagram. Also function mapping is the study of special type of relation i.e, f: x→y . Where, x is input data set and y is output datasets. A mapping diagram consists of two parallel columns. The first column represents the domain of a function f , and the other column for its range. Lines or arrows are drawn from domain to range, to represent the relation between any two elements.

- S be the whole system S= {I, P,O} Where,
  I =Input
  P =Procedure
  O =Output
  user u= {data distributor, service provider, data consumer}
- I= {I0, I1, I2, I3, I4}
  I0 = real id of user I1
  =user's raw data
  I2 =public/private key I3
  =Enter comments
  I4 =Service provider activities
- P ={P0, P1, P2, P3, P4, P5, P6, P7,P8}
  P0 = Login to registration server P1
  =create pseudo identity
  P2 =Data encryption(AES algorithm used)
  P3 =Identity base signature generation
  P4 =Data preprocessing
  P5 =Check data integrity and authenticity
  P6 =Check truthfulness of data
  P7 = Product review in data market
  P8 =Revoke user
- O= {O0,O1,O2}
  O0 =provide comments
  O1 =collection of truthful data O2 =Feedback to market place O3 =revocation of user

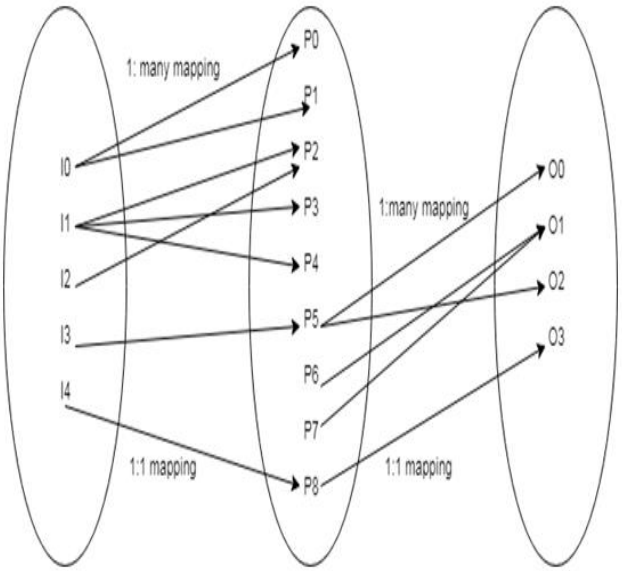

Fig.3: Function mapping

VII. RESULTS AND DISCUSSION

The final results of TPDM can be evaluated in terms of computation overhead and communication overhead. And finally discuss he/she practicality of TPDM in current data markets. The datasets which are used those are the real world datasets which is downloaded from data markets. Different types of datasents were present in huge number. The assesment results that reveal the TPDM can truly help to minimize the computation overheads of the service provider s. The system is executed using java platform with INTEL 2.8 GHz i7 processor and 16GB RAM. The graphical representation show the data distribution service.

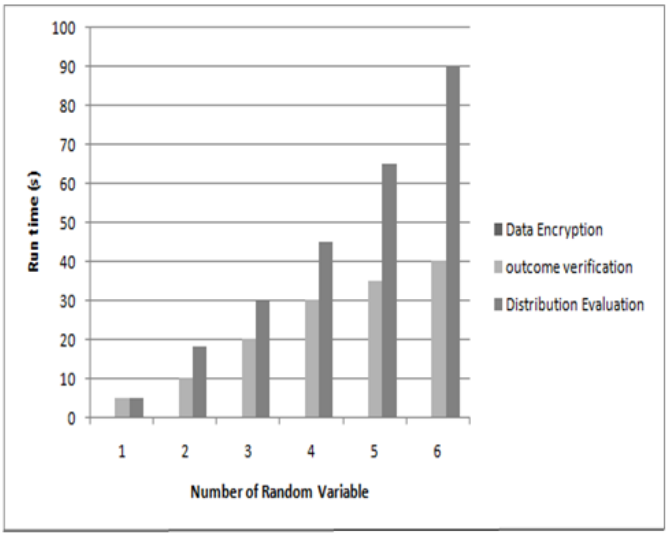

Fig.4: Data Distribution

Where, number of random variables will be taken which is

increases from 1 to 6 and also there is number of valid data subscribers were present, which is get fixed at 10000. The communication overhead is consider which get increase quadratically during each phase. The encryption algorithm is used from data encryption and decryption. The homomorphic addition and multiplication property were given. The above result will shows that the signature scheme that applied for data subscribers for security purpose in TPDM is efficient.

VIII. CONCLUSION

The TPDM has been proposed for truthfulness for data and providing the confidentiality preservation fir that data, the data subscribers have to truthfully submit their own data. Except, the service provider must collect true data . In TPDM, the data contribution have truthfully submit their own data, but cannot impersonate others. Moreover, both the personally identifiable information and the sensitive raw data of data subscribers are well protected. Service providers is forced to truthfully collect & process data. Both the personally identifiable information the sensitive raw data of data subscribers are well protected.

IX. REFERENCES

[1]. Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Xiaofeng Gao and Guihai Chen " Achieving Data Truthfulness and Privacy Preservation in Data Markets "‟ IEEE Transactions on Knowledge and Data Engineering (2018 Early Access Study on the ISCX Dataset." Data Intelligence and Security (ICDIS), 2018 1st International Conference on.IEEE, 2018.

[2]. 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.

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

[4]. 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.

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

[6]. Seung-Hyun Seo, Member, IEEE, Mohamed Nabeel, Member, IEEE, Xiaoyu Ding, Student Member, IEEE, and Elisa Bertino, Fellow,An Efficient Certificateless Encryption for Secure Data Sharing in Public system storage clouds,Vol.25, No.9, PP.2107.

[7]. Ricardo Mendes, Student member, And Joaqo P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications", Vol. 5, 2017

[8]. Z. Zheng, Member, IEEE, Y. Peng, Member, IEEE, F. Wu, S. Tang,Member, IEEE, and G. Chen, Member, IEEE "Trading data in the crowd: Profit- driven data acquisition for mobile crowdsensing," IEEE Journal on Selected Areas in Communications, vol. 35, no. 2, pp. 486–501, 2017..

[9]. M. Barbaro, T. Zeller, and S. Hansell,Fellows ―A face is exposed for AOL searcher no. 4417749, N Y Times, August 9, 2006.

[10]. C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in INFOCOM, 2010.

[11]. Wakchaure M. A., Sane S. S. ,An Algorithm for Discrimination Prevention in Data Mining: Implementation Statistics and Analysis. In2018 International Conference On Advances in Communication and Computing Technology (ICACCT) 2018 Feb 8 (pp. 403-409). IEEE.

[12]. Shitole M, Wakchaure M. A. ,Survey: Techniques Of Data Mining For Clinical Decision Support System. Vol-2 Issue-1 IJARIIE-ISSN (O)-2395-4396.;1571