International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)

ISSN: 0976-1353 Volume 14 Issue 1 –APRIL 2015.

MULTIPARTY PRIVATE DATA PUBLISHING USING EXPONENTIAL TRUST BASED MECHANISM IN MALICIOUS ADVERSARY MODEL

TAMILMANI.M
Department of computer science and engineering
Sona college of technology
Salem ,India

BALAMURUGAN.D
Department of computer science and engineering
sona college of technology
Salem, India

mtamil.2009@gmail.com
balamuruganD-81@yahoo.com

ABSTRACT
Differential privacy is a recent privacy definition that provides a strong privacy guarantee. It guarantees that an adversary learns not anything more about an individual, regardless of whether her record is present or absent in the data. In this paper, propose an algorithm to securely integrate person-specific sensitive information from two information providers, whereby the integrated data still retain the essential information for supporting data mining tasks. This protocol can be used as a sub protocol by any other algorithm that requires the exponential mechanism in a distributed setting. The propose a two party algorithm that releases differentially private data in a secure way according to the definition of secure multiparty computation. The proposed algorithm can be extended for more than two parties by modifying all the sub protocols while keeping the general top-down structure of the algorithm. The planned algorithmic rule will be extended for over two parties by modifying all the sub protocols whereas keeping the final top-down structure of the algorithmic rule. To increase the algorithmic rule for malicious parties, all sub protocols ought to be extended and should be secure underneath the malicious antagonist model. The general top-down structure of the algorithm provides the data privacy for more than the two parties. Experimental results on real-life knowledge recommend that the planned algorithmic rule will effectively preserve data for an information mining task.
Key words: Differential privacy, Exponential mechanism, Secure two party, secure data integration.

1. INTRODUCTION

Each information is in hand by a specific autonomous entity, as an example, medical knowledge by hospitals, financial gain knowledge by tax agencies, money knowledge by banks, and census knowledge by applied mathematics agencies. Moreover, the emergence of recent paradigms like cloud computing will increase the number of information distributed between multiple entities. This distributed knowledge will be integrated to modify higher knowledge analysis for creating higher choices and providing high-quality services. As an example, knowledge will be integrated to enhance medical analysis, client service, or Homeland Security. However, knowledge integration between autonomous entities ought to be conducted in such some way that no additional data than necessary is disclosed between the taking part entities.

At constant time, new data that results from the mixing method mustn't be misused by adversaries to reveal sensitive data that wasn't offered before the integration. During this paper, propose AN formula to firmly integrate person-specific sensitive knowledge from 2 knowledge suppliers, whereby the integrated knowledge still retain the essential data for supporting data processing tasks. The subsequent real-life state of affairs additional illustrates the requirement for synchronal knowledge sharing and privacy preservation of person-specific sensitive knowledge.

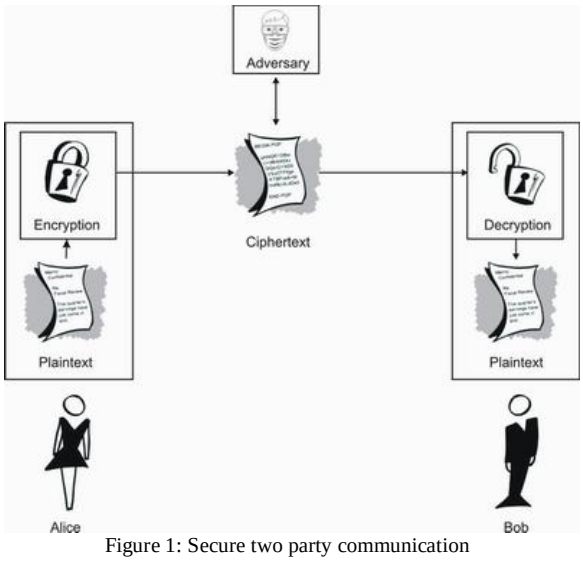

Figure 1: Secure two party communication

Data can be horizontally-partitioned among different parties over the same set of attributes. These distributed data can be integrated for making better decisions and providing high-quality services. However, data integration should be conducted in a way that no more information than necessary should be revealed between the participating entities. At the same time, novel knowledge that results from the integration process should not be misused by adversaries to reveal sensitive information that has not been available before the data integration.

The challenge in data privacy is to share data while protecting personally identifiable information. Differential privacy is a strong privacy definition. A two-party protocol for the exponential mechanism. It is a sub protocol of main algorithm. It uses the exponential mechanism in a distributed setting. Two-party data publishing algorithm for vertically partitioned data that generate an integrated data table satisfying differential privacy. Distributed data can be integrated to enable better data analysis for making better decisions and providing high-quality services. The first two-party differentially private data release algorithm for vertically partitioned data. It respect to a utility function while preserving differential privacy. A secure mechanism is required to compute the same output while ensuring that no extra information is leaked to any party.

The main contribution of our paper can be summarized as follows:

Present a two-party protocol for the exponential mechanism. Use this protocol as a sub protocol of main algorithm, and it can also be used by any other algorithm that uses the exponential mechanism in a distributed setting. To present the first two-party data publishing algorithm for vertically partitioned data that generate an integrated data table satisfying differential privacy. The algorithm also satisfies the security definition in the secure multiparty computation (SMC) literature.

2. RELATED WORK

In this paper [1]R.Bhaskar, S. Laxman, A.SmithandA.Thakurta they present economical algorithms for locating the foremost frequent patterns in an exceedingly knowledge set of sensitive records. The algorithms satisfy differential privacy, a recently introduced definition that pro-videos important privacy guarantees within the presence of absolute external data. Differentially non-public algorithms need a degree of uncertainty in their output to preserve privacy.

In this paper [6] A. Inan, M. Kantarcioglu, G. Ghinita, and E. Bertino they outlined personal matching may be solved firmly and accurately mistreatment secure multi-party computation (SMC) techniques, however such Associate in Nursing approach is prohibitively overpriced in apply. Previous work planned the discharge of sanitized versions of the sensitive datasets that permits interference, i.e., filtering out sub-sets of records that can't be a part of be part of result. This way, SMC is applied solely to a tiny low fraction of record pairs, reducing the matching price to acceptable levels.

In this paper [5] A. Frank and A. Asuncion . Based on the dearth of such a threshold, this paper presents a step by step guide for distinctive the dataset threshold for the performance estimators in supervised machine learning experiments. The identification of the dataset threshold involves playacting experiments exploitation four totally {different|completely different} datasets having different sample sizes from the University of CA Irvine (UCI) machine learning repository. The sample sizes area unit classified in respect to range|the amount|the quantity} of attributes and number of instances obtainable within the dataset.

In this paper [7] A. Friedman and A. Schuster consider the matter of knowledge mining with formal privacy guarantees, given an information access interface supported the differential privacy framework. Differential privacy needs that computations be insensitive to changes in any explicit individual's record, thereby limiting information leaks through the results. The privacy conserving interface ensures flatly safe access to the information and doesn't need from the information mineworker any experience in privacy. However, as we have a

tendency to show within the paper, a naive utilization of the interface to construct privacy conserving data processing algorithms may lead to inferior data processing results.

In this paper [4] **Fung, K. Wang, R. Chen, and P.S. Yu**. Data in its original type, however, usually contains sensitive data regarding people, and commercial enterprise such knowledge can violate individual privacy. These observe in knowledge commercial enterprise depends primarily on policies and pointers on what styles of knowledge are printed and on agreements on the employment of printed knowledge. This approach alone might cause excessive knowledge distortion or too little protection. Privacy-preserving knowledge commercial enterprise (PPDP) provides ways and tools for commercial enterprise helpful data whereas conserving knowledge privacy. Recently, PPDP has received tidy attention in analysis communities, and plenty of approaches are projected for various knowledge commercial enterprise eventualities.

3. PROPOSED SYSTEM

Differential privacy is a recent privacy definition that provides a strong privacy guarantee. It guarantees that an adversary learns nothing more about an individual, regardless of whether her record is present or absent in the data. A standard mechanism to achieve differential privacy is to add a random noise to the true output of a purpose. The noise is calibrated according to the sensitivity of the function. The sensitivity of a function is the maximum difference of its outputs from two data sets that differ only in one record.

Differential privacy is a strong privacy definition. A two-party protocol for the exponential mechanism. It is a sub protocol of main algorithm. It uses the exponential mechanism in a distributed setting. Two-party data publishing algorithm for vertically partitioned data that generate an integrated data table satisfying differential privacy.

3.1 Security Model

We present the security definition in the semi honest adversary model. Additionally, we introduce the required cryptographic primitives that are instrumented inside the proposed algorithm.

Many of the protocols, as in the case of the proposed algorithm in this paper, involve the composition of secure sub protocols in which all intermediate outputs from one sub protocol are inputs to the next sub protocol. These intermediate outputs are either simulated given the final output and the local input for each party or computed as random shares. Random shares are meaningless information by themselves. However, shares can be combined to reconstruct the result. Using the composition theorem, It can be shown that if each sub protocol is secure, then the resulting composition is also secure.

3.2 Two-Party Protocol for Exponential Mechanism

Privacy-preserving information publication addresses the matter of revealing sensitive information once mining for useful info. during this paper, address the matter of personal information publication, wherever totally different attributes for a similar set of people are command by 2 parties. Particularly, we tend to gift associate algorithmic program for differentially personal information unfairness for vertically-partitioned information between 2 parties within the semi-honest person model. to realize this, we tend to 1st gift a two-party protocol for the exponential mechanism. This protocol is used as a sub protocol by the other algorithmic program that needs the exponential mechanism in a very distributed setting. moreover, we tend to propose a two-party algorithmic program that releases differentially-private information in a very secure manner in keeping with the definition of secure multiparty computation.

The exponential mechanism chooses a candidate that's near optimum with relevancy a utility operate whereas protective differential privacy. within the distributed setting, the candidates are closely-held by 2 parties and, therefore, a secure mechanism is needed to reckon a similar output whereas making certain that no additional info is leaked to any party.

3.3 Two-Party Algorithm

We present our Distributed Differentially private anonymization algorithm based on Generalization (DistDiffGen) for two parties. The algorithm first generalizes the raw data and then adds noise to achieve differential privacy.

3.4 ADVANTAGES

These algorithms provide the security and the efficiency of the data access.

- It respect to a utility function while preserving differential privacy.
- A secure mechanism is required to compute the same output while ensuring that no extra information is leaked to any party.

4. IMPLEMENTATION RESULTS

To evaluate the impact on classification quality, we divide the data into training and testing sets. We observe two general trends from the experiments. First, the privacy budget has a direct impact on the classification accuracy.

A higher budget results in better accuracy because it ensures better attribute partitioning, and it lowers the magnitude of noise that is added to the count of each equivalence group. This observation also holds for DiffGen. Second, the classification accuracy is insensitive to the scaling (the number of the considered digits after the decimal points) for the Max function.

The value is large due to the score of the Max function, which is usually a large integer. Therefore, scaling has hardly any impact on the data utility for classification analysis.

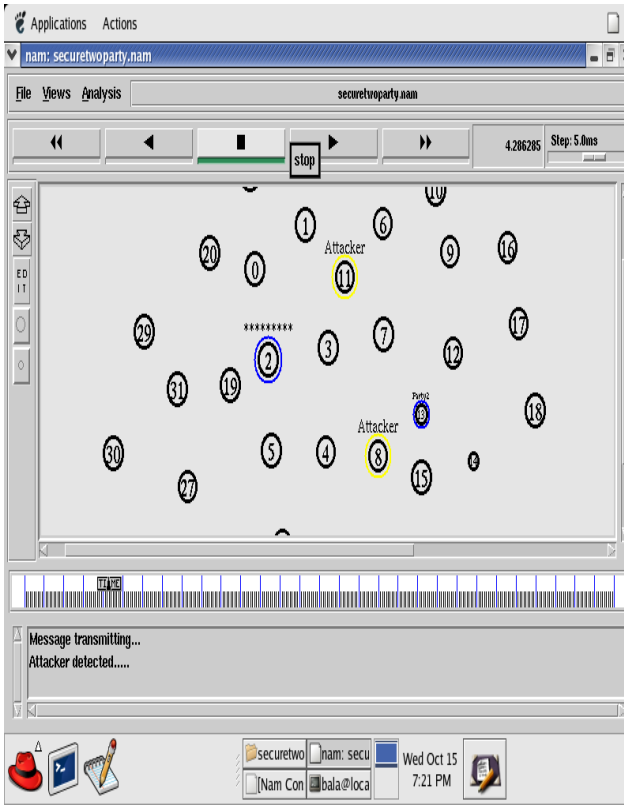

Figure 2: Detection of attackers

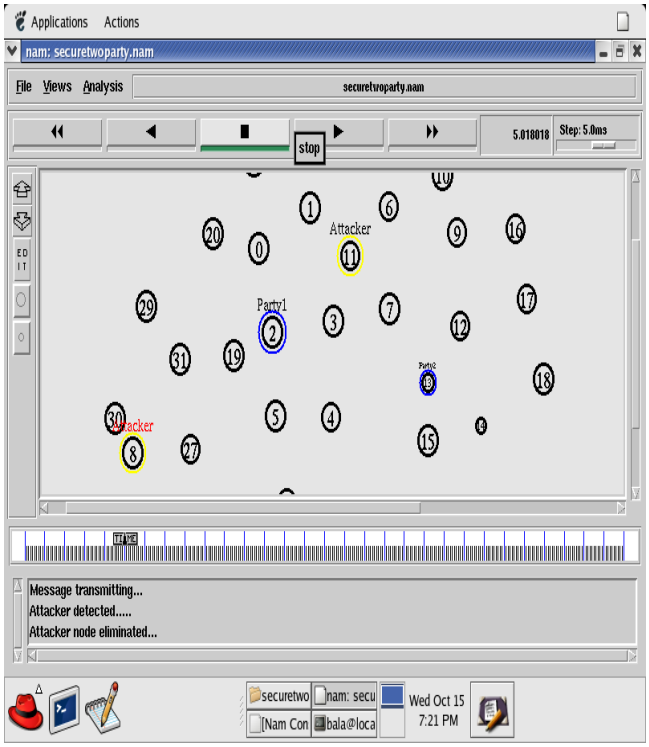

Figure 3: Eliminating the attacker's node

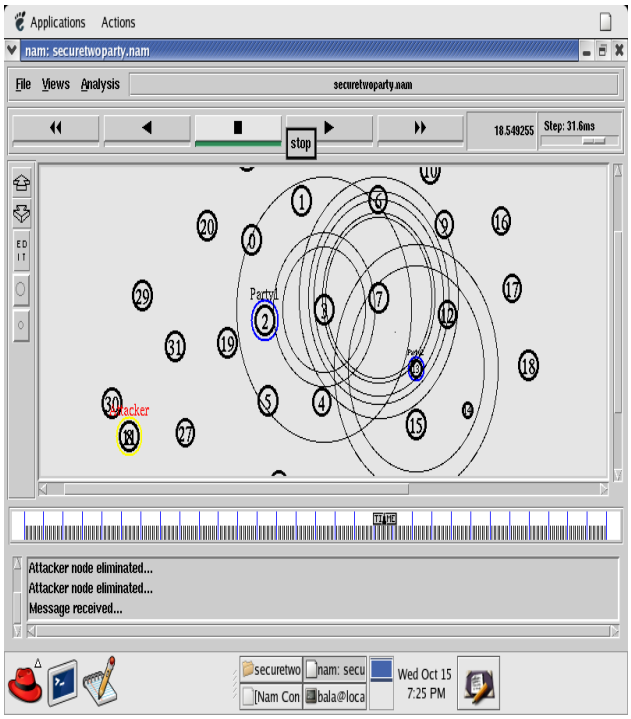

Figure 4: Secure Data Transmission

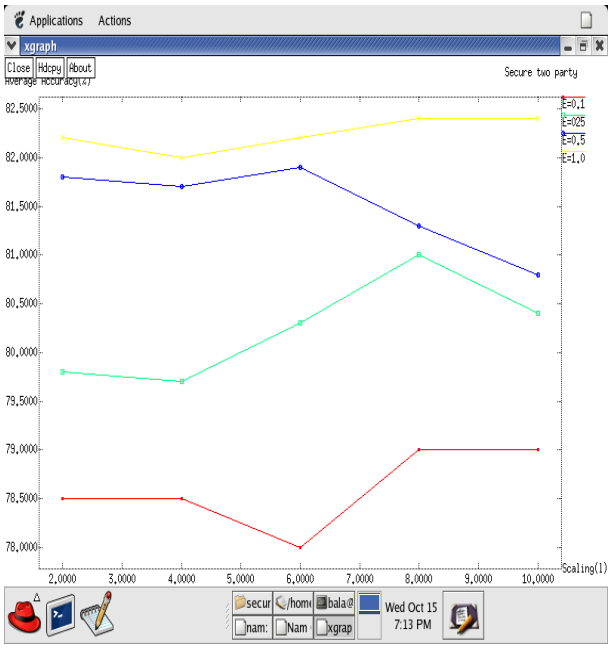

Figure 5: Classification accuracy

4. CONCLUSION AND FUTURE ENHANCEMENT

The two-party differentially private data release algorithm for vertically partitioned data. The proposed algorithm is differentially private and secure under the security definition of the semi honest adversary model. An experimentally evaluated the data utility for classification analysis. The proposed algorithm can effectively retain essential information for classification analysis. It provides similar data utility compared to the recently proposed single-party algorithm and better data utility than the distributed k-anonymity algorithm for classification analysis.

More than two parties: The proposed algorithm is only applicable for the two-party scenario because the distributed exponential algorithm and the other primitives are limited to a two-party scenario. The proposed algorithm can be extended for more than two parties by modifying all the sub protocols while keeping the general top-down structure of the algorithm.

Semi honest Adversary Model: This is the common security definition used in the SMC literature; it is realistic in our problem scenario because different organizations are collaborating to securely share their data for mutual benefits. Hence, it is reasonable to assume that parties will not deviate from the defined protocol. However, they may be curious to learn additional information from the messages they received during the protocol execution. To extend the algorithm for malicious parties, all sub protocols should be extended and must be secure under the malicious adversary model.

REFERENCES

[1] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta, "Discovering Frequent Patterns in Sensitive Data," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '10), 2010.

[2] K. Chaudhuri, C. Monteleoni, and A. Sarwate, "Differentially Private Empirical Risk Minimization," J. Machine Learning Research, vol. 12, pp. 1069-1109, July 2011.

[3] K. Chaudhuri, A.D. Sarwate, and K. Sinha, "Near-Optimal Differentially Private Principal Components," Proc. Conf. Neural Information Processing Systems, 2012.

[4] 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, June 2010.

[5] A. Frank and A. Asuncion, UCI Machine Learning Repository, http://mlearn.ics.uci.edu/MLRepository.html, 2010.[6] A. Inan, M. Kantarcioglu, G. Ghinita, and E. Bertino, "Private Record Matching Using Differential Privacy," Proc. Int'l Conf. Extending Database Technology (EDBT '10), 2010.

[7] A. Friedman and A. Schuster, "Data Mining with Differential Privacy," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '10), 2010.

[8] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M.Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 28-34, Dec. 2002.

[9] I. Dinur and K. Nissim, "Revealing Information while Preserving Privacy," Proc. ACM Symp. Principles of Database Systems (PODS '03), 2003.

[10] C. Dwork, "A Firm Foundation for Private Data Analysis," Comm. ACM, vol. 54, no. 1, pp. 86-95, 2011.