International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Privacy Preserving Utility Verification Security of Data Published by Non Interactive Differentially Private Mechanisms

Gouri Namdeo Kale[1], Dr. S. N. Kini[2]

[1]Department Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-28, Savitribai Phule Pune University, Pune, India

[2]Professor, Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-28, Savitribai Phule Pune University, Pune, India.

Abstract: ***Service providers have the ability to collect large amounts of user data. Sometimes, a set of providers may try to aggregate their data for specific data mining tasks. In this process, how to protect users' privacy is extremely critical. Number o f user write an own novel, own story and own dataset, every user want to keep this data safe on own site and with user internet facility user always search a digital publisher. This is the so-called privacy-preserving collaborative data publishing problem. In this paper, we consider the collaborative data publishing problem for anonymizing horizontally partitioned data at multiple data providers. We consider a new type of "insider attack" by colluding data providers who may use their own data records (a subset of the overall data) to infer the data records contributed by other data providers. The paper addresses this new threat, and makes several contributions.***

Keywords: privacy, security, frequent itemset mining

1. Introduction

In digital life every user are connected with internet and web related activity, Now very large user are want to deal with internet and internet data. Some thing now happen with the data writer they want to write data like novel,Story,etc and securely publish this data on Internet publishing site. Number of sites providing a data publishing feature, but now a day unethical activity are increase, so data privacy preserving is become a very important issue on every level. The data privacy is very big issue on publisher site because they want to create a trust model between writer and reader. In this trust some important point are consider that are data privacy, Data integrity, data security, SERVICE providers have the ability to collect large amounts of user data. Sometimes, a set of providers may try to aggregate their data for specific data mining tasks. For example, the hospitals nation-wide may outsource their medical records to a research group for mining the spreading patterns of influenza epidemics. In this process, how to protect users' privacy is extremely critical. This is the so-called privacy-preserving collaborative data publishing problem. A lot of privacy models and corresponding anonymization mechanisms have been proposed in the literature such as *k*-anonymity and differential privacy. *k*-anonymity and its variants protect privacy by generalizing the records such that they can not be distinguished from some other records. Differential privacy is a much more rigorous privacy model. It requires that the released data is insensitive to the addition or removal of a single record. To implement this model, the corresponding anonymization mechanisms usually have to add noise to the published data, or probabilistically generalize the raw data. Obviously, all these *data anonymization mechanisms have serious side effects on the data utility*. As a result, the users of the published data usually have a strong demand to verify the real utility of the anonymized data.

2. Related Work

For a discussion of the guarantees provided by differential privacy and their limitations, see [Kasiviswanathan and Smith 2008; Kifer and Machanavajjhala 2011].As the theoretical foundations of differential privacy become better understood, there is momentum to prove privacy guarantees of real systems.

Several authors have recently proposed methods for reasoning about differential privacy on the basis of different languages and models of computation, e.g. SQL-like languages [McSherry 2009], higher-order functional languages [Reed and Pierce 2010], imperative languages [Chaudhuri et al. 2011], theMapReducemodel [Roy et al. 2010],and I/O automata [Tschantz et al. 2011]. The unifying basis of these approaches are two key results: The first is the observation that one can achieve privacy by perturbing the output of a deterministic program by a suitable amount of symmetrically distributed noise, giving rise to the so-called Laplacian [Dwork et al. 2006b] and Exponentialmechanisms [McSherry and Talwar 2007]. The second result is theorems that establish privacy bounds for the sequential and parallel composition offerentially private programs, see e.g. [McSherry 2009].

In combination, both results form the basis for creating and analyzing programs by composing differentially private building blocks. While approaches relying on composing building blocks apply to an interesting range of examples, they fall short of covering the expanding frontiers of differentially private mechanisms and algorithms.

Examples that cannot be handled by previous approaches include mechanisms that aim for weaker guarantees, such as approximate differential privacy [Dwork et al. 2006a], or randomized algorithms that achieve differential privacy

without using any standard mechanism [Gupta et al. 2010]. Dealing with such examples requires fine-grained reasoning about the complex mathematical and probabilistic computations that programs perform on private input data.

Such reasoning is particularly intricate and error-prone, and calls for principled approaches and tool support. In this article we present a novel framework for formal reasoning about a large class of quantitative confidentiality properties, including (approximate) differential privacy and probabilistic non-interference.

3. Proposed Work

Apriori Algorithm for Data Processing:-

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

RSA algorithm for two level Encryption and Decryption:-
RSA is the algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two different keys. This is also called public key cryptography, because one of them can be given to everyone. The other key must be kept private.

RSA involves a public key and private key. The public key can be known to everyone, it is used to encrypt messages. Messages encrypted using the public key can only be decrypted with the private key.

Encryption:
Sender A does the following:-

- Obtains the recipient B's public key (n, e).
- Represents the plaintext message as a positive integer *m*, 1 < m < n [see note 4].
- Computes the cipher text $c = m^e \bmod n$.
- Sends the ciphertext *c* to B.

Decryption:
Recipient B does the following:-

- Uses his private key (n, d) to compute $m = c^d \bmod n$.
- Extracts the plaintext from the message representative *m*.

Advantages:

- It works as a fast verification algorithm
- Data is secure on publisher site.
- Provide Integrity, Security, Privacy
- Constraints remove all duplicate records information.
- Compare to previous approach its gives the better utility and efficiency results.

4. Architectural View

The architecture diagram of the system shown below helps us to understand the system.

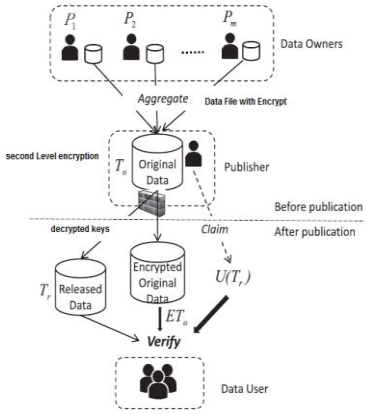

Figure 1: System Architecture

Sr No.	Paper	Technique	Author	Results
1	Differentially Private Data Release through Multidimensional Partitioning	Private Histogram	Yonghui Xiao	multidimensional partitioning algorithms for differentially private histogram release based on an interactive Differential privacy mechanism.
2	On the Complexity of Differentially Private Data Release	Encrypt algorithm	Cynthia Dwork	Data Privacy
3	Secure distributed framework for achieving -differential privacy	Algorithm for processing a new event at the Graph-based Inference Engine	D. Alhadidi, N. Mohammed	Privacy Enhance
4	Differential privacy and their limitations	foundations of differential privacy	Kifer and Machanavajjhala	there is momentum to prove privacy guarantees of real systems
5	The Algorithmic Foundations of Differential Privacy	this monograph is devoted to fundamental techniques for achieving differential privacy	Aaron Roth	A key point is that, by rethinking the goal, one can often obtain far better results

5. Conclusion

We develop new strategic module for data privacy for data on non-publishing sites, this project provide a very secure. Communication trust between Reader, Publisher and writer. We all know that is one prime level. This system provides a very reliable and easy way to protect data from unethical activity. Privacy maintains one prime level.

With user this system user fully aware of data security, privacy and data redundancy. So this system are fully satisfied our objective.

In future work we want to implement same system on multimedia content and data. There is big scope for use same architecture on multimedia content.

References

[1] L. Fan, L. Xiong, and V. Sunderam, FAST: Differentially private real-time aggregate monitor with filtering and adaptive sampling, in Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD), 2013, pp. 10651068
[2] R. Chen, B. C. M. Fung, and B. C. Desai. (2011). "Differentially private trajectory data publication." [Online]. Available: http://arxiv.org/abs/1112.2020.
[3] D. M. Freeman, Converting pairing-based cryptosystems from composite order groups to prime-order groups, in Proc. 29th Annu. Int. Conf. Theory Appl. Cryptogr. Techn. (EUROCRYPT), 2010, pp. 4461.
[4] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, Privacy-preserving data publishing: A survey of recent developments, ACM Compute. Surv., vol. 42, no. 4, 2010, Art. no. 14.
[5] Y. Hong, J. Vaidya, H. Lu, P. Karras, and S. Goel, Collaborative search log sanitization: Toward differential privacy and boosted utility, IEEE Trans. Dependable Secure Comput., vol. 12, no. 5, pp. 504518, Sep./Oct. 2015.
[6] W. Jiang and C. Clifton, A secure distributed framework for achieving k-anonymity, Int. J. Very Large Data Bases, vol. 15, no. 4, pp. 316333, Nov. 2006.
[7] J. Lee and C. Clifton, How much is enough? Choosing for differential privacy, in Proc. 14th Int. Conf. Inf., 2011, pp. 325340.
[8] N. Li, T. Li, and S. Venkatasubramanian, t-closeness: Privacy beyond k-anonymity and l-diversity, in Proc. 23rd Int. Conf. Data Eng. (ICDE), Apr. 2007, pp. 106115.
[9] J. Liu and K. Wang, Enforcing vocabulary k-anonymity by semantic similarity based clustering, in Proc. 10th Int. Conf. Data Mining (ICDM),Dec. 2010, pp. 899904.
[10] X. Zhang, X. Meng, and R. Chen, "Differentially private set-valued data release against incremental updates" in *Proc. 18th Int. Conf. Database Syst. Adv. Appl.*, 2013, pp. 392–406.

Author Profile

Miss. Gouri Namdeo Kale is currently pursuing M.E (Computer) from Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007. She received her B.E (IT) Degree from SVERI's college of engineering Pandharpur. Maharashtra, India. Her area of interest is Information Security

Dr. Prof. S. N.Kini. is currently working as Ass. Prof. (Computer) at Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India - 411007.