International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Privacy Preserving Utility Verification of Data Published

Madhushree M

CSE, AMC Engineering College/ VTU, India

Abstract: ***With modern collaborative data publishing techniques, the problem is that a central data publisher is liable for aggregating sensitive data from multiple parties then anonymizing it before publishing for data processing. In such scenarios, the user demands to know the utility of their published data since most anonymization techniques have side effects on data utility. Moreover, a corrupt data publisher is capable of misusing the collected data for their gains. We could call this an "insider attack". In this paper, we address this problem and briefly discuss a few proposed solutions.***

Keywords: Data Privacy, Encryption, Frequency, Security, Utility Verification

1. Introduction

It is evident that, at present, people are spending a considerable amount of time on the Internet, and this, in turn, results in a glut of data being shared over the internet. Although this use of the Internet has greatly increased the level of communication available, it has also had detrimental effects on the privacy of the data owners.

The central data aggregator is liable for the collection, maintenance, preserving privacy, preserving anonymization, and sharing of data and the emergence of new cloud computing technology allows the easier exchange of information for mutual benefits but also at the same time has also resulted in the rise of unethical activities related to misuse of data. For example, Facebook data leak scandal in 2018 about 87 million Facebook users' data were collected by a cloud - based Facebook quiz app and then paired with information taken from their social media profile including their gender, age, relationship status, and location. This same set of data could be repurposed in different ways to infer certain sensitive personal information about people causing an uproar about the importance of privacy preservation techniques. In such cases, 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 are proposed within the literature like k - anonymity and differential privacy. k - anonymity and its variants (e. g., l - diversity and t - closeness) protect privacy by generalizing the records such they will not be distinguished from another record. 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.

2. Related Work

Designing such a privacy - preserving data releasing structure is rather challenging. To date, a few related approaches related to k - anonymity, l - diversity and t - closeness have been proposed [3], [4], [5], [6], [7], [8], [9], [10], [11]. However, these approaches cannot fulfill all the privacy requirements needed in the cloud - based data - driven application scenario because such models cannot handle the curse of dimensionality. Dimensionality reduction - based approaches [8], [9], [10], [11] have been proposed to preserve privacy while maintaining most of the utility. However, despite their good experimental performance on several public data sets, those approaches didn't introduce any uncertainty to hide the sensitive information, which failed to show the needed guarantees on the privacy targets mathematically.

3. Proposed System

A privacy - preserving utility verification mechanism that works as a two parts system, a differentially private anonymization algorithm (DiffPart) designed for set - valued data is proposed. This adds anomaly to the frequencies of the records supported a context - free taxonomy tree and no items within the original data are generalized. This proposal solves the challenge to verify the utility of the published data supported by the encrypted frequencies of the data records rather than their plain values. As a result, it can protect the data from the verifying parties because they can't learn whether or what percentage times a specific record appears within the raw data - set without knowing its real frequency. In addition, since the encrypted frequencies are provided by the publisher, a scheme for the verifying parties to incrementally verify its correctness is presented. Then the above mechanism is extended to the second part, differential generalization (DiffGen), which refers to a deferentially private anonymization algorithm designed for relational data. Different from the former part, the latter may generalize the attribute values before confounding the frequency of every record. Information losses are caused by both the generalization and therefore the disturbance. These two kinds of information losses are measured independently of each other making use of the same utility metrics. We take both into consideration. This analysis shows that the utility verification for generalization operations is often administered with only the published data. As a result, this verification doesn't need any protection. The utility metric for the disturbance is analogous thereupon for DiffPart. We thus adopt the proposed privacy - preserving mechanism to this verification. A series of experiments are conducted on real - world relational data to evaluate the efficiency of the proposed mechanisms. The results show that these mechanisms are efficient enough as long as both publishing and utility verification of information is carried out offline.

RSA algorithm for two - level Encryption and Decryption: [2]

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 a private key. The public key is often known to everyone, it's utilized to encrypt messages. Messages encrypted using the public key can only be decrypted with the private key.

4. Figures and Tables

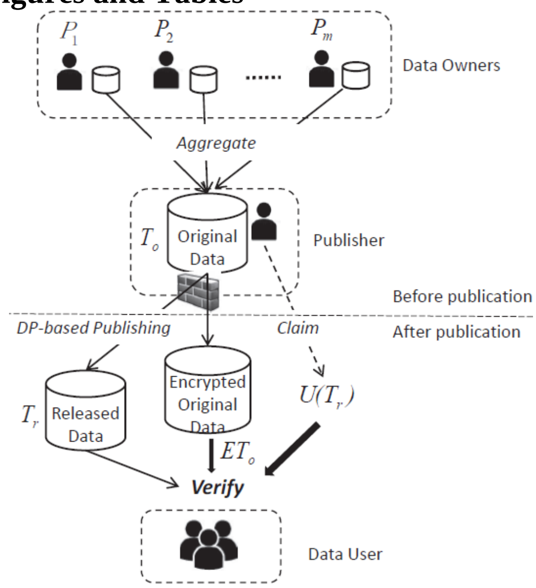

Figure 1: System Architecture

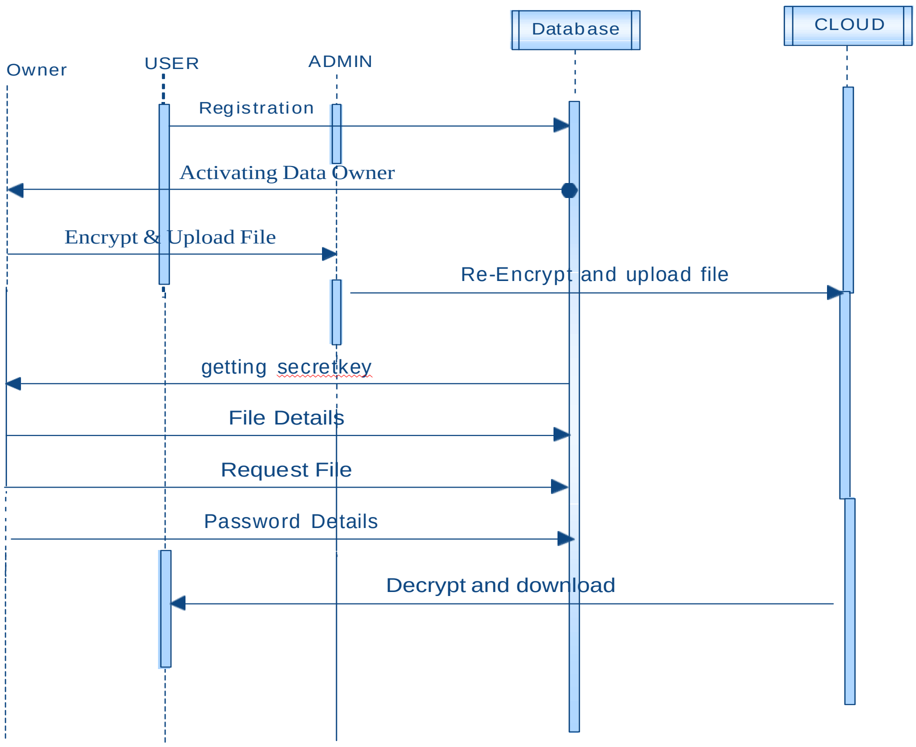

Figure 2: Sequence Diagram

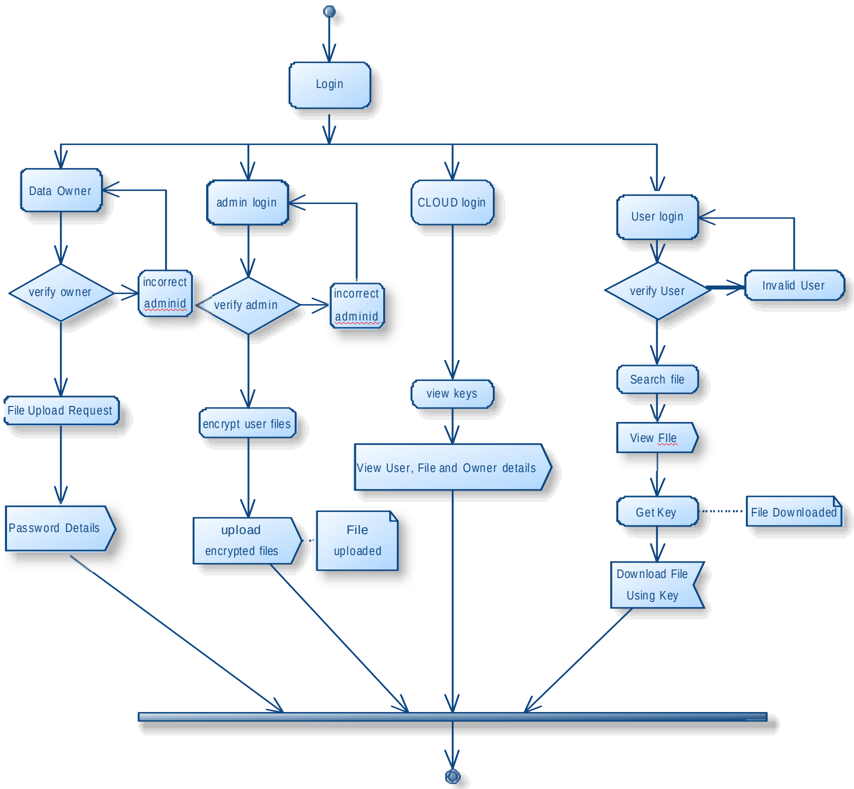

Figure 3: Activity Diagram

5. Conclusion

In conclusion, we develop a new strategic module for data privacy for data on non - publishing sites, this project provides high - level security. Preserves the communication trust between reader, publisher, and writer. With this technique users are fully conscious of data security, privacy, and data redundancy. So this system fully satisfied our objective. In future work, we would want to implement the same system on multimedia content and data. There is big scope for the use of similar architecture on multimedia content.

References

[1] Di Zhuang, J. Morris Chang, "Utility - aware Privacy - preserving Data Releasing" *2020 IEEE*

[2] Gouri Namdeo Kale1, Dr. S. N. Kini, "Privacy Preserving Utility Verification Security of Data Published by Non Interactive Differentially Private Mechanisms" *IJSR 2015*

[3] L. Sweeney, "k - anonymity: A model for protecting privacy, "*International Journal of Uncertainty, Fuzziness and Knowledge - Based Systems, vol.10, no.05*, pp.557–570, 200

[4] J. Kim and W. Winkler, "Multiplicative noise for masking continuous data, " Statistics, p.01, 2003.

[5] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, "Functional mechanism: regression analysis under differential privacy, " *Proceedings of the VLDB Endowment, vol.5, no.11*, pp.1364–1375, 2012.

[6] R. Shokri and V. Shmatikov, "Privacy - preserving deep learning, " *in Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. ACM, 2015*, pp.1310–1321.

[7] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy, " *in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016*, pp.308–318.

[8] S. - Y. Kung, "Discriminant component analysis for privacy protection and visualization of big data, " *Multimedia Tools and Applications*, pp.1–36, 2015.

[9] K. Diamantaras and S. - Y. Kung, "Data privacy protection by kernel sub - space projection and generalized eigenvalue decomposition, " *in Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on. IEEE, 2016*, pp.1–6.

[10] S. - Y. Kung, "Compressive privacy: From information/estimation theory to machine learning [lecture notes], " *IEEE Signal Processing Magazine, vol.34, no.1*, pp.94–112, 2017.

[11] D. Zhuang, S. Wang, and J. M. Chang, "Fripal: Face recognition in privacy abstraction layer, " *in Dependable and Secure Computing, 2017 IEEE Conference on. IEEE, 2017*, pp.441–448.

[12] M. Al, S. Wan, and S. - Y. Kung, "Ratio utility and cost analysis for privacy preserving subspace projection, " *arXiv preprint arXiv: 1702.07976, 2017.*

[13] 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

[14] R. Chen, B. C. M. Fung, and B. C. Desai. (2011). "Differentially private trajectory data publication. " [Online]. *Available: http: //arxiv. org/abs/1112.2020.*

[15] 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.
[16] 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.*
[17] 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.
[18] W. Jiang and C. Clifton, A secure distributed framework for achieving k - anonymity, *Int. J. Very Large Data Bases, vol.15, no.4, Nov.2006*, pp.316333
[19] J. Lee and C. Clifton, How much is enough? Choosing for differential privacy, in Proc.14th Int. Conf. Inf., 2011, pp.325340.
[20] 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.
[21] 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.
[22] 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.