RESEARCH ARTICLE OPEN ACCESS

PRIVACY PRESERVING FOR ONLINE USER BEHAVIOUR DATA

Dr. K. Ravi Kumar [1], K. Saranya [2]
1(Assistant Professor, Department of Computer Science, Tamil University, Thanjavur, Tamilnadu, India)
2 (Mphil (Research Scholar), Tamil University, Thanjavur, Tamilnadu, India)

Abstract:

Online users' privacy is thus under the risk of being exposed to third parties. The last decade has spectator a body of research works trying to perform data aggregation in a privacy-preserving way. Most of existing methods guarantee strong privacy protection yet at the cost of very insufficient aggregation operations, such as allowing only summation, which hardly satisfies the need of behavior analysis. We propose a scheme PPSA, which encrypts users' sensitive data to prevent privacy exposure from both outside analysts and the aggregation service provider, and fully supports selective aggregate functions for online user behavior analysis while guaranteeing the exponent privacy. We have implemented our method and evaluated its performance using a trace-driven evaluation based on a real online behavior dataset. Experiment results show that our scheme expertly supports both overall aggregate queries and various selective aggregate queries with acceptable computation and communication overheads.

Keywords — **fraudulent, PPSA, Web services, Trace Driven, Selective Aggregation.**

I. INTRODUCTION

Online user behaviour analysis studies how and why users of e-commerce platforms and web applications behave. It is widely applied in the real world, especially in commercial environments, political campaigns, and web application developments. Data aggregation is one of the most critical operations in behaviour analysis. Nowadays, the aggregation tasks for user data are outsourced to third-party data aggregators including Google Analytics, comScore, Quant cast, and Stat Counter. While this tracking scheme brings great benefits to analysts and aggregators, it also raises serious concerns on disclosure of users' privacy. Aggregators hold detailed data of users' online behaviours, from which demographics can be easily inferred. To protect users' privacy, government and industry regulations were established, e.g., W3C Do-Not-Track, which significantly restricts the analysis of users' online behaviours. To address the conflict between the utility of analysis results and users' privacy, much effort has been devoted to designing protocols that allow operations on user Data while still protecting users' privacy. Unfortunately, existing schemes guarantee strong privacy at the expense of limitations on analysis.
Most of them can only enumerate summation and mean of data over all users without filter or selection, i.e., overall aggregation. Some previous methods allow more complex computations, such as polynomial evaluation, yet still do not support selection. However, selective aggregation is one of the most essential operations for queries on databases. For example, "select avg (income) from database group by gender". It can be used to tell the difference of different user groups in a certain respect.

II. LITERATURE REVIEW

We present the first scheme PPSA that allows privacy preserving selective aggregation on user data, which plays a critical role in online user behaviour analysis. The key to achieve selective aggregation is counting in data items of target users by multiplying them by 1 and skipping the rest by multiplying them by 0. These calculations are done

in cipher text, and thus no privacy disclosure would occur. We combine homomorphism encryption and differential privacy apparatus to protect users' sensitive information from both analysts and aggregation service providers, and protect individuals' privacy from being inferred. We prove that differential privacy can be achieved by adding integer noise generated from the geometric distribution. To our knowledge, we are the first to utilize integer noise to achieve differential privacy, which meets the requirement of homomorphism cryptosystems. We extend PPSA to support aggregation selected by multiple Boolean attributes. We implement PPSA and do a trace-driven evaluation based on an online behaviour dataset. Evaluation results show that our scheme completely supports various selective aggregate queries with high accuracy and acceptable computation and communication overheads.

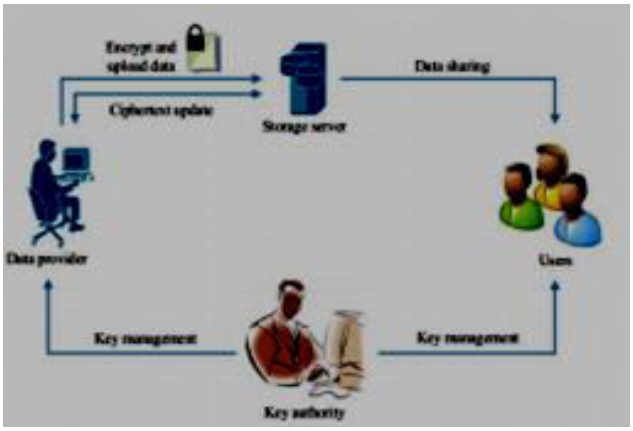
III. IMPLEMENTATION

Existing Process

Privacy-preserving aggregation on sensitive user data has raised much attention recently, including health care, time-series data, wireless sensor network data, and online behavior data for analysis an advertising. In general, there are two types of systems in previous work. Centralized Systems: - In a centralized system, all the user data are stored on the server. It is important that users encrypt or encode their data before sending them to the server. The server holds the encrypted data, but it can only compute answers to queries obliviously. Distributed Systems: - In a distributed system, clients need to proactively, or passively send required data to the aggregator in a private way. But both rely on the participation of clients. These systems all require online users, so analysis cannot go on when most of the users are offline.

Proposed System

In the proposed system, the system has described the challenges of making online user data aggregation while preserving users' privacy. Built on BGN homomorphic cryptosystem, we have designed the first system that is able to securely and selectively aggregate user data, making it practical in realistic data analytics. It guarantees strong powerful privacy preservation by utilizing differential privacy apparatus to protect individuals' privacy. The process has presented PPSA to evaluate aggregation selected by one Boolean attribute, and extended it to aggregation selected by multiple Boolean attributes and by one numeric attribute. Extensive inquiry have shown that PPSA supports various selective aggregate queries with acceptable overhead and high accuracy.



System Construction Module

In the first module, we develop the proposed system with the required entities for the evaluation of the proposed model. The data provider (e.g., David) first resolve the users (e.g., Alice and Bob) who can share the data. Then, David encrypts the data under the identities Alice and Bob, and uploads the cipher text of the shared data to the cloud server. Further Alice or Bob wants to get the shared data, she or he can download and decrypt the corresponding cipher text. However, for an unauthorized user and the cloud server, to access the plaintext of the shared data is not available.

Data Provider

In this module, we develop the Data Provider module. The data provider module is developed such that the new users will Signup initially and then Login for authentication. The data provider module provides the option of uploading the file to the Cloud Server. The process of File Uploading to the cloud Server is undergone with Identity based encryption format. Data Provider will check the progress status of the file upload by him/her. Data Provider provided with the features of Revocation and Cipher text update the file. Once after completion of the process, the Data Provider logouts the session.

Cloud User

In this module, we develop the Cloud User module. The Cloud user module is developed such that the new users will Signup initially and then Login for authentication. The Cloud user is provided with the option of file search. Then cloud user feature is added up for send the Request to Auditor for the File access. After getting decrypt key from the Auditor, he/she can access to the File. The cloud user is also enabled to download the File. After completion of the process, the user logout the session.

Key Authority (Auditor)

Auditor Will Login on the Auditor's page. He/she will check the pending requests of any of the above person. After accepting the request from the above person, he/she will generate master key for encrypt and Secret key for decrypt. After the complete process, the Auditor logout the session.

IV. CONCLUSIONS

We have described the provocation of making online user data aggregation while preserving users' privacy. Based on BGN homomorphism cryptosystem, we have draft the first system that is able to securely and selectively aggregate user data, making it practical in realistic data analytics. It guarantees strong privacy preservation by utilizing differential privacy mechanism protect individuals' privacy. We have presented PPSA to assess aggregation selected by one Boolean attribute, and enhance it to aggregation selected by multiple Boolean attributes and by one numeric attribute. Huge experiments have shown that PPSA supports various selective aggregate queries with acceptable overhead and high accuracy. In the future, we plan to study networks with multiple mobile sinks and adjustable trajectories. Furthermore, a network with fewer assumptions should be constructed and the communication delay between sensor nodes should be taken into consideration too.

REFERENCES

1. *B.S. Bhati, P.VenkataramPerformance analysis of privacy protection system during data transfer in manets Int. J. Wirel. Inf. Netw., 25 (2018), pp. 30-43,*
2. *J.Estrada-Jimenez,J.Parra-Arnau,A.Rodriguez-Hoyos, J. Forne, "Online advertising: Analysis of privacy threats and protection approaches", Computer Communications, 2017.*
3. *M. Nateghizad, Z. Erkin, R. L. Lagendijk, "An efficient privacy-preserving comparison protocol in smart metering systems", EURASIPJ. on Info. Security, 2016.*
4. *T. Jung, X.-Y. Li, and M. Wan, "Collusion tolerable privacy preserving sum and product calculation without secure channel," IEEE Transactions on Dependable and Secure Computing (TDSC), vol. 12, no. 1, pp. 45– 57, 2015.*
5. *D. Fiore, R. Gennaro, and V. Pastro, "Efficiently verifiable computation on encrypted data," in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS), 2014, pp. 844–855.*
6. *T. Jung, X. Mao, X.-y. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation," in Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM), 2013, pp. 2634–2642.*
7. *R. Chen, A. Reznichenko, P. Francis, and J. Gehrke, "Towards statistical queries over distributed private user data," in Proceedings of the 9th Symposium on Networked Systems Design and Implementation (NSDI), 2012.*
8. *M. Hardt and S. Nath, "Privacy-aware personalization for mobile advertising," in Proceedings of the ACM conference on Computer and Communications Security (CCS), 2012, pp. 662–673.*
9. *E. Shi, T.-H. H. Chan, E. G. Rieffel, R. Chow, and D. Song, "Privacypreserving aggregation of time-series data," in Proceedings of the Network and Distributed System Security Symposium (NDSS), 2011.*
10. *C. Dwork, "Differential privacy: A survey of results," in Proceedings of 5th International Conference on Theory and Applications of Models of Computation (TAMC), 2008, pp. 1-19*
11. *Jans M, van der Werf JM, Lybaert N, and Vanhoof K (2011) A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications 38, 13351-9.*
12. *Kirkos E, Spathis C, and Manolopoulos Y (2007) Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications 32, 995-1003.*
13. *Koh HC and Low CK (2004) Going concern prediction using data mining techniques. Managerial Auditing Journal 19, 462-76.*