ONLINE USER'S BEHAVIOR DATA PRESERVING USING RSA BASED SELECTIVE AGGREGATION

{ [1] M.Mahita,[2] J.Nandhini Devi,[3] M.Nivedhitha*

[1],[2],[3] Student Department of Computer Science and Engineering
M.Kumarasamy College Of Engineering, Karur

ABSTRACT

In this Paper, The people privacy are ensured by public key cryptography system by not letting the Data Aggregators to identify about the searches that we do in search Engines. By using RSA Algorithm, the Encryption is made such that the analysts can't know regarding your interest and likes. It involves a concept of public and private key cryptography on a client server basis. The padding up of bits is also done to ensure the security at the transport layer level in searching. The reliability and safety of the privacy is high on using RSA algorithm for Encryption. The Experimental result says that the privacy scheme effectively supports different selective aggregation queries.

Index Terms – Privacy preserving Selective aggregation, Differential privacy, RSA algorithm

I.INTRODUCTION

The Network security is the domain of our project and it provides security to contact with data files and file directories in a computer networking. An example of network security is an anti-virus system. The user is assigned by ID and password that allows the users to access the desired information and the programs within their authorized access. It protects the systems in the network from unwanted intruders. The Data aggregation is most significant operation in behaviour analysis. Aggregators seize detailed data of user's online behaviours and produce demographics. The Existing scheme security provides the strong privacy at the outlay of using data aggregation concept. To overcome those problems the noise is added to the aggregate result to achieve stronger privacy preservation.

The applications of this domain comprise of authentication Applications, Web Attacks, Email Security and IP Security.

WEB ATTACKS:

The Web security threats can be divided into two major types such as System and Web security. The Web Attacks are of two types such as

- Passive Attack
- Active Attack

The Passive attacks are the attack which access from the network traffic between the browser and the server. These attack also access the restricted information on a websites.

The active attacks are the attacks in which the another user alters the messages and it also alters the information from the website.

WEB APPLICATION SECURITY:

The web application security is a subdivision of information privacy that can be deals with security of web pages. This measure is taken to improve the security of an application. The various measures include finding, fixing and preventing security vulnerabilities. These procedures can be done at different phases such as design, development, upgrade and maintenance.

Kerberos – A faithful third party authentication protocol considered for TCP/IP networks. It acts as a trusted arbitrator in networks. It allows clients to access different

entities on the network. It keeps a database of clients and their secret keys at a high level.

EMAIL SECURITY:

The e-mail security means that the collective measures used here is to secure the access and text of an email account. It allows an individual or group of organizations to protect from the overall access to one or more email addresses and by accounts.

An email service provider implements by the email to be secured by the subscriber and email accounts by the data from the hackers. It is used in the banking sectors.

In this Research paper the application is mainly focus on Web Attacks.

There are mainly three challenges in the existing system:

First, the entrusted aggregator and the authority requirements to calculate the selective aggregation. The intermediatory cannot right of entry user data for privacy concerns.

Second, The privacy preserving selective aggregation requests to be achieved through the differential privacy in RSA algorithm. To safe guard those individual's privacy, we have to to add noise to aggregator. The existing differential privacy method generate noise to real numbers, but RSA algorithm requires the plain texts.

Third, The Selective aggregation have to be resistant in the circumstances where the clients can control between the online and offline repeatedly.

To address these challenges, we combine RSA based selective aggregation and differential privacy method to protect users perceptive information from both the analysts and aggregator. It protects the individuals privacy through RSA algorithm.

[1] The PPSA encoding scheme is used to encode the privacy and secure it from aggregators and analysts guaranteeing differential privacy, it's the combination of homomorphic and differential privacy.

[2] The existing first differentially private aggregation algorithm called as "**Fourier Perturbation algorithm**" is used. The major advantage of this algorithm is that with this we can scale with large number of users with an efficiency of O(n) for n queries. The disadvantage is that there is a absence of center server which actually increases the retrieval time of any particular data. To overcome this **Distributed Laplace Algorithm** is used to add noise. The major advantage is that it can scale large number of users and the computational power is also reduced from O(N) - O(1), where n is the total number of users. The disadvantage is that it is quite inaccurate for the data sets with high number of parameters.

[3] "Randomized Aggregatable privacy preserving Ordinal Response" merely called as **RAPPOR** technology . It is used for crowdsourcing from the end users. It provides Rapper algorithm for preserving data. This method provides high utility analysis in data collection. The advantage is that it uses differencial privacy and the utility guarantees for analyzing the original data. Using this procedure, each responder contains a strong ability for any "Yes" answers and also provides ability for the "Yes" or "No" answers.

[4] The **"clickstream"** algorithm is mainly used to record the internet usage through the web servers and the third party services. It improves the measurement of sizes and characteristics of media. It explains about the detailed description of electronic trace by clickstream records. The disadvantage is that the website is affected during investigation. The paid search be the service which offers with the online search engines in which the supporter picks the definite keywords. It acts as the main source for online publicity of growth for current years. In addition to that search engine interchange starts from a intentional activity than forced exposure. The another important advertising tool is email. The problems involved here is customization of online content through the click stream data analysis.

[5] GraphChi, a method for compute the graphs with millions of edges. By using this method, it breaks the large graphs into smaller parts. The GraphChi is capable to implement the numerous complex areas in data withdrawal, graph mining, and on the machine learn algorithms by using just a single consumer level computer. The advantage in using this algorithm is that disk based computation provides probabilistic graphic module and collaborative filtering where it recommends the goods

based on purchase of other user with parallel interests. The proposed system provides a new method called **Parallel Sliding Windows** (PSW) which exploits characteristics of sparse graphs which giving out from the disk analysis. PSW requires undersized number of sequential disk block transfers where it allows to complete on the areas of SSDs and from the habitual hard disks.

[6] The privacy is maintained through the single, propose storing user data in client computers and it stores in the cloud. It has the specific apps to use specific approved user's data. **Distributed differential privacy** is used to enable the analysis, but in the previous proposals the scale rate is hectic. Thus it faces the problems by placing tight bounds on the extent to which malicious activities can easily be distorted to the corresponding answers among the clients.

[7] This paper presents the technique for **Wireless Sensor Network** (WSN) for privacy preserving that have been categorized in terms of data oriented and through the context oriented. This application is mainly used in military, environmental purpose, health, home and in commercial applications. The privacy in WSN is classified into two forms called data privacy and in context privacy. The data privacy uses the data aggregation and data queries whereas in context privacy it supports location privacy and temporal privacy. The location security is then divided into two different types such as data source and base stations. The data aggregation provides the aggregate for **Cluster-based Private Data Aggregation (CPDA),** since it is the best way for reducing computational overheads.

[8] This paper mainly focus on the Public-Key models and the Common Reference String (CRS) model. This models brings the task of secure multiparty computations with n number of parties. It ensures the black box protocol, special purpose zero knowledge protocol, soundness of the stand alone protocol. The $IdealZK^{1:M}$ functionality is parameterized by relation of NP and RL which is the one to many extension protocols is used for securing the data to be private by encrypting the keys. This hybrid model run the recreation confirmation for all executions of cZKi except for the selected execution.

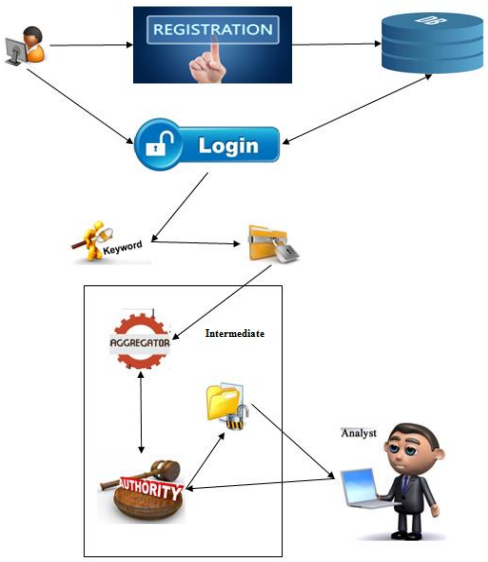
[9] This paper includes recent advances in techniques that combine and analyze data collected from multiple partners. Also, it led to many new promising distributed collaborative applications. Such collaborative computations could occur between trusted partners, between partially trusted partners, or between competitors. Here preserving privacy is an important issue in this context. This paper presents a distributed protocol for privacy-preserving aggregation with 68W15 Distributed algorihms to enable computing a class of aggregation functions that can be expressed as Abelian group. The proposed protocol is based on an overlay structure that enables secret sharing without the need of any central authority or heavyweight cryptography**.**

[10] SplitX is used for preserving online user privacy data and it allows aggregate queries for private data. This paper mainly stores the user data and it makes query the data through the differentially private produces the noisy result with the exception of affecting the individual user's data. It performs high performance analysis for building differentially private query larger than the distributed data. This idea presents the design implementation of SplitX and it examine the security and the performance level.

[11] Differential privacy preserves user's privacy in selective aggregation and it generally persistent on selective aggregation, differential privacy, homomorphic encryption, and in laplacian noise. It reduces the number of records by applying filters on user data values. The differential privacy intend to exploit the correctness of queries from the analytical databases and it reduce the probability of identify its record.

[12] The practical techniques are based on the hardware based and software based approaches over the encrypted data. It provides the code footprint in the trusted environment. It designs an efficient searchable encryption scheme with the queried ranges. This scheme has logarithmic complexity in the size of index and searches are performed in a nanoseconds. It is best known for the searchable encryption schemes. It provides an implementation of the performance scale to reduce the TrustedComputingBase.

SYSTEM ARCHITECTURE



The user has to register with the username and password. These data will be saved in the database. This database will be maintained by an admin. The admin uploads the products. Once the account is created, the user can login and search for keywords. The user types a keyword in the search bar and click Submit button. Once the submit button is clicked, the keyword will be encrypted using Homomorphic RSA algorithm and stored with the aggregator. The aggregator and the authority form an intermediate. The analyst will request the authority for a count on a product. The authority and aggregator combines to decrypt the keyword. The decrypted keyword will be stored with the authority. The authority will return with a count to the analyst as a response for the analyst's query.

SUMMARY

[1] The PPSA encoding scheme is used to encode the privacy and keeps secure from the aggregators and analysts guaranteeing differential privacy, it's the combination of homomorphic and privacy. [2] Differentially private aggregation algorithm is mainly focused to scale with large number of users with a efficiency of O[n] for n queries.[3] Randomized Aggregatable privacy preserving Ordinal Response utilities concentrate guarantees for analyzing the original data. [4] In this paper the "clickstream" algorithm is used and it is developed to record the internet usage through the web servers and the third party services. [5] Graphchi, a disk based computation provides probabilistic graphic module and collaborative filtering. [6] Distributed differential privacy is developed in this paper to enable the

analysis and it computes scale rate. [7] Cluster-based Private Data Aggregation (CPDA) is used mainly for reducing computational overheads. This application is mainly used in military, environmental purpose, health, home and in commercial purpose. [8] This paper developed multi party computation with honest minority and it concentrates on Public-key in model and the Common Reference String (CRS) bring the mission of secure multiparty computations with n number of parties. [9] 68W15 Distributed algorithms enables the computation of classes of aggregation function which is expressed in an Abelian group.[10] SplitX analysis method used for the most part to store user data at users procedure, and to query the data. [11] The privacy for preserving user privacy on selective aggregation mainly concentrate on privacy, homomorphic encryption, laplacian noise.[12] The practical techniques developes the hardware and software based approaches over the encrypted data and also it provides the code footprint in the trusted environment.

CONCLUSION

The proposed model ensures that the user's online behavior data is completely privacy preserved. By Using asymmetric key encryption concept like RSA algorithm and the differential privacy makes sure that no one can assume the user's behaviors. The limitation in the number of queries answered will be increased considerably. In future, we can use images instead of noise for encrypting the user's data.

With reference to the papers on detect and protecting against Third party track on web and the Splitx procedure ,its now trending to track the searches made by different users over the internet where the privacy of the users results to be in risk. So, To Overcome this issues, we propose PPSA schema combined with the RSA algorithm. This Preserves the privacy of the user from being at risk and it also improves the efficiency by allowing multiple users to get connected to a server. Even a predictive analysis of the users search can't be made on implementation of this algorithm.

REFERENCES

[1] Jianwei Qian, Fudong Qiu, Student Member, IEEE, Fan Wu, Member, IEEE, Na Ruan, Member,IEEE, Guihai Chen, Member, IEEE, and Shaojie Tang, Member, IEEE, "privacy preserving selective aggregation using onlune user behavior data",2016

[2]V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in Proceedings of the ACM International Conference on Management of Data (SIGMOD), 2010, pp. 735–746

[3] "RAPPOR:RandomizedAggregatable Privacy-Preserving Ordinal Response"Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security,pp 1054-1067

[4]R. E. Bucklin and C. Sismeiro, "Click here for internet insight: Advances in clickstream data analysis in marketing," Journal of Interactive Marketing, vol. 23, no. 1, pp. 35–48, 2009.

[5] Aapo Kyrola, Guy Blelloch, Carlos Guestrin, "Graphchi: large-scale graph computation on just a pc" in Carnegie Mellon University 1123–1126, Indianapolis, Indiana, USA, 2010. ACM.

[6] 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

[7] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "Pda: Privacy-preserving data aggregation in wireless sensor networks," in Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM), 2007, pp. 2045–2053.

[8] Rafael Pass , "Bounded Concurrent secure multiparty computation with a dishonest majority" in Massachusetts Institute of Technology, 2004

[9] Benkaouz Y, Erradi M. A Distributed protocol for privacy preserving aggregation with non- permanent participants. Computing [Internet]. 2015.

[10] R. Chen, I. E. Akkus, and P. Francis, "SplitX: high-performance private analytics," in Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM), 2013, pp. 315–326

[11] C. Dwork, "Differential privacy," in Proceedings of the 33rd Interna-tional Colloquium on Automata, Languages and Programming (ICALP), 2006, pp. 1–12.

[12] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proceedings of the 32nd IEEE Symposium on Security and Privacy (S&P), 2000, pp. 44–55.

[13] R. S. Katti and C. Ababei, "Differential Privacy for preserving user Privacy on Selective Aggregation" arXiv preprint arXiv:1204.2854, 2012.

[14] D. Beaver, S. Micali and P. Rogaway The Round Complexity of Secure Protocols. In 22'th STOC, pages 503–513, 1990.

[15] R. Canetti, O. Goldreich, S. Goldwasser and S. Micali. Resettable Zero-Knowledge. In 32nd STOC, pages 235–244, 2000.

[16]The pairing-based cryptography library. [Online]. Available: https://crypto.stanford.edu/pbc/

[17] R. Canetti, J. Kilian, E. Petrank and A. Rosen. Black-Box Concurrent Zero-Knowledge Requires (almost) Logarithmically Many Rounds. SIAM Jour. on Computing, Vol. 32(1), pages 1–47, 2002.

[18] R. Canetti, Y. Lindell, R. Ostrovsky and A. Sahai. Universally Composable Two-Party and Multy-Party Computation. In 34th STOC, pages 494–503,2002.

[19] B. Chor, M. Rabin. Achieving Independence in Logarithmic Number of Rounds. In 6th PODC, pages 260-268, 1987.

[20] I. Damgard. Efficient Concurrent Zero-Knowledge in the Auxiliary String Model. In EuroCrypt2000, LNCS 1807, pages 418–430, 2000.

[21] D. Dolev, C. Dwork and M. Naor. Non-Malleable Cryptography. SIAM Jour. on Computing, Vol. 30(2), pages 391–437, 2000.

[22] C. Dwork, M. Naor and A. Sahai. Concurrent Zero-Knowledge. In 30th STOC, pages 409–418, 1998.

[23] C. Dwork and A. Sahai. Concurrent Zero-Knowledge: Reducing the Need for Timing Constraints. In Crypto98, Springer LNCS 1462 , pages 442–457, 1998. [20] U. Feige and A. Sham

[24] M. P. Armstrong, G. Rushton, D. L. Zimmerman et al., "Geographically masking health data to preserve confidentiality," Statistics in medicine, vol. 18, no. 5, pp. 497–525, 1999.

[25]A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, Handbook of applied cryptography. CRC press, 1996.

[26]F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009, pp. 19–30.

[27]R. S. Katti and C. Ababei, "Secure comparison without explicit xor," arXiv preprint arXiv:1204.2854, 2012.