International Journal of Engineering & Technology, 7 (3.8) (2018) 87-91

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Efficiently Verifiable Computations on User Data for Behavior Analysis with Privacy Preservation

Miss. Pooja R. Kotwal[1], Prof. Mangesh M. Ghonge[2], Dr. Amol D. Potgantwar[3]

[1]PG Student ,Computer Engineering, SITRC, Nashik
[2]Assistant Professor, Computer Engineering, SITRC, Nashik
[3]Head Of Department, Computer Engineering, SITRC, Nashik
ashukotwal22@gmail.com
mangesh.ghonge@sitrc.org
amol.potgantwar@sitrc.org

Abstract

Along with development of internet and web, online social network are becoming important information propagation platform with hundreds of million users worldwide. Online social network attract thousands of million users to use it every day for different purpose. So that tons of user behavior data is generated on internet. Developing endeavors have been committed to mining the inexhaustible behavior data to extract significant information for research purposes to inquire about that, or analyst to develop better ecommerce strategies for business purpose. However the concern arises with this data is security, which is going to be presented to third parties. The most recent decade has seen an assortment of look into works endeavoring to perform information conglomeration in a privacy protecting manner. Most by far of existing techniques give protection to users information yet at the cost of very limited data aggregation operations like calculating sum and mean of particular query, which barely fulfill the requirement of behavior analysis. So that, proposed system mainly focuses on privacy preservation and behavior analysis of online user data. In this paper we use general accumulation and specific collection for behavior analysis. Using cryptographic algorithm we prevent privacy disclosure from both third party data aggregator and analyst. We have executed our technique and assessed its execution utilizing a relational dataset. The results of the experiment shows that this research scheme handle both overall queries and various selective aggregate queries with acceptable computation, privacy, and overheads of the communication effectively.

Keywords: Advance Encryption, Behavior Analysis, Data Aggregation, Privacy Preservation.

1. Introduction

Recently, we have seen unprecedented development of the informal community applications, like Facebook, Amazon, Linkdn and so on are picking up importance and considering essential part in each day of lifestyles. The most widely recognized way for people to discover the items or services online is to utilize web crawlers i.e. search engines, especially Google, Bing or Yahoo. These search engines have certain criteria for giving websites more or less opportunity to be returned in search results They have turned into a dominating method for interfacing, associating, conveying and sharing data on the web [1]-[3]. There are millions of presently estimated active users uses the online social networks (OSNs) [4]. According to the professions, living location or need of problem, the users can communicate or associated with each other across the economical, geographical, political, social borders. Due to this lots of user behavior data is being generated on internet every day [5]. As each user on the social network platforms stores and shows a large amount of personal data, the worry emerges with that individual information of user might be abused by unapproved access for various reason. Keeping in mind the end goal to cut back this security thought an assortment of system have arranged as an approach to do the information mining challenges in privacy preserving way [6].

Online behavior analysis plays an important role in data mining which will be helpful for research purpose as well as other consumer for business interest [8]. Behavior analysis is the scientific study of standards of learning and behavior, which is concerned with depicting, understanding, anticipating, and evolving behavior of people. What people do and say, how they work etc. In terms of internet, online behavior analysis studies how and why users use different e-commerce platform [8],[10]. In behavior analysis, data aggregation plays a vital factor. Data aggregation is process of extracting the summary of large amount of data with report based, summarized layout to reap precise business goals or human/statistical evaluation like data mining. Now a days, in online user behavior analysis data aggregation is outsource to third party that means separate data aggregator service provider is use for data aggregation [16], so that it may happens the users data is spread across the unauthorized access. So that it is imperative to maintain the privacy of user data from third party aggregator moreover. Basically Selective aggregation and Overall aggregation is two types of data aggregation. Overall aggregation is described as, evaluating the sum and mean of somewhat reliable value of every online user. For example, the aggregate sum of time of all users today. Maximum of the present system makes use of overall aggregation, unfortunately those scheme guarantee robust privateness at the rate of obstacles on analysis, because the majority of them would more able to adequately compute summation and mean of all information without filter or selection [7].

However Selective aggregation is one of the essential options for queries on databases. Selective aggregation is nothing but the selecting particular user who meet expectations or satisfies some condition before aggregating their significance, for instance the common quantity of time online of all lady customers, hear lady is condition that satisfies and pick out the target customer.

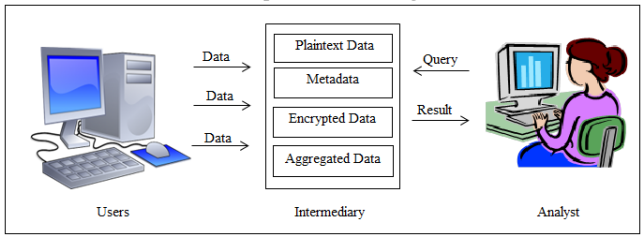

Fig. 1: There are various clients that uses the various social networking sites. Intermediary is only the information aggregator and in charge of encoding the information, and analyst gives the response to given query based on collected information.

2. Literature Review

Jianwei Qian [1] introduce a plan that encrypts users confidential information to keep privacy exposure from both outside analysts and aggregator which gives aggregated data and absolutely supports selective aggregate function for online user behavior analysis at the same time as it maintained the differential privacy.

Jemal Abawajy [2] develop a social network threat analysis framework for privacy preservation in which graph concept is used as well as K-Anonymity and state-of-the-art privacy preserving methodology is utilized for ensuring the identity of individual user when delivering or sharing confidential users data is to anonymize it.

Thripthi P. Balakrishnan [3] introduces a system based on heterogeneous abstract machine for encrypted and unencrypted computer program, by using order preserving Hash function and Homomorphic encryption.

Gul Calikli [4] describes the number of active users on social network is increases day by day as the use of OSN has grown, Due to the inappropriate sharing of information there is privacy violation of users sensitive information is take place on social networking sites so that the privacy of user has been disregarded.

Jun Du [5] design a game theoretic structure to demonstrate users interactions to steer users methodologies to take security assurance or not. In which Behavior analysis is based on community structures evolutionary game theoretic framework. By using these technique critical cost performance is obtain, which is an important parameter that can help to design incentive mechanism to facilitate the security assurance of behavior among various users. Which is effective theoretic scheme in modeling the users relationship behavior, but this framework is more complex for large scale application or large database.

Weihao Li [6] discuss about privacy concern of big data, in which a Local Record-Driving Mechanism (LRDM) is design for big data privacy to achieve privacy which contain privacy metric and a framework to optimize users privacy preserving method as well as methodology for organization or individual to find privacy.

Sum and product protocol like protocols define by the T. Jung, Xiang-Yang Li [7], according to that data aggregation scheme is carried out for data extraction.

Nihar VuppaJapati [10] in this paper a generic, secure OBA framework by using sensitivity analyzer algorithm. is design that can be applied to the industry, promoting the online market under the criteria of preserving the privacy of the users.

F. Chen [11] describes a scheme for multidimensional range queries which is scalable and efficient for privacy and integrity preservation. As well as they propose an order preserving hash based function to encrypt both data and queries so that a cloud supplier can efficiently proceed encrypted queries over the data which is in the coded form.

X. Yi, M. G. Kaosar [13] illustrates two protocols like PIR and PBR from Fully Homomorphic Encryption (FHE) for secure information retrieval.

S. Oh and P. Viswanath [15] introduces Composition theorem for differential privacy in which Laplacian mechanism or the staircase mechanism is used to add noise in the data so that it can't be spread over the unapproved access get to.

T. Jung [16] states that how aggregation service supplier and analyst can understand or learn users sensitive information over multiple online users and it uses data aggregation scheme. It is done by external aggregator or multiple parties. Homomorphic Encryption is used for encryption and decryption process. In this framework Secure Multiparty Computation is also used.

3. Key Contribution

In the proposed system, we use a secure and efficient privacy preserving data aggregation scheme for behavior analysis. We propose an effective privacy-preserving scheme that can process multi-dimensional range queries across the relational database. In order to achieve privacy in high speed network environment, we uses the digital signature algorithm that secures the user generated data which will store by client agent. We described some of the outstanding challenges that need to be addressed in future research. Toward the end, we had taken extensive experiments on synthetic dataset to assess the viability and productivity of our scheme.

4. Proposed Framework

In this section, we talk about the outline of our framework. We will introduce the working environment of proposed system, analyze the particular function requirements and indicates the design principal of our system.

4.1. Background and Environment

The principal goal of proposed system is to design a comprehensive privacy-preserving, accountable, and efficient authentication framework for behavior analysis which will hide or provide privacy to certain sensitive information so that they can't be revealing to unauthorized access. The core task to meet this goal is to develop a group oriented cryptosystem. Use of Advance Encryption Standard, for encryption which will support privacy preservation as well as maintaining the differential privacy.

- To give the privateness-retaining selective aggregation on users information.
- To combine encryption and differential privacy mechanism by using AES algorithm and ECDSA algorithm to protect users sensitive information from both aggregation service supplier and analysts.

4.2. System Components

1. User:
There are n numbers of users that accessing the social networking sites as per individuals need.
2. Client Agent:
Client agent is an entity which is authoritative for collecting the user sensitive data from users and updates the data as per need.
3. Aggregator:
The function of data aggregator module is to extract the information from webpage browsed or the data stored by the client agent. Aggregator extracts the information depending upon the attribute selected by authority. Although the data aggregator follows the prescribed operations and does not collude with analyst.
4. Authority:
The Central Authority (CA) is a completely trusted foundation that stores users organizes in its stockpiling. It is also responsible for system setup i.e. selection of collection of attributes and gener-

ating public/private key pairs. It also generates the query count and according to query updations CA updates the generated value and public key. Authority plays main role in behavior analysis as it send the response back as far as answer to query to analyst.

5. Analyst:

Analysts are individuals or institutions that want to query about user data. Analyst send the query Q to intermediary, in that case authority checks the query with collected data and send answer to the analyst. If the query is invalid then it send the error message.

4.3. System Design

In this clause, the solving approaches and efficiency issue are described. It explains the block diagram of system and steps of process. In the proposed system as mention earlier, data aggregation scheme such as overall aggregation and selective aggregation is used to study the behavior analysis of user. Following block diagram shows the component of proposed framework.

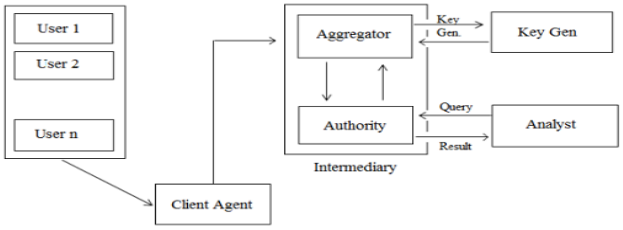

Fig. 2: Workflow of Proposed Approach

In the proposed system clients are installed on user side. Clients are nothing but the number of users which access the various social networking sites every day for different purpose. Due to the increasing use of social networking sites by various clients, enormous amount of user behavior data is generated on the server side. In data mining, behavior analysis plays an important role. Behavior analysis is nothing but the study of behavior of users. As mention above lots of data is being generated on internet every day, this data is managed by third party called as aggregation service provider. Aggregator is in charge of dealing with the gathered information [16]. Data aggregation is utilized for statistical analysis, by which data is collected and expressed in summary form. Primary goal of aggregation is to get more information about specific group based on the specific variable such as age, profession, income etc. the information about such group can then be used for website personalization. As mention above, for statistical analysis we use aggregation function. Statistical analysis is nothing but the study of numeric data. In aggregation process statistical analysis is used to perform numeric operations. In the propose system aggregator collects this data, perform some operations on this data and express this data in summary form.

Analyst plays an vital role in propose system and it is mainly used for data analysis. To analyse the important aspect or information from aggregated data analyst send the query to aggregator. Aggregator reply these queries based on aggregated data. The query fired from analyst is a SQL type of query which includes any type of operation such as select, insert, delete, update etc. Due to the analysis of behavioral data it can be helpful for business purpose, purpose as well as understanding the mentality of people in purchasing the product from different shopping sites [17]. As mention above, aggregator is responsible for data aggregation, so aggregator is aware about all the user behavior data. But the aggregator is third party untrusted entity [18], it can spread this data or unauthorized user can use this data from aggregator. To magnetize the adversity of existing system, Advance encryption standard algorithm is used. As shown figure there are various components involves in proposed framework.

- Data Collection: Data collected from multiple users when user access the social sites is done by client agent after collecting the all data client agent is responsible for performing the data encryption using Advance Encryption Standard algorithm.
- Data Aggregation: Aggregator expressed the tons of user behavior data in summary form.
- Setup Phase: All key generation operation and aggregate table creation is carried out by authority which comes under the intermediary. It creates the public and private key for third party data aggregator.
- Query Evaluation and Result Generation: Query evaluation process can be executed when the authority decides the attributes for collected data, before data collection is finished, because query evaluation does not consist all the data in table T. In this phase query processing is taking place where, analyst sends SQL type of query to authority, and based on this aggregated data by aggregator, authority reply back to the query from analyst by adding noise and result updating.

5. Mathematical Model

The System S can be mathematically defined as a collection of tuples. S can be written as,

$S= \{I, O, Q, U, A, T, K\}$

I = I is the set of input to system; $I= \{i_1, i_2,i_3, ...i_n\}$
O = O is the set of output O; $O= \{o_1, o_2,o_3, ...o_n\}$
Q = Set of Analyst query Q; $Q= \{q_1, q_2,q_3, ...q_n\}$
U = Set of client agent U; $U=\{u_1, u_2,u_3, ...u_n\}$
A = Set of attributes A; $A= \{a_1, a_2,a_3, ...a_n\}$
K = Set of generated key public/private pairs
$K = \{k_{pub} / k_{pri}\}$

Function f_1 : generate set of att. and public, private keys
$f_1 (A)=\{(a_1, a_2,a_3, ...a_n), k_{pub} / k_{pri} \} \in (A, K)$

Function f_2 : Read user data and perform encryption
$f_2 (U)=\{(u_1, u_2,u_3, ...u_n)-(E_1, E_2,E_3, ...E_n)\} \in E_U$

Function f_3 : Read encrypted data and update table entry
$f_3 (E_u)=\{(E_1, E_2,E_3, ...E_n)- T\}€T$

Function f_4 : Read analyst query and start aggregation
$f_4 (E_Q)=\{(a_1, a_2,a_3, ...a_n)\} A$

Function f_5 : Decrypt the message and send to analyst
S_{att} = Decrypt (S_k, $E(S_{att})$)
S_k= Private key of authority

6. Methodology of Evaluation

To evaluate the flow of proposed system the experiments conducted are applied to dataset for data aggregation which will supports the multidimensional range queries.. In our task we mainly concentrate to data analysis using privacy preserving way with high efficiency. Our experiments were implemented in Java, Eclipse is used as Integrated development environment and carried out on a PC with processor Pentium-IV and 2GB RAM.

6.1. Dataset

Proposed system can be work on any type of relational dataset or text. As we know that relational dataset establishes a well-defined relationship between database tables and it supports to the Structured Query Language (SQL), which is useful for database interaction, and standard user application that gives an easy programming interface.

6.2. Evaluation Parameter

In our experimental evaluation, we analyse the system performance under varying parameters, such as execution time vs sample size, Computational overhead vs. number of bits, Communica-

tion overhead. The test data and relevant parameter is show in table and corresponding sample size is listed. The time required to performing the whole process is measured in seconds.

6.3. Result and Discussion

Table 1: Sample size of data and relevant parameters for time comparison

Sample Size	Algorithm	
	Selective Aggregation Algorithm (Time in ms)	Our Contribution (Time in ms)
1000	12	4
2000	47	25
5000	85	32
10000	124	41
500000	254	125

Table 2: Sample size of data and relevant parameters for relative error

Sample Size	Relative Error	
	Selective Aggregation Algorithm	Our Contribution
100	0.5	0.4
200	0.35	0.33
300	0.25	0.21
400	0.22	0.19
500	0.2	0.17
600	0.19	0.18
700	0.17	0.14
800	0.15	0.13
900	0.12	0.10
1000	0.1	0.08

Table 3: Communication Overhead

Component	Communication Overhead		
	Extension 1	Extension 2	Extension 3
Authority	30S+107	30S+114	30S+104
Aggregator	30S+60	60ls+ 60	30S+40

Where,
I= No. of digits
S= Sample Size
n= No. of bits (max=4)
k= No. of chars (max=26)

Table 4: Computational overhead vs number of bits

No. of Bits	Computational Overhead	
	Runtime E2/E3	Our Contribution
7	2	1.8
8	3.5	1.85
9	3.9	1.9
10	4.2	1.86
11	5.8	1.94
12	8.2	1.97

Table l shows the Time comparison between overall aggregation algorithm and proposed algorithm by taking various sample set of data. Likewise, Relative error comparison between overall aggregation algorithm and proposed algorithm is shows in Table 2. Communication overhead is defined in table 3 and table 4 describes computational overhead vs number of bits.

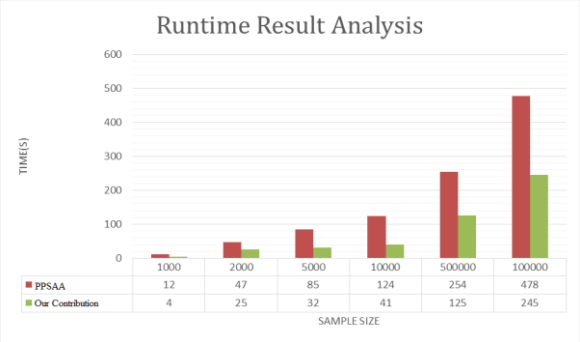

Fig. 3: Time Comparison

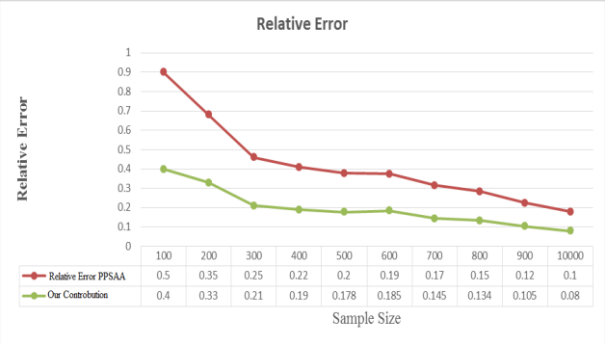

Fig. 4: Relative Error

The comparison of Time, Relative Error is given in Figure 3, 4. The Comparison of all parameter is done by taking various sample data. The results which we have got are on comparing the overall aggregation algorithm and proposed approach.

7. Discussion and Conclusion

The key advantage to find out the behavior analysis of online user is to study market analysis, customer retention, corporate analysis, production control etc. but there are various challenges of making online user data aggregation. In this paper, we have study the problem which will occur during the data aggregation in study of behavior analysis while preserving the privacy of online users data, and describe data aggregation using the AES algorithm and EC-DSA algorithm which fully supports to the selective aggregation scheme for behavior analysis of online user while maintaining the privacy, as well as we have studied that, aggregation based implementation is better solution for privacy issue in online user behavior.

8. Future Work

In this research, we have described the how data aggregation scheme is efficient for behavior analysis of user as well as the data aggregation with privacy preserving way gives assurity of protection of individuals privacy. We have presented, the data is being secured when the data is stored.
In the future, we can extend our study to present a secure file syncing and sharing service(FSS) for social networking sites based on digital forensics mechanisms against the abnormal or suspicious attackers. Using a group oriented cryptosystem for cloud data encryption which will supports the trailor tracing and revoking mechanism for digital forensics of the detected attacker. In short we intend to introduce a anomaly detection technique using pattern matching which will help in detection of suspicious user or abnormal player and will detect and renounce the authorities of user.

Acknowledgement

I Would sincierely like to thank our Prof. Mangesh M. Ghonge, Department of Computer Engineering, SITRC, Nashik for his guidance, encouragement and the interest shown in this project by timely suggestions in this work. His expert suggestions and scholarly feedback had greatly enhanced the effectiveness of this work.

References

[1] J. Qian, F.Qiu, Fan Wu, N. Ruan, G. Chen, S. Tang Privacy preservation Selective Aggregartion of Online user Behavior Data,IEEE Transaction On computers,vol 14, TC.2016.2595562, 2016.
[2] J. Abawajy, Mohd Izuan Hafez Ninggal, Tutut Herawan Privacy Preserving Social Network Data Publication, IEEE Transaction, vol 27, 2016.

[3] Thripthi. P. Balakrishnan , Mr.S.Vijayanand, T. Senthil Prakash, Efficiently Verifiable Computation On Encrypted Data, International Journal of Science and Engineering Research (IJ0SER), vol 4, Dec. 2016.
[4] G. Calikli, M.Law, A. Bandara, A. Russo, L. Dickens, B.Price, A. Stuart, M. Levine, B.Nuseibeh Privacy Dynamics: Learning Privacy Norms for Social Software,IEEE/ACM 11th International Symposium on Software Engineering for Adaptive and Self-Managing Sytems, vol 10, 2016.
[5] Jun Du, Chunxiao Jiang, Shui Yu, Kwang-Cheng Chen, Yong Ren Privacy Protection: A Community-Structured Evolutionary Game Approach,IEEE, vol 5, 2016.
[6] Weihao Li, Hui Li LRDM: Local Record-Driving Mechanism for Big data Privacy Preservation in Social Networks, IEEE first International conference on data.
[7] Taeho Jung, Xiang-Yang Li, Meng Wan, Collusion tolerable privacy preserving Sum and Product Calculation without Secure Channel, vol 14, 2015.
[8] Linke Guo, Chi Zhang, Yuguang Fang, A Trust-based Privacy Preserving Friend Recommendation Scheme for Online Social Networks, IEEE Transaction on Dependable and Secure Computing, vol 14, 2015.
[9] Borui Yang, Jianxin Li, Yingjie Cao, Hua Wei, Peiyuan Sun, Nannan Wu and Bo Li,Lu Liu, ShutterRoller: Preserving Social Network Privacy towards High Speed Domain Gateway, IEEE International Conference on Computer and Information Technology, vol 8, 2015.
[10] Nihar VuppaJapati and Joan S. Park, Online Behavioral Advertising (OBA) with Privacy Protection, IEEE, 2014.
[11] F. Chen, A. X. Liu, Privacy and Integrity preserving Multidimensional range queries for cloud computing,2014.
[12] B. Mood, D. Gupta, K. Butler, J. Feigenbaum, Reuse it or Lose it: More Efficient secure computation through reuse of encrypted values , 2014.
[13] Xun Yi, Mohammed Golam Kaosar, Russell Paulet, and Elisa Bertino, Single Database Private Information Retrieval from Fully Homomorphic Encryption,IEEE Transaction on Kowledge and Data Engineering, vol 25,May 2013.
[14] Luca Ferretti, Michele Colajanni, and Mirco Marchetti, Distributed, concurrent, and independent access to encrypted cloud databases,IEEE Transaction on Parallel and Distributed System vol 11, 2013.
[15] Peter Kairouz, Sewoong Oh, Pramod Viswanath, The Composition Theorem for Differential Privacy vol 10, 2013.
[16] Taeho Jung , XuFei Mao, Xiang-Yang Li, Shao-Jie Tang , Wei Gong ,and Lan Zhang, Privacy Preserving Data Aggregation without Secure Channel: Multi-variate Polynomial Evaluation, vol 9, 2013.
[17] J. C. Duchi, M.I. Jordan, M.J. Wainwright, Local Privacy and Statistical minmax rates,2013.
[18] Ruichuan Chen, Istemi Ekin Akkus, Paul Francis, SplitX: High Performance Private Analytics, vol 12, 2013.
[19] Ruichuan Chen, Alexey Reznichenko, Paul Francis, Johannes Gehrke, Towards Statistical Queries over Distributed Private User Data, vol 14, 2012.
[20] Istemi Ekin Akkus, Ruichuan Chen, Michaela Hardt, Paul Francis, Johannes Gehrke, Non-tracking Web Analytics, vol 12, 2012.
[21] A. Khalique, K. singh, S. Sood Implementation of Eliptical Curve Digital Signature Algorithm, IJCA, Vol 2, May 2010.