RESEARCH ARTICLE OPEN ACCESS

M-Privacy for Collaborative Data Publishing By Using Heuristic Approach

[1]Ms. Sheetal D. Shahare , [2]Mr.Sachin Barahate
[1]Dept name: Computer science ,Yadavrao Tasgaonkar college of Engg , Mumbai , India
[2]Dept name: Information Technology , Vasantdada Patil college of Engg, Mumbai, India

Abstract

According to the survey Maintaining and preserving privacy has become more significant problem. 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. For M-privacy several anonymization techniques have been used such as bucketization, generalization, perturbation which does not prevent privacy and fail to maintain a privacy constraint and results in loss of information. So we consider the collaborative data publishing for anonymizing horizontally partitioned data at multiple data provider. First, we provide the notion of *m*-privacy, which give guarantees that the anonymized data satisfies a given privacy constraint against any intruder attack. Then we present heuristic algorithms with effective pruning strategies and adaptive ordering techniques for efficiently checking m-privacy for a set of records i.e to breach the privacy. Here intruder are also detected which try to breach the privacy. This technique shows the better utility and efficiency than the previous techniques.We develop a truthful and efficient M-privacy for collaborative data publishing by using pruning strategy and providing them anonymized data in case of emergency.

Keywords — **m-privacy ,database, anonymizaton.**

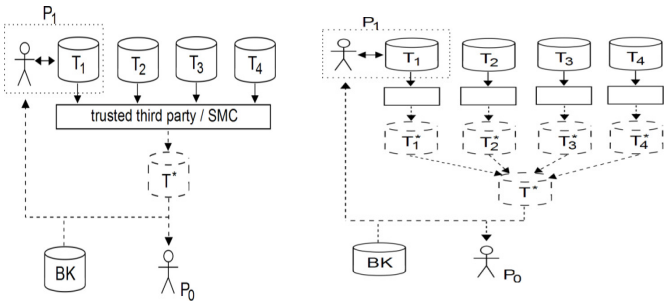

Fig 1. Distributed data publishing

I . INTRODUCTION

In today eras M-privacy is big issue. M-Privacy is to protect the large amount of data from external recipient or from any attacker. Every one try to make their data secure and confidential i.e promising approaches for sharing data while preserving individual privacy. But when data is distributed among multiple number of data provider then for anonymizaton took two approaches. First anonymized data independently which is time consuming and second using collaborative data publishing[1].Here we partitioned data horizontally. For example Health care domain in which patient data or product data is very sensitive. As the data of hospital is very sensitive then Anonymization act as TTP(TRUSTED THIRD PARTY PROTOCOL) to do the computations[3]. By using anonymous system the data is changed and after short time of period data is released. Anonymization means a type of information whose intent is privacy protection. It is the process of either encrypting or finding personally identifiable information from data sets, so that the people whom data is describe remain anonymous.

Our goal is to publish an anonymized view of the integrated data so that a data recipient including the data providers will not be able to compromise the privacy of the individual records given by other parties also we breach the privacy which can be done by anonymized data.(i.e nothing but pruning strategy). We also find the attacker which is also called intruder at what time try to access the data and what data he try to access. Here we give anonymized data to the

intruder in case of emergency for example. If a doctor is on leave and he has to follow up his patient and to call down is relatives then he can't access that patient data and no other can reach that doctor so in such case the intruder will get access for the third time and he will get anonymised data in the form of zip code. Our main goal is to breaching the privacy, maintaining privacy, to prevent from attacker or intruder and finding the intruder. The first two approach are addressed in existing work, the last one receives little attention and will be the focus of this paper.

II. MODULE DESCRIPTION

1.PATIENT AND PRODUCT REGISTRATION

In this module if patient has to take treatment then he or she must be register their details such as name, mobile no, disease, age , etc . these details are maintained in database by the hospital management and can only doctors see the details and admin. Similarly we need to the product registration.

2. Attack by external data by using anonymised data

A data recipient could be an attacker who try to infer additional information about the records which are using for the published data and some background knowledge which is publicly available that is external data.

3. Attacks By Data Providers Using Anonymized Data And Their Own Data

Each data provider can also use anonymized data and his own data to which infer additional information about other records. If we compare this attack by the external recipient in the first attack scenario, then we can see that each provider has additional data knowledge of their own records, which can help with the attack. This issue get further more worsened when multiple number of data providers collude with each other.

4. Doctor Login

In this module Doctor can see all the patients details and will get the background knowledge(BK),by the same time he will see horizontally partitioned data of distributed data base of the group of hospitals and can see how many patients are affected without knowing of individual records of the patients and sensitive information about the individuals.

5. Admin Login

In this module Admin acts as Trusted Third Party(TTP).He can see all individual records of the patient and their sensitive information among the overall hospital distributed data base other than this he can also see the unauthenticated user with all details along with time and which data he access and date. Anonymization can be done by this people. He/She collected information from various hospitals and grouped into each other and make them as an anonymised data.

I. M-PRIVACY DEFINITION

First of all we describe our problem setting Then we present our m-privacy definition with respect to a given privacy constraint so that we can prevent inference attacks by opponent or intruder. Let T = {t1,t2,...} be a set of records horizontally distributed among n number of data providers P = {P1,P2,...,Pn}, such that Ti ⊆ T is a set of records given by Pi. Our goal is to publish an anonymized table T∗ while preventing from intruder and to find out the intruder. To protect our data from external attacker with certain background knowledge BK, we assume a given privacy requirement C, which is defined by a conjunction of privacy constraints: C1 ∧ C2 ∧ ... ∧ Cw. If a set of records T∗ satisfies C, we say C(T∗) = true. We can say that privacy is maintained.

II. VERIFICATION OF M-PRIVACY

Now we are checking whether a set of records of data satisfies m-privacy and showing the ability to challenge the number of opponents or intruder. In this section we present heuristic algorithm to breach the privacy that is pruning strategy is applied.

A. HEURISTIC ALGORITHM

It's a technique designed for solving problem more quickly when some classic methods are too slow. There are two pruning strategies out of which we used set of search strategy to breach the privacy that is anonymised . Attack done by external attacker, a recipient for example P1, could be an attacker and try to infer sme additional information of the current data using the published data (T*) which is sore in data base and some background knowledge (BK) such as easily available external data. Now considering some another type of attack which is attack done by data providers using their own data; each data provider such as P2 in fig 1 can also use anonymized data T* and his own data T1 to infer additional information about other records. If we compare to other attack, then we can see that attack done by the internal attacker has more additional in data information of their own records, which is very helpful for attack. This issue get more degraded when multiple number of data providers mist with each other. So we not only save our data from any intruder attack but we find out the intruder or attacker who try to access or login into our system then at what time and what data he try to access , with all details are recorded into our database or we maintained that data .

B. PRUNING STRATEGY

Its the one which checks the coalition to breach the privacy or not. If its not possible to breach the privacy then its not to be checked else if its breach the privacy then we need to check or pruned. Here we have used top down approach method which will take data from top to bottom. The top-down algorithm checks the coalitions in a top-down fashion using downward pruning, starting from (nG1)-adversaries and moving down until a violation by an m-adversary is detected or all m-adversaries are pruned or checked.

- Count total values to search –n
- Then we count total records in dataset-N
- Fetch first record to search –i
- Continue search from N
- If found display that record
- Continue with i++

Else

i++

- Display values in record set

Above discussion shows for the admin login where admin has the authentication to fid out the attack of intruder.

Then we display anonymised records. Here attacker or intruder will get anonymised data

- Count records to anonymise –N
- Fetch first record –i
- Anonymise mobile no.
- Store this record in intruder table.
- Put record in data set
- If i=N

Display all records to intruder

Else

Go to step 2

Ta* TABLE 1.

Provider	name	Age	Zip	disease
P1	alice	20-30	***	flu
P2	john	20-30	***	cancer
P2	bob	31-35	***	astha ma
P1	alice	31-35	***	flu
P3	john	31-35	***	cancer

Tb* TABLE 2.

Provider	name	Age	Zip	disease
P1	alice	20-40	***	flu
P2	john	20-40	***	cancer
P2	bob	20-40	***	astha ma
P1	alice	20-40	***	Flu
P3	john	20-40	***	cancer

III. EXPERIMENTAL RESULT

In this project we have a new type of potential attacker for collaborative data publishing. We are giving first priority to maintain our privacy i.e. making m=1 true if it happens then our privacy is guarantying enough and intruder will get anonymised data. Here intruder will login to system for the first time then message will display invalid username or password. Similarly intruder login to the system 2nd time same message display i.e. invalid username or password. When intruder login for third time got access to system and will get anonymised data. Here we concluded three cases.

In first case we have shown the graph of larger records and smaller records and the time required to display while comparing the patient records. In second case we have shown the graph of authenticated user to display the number of patient records and time taken to display. In third case we have shown the graph of intruder attack, number of records access and time required to display i.e. what kind of records attacker attacks and how many number of a records at what time and period of a time require to display the records is displayed on admin side

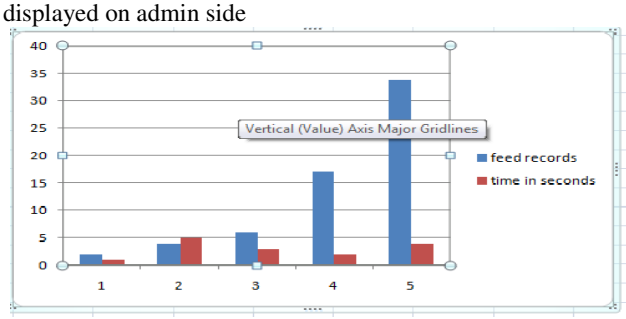

Fig .2: Graph for large and small records.

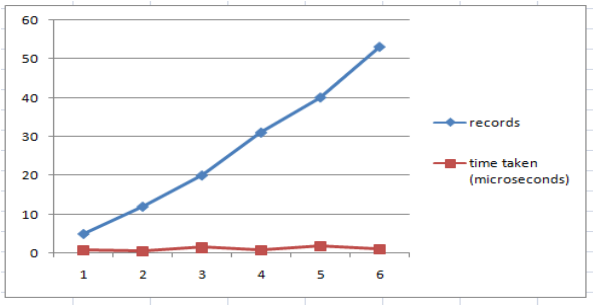

Fig 3: Time required to show data to authenticated user.

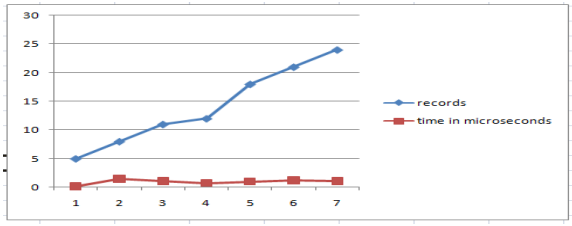

Fig 4. Time taken by intruder to display records anonymise

IV. RELATED WORK

Privacy preserving data analysis and publishing has received much attention in recent years [1], [2], [3]. Mostly the work has focused on a single data provider setting and

considered the data recipient as an attacker. A large body of literature [2] assumes some limited background knowledge of the attacker and defines privacy using adversarial notion [9] by considering specific types of attack . Some principles include k-anonymity [10], [11], l-diversity [9], and t-closeness [12]. Few recent works have done the level background knowledge as corruption and studied some perturbation techniques under these weak privacy notions [20].

In the distributed setting we studied, since each data holder knows there own records, so the corruption of records is an inherent element in our attack model and is further complicated by the collusion power of the data providers. On the other hand, differential privacy [1], [3] is an unconditional privacy gives the guarantee for statistical data release or some data computations. While giving some desirable unconditional privacy guarantee, non-interactive data release with differential privacy remains an open problem. Many different anonymization algorithms have been introduced so far including Datafly [13], Incognito[12], and Mondrian [16].

In our research we considered that binary search algorithm as a baseline because its efficiency and extensibility. There are some work which are focused on anonymization of distributed data [5],[6], [23] studied distributed anonymization for vertically partitioned data using k-anonymity. Zhong et al. [24] studied classification on data collected from individual data owners (each record is contributed by one data owner) to address high dimensional data. Our work is the first one which considers data providers as potential attackers in the collaborative data publishing setting and explicitly models the inherent instance knowledge of the data providers.

VI.CONCLUSION

In this project we considered a new type of attacker in a collaborative data publishing where multiple data providers comes together to perform particular task which is called as m-adversary. Our main focus is to maintain the privacy i.e. making m=1. To prevent a privacy disclosure by any m-adversary we provide anonymised data to the attacker or the intruder. We also introduce a pruning strategy to breach the privacy which is nothing but a Heuristic approach for checking strategies to ensure high utility and m-privacy of an anonymised data. To find out an Intruder we make a use of IDS technology (Intrusion Detection System) in which Intruder will get a third time access to a data which is an anonymised data. Our experiments confirm that our approach achieves better or comparable result ensuring data m-privacy.

Future Scope

- If number of machines are available then how u will come to know on which machine attacker was login ,so for that we can crack attacker by using IP address.(by networking)
- Image of intruder can be fetched.
- Big data hadoop can be implemented for bulk of data for health care domain.

References

[1] Olvi L. Mangasarian, “Privacy-Preserving Horizontally Partitioned Linear Programs”.2003

[2] P. Krishna Prasad and C. Pandu Rangan “Privacy Preserving BIRCH Algorithm for Clustering over Arbitrarily Partitioned Databases”.

[3] Ali Inan, Yücel Saygin, Erkay Sava, Ayça Azgin Hinto lu, Albert Levi “Privacy Preserving Clustering on Horizontally Partitioned Data”.

[4] Jaideep Vaidya and Chris Clifton “Privacy-Preserving Decision Trees over Vertically Partitioned Data”

[5] Zhiqiang Yang and Rebecca N. Wright, “Privacy-Preserving Computation of Bayesian Networks on Vertically PartitionedData” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 9, SEPTEMBER 2006.

[6] Khuong Vu and Rong Zheng Jie Gao “Efficient Algorithms for K-Anonymous Location Privacy in Participatory Sensing” 2012 Proceedings IEEE INFOCOM.

[7] Sebastian Schrittwieser, Peter Kieseberg, Isao Echizen, Sven Wohlgemuth, Noboru Sonehara, and Edgar Weippl “An Algorithm for k- anonymity-based Fingerprinting”

[8] S. Goryczka, L. Xiong, and B. C. M. Fung, “m-Privacy for collaborative data publishing,” in Proc. of the 7th Intl. Conf. on Collaborative Computing: Networking, Applications and Work sharing, 2011.

[9] W. Jiang and C. Clifton, “A secure distributed framework for achieving k-anonymity,” VLDB J., vol. 15, no. 4, pp. 316–333, 2006

[10] L. Sweeney, “k-Anonymity: a model for protecting privacy,” Int. J. Uncertain. Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557–570, 2002.

[11] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, “Incognito: efficient full-domain k-anonymity,” in Proc. of the 2005 ACM SIGMOD Intl. Conf. on Management of Data, 2005, pp. 49–60.

[12] N. Mohammed, B. C. M. Fung, K. Wang, and P. C. K. Hung, “Privacy preserving data mashup,” in Proc. of the 12th Intl. Conf. on Extending Database Technology, 2009, pp. 228–239. [24]

[13] N. Li and T. Li, “t-closeness: Privacy beyond k-anonymity and ldiversity,” in In Proc. of IEEE 23rd Intl. Conf. on Data Engineering (ICDE), 2007.

[14] R. Burke, B. Mobasher, R. Zabicki, and R. Bhaumik, “Identifying attack models for secure recommendation,” in In Beyond Personalization: A Workshop on the Next Generation of Recommender Systems, 2005.