ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

A Novel Method for Privacy Preserving Data Publishing

[I]G.Sivaranjani, [II]G.Vaira Suganthi
[I]PG Scholar, Dept. of CSE, Sethu Institute of Technology, Affiliated to Anna University, Kariappatti, Tamilnadu, India
[II]Assistant Professor, Dept. of CSE, Sethu Institute of Technology, Affiliated to Anna University, Kariappatti, Tamilnadu, India

Abstract
There are many different reasons that people put forward to support the proposition that privacy is important. People need private space and need to be free to behave in public data sharing. Privacy protection is the interests of one person or category of people. Collaborative data publishing is a method which is used to share some helpful information like blood donor details in hospital. Collaborative data publishing may face various attacks like potential loss of integrated data utility. This work proposes a solution to one such attack called "insider attack". m-privacy guarantees that the anonymized data satisfies a given privacy constraint. to solve the attacks provider- aware anonymization, high dimensional top down specialization algorithms are used. Provider aware anonymization algorithm is ensuring m-privacy methodology which is highest rated anonymized data with efficiency. Then, trusted third party would give anonymized data's to n-providers. High dimensional top down specialization achieves privacy.

I. Introduction

People think of privacy as some kind of right. Unfortunately, the concept of a 'right' is a problematical way to start, because a right seems to be some kind of absolute standard. It's very easy to get confused between legal rights, on the one hand and natural or moral rights on the other. Privacy is the interest that individuals have in sustaining a 'personal space' and free from interference by other people and organizations.Individuals claim that data about themselves should not be automatically available to other individuals and organisations, and that, even where data is possessed by another party, the individual must be able to exercise a degree of control over that data and its use. This is sometimes referred to as 'data privacy' and 'information privacy'. With the close coupling that has occurred between computing and communications, particularly since the 1980s, the last two aspects have become closely linked. This is the primary focus of public attention and of this document. It is useful to use the term 'information privacy' to refer to the combination of communications privacy and data privacy. The increased capability of information technology to support data collection, storage, processing, discovery, use and disclosure and the privacy interest runs emphatically counter to attempts to convert anonymous into identified transactions. Such case sensitive data's are listed as per them age and zipcode.in this case hacker or the insider hacker cannot identify the user sensitive data's who is related. So end user can hold the privacy with themselves with this project.

There are many different reasons that people put forward to support the proposition that privacy is important. Some will be provides a classification and brief overview. Psychologically, people need private space. This applies in public as well as behind closed doors and drawn curtains. We need to be able to glance around, judge whether the people in the vicinity are a threat, and then perform actions that are potentially embarrassing, such as breaking wind.

Sociologically, people need to be free to behave and to associate with others, but without the continual threat of being observed.The privacy interests of one person or category of people may conflict with some other interest of their own, and the two may have to be traded off (e.g. privacy against access to credit, or quality of health care);The privacy interest of one person or category of people may conflict with the privacy interests of another person, or another category of people (e.g.health care information that is relevant to multiple members of a family).

The privacy interest of one person or category of people may conflict withother interests of another person, category of people, organisation, or society as a whole (e.g. creditors, an insurer, and protection of the public against serious diseases). Privacy Protection is a process of finding appropriate balances between privacy and multiple competing interests. Because there are so many dimensions of the privacy interest, and competing interests, at so many levels of society, the formulation of detailed, operational rules about privacy protection is a difficult exercise. An anonymous record or transaction is one whose data cannot be associated with a particular individual, either from the data itself, or by combining the transaction with other data.

Some of the reasons that people use anonymity are of dubious social value, such as avoiding detection of their whereabouts in order to escape responsibilities. Other reasons are of arguably significant social value, such as avoiding physical harm, enabling 'whistle-blowing', avoiding unwanted and unjustified public exposure and keeping personal data out of the hands of intrusive marketers and governments. An identified record or transaction is one in which the data can be readily related to a particular individual. This may be because it carries a direct identifier of the person concerned or because it contains data which, in combination with other available data, links the data to a particular person.

A. Personal Privacy

Most people have a strong sense of privacy in relation to the exposure of their body to others. This is an aspect of personal modesty. A person will go to extreme lengths to protect this personal modesty, the main way being the wearing of clothes. Other ways include erection of walls, fences, screens, use of cathedral glass, partitions, by maintaining a distance, beside other ways. People who go to those lengths expect that their privacy will be respected by others. At the same time, people are prepared to expose themselves in acts of physical intimacy, but these are confined to exposure in circumstances and of persons of their choosing. Even a discussion of those circumstances is regarded as intrusive and typically unwelcome.

B. Informational Privacy

Information or data privacy refers to the evolving relationship between technology and the legal right to, or public expectation

of, privacy in the collection and sharing of data about one's self. Privacy concerns exist wherever uniquely identifiable data relating to a person or persons are collected and stored, in digital form or otherwise. In some cases these concerns refer to how data is collected, stored, and associated. In other cases the issue is who is given access to information. Other issues include whether an individual has any ownership rights to data about them, and/or the right to view, verify, and challenge that information.

Various types of personal information are often associated with privacy concerns. For various reasons, individuals may object to personal information such as their religion, sexual orientation, political affiliations, or personal activities being revealed, perhaps to avoid discrimination, personal embarrassment, or damage to their professional reputations.

Financial privacy, in which information about a person's financial transactions is guarded, is important for the avoidance of fraud including identity theft. Information about a person's purchases, for instance, can reveal a great deal about their preferences, places they have visited, their contacts, products (such as medications) they use, their activities and habits etc.

Internet privacy is the ability to determine what information one reveals or withholds about oneself over the Internet, who has access to such information, and for what purposes one's information may or may not be used. For example, web users may be concerned to discover that many of the web sites which they visit collect, store, and possibly share personally identifiable informationabout them. Similarly, Internet email users generally consider their emails to be private and hence would be concerned if their email was being accessed, read, stored or forwarded by third parties without their consent.

Medical privacy allows a person to withhold their medical records and other information from others, perhaps because of fears that it might affect their insurance coverage or employment, or to avoid the embarrassment caused by revealing medical conditions or treatments. Medical information could also reveal other aspects of one's personal life, such as sexual preferences or proclivity. A right to sexual privacy enables individuals to acquire and use contraceptives without family, community or legal sanctions

C. Organizational Privacy

Governments agencies, corporations, groups/societies and other organizations may desire to keep their activities or secrets from being revealed to other organizations or individuals, adopting various security practices and controls in order to keep private information confidential. Organizations may seek legal protection for their secrets. For example, a government administration may be able to invoke executive privilege or declares certain information to be classified, or a corporation might attempt to protect valuable proprietary information as trade secrets.

D. Spiritual and Intellectual Privacy

The earliest legislative development of privacy rights began under British common law, which protected "only the physical interference of life and property." Its development from then on became "one of the most significant chapters in the history of privacy law." Privacy rights gradually expanded to include a "recognition of man's spiritual nature, of his feelings and his intellect." Eventually, the scope of those rights broadened even further to include a basic "right to be let alone", and the former definition of "property" would then comprise "every form of possession -- intangible, as well as tangible."

E. Literature Survey

S. Zhong, et al. [11] in the paper 'Privacy enhancing k anonymization of customer data' , has been proposed that the miner extracts the anonymous part of the data. (i.e. the maximum subset of rows that is k anonymous). But does not learn extra information about the sensitive attributes of the rows outside the k anonymous part. We design a protocol that privately extracts the k anonymous part of a table. The basic idea of this design is that each customer encrypts their sensitive attributes using an encryption key. Each customer submits to the miner one share of the key(s) corresponding to her quasi identifiers. As a result, if and only if there are at least k customers whose quasi identifiers are equal, the miner is able to recover a key.

Robin Burke, et al. [1] in the paper 'A Identifying attack models for secure recommendation' has been proposed that the most typical formulation of the idea is the user to user comparison. The system compares the patterns of preference for each item across all users. In this paper they used the KNN technique The preservation of the trust is important for both users and site owners and it is dependent upon the perception of recommender systems. Because recommendation systems are dependent on external sources of information, such as user profiles which are vulnerable to attack. If a system generates recommendations collaboratively, that is by user to user comparison, users can generate profiles for the purpose of biasing the system's recommendations for against certain products.

Graham Cormode, et al. [12] in the paper 'Minimizing Minimality and Maximizing Utility: Analyzing Method based Attacks on Anonymized Data', has been analyzed that the increased probability with which the attacker can associate particular SAs with particular Qis. As each group is generated independently from the others, it is sufficient to analyze a single group at a time. If the group contains only items, there is no information for the attacker to use. We proposed the algorithms which consider only identifying attributes or sensitive attributes and providing k anonymity or the Anatomy algorithm.

N. Mohammed, et al. [13] in the paper 'Centralized and Distributed Anonymization for High Dimensional', has been proposed methods for protecting privacy. Initially all values in QID are generalized to the topmost value in their taxonomy trees and contains the topmost value for each attribute. We propose a new privacy model, in conjunction with data anonymization algorithms, to effectively preserve individuals' privacy and meet the information requirements specified by the BTS.

III. Existing System

In existing system, the collaborated data is to be masking with anonymized data. Through the collaboration function, end user can get more privacy about their data. In existing system the insider attack can be overcome with provider aware anonymization algorithm. With this algorithm provider also cannot act as an attacker and they cannot hack collaborated user data. Here, large amount of data cannot be processed in this system. There may be possibilities to hack small amount of user data. Time efficiency is low to produce result.

IV. Proposed System

In proposed system, Provider aware algorithm and a Novel High dimensional top down specialization (HDTDS) is addressed. Provider aware algorithm will work like an existing system. But HDTDS algorithm is an efficient and scalable algorithm. In this

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

proposed system, it can handle large amount of data when compare with the existing system. Attackers cannot guess user data because large amount of database is handled in proposed system. With use of this both algorithms users data could be protected and the data privacy also protected in this project.

High-Dimensional Top-Down Specialization (HDTDS) is an efficient and scalable algorithm HDTDS achieving LKC-privacy on high-dimensional relational data.LKC- privacy means, it is a general privacy model that stop both records linkage and attribute linkage that is the privacy model is applicable to anonymize data, with or without sensitive attributes.Here, trusted third party can handle and manage large number of databases and data sets. It computes and produces large number of anonymized data to provider and end user. Execution time is reduced and computation performance is increased. Result is efficient when compare with existing system.

A. Gathering User Details

In this module user details are collected by the various providers. For example if patients have to take treatment, he/she should register their details. Details of the customer like Name, Age, and Disease by which they get affected, Emai id etc. These details are maintained in a Database by the different kind of providers like hospital. Only concern providers can see all their details. Existing users can only see their own record.

B. Providers Control

In this module providers can see all the patients details and will get the background knowledge (BK) by the chance providers will see horizontally partitioned data of distributed data base of the group of providers. Providers can also see how many users are affected without knowing of individual records of the users and sensitive information about the individuals.

C. Attacks by Data Providers

Each data provider can also use anonymized data (T∗) and his own data (T1) to infer additional information about other records. Compared to the attack by the external recipient in the first attack scenario,each provider has additional data knowledge of their own records, which can help with the attack. This issue can be further worsened when multiple data providers collude with each other.

D. Collaborate Data

This module collaborates the data. The data gathered from all data providers. Here, trusted third party perform anonymization of the whole data and sensitive data are filtered here.

V. Experimental Results

We present two sets of experiment results with the following goals: (1) To compare and evaluate the privacy level for the given details. (2) To evaluate and compare the proposed algorithms for the given user details in terms of utility and efficiency. The proposed algorithm provide high privacy level than the existing binary and top-down algorithms.The HDTDS algorithm supports large number of providers than the existing algorithms in less time.

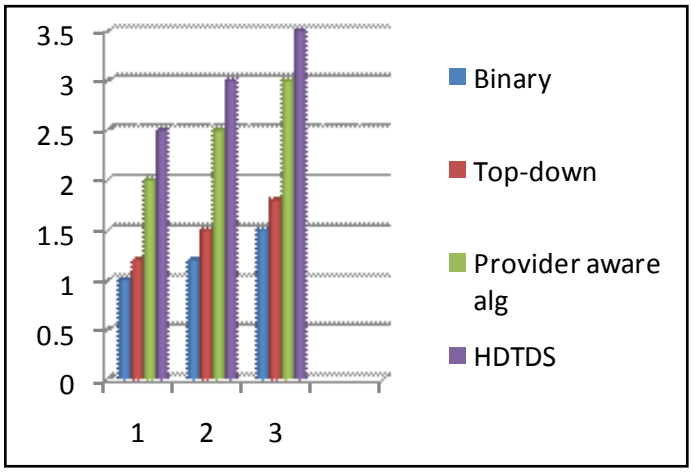
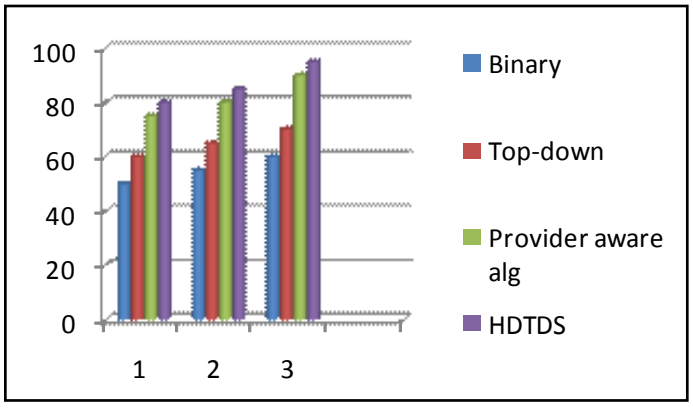


Fig. 1: Privacy Level Vs. Time



No.of.Providers vs. Time

VI. Conclusion and Future Work

In this project, we considered a new type of potential attackers in collaborative data publishing. A coalition of data providers, called adversary. To prevent privacy disclosure by any adversary we showed that guaranteeing privacy is enough. We presented heuristic algorithms exploiting equivalence group monotonicity of privacy constraints and adaptive ordering techniques for efficiently checking privacy. We introduced also a provider-aware anonymization algorithm with adaptive privacy checking strategies to ensure high utility and privacy of anonymized data. Our experiments confirmed that our approach achieves better or comparable utility than existing algorithms while ensuring privacy efficiently. There are many remaining research questions. Defining a proper privacy fitness score for different privacy constraints is one of them. It also remains a question to address and model the data knowledge of data providers when data are distributed in a vertical or ad-hoc fashion. It would be also interesting to verify if our methods can be adapted to other kinds of data such as set valued data.

References

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

[2] *Dwork, C. "Differential privacy: a survey of results," in Proc. of the 5th Intl. Conf. on Theory and Applications of Models of Computation, 2008,pp. 1–19.*

[3] *Dwork, C. "A firm foundation for private data analysis," Commun. ACM, vol. 54, pp. 86–95, January 2011.*

[4] *Fung, B.C.M. Wang, K. Chen, R. and Yu, P.S. "Privacy-preservingdata publishing: A survey of recent developments," ACM Comput. Surv.,vol. 42, pp. 14:1–14:53, June 2010.*
[5] *Goldreich, O. Foundations of Cryptography: Volume 2, Basic Applications. Cambridge University Press, 2004.*
[6] *Jiang, W. and Clifton, C. "Privacy-preserving distributed k-anonymity," in Data and Applications Security XIX, ser. Lecture Notes in Computer Science, 2005, vol. 3654, pp 924–924.*
[7] *Jiang, W. and Clifton, C. "A secure distributed framework for achieving k-anonymity," VLDB J., vol. 15, no. 4, pp. 316–333, 2006.*
[8] *Kifer, D. "Attacks on privacy and definetti's theorem," in Proc. of the 35th SIGMOD Intl.Conf. on Management of Data, 2009, pp. 127–138.*
[9] *Kifer, D. and Machanavajjhala, A. "No free lunch in data privacy," in Proc. of the 2011 Intl.Conf. on Management of Data, 2011, pp. 193-204.*
[10] *Li, N. and Li, T. "t-closeness: Privacy beyond k-anonymity and ldiversity," in In Proc. of IEEE 23rd Intl. Conf. on Data Engineering (ICDE), 2007.*
[11] *Zhong, S. Yang, Z. and Wright, R.N. "Privacy-enhancing k-anonymization of customer data",In Proc. Of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 2005, pp. 139-147.*
[12] *Cormode, G. Srivastava, D. Li, N. Li, T. "Minimizing minimality and maximizing utility: analyzing method-based attacks on anonymized data", Proc. VLDB Endow., vol. 3, Sept. 2010.*
[13] *Mohammed, N. Fung, B.C.M. Hung, P.C.K. and Lee, C. "Centralized and distributed anonymization for high-dimensional healthcare data", ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 4, no. 4, pp. 18:1-18-33, October 2010.*

Author's Profile

G.Sivaranjani had received her B.Tech degree in the department of Information Technology from Sethu Institute of Technology. She pursuing her M.E degree in Computer Science and Engineering from Sethu Institute of Technology, India. Her research interest in Datamining.