ISSN:0975-9646

Preserving Privacy and Data Security in Database

Adesh Chaudhary , Krishna Pratap Rao , Prashant Johri

School of Computing Science
Galgotias University
Greater Noida, India

Abstract - Database mining is defined as the process of extracting for useful, previously unknown, and indirect information from vast amount of data by efficient knowledge discovery techniques. Naturally such a process may open up new conclusion channels, detect new intrusion technique, and create new security issue. New security issues and research problems are detected and identified. Finally an individually well-developed theory, rough set theory, has been discussed and some needed applications to security problems are discussed

Keywords - Database mining, database security, rough sets, inferences, intrusion detection, association rule mining.

.

I. INTRODUCTION

The troupe of the information by governments, private organization, and individuals has created enormous chance for knowledge-based decision making. Controlled by interactive interest, or by continuity that require certain data to be published, there is a demand of exchange of data between different organizations. For exam-Kaislash hospitals in India are required to submit precise demographic data on every patient discharged from their ward. In the month of June 2004, the Information Technology Advisory Committee released a report a Revolution on Health Care with the help of Information Technology. One of its main points was to establish a worldwide system of electronic medical records that elaborate the exchange of medical knowledge through computer. Data publishing is equally especially at the same time in some other domains. For example, Bigflix, a popular online movie rental service, recently published a survey which contain a data set containing movie ratings of 300,000 subscribers, in a drive to examine the review of movie recommendations based on personal views. Complex person-specific data in its real format often contains important and sensitive information about individual one, and reveal such data immediately disclose individual privacy. The running task mainly lies on conditions and guidelines to block the types of reveal data and on lows on the use and storage of sensitive data. The restriction of this approach is that it either damage data vastly or need a trust level that is high in many data-sharing schemes. For example, license and agreements cannot assure that sensitive data will not be wrongly misplaced and end up in the fake hands. A task with the serious importance is to develop technique andtools for publishing information in a more combative environment, so that the published data remains useful while personal privacy remain safe. This agreement is called privacy-preserving data publishing (PPDP). In the recent few years, research communities became active to this issue and proposed many technique. While the research field is still curiously developing, it is a great time to handle the limitation and discuss feature of PPDP, justifying the differences and requirements that define PPDP from other revealed problems, and systematically summarize and evaluation of different approaches to PPDP. This survey aims to acquire these goals.

II. DATA MINING WITH SECURITY ASPECTS

This work does not have any security issues; truly it has just the other goal to give content-based metadata (The data about data) for the predict huge data holding of the Earth Observing System Data and Information System (EOSDIS) [1]. This pleased metadata would then be used to oblige scientists in searching data of interest from the EOSDIS data captured. Although this Project has no security issue, it provides an important example of data mining in a scientific-data domain. The problem labeled by the UAH data mining research is that the quantity of data in scientific data collection is growing. Some calculated project that projects such as EOSDIS will grow up to two terabyte of data per day Scientists need to be able to find data of interest. The problem of searching of data is typical since there is a short of content-based metadata. The typical metadata available in the currently operated Version 0f EOSDIS system is fixed to satellite, sensor and date captured. All of these present non- content-based metadata. The availability content-based metadata on some data values are browse picture which gives a limited resolution view of one of the channels of the data. But this cannot be automatically forwarded.

A. Inference Problem

Data mining is an approach to answer the long-standing problems "what does all this data mean?" Such study are existing attempt to start the "inference problem" in data security. Inference is generally the process of creating the connection between data sets, the common objective as data mining. That is, given that certain point apply to a set of data, we "know" that certain other points also apply to that set of data. This is equal to positioning that one set "implies" the other. As, in a multi-level secure (MLS) database, we are against of Low-classified data over High-classified data. Data mining processes cannot be used to settlement these kind of rules this is because each DM process should operate at a fixed level (i.e. Low) and must have eligibility over the High data in order to "create" the rule. Although, such Low-to-High rules may be "common information" but unknown to the database developer. Data mining can be used to correlate Low information until the

back of the common-information rule is derived. This is known as the inference. Willingly, data mining can be used effectively to ensure security. The simplest way is to search the rules which correlate Low and High data with each other. The security officer studying the advantages over a hackers, until he/she has reached to both the High and Low data. Generally in all the systems, there is less High data, so the number of rules correlate High data to Low data is much lesser than that of total number of actual possible rules.

B. Rough Sets And Data Mining

The theory of rough sets, given by Zdzislaw Pawlak, has been grow up fastly in the past few years, and has comes into the technology. This adoption of new technology bother the ordered study of indefinite, unsure or not complete information. In the recent years, it has been expressed that rough set theory [8] is a very accurate technique for data study and generating rules in the sets-value based domains. It is an effective tool for data mining in relational databases.

C. Privacy Preserving Data Publishing

A typical framework for data gathering and production is described in Figure 1. In the data gathering phase, the data publisher assemble data from defined owners [2]. In the data publishing phase, the data publisher leave the gathered data to a data collector or to the public, called the data receiver, who will then organize data mining on the published data. In this survey, data mining has a vast sense, not rarely restricted to frame mining. For example, a hospital gather data from sick person and publishes the sick person records to another medical center. In this example, the hospital is the data publisher, sick persons are record owners, and the medical center is the data collector. The data mining started at the medical center can be anything from an easy count of the number of person with cancer to a practiced group analysis. There are number of models of data publishers [3]. In the fake model, the data publisher is untrusted and may attempt to clarify sensitive information from record counter. Different cryptographic solutions [4]; unidentified communications [5, 6]; and s=graphical methods [7] were proposed to count records innominately from their counters without conceding the owners 'status. In the original model, the data publisher is trusty and record counter are to provide their personal knowledge to thedata publisher; although, the trust is not clear to the data collector.

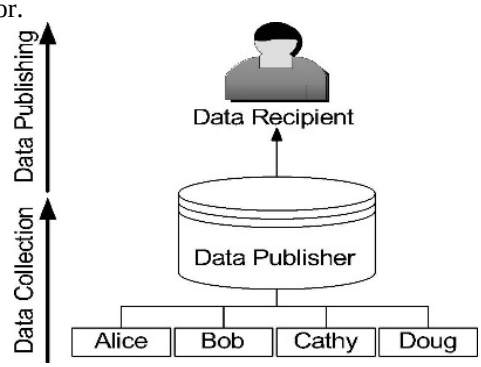

Fig. 1. Data collection and data publishing

III. PRIVACY SUGGESTION

We have to find out what is privacy before we take a close look at the privacy proposal of data mining and forward effective solutions. Generally various societies have different thinking of privacy. In the case of the medical department, privacy is around a patient finding out what treatment the doctor should dispense about him. Generally employees, sales and insurance companies can try to find out information about peoples. It is up to the person to check out the details to be given about him. In the fiscal society, a bank customer finds out what fiscal characteristics the bank should provide about him/her. In addition with, retail organization should not be providing the expenditure details about the persons unless the individuals have accede the release. In the case of the government organization, privacy may get a completely new seriousness. Let us consider an example, the students who take my classes at AFCAT have pointed out to me that IB would collect data about Indian citizens. As we know IB finds out what data about an Indian citizen it can able to say the RAW. That is, the IB has to be ensure the privacy of Indian citizens. Additionally, allow access to once travel and prodigal data as well as his/her web surfing activities should also be given upon receiving access from the individuals once. Now that we have defined what we classify by privacy, we will now checkup the privacy suggestion of data

Mining. Data mining provides us "actuality" that are not clear to human surveyors of the data. For can we extract out highly confidential relations from public data? In the previous case we require to secure the person data entries while notifying the collaborating or collecting while in the last case we need to protect the correlations between the data.

IV. GROWTH IN DATA PRIVACY

Various types of privacy problems have been examined by researchers. We will point out these problems and the solutions estimated [10].

A- *Problem***:**Privacy violation that outcome due to data mining: In this case the option is Privacy protecting data mining. It means, we conduct data mining and give out the results without notifying the data entries used to execute data mining**.**

B- *Problem***:** Privacy infringement resulted due to the speculation problem. Record that speculation is the way of perceive sensitive data entry from the legal replies awarded to user inquiries. The solution to this problem is PrivacyConstraint Processing.

C- *Problem:*Privacy transgression due to non-encrypted data: the solution of this problem is to make use of Encryption at various levels.

D- *Problem:*Simple Constraint: Some feature of a document is private. Content footed constraint: If document contain information about X, then it will be private**.**

V. CONCLUSION

In this paper we have studied data mining approaches in security and their intimation for privacy. We have studied the design of privacy and then discussed about the growth those on privacy preserving data mining. After it we proposed a conclusion for further study on privacy and data mining. Here are our report. There is no fixed definition for privacy, each firm must clear-cut what it shows by privacy and build accurate privacy techniques. Technology only is not satisfactory for privacy; we require Technician, expert, legalsurveyors and Social researchers to effort on Privacy. Some well-knownpeople confidently say 'Forget about privacy" Therefore, should we do deep study on Privacy? We believe that there are fetching research problems; therefore we need further study on this research. Furthermore, something is better than nothing. Another school of consideration is attempted to discard privacy demolition and if demolition take place then put on proceeding. We need to undertake privacy from all directions**.**

REFERENCES

1. Database Security IX Status and Prospects Edited by D. L. Spooner, S. A. Demurjian and J. E. Dobson ISBN 0 41272920 2, 1996, pp. 391-399.
2. ACM Computing Survey Vol 42, No. 4, Article 14 Publication Date June 2010.
3. GEHRKE, J. 2006. Models and methods for privacy-preserving data publishing and analysis. Tutorial at the 12th ACM SIGKDD
4. YANG, Z., ZHONG, S., AND WRIGHT, R. N. 2005. Anonymity-preserving data collection. In *Proceedings of the11th ACM SIGKDD Conference*. ACM, New York, 334–343.
5. CHAUM, D. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Comm. ACM 24*, 2, 84–88.
6. JAKOBSSON, M., JUELS, A., AND RIVEST, R. L. 2002. Making mix nets robust for electronic voting by ran-domized partial checking. In*Proceedings of the 11th USENIX Security Symposium*. 339–353
7. WARNER, S. L. 1965. Randomized response: A survey technique for eliminating evasive answer bias. *J.Am. Statistical Assoc. 60*, 309, 63–69.
8. WARNER, S. L. 1965. Randomized response: A survey technique for eliminating evasive answer bias. *J.Am. Statistical Assoc. 60*, 309, 63–69.
9. Pawlak, Z. (1990). Rough sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1992
10. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 1, ISSUE 7, AUGUST 2012