Indian J.Sci.Res. 17(2): 146-150, 2018 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

AN APPRAISAL OVER DATA MINING PRIVACY CONSTRAINTS AND SUGGESTIONS

[1]Venkatesh V, [2]ANKPrasannanjaneyulu, [3]V. Madhavi
[1]Senior Software Engineer, Progress Software Development PVT LTD.
[2] Faculty(IT), Institute of Insurance and Risk Management (IIRM), Gachibowli, Hyderabad
[3]Manager 1, TSCAB Bank, Hyderabad

Abstract- Data mining has been gaining popularity in knowledge discovery field, particularity with the increasing availability of digital documents in various languages from all around the world .In this paper we first look at data mining applications in safety measures and their suggestions for privacy. After that we then inspect the idea of privacy and give a synopsis of the developments particularly those on privacy preserving data mining. Data mining has several applications in protection, there are also serious privacy fears. Because of data mining, even inexperienced users can connect data and make responsive associations. Therefore we must to implement the privacy of persons while working on practical data mining. In this paper we will talk about the developments and instructions on privacy and data mining. We then present an outline for research on confidentiality and data mining.

Keywords **-** Data mining, security, safety, security suggestions, preserving data mining, data mining applications

I. Introduction

DATA mining is the procedure of posing questions and taking out patterns, often in the past mysterious from huge capacities of data applying pattern matching or other way of thinking techniques. Data mining has several applications in protection together with for national protection as well as for cyber protection. The pressure to national protection includes aggressive buildings, demolishing dangerous infrastructures such as power grids and telecommunication structures. Data mining techniques are being examined to realize who the doubtful people are and who is competent of functioning revolutionary activities. Cyber security is concerned with defending the computer and network systems against fraud due to Trojan cattle, worms and viruses. Data mining is also being useful to give solutions for invasion finding and auditing. While data mining has several applications in protection, there are also serious privacy fears. Because of data mining, even inexperienced users can connect data and make responsive associations. Therefore we must to implement the privacy of persons while working on practical data mining. In this paper we will talk about the developments and instructions on privacy and data mining. In particular, we will give a general idea of data mining, the different types of threats and then talk about the penalty to privacy. This paper is organized as follows. Section 2 talks about data mining for safety applications. Section 3 explains the overview of privacy. Section 4 discusses different aspects of data mining on. Directions are provided in section 5 and section 6 gives the conclusion of this paper or work done on thepaper. Dataminingis fitting a key technology for identifying doubtful activities. In this section, data mining will be discussed with respect to use in both ways for non-real-timeandfor real-time applications. In order to complete data mining for counter terrorism applications, one wants to gather data from several sources. For example, the subsequent information on revolutionary attacks is wanted at the very least: who, what, where, when, and how; personal and business data of the possible terrorists: place of birth, religion, education, ethnic origin, work history, finances, criminal record, relatives, friends and associates, and travel history; unstructured data: newspaper articles, video clips, dialogues, e-mails, and phone calls. The data has to be included, warehoused and mined. One wants to develop sketches of terrorists, and activities/threats.

II. DATA MINING FOR SAFETY APPLICATIONS

Data mining is fitting a key technology for identifying doubtful activities. In this section, data mining will be discussed with respect to use in both ways for non-real-timeandfor real-time applications. In order to complete data mining for counter terrorism applications, one wants to gather data from several sources. For example, the subsequent information on revolutionary attacks is wanted at the very least: who, what, where, when, and how; personal and business data of the possible terrorists: place of birth, religion, education, ethnic origin, work history, finances, criminal record, relatives, friends and associates, and travel history; unstructured data: newspaper articles, video clips, dialogues, e-mails, and phone calls. The data has to be included, warehoused and mined. One wants to develop sketches of terrorists, and activities/threats. The data has to be mined to take out patterns of possible terrorists and forecast future activities and goals. Fundamentally one wants to find the “needle in the haystack” or more suitably doubtful needles among probably millions of needles. Data integrity is essential and also the methods have to SCALE. For several applications such as urgent situation response, one needs to complete real-time data mining. Data will be incoming from sensors

[1]Corresponding Author

and other strategy in the form of nonstop data streams together with breaking news, videocassette releases, and satellite images. Some serious data may also exist in caches. One wants to quickly sift through the data and remove redundant data for shortly use and analysis (non-real-time data mining). Data mining techniques require to meet timing restriction and may have to stick the quality of service (QoS) tradeoffs amongfor the longer-term, we require a research and development diagrams. In summary, data mining is very helpful to resolve security troubles. Tools could be utilized to inspect audit data and flag irregular behavior. There are many latest works on applying data mining or cyber safety applications, Tools are being examined to find out irregular patterns for national security together with those based on categorization and link analysis. Law enforcement is also using these kinds of tools for predictable future? What are the standards for achievement? How do we assess the data mining algorithms? What test beds do we construct? We require both a near-term as well as longer-term resolutions. For the future, we require to influence present efforts and fill the gaps in a objective aimed way and complete technology transfer.

For the longer-term, we require a research and development diagrams. In summary, data mining is very helpful to resolve security troubles. Tools could be utilized to inspect audit data and flag irregular behavior. There are many latest works on applying data mining for cyber safety applications, Tools are being examined tofindoutirregular patterns for national security together with those based on categorization and link analysis. Law enforcement is also using these kinds of tools for fraud exposure and crime solving.

III. Privacy suggestions

We require finding out what is meant by privacy before we look at the privacy suggestions of data miningandrecommend efficient solutions. In fact different society-ties have different ideas of privacy. In the case of the medical society, privacy is about a patient finding out what details the doctor should discharge about him/her. Normally employers, marketers and insurance corporations may try to find information about persons. It is up to the individuals to find out the details to be given about him. In the monetary society, a bank customer finds out what financial details the bank should give about him/her. Additionally, retail corporations should not be providing the sales details about the persons unless the individuals have approved the release. In the case of the government society, privacy may get a whole new significance. For example, the students who attendmy classes at AFCEA have pointed out to me that FBI would gather data about US citizens. However FBI finds out what data about a US citizen it can provide to say the CIA. That is, the FBI has to make sure the privacy of US citizens. Additionally, permitting access to individual travel and spending data as well as his/her web surfing activities should also be provided upon receiving permission from the individuals. Now that we have explained what we signify by privacy, we will now checkupthe privacy suggestion of data mining. Data mining provides us “facts” that are not clear to human analysts of the data. For instance, can general tendency across individuals be calculated without enlightening details about individuals? On the other hand, can we take out highly private relations from public data? In the former case we require to protect the person data values while enlightening the associations or aggregation while in the last case we need to defend the associations and correlations between thedata.

IV. Growth Inprivacy

Different types of privacy problems have been considered by researchers. We will point out the various problems and the solutions projected.

A. Problem: Privacy contraventions that consequence due to data mining: In this case the way out is Privacy protecting data mining. That is, we perform data mining and give out the results without enlightening the data values used to perform data mining.

B. Problem: Privacy contraventions that result due to the Inference problem. Note that Inference is the procedure of realizing sensitive data details from the lawful answers received to user inquiries. The way out to this problem is Privacy ConstraintProcessing.

C. Problem: Privacy contravention due to un-encrypted data: the way out to this problem is to make use of Encryption at differentlevels.

D. Problem: Privacy contravention due to poor system design. Here the way out is to build up methodology for designing privacy-enhanced systems. Below we will observe the ways out projected for both privacy constraint/policy processing and for privacy preserving data mining. Privacy limitation or policy processing research was carried out by [8] and is footed on some of her prior research on security restriction processing. Instance of privacy restrictions include thefollowing.

E. Simple Constraint: an aspect of a document is private. Content footed constraint: If document holds information about X, then it is private.

F. Association-based Constraint: Two or more documents used together are private; individually each document ispublic.

G. Free constraint: After X is freed Y becomes private. The way out projected is to augment a database system with a privacy checker for constraint processing. During the inquiry process, the constraints are checked up and only the public information is freed unless certainly the

user is approved to obtain the private information. Our approach also contains processing constraints during the database update and design operations. For details we refer to[8].

Some early work on managing the privacy problem that consequence from data mining was performed by Clifton at the MITRRE Corporation [9].

The suggestion here is to avoid useful outcomes from mining. One could initiate "cover stories" to provide "false" outcomes. Another approach is to only build a sample of data existing so that a challenger is not capable to come up with helpful rules and analytical functions. However these approaches did not impression as itbeaten the idea of data mining. The objective is to perform effective data mining but at the same time guard individual data values and sensitive relations. Agrawal was the first to invent the word privacy preserving data mining. His early work was to initiate random values into the data or to bother the data so that the real data could be confined. The challenge is to initiate random values or agitate the values without touching the data mining results [1]. Another new approach is the Secure Multi-party Computation (SMC) by KantarciogluandClifton [3]. Here, each party knows its individual contribution but not the others' contributions.

Additionally the final data mining outcomes are also well-known to all. Various encryption techniques utilized to make sure that the entity values are protected. SMC was demonstrating several promises and can be used also for privacy preserving scattered data mining. It is probably safe under some suppositions and the learned models are correct; It is assumed that procedures are followed which is a semi truthful model. Malicious model is also investigated in some current work by KantarciogluandKardes [4]. Many SMC footed privacy preserving data mining algorithms contribute to familiar sub-protocols (e.g. dot product, summary, etc.). SMC does have any disadvantage as it's not competent enough for very large datasets. (E.g. petabyte sized datasets); Semi- honest model may not be reasonable and the malicious model is yet slower. There are some novel guidelines where novel models are being discovered that can swap better between efficiency and security.

Game theoretic and motivation issues are also being discovered. Finally merging anonimization with cryptographic techniques is also a route. Before performing an evaluation of the data mining algorithms, one wants to find out the objectives. In some cases the objective is to twist data while still preserving some assets for data mining. Another objective is to attain a high data mining accuracy with greatest privacy protection. Our current work imagines that Privacy is a personal preference, so should be individually adjustable. That is, we want to make privacy protecting data mining approaches to replicate authenticity. We examined perturbation based approaches with real-world data sets and provided applicability learning to the existing approaches [5]. We found that the rebuilding of the original sharing may not work well with real-world data sets. We attempted to amend perturbation techniques and adjust the data mining tools. We also developed a new privacy preserving decision tree algorithm [6]. Another growth is the platform for privacy preferences (P3P) by the World Wide Web association (W3C). P3P is an up-and-coming standard that facilitates web sites to convey their privacy practices in a typical format. The format of the strategies can be robotically recovered and appreciated by user agents. When a user comes in a web site, the privacy policies of the web site are communicated to the user; if the privacy policies are dissimilar from user favorites, the user is notified; User can then make a decision how to continue. Several major corporations are working on P3P standards.

V. Directions Forprivacy

Thuraisinghamverifiedin 1990 that the inference problem in common was unsolvable; therefore the suggestion was to discover the solvability features of the problem [7].We were able to explain comparable results for the privacy problem. Therefore we need to inspect the involvement classes as well as the storage and time complication. We also need to discover the base of privacy preserving data mining algorithms and connected privacy ways out. There are various such algorithms. How do they evaluate with each other? We need a test bed with practical constraints to test the algorithms.

Is it meaningful to observe privacy preserving data mining for each data mining algorithm and for all application? It is also time to enlarge real world circumstances where these algorithms can be used. Is it possible to build up realistic commercial products or should each association get used to products to suit their needs? Investigative privacy may create intelligence for healthcare and monetary applications. Does privacy work for Defense and Intelligence purposes? Is it even important to have privacy for inspection and geospatial applications? Once the image of my home is on Google Earth, then how much isolation can I have? I may wish for my position to be private, but does it make sense if a camera can detain a picture of me? If there are sensors all over the position, is it important to have privacy preserving surveillance.

This proposes that we require application detailed privacy. Next what is the connection between confidentiality, privacy and faith? If I as a user of Association A send data about me to Association B, then imagine I read the privacy policies imposed by Association B. If I agree to the privacy policies of Association B, then I will drive data about me to Association B. If I do not concur with the policies of association B, then I can bargain with association B. Even if the website affirms that it will not distribute private information with others, do I faith the website? Note:

while secrecy is enforced by the association, privacy is strong-minded by the user.

Therefore for confidentiality, the association will conclude whether a user can have the data. If so, then the association can additional decide whether the user can be trusted. Another way is how can we make sure the confidentiality of the data mining procedures and outcome? What sort of access control policies do we implement? How can we faith the data mining procedures and results as well as authenticate and validate the results? How can we join together confidentiality, privacy and trust with high opinion to data mining? We need to check up the research challenges and form a research schema. One question that RakeshAgrawal inquired at the 2003 SIGKDD panel on Privacy [2] "is privacy and data mining friends or rivals? We think that they are neither associates nor rivals. We need progresses in both data mining and privacy. We require planning flexible systems. For some applications one may have to hub entirely on "pure" data mining while for some others there may be a need for "privacy-preserving" data mining. We need flexible data mining techniques that can settle in to the changing environments. We consider that technologists, legal specialists, social scientists, policy makers and privacy advocates MUST worktogether.the idea of data mining. The objective is to perform effective data mining but at the same time guard individual data values and sensitive relations. Agrawal was the first to invent the word privacy preserving data mining. His early work was to initiate random values into the data or to bother the data so that the real data could be confined. The challenge is to initiate random values or agitate the values without touching the data mining results [1]. Another new approach is the Secure Multi-party Computation (SMC) by KantarciogluandClifton [3]. Here, each party knows its individual contribution but not the others' contributions. Additionally the final data mining outcomes are also well-known to all. Various encryption techniques utilized to make sure that the entity values are protected. SMC was demonstrating several promises and can be used also for privacy preserving scattered data mining. It is probably safe under some suppositions and the learned models are correct; It is assumed that procedures are followed which is a semi truthful model. Malicious model is also investigated in some current work by KantarciogluandKardes [4].

Many SMC footed privacy preserving data mining algorithms contribute to familiar sub-protocols (e.g. dot product, summary, etc.). SMC does have any disadvantage as it's not competent enough for very large datasets. (E.g. petabyte sized datasets); Semi- honest model may not be reasonable and the malicious model is yet slower. There are some novel guidelines where novel models are being discovered that can swap better between efficiency and security. Game theoretic and motivation issues are also being discovered.

Finally merging anonimization with cryptographic techniques is also a route. Before performing an evaluation of the data mining algorithms, one wants to find out the objectives. In some cases the objective is to twist data while still preserving some assets for data mining. Another objective is to attain a high data mining accuracy with greatest privacy protection. Our current work imagines that Privacy is a personal preference, so should be individually adjustable. That is, we want to make privacy protecting data mining approaches to replicate authenticity. We examined perturbation based approaches with real-world data sets and provided applicability learning to the existing approaches [5].

We found that the rebuilding of the original sharing may not work well with real-world data sets. We attempted to amend perturbation techniques and adjust the data mining tools. We also developed a new privacy preserving decision tree algorithm [6]. Another growth is the platform for privacy preferences (P3P) by the World Wide Web association (W3C). P3P is an up-and-coming standard that facilitates web sites to convey their privacy practices in a typical format. The format of the strategies can be robotically recovered and appreciated by user agents. When a user comes in a web site, the privacy policies of the web site are communicated to the user; if the privacy policies are dissimilar from user favorites, the user is notified; User can then make a decision how to continue. Several major corporations are working on P3P standards.

VI. Security Analysis

The evolution of the computer system and the development of new technologies, the attacks become increasingly e efficient. For these reasons, Kerberos has known several modifications to the levels of performance and functionality against these attacks. However Kerberos V5, the current version, with all its amelioration, was discussed by several security analysis, those show its weaknesses specifically against the dictionary attack only in the communication phase.

In this section, we evaluate the security of our proto-col by analyzing the level of influence using addition salt to the password, and the impact of the Diffie Hellman principle against different types of attacks. Further, we discuss the impact of adding dynamic salt per session to the password in both client side and KDC server side. In the client side, the addition of a dynamic salt per session to the password, and the application of virtualization function make the authentication process by password un-breakable. They reduce the chance of password divination attacks such as brute force and dictionary attacks. In the other hand, storing the password footprint (dynamic pass-word disturbed by salt in our case) is stronger than storing the clear password in KDC database server. This makes storage of password more reliable in the KDC server side.

A. Impact of Salt upon Password

The majority of the applications users are conscious of authentication by passwords. It requires the storage of simple passwords in most cases. In parallel, other authentication alternatives have been proposed. However, their use is too limited especially in web applications The description of Kerberos integrated a static salt (client address or the domain name) to disrupt the password used for the generation of encryption keys .

This technique does not solve the problem of dictionary attack that represents a real challenge against the Kerberos authentication techniques. To ad-dress this type of attack, our approach is based on theTable 1: Comparison between our protocol and previous versions of Kerberos

Parameters	Previous version of Kerberos	Our protocol
Mutual authentication	OK	OK
Portability	OK	OK
Use of ticket	OK	OK
Use of expiration time	OK	OK
Use of De e Helman		OK
salt	static and per user	dynamic and per session
Session key	Ks = H(pwd)	pwv = f(pwd) and Ks = H(pwvjjsalti)
Based key	based on string-to-key function	based on S2KexS function
Derived key	based on derived key function	based on DKexS function
N		New authentication number calculated from password

VII. Conclusion

In this paper we have examined data mining applications in security and their implications for privacy. We have examined the idea of privacy and then talked about the developments particularly those on privacy preserving data mining. We then presented an agenda for research on privacy and data mining. Here are our conclusions. There is no collective definition for privacy, each organization must clear-cut what it indicates by privacy and develop suitable privacy policies. Technology only is not adequate for privacy; we require Technologists, Policy expert, Legal experts and Social scientists to effort on Privacy. Some well acknowledged people have believed 'Forget about privacy" Therefore, should we follow research on Privacy? We trust that there are attractive research problems; therefore we need to carry on with this research. Additionally, some privacy is better than nil. One more school of consideration is tried to avoid privacy destructions and if destructions take place then put on trial. We need to put into effect suitable policies and checkup the legal aspects. We need to undertake privacy from all directions.

References

[1] Agrawal, R., Srikant, R.: Privacy-Preserving Data Mining. In: SIGMODConference, pp.439–450 (2000)

[2] Agrawal, R.: Data Mining and Privacy: Friends or Foes. In: SIGKDD Panel(2003)

[3] Kantarcioglu, M., Clifton, C.: Privately Computing a Distributed k-nn Classifier. In: Bou-licaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) PKDD 2004. LNCS, vol. 3202,279–290. Springer, Heidelberg(2004)

[4] Kantarcioglu, M., Kardes, O.: Privacy-Preserving Data Mining Applications in the Mali-cious Model. In: ICDM Workshops, pp. 717–722(2007)

[5] Liu, L., Kantarcioglu, M., Thuraisingham, B.M.: The applicability of the perturbation based privacy preserving data mining for real-world data. Data Knowl. Eng. 65(1), 5–21(2008)

[6] Liu, L., Kantarcioglu, M., Thuraisingham, B.M.: A Novel Privacy Preserving Decision Tree. In: Proceedings Hawaii International Conf. on Systems Sciences(2009)

[7] Thuraisingham, B.: One the Complexity of the Inference Problem. In: IEEE Computer Se-curity Foundations Workshop (1990) (also available as MITRE Report, MTP- 291)

[8] Thuraisingham, B.M.: Privacy constraint processing ina Privacy enhanced database management system. Data Knowl. Eng. 55(2), 159–188 (2005)

[9] Clifton, C.: Using Sample Size to Limit Exposure to Data Mining. Journal of Computer Security 8(4)(2000)