Privacy Preserving Over Bigdata Using Mongodb

S.Nivetha[1],S.Nivethapriyadharshini[2],S.S.Shanthini[3],S.Shifnas[4],Dr.S.A.Sahaaya Arul Mary[5]
[1, 2] Dept Of Computer Science Engineering
[1, 2] Saranathan College Of Engineering
Trichy

Abstract- *Data mining, or knowledge discovery from data (KDD), aims to discover interesting patterns and knowledge from big data. Although its application has been largely successful, data mining can set up compromising situations for sensitive information, which is a serious privacy threat. The response to this problem has been significant enough to lead to privacy-preserving data mining (PPDM), the goal of which is to safeguard information from unsolicited or unsanctioned disclosure while preserving the data's utility. PPDM approaches aim to avoid the direct use of sensitive raw data, such as an individual's ID and cell phone numbers and attempt to exclude sensitive patterns in mining results, such as clues to undisclosed personal information derived from a consumer's shopping behavior. PPDM models and algorithms focus on the prevention of information disclosure during specific mining operations. As such, they can sometimes fail to consider privacy issues in other KDD stages, such as data preparation and the use of extracted patterns. For example, data preparation could expose the original data owners' identities and create vulnerabilities that lead to deliberate abuse of data patterns or unintentional inappropriate use—both of which can compromise individual privacy and even national security.*

Keywords- Windows/XP/7.Tomcat 5.0/6.0 , HTML,JavaScript,Java Server Pages,MongoDB,Robomongo-0.8.5-i386.

I. INTRODUCTION

Database security is a critical aspect of information security. Access to enterprise databases grants a attackers great control over critical data. For example, SQL injection a-ttacks insert malicious code into the statements the application passes to the database layer. - is enables a attackers to do almost anything with the data, including accessing unauthorized data and altering, deleting, and inserting data. Although SQL injection exploitation has declined steadily over the years owing to secure frameworks and improved awareness, it remains a high-impact means to exploit system vulnerabilities. For example, Web applications receive four or more Web attack campaigns per month, and SQL injections are the most popular a attacks on retailers. 1 Furthermore, SQL injection vulnerabilities affect 32 percent of all Web applications.

NoSQL (not only SQL) is a trending term in modern data stores; it refers to non relational databases that rely on different storage mechanisms such as document store, key-value store, and graph. - e wide adoption of these databases has been facilitated by the new requirements of modern large-scale applications, such as Facebook, Amazon, and Twitter, which need to distribute data across a huge number of servers. Traditional relational databases don't meet these scalability requirements; they require a single database node to execute all operations of the same transaction.

As a result, a growing number of distributed, NoSQL key-value stores satisfy the scalability requirements of modern large-scale applications. - These data stores include NoSQL databases such as MongoDB and Cassandra as well as in-memory stores and caches such as Redis and Memcached. Indeed, the popularity of NoSQL databases has grown consistently over the past several years, and MongoDB is ranked fourth among the 10 most popular databases, as Figure 1 illustrates. In this article, we provide an analysis of NoSQL threats and techniques as well as their mitigation mechanisms.

II. LITERATURE SURVEY

A. REVIEW

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

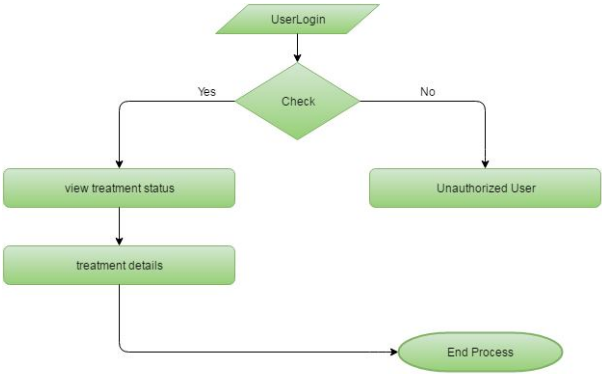
III. EXISTING SYSTEM

Most existing k-anonymity techniques apply the same privacy preservation for all individuals, ignoring the individual's privacy preferences. Some proposed k-anonymity methods aim to support personalized privacy preservation, either formulating a privacy preference as a specific parameter value the value of k in k-anonymity or having a specific node denote the preference in a domain-generalization hierarchy. Although these methods have a worthy goal, it is somewhat unrealistic to expect individuals to declare their privacy preferences in such a formal way. Researchers need to find practical ways to obtain personalized privacy preferences in k-anonymity techniques as well as in other PPDP algorithms

IV. PROPOSED SYSTEM

Data modification methods proposed for privacy-preserving classifications vary with the adopted classification models. One model considers the privacy threat of classification based on a support vector machine (SVM), which stems from the support vectors in the learned classifier. To destroy the sensitive information in support vectors, the model transforms the original decision function, determined by the support vectors, to an infinite series of linear monomial-feature combinations. Geometric transformations, such as translation, scaling, and rotation, are often applied to establish privacy-preserving clusters.

Modified data is often less useful, so data mining must balance privacy and data utility, thereby ensuring that nonsensitive information is still available. Because data types are becoming more complex and new applications continue to emerge, finding appropriate ways to quantify privacy and utility is still a high priority in future PPDM research. If the data miner must release the model learned from data (for example, an SVM classifier) to others, attackers might be able to infer sensitive information from the released model. Future work should look at what sensitive information can be inferred from the model's parameters, what background knowledge the attacker can use, and how to modify the learned model to prevent the sensitive-information disclosure. Individuals, not just data miners, can modify data to protect their privacy.



V. HARDWARE REQUIREMENTS

System	**- Pentium –IV 2.4 GHz**
Speed	- 1.1 Ghz
RAM	- 256MB (min)
Hard Disk	- 40 GB
Key Board	- Standard Windows Keyboard
Mouse	- Logitech
Monitor	- 15 VGA Color.

REFERENCES

[1] C.C. Aggarwal and P.S. Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," Privacy-Preserving Data Mining, C.C. Aggarwal and P.S. Yu, eds., Springer, 2008, pp. 11–52.

[2] B.C.M. Fung et al., "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, June 2010, pp. 14:1–14:53.

[3] K.-P. Lin and M.-S. Chen, "On the Design and Analysis of the Privacy-Preserving SVM Classifier," IEEE Trans. Knowledge and Data Eng., vol. 23, no. 11, 2011, pp. 1704–1717.

[4] S. Carter, "Techniques to Pollute Electronic Profiling," 26 Apr. 2007, US Patent App. 11/257,614; www.google.com /patents/US20070094738.

[5] Y. Guo, "Reconstruction-Based Association Rule Hiding," Proc. SIGMOD2007 PhD Workshop Innovative Database Research (IDAR 07), 2007, pp. 51–56.

[6] T. Mielikäinen, "On Inverse Frequent Set Mining," Proc. Workshop Privacy-Preserving Data Mining, 2003, pp. 18–23.

[7] Y.L. Simmhan, B. Plale, and D. Gannon, A Survey of Data Provenance Techniques, tech. report 47405, CS Dept., Indiana Univ., 2005.

[8] G. Barbier et al., "Provenance Data in Social Media," Synthesis Lectures on Data Mining and Knowledge

Discovery, vol. 4, Jan. 2013, pp. 1–84.

[9] A. Ghosh and A. Roth, "Selling Privacy at Auction," Proc. 12th ACM Conf. Electronic Commerce (EC 11), 2011, pp. 199–208.

[10] C. Riederer et al., "For Sale: Your Data by You," Proc. 10th ACM Workshop Hot Topics in Networks (HotNets 11), 2011, p. 13.

[11] H. Kargupta, K. Das, and K. Liu, "Multiparty, Privacy-Preserving Distributed Data Mining Using a Game Theoretic Framework," Proc. Knowledge Discovery in Databases (PKDD 07), 2007, pp. 523–531.

[12] R.K. Adl et al., "Privacy Consensus in Anonymization Systems via Game Theory," Proc. Data and Applications Security and Privacy XXVI (DBSec 12), 2012, pp. 74–89.