Volume 4, Issue 4, April 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Data Warehouse Security Using Log Based Analysis: A Review

Raj Rani
M Tech Student
Department of computer science and Application
Kurukshetra University, Kurukshetra
Haryana (India)

Abstract— Various techniques have been emerged in last few years to improve data privacy for security reasons. But these techniques involve extra overhead in case of large data warehousing system. Thus to remove extra overhead and to improve the security of sensitive information we involve the concept of masking and intrusion detection in data warehouse. Thus by introducing these approaches we can improve overall security strength of database. Intrusion detection in Log Analysis is the way to detect malicious attack in a particular environment using logs as the source of information. Thus in future there will be more scope in log based security techniques to increase the efficiency and security of data warehouse. Through the log analysis, it will help to recognize the unauthorized access in data warehouse. In this paper we discuss work done in the field of security in data warehousing system by log analysis. This paper provides a review on masking of hypersensitive data to prevent it from the unauthorized access and attacks.

Keywords— log analysis, DW, data intrusion, masking, log logic.

I. INTRODUCTION

The main source of historical data is mainly known as data warehouse. In log based security analysis log data have introduced which is collected and analyzed from legacy tools or manually [4]. A very large amount of data can be gathered in every second from various sources such as firewalls, routers, servers. The data warehouse store very large amount of valuable data which include sensitive information about any organization. Data Warehouse maintains very large amounts of credit card numbers, other personal information organization secrets and financial information. An efficient data warehouse makes us sure that sensitive data does not going into unauthorized hands [1]. Many security approaches have been emerged during last few years. In My SQLv5 the encryption algorithms provide the effective way of securing sensitive data but increase the data overhead problem [14]. After that Oracle has developed the Transparent Data Encryption in 11g versions [15]. The encryption has applied on column and table space. But these encryption schemes failed to provide efficient security with the increasing size of data warehouse. Log based security analysis have been emerged to overcome the problem of large overhead in the large data warehousing system. In log based security realistic-looking values based on masking rules have been replaced with sensitive information [2]. In this way the original data remains unchanged and protected. The main approach is used in log based security approach is modulus with masking formula. Masking formula introduced with modulus (division remainder) and simple arithmetic operations. This formula leads lowest computational overhead and, relatively minimum efforts in every response time, and provides highest level of security. As compare to encryption schemes log based scheme requires low overhead in storage space. Network bandwidth overflow is minimized, when user simply rewriting user queries. Sensitive information remains masked in database all the times, which allows only using the masked database for testing software production; directly querying the database will retrieve realistic data instead of real data. Our main goals in this paper are to deal with vulnerabilities and attacks in database [3].

In 2006 Intrusion detection systems provides a precious information on malicious behavior and detected attacks and intrusion prevention systems are used to stop malicious activities to take place Some intrusion detection systems, such as file integrity checking software, instead of running continuously progress periodically, so they generate log entries [5]. The log based security techniques involves three steps for the detection of malicious attacks in database which is a log collection, then collected logs analyzed and at end we check whether user is permitted to access data from the data warehouse. If not permitted then a message is generated for unauthorized attacks from malicious users [1].

II. TECHNIQUES USED IN DATA WAREHOUSE SECURITY

The main issue for data warehouse is security of the hypersensitive information that should be remains protected in warehouse. In data warehouse as quickly as the concerns evolve permissions on the warehouse must specify some constraints of the data owners, and information remains updated. Efficient protection of data warehouse information is very difficult task. Many techniques have been proposed for securing data warehouse. Oracle presents efficient mechanism to protect the sensitive data of data warehouses by using encryption techniques. It encrypts the stored information. The encryption can be applied only on column and table space. Source code does not need to be modified so

this technique is called transparent. My SQLv5 also provides an Advanced Encryption Standard [1]. The encryption mechanism proved efficient for securing sensitive information but fails to provide security in case of large database. The new enhancement that comes across the data security is the intrusion detection. It is based on two methods: attacks detection and misuse detection [10]. Normal and malicious behavior are difficult to distinguish. Detection accuracy is increased by data mining [11]. Data masking is efficient data security solution that secures the data by only changing the values not the original format. It replaces sensitive data with realistic data. Data Masking can be done by using many available techniques. The main objective of data masking is to prevent access or manipulation to sensitive data whatever the technique is used. Encryption can also be used as an advanced form of data masking. Sensitive data of data warehouse is replaced with some random values from available dataset in substitution technique. So an efficient mechanism based on mathematical modulus operator has been introduced to provide better execution times [7]. This new emerging technique in present day is log based security technique which provides the efficient mechanism for the detection of unauthorized access to sensitive information stored in data warehouse [1]. The main function in data security involved data warehouse is collection of data then analyzes the collected data on the basis of authentication and then alert the system case of attack or unauthorized access.

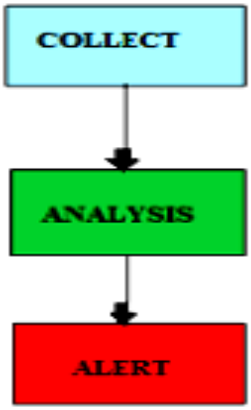

Fig 1: Flow of data.

For data warehouses that contain very private information, privacy preserving techniques will become more valuable.

III. LOG BASED SECURITY SYSTEM ARCHITECTURE

Various techniques have been used to prevent the sensitive information from malicious attacks in organizations. Data masking is new concept that is evolved in which format of original data remains same but only values changed this means that realistic data is replaced with sensitive information in database. By this means we prevent the original data from being changed [6].

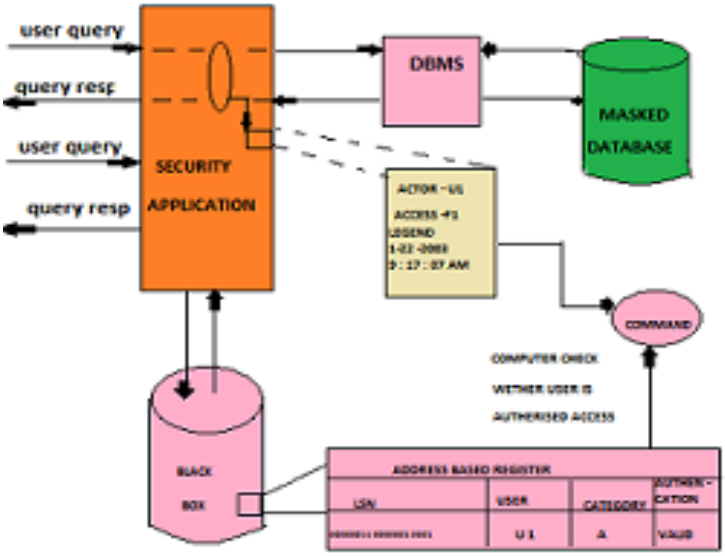

Fig 2: Log based system architecture

Data masking is mainly based on mathematical modulus operator. And the modulus based data masking techniques serves as the middle layer between sensitive information and users which make sure that data is manipulated securely [7].

Log based information is stored in black box dictionary of database server which includes set of files which is maintained separately for masked data [8]. In order to process the data in database, user applications need to send queries to security application. Only final results return to authorized users. In log based system architecture in order to process the data users need to send queries to middle ware layer MOBAT (modulus based security technique) and this security application stores the history logs. MOBAT application frequently keeps track and maintains history of the actions of every user and keeps the log created for each access in black box. In log based Security application generates three masking keys; one is public and two is private.

The user send request for authentication to access the sensitive data, the middleware security application receive the request from user, it rewrites the query and send it to process by DBMS and get the results, and at the end results send back to the requesting user. Thus in database, processed data remains protected at all times. Black Box maintains the predefined user policies which include access definitions.

IV. PROBLEMS EMERGING IN DATA SECURITY

In data warehouse there are various problems evolving with the increase in size of database. Thus the critical problem is how to automatically coordinate the access rights of the warehouse with those of the sources. To do so, one must be able to infer access rights across subsystems, without infringing on their local autonomy. This problem has not been addressed in systems to date [9]. The main issue related to the data warehouse is security of hypersensitive data which means to prevent the unwanted attacks from malicious users. Security techniques that we have been using during in last few years are facing many problems [7]. The encryption algorithms used for security of sensitive data is dealing with increased overhead problem with increasing size of data warehouse. When users query data, security becomes an issue. The data may be well protected in the data warehouse but a compromised user with full access to the data warehouse will certainly compromise all of the data. The data masking techniques overcome the overhead problem but it economically dealing with some difficulties [13].

V. CONCLUSION

Our review in data warehouse security will deal with several issues. With the increasing size of DWs containing very personal information, privacy preserving techniques will become more important. The main focus of this paper is on to distinguish the normal users from the malicious user. The masking techniques for security purpose becoming more efficient in order to preserve the sensitive data remains protected. In this review main focus is on the efficient security techniques that have been evolving in these days. And these security techniques provide efficient security mechanism for preserving the sensitive data remains unchanged. Data can be vulnerable when transmitted over non-secure networks or when appropriate access controls have not been enabled for stored data. It is important to implement appropriate controls to protect sensitive data. Thus the main objective is to discuss the security reasons that make the use of data warehouse efficient and provide the access to only authorized users and prevent the sensitive data from malicious attacks.

REFERENCES

[1] S. Amritpal, Nitin Umesh "Implementing Log Based Security in Data Warehouse", International Journal of Advanced Computer, 2013.
[2] Data Masking Best Practice Oracle White Paper June 2013.
[3] J. Horwath, "Setting Up a Database Security Logging and Monitoring Program", October, 2012.
[4] O. W. Rose., H. Albany, K. Beijing, Logs: Data Warehouse Style white paper 2007.
[5] Irad Ben-Gal, "Outlier Detection", Kluwer Academic Publishers, ISBN 0-387-24435-2, 2005.
[6] Edgar R. Weippl, Security in Data Warehouses, IGI Global, Data Ware housing Design and Advanced Engineering Applications, Ch 015, 2010.
[7] Santos, R.J., Bernardino J., Vieria, "Balancing Security and Performance for Enhancing Data Privacy in Data Warehouses", International Joint Conference Of IEEE TrustCom-11/IEEE ICESS- 11/FCST -11, 2011.
[8] P. Huey, "Oracle Database Security Guide 11g", Oracle Corp., 2008.
[9] Arnon Rosenthal, Edward Sciore, "View Security as the Basis for Data Warehouse Security", Proceedings of the International And Management of Data Warehouses Workshop, June 2000.
[10] M. Vieira, R.J. Santos and J. ernardino, "A Survey on Data Security in Data Warehousing".
[11] Lee, S. Y. Low, W. L., and Wong, P. Y., "Learning Fingerprints for a Database Intrusion Detection System", European Symposium on Research in Computer Security (ESORICS), 2002.
[12] Bockermann, C., Apel, M., and Meier, M., "Learning SQL for Database Intrusion Detection using Context Sensitive Modeling", Int. Conference on Knowledge Discovery and Machine Learning (KDML), 2009.
[13] N. Yuhanna, "Your Enterprise Database Security Strategy 2010", Forrester Research, 2009.
[14] Oracle Corporation, "Security and Data Warehouse", Oracle white paper,2005
[15] Oracle Corporation, "Data Masking Best Practices", Oracle White Paper, 2010.

