A Novel Survey on Intrusion Detection Using Data Mining

Purtata Bhoir
Computer Engineering Department
Saraswati College of Engineering
Kharghar, India

Shilpa Kolte
Computer Engineering Department
Saraswati College of Engineering
Kharghar, India

ABSTRACT
Database security is vital nowadays as database system contain valuable information. In today's computer world, attacks are used to disclose, destroy, alter, or steal information. The information security plays vital role to protect confidentiality, integrity and availability of information. Intrusion detection system (IDS) is one of the important components of strong information security system.IDS serve three security functions: they monitor, detect and respond to unauthorized activity. Researchers are working on various data mining techniques such as access patterns of users, data dependencies to detect malicious attacks. Data mining is widely used to find useful patterns from large volume of data. In this paper we have enlisted some existing ID approaches of data mining for detecting insider attacks and compared them with considering their advantages and disadvantages.

Keywords
Intrusion detection, Security, RBAC, Data dependency, weighted sequence mining, data mining.

1. INTRODUCTION
We live in an information age, where the volume of data processed by organization increases exponentially .According to International data corporation (IDC), the total amount of information worldwide will reach 35,000 Exabyte in 2020.In today's scenario, information plays an important role in any organization [1] and its protection against unauthorized disclosure (Secrecy) and improper modification (Integrity) while ensuring its availability is becoming of paramount importance. It is becoming necessary to protect information from various attacks.

1.1 Attacks
An attack is an attempt to gain unauthorized access to a resourses.An Attacks on information can be both internal and external. The external attack comes from skilled and sophisticated hackers. These attackers find the vulnerabilities of system or socially manipulate insiders to give access of system to them by using different technique such as SQL injection.In case of internal attacks the authorized users try to compromise the integrity, confidentiality or availability of resources. In organization many are disgruntled employee who uses their privileged access to damage their employer. Others are infiltrators who work for outside intelligence. Malicious insiders with full access are hard to stop. So each and every organization should have their own security policy solutions or approaches to fix those problems. According to [2] the development of database management system with high assurance security is central research issue and such development requires revision of architecture and techniques adopted by traditional DBMS. One of the most important parts of such new generation security aware DBMS is ID mechanism. IDS make data less vulnerable to attacks and enables early detection of it.

1.2 Intrusion detection system (IDS)
Intrusion is nothing but unauthorized access of data .A system which detects such intrusions is known as intrusion detection system. Although intrusion detection technology is immature and should not be considered as a complete defense, it plays a vital role in an overall security architecture .An IDS can be hardware device or application software which is responsible for monitoring network level or system level activity for malicious activity. IDS can be categorized as "Knowledge-based" (Signature based) and "Behavior-based" (Anomaly based) IDS.The signature based IDS detects intrusion by referring database of previous attack signature or footprints. Each intrusion leaves footprints. Theses footprints can be used to find and prevent same intrusion in the future. A behavior based or anomaly based IDS detects intrusion by learning a pattern of normal system activity. If IDS finds any change or deviation from discovered pattern then it triggers an alarm. The elements central to the IDS are 1. Resources to be protected 2. Models that characterize the 'normal' or 'anomalous' behaviors of the resources 3. Techniques that compare the actual system activities with established models and identify those that is 'intrusive'.

According to [3] IDS's at network or operating system level are not appropriate for detecting malicious attack in database system because it may be possible that the actions which are considered as a malicious for database system need not to be malicious for network or operating system. Hence, ID detection model and techniques specially designed for databases are becoming imperative in need.Fig.1 illustrates the database intrusion detection system.

There are several approaches present for intrusion detections which are based on **data mining** approaches.

1.3 Data mining
Data mining or knowledge discovery in database (KDD) is the process of extracting interesting information or patterns from data in large database. Data mining gaining popularity due to increased of huge amount of data and need for extracting only useful data from it. The steps of KDD process is shown in Fig.2.

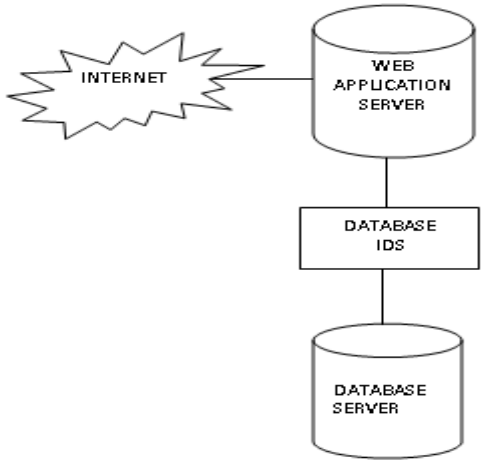

Fig 1.Database Intrusion Detection System

KDD process is defined with different steps such as,

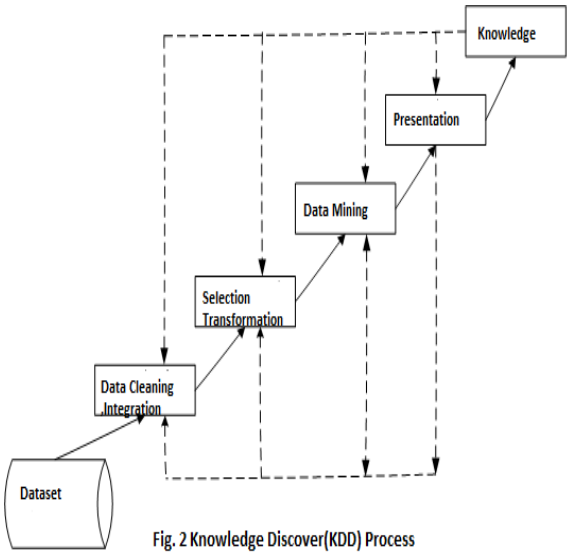

Fig. 2 Knowledge Discover(KDD) Process

1. Data Cleaning - Most of the time real world's databases consist of inconsistent and missing data due to large size of it and multiple sources. In order to improve the quality of data, data cleaning is the important step of mining.
2. Data Integration - It is nothing but collecting data from multiple required sources.
3. Data Selection – Basically it is the process of input reduction .In this step only relevant data required for analysis is retrieved from the database.
4. D. Data Transformation – It is also known as data consolidation. It is a phase in which the selected data from previous step i.e. data selection step is transformed into forms appropriate for the mining procedure.
5. Data Mining – In this step various useful techniques are applied to extract data patterns.
6. Pattern Evaluation – In data mining step different patterns are found but typically users are interested in specific pattern .In step, only interested patterns are to be discovered based on given measures.
7. Knowledge Presentation – It is the last step of data mining where discovered knowledge from all previous steps is visually represented to the user.

Different tasks that can be performed under data mining are classification, clustering, association rule discovery, sequential pattern discovery, and regression and deviation detection.

2. DETECTION TECHNIQUES

2.1 Signature based

In Signature based IDS, signature or structure of all known attacks are used as the basis for intrusion detection. It tries to identify activities by matching a signature stored in a database.

One approach proposed by Dr.M.A.Prabhakar [4], based on 'Aho – Corasick" pattern matching algorithm for static anomaly detection. The overall working of system is divided into static and dynamic phase. In static phase, they maintained a list of known anomaly patterns .The queries submitted by users are checked with stored pattern. In dynamic phase, if new anomaly is occur then alarm is generated and this new pattern is stored to update static pattern list.

Approach by William G.J.Halfont [5] works by identifying trusted string in an application which is unlike previous techniques based on negative tainting.

2.2 Profile based

In profile based ID's, pattern of normal user behavior is noted to create a profile for identifying intrusive behavior. Profile – based anomaly detection focuses on characterizing the past behavior of individual users or related groups of users and then detecting significant deviations. While creating profile, different parameters are considered and deviation on just a single or few parameters may not be sufficient to signal an alert when system detects any deviation from normal profile of stored activities, an alarm is generated so that appropriate action can be taken. The foundation of this approach is an analysis of audit recodes.

Cristina Yip Chung [6] has proposed a system for misuse detection called DEMID .DEMID uses audit log to derive profile of user's behaviors working with database system. The main problem of DEMID system is that, as the number of users of system increases, it becomes difficult to maintain the profile of individual users.

Sudam Kokane [7] has used an approach based on role based access model (RBAC) .In this technique they have built a profile of each role with respect to specific role .The working of their system is split into 3 different algorithms. "Automatic profile Generator" works as profile creator, "SQL Query Parsing" algorithm works as features selector and "Automatic malicious Transaction Detection" algorithm used as a detection engine .If online transaction profile doesn't match with authorized profile then system raises the alarm.

2.3 Dependency Mining

It is been observed that in real world database application ,although the transaction program changes often, the whole database structure and correlations rarely change [3].Because of such correlations ,data dependencies can be used to identify malicious transaction.

Data dependency refers to the access correlation that exists between items of transaction .To generate data dependency rule, one has to find out correlation i.e. to update any single

item of database, which other data item need to be read and as soon as first item is updated, which other data items are likely to get updated .Once the data dependency rules are generated, it can be used to detect malicious transaction .A transaction that do not have mined data dependency rule, marked as malicious transaction i.e. this approach identifies a new user transaction as anomalous if it does not conform to the pattern of normal transaction, mined by data dependency .

Finding dependency among attribute along with the corresponding {read, write} access operation is similar to problem of mining sequential pattern [3].

R.Agrawal [9] has proposed an algorithm for detecting pattern. The main problem with this technique is it treats all attribute at same level without considering the sensitivity of attribute. Sensitivity of attribute shows the importance of the attributes for tracking against malicious modification.

In recent year, size of database is getting larger in terms of number of records and the number of attributes present in the records hence it is becoming difficult for an administrator to determine whether the attribute is being accessed only by genuine transaction or not.IDS often raises an alarm and many of which are triggered because of modification of attribute of less sensitivity which hardly affect on performance of database. By categorizing the attributes into different type based on their relative importance or sensitivity ,it become comparatively easier to track only those attributes whose unintended or malicious modification can potentially have large impact on the database application security [8] .If sensitive attributes are to be tracked for malicious modification then generating data dependency rules for these attribute is essential because if there is not any rule for an attribute ,it cannot be checked[3].Once the rules get generated ,it can be used to verify whether incoming transaction is malicious or not.

Abhinav Srivastava [8] has proposed method that assigns different weights to the attribute present in the database. While assigning weights, it considers its sensitivity, for tracking against malicious transaction. This method assigns higher weight to high sensitivity data item. The problem with this approach is that it considers dependencies at single granularity level.

Yi Hu [10] have proposed an approach which consider multilevel and multi dimensional data dependency mining approach for profiling legitimate data access patterns from the database log directly.

W.Wang [11] have proposed weighted association rule mining techniques also referred as WAR .The problem with his approach is that the rules are generated by using frequent itemset only. There can be attributes present in transaction which are accessed less frequently then no rule generated for such attribute even though it is more sensitive and modification by intruder may cause serious loss to an organization. Table 1. shows the comparison of all above techniques.

Table 1.A Tabular Comparison of Various Intrusion Detection Approach

1. Signature based Intrusion Detection

Sr no.	Approach	Advantages	Limitations
1	Signature Based using Aho-Corasick Pattern Matching	Once the pattern set is built ,the matching is straight forward	Difficult to keep signature of attacks up to date and new attacks are hard to forecast

2. Profile Bases Intrusion Detection

Sr no.	Approach	Advantages	Limitations
1	Access patterns of users	Able to find unknown attacks in database.	It requires more training data set.
2	Based on profile of individual user (DEMIDS)	Ability to detect abuse of users privileges	For a system with large user bases such an approach would be extremely inefficient.
3	Based on profile of user roles (RBAC)model	Managing few roles are much more efficient than managing individual	Only suitable for databases which working is based on role based access control.

3. Dependency Mining based Intrusion Detection

Sr no.	Approach	Advantages	Limitations
1	Weighted Sequence Mining	It minimizes the number of false positive alarm.	Weights of attributes must be assigned manually
2	Weighted association rule mining	Improves the confidence of rules	The rules are generated by using frequent itemset only.

3. CONCLUSION

In this paper, we have surveyed three different approaches for database intrusion detection.A signature based approach is suitable when pattern of attacks are known. For finding unknown attacks, profile based and dependency mining approaches are suitable. A system with large number of user, creating profile according to the user's role is advantageous than individual user's profile.

4. REFERENCES

[1] Mohammad M., Javidi Mina Sohrabi and Marjan Kuchaki Rafsanjani: Intusion detection in databasesystem, Springer-Verlag Berlin Heidelberg 2010.

[2] Ashish Kamra, Evimaria Terzi, and Elisa Bertino: Detecting Anomalous Access Patterns in Relational Database.

[3] Mohammad M., Javidi Mina Sohrabi and Marjan Kuchaki Rafsanjani: An overview of anomaly based database intrusion detection system, Indian Journal of Science and Technology (2012).

[4] Dr. M. Amrutha, Prabhakar, M. KarthiKeyan, Prof. K. Marimuthu, An Efficient technique for preventing SQL injection attack using pattern matching algorithm,IEEE(2013) .

[5] William G.J.Halfond, Alessandro Orso, and Panagiotis Manolios, Using Positive Tainting and Syntax –Aware Evaluation to Counter SQL Injection Attack.

[6] Cristina Yip Chung, Michael Gertz, Karl Levitt, DEMIDS: A misuse Detection system for Database System, Springer 2000.

[7] Sudam Kokane,Aishwarya Jadhav,Nikita Mandhare,Mayur Darekar,Intrusion detection in RBAC Model,International Journal of Innovative Research & Studies May(2013).

[8] Abhinav Srivastava, Shamik Sural and A.K.Majumdar: Database Intrusion Detection using Weighted Sequence Mining, Journal of Computers, Vol.1 (2006).

[9] R.Agrawal, R.Srikant: Mining Sequential Patterns, International Conference Data Engineering (1995).

[10] Yi Hu, Alina Campan, James Walden, Irina Vorobyeva, Justin Shelton: An Effective Log Mining for Database Intrusion Detection, IEEE (2010).

[11] W.Wang, J.Yang, P.S.Yu: Efficient Mining of Weighted Association Rules, ACM SIGKDD Conference on Knowledge Discovery and Data mining (2000).