A Review On Intrusion Detection In Databases

Kandukuri Chandrasena Chary
Associate Professor, Dept of CSE
Sree Chaitanya Institute of Technological Sciences, Karimnagar (Dist), Telangana State, India.

Abstract- *The Database Management System's insider threat is a predominant security problem in recent information age. The rapid growth of Internet and Web application usage has increased the convenience to authorized users (i.e. insiders) of database to access data from anywhere. This also encourages the external users to motivate towards the masquerading as normal user and damage the database. The threat of Information security breaches has increased rapidly by hasty explosion of Internet and Web applications. Traditional Database Intrusion Detection Methods are not sufficient to defend against novel attacks. Most of the Database intrusion detection techniques proposed over the last decade was addressing efficient handling of external and internal attacks. There is currently need of up-to-date and thorough survey of the research in Database Intrusion Detection. In this article, we have presented an overview of efficient methods for insider threat detection and mitigation against Databases, and classification of surveyed systems based on their method of implementation.*

Keywords- Insider Threats, Database Intrusion Detection, SQL Injection, Data Dependency Analysis, Malicious Data Modification.

I. INTRODUCTION

Relational Databases are the key information storage and management repositories for all organizations in today's networked world. Since organization's future business strategies, employee details and other organization sensitive information is kept in the same repositories, ensuring privacy and security of such data in those repositories is a challenging problem. Broadly we can divide database attacks as *External* and *Internal attacks*. *External attacks* are the malicious activities carried out on database by persons those are not authenticated users of the database. These External Attacks can be detected by carefully crafted authentication mechanisms, such as RBAC (Role Based Access Control), controlled access to data by issuing stringent access privileges and roles. *Internal attacks*, these are the malevolent activities by an already trusted person with access to sensitive data and database systems. These personnel who perform internal threats are called Insiders. Many Definitions exist in literature for Insider. The Authors in [1] define the insider as a person who has privileges to access the underlying system. And the authors in [2] have given definition for insider as, "an individual who has the knowledge of the organizations information system structure to which he/she has authorized access". Another definition stated in [4] as, insider can be an employee that uses his/her knowledge of attacks are very serious and dangerous due to their nature which includes personnel who have privileges and authorization to access organizations' resources. These Insiders, working under the security perimeter of the database system will perform malicious activities on data and damages the database.

As per [7] there are three Intrusion Detection methodologies that have been used, they are: *Misuse Detection* – which is based on signatures of the known attacks, and it can't detect unknown attacks. *Anomaly Detection* – This stores all normal user operation profiles and is more effective in detecting unknown attacks. *Insider Misuse* – the sources which has access to critical data and perform malicious activities such way that outsider gain access to data. These types of insider attacks [5] are more dangerous because they are accessing system critical data.

The major issues that need to be addressed efficiently by Database Insider Threat Detection Techniques are: Efficient Profiling of User Actions, Improved Detection Rate, Scalability of Detection Method, Reduced False Positives & False negatives, and Fast Damage Mitigation through real time detection.

The major malicious activities that an insider can do on a database are: SQL Injection attacks, trying to gain access into system in which he/she unauthenticated, Masquerade attacks, Penetration of security characteristics of system, Data Leakage, Deny other users from using Database resources, Malicious attacks such as deletion of files.

II. REVIEW ON INTRUSIONS IN DATABASES

As per the extensive survey that we have done on Insider threat detection techniques in Databases, we have clustered them as following major categories:

1. Graph Theoretic Methods
2. Structural Methods
3. Data Mining Based Methods
4. Machine Learning Methods
5. Statistical Methods

2.1 Graph Theoretic Methods:

William Eberle et al. [9] has proposed a Graph based anomaly detection (GBAD) method to detect insider threats, in which the instance of anomalous structural pattern can be detected which includes entities, relationships and actions. In GBAD entities are represented by labeled vertices and relationships or actions are represented by labeled edges between entities. The GBAD uses three methods to detect changes to Graph: modifications, insertions and deletions. Each algorithm finds substructures that closely match with normal patterns, but which are structurally different. The anomaly is found by detecting "additional substructure" in the graph which is a deviation from the normal activity. The anomalous graph sub structure is defined in [9] as

Definition 1: A graph substructure S' is anomalous if it is not isomorphic to the graph's normative substructure S, but is isomorphic to S within X%.

X signifies the percentage of vertices and edges that would need to be changed in order for S' to be isomorphic to S, Where S is a Normative Substructure in graph.

By using Minimum Description Length Principle and Probabilistic methods the author has optimized the pattern to be matched, hence increased speed of detection

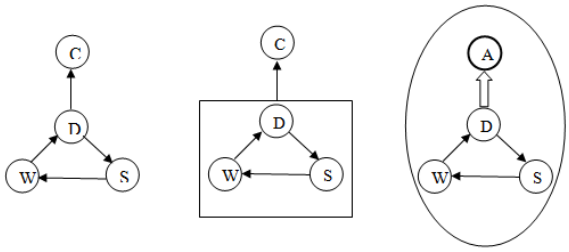

Figure 1: Example Instance with Normative pattern boxed and anomaly circled

State Transition based Methods: These methods are a Sub category of Graph Theoretic Models. In these methods, a model is build based on the system inherent Events, Conditions and Actions. These systems can be effectively modeled by Petri Nets.

Petri Nets are essentially weighted, labeled, directed graphs with token that move around the graph as reactions take place. There are two types of nodes in the Petri Net, they are Places and Transitions, where Place is represented by Circle and Transition is represented by Rectangle. An Arc may direct only from Place to Transition or from Transition to Place. Hence a Petri Net is always a bipartite graph.

According to Tadao Murata [23] a Petri Net is defined formally as below

Definition 2: A Petri net is a 5-tuple, PN = (P, T, F, W, M_o) where:

P={p1, p2, ….., pm} is a finite set of Places,
T={t1,t2,…….,tn} is finite set of Transitions,
F ⊆ (P x T) U (T x P) is a set of Arcs (flow relation),
W: F → {1,2,3…} is a Weight Function,
M_o: P → {0,1,2,3…..} is the initial marking,
P∩T = Φ and P∪T ≠ Φ

A Petri net structure N=(P, T, F, W) without any specific initial marking is denoted by N. A Petri Net with given initial marking is denoted by (N, M_o).

Most of the systems behavior can be described in terms of States and their changes. According to the dynamic behavior of system a state or marking is changed in Petri net. The following rules are applied while changing the marking as defined in [23]

- A transition t is said to be enabled if each input place *p* of *t* is marked with at least w(p,t) tokens, where w(p,t) is the weight of the arc from *p* to *t*.
- An enabled transition may or may not fire
- A firing of an enabled transition *t* removes w(p, t) tokens from each input place *p* of *t*, and adds w(t, p) tokens to each output place *p* of *t*, where w(t, p) is the weight of the arc from the *t* to *p*.
- *Application of Petri Nets in Intrusion Detection:* Yuan Ho et al. [24] has proposed a method to identify intrusions where incomplete behavioral data available. Authors have proposed architecture as shown in Figure 3, which combines partial order planning and executable Petri Nets to detect intrusions. This architecture includes planning agent, a searching agent, and a site security informing agent. The planning agent constructs intrusion scenarios using a first-order logic description of the known activities and goals of the intruder to specify an attack sequence. The searching agent takes a Petri Net representation of these intrusion scenarios and uses them to determine whether any of the specified intrusions are in progress. The informing agent processes final information and provides a user interface for the system.
- Xiaoou Li et al. [22] has applied Petri Nets for Active Database Systems to modeled the Events, Conditions and Actions. Authors have proposed a Conditional Colored Petri Nets (CCPN) based system to model

ECA rules can be effectively. They have also developed a software platform ECAPNSim to generate CCPN directly from the given text file of ECA rule description. For example consider the following ECA rule

ON update of amount on BONUS
IF BONUS_Amount > 100
THEN update EMP set rank = emp.rank + 1
Where emp_id = update.emp_id

Here Event E_o is update BONUS.Amount and Event E_1 is update Emp.rank then the generated Petri Net by ECAPNSim is as shown in Figure 4.

Review Conclusion: Graph Based approach is very effective in detecting malicious insider threats, but the complexity of this approach increases when the number of relationships and number of actions between entities are more. Hence the scalability of this approach is limited. We can apply this approach to applications with less number of entities, relationships and actions. Since it includes more computational complexity, we can't apply for online malicious activity detection.

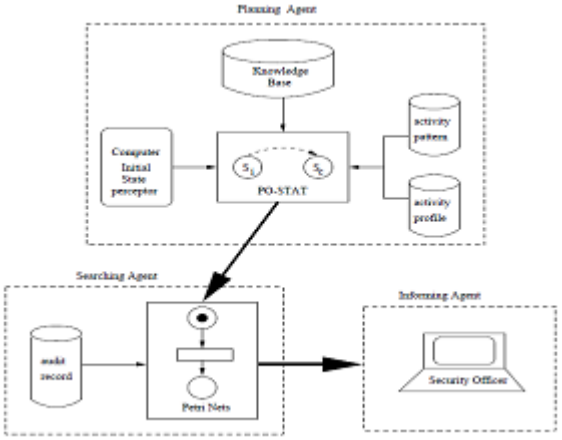

Figure 2: Organization of Planning, Searching and Informing agents for Intrusion Detection

2.2 Structural Methods: (User Normal Query Profile based methods)

Structural methods are based on the Profiling of Normal User Query execution sequence. These profiles will be matched with the incoming queries to the database, and if it matches with any one of existing profiles, then that query is normal otherwise it's a threat. The Ke Wei et al. [10] has defined the SQL Injection attack as

Definition 3: An *SQL Injection Attack* (SQLIA) is a subset of the unverified/unsanitized input vulnerability and occurs when an attacker attempts to change the logic, semantics or syntax of a legitimate SQL statement by inserting new SQL keywords or operators into the statement.

The above definition includes, attacks based on tautologies, injecting additional statements, exploiting untyped parameters, stored procedures, injection of "UNION SELECT", "ORDER BY" and "HAVING" clauses.

In [10], authors proposed SQLIA detection combining static and runtime analysis with the features: no code modification, optimized runtime analysis using SQL graphs and SQL query validation in parallel and SQLIAs using access control violations in the script. In static analysis, the application code is scanned to find out key information that help us to infer models for legitimate SQL queries. And an SQL injection attack should violate these models. In Static Analysis phase, they build Finite State Automata (FSA) by applying program analysis. These FSAs are called SQL-graphs. And in run time phase they check dynamically generated SQL queries with static data structures for marking normal or threat.

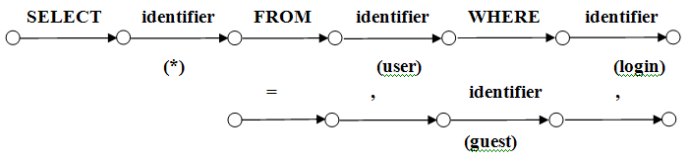

Figure 3: Example of SQL Finite State Automata

Ashish Kamra et al. [11] have proposed a new scheme to deal with SQL Injection attacks. This approach fingerprints the database application program, which includes the normal queries submitted to the database. They have used the Query Traces from Log table to build the profiles. Then they have applied Anomaly Detection technique based on Data mining approach to find out SQL injections. In the process of building normal profiles, they have used query encoding technique, to speed up the detection process.

Survey Conclusion: The above two techniques are well suited for insider attack detection, since insiders basically write and inject the SQL code that may include malicious statements which cause severe damage to the database or which can leak the sensitive data. Hence thorough scanning of application code and careful crafting of application profiles will lead to increase in detection of insider attacks.

2.3 Data Mining based Methods

Data Mining based models include application of Association Rule mining, Clustering, Classification methods and sequence Mining. These Methods are more effective in finding the hidden knowledge in data. Hence these methods are more effective in insider threat detection, where the user behavior is dynamic. Most of the Database intrusion detection methods [17] [18] in the literature are based on Data Mining techniques, which mine the anomalous access patterns and profile them as signatures. These signatures can be later used to detect malicious operations.

Y Zhong et al. [18] has proposed a method based on user query frequent item sets mining based on the Item constraints. In this method, they discover the most frequently accessed query sets by each individual user, and then Item level constraints will be imposed to reduce the number of frequent Item sets generated. The author has used a Query template to find frequent query item sets, which is a static structure defined for each individual user data access behavior.

Zhang Yanyan et al. [17] has proposed a scheme which mines the association rules between frequent item sets using Improved Apriori Algorithm. The contribution of the author in this scheme is, divides the dataset into Partitions and finds local frequent Item sets to each partition, and apply same Apriori algorithm on combined set of all those local frequent sets to find globally frequent Item sets. The improvement of this approach over base Apriori algorithm is, reduced number of candidate item sets to be considered at each iteration and scans transaction database only twice throughout the process.

Yinzhao Li et al. [19] have proposed a Event Sequence Clustering method to detect intrusions in Databases. They have proposed an improved Edit Distance method to find the similarity between different SQL Sequences. First in the training phase the database log records are preprocessed and clustered according to pre-defined cluster distances. Then the intrusive sequences are detected by calculating distance between user operation sequence and cluster center. If the difference is more than given threshold values then that user sequence is reported as intrusive and user are warned about the threat.

Survey Conclusion: The application of Data Mining methods in insider threat detection investigate deep into the access patterns of insiders and extract useful hidden knowledge which is helpful to identify malicious behavior of insider of database.

2.4 Machine Learning based Methods

These models include building a classifier based on automatic learning of behavior of the system. These models include Neural Networks and Support Vector Machines. Neural Networks are the efficient way of learning knowledge and inference. Neural Networks have fast inference and accurate result capability. Conventional methods of detecting insider threats will only work on known attack signatures and can't detect new and novel attack signature. Hence there is a enormous demand of self learning and inference methods for intrusion detection.

Susan C. Lee et al. [14] and Andrew Sung et al. in [12] has proposed a method to detect novel intrusions by training a neural network. The normal problem space has been used to train the neural network, which then be used to detect known attacks and novel unknown attacks. The method in [14] is based on network protocol anomaly detection, but this method can be extended to isolate the insider threats in databases effectively. Since it is very difficult to trace the malicious actions performed by a insider, Neural Networks play a critical role in learning dynamically changing user behaviors.

Sahani et al. in [13], proposed a Machine learning based approach at transaction level to detect anomalous behavior in databases. This method is enabled with Role based access control mechanism (RBAC). In this approach the number of profiles generated are very less compared to other methods and it can easily detect the malicious activities of the authorized users. Based on correlation between attributes accessed by set of queries in a given transaction the anomalies can be easily detected. Consider the following example transaction:

Begin transaction
Update emp set sal = sal + 100;
Select eno,deptno, sal from emp, dept;
End Transaction

In the above transaction the emp, dept are the relations and <eno, deptno, sal> are the attribute of the relations. This approach is using the following data structures to profile the relations and attributes accessed in a transaction, they are

(Read, TB-Acc[ ], Attr-Acc[ ][ ]) (Write, TB-Acc[ ],Attr-Acc[ ][ ])

The above binary data structures stores the presence or absence of the relation or attribute in the transaction. Then based on the correlation between operations anomalies are defined and detected.

Jinfu Chen et al. [20] have proposed an Auto Generation approach for legal transaction profiles in Database system. This method uses Database Audit Log and builds the User normal transaction profile graph for a fixed number of transactions that a user can perform. This method saves the DBA time in building normal transaction profiles by scanning entire audit log manually. Then these Transaction Profile Graphs are used to find the deviations of user operations compared to profiled operation sequences. Figure 3 depicts the model of Legal Transaction profile graph generation.

Review Conclusion: Machine Learning methods are useful to learn the dynamic behavior of malicious insiders and protect database against their attacks. Once we have built and trained a classifier it will detect known attacks and if it gets a new novel attack, infer new signatures from already existing signatures. By which it reduce the explicit recurring training of classifier, hence improves the speed of detection and accuracy.

2.5 Statistical Methods

Gand Lu et al. in [21] has proposed a statistical and fuzzy logic based method which is based on Cumulated Anomaly. The Database Anomaly can be detected by addressing the concept of Cumulated Anomaly. Author has proposed a Detection Model called *Dubiety Determining Model (DDM)* which is based on statistical and fuzzy theories for Cumulated Anomalies. This Model measures the Dubiety Degree for each Database Transaction to identify Cumulated Anomaly.

Survey Conclusion: As the malicious insider never inject the intrusive operations in a single transaction, it is necessary to scan the sequence of transactions issued by the same insider to identify the malicious intent of the attacker, which can be effectively done by Cumulated Anomaly Detection Method.

III. CONTRIBUTION AND CURRENT WORK

Currently we are working on the Self Learning Petri Net Models for automatic learning of user behavior patterns in database access. Since it is very difficult to identify the malicious operations issued by an insider, scanning and learning the user transactions dynamically is a key challenge. Our model effectively learns unknown user behaviors and detects the insider exploitations. Since our model is based on Anomaly as well as Misuse intrusion detection techniques, which reduces the number of false positives, false negatives and hence increased detection rate.

IV. CONCLUSION AND FUTURE WORK

In this paper we have presented a comprehensive study of different Database Intrusion Detection Methods and discussed their applicability to detect Insider threats in Database Systems. Based on our study, we have proposed five broad categories of methods those are more appropriate to Insider threat detection in databases, though there were so many methods in literature, effectively applied to Database IDS. We have narrowed down our study to insider malicious activity detection.

Since the attack behavior of Insider has been changing dynamically, it is required to devise methods which can learn, infer and detect novel insider attacks efficiently. As the future work of above mentioned methods, one can work towards automatic learning of user behavior using models such as Fuzzy Neural Networks, Support Vector Machines, and Fuzzy Petri Nets. There is also a scope of implementing methods to make Database Intrusion Detection systems resistant itself against attacks using genetic algorithms and Artificial Learning Techniques.

REFERENCES

[1] Nam Nguyen, Peter Reiher, Geoffrey H. Kuenning. Detecting insider threats by monitoring system call activity. IEEE CS digital library 2003.

[2] Qutaibah Althebyan. Design and analysis of knowledge-base centric insider threat models [dissertation]. University of Arkansas; July 2008

[3] Cappelli, D.: Preventing insider sabotage: Lessons learned from actual attacks (2005), http://www.cert.org/archive/pdf/InsiderThreatCSI.pdf

[4] Mark Maybury, Penny Chase, Brant Cheikes, Dick Brackney, Sara Matzner et al. Analysis and detection of malicious insiders. International Conference on Intelligence Analysis; McLean, VA. 2005.

[5] Asish Kamra, Department of Defense. DoD insider threat mitigation: report of the insider threat integrated process team. Washington DC, USA: Technical report; 2000.

[6] Nahla Shatnawi, Qutaibah Althebyan, and Wail Mardini, "*Detection of Insiders Misuse in Database Systems",* International Multi Conference of Engineers and Computer Scientists 2011, Hongkong.

[7] Michael Gertz, Sushil Jajodia. *Handbook of database security*. Berlin: Springer- Verlag; 2007.

[8] W. Eberle and L. Holder. "Insider Threat Detection Using Graph-Based Approaches," *Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH),* March 3-4, 2009.

[9] Eberle, W.; Holder, L.; Mining for insider threats in business transactions and processes IEEE Symposium on Computational Intelligence and Data Mining, 2009. CIDM '09

[10] M. Muthuprasanna, Ke Wei, Suraj Kothari, Eliminating SQL Injection Attacks – A Transparent Defense Mechanism, Eighth IEEE International Symposium on Web Site Evolution (WSE'06).

[11] Bertino, Elisa; Kamra, Ashish; Early, James P, **Profiling Database applications to detect SQL injection attacks, IEEE** International Conference on Performance, Computing, and Communications Conference, 2007. IPCCC 2007.

[12] Mukkamala S, Janoski G, Sung A, Intrusion detection using neural networks and support vector machines, International Joint Conference on Neural Networks, 2002. IJCNN '02.

[13] Udai Pratap Rao, G. J. Sahani, Dhiren R. Patel, *Machine Learning Proposed Approach for Detecting Database Intrusions in RBAC Enabled Databases*, Second International conference on Computing, Communication and Networking Technologies, 2010.

[14] Susan C. Lee and David V. Heinbuch, *Training a Neural-Network Based Intrusion Detector to Recognize Novel Attacks*, IEEE Transactions On Systems, MAN, and Cybernetics—Part A: Systems And Humans, VOL. 31, NO. 4, JULY 2001

[15] Yi Hu and Brajendra Panda, *Design and Analysis of Techniques for Detection of Malicious Activities in Database Systems*, Journal of Network and Systems Management, Vol. 13, No. 3, September 2005.

[16] Peng Liu, Paul Ammann, Shushil Jajodia, Rewriting Histories: Recovering from Malicious Transactions.

[17] Zhang yanyan,Yao Yuan, *Study of Database Intrusion Detection Based on Improved Association Rule Algorithm*, 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010.

[18] Yong Zhong, Xiao-lin Qin, *Database Intrusion Detection Based on User Query Frequent Itemsets Mining with Item Constraints*, InfoSecu '04 Proceedings of the 3rd international conference on Information security, 2004.

[19] Yinzhao, Li,Dongxu, Yang,Jiadong, Ren,Changzhen Hu, *An Approach for Database Intrusion Detection Based on the Event Sequence Clustering*, IEEE Fifth International Joint Conference on INC, IMS and IDC 2009.

[20] Jinfu Chen, Yansheng Lu, and Xiaodong Xie, *An Auto-Generating Approach of Transactions Profile Graph in Detection of Malicious Transactions,* In *Proceedings of the Third International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007) - Volume 01* (IIH-MSP '07), Vol. 1.

[21] Gang Lu, Junkai Yi, *Statistical and Fuzzy Approach for Database Security*, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), Vol. 3, Page. 670 – 674.

[22] Xiaoou Li, Joselito M. Medina, Sergio V. Chapa, *Applying Petri Nets in Active Database Systems*, IEEE Transactions on Systems, MAN, and Cybernetics— part c: Applications and Reviews, Vol. 37, No. 4, July 2007.

[23] Tadao Murata, *Petri nets: Properties, Analysis and Applications*, Proceedings of the IEEE, 77(4), 1989.

[24] Yuan Ho, Deborah Frincke, Donald Tobin Jr., Planning, Petri Nets, and Intrusion Detection, Proceedings of the 21st National Information Systems Security Conference 1998, 348-360.