International Journal of Emerging Engineering Research and Technology
Volume 5, Issue 10, 2017, PP 30-36
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Design and Implementation Data Base Intrusion Detection System

Jamal Mohammed Kadhim, Hadeel Alaa

Department of Computer Science, College of Science, AL-Nahrain University, Baghdad, Iraq

***Corresponding Author:** Hadeel Alaa, Department of Computer Science, College of Science, AL-Nahrain University, Baghdad, Iraq.*

ABSTRACT

In this paper, We proposed the detection mechanism of the database intrusion to enhance the security of DBMS. In a typical database environment, it is possible to define the transaction profile that each user is allowed to perform. On our way we will work on two branches to ensure the protection of the database.

- *The first phase is through verification of user access validity in case if authorized access to the system or an intruder is trying to illegally access.*
- *The second phase by learning the system to guess the user if he is an intruder according to his behavior and operation inside the system and alert the admin about it.*

Keywords: *Data Mining, Machine Learning Technique, and Intrusion Detection Model.*

INTRODUCTION

Data mining is a very obvious correction in the world. Mostly, we never even notice it happening. But when we want to sign up for a grocery store shopping card, put up a purchase using a credit card, or browse the Internet, we create the data. This data is stored in a wide range of collections on large computers owned by the companies we contract with each day. Lying within those data sets are patterns of indicators of our behaviors, interests, and habits. DM allows people to develop and deduce those patterns, and help them make better informed decisions and better serve their customers.

However, there is also interest on data mining performance. Monitoring groups express isolation in particular from organizations that accumulate huge amounts of data, some of which can be very special in nature [1]. DM techniques are increasingly being used in disciplines of traditional scientific discoveries, such as chemical, biological, social sciences, medical, physical and, and a set of other knowledge manufacture, such as government and education, aimed at discovering previously unknown patterns and links, as well as predicting trends and behaviors .Thus, DM plays a very significant part in building and modeling future knowledge-based on manufactures and businesses [2]. This is due to the rapid growth of interest in DM [3][4]:

1. The advancement of Internet technology and the wide interest in multimedia applications in this domain.
2. Low cost of large storage devices and increased no difficulty of data collection across networks.
3. The sharing and distribution of data over the network, along with adding of new data in existing data repository.
4. The development of powerful and effective machine learning algorithms to process this data.
5. Computer engineering offers lower cost of computational power, allowing the use of computationally intensive methods to analyze data.
6. The inadequate scaling of conventional querying or analysis methods, prompting the need for modern ways of interaction.
7. Strong competitive pressures from commercial products available
8. The explosive growth in data collection.
9. The keep of data in data repository, so that the whole project has access to a authoritative current DB.

DATA MINING

Data mining is often described as a multi-stage replication process involving data selection; data clean up, data mining and evaluation algorithms,

and so forth. Here we adopt a somewhat different. An action-oriented view and breaking it into five basic steps:[5]

1. Exploring and Processing: The initial steps to explore, visualize, and query data, to get an prudence into data in reactive way. Pre-processing steps like variable chosen converge on data, and data effectiveness can also be involved in these first steps.
2. Modeling: The steps involved (a) to choose the representation of the model we seek to fit with the data (e.g tree, linear function, probability density model, etc.), (b) selecting the score functions that score different models with respect to data, and(c) Computational methods and algorithms to improve the result function (for example, local greedy search). Together these components determine the data mining algorithm that can be used. The components can be pre-assembled into a certain algorithm (such as the Kart or C4.5 decision tree applications) or can be merged in a custom way "(more common in science).
3. Mining: the step (often repeated) of indeed working a special data mining algorithm on a special data set.
4. Evaluation: stage (often ignored) to critically evaluate the kind of output of the data mining algorithm results from step 3, both from the model predictions and the interpretation of the model itself.
5. Deployment: Step (seldom achieved) from model development of data mining algorithm to routine predictive use, for example, using the model continuously in real time to register customers visit the e-commerce website. A technical issue facing a challenge (and under-appreciation) in this context is how and when models for such applications should be updated "continuous data stream".

MACHINE LEARNING

Machine learning is a subfield of computer science that award computers the capability to learn without being clearly programmed. Developed from a computational pattern survey and recognition of learning theory in artificial intelligence, machine learning search study and construct algorithms that can learn from and predict data such algorithms overcome after program orientation is completely fixed by making driven data decisions or predictions, by building a model of Sample input. Machine learning is used in the field of computational tasks where the design and programming of explicit algorithms are not usable; applications include instance spam filtering, detecting intruders or malicious network insiders working towards data breaches, search engines and computer vision [6].

Machine Learning Techniques

There are various ways the algorithm can pose a problem based on its interaction with the environment or experiment or whoever we need to call input data. It is popular in machine learning and textbooks artificial intelligence to first believe the types of learning that the algorithm can assume.

Supervised Learning

Input data is called training data and have a known label or result such as spam /not-spam or stock price at a time. The sample is to get out of the training operation that is asked to make prediction and correct when these prediction are incorrect. The training operation continues while the model achieves the desired level of precision in the training data. Instance algorithms include logistic Regression and Back Propagation Neural Network.

Unsupervised Learning

Input data is not labeled and does not contain a know result. The sample is obtained by deducing the current builds into input data. This may be to extract general rules. It may be out of a mathematical operation to systematically reduce redundancy, or it may be to control data by similarity. Instances include clustering, reduction of dimensions, and association rule learning. Instances algorithms include: k-mines and the **Apriori algorithm.**

Semi-Supervised Learning

Input data is a combination of labeled and unlabelled situations. There is a desired prediction problem but the model must learn build to organize data as well as make predictions. Instance problems are regression and classification. Instance algorithms are expansion to other supple techniques that make hypothesis around how to model the unlabeled data.

Intrusion Detection

Intrusion Detection is a security mechanism that monitors and tests computer and network events to provide real-time warnings for unauthorized access to system resources or archiving log information and traffic information for later

analysis [7] [8]. Intrusion detection works on logs or other information available from your computer / network. ID is an important component in infrastructure protection mechanisms [9].An Intrusion Detection System is a program, device or combination of both that monitors and collects computer / network information and analyzes them to determine if an intrusion has occurred .Snort is an open-source IDS available to the general public, see Appendix A. IDS may have various capabilities depending on the complexity and requirements of the components [10]. Inevitably, the better intrusion preventing systems will failure. Thus the second defense system is IDS, and this has been localized a lot of research in the new years.

SECURITY ATTACKS

Computer Security is the protection given to the computer system in order to realize the established objectives of maintaining the security devices of computer resources (including hardware, software, information / data, and telecommunications).

Network Security includes protecting all network resources from threats. Do not just look at computers on the network, but other network devices, network transports media,, and data transmitted over the network.

A Security Service is a processing or communication service that enhance the security of data processing systems and the transmission of information from an organization. These services aim to address security attacks, and benefit from one or several security mechanisms to supplies service. Security devices, as described in [11]:

Integrity includes that only authorized parties are capable to modify computer assets and transferred information. Modification includes typing, changing, deleting, creating, delaying or responding to sent information. A loss of integrity is unauthorized modification or devastation of information.

Availability ensures timely and credible access to and utilize of information. A loss of availability is the disturbance of access to or use of information or an information system.

Confidentiality includes that information in the computer and information transferred is accessible only for reading by the authorized parties. A loss of confidentiality is the unauthorized detection of information.

TYPE OF ATTACKS

Passive Attack

Is a try to learn or make use of information from the system but does not affect the system resources. The passive attacks are in the nature of transmission, or monitoring, and tapping. The object of the opponent is to obtain the information that is being transferred. It is very difficult to detect passive attacks, because they do not involve any change in information. Normally, messages are sent and received in a normal way, and the sender and receiver do not realize that a third party has read the messages or observed the traffic pattern. However, it is possible to prevent the success of these attacks, usually by encryption. Thus, the focus of dealing with passive attacks is on prevention rather than detection. There are two kinds of passive attacks: "Release message contents" and "Traffic analysis" [11]:

Release of Message Content

The information you send can be e-mail messages, important files, or any important data. The idea is to deny the opposition to learn the import of these transmissions.

Traffic Analysis

Even if the information sent is preserved by encryption or otherwise, the opponent could determine the identity and location of the communications computers, the frequency and length of the mutual massage that may be useful in guessing the nature of communications.

Active Attack

An active attacks attempt to change system resources or effect their operation. Active attack include some modification of the data stream or the creation of false stream and can be divided into four groups: "Masquerade", "Modification of Messages", "Replay", and "Denial of Service (DoS)" [12][13][14]:

Masquerade

occurs when one entity pretends a various entity. A masquerade attack usually involves disguising one of the other forms of active attack.

Modifying Messages

simply means that part of a legal message has changed, or the message has been delayed or rearranged, producing an unauthorized effect.

Replay

include the passive take of a data unit and its

next retransmission to product an unauthorized effect.

DoS: deny or inhibits the normal utilize or department of communications facilities and computer resources. DoS attack may have a specific target. In general, DoS attacks grow on and on, doing little harm besides wasting users' time and bandwidth and occasionally crashing a computer. In the vast majority of these attacks, the source address is faked or "spoofed". A DoS attack is designed to turn a computer/network down by flooding it with useless traffic usually generated by a Trojan application.

CHARACTERISTICS OF TCP

TCP supply a communication channel between operations on each host system. Reliable channel, full duplex, flow. To realize this function, the TCP driver separates the session data current into separate pieces, and attaches a TCP header to each segment. An IP header is attached this TCPC packet, and the composite packet is then passed to the network for transmission . The TCC header contains many fields that are used to support the intended TCP functionality. TCP has the following practical features:

Unicast Protocol

TCP is based on a unicast network model, and supports data commutation between two parties specifically. It does not support broadcast or multicast formats.

Connection State

Instead of forcing a case within the network to supports the connection, TCP uses a synchronized case between two endpoints. This synchronized case is set up as portion of the first connection process, so TCP can be considered as connection-oriented protocol. Much of the design of the protocol is intended to include that each local transmission case is transferred to, and knowledge by, the remote party.

Reliable

Means that the octet flow push through to the TCP driver on one end of the connection will be transmitted over the network so that the flow is presented to the remote process in the same octet series in the same order as the sender. This means that the protocol discover when portions of the data flow are ignored by the network duplicating, reordering, or corrupting. Where needful, the sender resends the damaged segments to allow the receiver to rebuild the original data train. This implicate that the TCP sender must ensure a local copy of all data sent while it receives an reference that the receiver has full the data transmission accurately[14].

INTRUSION DETECTION IMPLEMENTATION

The Intrusion Detection Model is a "hybrid IDM" (Network-based and Host-based IDM) that it considers features of network by using TCP features to create a connection session from client to server and some critical features of host that are directly affected by the intruder. The proposal is a DB system-based IDM in which anomaly detection techniques depend on in the detection of intrusion, where anomaly technique should first learn the characteristics of normal and abnormal (intrusive) activities of the system, and then the IDM detects the activity that deviates from normal activities.

SYSTEM MODEL

1. At first the system should verifying user access by the main gate of the system which is the authentication step, here he has to enter his account or create a new one in case he doesn't have an account. If the input information matched with the database then he can enter directly to the system else he has to retry by using true information. When the user enters correctly to the system, directly a new profile will created for him. The user profile contains the main data which are required to ensure the security for the system such as user name, date and time, table name, user_response, and transaction id. The user_response flag is to identify is it the first time for the user enter to the system or he has already entered the system**.**
2. This log file is updated directly whenever the user enters to the system. For each operation and process the system calls the log file for that user after bring his information from the database. The whole information of the user will saved temporarily in the cache of the IDM, also the log file after calling will saved temporarily in the cache of the IDM to make the comparison to detect the intrusion attempts and prevent the intruders later or take action for them**.**
3. When the user want to execute any transaction, the first step is fetching the log file from the database and put it into the cache of IDM as a temporary file to make the

process of comparison easier, it will removed automatically whenever the process is finished . The system at the same time is trying to analyze the behavior of the user and identify the transactions either it is normal or up normal. The normal transaction will comparing with the log file and execute it by the system after enter all the required inputs for the database system. The system stores all the information and updates the database with the new input data. In case of there are some up normal actions extracts from the user behavior then the system has to prevent that action to protect the whole system and the data too.

4. When the system detects an intrusion attempt then it has to activate the second phase of security levels in order to protect the data and information. The second level of protection is divided into 2 approaches, in each one the system will take a different action to keep security for the whole system. The first way depends on recording the information of the user which did the up normal behavior or the anomaly from entering the system and deny him, also directly the system will include the name of this user in the intrusion attempts' list in order to display his name in the archive list when the admin logs in, this process is very important to alert the administrator about all the attempts. The behavior of the user is considered as a normal or up normal according to the ability of the system when it learned about the kinds of behaviors by identify the normal behaviors and the acceptable transactions, so then the IDM can automatically recognize and decide if this behavior is to ensure as much as security for the system it will alert the administrator of the system and indicates him by the names attempted to break through the system and tried to execute some of transactions illegally.

5. When the administrator sign in to the system, he has the capability to view and see the archive of all attack attempts and the action which taken in order to prevent them. The information table has the main data and the necessary information about the intruders that gives the ability to recognized by the administrator. These information include name, time, date, transaction id, transaction type, and the table name.

6. To apply more security to the system there is extra protection mechanism. This mechanism doesn't depend on the intrusion or the behavior is it normal or up normal. This mechanism is applied for all users on the system. It gives them slots of time to execute their transactions, these slots of time are decided by the user himself according the needing of him to finish all of his transactions. The user has to finish and execute his transactions completely within this duration in order to get the results because he will be unable to execute any more transaction in case of disconnection from the server, so he has to finish all of his operations before ending the session and disconnect himself from the server, otherwise he must reconnect again to the server. For example, when the user wants to doing some transactions to update the database just like insert some records. The system will directly start the session whenever he connects to the server, the user must finish all of his transactions and operations before he ends the session and be disconnected, else he has to start a new session and connect again to the server. This simple technique is by limiting the time of execution for each user in order to ensure extra security for the system generally and without looking at the intrusion detection.

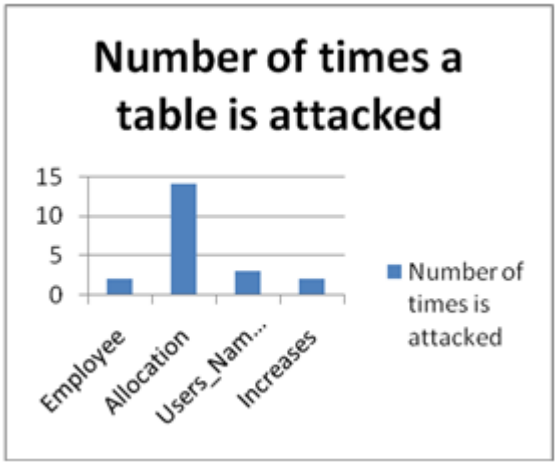
7. The second part of the system gives the ability to the administrator to detect intrusion to take action for them in order to prevent them by using coding, when we say coding that means the system should has the ability to identify programming codes and recognize it to analyze each statement and take the right decision about the behavior is it normal or up normal.

THE PERFORMANCE OF THE IDM

Performing the Attack When sending an attack notification to the server, the server is successfully able to prevent the attack, log the attack entry in the database, and notified the administrator of the server about the attack. As it explained in the last sections and figures.

Most Targeted Tables

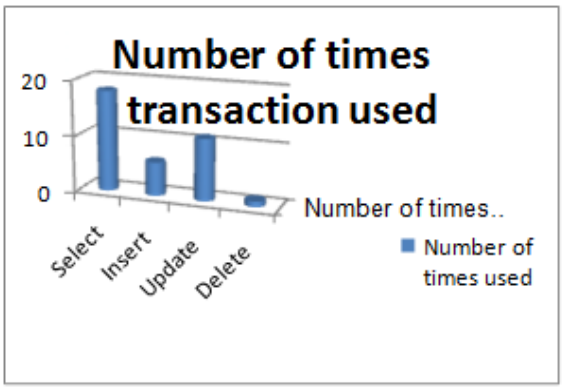
This experiment consists of examine the collected data to recognize which table is most attacked by the intruders, a sample of the collected data examined in this experiment to recognize the tables as it shown in figure.



In this experiment a sample of attack attempts collected to examined in order to get the information about which table is the most attacked and why? Of course this discussion is very useful in order to have an idea about the attackers thinking. As it shown in figure (1) **allocation** table has the highest rate of attempts maybe that is because it related to the salaries of the employees and thin table is very important and attractive and it is more interested than the others, secondly the **Users_Names** table has the second rate of attack attempts in order to break through the system by legally way if they know any information about users. Lastly there are two tables get the same rate of attacking attempts which are **Employees** and **Increases**, they are less importance than the salaries in all database systems.

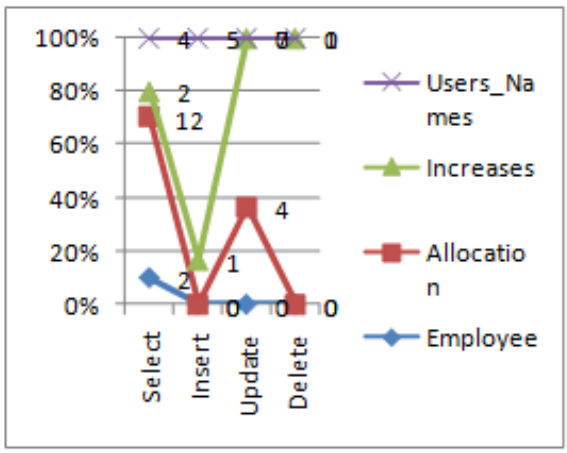
The Most Used Transactions

Here as it shown in figure (2) which elicited from a simple sample of an experiment the main instructions and transactions were applied onto the system were **Select, Insert, Update**, and **Delete**. each of these transactions has its rate and its reasons that made this transaction takes this rate of used by attackers. The maximum uses **Select** is come from the importance of these information which makes the attackers very interested to snoop on the data. On the other hand we have the minimum rate of attack attempts was recorded in **Delete** transaction, that came from that the attackers knowing perfectly that there is a backup for whole the system and their chance of deleting data will be very poor and it will be detected directly. If we take a look for **Update** transaction we can see that has not a few rate of attempts because it's easy to change the information as they want without deleting it, that will not be noticed by the workers over the system. **Insert** also has a middle rate of attempts because it just like the **Delete** transaction, it increases the number of the records which is easy to be noticed by the administrator.



The Relation between Table and Transaction

When we try to make a relationship between the tables and the applied transactions over it we found that there is a clear proportionality among them, for example we can see that there is a marked increase when using **Select** transaction on **Allocation** table that refers to the relation of allocation with the salaries as it mentioned previously. However, it is same for **Employees** and **Increases** in time it is increased slightly when talking about **Usere_Names**. Even though there is a big variation of using transactions over the tables but still the rates are nearly between tables expect of **Allocation** because of its strength relation with salaries amounts, also there are some tables have no relation with the transactions just like **Delete** over **Users_Names**.



CONCLUSION

In this paper, Constructing IDM include new approach to detect intrusion attempts depending on the user behavior, and this system is a view of a proposed hybrid IDM to detect intruders, since intruders have a clear effect on performance host resources and data.

REFERENCES

[1] M.A. North, "Data Mining for the Masses", Global Text, New York, 2012.

[2] W. Dubitzky, "Data Mining Techniques in Grid Computing Environments", Wiley-Blackwell. John Wiley & Sons, Ltd. Publication, 2008.

[3] S. Mitra, and T. Acharya, "Data Mining-Multimedia, Soft Computing, And Bioinformatics", A John Wiley & Sons, Inc. Publication, 2003.

[4] D.T. Larose, "Discovering Knowledge in Data: an Introduction to Data Mining", Wiley InterscienceA John Wiley & Sons, Inc. Publication, 2005.

[5] J.M. Żytkow, and J. Rauch, "Principles of data mining and knowledge discovery", 2010".

[6] H. Bensefia, and N. Ghoualmi, "A New Approach for Adaptive Intrusion Detection", Seventh International Conference on Computational Intelligence and Security, pp. 983-987, 2012.

[7] D. Gollmann, "Computer Security", WILEY A John Wiley and Sons, Ltd. Publication, 2011.

[8] E. Cole, R. Krutz, "Network Security Fundamentals", John Wiley & Sons, Inc., 2008.

[9] H. Al-Anie, "A Simulated Intrusion Detection System Using Packet Header", Ph.D. Thesis, Department of Computer Science, University of Technology, 2003.

[10] D. Moore, and C. Shannon, Code-Red: a case study on the spread and victims of an Internet worm. In Proceedings of the Internet Measurement Workshop (IMW), pages 273–284, 2002.

[11] W. Stallings, and L. Brown, "Computer Security Principles and Practice", Pearson, 2012.

[12] W. Easttom, "Computer Security Fundamentals", Pearson, 2012.

[13] S. Northcutt, and J. Novak, "Network Intrusion Detection", New Riders Publishing, 2003.

[14] G. Miao, and S. Guocong, "Energy and spectrum efficient wireless network design". Cambridge University Press, 2014.

Citation: *M. Jamal and A. Hadeel, "Design and Implementation Data Base Intrusion Detection System", International Journal of Emerging Engineering Research and Technology, vol. 5, no. 10, pp. 30-36, 2017.*