Design and Development of Data Mining Based Intrusion Detection and Intrusion Tolerant System

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

***Abstract*— In today information world, every mission critical information system deals with database products. Also almost every electronic commerce site has one (or more) database systems at the back office. Database systems motivated 32% of the hardware server volume in 1995 [2], and 39% of the server volume in 2000. Securing data is very important component for many of the database applications. Database security concerns the confidentiality, integrity, and availability of the data stored in a database is very critical for database security. A broad span of research from authorization, to inference control, to multilevel secure databases, and to multilevel secure transaction processing, addresses primarily how to protect the security of a database, especially its confidentiality. However, very limited research has been done on how to survive the set of successful database attacks that can seriously impair the integrity and availability of a database, and the ability of existing database systems to survive attacks is very poor**

***Index Terms*— Intrusion Detections, IDS, Intruder, Anomaly, Signature based, Data Mining**

I. INTRODUCTION

The visions of internet based applications and pervasive computing not only pushes computations from a computer into everywhere, but also maximizes the dependence on networked computer systems. Quickly increased complexity, openness, inter-connection, and inter-dependence has made these systems more vulnerable and difficult to protect. Current security mechanisms are inadequate to prevent every attack and the current traditional prevention-centric security is not enough and there is an urgent need for intrusion tolerant or attack resilient system [1]. The focus of such intrusion tolerant systems is the ability to continue delivering essential services in the face of attacks

Recently in [1], Liu has proposed four different types of architectures for intrusions tolerant Database systems [13]. Substantial technologies have been developed to detect operating system and network intrusions, but very few can be directly used to detect database intrusions, i.e., malicious transactions.

A significant challenge in providing an effective defense mechanism to a network perimeter is having the ability to detect intrusions and implement countermeasures. Components of the network perimeter defense capable of detecting intrusions are referred to as Intrusion Detection Systems (IDS) [7]. Intrusion Detection techniques have been investigated since the mid-80s and depending on the type and source of the information used to identify security breaches, they are classified as host-based or network based [12]. Host-based systems use local host information such as process behaviour; file integrity and system logs to detect events. Network-based systems use network activity to perform the analysis. Depending on how the intrusion is detected an IDS is further classified as signature-based (also known as misuse system) or anomaly-based [7]. Signature-based systems attempt to match observed activities against well-defined patterns which also called signatures [7]. Anomaly-based systems look for any evidence of activities that deviate from what is considered normal system use [6]. These systems are capable of detecting attacks for which a well-defined pattern does not exist

II. IMPORTANCE WITH INDUSTRY APPLICATION

The research in the proposed area of research is an important issue and it is very much essential for current business needs. There isseveral new attack techniques are coming out frequently. In order to protect the business systems, there is an urgent need to enhance the functionality existing IDS systems for protecting from these attacks. Most of the Intrusion Detection Systems that are used by industry are based on signature-base and anomaly detection. The advantage of Anomaly based Detection is that it will easily identify the patterns which are not matching with any of the existing patterns, hence classify as Intrusion. The proposed work implements Anomaly based Intrusion Detection [5], to provide worthwhile information about malicious network traffic and help identify the source of the incoming probes or attacks. This can also deals with the Host based Intrusions, where most organizations are now concentrating on. If any industry depends on networking, IDS is useful for protecting their information and reduce the damage. Most of the Intrusion Detection systems are classifying known attacks [14], but in our proposed system we are trying to classify

unknown attacks by applying some efficient clustering techniques, hence improved security.

This proposal will aim how existing intrusion detection techniques can be adapted to detect malicious transactions. The key challenge is how to capture and exploit transaction semantics

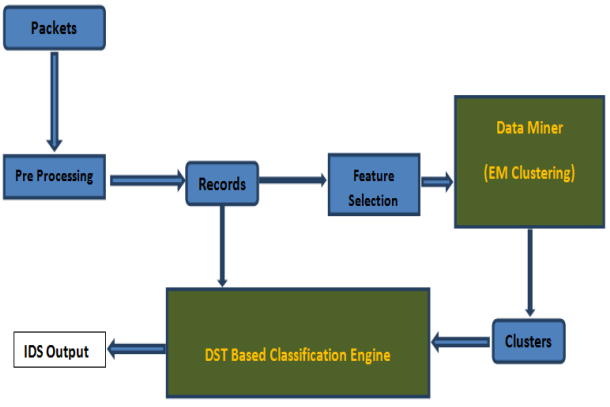
III. PROBLEM DEFINITION

With the rapid growth of the Internet and its related network infrastructure, timely detection of intrusions and appropriate responses has become extremely important. A security breach can cause mission-critical systems to be unavailable to end users causing millions of dollar worth of damage. If the next generation of the Internet and network technology is to operate successfully, it will require a set of tools to analyse the networks and detect and prevent intrusions. The *Dempster-Shafer theory in association with Data mining algorithms* provides a new method to analyse data from multiple nodes to estimate the likelihood of an intrusion. Dempster-Shafer theory can be used as an alternative to Fuzzy logic in dealing with vagueness of information.

IV. PROPOSED WORK

In the proposed work the work [8] on Fuzzy logic based reasoning can be replaced by Dempster-Shafer theory of evidence [5]. This Dempster-Shafer theory approach considers sets of propositions and assigns to each of them an interval [Belief, Plausibility] in which the degree of belief lies. Belief is a measure that gives the strength of evidence in favour of set propositions. It ranges from 0 to 1, where 0 indicating no evidence and 1 indicating certainty. The plausibility measures the extent to which evidence in favour of negation of S leaves no room for belief in the attribute S. Plausibility also ranges from 0 to 1. The belief – Plausibility interval measures not only the level of belief in some propositions but also the amount of information it has.

Based on the literature survey made on the proposed topic, tentatively the following architecture is proposed. There are four different module present in this architecture and are briefly described their functionality below.



Pre Processing: The Pre Processing Step is responsible for accepting packets either from Physical Media or tcpdump file and produces the records which contain the aggregate information about the packet groups. After the records are generated we will apply Feature Selection algorithm (i.e. Genetic Algorithms) to improve the Information Gain

Data Miner: In Data Miner module the Records will be clustered based on the selected features. Since we are dealing with Anomaly based Intrusion Detection System [1] the normal data records (i.e. attack free) are classified into clusters. Here we are using Expectation maximization clustering to get increased Detection rate and reduce false positives.

DST Based Classification Engine: The D-S Theory [5] is used to classify records with uncertain data into any one of the normal cluster [8]. Then we will classify the new records based on the Similarity Measure with respect to Normal Clusters. If the similarity measure of new record is less than or equal to the given Threshold value then give output as "NULL" otherwise "DETECTED" [8].

V. PROPOSED OBJECTIVES

In this research proposal, the following issues with reference to Intrusion detection system will take into account:

a) Proposed to design and development of a learning system to support Intrusion Detection System by using standard machine learning/ data mining techniques and their comparison. Building such a learning system will help to detect malicious transactions.

b) Enhancement of the functionality of database security for intrusion tolerant database systems (ITDB): The proposed learning system also throws some directions to decide the framework for masking to distinguish between database security and database consistency. That is to what extent the functionality of database security can be enhanced for ITDB can be determined.

c) Automatic discovery of knowledge using DST method or their associated data mining techniques and study of their impact for Intrusion Tolerant Database systems

d) To Measure and Increase the amount of Information contained in the Propositions using Dempster-Shafer Theory.

e) To be able to deal with Network based and Host based Intrusions effectively.

f) Collection of real world data from one or two business organizations for testing the proposed system.

g) To increase Detection Rate and reduce False Positives. References

VI. CONCLUSION

In today world, data is essential for analysis to make right decision. So, it is necessary for any organization to protect their data. Now a day every organization designing their own

data miming based intrusion detection system to secure their data. Techniques like DST, Data mining and Artificial Intelligence to make IDS robust.

ACKNOWLEDGMENT

I thank my organization support in all aspects for successful completion of my paper.

REFERENCES

[1] P. Liu "Architectures for Intrusions tolerant Database Systems" IEEE Proceedings of the 18th Annual Computer Security Applications Conferences (ACSA02), 2002.

[2] P. Stenstrom et al "Trends in shared multi-processing", IEEE Computer Vol 12:44-50, 1997

[3] YIonnaidis et al "Conceptual Learning in Database Design", ACM Trans on Database Systems, 265-293, 1991

[4] J Gomez et al "Evolving Fuzzy Classifiers for Intrusions Detection", IEEE Proc. on Workshop on Information Assurance NY, 2002

[5]. MrutyunjayaPanda andManasRanjanPatra,*A Novel Classification via Clustering Methodfor Anomaly Based Network Intrusion Detection System.* International Journal of Recent Trends in Engineering, Vol 2, No. 1, November 2009

[6]. E. Eskin. *Anomaly detection over noisy data using learned probability distribution.* In proceedings of the International Conference of the Machine Learning, 2000.

[7]. D.E. Denning. "An Intrusion Detection Model".IEEE Transaction on Software Engineering. SE13:222-232, 1987.

[8]. D. Barbara, N. Wu, S. Jajodia, Detecting Novel NetworkIntrusions Using Bayes Estimators, *First SIAMConference on Data Mining*, Chicago, IL, 2001.

[9]. Karl Sentz and Scott Ferson, *Combination of Evidence in Dempster-Shafer Theory*

[10].EvgeniyaPetrovaNikolova,VeselinaGospodinovaJechev a.*Anomaly Based Intrusion Detection Based on theJunction Tree Algorithm.*Proceedings of the International Multiconference on ISSN 1896-7094Computer Science and Information Technology, pp. 641 – 649 © 2007.

[11]. John Yen. A Reasoning Model Based On An Extended Dempster-Shafer Theory. AAAI-86 proceedings.

[12]. Y. Danalaxmi and Dr. I. Ramesh Babu. *Intrusion Detection Using Data Mining Along Fuzzy Logicand Genetic Algorithms.*International Journal of Computer Science and Network Security, VOL.8 No.2, February 2008.

[13]. *Data Mining Approach for Database Intrusion Detection*, in the national conference on data mining and applications held at Annamalai University, Chidambaram, Chennai in March 2006.

[14]. *Database Intrusion Detection System using Data Mining* in the 2nd International conference on Advanced Databases held at National University, US Educational services, San Diego, California during June 27-29, 2006.(pp 144-149).

K. Chandrasena chary, completed my M.Tech in Computer Science and Engineering from JNTUH, Hyderabad and currently working as Associate Professor in CSE department, Sree Chaitanya Institute of Technological Sciences, Karimnagar, Telanagana. India. My Interest areas include Data Mining, Information Security, Mobile computing and MANETS.