Data mining with Improved and efficient mechanism to detect the Vulnerabilities using intrusion detection system

Awan Dhawan

,

Abstract--Intrusion detection system must be capable of known and unknown vulnerabilities. We already studied the previous problems which includes detection of known vulnerabilities and unknown vulnerabilities. In order to obtain good accuracy a relevant or efficient dataset should be there to detect the known attacks and unknown attacks. Therefore, there are numerous security systems and intrusion detection systems that address different aspects of computer security. In this research work we proposed an approach that is sequential multilevel misuse detection model with the fuzzy rules for the detection of known and unknown attacks on the efficient intrusion dataset either kdd dataset, Nsl-dataset, tcp dump etc. Empirical studies show not much performance or accuracy of detecting the known and unknown attacks. We discuss the generation of a misuse detection models from pure normal data, and also discuss the generation of sequential multilevel misuse detection models along with fuzzy rules from data that contains known classes. We apply the proposed approaches to network-based intrusions. And the simulation of results is implemented on the good platform.

Keywords: fuzzy rule, data mining, intrusion detection, multilevel misuse detection model

I. INTRODUCTION

A. Data Mining: It is the process of determining and identifies the patterns from data and also called knowledge discovery in databases, the process of discovering or finding interesting and useful patterns and relationships in large volumes of data. This field also combines tools from statistical environment and artificial such as neural network and machine learning with database management to analyse large digital collections, known as data sets. For more information it can be seen on this [1]. Data mining is widely used in business insurance, banking, retail, science research astronomy, medicine, and government security detection of criminals and terrorists which is the best technology for finding the knowledgeable patterns.

The process of collecting, searching through, and analysing an amount of data in a database, as discover patterns relationships the use of data mining to detect fraud.

Data mining Task:

Data mining involves six common classes of task:

(i) **Anomalies and Attacks detection**: The identification or prediction of unusual data records that might be interesting or data errors and require further investigation.

(ii) **Association and combination rules**: Searches for relationships between variables or other values.

(iii) **Clustering Task**: Is the task of discovering groups and structures in the data that are in some way or another similar, without using known structures in the data.

(iv) **Classification criteria**: Is the task of generalizing known structure to apply to new data.

(v) **Regression:** Attempts or efforts to find a function which models the data with the least error.

(vi) **Summarization**: Providing a more compact representation of the data set, including visualization and report generation.

B. Intrusion Detection System (IDS) is very necessary because traditional firewalls cannot provide the complete security against the intrusion. Intrusion Detection (ID) is an active and important research area of network security. IDS make a real time response to intrusion events and intrusion processes.

The role of Intrusion Detection Systems (IDSs), as special-purpose devices to detect anomalies and attacks in the network, is becoming more important. Intrusion Detection Systems (IDS) is a combination of software and hardware that attempts to perform intrusion detection facilities. For more information it can be seen on the website [2].Intrusion detection is a process of gathering intrusion related knowledge occurring in the process of monitoring the events and analysing them for sign or intrusion. It raises the alarm when a possible intrusion occurs in the system.

C. Classification of Intrusion Detection System

Host based intrusion detection system: A host-based ID monitors all or parts of the dynamic behaviour and the state of a computer system. The information about whole system is in [3].HIDS might detect which program accesses what resources and discover that, for example, a word-processor has suddenly and inexplicably started.

Network based intrusion detection system: It is an intrusion detection system that attempts to discover unauthorized access to a computer network by traffic on the network for signs of malicious activity [3].NIDS are also referred as "packet-sniffers", because it captures the packets passing through the of communication mediums. This is based on the network. Stack based intrusion detection system: It is integrated closely with the TCP/IP stack, allowing packets to be watched as they traverse their way up the OSI layers [3]. This allows the IDS to pull the packets from the stack before the OS or the application has a chance to process the packets. This is purely based on stack implementation i.e., Last In First Out.

Application protocol based intrusion detection system: An application protocol-based intrusion detection system (APIDS) is an intrusion detection system focuses its monitoring and analysis on a specific application protocol or protocols in use by the computing system [3] modifying the system password database. So APIDS is basically based on the only particular system. It is based on the security, So no one intruders intrude the data without knowing the knowledge of the authentication process of the system.

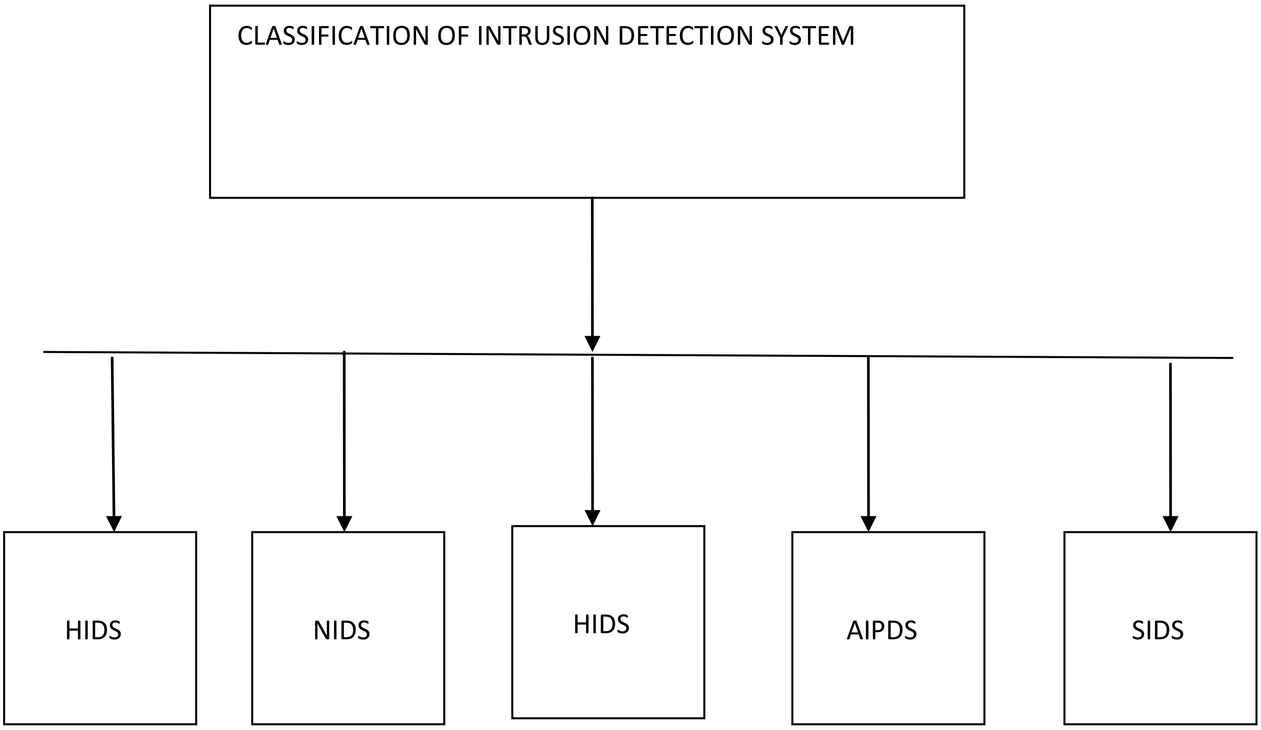

Figure 1Classification for intrusion detection system

D. Intrusion Detection through Data mining:

There are following areas where data mining is or can be employed: misuse/signature detection, anomaly detection, scans detection, etc. This is very efficient and fast way to find knows threats. The false positives happen, when normal network flow or system calls are marked as a threat[4] For example, a user can fail to provide the correct password for three times in a row or start using the service which is deviation from the standard profile. Novel attack can be define as an attack not seen by the system, meaning that signature or the pattern of such attack is not learned and the system will be penetrated without the knowledge of the administrator. Misuse detection can be built on your own data mining techniques.

II. PREVIOUS WORKS

Shan Suthaharan *et.al* (2012) in their study [5] evaluates the machine learning techniques for Intrusion Detection Systems. Tejaswi Panchagnula *et.al* (2012) They have worked on the Labelled datasets, play a major role in the process of validating and evaluating machine learning techniques in [6] intrusion detection system. Sandip Sonawane *et.al* (2012) Intrusion detection system (IDS) is a security layer that is used to discover ongoing intrusive attacks and anomaly activities in information systems and is usually working in a dynamically [7] changing environment. Asmaa Shaker Ashoor *et.al* (2011) Intruders computers, who are spread across the Internet have become a major threat in our world, the researchers proposed a number of techniques such as (firewall, encryption) to prevent such penetration and protect the infrastructure of computers, but with this, the intruders managed to penetrate the computers.[8]

III. PROPOSED METHODOLOGY

The overall goal of this work is to enhancement of the performance of the above issues or problems. Basic methodology consisted of first there will be a selection of any intrusion dataset like KDD'99 dataset, NSL-dataset, TCPdump etc**.** then there is a task of prediction and monitoring of the attacks like known attacks and unknown attacks by using the enhancement on the misuse detection model which does not monitor the unknown attacks, So The proposed methodology for the detection of the known and unknown attacks the sequential multilevel misuse detection model along with fuzzy logic if then else rules. With respect to the factors like performance, accuracy, speed and enhancement of detection of the known attacks and unknown attacks and detect either there is any new attack with the help of decision rules and fuzzy rules.

Then defining and simulating the basic scenario will be represented by the visualization criteria which will be implemented in the .net framework to develop a simulator and matlab be used for the performance analysis

Overall Research Design:

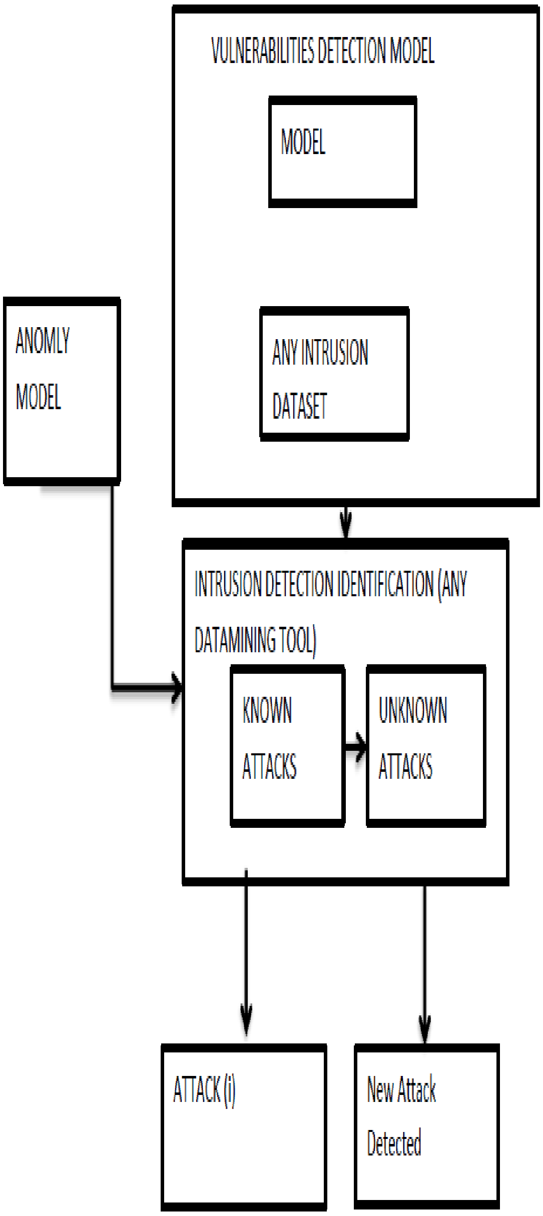

Figure 2. Overall Design

The overall design of a research consist of the intrusion Datasets Like KDD dataset, NSL-KDD dataset, TCP dump etc. that must be checked under the attacks .From These Datasets We just choose any dataset and apply attacks on that dataset with the help of proposed intrusion detection model that the sequential multilevel misuse detection model with fuzzy-rules.

ATTACK- N=1…..N
N- Total kinds of attacks already known
NEW ATTACK DETECTED- Give information about new attack.

IV. CONCLUSION

The earlier studies have shown the detection of some of the known attacks and to select the relevant features with the help of intrusion detection system and to do the visualization of results to enhance the performance and to increase the accuracy of datasets .Now we are interested to detailed analysis of the dataset like KDD dataset , NSL dataset, any TCP dump and to check the datasets and apply the relevant method or model to detect the both types of attacks like known attack or unknown attacks to provides the more accuracy ,speed ,performance enhancement ,user interaction criteria. The overall goal of the work is to just to measure the performance of the detection of the known attacks and unknown attacks on the different datasets.

This research is beneficial in following points:

I. To need for the user interaction.
II. To remove the problem of inaccuracy
III. To improvement the computational issues due to the non-contributing features.
IV. To enhance the speed and accuracy of detection of attacks.
V. To reduce the complexity of overall work by using decision criteria.
VI. To enhance the performance when suddenly unknown attacks comes into picture with the known attacks side by side.

REFERNCES

[1] http://www.britannica.com/EBchecked/topic/1056150/data-mining.

[2] http://en.wikipedia.org/wiki/Intrusion_detection_system.

[3] http://en.wikipedia.org/wiki/Intrusion_detection_system.

[4] http://www.r-bloggers.com/data mining-for-network-security-and-intrusion Science, vol. 1, Oct-2010.

[5] Shan Suthaharan and Karthik Vinnakota "An Approach for Automatic Selection of Relevance Features in Intrusion Detection Systems"in 2012(3):127-13ISSN: 2231 – 2587.

[6] Tejaswi Panchagnula." Relevance Feature Selection with Data Cleaning for Intrusion Detection System "International conference of computer science and information technology 2012.

[7] Sandip Sonawane , Shailendra Pardesh and Ganesh Prasad," A survey on intrusion detection techniques" World Journal of Science and Technology 2012, 2(3):127-13ISSN: 2231 – 2587.

[8] A. S. Ashoor and S. Gore, "Importance of Intrusion Detection system (IDS)". International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp.1-4 Jan-2011.

[9] Aleksandra Lazarevic, Vipin Kumar, Jaideep Srivastava." Intrusion detection survey" Computer Science Department, University of Minnesota.

[10] S. Suthaharan and K. Vinnakota, "An approach for automatic selection of relevance features in intrusion detection systems," in Proc. of the 2011 International Conference on Security and Management (SAM'11) pp. 215-219, July 18-21, 2011, Las Vegas, Nevada, USA.

[11] Duanyang Zhao," Analysis of an intrusion detection system"" in second interational conference on security and management 2012(3):127-13ISSN.

[12] A.A.Olusola, A.S.Oladele and D.O.Abosede, "Analysis of KDD '99 Intrusion Detection Dataset for Selection of Relevance Features". Proceedings of the World Engineering and Computer Science,vol. 1, Oct-2010.

[13] Mrudula Gudadhe, Prakash Prasad" A New Data Mining Based Network Intrusion Detection Model" International conference on computer science and technology.

[14] Lei Li, De-Zhang Yang, Fang-Cheng Shen," A Novel Rule-based Intrusion detection System Using Data Mining "international conference 2010.

Author Description

Awan Dhawan, Research Scholar, Done B.TECH (ECE) from RIMT-Maharaja Aggrasen Engineering College (**RIMT-MAEC**), mandi gobindgarh. Now doing M.Tech(ECE) from Al-Falah School of Engineering & Technology (AFSET), India,