

# Network Analysis System for Detecting Unknown Attack using SVM and Network Intrusion: A Review

S.P.Bhende<sup>1</sup> P.Kullurkar<sup>2</sup>

<sup>1</sup>M.Tech. Research Scholar <sup>2</sup>Assistant Professor

<sup>1,2</sup>Department of Computer Science & Engineering

<sup>1,2</sup>Vidarbha Institute of Technology, Nagpur, Maharashtra, India

**Abstract**— For network administration, one of the most critical component is intrusion detection due to large number of unknown attacks. These unknown attacks are not able to detect system so persistently damage our existing system. Past network-attacks had simple purposes of leaking personal information by attacking the PC or destroying the system or existing technologies to detect these attacks are based on pattern matching methods which are very limited. Because of this fact, detection rate becomes very low in the event of new and previously unknown attacks. To defend against these unknown attacks, in this paper, we plan to develop a big data based system for detecting attacks which are unknown to the existing system. Big data analysis techniques can extract information from a variety of sources to detect future attacks. This is done using previous learning about the attacks on which the system is trained previously and finding patterns about these attacks by using SVM method.

**Key words:** Intrusion Detection, Network Administration, Big Data, Network-Attack

## I. INTRODUCTION

Nowadays, there exists an extensive growth in using Internet in social networking. So network security is critical as society becomes increasingly dependent on computerized systems for its finances, industry, medicine, healthcare, e-commerce, bank transactions and other important aspects. Intrusion Detection is one of the most important parts in network security which is closely linked to the safe use of network services. Awareness of an attack is essential to being able to react and defend against attackers, In order to prevent attacks. To look for hidden attack patterns and trends, Network Defences can be further improved by utilizing two important concepts as Security Analytics and Intrusion Detection data. For forensic purposes, Intrusion Detection is also important in order to identify successful braches even after they have occurred. For example, it is important to know afterwards if information such as credit card data has already been stolen, in order to take additional precautions or possibly take law enforcement. Beyond detecting network-attacks, Attack Detection can very helpful in noticing abnormal system behaviour to detect undesired conditions

An intrusion can be any set of actions that threaten the security requirements (e.g., integrity, confidentiality, availability) of a computer/network resource (e.g., user accounts, file systems, and system kernels). Intruders have promoted themselves and invented innovative tools that support various types of attacks. So, efficient methods for intrusion detection (ID) have become an insisting need to protect our computers from intruders. Network Intrusion Detection System (NIDS) is one traditional IDS product which monitors for cyber threats at the network layer by

evaluating network traffic. Host-based Intrusion Detection System (HIDS) is second IDS product which monitors for cyber threats directly on the computer hosts by monitoring a computer host's logs of system, processes of system, files, or network interface. An IDS can monitor specific protocols like a web server's Hyper Text Transfer Protocol (HTTP); this type of IDS is called a Protocol-based Intrusion Detection System (PIDS). Like an Application Protocol-based Intrusion Detection System (APIDS), IDSs can also be specialized to monitor application-specific protocols.

For detecting pattern, the best pattern detection method is SVM which is a set of supervised learning method used for detection mechanism. From various advantages we use this method, SVM advantages are:1) In high dimensional spaces, it is very efficient.2)In number of cases, where number of samples still it is very effective.3)In the decision function also called as support vector, it uses a subset of training point so it is very efficient for memory. 4)It is versatile, because for decision function different kernel functions can be specified. Common kernels are provided, but it is also possible to specify custom kernels.

Intrusion Detection Systems (IDSs) are designed to defend computer systems from various network attacks and viruses of computer. Effective classification models or patterns is build by IDSs to distinguish normal behaviours from abnormal behaviours that are represented by network data. There are two primary assumptions in the research of intrusion detection: (1) user and program activities are observable by computer systems (e.g., via system auditing mechanisms), and (2) normal and intrusion activities must have distinct behaviours .The main function of IDS (Intrusion Detection System) is to protect the system, analyse and predict the behaviours of users. Then these behaviours will be considered an attack or a normal behaviour. For many years, Though IDS has been developed, managers maintain system inefficiently due to the large number of return alert messages Intrusion detection systems (IDSs) identify activities that violate the security policy of a computer system or network. IDS are a necessary complement to preventive security mechanisms because IDS detect attacks that exploit system design flaws or bugs.

Hence, intrusion detection (ID) is one of the effective method to become an insisting need to protect our computers from intruders.

There are two main detection strategies (approaches): that are currently used by network or host-based in-trusion detection systems (NIDS/HIDS):

### A. Signature-based

Signature-based is still the most common technique and focuses on the identification of known worst patterns. system that uses a blacklist approach which is liable to attacks for which the signature is unknown.

Most commercial IDS employ the misuse strategy in which known intrusions are stored in the systems as signatures so it is also known as misuse detection. The main idea behind misuse detection is to represent attacks in a form of a pattern or a signature in such a way that even variations of these attacks can be detected. Based on these signatures, this approach detects attacks through a large set of rules describing every known attack. The main disadvantage of the signature based approach is its difficulty for detecting unknown attacks

In signature-based, The system searches network traffics for patterns or user behaviours that match the signatures, if a pattern matched a signature; an alarm is raised to a human security analyst who decides what action should be taken based on the type of attack. In such systems, known intrusions (signatures) are provided and hand-coded by human experts based on their extensive experience in identifying intrusions. Current misuse IDS are built based on: expert systems. which use a set of rules to describe attacks, signature analysis where features of attacks are captured in audit trail, state-transition analysis which uses state-transition diagrams, coloured petri nets or case-based reasoning.

### B. Anomaly-based

Anomaly-based, which consists of monitoring system activity to determine whether an observed activity is normal or anomalous, according to a heuristic or statistical analysis, can be used to detect unknown attacks, but despite the significant research effort, such techniques still suffer from a high number of false positives. Furthermore, it is not fool-proof, as multiple malware samples use a communication channel that resembles legitimate traffic (e.g. over an SSL/TLS connection) and, thus, can easily evade such systems.

The main goal of the anomaly detection approach is to build a statistical model for describing normal traffic. Then, any deviation from this model can be considered an anomaly, and recognized as an attack. Notice that when this approach is used, it is theoretically possible to detect unknown attacks, although in some cases, this approach can lead to a high attack rate. This capability to detect unknown attacks has been the cause of the increasing interest in developing new techniques to build models based on normal traffic behaviour in the past years.

Anomaly-based strategy can identify novel intrusions. It builds models for normal network. Behaviour (called profiles) and uses these profiles to detect new patterns that significantly differentiate from them. This suspicious patterns may represent actual intrusions or could simply be new behaviours that need to be added to profiles. Current anomaly detection systems use statistical methods such as multivariate and temporal analysis to identify anomalies;

Two types of intrusion detection are compared as follows:

#### 1) Misuse Detection:

Characteristics of misuse detection is to use patterns of well-known attacks (signatures) to identify intrusions, Any match with signatures is reported as a possible attack. Drawbacks of misuse detection is:

- Not able to detect new attacks
- Obligation(want)signatures update

- Known attacks has to be hand-coded
- Strongly affect security analysts

#### 2) Anomaly Detection:

Characteristics of Anomaly Detection is to use deviation from normal usage patterns to identify intrusions, any significant deviations from the expected behaviour are reported as possible attacks.

Drawbacks of Anomaly Detection is :

- Selecting the right set system features to be measure is ad hoc and based on experience
- Between transactions it has to study sequential interrelation
- Strongly affect security Analysts

From the above discussion, we conclude that traditional IDS face many limitations. So by Applying Data Mining (DM) techniques on network traffic data is a promising solution that helps improve IDS and We plan to develop a big data based system for detecting attacks which are unknown to the existing system. This is done using previous learning about the attacks on which the system is trained previously and finding patterns about these attacks. Once the pattern learning process is done, we would apply these learned patterns to the new input stream in order to detect any unknown attacks. Hadoop will be used to process the data; it will first map the input dataset into code understandable patterns, and then reduce these patterns to get information about the intrusion type.

## II. RELATED WORK

A. Shengyi Pan, Thomas Morris, Uttam Adhikari, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems",

Author in this paper use signature based and specification approach to build a hybrid IDS that learns temporal state-based specifications for power system scenarios including disturbances, normal control operations, and cyber-attacks. A data mining technique called common path mining is used to automatically and accurately learn patterns for scenarios from a fusion of synchrophasor measurement data, and power system audit logs. As a proof of concept, an IDS prototype was implemented and validated. The IDS prototype accurately classifies disturbances, normal control operations, and cyber-attacks for the distance protection scheme for a two-line three-bus power transmission system. In this way author develope a hybrid IDS and used common path mining.

B. Ahmed Youssef, Ahmed Emam, " Network intrusion detection using data mining and network behaviour analysis",

Author in this paper uses combination of most two powerful approaches as data mining and network behaviour analysis to detect intrusion in network .Firstly In traditional network systems, NBA can help to cover the gap and significantly enhance the value of the data generated from IDS . Secondly, Using DM as intrusion detection technique by analyzing and correlating large amount of sequence. In this way author uses two approaches to detect network intrusion.

C. Vahid Golmah, "An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM",

Author in this paper use hybrid method C5.0 and SVM to improve the accuracy of the intrusion detection system. The method C5.0 is used in a large number of input fields and for handling missing data. For detecting intrusion, firstly the data set is passed through the C5.0 and generated node information. This information is determined according to the rules generated by the C5.0 and represent the particular class. This node information (as an additional attribute) along with the original set of attributes is passed through the SVM to obtain the data intrusion free. In this way author uses two hybrid approaches C5.0 and SVM for detecting intrusion.

D. Shaohua Teng, Hongle Du, Naiqi Wu, Wei Zhang, Jiangyi Su, "A Cooperative Network Intrusion Detection Based on Fuzzy SVMs",

Author in this paper firstly pre-processes the data and then the fuzzy membership function are used called v-FSVM then according to TCP, UDP and ICMP protocol, three types of detecting agents are generated. Finally used KDD CUP 1999 data set then intrusion should be detected from network. In this way author uses Fuzzy SVMs that reduce training time and storage space and improve the classification accuracy.

E. Vikas Belwal, Sandip Mandal, "An Improved Approach for Detecting Unknown Attacks Using Feature Extraction Scheme and Fuzzy-Neural Networks"

Author in this paper, proposing a novel approach for detection of unknown attacks which combines Feature Extraction Scheme and Fuzzy-Neural Networks (FNN) along with K-means Clustering and Support Vector machine. Which can be implemented in modern IDSs to reduce the false positive rate to an extent and thus improve the efficiency of the IDSs.

### III. PROBLEM DEFINITION

Existing systems uses number of developing method. In existing system use algorithms like common path mining for pattern matching but for power system not for network window's. Real time data mining technique, and SVM use in combination with K-means clustering, C5.0 are used for developing IDS. In some existing system, detect intrusion using technique data mining and with various approaches called network behavior analysis, signature based and specification, Extraction Scheme. In all existing system, Not detected unknown attack on any network windows using SVM method.

### IV. PROJECT OBJECT

The objective of proposed techniques is,

- To read dataset, map technique is developing.
- To reduce datasets and For finding patterns, Reduce technique is Developing.
- For creating a database of these patterns, learning algorithm is developing.
- For matching pattern "support vector machine" is used.
- Immediately unknown attack is detected by using SVM.

- Unknown attack is detected on any network window.

### V. INVESTIGATIONAL OUTCOME

To achieve the objective of this project, we have proposed technique called SVM ("support vector machine") for mapping pattern and to develop a big data based system for detecting attacks which are unknown to the existing system. This is done using previous learning about the attacks on which the system is trained previously and finding patterns about these attacks. Once the pattern learning process is done, we would apply these learned patterns to the new input stream in order to detect any unknown attacks. Hadoop will be used to process the data; it will first map the input dataset into code understandable patterns, and then reduce these patterns to get information about the intrusion type.

### VI. CONCLUSIONS

This review paper proposes a technique to prevent the unknown attack on any network windows. We will use an SVM ("support vector machine") algorithm for detecting unknown attack immediately. The SVM algorithm will be a set of supervised learning method used for detection mechanism and mapping process. Firstly big data based system will develop that detecting attacks which are unknown to the existing system. This is done using previous learning about the attacks on which the system is trained previously and finding patterns about these attacks. Once the pattern learning process is done, we would apply these learned patterns to the new input stream in order to detect any unknown attacks. SVM ("support vector machine") will be used to process the data, it will first convert the input dataset into code understandable patterns, and then process these patterns to get information about the intrusion type and finally detect unknown attack for any network windows.

### REFERENCES

- [1] Shengyi Pan, Thomas Morris, Uttam Adhikari" Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems" IEEE TRANSACTIONS ON SMART GRID
- [2] Ahmed Youssef and Ahmed Emam." Network intrusion detection using data mining and network behaviour analysis" International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011.
- [3] Vahid Golmah," An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM" International Journal of Database Theory and Application Vol.7, No.2 (2014), pp.59-70 <http://dx.doi.org/10.14257/ijdta.2014.7.2.06>.
- [4] Shaohua Teng, Hongle Du, Naiqi Wu, Wei Zhang, Jiangyi Su," A Cooperative Network Intrusion Detection Based on Fuzzy SVMs" journal of networks, vol. 5, no. 4, april 2010
- [5] Vikas Belwal, Sandip Mandal," An Innovative and Efficient Approach for Detecting Unknown Attacks Using Feature Extraction Scheme and Fuzzy-Neural Networks" International Journal of Computer Trends and Technology (IJCTT) – volume 23 Number 2 – May 2015