A Review on IDS By Correlation & KPCA with Neural Network Optimized By Genetic Algorithm

Harpreet Kaur
M. Tech. Scholar
Rayat Bahra University,
Mohali, Punjab, India,
Harpreetmultani6@gmail.com

Gaganpreet Kaur Bhalla
Assistant Professor, Rayat
Bahra University, Mohali,
Punjab, India,
gaganb6@gmail.com

ABSTRACT
With the tremendous growth of the usage of computers over network and development in application running on various platform captures the attention toward network security. This paradigm exploits security vulnerabilities on all computer systems that are technically difficult and expensive to solve. Hence intrusion is used as a key to compromise the integrity, availability and confidentiality of a computer resource. The Intrusion Detection System (IDS) plays a vital role in detecting anomalies and attacks in the network.In this paper, I will discuss about data mining concept is integrated with IDS with their different features, feature extraction, advantages of data mining with IDS , some results and tools used for data mining with IDS and their applications.

Keywords
Feature Extraction, Intrusion Detection System(IDS), Machine Learning

INTRODUCTION
The network security is becoming an essential need of modern society to protect the confidential information flowing over the networks. Detection of Intrusion over the network is one the most extremely important task to prevent their unlawful use by the attackers. Efficient intrusion detection is needed as a defense of the network system to detect the attacks over the network. A feature selection and classification based Intrusion Detection model is presented, by implementing feature selection, the dimensions of NSL-KDD data set is reduced then by applying machine learning approach, we are able to build Intrusion detection model to find attacks on system and improve the intrusion detection using the captured data. With the increasing number of new unseen attacks the purpose of this model is to develop a system for intrusion detection, and the model will be capable of detecting new and previously unseen attacks using the basic signatures and the features of known attacks.

The important and valuable information always attract attackers and is always liable to maximum attacks over the network. Intrusion is getting into the system or system server by the attacker by sending the malicious packet to the user system and then stealing, corrupting or modifying any confidential information or important information, the sending of network packet over the network for illegal purpose is known as attack. The intrusion can occur over the system or server due to any existing system weakness or vulnerability, such as system misconfiguration, user misuse or program defects. An intelligent intrusion can also be made by putting multiple vulnerabilities together. In a global network there are millions of big servers and large number of on-line services running in the system while such networks attract more attackers and need intelligent intrusion detection model as a defense for their network system.

An Intelligent intrusion or system attack includes following step:

- **Collecting Information**: Collecting information about the target getting all the knowledge and details about the user who is attacked. This can be done by using the query tools like ‘‘whois”, ‘‘nslookup” or by using network commands in command prompt to get IP addresses, domain name server etc.[3]

- **Probing and scanning:** Scanning the Target host and check the unguarded or unprotected area on system and seek for the sensitive information in them.
- **Remote to Local access**: It is gaining the access of user system by R2L (Remote to Local) type of attack, like password guessing, network sniffing, buffer overflow attack, etc. An R2L attack means a person who is unknown to user machine send the network packet to get local access of user machine to execute command on a target. This attack can be done by using the system vulnerabilities, using open ports of the target machine, password guessing etc.

- **User to Root access**: In this attack a normal user of system tries to gain a root access of the system by using system vulnerabilities. These attacks are quite similar to R2L attack but in this attacker are already a normal user of machine and try to gain root access of machine.

- **Launch attacks**: Finally attacks are made like stealing confidential information, modifying web pages, accessing another person accounts and creating a backdoors for future attacks.

An Intrusion Detection System (IDS) is security technique to detect the attacks over the network. Intrusion detection has

been classified under two categories, namely misuse detection and anomaly detection.

Types of Intrusion Detection:

Intrusion detection systems (IDS) can be classified into different ways. The major classifications are:

- Active and passive IDS
- Network Intrusion detection systems (NIDS) and host Intrusion detection systems (HIDS)

Active and passive IDS

Active IDS: An active Intrusion Detection Systems (IDS) is also known as Intrusion Detection and Prevention System (IDPS). Intrusion Detection and Prevention System (IDPS) is configured to automatically block suspected attacks without any intervention required by an operator. Intrusion Detection and Prevention System (IDPS) has the advantage of providing real-time corrective action in response to an attack.

Passive IDS: A passive IDS is a system that's configured to only monitor and analyze network traffic activity and alert an operator to potential vulnerabilities and attacks. A passive IDS is not capable of performing any protective or corrective functions on its own.

Network Intrusion detection systems (NIDS) and Host

Intrusion detection systems (HIDS)

1. Network based: Network based intrusion detection system (NIDS) monitors is used to monitor the information flowing over the internet network and detect the intrusions. In a network-based system, or NIDS, the individual packets flowing through a network are analyser. In this method NIDS is applied before the Firewall, so that it can examine all the data packets flowing through the network.

2. Host based: A host-based intrusion detection system (HIDS) is an intrusion detection system that monitors and analyses the internals of a computing system as well as (in some cases) the network packets on its network interfaces. This was the first type of intrusion detection software to have been designed, with the original target system being the mainframe computer where outside interaction was infrequent.

Intrusion Detection System is always work before the firewall. Because if IDS finds attacks it alerts the administrator for attacks.

The drawbacks of Host Intrusion Detection Systems (HIDS) are

• Difficult to analyses the intrusion attempts on multiple computers.

• Host Intrusion Detection Systems (HIDS) can be very difficult to maintain in large networks with different operating systems and configurations

• Host Intrusion Detection Systems (HIDS) can be disabled by attackers after the system is compromised.

Knowledge-based (Signature-based) IDS and behavior-based (Anomaly-based) IDS

A knowledge-based (Signature-based) Intrusion Detection Systems (IDS): A knowledge-based (Signature-based) IDS references a database of previous attack signatures and known system vulnerabilities. The meaning of word signature, when we talk about Intrusion Detection Systems (IDS) is recorded evidence of an intrusion or attack. Each intrusion leaves a footprint behind (e.g., nature of data packets, failed attempt to run an application, failed logins, file and folder access etc.). These footprints are called signatures and can be used to identify and prevent the same attacks in the future. Based on these signatures Knowledge-based (Signature-based) IDS identify intrusion attempts.

The disadvantages of Signature-based Intrusion Detection Systems (IDS) are signature database must be continually updated and maintained and Signature-based Intrusion Detection Systems (IDS) may fail to identify a unique attacks.

A Behavior-based (Anomaly-based) Intrusion Detection Systems (IDS): A behavior-based (Anomaly-based) IDS references a baseline or learned pattern of normal system activity to identify active intrusion attempts. Deviations from this baseline or pattern cause an alarm to be triggered. Higher false alarms are often related with Behavior-based Intrusion Detection Systems (IDS).

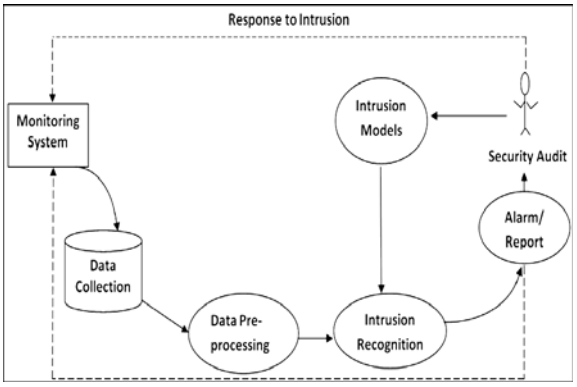

Figure 1 Overall structure of Intrusion Detection System

Tools of Data Collection & Analysis

Various tools are needed for that project some for analyzing data, some for designing, implementation and some developing software tool these are:

Matlab 7.1: Matlab is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C,C++,Java, Fortran.

IDS: Intrusion Detection System is a device or software application that monitors network or system activities for malicious activities or policy violations and produces electronic reports to a management station. IDS come in a variety of "flavors" and approach the goal of detecting suspicious traffic in different ways.

KPCA : Kernel Principal Component Analysis is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping.

Neural network: Artificial neural networks are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning.

Genetic Algorithm: A genetic algorithm is a search heuristic that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions to optimization and search problems.[1]Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover.

Advantage of Data Mining In Academics

Data mining tells us following things like:

- Tells us about the features of IDS.
- Helps in improving the performance.
- Helps in error detection.
- Helps for detecting the attacks in feature extraction.
- Improving the model of IDS.

CONCLUSIONS

The current IDS system doesn't work in signature based IDS system. The system doesn't deal with the misuse detection concept. The IDS system has no efficient method for decreases the complexity due the features of IDS. The current IDS system doesn't used the kernel based feature extraction and ranking based feature selection. Another main problem in IDS is number of features, as the number of features increases complexity increases, so it degrades the performance. The complexity also increases errors in the security system. The proposed method used the signature based IDS system. The system uses the feature selection and feature extraction method for the misuse detection. The proposed method will be able to detect the attack on the basis of the behavior of the basic features of network. The proposed method used the kernel based feature extraction and ranking based feature selection. Errors will be removed in proposed method by using neural network and genetic algorithm.

REFERENCES

[1] G.V. Nadiammai, M.Hemalatha "Effective approach toward Intrusion Detection System using data mining techniques," Egyptian Informatics Journal, pp.15,37-50, December 2013.

[2] L.Dhanabal1, Dr. S.P. Shantharajah "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2015.

[3] Jayveer Singh1, Manisha J. Nene "A Survey on Machine Learning Techniques for Intrusion Detection Systems," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2013.

[4] Tavallaee,M, Bagheri, E, W, Lu, and Ali A,G, "A Detailed Analysis of the KDD CUP 99 Data Set" IEEE Symposium on computational intelligence in security and defense application, pp. 1-6, 8-10 July 2009.

[5] Anup K. Ghosh, Schwartzbard A, Schatz M, "Workshop on Intrusion Detection and Network Monitoring," Santa Clara, California, USA.,1999.

[6] Kumar J,D, "Attack Development for Intrusion Detection Evaluation Attack Development for Intrusion Detection Evaluation", Massachusetts Institute of Technology,2000.

[7] Shrivastava, G, Sharma, K,R,S, "The Detection & Defense of DoS & DDoS Attack: A Technical Overview," In Proc. of ICC, 27-28, pp. 274-282 ,December 2010.

[8] Sharmila, K, Wagh, Vinod K, Pachghare, Satish R, Kolhe, "Survey on Intrusion Detection System Using Machine Learning Techniques," IJCA International Journal of computer Applications, vol.78, no.16, pp. 30-37, Sept 2013.

[9] Bajaj, K. Arora, A., "Dimension Reduction in Intrusion Detection Features Using Discriminative Machine Learning Approach," IJCSI International Journal of Computer Science Issues, Vol. 10, no. 4, pp. 324-328, July 2013.

[10] Chen, M. C, Chen, Y. L, Chung L. H, "An efficient network intrusion detection," Elsevier, Computer Applications, vol.33, no.4, pp. 477-484, March 2010.

[11] Qiang Wang, Vasileios Megalooikonomou. A clustering algorithm for intrusion detection. In: International conference on data mining, intrusion detection, information assurance, and data networks, security, 5(12), 2005, p. 31–8.

[12] Ching-Hao, Hahn-Ming L, Devi P, Tsuhan C, Si-Yu H. Semi supervised co-training and active learning based approach for multi-view intrusion detection. In: ACM symposium on applied computing, no. 9; 2009. p. 2042–7.

[13] Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang. Semisuper-vised learning for false alarm reduction. In: Industrial conference on data mining, no. 10; 2010. p. 595–605.

[14] Li Jimin, Zhang Wei, KunLun Li. A novel semi-supervised SVM based on tri-training for intrusion detection. J Comput 2010;5(4): 638–45.

[15] Monowar H. Bhuyan, Bhattacharyya DK, Kalita JK. An effective unsupervised network anomaly detection method. In: International conference on advances in computing, communications and informatics, no. 1; 2012. p. 533–9.

[16] Lane T. A decision-theoretic, semi-supervised model for intrusion detection. In: International conference on machine learning and data mining for computer security; 2006. p. 157–77.

[17] Zhang Fu, Marina Papatriantafilou, Philippas Tsigas. Off-thewall: lightweight distributed filtering to mitigate distributed denial of service attacks. In: IEEE international symposium on reliable distributed systems, no. 31; 2012. p. 207–12.