International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 4, Volume 2 (April 2015) www.ijirae.com

Anomaly Network Intrusion Detection: A review

Vidhya N. Gavali*
Computer Dept, Pune University,

Sunil Sangve
Computer Dept, Pune University,

Abstract— the Intrusion Detection System (IDS) is tool which detects an unauthorised, misuse of computer system and provides information security. An intrusion detection system (IDS) is combined with hardware and software elements that work together to find unexpected events which may indicate an attack will happen, is happening, or has happened. Network intrusion detection based on anomaly detection procedures has a important part in securing systems and networks against damaging behavior. Distinctive metaheuristic strategies have been utilized for anomaly detector generation. Here, an integrated approach is studied for anomaly detection in extensive scale datasets utilizing indicators produced focused around multi-start metaheuristic strategy and Genetic algorithms. The proposed methodology has taken some motivation of negative selection based detection generation. The assessment of this methodology is performed utilizing NSL-KDD dataset which is an altered version of the broadly utilized KDD CUP 99 dataset. It also to increase its adaptability and flexibility the studied parameter value selected automatically according to the used training dataset. And also decrease the detection generation time by enhancing the clustering.

Keywords— Intrusion Detection System (IDS) ,Anomaly Detection, NSL-KDD, Metaheuristic Strategies.

I. INTRODUCTION

Now a day, security problem becomes a major issue due to large amount of use of internet and computer system. Any network attacks on a system violets integrity, confidentiality, and availability. To decrease such an influence on a network we need intrusion detection system (IDS). There are various types of intrusion detection system like host based IDS, Network based IDS. The Host based IDS run on separately on system. The Network based IDS monitors traffic on a network for any suspicious activity.

Signature-based schemes search for defined patterns, or signatures [6]. So, its use is preferable in known attacks but it is incapable of detecting new ones even if they are built as minimum variants of already known attacks. On the other hand, anomaly-based detectors try to learn system's normal behavior and generate an alarm whenever a deviation from it occurs using a predefined threshold [6]. Anomaly detection can be represented as two-class classifier which classifies each sample to normal or abnormal .It is capable of detecting previously unseen intrusion events but with higher false positive rates compared to signature-based systems.

Metaheuristics are nature inspired algorithms based on some principles from physics, biology or ethology. Metaheuristics are categorized into two main categories, single solution- based and population-based metaheuristics. Population-based metaheuristics are more appropriate in generating anomaly detectors than single-solution-based metaheuristics because of the need to provide a set of solutions rather than a single solution through recombination and mutation operators. Genetic algorithms, evolutionary programming, genetic programming, scatter search and path relinking, coevolutionary algorithms and multi-start framework are examples of EC algorithms [5] .Genetic algorithms (GAs) [5] are widely used as searching algorithm to generate anomaly detectors. Generating anomaly detectors requires a high-level solution methods (metaheuristic methods) that provide strategies to escape from local optima and perform a robust search of a solution Multi-start procedures, as one of these methods, were originally considered as a way to exploit a local or neighborhood search procedure (local solver), by simply applying it from multiple random initial solutions. [7]NSA detectors are formed with different geometric shapes such as hyper-rectangles, hyper spheres, hyper-ellipsoids or multiple hyper-shapes. The size and the shape of detectors are selected according to the space to be covered.

This paper exhibits a hybrid methodology to anomaly detection utilizing a genuine esteemed negative choice based generation detector. The arrangement particularly addresses issues that emerge in the connection of extensive scale datasets. It utilizes k-means clustering to lessen the extent of the preparation dataset while keeping up its differing qualities and to recognize great beginning stages for the identifier generation focused around a multi-begin metaheuristic system and a genetic algorithm. It utilizes a lessening venture to evacuate excess finders to minimize the quantity of created locators and accordingly to decrease the time required later for online anomaly detection.

II. LITERATURE SURVEY

The Anomaly detection methods are classified into several types. One of the methods among them is Statistic-based method. It identifies the intrusion by using the predefined threshold, standard deviation, mean, and the probabilities [9], [10]. Another category is Rule-Based methods. It uses the If-Then and If-Else rules, in order to construct the model of detection for some previously known intrusions [12] [11]. Additionally, the State-Based approach is also there. It makes

the use of Finite state machine, which is derived from the network topologies to determine the attacks [13]. In addition, heuristic-based approach is also a category for use [14].

A. Intrusion Detection System using Support Vector Machines and feature selection method :

Yinhui Li et al. [1], proposed the intrusion detection system using support vector machine and feature selection method. The feature selection method chooses 19 critical features. The KDD Cup 99 data set is used to calculate the results. The features of KDD Cup 99 data set are:

From 1999, this dataset is mostly used in network based IDS to detect intrusion or anomaly. The KDD Cup 99 was invented by Stolfo et al. and implemented based on DARPA 98.

KDD Cup 99 contains:

1. 4,900,000 single connection vectors
2. Each single vector has 41 features which is labeled as normal or abnormal
3. Test data set with 300,000 samples and 24 training attack types.
4. 14 additional attack types in test set.

Steps to detect intrusion:

1. Data Preprocessing: To reduce data redundancy from KDD Cup 99 data set and to create more efficient dataset clustering technique is used. The K-mean clustering is used to make 5 data cluster and referred as compact data set.
2. Small data set construction: Small sample data set is chosen to represent whole data set and each sample has ability to represent whole data set. Ant colony optimization (ACO) method is used to select proper small training data set.

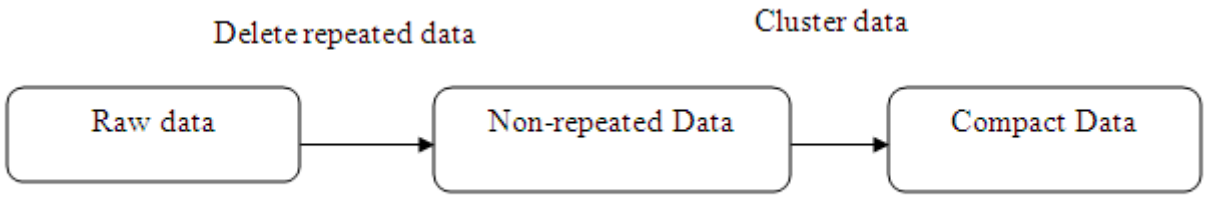

Fig.1 Data Pre-processing

3. Feature reduction strategy: feature removal is done with four different strategies, the feature removal method, the sole feature method, the hybrid method and the GFR method.
4. The data set is trained and tested with 41 features.

Yinhui Li et al. [1] extended the problem of Current strategy of small training data setting is not adaptive to complex program in multiple classification problems, especially in the unbalanced circumstances.

B. intelligent algorithm with feature selection and decision rules for anomaly detection [2]

Anomaly intrusion detection system is used to detect new attacks on a network. Anomaly detectors learn system normal behavior and generate alarm when there is deviation from normal behavior. Shih-Wei Lin et al. [2], proposed an intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection. They take the advantages of support vector machine (SVM), decision tree (DT), and simulated annealing (SA). The results were getting from KDD 99 dataset. The Support vector machine and simulated annealing is used to find best feature and evaluate the accuracy of anomaly detection.

Support vector machine is based on principal of risk minimization and learning machine to map training set using some nonlinear mapping. The Decision Tree (DT) algorithm is used for classification. DT is one of the artificial intelligence algorithms. The decision rules are used to detect new attacks. KDD data set consists of four types of attack.

C. Intrusion detection system based on hierarchical clustering and support vector machines [3]

Shi-Jinn Horng et al. [3], proposed the IDS with the combination of hierarchical clustering and support vector machines to increase detection accuracy. The experimental results are based on KDD 99 dataset. The result has shown better performance for detection of DoS and Probe attacks. The BIRCH Hierarchical clustering algorithm is used and it was firstly introduced by Zhang et al. (1996) [4]. A BIRCH algorithm is used to reduce training data set so that training data set processing complexity is reduced. The SVM algorithm has high detection accuracy and low false positive rate. The following steps are followed:

Table 1: KDD dataset attack types [2]

Sr. No	Attack Type	Attack name
1.	Denial of service	1. Back 2. land 3. neptune 4. pod 5. smu rf 6. teardrop
2.	Probing	1. ipsweep 2. nmap 3. por tsweep 4. satan
3.	User to Root (U2R)	1. rootkit 2. perl 3. load module 4. buffer-overflow
4.	Remote to Local (R2L)	1. ftp-write 2. spy 3. phf 4. guess-passwd 5. imap 6. warezclient 7. warezmaster 8. multihop

1. Transformation of non-continuous attributes into continuous attributes.
2. Construct tree for U2R, R2L, probing and DoS and one tree for Normal packets.
3. Do feature selection for each type of attacks.
4. Train four SVM classifiers separately corresponding to the four kinds of attacks, by the centroid of all entries in leaf nodes of CF trees.
5. Combine the four SVMs classifiers to build an intrusion detection system.

The following figure describes the classification using SVM. It classifies data into two categories. Labeled pairs {(x, y)}, where y is the label of instance x, SVM works by maximizing the margin to obtain the best performance in classification [3].

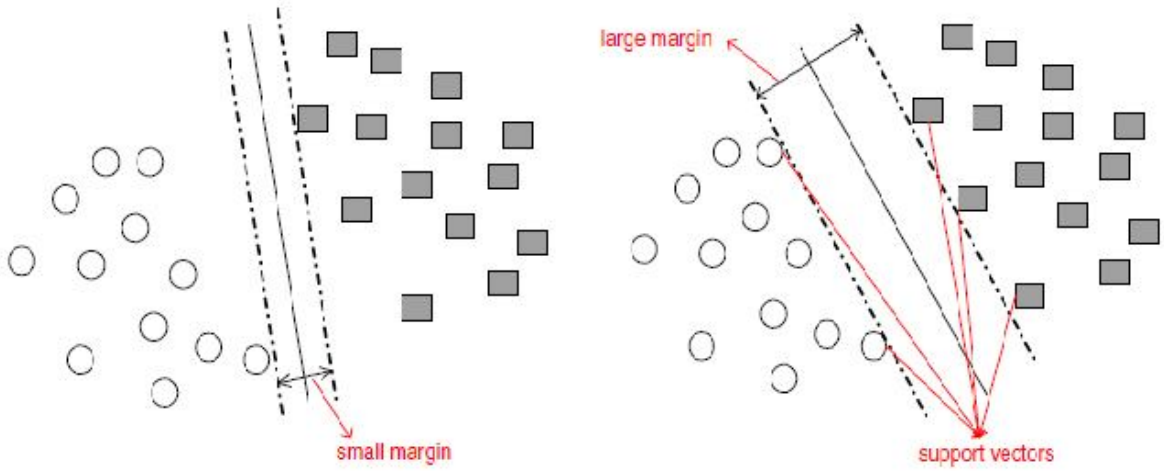

Fig.2: Support Vectors and margin maximization [3]

D. Anomaly detection using a metaheuristic algorithm [5]

A new anomaly detector generation approach is proposed based on negative selection algorithm concept. As number of detectors is playing a vital role in the efficiency of online network anomaly detection, the studied metaheuristic approach aims to generate a suitable number of detectors with high detection accuracy. The main idea is based on using K-means clustering algorithm to select a reduced training dataset in order to decrease time and processing complexity. Also, k-means is used to provide a way of diversification in selecting initial start points used by multi-start methods. Moreover, the radius of hyper-sphere detectors, generated using multi-start, is optimized later by genetic algorithm. Finally, rule reduction is invoked to remove unnecessary redundant detectors. Detector generation process is repeated to improve the quality of detection. Preprocessing, clustering and training dataset selection, detectors generation and optimization, rules reduction, training and test dataset evaluation are the main stages of the proposed approach.

Here, training data source (DS) is normalized to be ready for processing after the preprocessing clustering and training dataset is selected in order to decrease time complexity and number of detectors to be generated in later stages, small sample training dataset (TR) should be selected with a good representation of the original training dataset. After that detector generation and optimization process is done here Multi-start searching algorithm focuses on strategies to escape from local optima and perform a robust search of a solution space. So, it is suitable for generating detectors which is used later to detect anomalies. Reducing the number of detectors is a must to improve effectiveness and speed of anomaly detection that's why detection reduction is done and evaluate training and test dataset.

The Intrusion Detection System plays a very significant role in identifying attacks in network. There are various techniques used in IDS like signature based system, anomaly based system. But Signature based system can detect only known attack, unable to detect unknown attack but anomaly based system is able to detect attack which is unknown. Here Anomaly based system with integrated approach using multi-start metaheuristic method is defined. There are different metaheuristic techniques introduced but no one has mentioned the use of metaheuristic method for detector generation (Tamer F. Ghanem et al., 2014). Thus to increase not only detection accuracy but also to reduce false alarm rate and detector generation time metaheuristic technique have a significant role.

Table 2: Types of attack in Network

Sr.No.	Intrusion attack	Description with example
1	Worms	It is a program which replicates themselves from computer to computer without use of host file. The entire document such as word or excel vary from one system to another system should consider as worm. E.g. PrettyPark
2	Virus	It is a small program written to a change the way of computer operates without permission of user. Virus executes itself and replicates itself. E. g virus replaces the executable file with copy of infected file. Viruses affects on computer by damaging program, deleting files, and reformatting the disk. There are also types of viruses File infector viruses, Boot sector viruses, Master boot record viruses, Macro viruses. E. g Polyboot.B
3	Trojans	Trojans have a malicious code, when triggered it causes loss or even theft of data. E. g Mail Bomb
4	Physical Attack	It is an attempt to damage the physical components of networks or computers. E. g Cold boot
5	Password Attack	In this attack, password is obtained within a short period of time which is indicated by a sequence of login failure. E. g SQL injection attack
6	Remote to Local (R2L) attack	Packets are sent to remote place through a network and while sending packet it does not have any account on the system. The access to the system as a user or root and perform the harmful activity. The attacks on public services like HTTP or FTP. E. g spy
7	User to Root (U2R) attack	Able to violet privilege levels of the system to gain access of super user (sniffing passwords, dictionary attack, or social engineering) E. g Perl
8	Information Gathering attack	Able to gather information to find out intrusion or vulnerabilities. These activities are done by scanning computers or networks. E. g XMAS scan
9	Probe	Able to scan network to get valid Internet Protocol address (IP) of the system. Then provide the information to the system about list of vulnerabilities so that they able to do attack on system. E. g Portsweep

10	Bots	Bots can be used for either good or malicious purpose. A malicious bot is self-propagating malware. It is designed to infect a host. Bots connect back to a central server or servers that act as a command and control (C&C) center for a whole network of compromised devices. With the help of a botnet, attackers can launch broad-based, remote-control, flood-type attacks against their targets.
11	DoS	An attempt to prevent genuine users of a service from using that service. When this attack comes from a single host or network node, then it is referred to as a DoS attack.
12	Malware	Malware is a program written intentionally to harm system. It is not buggy software or programs.
13	Vulnerability	Program written by humans. Programmers are sometimes forget to cross t's and dot i's and those mistakes create strange behavior in programs, it create a hole that malware or attacker could use to access system more easily, known as a vulnerability.
14	Exploit	The strange behavior that can be used to create a hole for attacker or malware to get a particular sequence of actions or text to cause the right (or is that wrong) condition. This text needs to be put into code form, which is also called exploit code.
15	Phishing	The attacker creates a fake web site which looks like a exactly similar to original website. In phishing, the attacker sends an email to the computer user with a link. When user click on link to log on with their username and password, the attacker records username and password and tries to access information on real site.
16	Hijack attack	The attacker is places between two computer users. When these two user are communicating with each other, at that time attacker disconnects second individual person and the first user feels like that he still communicate to the original person and may send private data to attacker.
17	Spoof attack	In spoofing, the attacker alters the source address of the packets. When router routes the packet to destination address it does not consider the source address and send the packet to the destination. if destination system sends the reply then that reply got by attacker instead of original user.
18	Buffer overflow	When attacker sends more data to an application than its capacity, the buffer overflow takes place and results in attacker got access to the system in command prompt or shell.
19	Close-in Attack	The attacker is close to network component, data and computer physically and tries to learn more about the network. After getting information they try for denying access to information. This will happen due to open access.
20	Insider Attack	Someone from the inside organization or company employee is an attacker. The insider attacks can be harmful or no harmful. The harmful attacks for damaging information or deny access to authorized user. No harmful attacks due to lack of knowledge or carelessness.

Table 3: Terms regarding IDS [8]

Sr.no	Term	Description
1	Alarm	A signal is generated to specify whether system has been or being attacked.
2	Detection Rate	It is a rate to specify numbers of intrusions are detected by the system
3	False Alarm rate	If entity is normal and detected as intrusion.
4	True positive rate:	If entity is intrusion and detected as intrusion.
5	False negative rate	No alarm is generated even though attack takes place.
6	True negative	No attack takes place and no detection takes place
7.	Noise	It is an interference that can trigger a false positive or true positive.
8	Site policy	These are the guidelines provided by organization that controls the rules and configuration of an IDS.
9	Site policy awareness	Ability to dynamically modify rules and configuration of an IDS

III. CONCLUSIONS

The paper presents the different anomaly network intrusion detection systems techniques. The signature based detection gives higher detection accuracy and lower false positive rate but it detects only known attack but anomaly detection is able to detect unknown attack but with higher false positive rate. The various detection techniques introduced but till the main issue is regarding detection accuracy and false positive rate. The various types of attacks are also described and also terms regarding Intrusion detection system are also described.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] Yinhui Li, Jingbo Xia, Silan Zhang, Jiakai Yan, Xiaochuan Ai,, Kuobin Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method ", Expert Systems with Applications 39 (2012) 424–430.
[2] Shih-Wei Lin, Kuo-Ching Ying, Chou-Yuan Lee, Zne-Jung Lee," An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection" , Applied Soft Computing 12 (2012) 3285–3290.
[3] Shi-Jinn Horng , Ming-Yang Su , Yuan-Hsin Chen , Tzong-Wann Kao , Rong-Jian Chen , Jui-Lin Lai , Citra Dwi Perkasa ," A novel intrusion detection system based on hierarchical clustering and support vector machines ",Expert Systems with Applications 38 (2011) 306–313
[4] Zhang, T., Ramakrishnan, R., & Livny, M., (1996). BIRCH: An efficient data clustering method for very large databases. In Proceedings of the ACM SIGMOD (SIGMOD'96) (pp. 103–114).
[5] Tamer F.Ghanem,Wail S. Elkilani, Hatem,"A hybrid approach for efficient anomaly detection using metaheuristic methods", Journal of advanced research,2014.
[6] Garcı´a-Teodoro P, Dı´az-Verdejo J, Macia´ -Ferna´ ndez G, Va´ zquez E. Anomaly-based network intrusion detection: techniques, systems and challenges. Comput Secur 2009;28(1–2):18–28.
[7] Pourhabibi and R. Azmi Anomaly Based IDS Using Variable Size Detector Generation in AIS: A Hybrid Approach.
[8] http://en.wikipedia.org/wiki/Host-based_intrusion_detection_system.
[9] Assis MVOD, Rodrigues JJPC, Jr. MLP, "A hybrid approach for anomaly detection on large-scale networks using HWDS and entropy", in: 21st international conference on software, telecommunications and computer networks (SoftCOM 2013).
[10] Xu X, "Sequential anomaly detection based on temporal difference learning: principles", models and case studies. Appl Soft Comput 2010.
[11] Wang SS, Yan KQ, Wang SC, Liu CW, "An integrated intrusion detection system for cluster-based wireless sensor networks", Exper Syst Appl 2011.
[12] Kartit A, Saidi A, Bezzazi F, El Marraki M, Radi A, "A new approach to intrusion detection system", JATIT 2012.
[13] Garcia-Teodoro P, Diaz-Verdejo J, Macia -Fernandez G, Va zquez E, "Anomaly-based network intrusion detection: techniques, systems and challenges. Comput Secur, 2009.
[14] Abadeh MS, Mohamadi H, Habibi J., "Design and analysis of genetic fuzzy systems for intrusion detection in computer Networks", Expert Syst Appl 2011.

