Engineering Universe for Scientific Research and Management
(International Journal)

Vol. 7 Issue 8
August 2015

A review on Attribute Selection for Intrusion Detection System with Evolutionary Algorithm

Rakesh Singh Thakur[1], Gaurav Shrivastava[2]
M.Tech Scholar, Department of IT SVITS, Indore India[1]
AP, Department of IT SVITS, Indore India[2]
rkt2583@gmail.com[1], gaurav2086@gmail.com[2]

Abstract
The process of clustering technique plays an important role in intrusion detection system. The processes of clustering technique grouped the network traffic data on the basis of similarity and validate the traffic data. The process of clustering suffered from the problem of large number of iteration and loss of data. Now a day's various authors used various optimization technique for the controlling the number of iteration and selection of seed. In this paper present review of intrusion detection techniques for clustering data using KDDCUP dataset which include both normal and abnormal data.

Keywords: Clustering, Intrusion Detection, PSO, KDDCUP.

1. Introduction

The performance of intrusion detection system depends on classification of unknown types of attacks. The detection of unknown types of attack is very difficult due to large number of attribute and huge amount of network data. For the improvement of unknown attack feature reduction is important area of research. The reduction process reduces the large number of attribute and improved the detection of intrusion detection system. In the process of feature reduction various algorithm are used such algorithm are principle of component analysis and neural network. The reduction process used PCA method this method is static reduction technique, reduces only fixed number of attribute. The fixed number of feature reduction process not justify the value of feature it directly reduces the feature [8]. On the consideration of computational time feature reduction is also an important aspects, the reduces feature increase the processing of detection ratio. Many methods have been proposed in the last decades on the designs of IDSs based on feature reduction technique. With the tremendous growth of network-based services and sensitive information on networks, network security is becoming more and more importance than ever before [10]. Intrusion detection techniques are the last line of defenses against computer attacks behind secure network architecture design, firewalls, and personal screening. Despite the plethora of intrusion prevention techniques available, attacks against computer systems are still successful. Thus, intrusion detection systems (IDSs) play a vital role in network security [5]. Symantec in a recent report uncovered that the number of fishing attacks targeted at stealing confidential information such as credit card numbers, passwords, and other financial information are on the rise, going from 9 million attacks in June2013 to over 33 millions in less than a year. One solution to this is the use of network intrusion detection systems (NIDS) that detect attacks by observing various network activities. It is therefore crucial that such systems are accurate in identifying attacks, quick to train and generate as few false positives as possible [3]. Internet has rapidly become one of the main communication methods in our society. Various types of internet application and usage are available more and more. Increasing usages of network applications also increase security risks to internet users, to prevent unwanted or dangerous threats.

There are two primary approaches to analyze events to detect attacks, namely misuse detection and anomaly detection. Misuse detection is based on the extensive knowledge of known attacks and system vulnerabilities provided by a human expert, looking for hackers who attempt to perform these attacks and/or to exploit known vulnerabilities [12]. Although misuse detection can be very accurate in detecting known attacks, it cannot detect unknown and emerging cyber threats this shortcoming makes them vulnerable to the reactivity of attackers. In other words, when attackers change their behavior in response to detection techniques, these techniques become useless and need major redesign. One solution for this problem would be to use adaptive approaches which are inherently designed to be resilient to small changes in the environment and adapt easily. On the other hand, anomaly detection is based on the analysis of profiles that represent normal behavior of users, hosts, or network connections. Anomaly detectors characterize normal "legitimate" computer activity using different techniques and then use a variety of measures to detect deviations from defined normal behavior. The major benefit of anomaly detection algorithms is their potential to recognize unforeseen

attacks. However, the major limitation is the possibly high false alarm rate. Note that deviations detected by anomaly detection algorithms may not necessarily represent actual attacks as they may simply be new or unusual but still legitimate network behavior [6]. Anomaly detection techniques fall into the following five groups: statistical methods, rule-based methods, distance-based methods, profiling methods, and model-based approaches. It should be mentioned that many IDSs, such as snort, use both misuse detection and anomaly detection to benefit from their respective advantages. Section II gives the information related work. In section III discuss the problem states and formulation. In section IV discuss our proposed approach for solution and finally in section-V conclusion and future scope.

2. Related Work

In this section we discuss the about literature survey work. In survey, numbers of anomaly detection systems are study based on many different machine learning techniques. Some studies apply single agent learning technique, such as neural networks, genetic algorithms, support vector machines, etc. On the other hand, some systems are based on combining different learning techniques, such as hybrid or ensemble techniques. In particular, these techniques are developed as classifiers, which are used to classify or recognize whether the incoming Internet access is the normal access or an attack.

[1] Network Anomaly Intrusion Detection based on Genetic Clustering (NAIDGC) algorithm is proposed in this paper. The cluster centers are binary encoded. The sum of the Euclidean distances of the points from their respective cluster centers is adopted as the similarity metric. The optimal cluster centers are chosen by the genetic algorithm. Hence, self-identification of invasions is achieved. The experimental results demonstrate that this method can detect intrusion data efficiently in the network environment. In this paper, a new detection algorithm, the Network Anomaly Intrusion Detection based on Genetic Clustering (NAIDGC) algorithm is proposed, which is an unsupervised learning algorithm. The sum of the Euclidean distances of the points from their respective cluster centers is adopted as the similarity metric. The optimal cluster centers are searched by the genetic algorithm. With the help of preprocessing the data, the effect is better.

[2] In this paper, an improved algorithm based on the BM algorithm: BMD is proposed. BMD algorithm can reduce the space complexity and maintain the time complexity by reducing a pretreatment function and recording the number of times that a bad char found in the pattern. Experiments indicate that the space complexity is reduced by 36% at most. Therefore, the improved algorithm can provide significant improvement in pattern matching performance when using in an IDS. This paper proposed an effective algorithm: BMD, which solves the efficient problem of pattern matching in high-speed network environment by reducing the space complexity of BM algorithm.

[3] In this paper, they designed a new TCP scheme called TCP NJ Plus, which is capable of distinguishing non-congestion losses from packet reordering by gathering information from the current status of the network at the time of receiving three duplicate acknowledgments and react accordingly. The simulation results using Qualnet 4.5 confirm that, TCP NJ Plus achieves more than 20% throughput improvement over existing TCP schemes when the network co-existed with congestion, non-congestion loss and packet reordering.

[4] Pattern matching is an important detecting method of a misuse intrusion detection system. With the increase in the number of rules, the performance of pattern matching algorithm has been a gradually decline and has been a bottleneck. A new type of pattern matching based on suffix tree is proposed. The method mines rule's data structure and prunes rule set based suffix tree, the model size of the search space has been reduced. Experiments show, compared to the conventional pattern matching method, it is an effective method to reduce the time of pattern matching, and improves the detection efficiency of intrusion detection systems. A simple suffix tree-based pattern matching algorithm was presented. It adopted rule set's mining and pruning mechanism based on the suffix, shown an effective rules item pruning, and generated a clean association rule set with avoiding duplication of data sets.

[5] In this paper, they propose a novel supervised network intrusion detection method based on TCM-KNN (Transductives Confidence Machines for K-Nearest Neighbors) machine learning algorithm and active learning based training data selection method. It can effectively detect anomalies with high detection rate, low false positives under the circumstance of using much fewer selected data as well as selected features for training in comparison with the traditional supervised intrusion detection methods.

[6] In this paper, they effectively introduced feature selection methods to intrusion detection domain. They propose a wrapper-based feature selection algorithm

aiming at building lightweight intrusion detection system by using modified random mutation hill climbing (RMHC) as search strategy to specify a candidate subset for evaluation, as well as using modified linear Support Vector Machines (SVMs) iterative procedure as wrapper approach to obtain the optimum feature subset. they verify the effectiveness and the feasibility of our feature selection algorithm by several experiments on KDD Cup 1999 intrusion detection dataset. The experimental results strongly show that our approach is not only able to speed up the process of selecting important features but also to yield high detection rates.

[7] In this paper, they proposes a new modification of TCP Reno based on monitoring the wireless packet loss rate in real time. When the modified Reno cooperates with the router configured with explicit congestion notification (ECN), it is capable of distinguishing the wireless packet losses from the congestion packet losses, and reacting accordingly. At the same time, the sender takes advantage of the monitor result to adjustment the TCP segment size. The simulations in this paper show that the modification of TCP is feasible, and the performance of TCP is improved actually. They proposed a new TCP scheme to improve the TCP performance in wireless link.

[8] This paper proposes an efficient intrusion detection architecture which named NIDERC (Network Intrusion Detection based on Ensemble Rough Classifiers). The NIDERC contains a new algorithm of attribute reduction which combined Rough Set Theory with Quantum Genetic Algorithm, a method of establishing multiple rough classifications and a process of identifying intrusion data. The experimental results illustrate the effectiveness of proposed architecture.

3. Problem Formulation

In this section we discuss the problem of intrusion detection system. The environment in which the feature extraction is done is a mobile operator's network with real people using it. This means that the network traffic contains user confidential information. For example in Finland user network traffic is protected by the data protection law. Because of this, only a limited analysis for the network traffic can be done, meaning that a deep packet analysis cannot be done. In general, only the header fields of the packets can be checked but not the user data in the payload. Scalability is an issue with IDS. Because of the huge amount of data flowing through the mobile operator's network, it is not an easy task to find out the right information needed for IDS. The problem is to find an answer to the question: "What features need to be taken into account when calculating or analyzing whether the activity is malicious or not?"Based on prior research on IDS it is clear that either one of the techniques alone cannot detect everything but the combination of the both is the most promising approach. For example misuse detection can be used to filter known threats from the traffic to make it easier for the anomaly detection system to focus on the unknown. Even though IDS have been researched over 20 years, we still do not have an answer to the question of what features should be monitored. So far different kinds of methods and algorithms have been developed for anomaly detection but the focus has been on making them more efficient. Almost all of them are lacking the same information; what features are important for IDS, especially in telecommunications networks? For some reason information on the used features is not easily found from IDS research publications. No matter what the reason is the result is the same; every researcher has to figure out by themselves which features should be used for the monitoring.

1. The pre-processing of KDDCUP99 takes more time.
2. The rate of false alarm generation is high.
3. Some clustering technique is used such as k-means and genetic algorithm
4. Entropy based intrusion detection system suffered by high false rate
5. The detection of dynamic feature evaluation as normal data.

4. Our Approach

In this section we discuss the improve efficiency of intrusion detection system with the help of KDDCUP dataset. Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy [16] in 1995, inspired by social behavior of bird flocking or fish schooling. The PSO method is a member of the wide category of Swarm Intelligence methods. PSO shares many similarities with evolutionary computation techniques such as genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. PSO can be easily implemented and is computationally inexpensive sine its memory and CPU speed requirements are low. Also, it does not require gradient information of the objective function being considered, only its values. PSO is proving itself to be an efficient

method for several optimization problems, and in certain cases it does not suffer from the problems encountered by other Evolutionary Computation techniques. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other areas where GA can be applied. Even though, PSO typically moves quickly towards the best general area in the solution space for a problem, it often has difficulty in making the fine grain search required to find the absolute best point. Many control system applications, such as satellite altitude control, fighter aircraft control, model-based predictive control, control of fuel injectors, automobile spark timer, possess a mathematical model of the process with higher order, due to which the system defined becomes complex. The population consists of potential solutions, named particles, which are metaphor of birds in flocks. These particles are randomly initialized and freely fly across the multi dimensional search space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population. PSO utilizes several searching points and the searching points gradually get close to the global optimal point using its pbest and gbest. Initial positions of pbest and gbest are different. However, using thee different direction of pbest and gbest, all agents gradually get close to the global optimum solution.

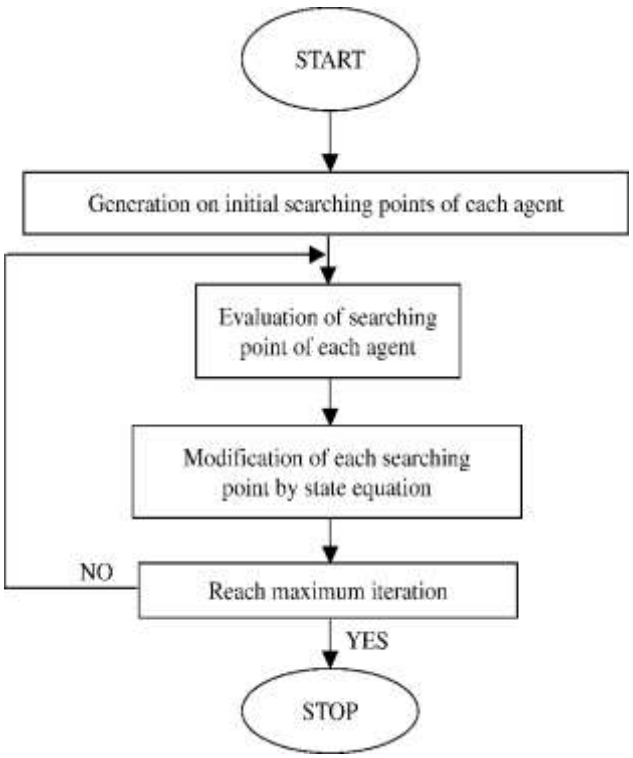

Figure 1: Shows that block diagram of working principle of POS algorithm.

5. Conclusion & Future Work

In this paper presents the review of intrusion detection system technique in terms of clustering and classification process. The similar patterns of data matching into a same group for clustering process, all the data are organized in the forms of group for better result in the form of improve efficiency and accuracy of grouped data. The data is combination of normal and abnormal, here we used KDDCUP dataset for the experimental process. In this paper basically discuss the PSO algorithm for the optimization of data and improve the efficiency for clustered dataset. In future implement this concept and used some standard optimization methods data for the evaluation of our proposed method.

References

[1] Huiling Guo, Weichen, Fang Zhang"Research of Intrusion Detection based on genetic clustering algorithm" IEEE, 2012, Pp 1204-1207.
[2] Feng Du "An effective pattern matching algorithm for intrusion detection" ICCSE 2012, Pp 34-38.
[3] Prasanthi S, Sang-Hwa Chung and Won-Suk Kim "An enhanced TCP scheme for distinguishing non-congestion losses from packet reordering over wireless mash networks" IEEE, 2011, Pp 440-447.
[4] Xie Yong He Fubao, Zhang Yilai" A descending suffix tree based pattern matching algorithm for intrusion detection" IEEE, 2012, Pp 21-28.
[5] Yang Li, Li Guo" An active learning based TCM-KNN algorithm for supervised network intrusion detection" IEEE, 2007, Pp 459-46.
[6] Yang Lia,d, Jun-Li Wangb, Zhi-Hong Tiand, Tian-Bo Luc, Chen Youngc "Building lightweight intrusion detection system using wapper based feature selection mechanisms" IEEE, 2009, Pp 466-475.
[7] Hu Han "performance improvement of TCP reno based on monitoring the wireless packet loss rate" IEEE, 2011, Pp 469-472.
[8] Shen li "An efficient architecture for network intrusion detection based on ensemble rough classifiers" ICCSE, 2013, Pp 1411-1415.
[9] Zhenwei Yu, Jeffrey J.P.Tsai "An automatically tuning intrusion detection system" IEEE,2007, Pp 373-384.
[10] Mohammad saniee Abadeh and Jafar Hakibi "Computer intrusion detection using an interactive fuzzy rule learning approach"IEEE,2007, Pp 2345-2351.
[11] Cichen Shingo mabu, Chuan Yue, Kaoru Shimeda and kotoro Hirasawa "Network Intrusion Detection using fuzzy class association rule mining based on

genetic network programming" IEEE,2009, Pp 60-67.

[12] Deepak Rathore and Anurag Jain "A novel method for intrusion detection based on ecc and radial bias feed forword network, in IJACR, Vol. 2, Issue 3: July-Sep.: 2012.

[13] Ambareen siraj, Susanm Bridges, Rayford B.Vaughu"Fuzzy cognitive maps for decision supporting an intelligent intrusion detection system" IEEE, 2012.

[14] Mansour Sheikhan , Zahra Jadidi "Misuse detection using hybrid of association rule mining and connectionist modeling" world applied science journal,2009, Pp 31-38.

[15] R. Shanmugavadivu, Dr. N Nagarajan "Network intrusion detection system using fuzzy logic" IEEE 2011, Pp 101-111.

[16] Sunita Patel, Jyoti Sondhi, Anand Motvani, Anurag Shrivastava "Improved Intrusion Detection Technique based on Feature Reduction and Classification using Support Vector Machine and Particle of Swarm Optimization" International Journal of Computer Applications, Vol-100, 2014. Pp 34-37.