Support Vector Machine Used in Network Intrusion Detection

Nilamadhab Mishra[1], Sarojananda Mishra[2]
[1](Research Scholar,Biju Patnaik University of Technology, Odisha, India)
[2](CS&A, Indira Gandhi Institute of Technology, Odisha, India)

Abstract: *Computer behavior in Intrusion Detection System can be classified into two important categories i.e. normal and abnormal activities. To achieve the categories, Intrusion detection adapt a new machine learning based data classification algorithm which is applied to network intrusion detection. The basic work is to classify network activities as normal or abnormal while minimizing classification. Different models for Network Intrusion detection already developed , each of them has its own strengths and weaknesses. By applying the Support Vector Machine we have to classify the data and do the clustering using Ant Colony Network. To avoid the weakness , we combine both SVM and ACN . In this approach , we need to use KDD cup 1999 dataset. IDS will be detected its rate, speed attack types and false alarm rate.*

I. Introduction

The concepts of network and information security related to a large-scale cyber-attack in the US have continued to dominate headlines and surpass any probability of a land-based terror attack in the US. The FBI Director testified in an interview before the Senate Homeland Security Committee in 2013 that cyber-attacks had surpassed land-based terror attack threats in the US and have become a major threat which has continued to rise [1]. In the area of cyber security, effective and efficient situational awareness often requires knowledge of current and historical cyber (i.e. host or network) activities to detect and respond to threatening behaviours [2]. The analysis of cyber threats could be improved by correlating security events from numerous heterogeneous sources [3]. Organizations must implement intrusion detection and prevention systems (IDPS) to protect their critical information against various kinds of attacks because anti-virus software and firewalls are not enough to provide full protection for their systems [4].

Intrusion detection systems (IDSs) can be categorized into three types: a network-based intrusion detection system (NIDS), a host-based intrusion detection system (HIDS), and a hybrid-based intrusion detection system (hybrid IDS). An HIDS detects malicious activities on a single computer while an NIDS identifies intrusions by monitoring multiple hosts and examining network traffic. In an NIDS, sensors are located at choke points of the network to perform monitoring, often in the demilitarized zone (DMZ) or on network borders and capture all the network traffic. Hybrid-based IDSs detect intrusions by analyzing application logs, system calls, file-system modifications (password files, binaries, access control lists, and capability databases, etc.) and other host states and activities [5]. IDSs are often used with other technologies (e.g., routers and firewalls). IDS technologies such as HIDS, NIDS, network behaviour anomaly detection (NBAD), and wireless local area network (WLAN) IDS are used together to correlate data from each device and make decisions according to what these IDSs monitor [5].

In today's information system management, large-scale data clustering and classification have become increasingly important and a challenging area. Although various tools and methods have been proposed, few are sufficient and efficient enough for real applications due to the exponential growing-in-size and high dimensional data inputs. As a particular application area, Intrusion Detection Systems (IDSs) are designed to defend computer systems from various cyber attacks and computer viruses. IDSs build effective classification models or patterns to distinguish normal behaviors from abnormal behaviors that are represented by network data. [6]

Intrusion detection systems classify computer behavior into two main categories: normal and abnormal activities. In order to achieve the categorization, Intrusion detection. we introduce a new machine-learning-based data classification algorithm that is applied to network intrusion detection.[6]

II. Related Work

Issues related to intrusion detection can be categorized into two broad areas: [7] network security and intrusion detection models, and [8] intrusion detection methods and algorithms based on artificial intelligence (mostly machine learning) techniques. In this section we shall briefly review some related work in the second area, and leave area [7] to the next section, when we discuss the background of IDSs. Intrusion detection as a classification problem has been studied for decades using machine learning techniques, including traditional classification methods (single classifier) such as K-Nearest Neighbor (K-NN), Support Vector Machines (SVMs), Decision Trees (DTs), Bayesian, Self-Organized Maps (SOMs), Artificial Neural Networks(ANNs), Generic Algorithms (GAs), and Fuzzy Logic, as well as hybrid classifiers that combine multiple machine learning techniques to improve the performance of the classifier. A review of using these approaches was given , which also included statistics of the use of these techniques reported in 55 research articles during the period 2000–2007. The review indicates that SVM and K-NN were the most commonly used techniques while the use of a hybrid increased significantly after 2004 and became mainstream. Another more recent review provided a thorough survey of intrusion detection using computational intelligence. It presented the details of the classification algorithms and swarm intelligence methods to solve problems using the decentralized agents. Most recently, IDS was introduced by integrating On Line Analytical Processing (OLAP) tools and data mining techniques. It is shown that the association of the two fields produces a good solution to deal with defects of IDSs such as low detection accuracy and high false alarm rate. As stated, as one of the swarm intelligence approaches, Ant Colony Optimization (ACO),has been applied in many fields to solve optimization problems, but its application to the intrusion detection domain is limited. Several methods were reported using ACO for intrusion detection. For example, an ant classifier was proposed that used more than one colony of ants to find solutions in multiclass classification problem. Another ant-based clustering algorithm applied to detect intrusions in a network presented is showed that the performance was comparable to some traditional classification methods like SVM, DT, and GA the authors evaluated the basic ant-based clustering algorithms and proposed several improvement strategies to overcome the limitations of these clustering algorithms that would not perform well on clustering large and high-dimensional network data. The work presented also used ACO for intrusion detection in a distributed network. The basic ingredient of their ACO algorithm was a heuristic for probabilistically constructing solutions. All these ACO-based intrusion detection approaches are single classifiers as categorized. Hybrid intrusion detection approaches involving SVM have been studied in the past, such as the one reported that uses the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering to help in finding the most qualified points to train the SVM classifier. It starts with an initial training set and expands the set gradually so that the training time for the SVM classifier is significantly reduced. Another hybrid intrusion detection approach was recently reported that combines hierarchical clustering and SVM.

III. Proposed Technique and Algorithm

Support Vector Machines have been widely accepted as a powerful data classification method. On the other hand, the Self-Organized Ant Colony Network has been shown to be efficient in data clustering. Our work aims to develop an algorithm that combines the logic of both methods to produce a high performance IDS. One challenge of developing IDSs is to realize real-time detection in high-speed networks. In this paper we are using 2 algorithms:

a) Support vector machines (SVM)

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. In two dimensional space this hyper plane is a line dividing a plane in two parts where in each class lay in either side.

To construct an optimal hyper plane, SVM employs an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVM models can be classified into four distinct groups:

- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)
- Regression SVM Type 1 (also known as epsilon-SVM regression)
- Regression SVM Type 2 (also known as nu-SVM regression)

b) Ant Colony Network (ACN)

ACO algorithms can be applied in the network routing problems to find the shortest path. In a network routing problem, a set of artificial ants (packets) are simulated from a source to the destination. The forward ants are selecting the next node randomly for the first time taking the information from the routing table and the ants who are successful in reaching the destination are updating the pheromone deposit at the edges visited by them by an amount (C/L), where L is the total path length of the ant and C a constant value that is adjusted according to the experimental conditions to the optimum value. The next set of the ants can now learn from the pheromone deposit feedback left by the previously visited successful ants and will be guided to follow the shortest path.

IV. Pseudo Code

A) SUPPORT VECTOR MACHINES (SVM):

Algorithm Steps:

Input: A training set with each data point labeled as positive or negative (class labels).

Output: A classifier.

```
Begin
    Randomly select data points from each class.
    Generate a SVM classifier.
    While more points to add to training set do
        Find support vectors among the selected points;
        Apply CSOACN clustering around the support vectors;
        Add the points in the clusters to the training set;
        Retrain the SVM classifier using the updated training set;
    End
End
```

B) Combination of Support vector machines (SVM) and Ant Colony Network (ACN):

Algorithm Steps:

Input: A training data set.

Input: N – number of training iterations.

Input: RR – detection rate threshold.

Output: SVM and ACN Classifiers.

```
Begin
    Normalize the data;
    Let r be the detection rate, initially 0;
    While r <RR do
        for k = 1, · · · , N do
            SVM training phase;
            Ant clustering phase;
        End
        Construct classifiers;
        do testing to update r;
    End
End
```

V. Conclusion and Future Work

In real-world applications managing and mining Big Data is Challenging task. The latest representation-learning technique and support vector machine to predict network intrusions through Big Data classification strategy. Additionally it suggested adopting machine learning framework for solving the problems associated with the continuity parameter. It also discussed the problems and challenges that the Big Data classification system for network intrusion prediction have to experience during the Big Data analytics. Research on Big Data techniques and technologies evolving and at the same time new problems and challenges are emerging, hence the hope is to develop better and better techniques and technologies towards finding solutions for Big Data classification problem.

References

[1]. Conteh, N. Y., & Schmick, P. J. (2016). Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. *International Journal of Advanced Computer Research*, 6(23), 31-38.

[2]. Streilein, W. W., Truelove, J., Meiners, C. R., & Eakman, G. (2011, November). Cyber situational awareness through operational streaming analysis. In Military Communications Conference, 2011-MILCOM 2011 (pp. 1152-1157). IEEE.

[3]. Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion detection and big heterogeneous data: a survey. Journal of Big Data, 2(3), 1-41.

[4]. Sandhu, U. A., Haider, S., Naseer, S., & Ateeb, O. U. (2011). A survey of intrusion detection & prevention techniques. In 2011 International Conference on Information Communication and Management, IPCSIT (Vol. 16). 66-71.

[5]. Tyler, G. (2009). Information Assurance Tools Report Intrusion Detection Systems. Information Assurance Technology Analysis Center (IATAC).

[6]. Sandeep Ranode, Prof.Ramesh G. Patole (2015), Survey on Network Intrusion Detection Using Ant Colony Networks and SVM Classification . International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 12 .

[7]. W. Lee, S.J. Stolfo, K.W. Mok, A data mining framework for building intrusiondetection models, in: Proceedings of IEEE Symposium on Security and Privacy, 1999, pp. 120–132.

[8]. W. Lee, S.J. Stolfo, K.W. Mok, Mining audit data to build intrusion detection models, in: Proceedings of the 4thInternational Conference on Knowledge Discovery and Data Mining, AAAI Press, 1998, pp. 66–72.