ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 15, No. 1, 2020, *pp*.022-030

CRF BASED INTRUSION DETECTION SYSTEM USING GENETIC SEARCH FEATURE SELECTION FOR NSSA

Azhagiri Mahendiran [1], Rajesh Appusamy [2] , Rajesh Prabhakaran[3] and Gowtham Sethupathi[4]
[1] Computer [1]Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[2] Computer Science and Engineering , C.Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamilnadu, India.
[3] Computer Science and Engineering , Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India.
[4] Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
(Received December 16, 2019, accepted January 25, 2020)

Abstract - Network security situational awareness systems helps in better managing the security concerns of a network, by monitoring for any anomalies in the network connections and recommending remedial actions upon detecting an attack. An Intrusion Detection System helps in identifying the security concerns of a network, by monitoring for any anomalies in the network connections. We have proposed a CRF based IDS system using genetic search feature selection algorithm for network security situational awareness to detect any anomalies in the network. The conditional random fields being discriminative models are capable of directly modeling the conditional probabilities rather than joint probabilities there by achieving better classification accuracy. The genetic search feature selection algorithm is capable of identifying the optimal subset among the features based on the best population of features associated with the target class. The proposed system, when trained and tested on the bench mark NSL-KDD dataset exhibited higher accuracy in identifying an attack and also classifying the attack category.

Index terms: Network Security Situational Awareness (NSSA), Intrusion Detection System (IDS), Network Security, Intelligent Systems, Conditional Random Fields(CRF), Feature selection, Machine learning.

1. Introduction.

The term situational awareness is used in military combat operations to denote "the ability to identify, process, and comprehend the critical elements of information about what is happening to the team with regards to the mission" [1]. Network security situational awareness (NSSA) is the ability to assess the current state of a network based on inputs provided by various sensors at different levels of the network [2]. This is quite a difficult task considering the volume of transactions done on any kind of network.
The NSSA operates at four different levels as in [4]:

- Acquiring information from intrusion detection systems (IDS), firewall logs, scan reports etc.
- Analyze the received information for evidences of any threat.
- Predict future threats based on the information learned from inputs such as IDS, firewall logs, scan reports etc.
- Recommend remedial actions to address a security event when it happens.

In order for the NSSA to function effectively, identification of anomalies in a network is of great importance. Intrusion detection is the process of identifying activities on a network that are violating the security policies of the network [3]. Intrusions affect the integrity, confidentiality of the information on the network and prevent accessibility of the information sources on the network [5, 6, 7]. An IDS with high accuracy will aid in better functioning of Network Security Situational Awareness (NSSA) System. Hence, in this paper we have proposed an IDS that is capable of detecting attacks accurately so that it can be effectively used in a NSSA system.

Our contributions in this research,

[1] Corresponding author. Tel.: +91-9865958062
E-mail address: azhagiri1687@gmail.com.

- An IDS using Conditional Random Field (CRF), capable of detecting various attack categories with high accuracy.
- Identification of a feature selection method for selecting the features that result in optimal operation of the CRF classifier.

The system proposed in [17] also uses CRF based classifier. The proposed system differs from the system in [17] as follows:

The system in [17] uses 4 layers of binary CRF classifier each capable of predicting one of the 4 attack categories whereas our system comprises of a single multi class CRF classifier capable of predicting all 4 attack categories. The system in [17] uses manual feature selection whereas our system uses an automatic feature selection method.

The rest of the paper is organized as follows: Section II describes several state of the art IDS in the literature. Section III describes the proposed system. Section IV discusses the results obtained by the proposed system and Section V concludes this research.

2. Related Work

In this section a brief discussion of some of the state of the art IDS researched in the literature are given.

In [8] the authors have used multiclass support vector machine to identify the various attacks on a network. The chi-square feature selection method was used to reduce the dimensionality of the dataset and choose appropriate attributes for building the model.

In [9] the authors have used a fuzzy based semi-supervised learning approach to efficiently utilize the unlabeled samples and used supervised learning algorithm to improve the performance of the IDS. A single hidden layer feed forward neural network is used for building the model. In the first stage, the unlabelled samples are categorized using a fuzzy quantification process. The categorized output from the first stage is then used to retrain the neural network.

In [10] an anomaly based network intrusion detection system using feature correlation analysis and association impact scale to predict intrusions has been proposed. The usage feature correlation significantly minimized the computational time of measuring association impact.

In [11] the authors have proposed a multi-level hybrid intrusion detection model using support vector machine and extreme learning machine. A modified K means algorithm have been used to significantly improve the quality of the training dataset. This has resulted in reduced training time of the classifiers and also resulted in improved performance of the IDS.

In [12] a modified optimum path forest algorithm [OPF] has been used. The training samples were divided into homogeneous subsets using k-means clustering algorithm. This has resulted in improved scalability, accuracy, detection rate, false alarm rate and execution time than traditional OPF.

In [13] the authors propose a fuzzy membership function which reduces considerably the computational complexity of the intrusion detection process and at the same time increases the accuracies of the classifier algorithms.

In [14] an anomaly based intrusion detection system using hierarchically structured learning automata has been proposed. The automaton learns to choose the optimal action through repeated interactions with the environment thereby resulting in a highly resilient approach that excels in detecting unknown attacks.

In [15] a hybrid feature selection method for intrusion detection has been proposed. The authors have used binary gravitational search algorithm with mutual information based filter for pruning the subset of features. The search direction is controlled using a two objective fitness function to maximize detection rate and minimizing false positive rate. This led to a increase in accuracy and detection rate compared to other wrapper based and filter based methods.

In [16] a hybrid approach integrating evolutionary algorithm with neural networks has been proposed. The authors have come up with two hybrids - gravitational search and gravitational search along with particle swarm optimization to train artificial neural networks. They have shown that these hybrid approaches have out run traditional IDS.

In [17] a layered approach for intrusion detection using conditional random fields has been proposed. The conditional random field achieves high detection accuracy and layered approach helps in improving the efficiency of the detection process. The authors have conducted statistical tests to prove the higher detection accuracy of their method.