Intrusion Detection Model Based on Ensemble Learning for U2R and R2L Attacks

Ploypan Sornsuwit
Department of Computer Science, Faculty of Science
King Mongkut's Institute of Technology
Bangkok, Thailand
ploypan@kpru.ac.th

Saichon Jaiyen
Department of Computer Science, Faculty of Science
King Mongkut's Institute of Technology
Bangkok, Thailand
kjsaicho@kmitl.ac.th

Abstract— **Intrusion Detection System (IDS) is a tool for anomaly detection in network that can help to protect network security. At present, intrusion detection systems have been developed to prevent attacks with accuracy. In this paper, we concentrate on ensemble learning for detecting network intrusion data, which are difficult to detect. In addition, correlation-based algorithm is used for reducing some redundant features. Adaboost algorithm is adopted to create the ensemble of weak learners in order to create the model that can protect the security and improve the performance of classifiers. The U2R and R2L attacks in KDD Cup'99 intrusion detection dataset are used to train and test the ensemble classifiers. The experimental results show that reducing features can improve efficiency in attack detection of classifiers in many weak leaners.**

Keywords— Intrusion detection, Ensemble, Adaboost, KDD Cup'99, Feature Reduction

I. INTRODUCTION

Currently, intrusions have various patterns of attack behaviors that are more difficult to detect than in the past. For example, some attack patterns require long periods for analyzing packets, or some attacks have a few amount of traffic. Therefore, the efficiency of traditional methods may be poor to detect intrusions accurately. IDS are divided into 2 types which are signature-based detection and anomaly-based detection. Signature-based detection will match the unknown pattern with known pattern and then consider whether it is normal or abnormal. However, anomaly detection is to identify the behaviors that are deviated from normal patterns[1]**.** Both two types have difference advantages and drawbacks. Signature-based detection can give high accuracy because it can match predefined attack behavior in database, but it cannot detect novel attack. On the other hands, anomaly-based detection has low accuracy and high false alarm because anomaly-based method uses statistical methods to analyze packets, and it can detect novel attacks.

Anomaly detections using machine learning method have been investigated in a number of researches. Ensemble methods are adopted to detect anomaly patterns [2] and show the better performance when they use multiple classifiers [3-5]. Some researches include preprocessing method in order to reduce redundant features. Consequently, they have only relevant features and produce the higher performance. However, some researches may be poor efficient to detect difficult attack types in datasets such as U2R and R2L types because both attack types have a few number of instants and more complicated behaviors.

In this paper, we present an algorithm that can overcome these problems, increase the accuracy, and decrease the false alarm rate of U2R and R2L attacks by using Correlation-based feature selection and multiple weak classifiers such as Naïve Bayes, Decision Tree, MLP, k-NN and SVM based on Adaboost algorithm. The rest of paper is organized as follows. In section 2, we present the related works. In Section 3, we present the proposed method. In section 4, the experimental results are described. In last section, conclusions and future work are presented.

II. RELATE WORKS

There are many research topics in intrusion detection system with several algorithms that improve accuracy and decrease false alarm rate. One of the topics to classify anomaly is to use ensemble methods to improve the performance of classifiers. Zhaza Merghani AbdElrahaman and Ajith Abraham [8] presented the comparisons of performance in many algorithms using Boosting and Bagging techniques with KDD Cup'99 datasets that contained 4 types of DoS Probe U2R and R2L. They compared the performance of these methods with Adaboost algorithm and Bagging technique. Then, they presented a new hybrid ensemble algorithm for detecting intrusion based on Error Collecting Output Code (ECOC). From their experimental results, they found that the proposed methods improved accuracy and false alarm rate.

Te-Shun Chou1, Jeffrey Fan, Sharon Fan, and Kia Makki2 [5] presented three layers of multiple classifiers for intrusion detection that was able to improve the overall accuracy. They applied three different patterns of learning including Naive Bayes, fuzzy k-NN, and back-propagation neural network for generating the decision boundary. The experiment used KDD cup'99 dataset with 30 features to test the model. The

978-1-4673-7863-5/15/$31.00 ©2015 IEEE