Chapter 3.10
Machine Learning Techniques for Network Intrusion Detection

Tich Phuoc Tran
University of Technology, Australia

Pohsiang Tsai
University of Technology, Australia

Tony Jan
University of Technology, Australia

Xiangjian He
University of Technology, Australia

ABSTRACT

Most of the currently available network security techniques are not able to cope with the dynamic and increasingly complex nature of cyber attacks on distributed computer systems. Therefore, an automated and adaptive defensive tool is imperative for computer networks. Alongside the existing prevention techniques such as encryption and firewalls, *Intrusion Detection System* (IDS) has established itself as an emerging technology that is able to detect unauthorized access and abuse of computer systems by both internal users and external offenders. Most of the novel approaches in this field have adopted *Artificial Intelligence* (AI) technologies such as *Artificial Neural Networks* (ANN) to improve performance as well as robustness of IDS. The true power and advantages of ANN lie in its ability to represent both linear and non-linear relationships and learn these relationships directly from the data being modeled. However, ANN is computationally expensive due to its demanding processing power and this leads to *overfitting* problem, i.e. the network is unable to extrapolate accurately once the input is outside of the training data range. These limitations challenge IDS with low detection rate, high false alarm rate and excessive computation cost. This chapter proposes a novel *Machine Learning* (ML) algorithm to alleviate those difficulties of existing AI techniques in the area of computer network security. The Intrusion Detection dataset provided by Knowledge Discovery and Data Mining (KDD-99) is used as a benchmark to

DOI: 10.4018/978-1-60960-818-7.ch3.10

compare our model with other existing techniques. Extensive empirical analysis suggests that the proposed method outperforms other state-of-the-art learning algorithms in terms of learning bias, generalization variance and computational cost. It is also reported to significantly improve the overall detection capability for difficult-to-detect novel attacks which are unseen or irregularly occur in the training phase.

INTRODUCTION

Current security systems offer a reasonable level of protection; however, they cannot cope with the growing complexity of computer networks and hacking techniques. They have to face continuous environmental changes both with respect to what constitutes normal behavior and abnormal behavior. As the result, security systems suffer from *low detection rates* (missing out serious intrusion attacks) and *high false alarm rates* (falsely classifying a normal connection as an attack and therefore obstructing legitimate user access to the network resources). In order to overcome such challenging problems, there has been a great number of research conducted to apply *Machine Learning* (ML) algorithms to achieve a generalization capability from limited training data. In recent years, ML algorithms such as Artificial Neural Network (ANN), which is generally well regarded as the universal function approximator, have demonstrated successes in many network security applications. As a flexible "model-free" approach, ANN can fit the training data very well and thus provide a low learning bias. However, they are also susceptible to the overtting problem, which can cause instability in generalization. Some models of ANN also suffer from highly demanding computation power due to their large model complexity. For an ANN model to be useful, it should perform well on the training data and generalize reliably on the unseen data. Unfortunately, learning bias, generalization variance and model complexity are somewhat incompatible, i.e. reducing one element will inevitably increase the others. Therefore, a good tradeoff of these elements should be sought.

In this chapter, an innovative ML algorithm is proposed to alleviate the limitations of currently existing IDS, enhancing the performance of intrusion detection for rare and complicated attacks. By implementing *Adaptive Boosting* and *Semi-parametric* radial-basis-function neural networks (RBFNN), the proposed model can minimize learning bias (how well the model fits the available sample data) and generalization variance (how stable the model is for unseen instances) at an affordable cost of computation.

This chapter starts with the related works of ML approaches for Network Security domain, followed by an extensive review of ANN models. Particularly, emphasis is put on the Generalized Regression Neural Network (GRNN) and vector-quantized GRNN. These models belong to the RBFNN family which has been reported for great successes in many applications. We also provide an overview of Ensemble Learning methods in which multiple classifiers are trained to solve the same problem and their decisions are then aggregated in some manner. It is theoretically and experimentally proved that such an ensemble model can achieve superior performance compared with individual classifiers. Next, the research proposal and its features are presented. The usefulness of this model will be illustrated through its application to the Network intrusion detection problem.

RELATED WORKS

An intrusion detection system (IDS) is defined as a protective system that monitors computers or networks for unauthorized activities based on network traffic or system usage behaviors, thereby detecting if a system is targeted by a network attack such as a denial of service attack. There are two types of IDS: (1) *misuse-based detection* in which