An Improved Ant-IS Algorithm for Intrusion Detection

Amal Miloud-Aouidate, University of Science and Technology Houari Boumediene, Algiers, Algeria
Ahmed Riadh Baba-Ali, University of Science and Technology Houari Boumediene, Algiers, Algeria

ABSTRACT

During recent years, the number of attacks on networks has dramatically increased. Consequently, the interest in network intrusion detection has increased among the researchers. This paper proposes a clustering Ant-IS and an active Ant colony optimization algorithms for intrusion detection in computer networks. The goal of these algorithms is to extract a set of learning instances from the initial training dataset. The proposed algorithms are an improvement of the previously presented Ant-IS algorithm, used in pattern recognition. Results of experimental tests show that the proposed algorithms are capable of producing a reliable intrusion detection system.

Keywords: Active Learning, Ant Colony Optimization, Clustering, Detection Rates, False Alarm Rates, Instance Selection, Intrusion Detection Systems (IDS), Intrusion Detection, KDDCup99

INTRODUCTION

The need of protecting the computing resources from attacks is growing day by day. Hence, many technologies have been developed; among them the Intrusion Detection Systems (IDS) are the most used ones.

An IDS is an active device or process, that analysis network activity to detect any unauthorized entry and malicious activities. Detection approaches can be classified into two main categories: signature-based and anomaly-based detection. Signature-based IDS operate in much the same way as a virus scanner. They search for a known signature for each specific intrusion event. Behavioral IDS appeared much later than the signature-based IDS and are not yet mature. Thus, the use of such IDS can be difficult in the sense that the generated alarms may contain a significant amount of false alarms. This problem can be solved by using a powerful classifier which allows avoiding false positive alarms.

The K-nearest neighbor classification rule (KNN) is a powerful classification method that allows the classification of an unknown instance using a set of classified training instances (Cover and Hart, 1967). The purpose of this algorithm is to classify a new instance according to the classes of the instances of the training set. But,
the computing of a consistent training subset with a minimal cardinality for KNN rule turns out to be hard.

Ant-IS algorithm is an instance selection algorithm based on Ant Colonies Optimization principles. This algorithm improves the standard 1NN performances, by selecting the significant instances to create an effective training set. Ant-IS has been proven effective in the field of classification by learning. Active learning is a form of supervised machine learning in which a learning algorithm is able to interactively query the information source to obtain the desired outputs at new data points (Dasgupta and Langford, 2009).

In this paper, we present a framework that aims to involve the active learning in the management of the execution of the Ant-IS for intrusion detection, and therefore, to reach a better false positives rate. In the next section, we briefly review some related works. The two sections following introduce respectively Ant-IS and Active Learning. The new algorithm is presented in the section after that. The done first tests are shown in the following section. Finally some conclusions and future work are discussed in the last section.

RELATED WORK

Recently, data mining is becoming an important component in intrusion detection systems (Li, 2009). IDES (Denning, 1987) is the first model of IDS. This model uses statistical techniques to characterize abnormal behavior and is based on rules to detect violations. NIDES (Javits and Valdes, 1993) are the successor of the IDES project. It has a strong base of anomaly detection, complemented with a component-based expert system signature. MADAM ID (Lee et al., 1998) is one of the best known data mining projects in intrusion detection. It is an off-line IDS using Association rules and frequent episodes. MADAM ID permits to replace hand-coded intrusion patterns and profiles with the learned rules. ADAM (Audit Data Analysis and Mining) (Barbara et al., 2001) is an on-line network based IDS. It allows detecting known attacks as well as unknown attacks. ADAM uses association rules and classification. IDDM (Abraham, 2001) is a real-time NIDS for misuse and anomaly detection. It applies association rules, Meta rules, and characteristic rules.

Common representations for data mining techniques have been proposed for intrusion detection and made great success. These techniques include neural networks (Lippmann, 2000), Support Vector Machine (SVM) (Eskin, 2002; Mukkamala, 2002) and neighbor-hood based classification (Law and Kwok, 2005; Li and Guo, 2007; Liao and Vemuri, 2002; Li, 2009; Xiang et al., 2008; Kuang and Zulkernine, 2008; Shirazi, 2009; Shirazi and Kalaji, 2010; Shirazi et al., 2012; Tsai and Lin 2010; Muda et al., 2011; Wang, 2011; Deepika and Richhariya, 1999; Natesan and Balasubramanie, 2012).

Liao and Vemuri (2002) have introduced a new approach based on the KNN classifier to classify program behavior as normal or intrusive. Their approach employs the KNN classifier to categorize each new program behavior either into normal or intrusive class. The frequencies of system calls used by a program, instead of their local ordering, are used to characterize the behavior of the program.

The method proposed by Law and Kwok (2005) finds whether a sequence of incoming alarms is different from a normal situation. To detect abnormal patterns of newly incoming alerts, new data points are created for these new alarms. The distance of these points from the normal points indicates their deviation from the normal situation.

Li and Guo (2007) have proposed a novel supervised network intrusion detection method based on TCM-KNN (Transductive Confidence Machines for K-Nearest Neighbors) machine learning algorithm and active learning based training data selection method. TCM-KNN algorithm is a commonly used machine learning and data mining method; thus, effective in fraud detection, pattern recognition and outlier.


Hybridizing Harmony Search Algorithm with Multi-Parent Crossover to Solve Real World Optimization Problems
