Information Sciences 378 (2017) 484–497

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Fuzziness based semi-supervised learning approach for intrusion detection system

Rana Aamir Raza Ashfaq[a], Xi-Zhao Wang[a,*], Joshua Zhexue Huang[a], Haider Abbas[b], Yu-Lin He[a]

[a] College of Computer Science & Software Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China
[b] King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 21 August 2015
Revised 17 March 2016
Accepted 6 April 2016
Available online 3 May 2016

Keywords:
Fuzziness
Divide-and-conquer strategy
Semi-supervised learning
Intrusion detection
Random weight neural network

ABSTRACT

Countering cyber threats, especially attack detection, is a challenging area of research in the field of information assurance. Intruders use polymorphic mechanisms to masquerade the attack payload and evade the detection techniques. Many supervised and unsupervised learning approaches from the field of machine learning and pattern recognition have been used to increase the efficacy of intrusion detection systems (IDSs). Supervised learning approaches use only labeled samples to train a classifier, but obtaining sufficient labeled samples is cumbersome, and requires the efforts of domain experts. However, unlabeled samples can easily be obtained in many real world problems. Compared to supervised learning approaches, semi-supervised learning (SSL) addresses this issue by considering large amount of unlabeled samples together with the labeled samples to build a better classifier. This paper proposes a novel fuzziness based semi-supervised learning approach by utilizing unlabeled samples assisted with supervised learning algorithm to improve the classifier's performance for the IDSs. A single hidden layer feed-forward neural network (SLFN) is trained to output a fuzzy membership vector, and the sample categorization (low, mid, and high fuzziness categories) on unlabeled samples is performed using the fuzzy quantity. The classifier is retrained after incorporating each category separately into the original training set. The experimental results using this technique of intrusion detection on the NSL-KDD dataset show that unlabeled samples belonging to low and high fuzziness groups make major contributions to improve the classifier's performance compared to existing classifiers e.g., naive bayes, support vector machine, random forests, etc.

1. Introduction

Intrusion detection (ID) is a process of monitoring, detecting, and analyzing the events that are considered as violation to the security policies of a networked environment [45]. Denning [12] introduced the concept of detecting cyber-based attacks on computer networks by providing a framework for intrusion detection system (IDS), which is based on the hypothesis that security violations can be detected by monitoring system audit records for abnormal patterns of system usage. Organizations deploy their own access controls to grant or restrict the level of access for their assets but this approach

* Corresponding author.
E-mail addresses: aamir@szu.edu.cn (R.A.R. Ashfaq), xizhaowang@ieee.org, xwang@szu.edu.cn (X.-Z. Wang), zx.huang@szu.edu.cn (J.-Z. Huang), hsiddiqui@ksu.edu.sa (H. Abbas), yulinhe@szu.edu.cn (Y.-L. He).

http://dx.doi.org/10.1016/j.ins.2016.04.019
0020-0255/

does not guarantee the appropriate assurance and protection level for a particular resource [10]. This problem is evident through various security incidents around the world, for example, the compromise of Yahoo's and Amazon's websites after some sophisticated persistent attacks [6]. Intruders and attackers are always seeking to disrupt network traffic and degrade network performance with different types of attacks or intrusions. A network intrusion refers to a suspicious and sudden deviation from the normal behavior of the system, which destabilizes the security of the network system. According to Qui et al. [40], Hernndez-Pereira et al. [16] and Yan and Yu [56], intrusion can be depicted as the set of actions that attempt to compromise the confidentiality, integrity, or availability (CIA) of information resources; therefore, it is necessary to take different measures to minimize such risks.

The Internet has turned into an indispensable wellspring for exchanging information among users and organizations; therefore, security has become an essential aspect in this type of communication. IDSs are often used to sniff network packets by providing a better understanding of what is happening in a particular network. Two mainstream preferences for IDSs are (1) host-based IDSs, and (2) network-based IDSs. Correspondingly, the detection methods used in IDS are anomaly based and misuse based (also called signature or knowledge based), each having their own advantages and restrictions. In misuse-based detection, data gathered from the system is compared to a set of rules or patterns, also known as signatures, to describe network attacks. The core difference between these two techniques is that anomaly-based IDS uses collections of data containing examples of normal behavior and builds a model of familiarity, therefore, any action that deviates from the model is considered suspicious and is classified as an intrusion [20]. According to Mukkamala et al. [31], in misuse-based detection, attacks are represented by signatures or patterns. However, this approach does not contribute much in terms of zero-day attack detection. The main issue is how to build permanent signatures that have all the possible variations and non-intrusive activities to lower the false-negative and false-positive alarms.

The KDDCUP'99 [18] was derived in 1999 from the DARPA98 network traffic dataset and a very popular benchmark dataset used in the International Knowledge Discovery in Databases (KDD) competition. From the literature, one can study that this dataset is widely used for the evaluation of anomaly based IDS. Many machine learning techniques, which may be either supervised or unsupervised, have been used to increase the efficacy of IDSs. Supervised learning techniques are applied to obtain the training data in which instances are tagged with labels and each label indicates the class of a particular instance. Many supervised algorithms, such as *k*-nearest neighbor (KNN) [24], neural network (NN) [29], and support vector machine (SVM) [30] have been extensively used to detect the intrusions. These algorithms build the model that separates a new unseen example or instance with the correct label. Many advantages and disadvantages related to supervised learning with IDS have been reported by many researchers. One of the shortcomings of supervised learning is the need for labeled instances. The only dataset is available for ID is the KDDCUP'99 dataset [18], and many new types of attacks have been developed. Therefore, this dataset is considered as obsolete, and for new types of examples its accuracy drops [22]. Many researchers are widely using the KDDCUP'99 dataset because it is the only dataset that is publically available for ID problem and useful information can still be extracted from it. Apart from its disadvantage, supervised learning has the advantage to achieve better accuracy to classify similar examples [22]. Unsupervised learning techniques deal with the learning tasks with unlabeled or untagged data. Clustering is the most popular unsupervised learning technique [25]. In clustering, the learning algorithm finds similarities among instances to build the clusters (i.e. group of instances). Instances that belong to the same cluster are assumed to having similar characteristics or properties and then are assembled into the same class. The disadvantage of unsupervised learning is the manually assignment of cluster numbers, which results in low accuracy in predictions. However, it has the advantage of detecting new examples better than supervised learning techniques, and considered to be more robust in IDSs. According to Laskov et al. [22], many new attacks have been developed, and the improper labeling of examples could make the unsupervised learning and SSL techniques the best choices for improving the accuracy of IDSs.

Regarding the aforementioned development in this area, the main objective behind our work is not just to seek for the smallest classification error but also to try to find a model that must be capable of incorporating new data that keeps its good generalization ability. We compute the fuzziness of every unlabeled sample outputted by the classier, and try to discover its relationship with misclassification. From literature, except for [51,53], we have not found any studies on generalization based on the fuzziness of a classifier. Therefore, based on our preliminary work [51] in which the sample categorization is performed according to the fuzziness quantity, we propose a new algorithm for the IDS. The experimental results demonstrate that samples belonging to the low and high fuzziness categories play an important role in improving the accuracy of IDSs.

The rest of the paper is organized as follows. Section 2 presents a prologue to the background of semi-supervised learning (SSL). Section 3 details the proposed fuzziness based algorithm using the neural network with random weights (NNR_w). The performance evaluation is presented in Section 4. Finally, Section 5 ends this paper with concluding remarks, and provides future directions for this research.

2. Semi-supervised learning (SSL)

SSL is an amalgamation of supervised and unsupervised learning techniques. The SSL technique deals with the learning tasks by utilizing both labeled and unlabeled data [65]. Labeled instances are, however, expensive and time-consuming to obtain and require the efforts of domain experts. Apart from this concern, unlabeled data can easily be obtained in many real world applications. SSL methods assign labels by considering unlabeled instances, together with the labeled instances,