Expert Systems With Applications 67 (2017) 296–303

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system

Wathiq Laftah Al-Yaseen[a,b,*], Zulaiha Ali Othman[a], Mohd Zakree Ahmad Nazri[a]

[a] Data Mining and Optimization Research Group (DMO), Centre for Artificial, Intelligence Technology (CAIT), School of Computer Science, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), 43600 Bandar Baru Bangi, Malaysia
[b] Al-Furat Al-Awsat Technical University, Iraq

ARTICLE INFO

Article history:
Received 30 August 2015
Revised 29 September 2016
Accepted 30 September 2016
Available online 30 September 2016

Keywords:
Intrusion detection system
Support vector machine
Extreme learning machine
K-means
Multi-level
KDD Cup 1999

ABSTRACT

Intrusion detection has become essential to network security because of the increasing connectivity between computers. Several intrusion detection systems have been developed to protect networks using different statistical methods and machine learning techniques. This study aims to design a model that deals with real intrusion detection problems in data analysis and classify network data into normal and abnormal behaviors. This study proposes a multi-level hybrid intrusion detection model that uses support vector machine and extreme learning machine to improve the efficiency of detecting known and unknown attacks. A modified K-means algorithm is also proposed to build a high-quality training dataset that contributes significantly to improving the performance of classifiers. The modified K-means is used to build new small training datasets representing the entire original training dataset, significantly reduce the training time of classifiers, and improve the performance of intrusion detection system. The popular KDD Cup 1999 dataset is used to evaluate the proposed model. Compared with other methods based on the same dataset, the proposed model shows high efficiency in attack detection, and its accuracy (95.75%) is the best performance thus far.

1. Introduction

Dependence on convenient security systems to protect computers and networks against intrusions is a crucial issue in computer science because of the significant development of network-based computer services. Intrusion detection system (IDS) is one of the systems used to monitor and analyze the events in a computer or network to identify any deviation from normal behavior. IDSs can be categorized in several ways, but the most common are misuse-based and anomaly-based categories (Lee, Stolfo, & Mok, 1999). Misuse-based IDS can efficiently detect known attacks, such as Snort (Roesch, 1999). This type of IDS has a low false alarm rate, but it fails to identify new attacks that do not embody any rules in the database. Anomaly-based IDS builds a model of normal behavior and then distinguishes any significant deviations from this model as intrusions. This type of IDS can detect new or unknown attacks but features a high false alarm rate.

To reduce the false alarm rate of anomaly-based IDS, many machine learning techniques, including support vector machine (SVM) (Feng, Zhang, Hu, & Huang, 2014; Horng et al., 2011; Kuang, Xu, & Zhang, 2014) and extreme learning machine (ELM) (Cheng, Tay, & Huang, 2012; Singh, Kumar, & Singla, 2015), have been applied, along with models combining several techniques (Hasan, Nasser, Pal, & Ahmad, 2013; Panda, Abraham, & Patra, 2012). Each model offers specific strengths and weaknesses, with overall generic detection rates steadily increasing.

SVMs exhibit good detection performance with IDSs in terms of classifying the flow of a network into normal or abnormal behaviors. Horng et al. (2011) proposed an IDS based on a combination of BIRCH hierarchical clustering and SVM technique. Their proposed method achieved a good accuracy of up to 95.72% with a false alarm rate of 0.7%. Feng et al. (2014) introduced an approach combining SVM with self-organized ant colony network. This approach exhibited a good detection rate of up to 94.86% and a false positive rate of 6.01%. Kuang et al. (2014) proposed an IDS based on a combination of the SVM model with kernel principal component analysis (KPCA) and genetic algorithm (GA). KPCA was used to reduce the dimensions of feature vectors, whereas GA was employed to optimize the SVM parameters. The average detection rate was 95.26%, whereas the average false alarm rate was 1.03%.

ELMs exhibit performance comparable with that of SVMs in terms of classifying instances of IDS. Cheng et al. (2012) applied a kernel-based ELM for multi-class classification of IDS. Their results showed the ELM exhibited high data classification accuracy,

* Corresponding author.
E-mail addresses: wathiqpro@gmail.com (W.L. Al-Yaseen), zao@ukm.edu.my, zakree@ukm.edu.my (M.Z.A. Nazri).

http://dx.doi.org/10.1016/j.eswa.2016.09.041
0957-4174/