Computers and Electrical Engineering 000 (2017) 1–20

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Fuzzified Cuckoo based Clustering Technique for Network Anomaly Detection☆

Sahil Garg*, Shalini Batra

Computer Science & Engineering Department, Thapar University, Patiala (Punjab), India

ARTICLE INFO

Article history:
Received 24 April 2017
Revised 8 July 2017
Accepted 11 July 2017
Available online xxx

Keywords:
Anomaly detection
Feature selection
Decision Tree
Nature inspired algorithm
Cuckoo-search
K-means clustering
Fuzzy theory

ABSTRACT

With the increasing penetration of security threats, the severity of their impact on the underlying network has increased manifold. Hence, a robust anomaly detection technique, Fuzzified Cuckoo based Clustering Technique (F-CBCT), is proposed in this paper which operates in two phases: training and detection. The training phase is supported using Decision Tree followed by an algorithm based on hybridization of Cuckoo Search Optimization and K-means clustering. In the designed algorithm, a multi-objective function based on Mean Square Error and Silhouette Index is employed to evaluate the two simultaneous distance functions namely-Classification measure and Anomaly detection measure. Once the system is trained, detection phase is initiated in which a fuzzy decisive approach is used to detect anomalies on the basis of input data and distance functions computed in the previous phase. Experimental results in terms of detection rate (96.86%), false positive rate (1.297%), accuracy (97.77%) and F-Measure (98.30%) prove the effectiveness of the proposed model.

1. Introduction

Anomaly detection is an important component of data analysis that deals with the discovery of rare or unusual patterns which deviate significantly from the normal behaviour of data and possess unusual distributions. The rapid growth of information technology has resulted in the enhancement of security threats in computer networks. Since any malicious activity on network may lead to serious consequences, the importance of information carried out in these networks makes the task of anomaly detection very crucial [1].

On the basis of taxonomy, detection techniques are classified into two categories: Signature Based Detection and Anomaly Detection [2]. Signature-based detection techniques operate in the same way as virus scanners detect the malicious events. They examine the network traffic by comparing it with the signatures of known attacks to identify anomalous events which exploit vulnerabilities. These techniques are capable of detecting only those vulnerable instances for which signatures have been defined, but fail to detect unfamiliar attacks.

On the contrary, the network anomaly detection classifies network traffic as normal or anomalous without any knowledge about their signatures. These techniques lay on the baseline model for identifying network vulnerabilities, where baseline describes the minimal criteria required for acceptable network behaviour [3]. The event that falls outside the range of this

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. A. K. Sangaiah.

* Corresponding author.
E-mail addresses: garg.sahil1990@gmail.com (S. Garg), sbatra@thapar.edu (S. Batra).

http://dx.doi.org/10.1016/j.compeleceng.2017.07.008
0045-7906/

acceptable baseline model is considered as anomalous. Thus, anomaly detection techniques clearly have an advantage over signature-based techniques, where attacks with known signatures can be identified by considering the baseline model [4].

Anomaly detection techniques are used in many application domains and each domain requires a different approach to serve it [5]. In the context of data mining, anomaly detection algorithms are broadly classified into three categories: Supervised, Semi-Supervised, and Unsupervised. Supervised anomaly detection techniques learn a classifier using labeled data instances in order to classify them as either normal or anomalous. Since these techniques can detect only those vulnerabilities for which the labels are defined, they are relevant only in domains where labeled data is available to train the classifier. Semi-supervised anomaly detection techniques, on the other hand, make use of labeled data as well as unlabeled data for making the detections. Unsupervised anomaly detection techniques detect the anomalies in unlabeled dataset without having prior knowledge about the dataset labels. Thus, these techniques are believed to be of great relevance with respect to anomaly detection [6].

The remainder of the paper is organized as follows: In Section 2, related work is presented. Section 3 gives a brief overview about the contributions of the proposed work. In Section 4, F-CBCT is explained in detail. The results and discussions are presented in Section 5, followed by conclusion and future scope in Section 6.

2. Related work

The relevance of anomaly detection and its analysis has attracted researchers to contribute in this field. The success rate of these techniques depend on their high detection accuracies along with low false positive rates. To have a broader view of this domain, some of the anomaly detection techniques that have been proposed in the literature are discussed in this section.

Ghanem et al. [7] introduced a novel anomaly detection technique using meta-heuristic method and genetic algorithm to escape from the problems of local optima and robust search. The proposed technique used a selection-based detector generation methodology to detect anomalies in large scale datasets. The work done by the authors to generate the detectors with such a high accuracy is quite convincing but in order to increase the adaptability and flexibility of the proposed model, the values of the parameters used by the multi-start searching method and genetic algorithm should be decided dynamically instead of setting them apriori. Aburomman and Reaz [8] combined Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) along with Particle Swarm Optimization (PSO) to propose a hybrid intrusion detection technique. In the proposed approach, KNN and SVM experts are trained using binary classifiers and these classifiers are combined using PSO and Weighted Majority Algorithm (WMA). Ensembles are created to combine the expert opinions to reach the final classification decision. Similarly, Shahreza et al. [9] combined Self-Organized Maps (SOM) and PSO for the purpose of anomaly detection. In order to validate the proposed unsupervised approach, a case study was performed on forest fire detection.

Data clustering techniques help to evaluate the dataset instances which are homogeneous in nature. The K-means technique is the simplest and most intuitive partitioning technique but it suffers from local convergence and initialization of clusters like problems. Inspired from K-means technique, a number of partitioning algorithms have been developed by researchers like K-medoid, fuzzy C-means (FCM), etc. K-medoids algorithm divides a dataset into number of partitions by minimizing the absolute distance between the points and the centroid rather than minimizing the square distance. K-medoid is found to be more robust than K-means algorithm, but it is more expensive as it involves pairwise distance computation. Both K-means and K-medoids tend to give inaccurate results during cluster overlapping scenarios. In such cases, better accuracy can be achieved using FCM, in which each element is assigned to the cluster having highest membership grade. Thus, the nature of the patterns associated with the dataset should be evaluated carefully before selecting the appropriate clustering technique [10].

Dash and Liu [11] demonstrated the importance of feature selection in clustering. It has been shown that clustering algorithms are highly sensitive to the dimensionality of data. A relevant subset of features is thereby required for improving the performance of clustering algorithms. Thus, the authors employed RANK algorithm to order the features according to their importance in the cluster. Experimental evaluation of their work validated the effectiveness of the designed feature selection algorithm. Similarly, Fong et al. [12] proposed Accelerated PSO based feature selection algorithm for mining streams in big data. The related performance evaluation was done by collecting big data with high dimensionality. Duan et al. [13] also propounded clustering based technique which was facilitated by hierarchical feature selection scheme. The author employed partial distance strategy to reduce the dimensionality of the dataset. The experimental analysis of the proposed technique along with various other state-of-the-art algorithms proved the effectiveness of the same.

Karami and Guerrero-Zapata [14] developed a hybrid technique for proactive prediction of DoS attacks in named data networking environment. In this approach, a combination of multi-objective optimization and PSO was used to resolve the hybrid learning problem of Radical Basis Function (RBF) Network. This RBF-based network classifier was then utilized to improve the prediction accuracy of DoS attacks. Another work by Karami and Guerrero-Zapata [15] combined PSO and K-means algorithms for anomaly detection in content-centric networks. The proposed detection system operates in two phases: training and detection. The training phase employs two simultaneous cost functions, i.e., Davies–Bouldin Index (DBI) and Mean Square Error (MSE), whereas detection phase utilizes two distance-based approaches, i.e., classification and outlier. In training phase, the hybridization of PSO and K-means determines the optimal number of clusters while the fuzzy approach employed in detection phase detects the anomalies. Further, the performance was evaluated in terms of sensitivity, specificity and accuracy. Motivated from this work, we present a Fuzzified Cuckoo based Clustering Technique (F-CBCT) for network

Please cite this article as: S. Garg, S. Batra, Fuzzified Cuckoo based Clustering Technique for Network Anomaly Detection, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.07.008