Chapter 17
A Novel Fuzzy Anomaly Detection Algorithm Based on Hybrid PSO-Kmeans in Content-Centric Networking

Amin Karami
Universitat Politecnica de Catalunya, Spain

ABSTRACT

In Content-Centric Networks (CCNs) as a promising network architecture, new kinds of anomalies will arise. Usually, clustering algorithms would fit the requirements for building a good anomaly detection system. K-means is a popular anomaly detection method; however, it suffers from the local convergence and sensitivity to selection of the cluster centroids. This chapter presents a novel fuzzy anomaly detection method that works in two phases. In the first phase, authors propose an hybridization of Particle Swarm Optimization (PSO) and K-means algorithm with two simultaneous cost functions as well-separated clusters and local optimization to determine the optimal number of clusters. When the optimal placement of clusters centroids and objects are defined, it starts the second phase. In this phase, the authors employ a fuzzy approach by the combination of two distance-based methods as classification and outlier to detect anomalies in new monitoring data. Experimental results demonstrate that the proposed method can yield high accuracy as compared to preexisting algorithms.

INTRODUCTION

Content-Centric Networking (CCN, also referred to as Data-Centric Networking or Named Data Networking) has emerged to overcome the inherent limitations of the current Internet regarding content security and privacy, and to provide a better trust model (Ahlgren et al., 2011; Jacobson et al., 2009). Unlike the current Internet (host-centric approach) in which security mechanisms are based on the communication channels between hosts, in the content-centric network, security mechanisms must be applied to the Information Objects (IOs) themselves independent of its storage location and physical

DOI: 10.4018/978-1-4666-9474-3.ch017

representation. Consequently, new information-centric security concepts based on the information itself are required (Ahlgren et al., 2011). With this new paradigm, new kinds of attacks and anomalies –from Denial of Service (DoS) to privacy attacks– will arise (Karami & Guerrero-Zapata, 2015). Attacks and anomalies are deliberate actions against data, contents, software or hardware that can destroy, degrade, disrupt or deny access to a computer network (Louvieris et al., 2013). Hence, the contents should be resilient against both DoS and new forms of (unknown) attacks or at least limit their effectiveness (Gasti et al., 2013). In order to disarm new kinds of attacks, anomalous traffics, and any deviation, not only the detection of the malevolent behavior must be achieved, but also the network traffic belonging to the attackers should be also blocked (Liao et al., 2013; Kolias et al., 2011; Peddabachigari et al., 2007). In an attempt to tackle with the new kinds of anomalies and the threat of future unknown attacks, many researchers have been developing Intrusion Detection Systems (IDS) to help filter out known malware, exploits and vulnerabilities (Louvieris et al., 2013, Patcha & Park, 2007). Anomaly detection systems are becoming increasingly vital and valuable tools of any network security infrastructure in order to mitigate disruptions in normal delivery of network services due to malicious activities, Denial of Service (DOS) attacks and network intrusions (Palmieri & Fiore, 2010; Perdisci et al., 2009). An IDS dynamically monitors logs and network traffics, applying detection algorithms to identify potential intrusions and anomalies within a network (Faysel & Haque, 2010; Krawczyk & Woźniak, 2014). In recent years, data mining techniques specially unsupervised anomaly detection have been employed with much success in the area of intrusion detection (Krawczyk & Woźniak, 2014; Fiore et al., 2013; Chandola et al., 2009). Generally, unsupervised learning or cluster analysis algorithms have been utilized to discover natural groupings of objects and find features inherent and their deviations with similar characteristics to solve the detection problems of the abnormal traffics and unknown forms of new attacks (Corral et al., 2009; Wang & Megalooikonomou, 2010). Data clustering algorithms can be either hierarchical or partitioning (Jain et al., 1999; Karami, 2013). In this paper, we focus on the partitioning clustering and in particular, a popular method called K- means clustering algorithm. The K-means algorithm is one of the most efficient clustering algorithms (Kao et al., 2008; Laszlo & Mukherjee, 2007; Zalik, 2008). This algorithm is simple, easy to implement, straightforward, suitable for large data sets, and very efficient with linear time complexity (Chen & Ye, 2004). However, it suffers from two main drawbacks: (1) the random selection of centroid points and determining the number of clusters may lead to different clustering results, (2) the cost function is not convex and the K-means algorithm may contain many local optimum (Selim & Ismail, 1984). In the previous work (Karami, 2013), we employed K-means clustering in our anomaly detection system over CCN. But, the results were not appropriate due to the large number of clusters, trapping in the local optimum solution, and changing results by running the algorithm with the constant parameters in several times. However, if good initial clustering centroids can be assigned by any of other global optimal searching techniques, the K-means would work well in refining the cluster centroids to find the optimal centroids (Naldi & Campello, 2014; Anderberg, 1973).

To overcome these drawbacks, we present a fuzzy anomaly detection system in two phases: training and detection. In the training phase, we apply a meta-heuristic algorithm called PSO (Particle Swarm Optimization) which can find the optimal or the near optimal solution by the least iterations (Quan et al., 2014; Carlisle & Dozier, 2001; Kennedy & Eberhart, 2001). We employ the combination of the ability of global search of the PSO with a novel boundary handling approach and the fast convergence of the K-means to avoid being trapped in a local optimal solution. On the other hand, the most clustering