Chapter 1
Evaluation of Kernel Based Atanassov's Intuitionistic Fuzzy Clustering for Network Forensics and Intrusion Detection

Anupam Panwar
Symantec Corporation, Bangalore, India

ABSTRACT

Malware or virus is one of the most significant security threats in Internet. There are mainly two types of successful (partially) solutions available. One is anti-virus and other is backlisting. This kind of detection generally depends on the existing malware or virus signature database. Cyber-criminals bypass defenses by generating variants of their malware program. Traditional approach has limitations such as unable to detect zero day threats or generate so many false alerts et al. To overcome these difficulties, a system is built based on Atanassov's intuitionistic fuzzy set (AIFS) theory based clustering method that takes care of these problems in a robust way. It not only raises an alert for new kind of malware but also decreases the number of false alerts. This is done by giving it decision-making intelligence. There is not much work done in the field of network forensics using AIFS theory. Some clustering techniques are used in these fields but those have limitations like accuracy, performance or difficulty to cluster noisy data. This method clusters the malwares/viruses with high accuracy on the basis of severity. Experiments are performed on several pcap files with malware traffic to assess the performance and accuracy of the method and results are compared with different clustering algorithms.

DOI: 10.4018/978-1-7998-3025-2.ch001

1. INTRODUCTION

In this digital era, Internet has become a part of daily life and an essential component of business, education and entertainment today. It is widely used for personal and business purposes. But along with its widespread application, there is always a risk of getting infected by a malware or virus.

First step for an attacker is to install their malware programs on as many victim machines as possible. There are two widely used approaches to do that. First is via email with malware attached to it and other is to tempt users onto malicious web pages and the malware is secretly installed and launched on the victim's machine. There are various kind of techniques designed to detect and block the Internet-based attacks. Intrusion detection systems (IDSs) are one of them. IDSs provide a wall of defense to challenge the attacks of computer systems on Internet. Most of them are based on data mining or machine learning techniques. Different types of malicious network communications and computer systems usage can be detected using IDSs, whereas the conventional firewall cannot perform this task. Intrusion detection is based on the assumption that malware behavior is different from uninfected file behavior (Invernizzi et al., 2014).

In general, there are two types of IDSs: Statistical anomaly based IDSs and Signature based IDSs. Anomaly based IDS tries to figure out whether deviation from normal behavior can be marked as intrusion. On the other hand, signature based IDS will monitor packets on the network and compare them against a database of signatures or attributes from known malicious threats.

Different machine learning techniques are used for the development of various anomaly based detection systems (Tsai et al., 2009). For example, some studies apply different supervised learning techniques, such as neural networks, support vector machines, etc. (Vapnik, 1998; Haykin, 1999) . On the other hand, some are based on unsupervised techniques like genetic algorithm (Koza, 1992; Abadeh et al., 2007). However, very less attention has been given to Atanassov's intuitionistic fuzzy set (AIFS) theory (Zadeh, 1965) in the field of intrusion detection systems.

AIFS theory is rarely used in clustering algorithms. AIFS is built on top fuzzy set theory (Zimmermann, 2001). In fuzzy set theory, membership is the degree of belongingness of an element in a set or the membership function. But the membership function is not precise, as there is always hesitation present while defining the membership function. Due to this hesitation, non-membership degree is not the complement of the membership degree as in fuzzy set, rather less than or equal to the complement of membership degree. Atanassov (1986) introduced intuitionistic fuzzy set (AIFS) theory that considers the hesitation in the membership function. A new intuitionistic fuzzy based kernel clustering was suggested by Chaira and Panwar (2014) where another function is introduced that is the intuitionistic fuzzy entropy in the objective function. In most of the signature based IDSs, detection is done by analyzing series of bytes in the file. It could also be a cryptographic hash of the file or its sections. But they don't consider other details of network packets like Source IP, URLs etc. May be they are marked green by signature based IDS but they are actually coming from malicious server or through a malicious URL (iMPERVA; Gostev). In the approach, pcap attributes were narrowed down to three attributes (Source IP, File hash SHA256, URL) as part feature extraction. They are important red flag indicators.

In this paper, Kernel based A-intuitionistic fuzzy kernel based intrusion detection algorithm is evaluated and compared. Sugeno type fuzzy complement (Sugeno, 1977) is used to calculate the non-membership values and then hesitation degree. In the algorithm, trivariate dataset of three features – source IP, file hash SHA256, and URL is used. Experiments are performed on several malicious pcap files.