Journal of Network and Computer Applications 60 (2016) 19–31

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Review

A survey of network anomaly detection techniques

Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu

School of Engineering and Information Technology, UNSW Canberra, ACT 2600, Australia

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form
29 October 2015
Accepted 19 November 2015
Available online 11 December 2015

Keywords:
Intrusion detection
Computer security
Anomaly detection
Classification
Clustering
Information theory

ABSTRACT

Information and Communication Technology (ICT) has a great impact on social wellbeing, economic growth and national security in todays world. Generally, ICT includes computers, mobile communication devices and networks. ICT is also embraced by a group of people with malicious intent, also known as network intruders, cyber criminals, etc. Confronting these detrimental cyber activities is one of the international priorities and important research area. Anomaly detection is an important data analysis task which is useful for identifying the network intrusions. This paper presents an in-depth analysis of four major categories of anomaly detection techniques which include classification, statistical, information theory and clustering. The paper also discusses research challenges with the datasets used for network intrusion detection.

Contents

http://dx.doi.org/10.1016/j.jnca.2015.11.016
1084-8045/

1. Introduction

Computer security has become a necessity due to proliferation of information technologies in everyday life. The mass usage of computerized systems has given rise to critical threats such as zero-day vulnerabilities, mobile threats, etc. Despite research in the security domain having increased significantly, are yet to be mitigated. The evolution of computer networks has greatly exacerbated computer security concerns, particularly internet security in today's networking environment and advanced computing facilities. Although Internet Protocols (IPs) were not designed to place a high priority on security issues, network administrators now have to handle a large variety of intrusion attempts by both individuals with malicious intent and large botnets (Papalexakis et al., 2012). According to Symantecs Internet Security Threat Report, there were more than three billion malware attacks reported in 2010 and the number of denial of service attacks increased dramatically by 2013 (Symantec internet security threat report, 2014). As stated in Verizon's Data Breach Investigation Report 2014, 63,437 security breaches carried out by hackers (Verizon's data breach investigation report, 2014). The Global State of Information Security Survey 2015 (The Global State of Information Security Survey, 2015) found an increase in great rise of incidents. Figure 1 shows the security incidents growth from 2009 to 2014. Therefore, the detection of network attacks has become the highest priority today. In addition, the expertise required to commit cyber crimes has decreased due to easily available tools (Hacking and cracking tools, 2014).

Anomaly detection is an important data analysis task that detects anomalous or abnormal data from a given dataset. It is an interesting area of data mining research as it involves discovering enthralling and rare patterns in data. It has been widely studied in statistics and machine learning (Ahmed et al., 2014), and also synonymously termed as outlier detection, novelty detection, deviation detection and exception mining. Although an anomaly is defined by researchers in various ways based on its application domain, one widely accepted definition is that of Hawkins (Hawkins, 1980): '*An anomaly is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism*'. Anomalies are considered important because they indicate significant but rare events and can prompt critical actions to be taken in a wide range of application domains; for example, an unusual traffic pattern in a network could mean that a computer has been hacked and data is transmitted to unauthorized destinations; anomalous behavior in credit card transactions could indicate fraudulent activities, and an anomaly in a MRI image may indicate the presence of a malignant tumor (Ahmed et al., 2015a). Anomaly detection has been widely applied in countless application domains such as medical and public health, fraud detection, intrusion detection, industrial damage,

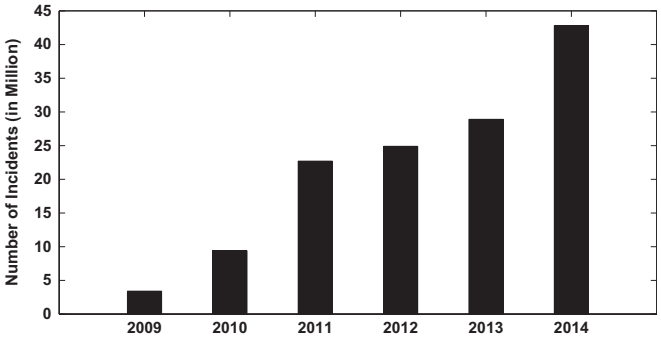

Fig. 1. Growth of information security incidents (The Global State of Information Security Survey, 2015).

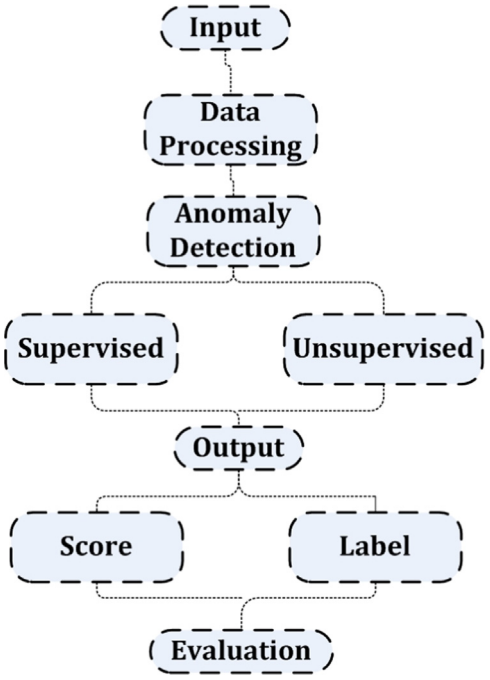

Fig. 2. Generic framework for network anomaly detection.

image processing, sensor networks, robots behavior and astronomical data (Mahmood et al., 2010; Ahmed et al., 2015b).

Figure 2 displays a generic framework for network anomaly detection. The input data requires processing because the data are of different types, for example, the IP addresses are hierarchical, whereas the protocols are categorical and port numbers are numerical in nature (Mahmood et al., 2008). Processing techniques are based on the individual anomaly detection techniques. Then, the anomaly detection techniques (broadly categorized in two: supervised and unsupervised) are applied on the data. For evaluation of the output, either scores or labels are used (discussed in Section 2.2).

Although network anomaly detection seems very straightforward, we need to find the data that do not follow normal behavioral patterns. Despite the many techniques available, following are the research challenges.

- A lack of universally applicable anomaly detection technique; for example, an intrusion detection technique in a wired network may be of little use in a wireless network.
- Data contains noise which tends to be an actual anomaly and, therefore, is difficult to segregate.
- A lack of publicly available labeled dataset to be used for network anomaly detection.
- As normal behaviors are continually evolving and may not be normal forever, current intrusion detection techniques may not be useful in the future (Qin et al., 2011). A need for newer and more sophisticated techniques because the intruders are already aware of the prevailing techniques.

Due to the aforementioned challenges, network anomaly detection has been more challenging than it was before. As most existing supervised techniques are based on knowledge provided by an external agent, they require labeled data and are unable to detect zero-day vulnerabilities. The research community has increased interest about proactive network security systems.

During the last decade several surveys of network intrusion detection have been conducted. One of the earliest was that of Towards a taxonomy of intrusion-detection systems (1999) who