Principal Component-based Anomaly Detection Scheme*

Mei-Ling Shyu

Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL, USA
shyu@miami.edu

Shu-Ching Chen

School of Computer Science
Florida International University
Miami, FL, USA
chens@cs.fiu.edu

Kanoksri Sarinnapakorn

Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL, USA
ksarin@miami.edu

LiWu Chang

Center for High Assurance Computer Systems
Naval Research Laboratory
Washington, DC, USA
lchang@itd.nrl.navy.mil

* The preliminary version of the current draft was published in [23].

Abstract. *In this chapter, a novel anomaly detection scheme that uses a robust principal component classifier (PCC) to handle computer network security problems is proposed. An intrusion predictive model is constructed from the major and minor principal components of the normal instances, where the difference of an anomaly from the normal instance is the distance in the principal component space. The screening of outliers prior to the principal component analysis adds the resistance property to the classifier which makes the method applicable to both the supervised and unsupervised training data. Several experiments using the KDD Cup 1999 data were conducted and the experimental results demonstrated that our proposed PCC method is superior to the k-nearest neighbor (KNN) method, density-based local outliers (LOF) approach, and the outlier detection algorithm based on the Canberra metric.*

Keywords: *Anomaly detection, principal component classifier, data mining, intrusion detection, outliers, principal component analysis.*

1 Introduction

A rapid technological progress has brought a new era to the way people communicate. With the merging of computers and communications, we now find ourselves highly depend on the digital communication networks in everyday life. For example, people can look for information from everywhere on the Internet and check e-mails or messages at any place, either from a desktop personal computer, a laptop, or even a mobile phone. While we treasure the ease and convenience of being connected, it is also recognized that an intrusion of malicious or unauthorized users from one place can cause severe damages to wide areas. This introduces a serious issue in computer network security. Heady et al. [9] defined an intrusion as "any set of actions that attempt to compromise the integrity, confidentiality or availability of information resources." The identification of such a set of malicious actions is called intrusion detection problem that has received great interest from the researchers.

The existing intrusion detection methods fall in two major categories: *signature recognition* and *anomaly detection* [11,18,19]. For signature recognition techniques, signatures of the known attacks are stored and the monitored events are matched against the signatures. When there is a match, the techniques signal an intrusion. An obvious limitation of these techniques is that they cannot detect new attacks whose signatures are unknown. In contrast, anomaly detection algorithms build a model from the normal training data and detect the deviation from the normal model in the new piece of test data, where a large departure from the normal model is