Journal of Network and Computer Applications 32 (2009) 31–44

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Fast intrusion detection based on a non-negative matrix factorization model ☆

Xiaohong Guan [a,b], Wei Wang [a,*], Xiangliang Zhang [a]

[a] MOE Key Lab for Intelligent Networks and Network Security (KLINNS) and State Key Lab for Manufacturing Systems (SKLMS), Xi'an Jiaotong University, Xi'an 710049, China
[b] Center for Intelligent and Networked Systems, TNLIST Lab, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 7 November 2007
Received in revised form
2 April 2008
Accepted 23 April 2008

Keywords:
Computer security
Intrusion detection system
Anomaly detection
Non-negative matrix factorization

ABSTRACT

In this paper, we present an efficient fast anomaly intrusion detection model incorporating a large amount of data from various data sources. A novel method based on non-negative matrix factorization (NMF) is presented to profile program and user behaviors of a computer system. A large amount of high-dimensional data is collected in our experiments and divided into smaller data blocks by a specific scheme. The system call data is divided into blocks by processes, while command data is divided into consecutive blocks with a fixed length. The frequencies of individual elements in each block of data are computed and placed column by column as data vectors to construct a matrix representation. NMF is employed to reduce the high-dimensional data vectors and anomaly detection can be realized as a very simple classifier in low dimensions. Experimental results show that the model presented in this paper is promising in terms of detection accuracy, computation efficiency and implementation for fast intrusion detection.

1. Introduction

Computer network security is gaining worldwide attention as attacks on computer network systems have become more and more widespread in recent years. In many cases, sophisticated

☆ The research presented in this paper was supported in part by the NSFC (60736027, 60574087), 863 High Tech Development Plan (2007AA01Z475, 2007AA04Z154, 2007AA01Z480, 2007AA01Z464) and 111 International Collaboration Program, of China.

* Corresponding author.
E-mail addresses: xhguan@tsinghua.edu.cn (X. Guan), wei.wang.email@gmail.com (W. Wang), xlzhang@lri.fr (X. Zhang).

1084-8045/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2008.04.006

hackers successfully penetrated many conventional peripheral protection mechanisms such as firewalls and various authentications, and caused enormous security and economic damages (Zou et al., 2002; Moore and Shannon C, 2002; Kruegel et al., 2005; Moore et al.). Moreover, distributed and coordinated attacks launched mostly from mal-codes and spam emails may cease normal functions of a large portion of Internet in less than 15 min (Moore et al.). Therefore, fast-automated and integrated detection mechanisms and control strategies are required, and intrusion detection systems (IDS) together with intelligent control of network infrastructure consisting of routers, switches, firewalls, hosts, etc., constitute the defense-in-depth or layered framework for securing a computer network system. The concepts of system modeling and control can play an important role in defending attacks and preventing intrusions. The framework of the integrated network defense system is shown in Fig. 1, where the information gathered from the IDSs with different data resources (host and network) and the flow analyzer for monitoring bursts of large-scale distributed denial of service attacks caused by worms, etc., is fed into control centers. The decisions are made and control actions are taken at the control centers with coordination and then sent to firewall, switches, routers, etc., to prevent intrusions and attacks in real time. Fast intrusion detection, therefore, is very important so that appropriate response action can be taken as soon as possible for intrusion prevention.

Intrusion detection is a technology for detecting hostile attacks against computer systems from both outside and inside. In terms of detection mechanism, the techniques for intrusion detection can be classified into two categories: signature-based detection and anomaly detection. Signature-based detection looks for evidences of malicious behaviors matched against pre-defined descriptions of attacks or signatures. Although signature-based detection is effective for detecting known attacks, it generally cannot detect the new attacks that are not predefined. Anomaly detection, on the other hand, builds the profile of normal behaviors and attempts to identify the patterns or activities that deviate from the normal profile. Based on the concept of profiling normal behaviors, a salient feature of anomaly detection is that it can detect unknown attacks. However, it may also cause significant number of false alarms since the model assumed to describe complete normal behaviors may not be accurate, and obtaining such a model usually by machine learning is difficult. The research focus on anomaly detection is to find more effective and accurate methods.

Anomaly detection is an active research area and has been widely studied for more than a decade since it was originally proposed by Denning (1987). There are multiple levels that anomaly detection can be built upon in an actual computer network system. A great challenge is to select features that best characterize patterns of a subject (e.g., a program, a user, etc.), so that abnormal activities can be

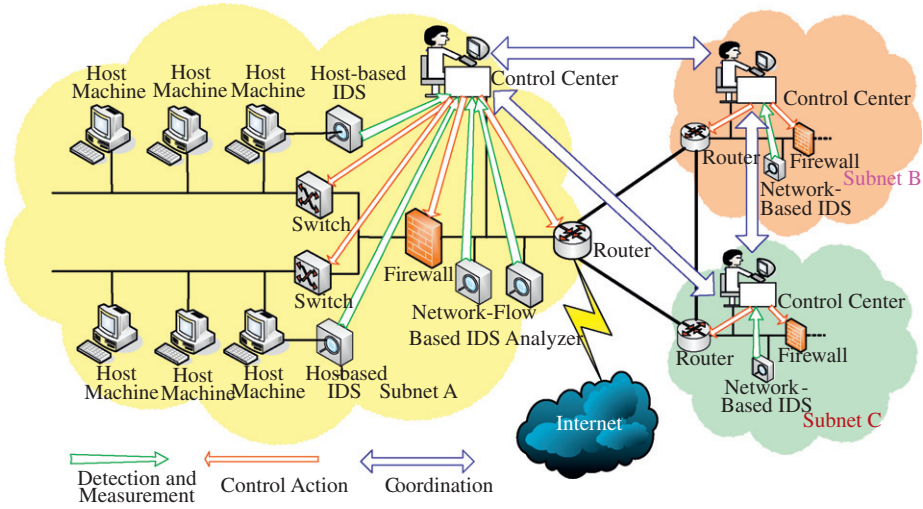

Fig. 1. The framework of the integrated network defense system.