Available online at www.sciencedirect.com
ScienceDirect
Journal of Network and Computer Applications 30 (2007) 114–132
Journal of NETWORK and COMPUTER APPLICATIONS
www.elsevier.com/locate/jnca

Modeling intrusion detection system using hybrid intelligent systems

Sandhya Peddabachigari[a], Ajith Abraham[b,*], Crina Grosan[c], Johnson Thomas[a]

[a]*Computer Science Department, Oklahoma State University, OK 74106, USA*
[b]*School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea*
[c]*Department of Computer Science, Babes-Bolyai University, Cluj-Napoca 3400, Romania*

Received 28 June 2005; accepted 28 June 2005

Abstract

The process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusions is known as intrusion detection system (IDS). This paper presents two hybrid approaches for modeling IDS. Decision trees (DT) and support vector machines (SVM) are combined as a hierarchical hybrid intelligent system model (DT–SVM) and an ensemble approach combining the base classifiers. The hybrid intrusion detection model combines the individual base classifiers and other hybrid machine learning paradigms to maximize detection accuracy and minimize computational complexity. Empirical results illustrate that the proposed hybrid systems provide more accurate intrusion detection systems.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Intrusion detection system; Hybrid intelligent system; Decision trees; Support vector machines; Ensemble approach

*Corresponding author.
E-mail addresses: ajith.abraham@ieee.org (A. Abraham), crina.grosan@ieee.org (C. Grosan), jpt@cs.okstate.edu (J. Thomas).

1084-8045/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2005.06.003

1. Introduction

Traditional protection techniques such as user authentication, data encryption, avoiding programming errors and firewalls are used as the first line of defense for computer security. If a password is weak and is compromised, user authentication cannot prevent unauthorized use, firewalls are vulnerable to errors in configuration and suspect to ambiguous or undefined security policies (Summers, 1997). They are generally unable to protect against malicious mobile code, insider attacks and unsecured modems. Programming errors cannot be avoided as the complexity of the system and application software is evolving rapidly leaving behind some exploitable weaknesses. Consequently, computer systems are likely to remain unsecured for the foreseeable future. Therefore, intrusion detection is required as an additional wall for protecting systems despite the prevention techniques. Intrusion detection is useful not only in detecting successful intrusions, but also in monitoring attempts to break security, which provides important information for timely countermeasures (Heady et al., 1990; Sundaram, 1996). Intrusion detection is classified into two types: misuse intrusion detection and anomaly intrusion detection.

Misuse intrusion detection uses well-defined patterns of the attack that exploit weaknesses in system and application software to identify the intrusions (Kumar and Spafford, 1995). These patterns are encoded in advance and used to match against user behavior to detect intrusions. Anomaly intrusion detection identifies deviations from the normal usage behavior patterns to identify the intrusion. The normal usage patterns are constructed from the statistical measures of the system features, for example, the CPU and I/O activities by a particular user or program. The behavior of the user is observed and any deviation from the constructed normal behavior is detected as intrusion.

Several machine-learning paradigms including neural networks (Mukkamala et al., 2003), linear genetic programming (LGP) (Mukkamala et al., 2004a), support vector machines (SVM), Bayesian networks, multivariate adaptive regression splines (MARS) (Mukkamala et al., 2004b) fuzzy inference systems (FISs) (Shah et al., 2004), etc. have been investigated for the design of IDS. In this paper, we investigate and evaluate the performance of decision trees (DT), SVM, hybrid DT–SVM and an ensemble approach. The motivation for using the hybrid approach is to improve the accuracy of the intrusion detection system when compared to using individual approaches. The hybrid approach combines the best results from the different individual systems resulting in more accuracy. The rest of the paper is organized as follows. The Literature review is presented in Section 2 followed by a short theoretical background on the machine-learning paradigms used in this research. Experimental results and analysis is presented in Section 4 and conclusions presented at the end.

2. Related research

With the proliferation of networked computers and the Internet, their security has become a primary concern. In 1980, James Anderson proposed that audit trails