*International Journal of Innovative Computing, Information and Control*
Volume **4**, Number **2**, February **2008**

ICIC International ©2008 ISSN 1349-4198
pp. 413–424

# INTRUSION DETECTION USING A HYBRID SUPPORT VECTOR MACHINE BASED ON ENTROPY AND TF-IDF

Rung-Ching Chen and Su-Ping Chen

Department of Information Management
Chaoyang University of Technology
168 Gifeng E. Rd., Wufeng, Taichung, Taiwan
crching@cyut.edu.tw

Received December 2006; revised April 2007

Abstract. *The main functions of an Intrusion Detection System (IDS) are to protect computer networks by analyzing and predicting the actions of processes. Though IDS has been developed for many years, the large number of alerts makes the system inefficient. In this paper, we proposed a classification method based on Support Vector Machines (SVM) with a weighted voting schema to detect intrusions. First, the entropy and TF-IDF (term frequency and inverse documents frequency) features are extracted from processes. Next, entropy and TF-IDF features are sent to the SVM model for learning and testing. Finally, we use a voting schema named Weighted Voting SVM (WV-SVM) to determine whether a process is an intrusion. Our experiments demonstrate improved efficiency.*
**Keywords:** Intrusion detection system, Entropy, TF-IDF, Support vector machine

1. **Introduction.** The security of computer networks is vital for both users and owners. While firewalls are a good way to prevent attacks, they cannot perform real-time monitoring of the network. In general, Intrusion Detection Systems (IDS) produce a high rate of false positives, creating extra work for information technology managers [1, 2, 3]. Thus, lowering the false positive rate is an important goal for IDS.

The major difference between a firewall and an IDS is that a firewall is a manual passive defense system. In contrast, IDS collects packets from the network, monitors and analyzes them and then provides the results of its analysis to human managers. These results could show either an attack or normal behavior. Thus, the IDS acts as the "second line of defense".

An ideal IDS system offers a high attack detection rate and low false positive rate, but in practice this is hard to achieve. The primary goal of IDS is to detect abnormal behaviors on the hosts or networks by monitoring various activities, such as log on attempts. The IDS will send a warning message to the managers if it detects an attack. There are two main detection methods of IDS: misuse and anomaly [4].

The idea of misuse detection is to establish a pattern or a signature form so that the same attack can be detected. Thus, the main drawback of misuse detection is it cannot detect new types of attacks. The IDS has a pattern database that includes signatures of possible attacks. If the system matches the data with the attack pattern, the IDS regards it as an attack. Consequently, misuse detection provides a low false positive rate. At present, some misuse detection systems are being produced for network security. The misuse detection IDS Snort, based on Libpcap, was developed by Marty Roesch [5]. It consists of four principal elements: Packet Decode, Preprocessor, Detection Engine and Output. Its basic functions are to monitor, collect, and display packets. Snort matches