Chapter 1.12
Intrusion Detection Using Modern Techniques:
Integration of Genetic Algorithms and Rough Set with Neural Nets

Tarun Bhaskar
Indian Institute of Management, Calcutta, India

Narasimha Kamath B.
Indian Institute of Management, Calcutta, India

ABSTRACT

Intrusion detection system (IDS) is now becoming an integral part of the network security infrastructure. Data mining tools are widely used for developing an IDS. However, this requires an ability to find the mapping from the input space to the output space with the help of available data. Rough sets and neural networks are the best known data mining tools to analyze data and help solve this problem. This chapter proposes a novel hybrid method to integrate rough set theory, genetic algorithm (GA), and artificial neural network. Our method consists of two stages: First, rough set theory is applied to find the reduced dataset. Second, the results are used as inputs for the neural network, where a GA-based learning approach is used to train the intrusion detection system. The method is characterized not only by using attribute reduction as a pre-processing technique of an artificial neural network but also by an improved learning algorithm. The effectiveness of the proposed method is demonstrated on the KDD cup data.

INTRODUCTION

The need for secured networked systems is now well established. With the widespread use of the Internet and other computer networks, both e-commerce and data communications depend on secure networks. Organizations have embraced information technology to share information and streamline their business operations. This makes it critical to have networks that function efficiently and reliably. E-business mode of operation is a necessity for today's global organizations. This also has brought with it the unwanted effect of security breaches.

A good intrusion detection system (IDS) is important to ensure the survivability of network systems. Intrusion detection is based on the fact that an intruder's behavior will be significantly different from that of a legitimate user. The number of intrusions is dramatically increasing and so are the costs associated with them. The number of incidents reported to Carnegie Melon's Computer Emergency Response Team/Coordination Center (CERT/CC) has increased from the range of 2,000 to 3,000 in the early and mid 1990s, to 52,658 in 2001, 82,094 in 2002, and 137,529 in 2003. It also reports that e-crime has cost the organizations approximately $666 million in 2003. The data published by the U.S. General Accounting Office shows about 250,000 attempts to attack the system and only 1 to 4% of those are detected. Rule mining can be used on large databases to generate learning algorithm to detect the attack on the site. This approach has tremendous thrust in numerous business applications such as e-commerce, recommender systems, supply chain management, group decision support systems, and so forth.

IDS helps network administrators prepare for and deal with network security attacks. It collects information from a variety of systems and network sources, which is then analyzed for signs of intrusion and misuse. Identifying the appropriate method is important in network intrusion since performance in terms of detection accuracy, false alarm rate, and detection time become critical for near real-time monitoring. Data mining is a very useful technique to extract meaningful information and improve the decision-making process. The extracted information is refined to gain useful knowledge, which is then used to predict, classify, model, and summarize the data being mined. Rule induction, neural networks, genetic algorithms, fuzzy logic, and rough sets are the widely used data mining techniques for industry classification and pattern recognition. The output of the IDS can be represented as shown in Table 1.

The traffic at a site can be broadly classified into normal and abnormal. We refer to abnormal traffic here as *attack*. A deployed IDS evaluates all the traffic that passes through it and classifies it as normal or attack. Proper detection of the attack situation is more important than a normal situation being classified as an attack. So, the emphasis here is on proper detection of the attacks, while keeping false alarms within acceptable level.

In order to develop better security and defense mechanisms against network attacks, it is important to investigate the patterns of attacks on network systems and sites. Data mining techniques have shown promising results when applied to such problems. In this chapter, we build on existing methods of IDS and evaluate the applicability of artificial neural networks and rough sets for the purposes of intrusion detection. A hybrid model is constructed with the integration of rough sets with neural networks by utilizing their complementary nature. Rough set is used as a preprocessing tool to eliminate the redundant data from the huge database. This reduces the learning time of the

Table 1. Output of IDS

		Actuality	
Prediction by the IDS		Attack	No Attack
	Attack	Detected	False Alarm
	No Attack	Not Detected	Correct Prediction