Neural Networks for Intrusion Detection

Rui Ma
Beijing Institute of Technology, China

INTRODUCTION

With the rapid expansion of computer networks, network security has become a crucial issue for modern computer systems. As an important and active defense technology, the intrusion detection system (IDS) plays an important role in defensive systems. IDSs provide real-time protection from interior attacks, exterior attacks, and invalid operations, and it can intercept intrusions and respond whenever the network system integrity is violated (Ma, 2004). Many intrusion detection approaches have been deeply researched and some widely deployed. But the diversification, complexity, and scale of intrusions raise new demands for IDSs. Neural networks are tolerant of imprecise data and uncertain information. With their inherent ability to generalize from learned data they seem to be an appropriate approach to IDSs (Hofmann, Schmitz, & Sick, 2003). This article discusses the detection of distributed denial-of-service (DDoS) attacks using artificial neural networks techniques. The implementation of a distributed intelligent intrusion detection system (DIIDS) is described, including both the data processing technique and neural networks approaches adopted.

BACKGROUND

Intrusion Detection System

Many IDSs are based on the general model proposed by Denning (1987). This model is independent of platform, system vulnerability, and type of intrusion. It maintains a set of historical profiles for users, matches an audit record with the appropriate profile, updates the profile whenever necessary, and reports any attacks detected.

IDSs can be divided into two types: (1) host-based IDSs and (2) network-based IDSs. Host-based IDSs evaluate information found on a single or multiple host systems, including contents of operating systems, system logs and application files. Network-based IDSs evaluate information captured from network communications, by analyzing the stream of packets traveling across the network. Packets are captured through a set of sensors placed at strategic points in the network (Jean & Philippe, 2001).

Intrusion detection schemes can be classified into two general categories (Ghosh, 1999a): (1) misuse detection and (2) anomaly detection. Misuse detection techniques assume that all kinds of intrusion behavior can be described as specific patterns, thus allowing the identification of intrusive behavior by comparing current user activity with specific patterns that have been observed previously during an attack. The most significant advantage of misuse detection techniques is that known attacks can be detected fairly reliably and with a low false positive rate. However, the key limitation of misuse detection techniques is that they cannot detect novel attacks.

Anomaly detection techniques assume that all kinds of intrusion behavior differ from normal user activities. Any current user behavior sufficiently deviant from the normal user activities will be flagged as anomalous and hence considered as a possible attack. The most significant advantage of anomaly detection techniques is that it directly addresses the problem of detecting novel attacks against systems. However, the most notable disadvantage of anomaly detection techniques is the high rates of false alarm.

In order to detect known attacks, subtle variations of known attacks, and novel attacks efficiently, IDSs should selectively combine aspects of both misuse detection techniques and anomaly detection techniques (Chen, 2004).

Neural Networks

An artificial neural network consists of a collection of processing elements that are highly interconnected and that transform a set of inputs to a set of desired outputs. The result of the transformation is determined by the characteristics of the elements and the weights associated with the interconnections among them. By modifying the connections between the nodes, the network is able to adapt to the desired outputs.

The neural network gains the necessary experience initially by being trained to correctly identify preselected examples of a problem. The response of the neural network is reviewed and the configuration of the system is refined until the neural network's analysis of the training data reaches a satisfactory level. In addition to the initial training period, the neural network also gains experience over time as it conducts analyses on data related to the problem.

The training algorithms of neural networks can be classified into two general categories: (1) supervised and (2) unsupervised. In the learning phase, a supervised algorithm

learns the desired output for a given input or pattern. Whereas an unsupervised algorithm learns without specifying the desired output (Jean & Philippe, 2001).

Neural Network Intrusion Detection Systems

Neural networks have capabilities of self-learning, self-organization, and self-adaptivity; as well as a capability to analyze fuzzy, nonstructured, imprecise, incomplete data and to generalize from previously observed behaviors. With the continuous development of network technologies and increasing multiplicity and novelty of network attacks, IDSs must be more flexible and efficient. Artificial neural networks offer the potential to resolve a number of the problems encountered by the other approaches to intrusion detection.

Various approaches to using neural networks for intrusion detection have been advocated. One approach is to create keyword, count-based, misuse detection systems with neural networks (Lippmann & Cunningham, 1999; Ryan, Lin, & Miikkulainen, 1998). The data that are presented to the neural network consist of attack-specific keyword counts in network traffic. Such an approach is close to a host-based IDS. Some researchers created a neural network to analyze program behavior profiles instead of user behavior profiles (Ghosh, 1999b). This method identifies the normal system behavior of certain programs and compares it to the current system behavior. Cannady (1998) developed a network-based neural network IDS in which packet-level network data were retrieved from a database and then classified according to the packet characteristics before being presented to a neural network.

In the case of DIIDS, three basic artificial neural networks approaches are described: (1) neural network expert system, which is mainly used to detect known attacks by using forward parallel inference which improves inference efficiency; (2) back-propagation neural network; and (3) adaptive resonance theory neural network, which are used to detect not only known attacks, but variations of them or indeed, unknown attacks. Finally, based on the neural network approaches just described, cooperative intrusion detection approaches are discussed.

NEURAL NETWORKS IN INTRUSION DETECTION SYSTEMS

The Distributed Intelligence Intrusion Detection System

DIIDS is based on neural networks techniques and is characterized by the following constituent components: data collection; detection and analysis; console; and database. DIIDS can be deployed in a distributed or integrated manner. In particular, DIIDS combines the capability of host-based and network-based IDSs and uses artificial neural networks as detective techniques.

The console is responsible for the management, monitoring, and reporting status of the system. Storage of feature data and detection results is undertaken by the database. Data collection assembles and processes both network-based and host-based data before identifying feature data. Detection and analysis then determines if the attack occurs through three basic neural networks approaches: (1) neural network expert system; (2) back-propagation neural network; and (3) adaptive resonance theory neural network. In addition, the cooperative intrusion detection technique is also used.

Data Processing

Three levels of data processing are conducted. Initially, data is selected from the available data set by the data collection component. The data collection component comprises a network engine, a host agent, and a data fusion facility. The network engine collects all data packages flowing in the network and then resolves them into the feature data set for the network. Eight elements of the feature data garnered from the network are: (1) protocol type, (2) source IP address, (3) destination IP address, (4) source port, (5) destination port, (6) sequence number, (7) acknowledgement number, and (8) raw data. The host agent collects the data that characterizes system performance, for example, network traffic, memory, and CPU usage. These form the original feature data. A second phase converts the data elements such as protocol type, source, and destination IP address into a standardized numeric representation. The third part is that the process concerns the association processing.

By analyzing the feature data, both the spatial and temporal associate relationships between intrusion behaviors can be determined. Some attacks, which originate from different sources and try to attack the same specific goal, send many packages which have identical destination IP addresses and different source IP addresses. The relationship of these packages is called the *spatial associate relationship*. The *temporal associate relationship* points to many packages which have same source IP address but try to attack the same specific goal in a certain time period. In order to embody the spatial and temporal associate relationship in a concrete detective record, the associate detection algorithm (Figure 1) must be used. This algorithm processes associate information of feature data and converts the stochastic associated attributes into the detection associated attributes. Based on a variable, continuous, and sliding temporal window, the algorithm calculates the count of any distinct value of stochastic associated attributes during this time. Then the algorithm compares the maximum count with the threshold to determine the value of the associate attributes (Lee, 1999). This approach can