We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

5,000 Open access books available
125,000 International authors and editors
140M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

Over the last 30 years, ubiquitous and networked computing has increasingly gained importance in our life. With the increase in complexity of computer networks, cybersecurity threats have also manifested in a variety of which was unimaginable even a decade back. While the rule-based intrusion detection systems (IDSs) can accurately detect already known attacks on a cyberinfrastructure, these systems are not capable to detect novel, unknown, and polymorphic cyber threats. Moreover, the computational overheads including CPU cycles and memory overheads are unacceptably high for most of the detection systems. Hence, it has been a constant challenge for security researchers to design automated, fast, and yet accurate IDSs for deployment in real-world cyberinfrastructures. From expert-crafted rules to sophisticated machine learning and deep learning algorithms, researchers have explored and attempted to push the boundary of the detection accuracy while minimizing the false alarm rates.

Applications of machine learning and data mining algorithms in both signature and anomaly detection systems have been widely proposed in the literature. In misuse detection systems, following approaches of machine learning are quite popular: (1) classification using association rules [1–3], (2) artificial neural networks [4], (3) support vector machines [5], (4) classification and regression trees [6, 7], (5) Bayesian network classifier [8–10], and (6) naïve Bayes method [11]. While the signature detection systems require labeled training data in order to learn the features of the attack and the normal traffic, anomaly detection systems are based on identifying any significant changes in the system from its normal state. Various approaches to machine learning in anomaly detection have been proposed in the literature. Some of these approaches are as follows: (1) association rule mining [12–14], (2) fuzzy association rule mining [15], (3) artificial neural network [16–18], (4) support vector machines [19, 20], (5) nearest neighbor [21], (6) hidden Markov model [22–24], (7) Kalman filter [25], (8) clustering [26], and (9) random forest [27, 28]. Other machine learning methods have been proposed for learning the probability distribution of data and in applying statistical tests to detect outliers [29–35].

The hybrid detection approach combines the adaptability and the powerful detection ability of an anomaly detection system with the higher accuracy and reliability of the misuse detection approach [28, 36–43]. The selection of misuse and anomaly detection systems for designing a hybrid detection system is dependent on the application in which the detection system is to be deployed. Following a combinational approach, the integration of an anomaly detection system with a misuse detection counterpart has been classified into four categories [28, 36]. These types
are: (1) anomaly–misuse sequence detection, (2) misuse–anomaly sequence detection, (3) parallel detection, and (4) complex mixture detection. The complex mixture model is highly application-specific. Figure 1 depicts the first three categories of hybrid detection systems. Complex detection systems are application-specific, and these systems cannot be represented by any generic architecture.

2. Conclusion

A fundamental challenge in designing an intrusion detection system is the limited availability of appropriate data for model building and testing. Generating data for intrusion detection is an extremely painstaking and complex task that mandates the generation of normal system data as well as anomalous and attack data. If a real-world network environment, generating normal traffic data is not a problem. However, the data may too privacy-sensitive to be made available for public research.

Classification-based methods require training data to be well balanced with normal traffic data and attack traffic data. Although it is desirable to have a good mix of a large variety of attack traffic data (including some novel attacks), it may not be feasible in practice. Moreover, the labeling of data is mandatory with attack and normal traffic data clearly distinguished by their respective labels.

Unlike classification-based approaches that are mostly used in misuse detection, unsupervised anomaly detection-based approaches do not require any prior labeling of the training data. In most of the cases, the attack traffic constitutes the sparse class, and hence, the smaller clusters are most likely to correspond to the attack traffic data. Although unsupervised anomaly detection is a very interesting approach, the results produced by this method are unacceptably low in terms of their detection accuracies.

In a pure anomaly detection approach, the training data are assumed to be consisting of only normal traffic. By training the detection model only on the normal traffic data, the detection accuracy of the system can be significantly improved. Anomalous states are indicated by only a significant state change from the normal state of the system.
In a real-world network that is connected to the Internet, an assumption of attack free traffic is utopian. A pure anomaly detection system can still be trained on a training data that include attack traffic. In that case, those attack traffic data will be considered as normal traffic, and the detection system will not raise an alert when such traffic is encountered in real-world operations. Hence, in order to increase the detection accuracy, attack traffic should be removed from the training data as much as possible. The removal of attack traffic from the training data can be done using updated misuse detection systems or by deploying multiple anomaly detection systems and combining their results by a voting mechanism.

For an intrusion detection system that is deployed in a real-world network, it is mandatory to have a real-time detection capability under a high-speed, high-volume data environment. However, most of the cluster techniques used in unsupervised detection require quadratic time. This renders their deployment infeasible in practical applications. Moreover, the cluster algorithms are not scalable, and they need the entire training data to reside in the memory during the training process. This requirement puts a restriction on the model size. The future direction of research may include studies on scalability and performance of anomaly detection algorithms in conjunction with the detection rate and false positive rate. Most of the currently existing propositions on intrusion detection have not paid adequate attention to these critical issues.

In this book, the following chapters deal with various aspects of network security and cryptography. While the chapters belonging to the network security section broadly discuss different aspects of applications and deployment of security protocols and secure system architecture, the cryptography section discusses various theoretical algorithms and their complexities.
References


Knowledge Discovery in Databases and Data Mining. Portland, OR: ACM; 1996. pp. 146-151


[41] Sen J, Sengupta I. Autonomous agent-based distributed fault-tolerant
