ISSN No. 0976-5697

Volume 8, No. 5, May-June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Intrusion Detection Systems based on Artificial Intelligence

Aafreen Jahan
Department of computer science
Jamia hamdard
New Delhi, India

Prof. M. Afshar Alam
Department of computer science
Jamia hamdard
New Delhi, India

Abstract: There are many attacks over the internet or network, to detect those attacks intrusion detection can be implemented for defencing of any network system. An Intrusion detection model can be based on classification and feature selection based techniques. We can build Intrusion detection system model to find attacks on system and can improve the system using captured data. By applying feature selection approach in machine learning, the NSLKDD data set obtained can be reduced and also can improve the intrusion detection using the captured data. By machine learning techniques, we can increase number of new unseen attacks the system of intrusion detection can be developed. They can learn the preferences of the security officers and show the kind of alerts first that the security officer has previously been more interested.

Keywords: IDS, Data mining, classification, machine learning

1. INTRODUCTION

Different methods of artificial intelligence (AI) have been deployed in intrusion detection, for example, artificial neural networks (ANNs), fuzzy logic, and genetic algorithms (GA). In addition, hybrid intelligent IDSs, such as evolutionary fuzzy neural networks (EFuNN) and evolutionary neural network (ENN)–based IDSs, are also used [1–3].

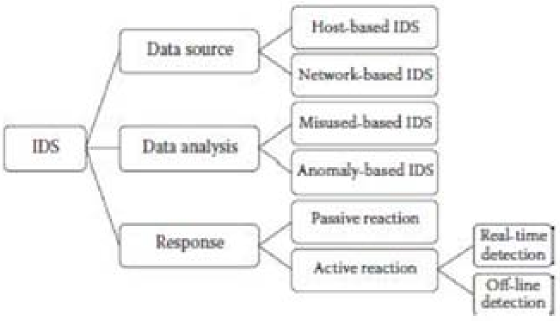
Most of the problems are related to wired networks. There are other types of environments, all of which suffer from a number of security threats, for example, grid computing, cloud computing, and wireless networks. Grid computing shares tasks over different machines without any knowledge about their locations [4]. Cloud computing provides a pool of resources as services on the Internet [4]. Wireless local area networks [5], mobile ad hoc networks [6], and wireless sensor networks are different types of wireless networks [7]. Several methods have been proposed to detect intrusions in wireless networks.

2. INTRUSION DETECTION IN WIRED NETWORKS

An IDS should monitor traffic and detect malicious activities. IDSs can be categorized based on different modules. Figure shows IDS classes upon three modules [8]: data source, data analysis, and response. A data source can be gathered from either an individual computer (host-based IDS, HIDS) or network traffic (network-based IDS, NIDS).

IDS classes

There are two methods for analyzing the collected data: anomaly-based detection and misuse-based detection, which will be explained in the following sections. Anomaly- and misuse-based detection has general meaning for all environments. The third module specifies a suitable response for the suspicious data. This response can be passive or active in terms of behavior. As opposed to passive methods, active IDSs detect and respond to attacks [4, 5].



3. ANOMALY BASED DETECTION

The pattern of normal behavior is used in anomaly-based detection, which can be either self-learned or programmed [12]. In the self-learned anomaly detection, the normal behavior of a system is built automatically. On the other hand, in the programmed detection, a system developer provides the model of normal behavior. Although an anomaly-based IDS is able to detect unknown attacks, it has a high false alarm rate and cannot distinguish between different types of attacks. A number of anomaly detection methods are listed below [11– 12], some of which are machine learning methods, such as neural networks:

Statistical techniques: This technique uses a statistical model for defining normal behavior of the components of the system. All traffic out of this normal behavior will be known as anomalies. This technique assumes that the probability of normal data instances is higher in a stochastic model in comparison with the probability of an anomaly occurrence [10].

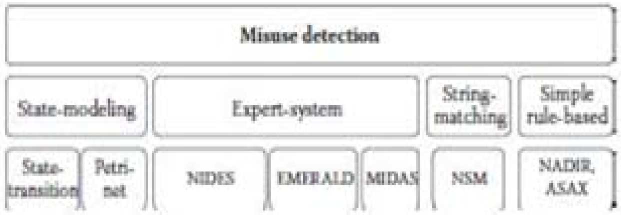
- **Clustering-based methods:** In this method, normal data belong to a cluster, and data not included in any cluster is detected as anomalies [10].
- **Information theoretic**: This method assumes that anomalies cause irregularities in information content of the data set. Different information theoretic measures are used

to analyze the information content, for example, entropy and Kolmogorov Complexity [10].

- **Bayesian networks:** The probabilistic relationships among variables are encoded in the Bayesian method. The combination of this method with a statistical scheme offers better detection capability [12].
- **Data mining methods:** Data mining is the application of machine learning in large databases to provide simple models [13].
- **Neural networks:** This method is inspired by the human brain. Neural networks are flexible and adaptable to environmental changes. They can be deployed to create a user profile, to detect intrusion, and to predict the future behavior of the traffic [12].
- **Support vector machines (SVMs):** The SVM is based on statistical learning theory. One class SVMs and multi-class SVM are well-known methods in classification and regression [14].
- **Nearest neighbor-based techniques:** In this method, the distance or similarity between two data instances is measured. Although normal data instances occur in dense neighborhoods, anomalies occur far from their closest neighbors [10].
- **Pattern matching**: Online learning is used in the pattern matching method to generate a traffic profile for each network. The profiles are used for anomaly detection. However, pattern matching may need to build traffic profiles for new networks, which results in a time-consuming process [11].

4. MISUSE BASED DETECTION SYSTEM

A misuse-based IDS, which is a programmed method, compares the user's activities with predefined signatures to find malicious traffic. Although this method is very accurate in detecting known attacks, it is not able to detect unknown attacks. Based on our knowledge, different misuse detection methods are shown in Figure 12.2 with a number of examples [26].



Misuse detection methods

- **State modeling:** Intrusions are defined as states in this method. State-transition [7] and Petri-net [8] are two important methods of state modeling. These methods are different in the type of states that make up the intrusions. The state transition analysis technique (STAT) is an example of the state-transition method and is able to detect new attacks [9]. In this technique, the sequences of actions performed by attackers are specified to describe computer penetrations. There are different types of STAT techniques: host based (USTAT) [7], network based (NETSTAT) [3], and distributed multi-host (NSTAT). Analysis tools use a system's audit trail or network traffic to obtain required information.
- **Expert system:** There are a set of rules in an expert system in which rules describe attack behavior. The expert system can be used to consider the security state of the system. Next generation intrusion detection expert systems (NIDES) [2], EMERALD [3], and MIDAS [4] are expert system-based IDS techniques.
- **String matching:** Numbers of misuse- or signature-based IDSs use this technique, which is a substring matching of characters in texts. If there is a change in an attack signature, this method is unable to detect the attack. NSM is a model that is proposed [5].
- **Simple rule-based:** Expert knowledge about attacks can be modeled by the rule-based method. NADIR [6, 7] and ASAX [48] are two methods that use a rule-based method.

5. INTRUSION DETECTION IN HIGH-SPEED NETWORKS

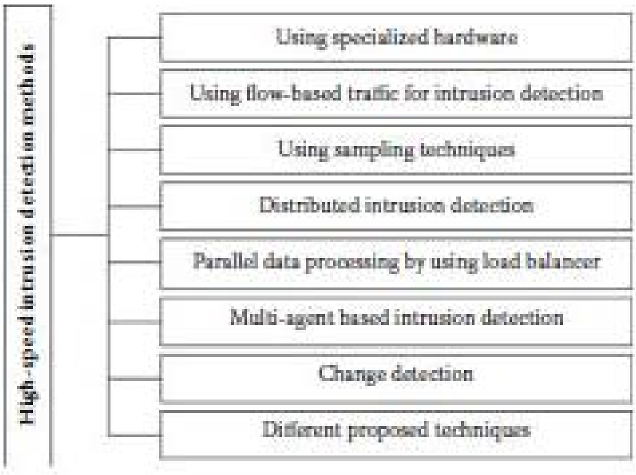
Using Specialized Hardware

Hardware-based intrusion detection is a scalable method as it is able to inspect packets in high-speed networks. Most of the hardware-based NIDSs have been proposed to improve deep packet inspection (DPI), using some specialized hardware, such as a field-programming gate array (FPGA), an application-specific integrated circuit (ASIC), and ternary content addressable memories (TCAM) [9–10]. For example, a regular expression method is proposed using TCAM [11]. It is shown that this method is useful for throughput up to 18.6 Gbps.

Using Flow-Based Traffic for Intrusion Detection

A flow record is defined as a group of packets with a number of common properties, which pass a monitoring point in a certain time interval. Flow-based traffic contains only packet headers, and hence it reduces data. Flow-based IDS (FIDS) cannot detect attacks related to packet payload.

Therefore, it is not a replacement for packet-based IDS. The FIDS can detect attacks such as denial of service (DoS), scans, worms, and botnets. DoS attacks caused by payload contents cannot be detected by FIDS [9]. For example, FIDS is not beneficial for a ping of death attack because this attack does not make a change in flow frequency and traffic volume. Several studies consider flow-based intrusion detection [7, 8]. HiFIND is an online DoS–resilient flow-level intrusion detection system for a high-speed network [2]. Sketches are used in this technique to detect anomalies. A sketch is a one-dimensional hash table used for storing information. It records traffic for specific keys. HiFIND uses 2-D sketches, which hashes a set of flow-derived fields for each dimension. This method is employed to detect SYN flooding and port scans.


High-speed intrusion detection methods
Using specialized hardware
Using flow-based traffic for intrusion detection
Using sampling techniques
Distributed intrusion detection
Parallel data processing by using load balancer
Multi-agent based intrusion detection
Change detection
Different proposed techniques

Intrusion detection methods in high-speed networks

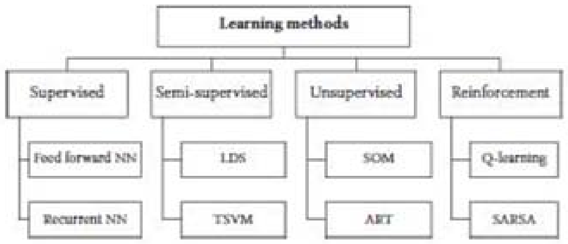
Using Sampling Techniques

Sampling data is a method used for anomaly detection [5–6] and change detection, for example, DoS attack detection. Cisco NetFlow [6] is a sampling technique to decrease the heavy load on router CPU in high-speed networks. However, sampling has negative impacts on the statistical characteristics of traffic and hence on the performance of intrusion detection. There are different typesof sampling [9, 6, 6]:

- **Packet sampling:** There are two types of packet sampling methods: systematic and random. In systematic packet sampling, a time interval, or a sequence of packet arrival, is chosen to select a packet. In random packet sampling, the probability distribution function is used as a basis of sampling.
- **Flow sampling:** This method is more accurate than packet sampling. Random probability is used in random flow sampling to select flows.
- **Smart sampling:** This method is proposed to control the size of sampling data. Both the smart and the sample-and-hold sampling are flow-sampling methods proposed to reduce required memory.
- **Sample-and-hold:** The smart and sample-and-hold sampling methods try to provide precise traffic estimation for larger flows.
- **Adaptive packet sampling:** In order to have an accurate traffic statistic, this method identifies the current traffic load to adjust the sampling rate.
- **Selective flow sampling:** Although sampling techniques address the scalability problems, they affect anomaly detection efficiency. Selective flow sampling provides an appropriate balance between the performance and the amount of sampled information. Using small flows in selective flow sampling helps this method to improve the performance with less selected flow [5, 6].

6. ARTIFICIAL INTELLIGENCE IN INTRUSION DETECTION

Artificial intelligence is a well-studied approach for intrusion detection. Flexibility, learning ability, and adaptability are characteristics of AI–based IDS. ANNs, fuzzy systems, artificial immune systems, GA, and swarm intelligence have been widely applied to intrusion detection [4, 5]. A brief overview of the applications of AI methods to intrusion detection is discussed in the following sections. AI–based, high-speed intrusion detection will also be considered.



Learning methods

Supervised Learning

The supervised method learns to map inputs to outputs using the correct values defined by the supervisor. A feed forward neural network (FFNN) and recurrent neural network (RNN) are two important methods that use supervised learning. A multi-layered feed-forward (MLFF) NN and radial basis function (RBF) are two examples of FFNNs. The calculation of the distance between inputs and the centers of hidden neurons is the basis of RBF classification. In comparison with the MLFF back-propagation (BP), the RBF is better for large data because it is faster [4].

Unsupervised Learning

There is no supervisor in unsupervised learning, and it is trained using unlabeled data only. Unsupervised learning is similar to a statistical clustering, in which they identify various groups of inputs using their similarity [2]. The self-organizing maps (SOM) and the adaptive resonance theory (ART) are two examples of unsupervised learning. The SOM is an important neural network method used for the anomaly and misuse detection [4]. However, the performance of ART and SOM– based intrusion detection are compared in [9], which shows the higher detection performance of ART on both offline and online data.

Semi-Supervised Learning

The semi-supervised learning method combines the supervised and the unsupervised learning capabilities. In this method, often some unlabeled data is provided in a data set besides labeled data. When labeling is expensive, this method can be useful even with less labeled data [8]. The semi-supervised method can act as a supervised or an unsupervised learning according to the availability of labeled data. This method is employed for intrusion detection in several studies [8, 8, 9]. A semi-supervised learning-based method can be trained by both labeled and unlabeled data and has more accurate prediction. Low density separation (LDS) [9] and transductive SVM (TSVM) [9] are examples.

Reinforcement Learning

Reinforcement learning (RL) is a combination of the supervised and unsupervised learning. In reinforcement learning, there is an agent acting upon the environment. The state of the environment changes with the agent's actions, and the environment, in return, gives feedback for those actions. The feedback is either a reward or a punishment;

therefore, RL is trained by rewards and punishments. For instance, when the system acts well, the teacher gives a reward. On the other hand, while getting a punishment, the system should improve itself. Thus, due to the existence of the feedback from the environment, RL is a form of supervised learning, but it is known as weak supervised learning because RL never presents the correct input/output pairs.

7. RESULTS AND DISCUSSION

Describes common artificial intelligence–based IDSs as well as a number of high-speed intelligent IDSs. Traditional IDSs were evaluated using IPv4 traffic. Migration to IPv6 provides new security challenges. Although, IPv6 is, in general, more secure than IPv4 due to the mandated IPsec, the transition process introduces a number of security problems. These security threats were considered in this chapter. Intrusion detection in other networks is as important as in wired networks. Artificial intelligence–based IDSs are applicable to many other environments as well. In this chapter, the application of artificial intelligence in grid computing, cloud computing, and wireless networks was also considered.

VIII. CONCLUSION

Our dependence on the Internet is increasing day by day. Attacks and malicious activities are very common in this cyber world. An intrusion detection system is an essential mechanism to protect computers and networks from attacks. While the Internet service providers can offer high bandwidth, detecting intrusions in the high-speed networks is a challenge for researchers. High-speed IDSs are required to handle this huge amount of traffic. Different high-speed intrusion detection techniques were described in this chapter. Use of artificial intelligence techniques has numbers of advantages due to their learning ability and adaptability. An artificial intelligence–based IDS is adaptable to environmental changes and is trained to detect even unknown attacks. The intelligent IDS may also be able to work in high-speed networks.

IX. REFERENCES

[1.] Pathan, A.-S. K., Lee, H.-W., and Hong, C. S. Security in wireless sensor networks: Issues and challenges. In Proc. Of IEEE ICACT '06, Vol. II, Phoenix Park, Korea, (February 20-22, 2006)

[2.] Cam, H., Ozdemir, S., Muthuavinashiappan, D., and Nair, P. Energy efficient security protocol for wireless sensor networks. In IEEE 58th VTC 2003 Fall, 2003, 5 (October 6–9, 2003)

[3.] Cam, H., Ozdemir, S., Nair, P., Muthuavinashiappan, D., and Sanli, H. O. Energy-efficient secure pattern based data aggregation for wireless sensor networks. Com. Commun., 29, I.4, (2006)

[4.] Yin, C., Huang, S., Su, P., and Gao, C. Secure routing for large-scale wireless sensor networks. In Proceedings of IEEE ICCT 2003, 2 (April 9–11, 2003)

[5.] Hass, Z. J. Design methodologies for adaptive and multimedia networks. IEEE Communications Magazine, 39(11), (November 2001)

[6.] Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wire. Commun., 1(4) (2002)

[7.] M. Almgren and E. Jonsson. Tuning an ids - learning the security officer's preferences. In 11th Nordic Workshop on Secure IT Systems - Nordsec 06, 2006.

[8.] P. P. Bonissone. Soft computing: the convergence of emerging reasoning technologies. Soft Computing— A Fusion of Foundations, Methodologies and Applications, 1(1):6–18, 1997.

[9.] J. Cannady. Artificial neural networks for misuse detection. In Proceedings of the 1998 National Information

[10.] J. Frank. Artificial intelligence and intrusion detection: Current and future directions. In Proceedings of the 17th National Computer Security Conference, Baltimore, MD, 1994.

[11.] A. H. M. Lichodzijewski, P.; Nur Zincir-Heywood. Host-based intrusion detection using self-organizing maps. In Proceedings of the 2002 International Joint Conference on Neural Networks, 2002.

[12.] M. Moradi and M. Zulkernine. A neural network based system for intrusion detection and classification of attacks. In 2004 IEEE International Conference on Advances in Intelligent Systems.

[13.] A. Mounji. Rule-Based Distributed Intrusion Detection. PhD thesis, University of Namur, 1997.

[14.] V. G. C. F.M. Valtorta. Paid: A probabilistic agent-based intrusion detection system. In Computers & Security, pages 529–545, 2005.