Expert Systems with Applications 41 (2014) 1690–1700

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A novel hybrid intrusion detection method integrating anomaly detection with misuse detection

Gisung Kim [a], Seungmin Lee [b,*], Sehun Kim [a]

[a] Institute for IT Convergence, KAIST, Guseong-dong, Yuseong-gu, Daejeon 305-701, South Korea
[b] Future Research Creative Laboratory, ETRI 218 Gajeong-ro, Yuseong-gu, Daejeon 305-700, South Korea

ARTICLE INFO

Keywords:
Hybrid intrusion detection
One-class SVM
Anomaly detection
Decision tree

ABSTRACT

In this paper, a new hybrid intrusion detection method that hierarchically integrates a misuse detection model and an anomaly detection model in a decomposition structure is proposed. First, a misuse detection model is built based on the C4.5 decision tree algorithm and then the normal training data is decomposed into smaller subsets using the model. Next, multiple one-class SVM models are created for the decomposed subsets. As a result, each anomaly detection model does not only use the known attack information indirectly, but also builds the profiles of normal behavior very precisely. The proposed hybrid intrusion detection method was evaluated by conducting experiments with the NSL-KDD data set, which is a modified version of well-known KDD Cup 99 data set. The experimental results demonstrate that the proposed method is better than the conventional methods in terms of the detection rate for both unknown and known attacks while it maintains a low false positive rate. In addition, the proposed method significantly reduces the high time complexity of the training and testing processes. Experimentally, the training and testing time of the anomaly detection model is shown to be only 50% and 60%, respectively, of the time required for the conventional models.

1. Introduction

An intrusion detection system (IDS) has been developed that is capable of detecting all types of network attacks in the available environments. The IDS is placed inside the network that it protects and it collects network packets promiscuously in the same manner as a network sniffer. The IDS detects malicious network activities by analyzing the collected packets, alarms to system administrator, and blocks attack connections in order to prevent further damage from attacks. It also connects with the firewall as a fundamental technology for network security.

Generally, intrusion detection algorithms are categorized into two methods: misuse detection and anomaly detection (Depren, Topallar, Anarim, & Ciliz, 2005). Misuse detection algorithms detect attacks based on the known attack signatures. They are effective in detecting known attacks with low errors. However, they cannot detect newly created attacks that do not have similar properties to the known attacks. In contrast, anomaly detection algorithms analyze normal traffic and profile normal traffic patterns. The anomaly detection method is based on the hypothesis that the attacker behavior differs to that of a normal user. They classify traffic as an attack if the characteristics of the traffic are far from those of normal traffic patterns. Anomaly detection algorithms can be useful for new attack patterns, but they are not as effective as misuse detection models in the detection rate for known attacks and false positive rates, which is a ratio of misclassified normal traffic.

In order to resolve the disadvantages of these two conventional intrusion detection methods, hybrid intrusion detection methods that combine the misuse detection method and the anomaly detection method have also been proposed (Depren et al., 2005). Because none of the misuse and anomaly detection methods are better than any other, a hybrid intrusion detection system uses both the misuse detection method and anomaly detection method. The detection performance of the hybrid intrusion detection system depends on the combination of these two different detection methods. Most hybrid detection systems independently train a misuse detection model and an anomaly detection model, and then simply aggregate the results of the detection models. For example, hybrid intrusion detection systems regard a traffic connection as an attack if at least one of the two models classifies the traffic connection as an attack. In this case, the detection rate will be improved but the IDS will still have a high false positive rate. In contrast, if the hybrid method regards a traffic connection as an attack only if both models classify the connection as an attack, false alarms will be reduced but it may overlook many attack connections.

* Corresponding author. Tel.: +82 42 860 1775; fax: +82 42 860 6504.
E-mail addresses: todtom@etri.re.kr, brightdad@gmail.com (S. Lee).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.066

In this research, a new hybrid intrusion detection method is proposed that hierarchically integrates a misuse detection model and an anomaly detection model, rather than just combining the their results as in previous hybrid methods (Depren et al., 2005; Zhang & Zulkernine, 2006). In the proposed method, the anomaly detection model can indirectly use the known attack information throughout the integration in order to enhance its ability to build profiles of normal behavior. Generally, only the misuse detection method uses the known attack information to build a classifier and the anomaly detection method builds a classifier only based on normal traffic information. The proposed hybrid method also follows this general principle, but it is proposed that the normal training data is decomposed into disjoint subsets using the misuse detection model and then an anomaly detection model is built for each disjoint normal training data subset.

The entire normal data set has various types of normal connections, so an anomaly detection model cannot profile it precisely, which leads performance deterioration (Song, Takakura, Okabe, & Kwon, 2009). In the proposed hybrid intrusion detection process, each area for the decomposed normal data set does not have known attacks and includes less variety of connection patterns than the entire normal data set. An anomaly detection model for each normal training data subset can profile more innocent and concentrated data so that this decomposition method can improve the profiling performances of the normal traffic behaviors.

In this paper, the C4.5 decision tree (DT) is used to create the misuse detection model and the one-class support vector machine (1-class SVM) is used to create multiple anomaly detection models. In order to implement the concept described above, the DT model is first trained based on a training data set consisting of normal traffic and known attack traffic information, and then a 1-class SVM model is trained for each normal training data subset decomposed by the DT model. The proposed hybrid intrusion detection method was evaluated by conducting experiments using the NSL-KDD data set, which is a modified version of the famous KDD Cup 99 data set (Tavallaee et al., 2009). The experiment results demonstrate that the proposed method is better in terms of detection rates for unknown and known attacks than the conventional methods that independently train the DT model and 1-class SVM model without the proposed decomposition technique.

In addition to improving the detection rates, through training the anomaly detection model in each decomposed data set, the proposed method can reduce the high time complexity of the training and testing processes. Time consumption is the cost for updating the detection model. In particular, the testing time should be minimized in order to reduce the overhead of the detection algorithm in order to operate the detection model in real time. As the training data set is decomposed into smaller subsets, the training and testing times are significantly reduced. The experiments in this research demonstrate that the training and testing times of the anomaly detection model are only 50% and 60%, respectively, of that of the conventional models.

The remainder of this paper is organized as follows. In Section 2, the existing hybrid intrusion detection methods are reviewed. The detailed process of the proposed method is presented and the properties of the method are discussed in Section 3. In Section 4, the performance of the proposed method is evaluated in terms of the detection rate and time required for training and testing the anomaly detection model. The study concludes in Section 5 with a summary of the research undertaken and plans for future research.

2. Related works

There has been much research on hybrid intrusion detection methods in attempts to overcome the limits of the anomaly detection and misuse detection methods. The research has used three different methods to combine the anomaly detection model and misuse detection model: anomaly detection followed by misuse detection, parallel use of anomaly detection and misuse detection, and misuse detection followed by anomaly detection.

Barbara, Couto, Jajodia, Popyack, and Wu (2001) proposed the Audit Data Analysis and Mining (ADAM) method where the anomaly detection is followed by the misuse detection. ADAM uses a combination of association rule mining and a classification method to detect attacks. First, the anomaly detection model that uses association rule mining locates suspicious traffic connections and passes the connections to the misuse detection model. Then, the misuse detection model classifies the suspicious connections as normal (false alarm of the anomaly detection model), known attacks connections, and unknown attack connections. It is unusual that ADAM uses a misuse detection method to detect unknown attack connections. In the ADAM method, connections that cannot be confidently classified as normal or known attacks are classified as unknown attacks. In order to use the anomaly detection followed by the misuse detection, the anomaly detection model should have a high detection rate and the misuse detection model should remove the false alarms of the anomaly detection model by distinguishing the normal and unknown attacks. However, most misuse detection methods are not suitable for reducing false alarms.

Parallel hybrid approaches use an anomaly detection model and a misuse detection model in parallel. Depren et al. (2005) suggested an intelligent hybrid intrusion detection system that consists of an anomaly detection model, a misuse detection model, and a decision support system. They modeled the anomaly detection model with a self-organization map (SOM) and the misuse detection model with a decision tree. Each model is trained independently, and then the decision support system simply combines the classification results of both models.

Anderson, Frivold, and Valdes (1995) developed the Next Generation Intrusion Detection Expert System (NIDES) and it uses a rule-based analysis model and statistical analysis model. The rule-based misuse detection model employs expert rules to define the known attacks and the statistical analysis anomaly detection model detects connections that depart from the established patterns of normal behavior. In this method, the detection rate of known attacks and unknown attacks is enhanced. However, the high false positive property of the anomaly detection method remains because it regards an incoming connection as an attack if any detection model classifies the connection as an attack. Also, there is a detection overhead problem because every connection should be checked using both the anomaly detection model and the misuse detection model, and this can cause increases in the detection overhead.

Zhang and Zulkernine (2006) and Hwang, Chen, and Qin (2007) used the misuse detection method followed by the anomaly detection method to design a hybrid intrusion detection system. Because the misuse detection model can detect known attacks with a low false positive rate and can operate faster than the anomaly detection model, the misuse detection model is used first to detect the known attacks and then the anomaly detection model is only applied to the remaining uncertain connections. The anomaly detection model detects outliers that depart from the normal data patterns and classifies them as unknown attacks. Zhang and Zulkernine (2006) also developed a weighted signature generation scheme that extracts the attack signatures from the attack connections detected by the anomaly detection model and adds those signatures to the misuse detection model for fast and accurate operation. However, as with the parallel hybrid method, the anomaly detection model and the misuse detection model are trained independently, which nevertheless results in a high false positive rate.