Feature Selection in Hybrid Intrusion Detection System

[1]Khin Khattar Myint, [2]Nang Saing Moon Kham
[1, 2] *University of Computer Studies, Yangon.*
[1]*khinnkhattarrmyint@gmail.com*, [2]*moonkhamucsy@gmail.com*

Abstract

With the enormous growth of computer networks, network security is gaining increasing importance. Therefore, the role of Intrusion Detection Systems (IDSs) is becoming more important. There are many techniques available for intrusion detection. In this paper, a hybrid intrusion detection method that integrates an anomaly detection model and a misuse detection model using the one-class SVM and C4.5 decision tree is proposed. Despite the inherent potential of hybrid detection, there are many issues that highly affect the performance of the hybrid systems such as detection rate, false positive rate, memory overhead, time overhead and so on. Moreover, most of the existing IDSs use all of the features available in the dataset to detect the attack while some of the features are redundant. It is time-consuming and may degrade the performance of IDSs. Therefore, rough set theory is used in the proposed hybrid intrusion detection system to select the most significant features. The experimentation is implemented in ROSETTA and WEKA tools using NSL-KDD dataset.

1. Introduction

Network Security is very important as networks are exposed to an increasing number of security threats. An intrusion detection system is a software tool that monitors network or computer system for malicious activities .Intrusions are defined as attempts to compromise the confidentiality, integrity or availability of a computer or network. They are caused by attackers accessing a system from the Internet, by authorized users of the systems who attempt to gain additional privileges for which they are not authorized or by authorized users who misuse the privileges given to them [1].

In the intrusion detection field two different approaches can be observed: misuse detection and anomaly detection. The main idea behind misuse detection is to represent attacks in a form of a pattern or a signature in such a way that even variations of these attacks can be detected. Based on these signatures, this approach detects attacks through a large set of rules describing every known attack. The main disadvantage of the signature based approach is its difficulty for detecting unknown attacks. The main goal of the anomaly detection approach is to build a statistical model for describing normal traffic. Then, any deviation from this model can be considered an anomaly, and recognized as an attack. Notice that when this approach is used, it is theoretically possible to detect unknown attacks, although in some cases, this approach can lead to a high false attack rate. This ability to detect unknown attacks has been the cause of the increasing interest in developing new techniques to build models based on normal traffic behavior in the past years [2]. 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. 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.

The evaluation of the proposed IDS is used with NSL-KDD data set which is an improved version of KDD99 data set. Each NSL-KDD connection record contains 41 features and is labeled as either normal or attack type [3].Though NSL-KDD dataset is enhanced version of the KDD99 dataset, there are ambiguities in some records of the testing dataset. That is some records have same value for all the 41 features, however they are labeled to different classes (one as normal and the other as attack).In this work, we aim to filter out redundant, worthless information, which leads significantly to reduce the amount of computer resources, both memory and CPU time, required to detect attacks.

The rest of this paper is organized as follow: Section 2 describes some related work for this research based on feature reduction and hybrid intrusion detection system. Section 3 describes the NSL-KDD dataset. Section 4 describes the theoretical background. Section 5 describes the detail analysis of the proposed hybrid intrusion detection system. Section 6 describes experiments and results followed by a conclusion in Section 7.

2. Related Work

There has been much research on hybrid intrusion detection system by researchers using different techniques and approaches to overcome the limitations of both misuse detection and anomaly detection.

Depren [7] proposed 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.

Goel [8] also proposed a novel hybrid model for misuse and anomaly detection. C4.5 based binary decision trees are used for misuse and Classification Based Association based classifier is used for anomaly detection. Model's performance is evaluated on DARPA KDD CUP99 benchmark data set.

Lakhina and Joseph [6] proposed a hybrid algorithm using principal component analysis and neural network algorithm. The PCA transform used to reduce the feature and trained neural network is used to identify the any kinds of new attacks. Test and comparison are done on NSL-KDD dataset.

3. NSL-KDD data set

KDD Cup 99 is the mostly widely used data set for the evaluation of the intrusion detection system. But it has the huge number of redundant records. In KDD train and test set, 78% and 75% of the records are duplicated. This large amount of redundant records in the train set will cause learning algorithms to be biased towards the most frequent records. The existence of these repeated records in the test set will cause the evaluation results to be biased by the methods which have better detection rates on the frequent records. NSL-KDD is the enhanced version of the KDD CUP 99 intrusion detection data set. It is generated by removing redundant instances in both the training and testing data of the KDD 99 dataset. NSL-KDD consists of the same features as KDD 99. The KDD99 data set consists of 41 features and one class attribute. The data set

contains 24 attack types. All these attacks fall into four main categories.

1. **Denial of Service (DOS):** In this type of attacks an attacker makes some computing or memory resources too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. Examples are Apache2, Back, Land, Mailbomb, SYN Flood, Ping of death, Process table, Smurf, Teardrop.

2. **Remote to User (R2L):** In this type of attacks an attacker who does not have an account on a remote machine sends packets to that machine over a network and exploits some vulnerability to gain local access as a user of that machine. Examples are Dictionary, Ftp_write, Guest, Imap, Named, Phf, Sendmail, and Xlock.

3. **User to Root (U2R):** In this type of attacks an attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. Examples are Eject, Loadmodule, Ps, Xterm, Perl, Fdformat.

4. **Probing:** In this type of attacks an attacker scans a network of computers to gather information or find known vulnerabilities. An attacker with a map of machines and services that available on a network can use this information to look for exploits. Examples are Ipsweep, Mscan, Saint, Satan, Nmap.

Table 1. Data Distribution in NSL-KDD data set

NSL-KDD dataset	Normal	Attack	Total
Training	67343	58630	125973
Testing	9711	12833	22544

4. Theoretical Background

In this section, Rough set theory, decision tree and one-class support vector machine algorithms that are required in order to build the misuse detection model and anomaly detection model are briefly introduced. Then, the integration of these models is explained.

4.1. Rough Set Theory

Feature selection aims to reduce the number of irrelevant and redundant features of the intrusion data set to improve the classification detection accuracy. Moreover, Effective features selection is very important for constructing a high performance IDS. Rough set theory (RST) is a useful mathematical tool to deal with imprecise and insufficient knowledge, find hidden patterns in data, and reduce dataset size [4]. Also, it is used for evaluation of significance of data and easy interpretation of results. RST contributes immensely to the concept of reducts. Rough set theory deals with inconsistencies, uncertainty and incompleteness by imposing an upper and a lower approximation to set membership. It has been successfully used as a selection tool to discover data dependencies and find out all possible feature subsets and remove superfluous information. Hence, a reduct is a minimal subset of attributes with the same capability of objects classification as the whole set of attributes. The following definitions as given in [4] show the reduct derivation for Rough set theory.

Definition 1:

Knowledge is represented by means of a table called an Information System given by $S = <U,A,V,f>$; where $U = \{x_1, x_2, \ldots, x_n\}$is a finite set of objects of the universe (n is the number of objects); A is a non-empty finite set of features, $A=\{a_1, a_2, \ldots, a_m\}$; $V = \cup_{a \in A} V_a$ and V_a is a domain of feature a; $f: U \times A \rightarrow A$ is a total function such that $f(x,a) \in V_a$ for each $a \in A$, $x \in U$. If the features in A can be divided into condition set C and decision feature set D; i.e. $A=C \cup D$ *and* $C \cap D = \Phi$. The information system A is called decision system or decision table.

Definition 2:

Every $B \subseteq A$ yields an equivalence relation up to indiscernibility, $IND_A(B) \subseteq (U \times U)$, given by: $IND_A(B) = \{(x,x') : \forall a \in B\ a(x) = (x')\}$a reduct of A is the least $B \subseteq A$ that is equivalent to A up to indiscernibility. i.e., $IND_A(B) = IND_A(A)$.

4.2. C4.5 decision tree

Decision tree learning is one of the most widely used techniques for classification. Its classification accuracy is competitive with other learning methods, and it is very efficient. It is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label. It operates in a divide and conquer manner, which partitions the training data set based on its attributes until the stopping conditions are satisfied [5]. A decision tree node has its corresponding data set; this specifies the attribute to best divide the dataset into its classes. Each node has several edges that specify possible values or value ranges of the selected attributes on the node. The primary issue of the decision tree algorithm is to locate the attribute that best divides the data into their corresponding classes.C4.5 builds decision trees from training data sets using the concepts of information entropy. It is based on the highest gain of each attribute. The gain is calculated using the formula:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} f_s(A_i) \times Entropy(S_{A_i}) \quad (1)$$

$$Entropy(S) = -\sum_{j=1}^{m} f_s(j) \times log_2 f_s(j) \quad (2)$$

4.3. One-class Support Vector Machine

One-class SVMs attempt to learn a decision boundary that achieves the maximum separation between the points and the origin. Interestingly this was the initial idea from which traditional super-vised SVMs emerged. The idea was hindered by the inability to learn non-linear decision boundaries as well as the inability to account for outliers. Both of these problems were solved by the introduction of kernels and the incorporation of soft margins. A one-class SVM uses an implicit transformation function $\emptyset(.)$ defined by the kernel to project the data into a higher dimensional space. The algorithm then learns the decision boundary (a hyperplane) that separates the majority of the data from the origin [5]. Only a small fraction of data points are allowed to lie on the other side of the decision boundary. Those data points are considered as outliers.

5. Proposed Hybrid Intrusion Detection System

The proposed hybrid intrusion detection system has three parts such as feature selection, misuse-detection and anomaly-detection. Figure 1 shows the proposed hybrid intrusion detection system architecture.

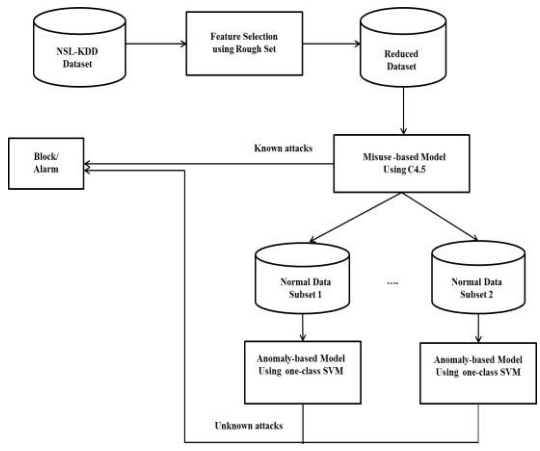

Figure 1. Proposed Hybrid IDS

In the training phase, rough set theory based feature selection method is used to select the most significant feature in order to reduce the time and space complexity of classifier. Then, a misuse detection model is built using the decision tree algorithm. The resulted reduced data set from the feature selection phase is used as an input to the misuse-based model. The misuse-based model can not only detect the known attack but also divide the normal data connections into subsets using the divide and conquer strategy. A one-class SVM is used to build anomaly detection models for each normal training data set. The reason for dividing the normal training data set into normal training data subsets is that the data patterns of each decomposed subset are less complex than those of the whole data set and training the small data subsets can reduce the training time of the classifier.

In the testing phase, the misuse-based model checks whether the data is known attack or normal. If the data is known attack, the system may block the data connection. In the anomaly-based model, a one-class SVM classifier is trained with normal data only and so it check the data connection whether it is normal or not. If the data is not normal, the system classify that the data is unknown attack and block the data connection.

6. Experimental Result

The experiment is run with a 3.40GHz processor and 4GB RAM running Windows XP. For Feature selection, rough set operations were done in ROSETTA. It is a toolkit used for data analysis using Rough Set theory. By applying the reduction algorithm defined by Rough Set Theory, using ROSETTA for the 5000 samples. The resulting reduced features are only six given in table 2. This gives 84% reduction in input data. After the reduct is produced, the experiment for misuse detection is carried out in WEKA tool. J48 in WEKA tool is used as a C4.5 decision tree for misuse detection. Table 3 shows the optimal experimental results for running the misuse detection system before feature selection and after feature selection.

Table 2. Features induced by Rough Set Theory

No.	Network Data Features	Description
1	flag	Status of the connection –Normal or Error
2	src_bytes	Number of data bytes transferred from source to destination in single connection.
3	count	Number of connections to the same destination host as the current connection in the past two seconds.
4	srv_count	Number of connections to the same service (port number) as the current connection in the past two seconds.
5	dst_host_srv_count	Number of connections having the same port number.
6	dst_host_diff_srv_rate	The percentage of connections that were to different services, among the connections aggregated in dst_host_count.

Table 3. Experimental results before and after feature selection

	Before feature selection (41 features)	After feature selection (6 features)
Number of Leaves	180	78
Size of the tree	221	128
Time taken to build model	0.42 seconds	0.05 seconds
Correctly Classified Instances	98.7 %	97.2 %

The results in table 3 show that the proposed model gives better and robust representation of data as it was able to reduce the number of attributes resulting in a 84% data reduction and time overhead is reduced eight times . The number of nodes before feature selection is 221 and that number of nodes after the feature selection is 128. Therefore our proposed method can effectively reduce the memory overhead of C4.5 method. Even the accuracy of the reduction

is a little less than the accuracy of the whole data set; the proposed system can reduce the building time of the model and also reduce the memory overhead.

There are various kinds of feature selection methods for intrusion detection system. This paper also compares the proposed system using rough set theory (RST) with the principal component analysis (PCA). Figure 2 shows the accuracy comparison between the two methods.

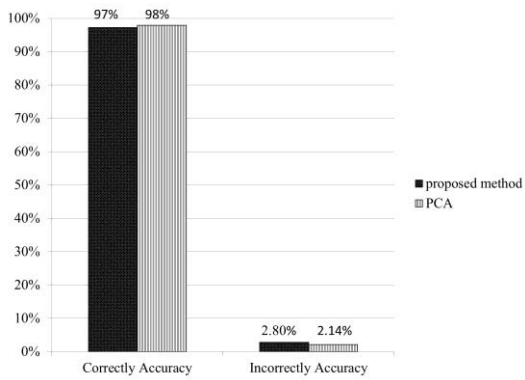

Figure 2. Accuracy of two methods

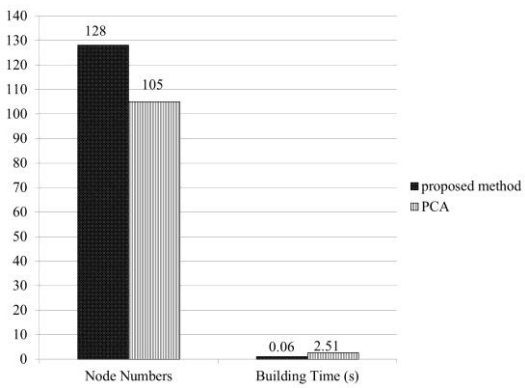

Figure 3. Nodes numbers and building time of two methods

Figure 3 shows the comparison between the two methods, RST and PCA. The number of nodes in our proposed method using RST is almost similar to PCA method. Because of we use rough set theory to reduce the dimension of the data, our proposed method just need 0.06 seconds to build the intrusion detection model while the system using the PCA took 2.51 seconds. Therefore, after comparing the training time of two methods we can conclude that rough set theory can be effective to reduce the time overhead of building Intrusion Detection Model.

In this experiment, false alarm rate comparison can't describe.

7. Conclusion

The proposed hybrid intrusion detection system combines both misuse and anomaly intrusion detection systems in a decomposition structure. Rough set theory is used to select out feature to cope the problem of dimensionality reduction as a feature extraction. Then, C4.5 is used as a misuse detection model not only to detect the known attack but also to decompose the normal training data into smaller subsets. One-class support vector machine is used to create an anomaly detection model in each decomposed region. It is expected to get a high detection rate and low false alarm rate in comparison with the existing individual and hybrid IDSs. It is also expected to reduce both memory and CPU time for building the hybrid IDS. In our research, we found that the computation time is reduced eight times than the model without using feature extraction. The ongoing research will be implementing the anomaly detection system using one-class support vector machine and combining it with the above misuse detection system.

References

[1] K.Scarfone, P.Mell, "Guide to Intrusion Detection and Prevention Systems", National Institute of Standards and Technology, Gaithersburg, 2007.

[2] V.Golmah, "An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM,"

International Journal of Database Theory and Application, 2014.
[3] S.Revathi, Dr.A.Malathi, "A Detailed Analysis on NSL-KDD Data set Using Various Machine Learning Techniques for Intrusion Detection", International Journal of Engineering Research and Technology, 2013.
[4] S.Rissino, G.L.Torres, "Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications", Brazil, 2009.
[5] X.Wu, V.Kumar, "The Top Ten Algorithms in Data Mining",Taylor & Francis Group, Minnesota, 2009.
[6] S.Lakhina, S.Joseph, B.Verma, "Feature Reduction using Principal Component Analysis for Effective Anomaly-Based Intrusion Detection on NSL-KDD", International Journal of Engineering Science and Technology, 2010.
[7] O.Depren, M.Topallar, E.Anarim, M.K.Ciliz, "An Intelligent Intrusion Detection System for Anomaly and Misuse Detection in Computer Network", Expert Systems with Applications, 2005.
[8] R.Goel, A.Sardana, R.C.Joshi, "Parallel Misuse and Anomaly Detection Model", International Journal of Network Security, 2012.