Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 1658 – 1664

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Modeling Optimisation and Computing

A Rule Based Approach for Attribute Selection and Intrusion Detection in Wireless Sensor Networks

K.Anand[a], S.Ganapathy[b], K.Kulothungan[b], P.Yogesh[b], A.Kannan[b]

[a]*Department of Electricals and Computers, KTH Technical University, Swedan.*
[b]*Deprtment of Information Science and Technology, Anna University, Chennai-600 025, Tamil nadu, India.*

Abstract

In this paper, we propose a new rule based attribute selection algorithm for removing the redundant attributes which are used in decision making on intrusions in wireless sensor networks. This work focuses mainly on finding important attributes to find Denial of Service attacks. In addition, we used an enhanced MSVM classification algorithm that was developed by extending the existing MSVM algorithm. The experimental results show that the proposed methods provide high detection rates and reduce false alarm rate. This system has been tested using KDD'99 Cup data set.

Keywords: Intrusion Detection System (IDS); Attribute Selection; Information Gain Ratio (IGR);

1. Introduction

Wireless Sensor Networks (WSN) can be defined as dynamic multi-hop networks that consist of a collection of nodes. The nodes employ multi-hop information transfer without requiring an existing infrastructure. Although WSN are characterized by great flexibility and are employed in a broad range of applications, they also present much inherent vulnerability that increases their security risks. Due to their energy dependent nature, WSN demand efficient and effective security mechanisms in order to be

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: author@institute.xxx .

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.06.202

safeguarded. In addition, as network based computer systems play increasingly vital role in modern society, security of network systems has become more important than ever before. It is difficult to keep the system safe by static safeguards like firewall. As an active defense technology, an Intrusion Detection System (IDS) attempts to identify existing attack patterns and recognize new intrusion, and hence becomes an indispensable component in security architecture. Intrusion detection and prevention techniques can be used as a first line of defense in this network in order to reduce the possible intrusions but undoubtedly, it cannot eliminate them. Intrusion detection using classification algorithms can help the administrators to effectively discriminate "normal" nodes from "abnormal" nodes due to their abnormal behavior and thus can detect possible intrusions effectively. Therefore, intrusion detection, serving as a second line of defense, is an indispensable part of reliable communication in WSNs. However, the existing intrusion detection methods, including misuse detection and anomaly detection [2], are generally incapable of adapting detection systems to the change of circumstances that is common in WSNs due to the change in energy of nodes. Moreover, traditional Intrusion Detection methods can only detect known intrusions since they classify instances by what they have learned. However, the necessity to build adaptive IDS with self learning abilities has become a hot spot in security field.

In this paper, a rule based new attribute selection algorithm for detecting the intruders in WSNs and various types of attacks happen in WSNs. We have focused mainly on the effective detection of DoS attacks by using rule based EMSVM algorithm, since DoS attacks are more serious than other attacks with respect to energy consumption.

The subsequent sections are organized as follows: Section 2 presents a general survey in field of misuse detection, anomaly detection and data reduction. Section 3, describes the architecture of the system proposed in this paper for implementing a new attribute selection algorithm. Section 4 gives a brief explanation about the proposed algorithm and implementation.. Section 5 discusses the results and its possible implications. Conclusions and plans for future works are given in section 6.

2. Literature Survey

A number of approaches to feature selection have been proposed in the literature Ref [2,3,11]. A simplified decision table method was proposed by Chuanjian Yang et al [11] for improving the efficiency of attribute reduction and to obtain a minimal attribute reduction which can be used for effective decision making. Discernibility matrix and discernibility function were proposed by Skowron [2]. The method of discernibility matrix is widely applied to the attribute reduction.

A new fitness function was introduced by Zhangyan Xu et al, and it was proved that the optimization of candidate attribute reduction helps in achieving effective attribute reduction. Moreover, it can be used to delete the candidate attributes for reduction of attributes using crossover and mutation process [3]. Wei Wang et. al [4] introduced four different methods of attribute normalization to preprocess the data for anomaly intrusion detection. In addition, three methods namely K- NN, PCA and SVM were employed on the normalized data for comparison of the decision results.

In addition, there are many classification algorithms that are found in the literature [5,6]. For example, an algorithm called Tree structured Multiclass SVM was proposed by Snehal A.Mulay et. al [6] for classifying the data effectively. Their paper proposed a decision tree based algorithm to construct multiclass IDS which was focusing on improving the training time, testing time and accuracy of IDS. However, the detection rate is not sufficient in the current internet scenario. Dewan Md. Farid et al [5] proposed a new learning approach for network intrusion detection that performs data reduction by

selecting important subset of attributes. Their system reduces the false positives compared to other classifiers like ID3 algorithm and Naïve Bayesian algorithm. A novel architecture of Support Vector Machine classifiers utilizing binary decision tree (SVM-BDT) for solving the multiclass problems was provided by Gjorgji Madzarov et. al [7]. This architecture provides techniques for achieving better classification accuracy. In Reference [8], Redundant and irrelevant attributes of intrusion detection data set were removed in a complex intrusion detection model to increase the detection accuracy.

3. System Architecture

The Architecture of the intrusion detection system proposed in this paper is shown in figure 1. This intrusion detection system consists of six modules namely User Interface Module, Information Gain Ratio Module, Attribute Selection Module, Sub divide Module, Classification Module and Decision Making module.

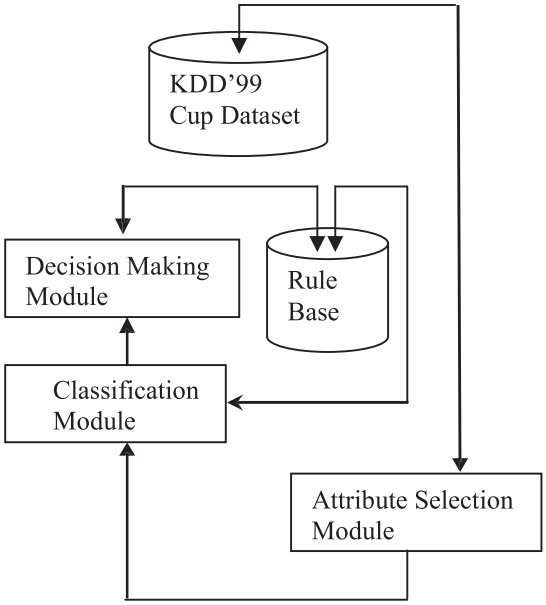

Fig. 1. System Architecture

The attribute selection module collects the networks data from the KDD cup data set. It also selects the necessary attributes based on the information gain ratio value. The clustering module is used to split the original data set in to sub data sets. The Classification module is used to classify the data by using the EMSVM Classifier [1]. The decision making module decides whether the particular node sends normal packets or malicious packets.

4. Implementation Details

In this paper, we propose a new attribute selection technique for effective preprocessing of the KDD cup data set. The rule based decision manager is responsible for providing effective decisions in clustering fault tolerance and security. Therefore, this component assists the administrator module for decision

making by applying suitable rules which are present in the rule base. The rule based decision manager uses forward chaining and backward chaining control flows. The Rule Manager applies special techniques for rule firing and rule matching. The rule based decision manager is responsible for maintaining integrity and security of the system in coordination with the administrator module. The rule based decision manager is capable of storing rules in the rule base and is responsible for manipulating them.

The rule base consists of security rules, cluster information rules and fault tolerance rules. All these rules are stored in the form of IF .. THEN rules. The rules are also created by the decision tree and support vector machine algorithms through training. Such rules are used for effective testing. The rule base consists of special rules for identifying the behavior of intruders acquired through the training phase. Rules can be added, deleted and modified by the rule based decision manager.

The data set consists of 41 features. However, all these features are not necessary to detect the DoS attacks. Therefore, we propose a new attribute selection algorithm that selects only the valuable and important attributes from the data set using projection. Moreover, Data cleaning, Data integration and Data transformation are carried out for performing effective preprocessing.

The attribute selection algorithm uses Information Gain Ratio for attribute selection

$$Info(D) = - [freq(Cj , D) / |D|] \log2 [freq (Cj, D) / |D|] \quad (1)$$

$$Info(T) = [|T_i| / |T|] * info(T_i) \quad (2)$$

$$IGR(A_i) = [Info (D) – Info (T) / Info (D) + Info (T)] * 100 \quad (3)$$

Attribute Selection and Classification Algorithm

Steps of the algorithm:

1. Compute the Information Gain Ratio for each attribute $A_i \in D$ using equation 3.
2. Choose an attribute Ai from D with the maximum Information gain value.
3. Split the training data D into K sub-data sets using rules to get $\{D_1, D_2, \ldots. D_k\}$ depending on the attribute values of A_i.
4. Classify the examples of each sub-dataset Di using rule based EMSVM.
5. If any example of sub-dataset Di is misclassified then
   Calculate the Information Gain Ratio of corresponding attributes.
   Classify the sub/sub-sub data set examples of their prior and EMSVM.
   Continue this process until all the examples of sub | sub- sub-datasets are correctly classified.
6. Preserve all the prior conditional probabilities for each sub-dataset D_i or sub-sub-dataset D_{ij} for future classification of unseen examples.

5. Experimental Results

5.1. Training and Test data

The dataset used in the experiment was taken from the Third International Knowledge Discovery and Data Mining Tools Competition (KDD Cup 99) [10]. Each connection record is described by 41 attributes. The list of attributes consists of both continuous-type and discrete type variables, with statistical distributions varying drastically from each other, which makes the intrusion detection a very challenging task.

5.2. Experimental Results

Table 1 shows the overall results after 3 stages of classification, excluding the 4th stage where the known attacks are separated into specific attacks. For old attacks we have some methodology to defend, so by using any of this method, we can handle and we can have high detection rate. That is not the case for new attacks.

Obviously, using this classifier has achieved the highest detection rates for old DoS, PROBE and R2L attacks. As to the new attacks, it also has the highest detection rates for DoS and U2R attacks. However, a weakness observed is that it did not perform so well for detecting R2L attacks. This is the trade off for the high detection rate of other attack types. Furthermore the low number of instances for R2L connections in the training and testing data makes the detection rate of R2L negligible compared to other attacks.

Table 1 Detection Rates (%) Using 41 attributes

Classes	ID3 Algorithm	Proposed Algorithm
Normal	99.71	99.89
Probe	98.22	99.83
DoS	99.63	99.84
U2R	86.11	99.63
R2L	97.79	99.41

Table 2 Detection Rates (%) Using 19 attributes

Classes	ID3 Algorithm	Proposed Algorithm
Normal	99.63	99.74
Probe	97.85	99.29
DoS	99.51	99.83
U2R	49.21	99.26
R2L	92.75	99.33

Table 3 False Positive (%) Using 41 attributes

Classes	ID3 Algorithm	Proposed Algorithm
Normal	0.10	0.06
Probe	0.55	0.40
DoS	0.04	0.03
U2R	0.14	0.11
R2L	10.03	7.80

Table 4 False Positive (%) Using 19 attributes

Classes	ID3 Algorithm	Proposed Algorithm
Normal	0.06	0.04
Probe	0.51	0.26
DoS	0.04	0.02
U2R	0.12	0.09
R2L	7.34	6.12

Above the Tables were shown the performance of the proposed algorithm. From the above results that significant attribute selection improves the performance of detection model.

6. Conclusion

In this paper, we proposed a new rule based attribute selection technique and used the rule based EMSVM learning method for performing network intrusion detection. This attribute selection algorithm performs data reduction by selecting important subset of attributes. The experimental results show that the proposed algorithm achieved high detection accuracy and reduced the false alarm rate with respect to DoS attacks in wireless sensor networks. The main advantage of this algorithm is that it helps to reduce the power consumption in WSNs by reducing the number of packets transmitted. Further works in this direction could be the proposal of a new secure routing algorithm that can use this intrusion detection model for providing secure communication.

References

[1] Ganapathy S, Yogesh P, Kannan A, "An Intelligent Intrusion Detection System for Mobile Ad-Hoc Networks Using Classification Techniques", Springer-CCIS, Vol. No. 148, pp.117-122 , 2011.

[2] Skowron A.Rauszer C, "The discenibility Matrices and Functions in Information Systems, In: Slowinski. Intelligent Decision Support hand Book of applications and advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publisher, 1991, pp. 331-362.

[3] Zhangyan Xu, Dongyuan Gu, Bo Yang," Attribute reduction algorithm based on genetic algorithm", Second International Conference on Intelligent Computation Technology and Automation, IEEE, 2009.

[4] Wei wang, Xiangliang Zhang, sylvain Gombault and Svein J. Knapskog, " Attribute Normalization in Network Intrusion Detection", 10th International Symposium on Pervasive Systems, Algorithms, and Networks, IEEE,2009.

[5] Dewan Md. Farid, Jerome Dormant, Nouria Harbi, Nguyen Huu Hoa and Mohammad Zahidur Rahman, "Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification", International Conference on Computer Systems Engineering, Thailand, 2009.

[6] Snehal A.Mulay, P.R. Devale, G.V. Garje, "Intrusion Detection System using Support Vector Machine and Decision Tree", International Journal of Computer Applications, Volume3-No.3,pp.0975-8887, June 2010.

[7] Gjorgji Madzarov, Dejan Gjorgjevikj and Ivan Chorbev,"A Multiclass SVM Classifier Utilizing Binary Decision Tree", Informatica33, pp. 233-241, 2009.

[8] Lee W K and S J Stolfo, " A Data Mining framework for Building Intrusion Detection Models", In Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA: IEEE Computer Society Press, pp. 120-132,1999.

[9] Denning D E, "An Intrusion Detection Model",. IEEE Transactions on Software Engineering, Vol. 51, no. 8, pp. 12-26, Aug. 2003.

[10] KDD Cup 1999 Data, Information and Computer Science, University of California, Irvine. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[11] Chuanjian Yang, Hao GE, Guangshum Yao, Lisheng Ma,"Quick complete Attribute Reduction Algorithm", Sixth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, pp.576-580, 2010.