ANOMALY DETECTION IN WIRELESS LAN USING ROUGH SET THEORY COMBINED CLASSIFIER MODEL

P. Kavitha[1] and M. Usha[2]
[1]Department of Information Technology, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India
[2]Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamilnadu, India
E-Mail: paranjothi.kavitha@gmail.com

ABSTRACT

In this paper, we suggest to exploit the framework for detecting anomalies in Wireless Local Area Networks (WLAN) using Rough Set Theory (RST). With the expansion of wireless network there is a challenge to compete with the intruders who can easily break into the system. So it becomes a necessity to device systems or algorithms that can not only detect intrusion but can also improve the detection rate. In this paper we propose an intrusion detection system that uses rough set theory for feature selection, which is extraction of relevant attributes from the entire set of attributes called minimal set. The extracted features are used by Naïve Bayesian classifier model to learn and test respectively. The simulation results with Kyoto2006+ data set demonstrate that our proposed method achieves the increasing performance for intrusion detection.

Key words: intrusion detection, wireless network, naïve bayesian classifier, feature selection, rough set theory.

1. INTRODUCTION

Intrusion detection is one of core technologies of computer security. The goal of intrusion detection is identification of malicious activity in a stream of monitored data which can be network traffic, operating system events or log entries. An Intrusion Detection System (IDS) is a hardware or software system that monitoring event streams for evidence of attacks. A majority of current IDS follow a signature-based approach in which, similar to virus scanners, events are detected that match specific predefined patterns known as signatures. The main limitation of these signature-based IDS is their failure to identify novel attacks, and sometimes even minor variations of known patterns. Machine learning is a valuable tool for intrusion detection that offers a major opportunity to improve quality of IDS [1][2]. As a broad subfield of artificial intelligence, machine learning is concerned with the design and development of algorithms and techniques that allow computers to "learn". At a general level, there are two types of learning: inductive, and deductive. Inductive machine learning methods extract rules and patterns out of massive datasets. Feature selection is useful to reduce dimensionality of training set; it also improves the speed of data manipulation and improves the classification rate by reducing the influence of noise. The goal of feature selection is to find a feature subset maximizing performance criterion, such as accuracy of classification. Not only that, selecting important features from input data lead to a simplification of the problem, faster and more accurate detection rates. Thus selecting important features is an important problem in intrusion detection. The algorithm Reduct set computation is used to extract the features

2. RELATED WORK

Denning [10] first proposed an intrusion detection model in 1987. Yuk Ying Chung, Noorhaniza Wahid have proposed a hybrid network intrusion detection system using simplified swarm optimization. IDS-Rough Set Theory is proposed to select the most relevant features that can represent the pattern of the network traffic to improve the classification accuracy of KDDCUP99 data set. Feng Jiang, Yuefei Sui, Cungen Cao, suggested information entropy (IE) model for the measurement of uncertainty in rough sets. They demonstrate the application of this model for the study of outlier detection and an algorithm to find outliers is also given. The effectiveness of IE-based method for outlier detection has been demonstrated on two publicly available data sets. Shi-Jinn Horng et.al proposed an Support Vector Machine (SVM)-based intrusion detection system, which combines a hierarchical clustering algorithm. The hierarchical clustering algorithm was used to provide a high quality, abstracted and reduced dataset for the SVM training, instead of the original enormous dataset. Thus, the system could greatly shorten the training time, and also achieves better detection performance in the resultant SVM classifier for KDDCUP99 data set. Feng Jiang et.al proposed a novel definition of outliers in information systems of rough set theory – sequence-based outliers and developed an algorithm to find such outliers in rough set theory. The effectiveness of sequence-based method for outlier detection was demonstrated on two publicly available databases. They introduced traditional distance-based outlier detection through set theory and discussed the definitions of distance metrics for distance-based outlier detection in rough set theory. Nandita Sengupta et.al designed an online intrusion detection system using roughest theory and Q-learning algorithm. The objective of the work is to achieve maximum classification accuracy while detecting intrusions by classifying NSL-KDD network traffic data either ‘normal’ or ‘anomaly’. Since RST processes discrete data only, by applying cut operation attributes in training data are discretized [16].

Using indiscernibility concept of RST, reduced attribute sets [15], called reducts are obtained and among the reducts a single reduct is chosen which provides highest classification accuracy of 98%.

3. WIRELESS ATTACKS

Wireless intrusions belong to four broad categories, namely: (1)Passive attacks (2) Active attacks (3) Man-in-the- middle (MITM) attack (4) Jamming attacks.[2][4].

A Passive attack (e.g., war driving) occurs when someone listens to (or eavesdrops) on network traffic. Armed with a wireless network adaptor that supports promiscuous mode, the eavesdropper can capture network traffic for analysis using easily available tools, such as Network Monitor. War driving is the act of searching unsecured Wi-Fi networks by a person with a Wi-Fi equipped computer. As long as somebody is sniffing the network packets and trying to discover some useful information from gathered packets (e.g., WEP key used in the network or available open ports), we classify these activities as passive attacks. Once this information is discovered through passive attacks, then hackers can launch some active attacks [19].

Active attacks launched by hackers who access the network to launch these active attacks include unauthorized access, Denial of Service (DoS) and Flooding attacks like (SYNchronized) SYN Flood attacks and (User Datagram Protocol) UDP Flood attacks. DoS attack attempts to engage a host of computer resources so that these resources are not available to other users. DoS is an attack in which the attacker keeps the resource too busy or too full to handle other legitimate requests, and thus, it denies legitimate users access to a machine. The attacker's IP address is fake and destination IP address is the server victim's address. Receiving so many packets from attacker prevents victim from accepting new legitimate requests and may crash the victim server [3].

Man-In-The-Middle (MITM) attack entails placing a rogue AP (Access Point) within range of wireless stations. If the attacker knows the SSID in use by the network (which is easily discoverable) and the rogue AP has enough strength, wireless users have no way of knowing that they are connecting to an unauthorized AP. Because of their undetectable nature, the only defense against rogue APs is vigilance through frequent site surveys using tools such as Netstumbler and AiroPeek, and physical security.

Jamming is a special kind of DoS attack specific to wireless networks. Jamming occurs when spurious RF (Radio Frequency) frequencies interfere with the operation of the wireless network. Intentional and malicious jamming occurs when an attacker analyzes the spectrum being used by wireless networks and then transmits a powerful signal to interfere with communication on the discovered frequencies. Fortunately, this kind of attack is not very common because of the expense of acquiring hardware capable of launching jamming attacks and it leads to a lot of time and effort being expended merely to disable communication.

4. ROUGH SET THEORY

Rough set theory is a new mathematical approach to data analysis and data mining [15]. After 15 year of pursuing rough set theory and its application the theory has reached a certain degree of maturity. In recent years we witnessed a rapid grow of interest in rough set theory and its application, worldwide. The connection of rough set theory and many other theories has been clarified. Particularly interesting is the relationship between fuzzy set theory and Dempster-Shafer theory of evidence. The concepts of rough set and fuzzy set are different since they refer to various aspects of imprecision, whereas the connection with theory of evidence is more substantial. Besides, rough set theory is related to discriminant analysis, Boolean reasoning methods and others. The relationship between rough set theory and decision analysis is presented. Various real life-applications of rough set theory have shown its usefulness in many domains. Very promising new areas of application of the rough set concept seems to emerge in the near future. They include rough control, rough data bases, rough information retrieval, rough neural network and others. Rough set Theory can be approached as an extension of the classical set theory, for use when representing incomplete knowledge. Concepts are represented by lower and upper approximations, according to which rough set methodology focuses on approximate representation of knowledge derivable from data.

a. Indiscernibility and set approximation

Let U be the universe of the discourse and A be the finite and nonempty set of attributes, then S=(U,A) is an information system .Let B a subset of A. With every subset of attributes B ⊆ A, an equivalence relation IB on U can be easily associated.IB = {(p,q) ∈ U x U/ ∀a ∈ B, a(p)=a(q)}

Where IB is called B-indiscernibility relation.

If (p,q) ∈ IB, then objects p and q are indiscernible from each other by attributes B.The equivalence classes of the partition induced by the B-indiscernibility relation are denoted by [p]B. These are also known as granules.We can approximate any subset X of U using only the information contained in B by constructing the lower and upper approximations of X. The sets $\{ p \in U: [p]_B \subseteq X \}$ and$\{ p \in U: [p]_B \cap X \neq \phi\}$, where $[p]_B$ denotes the equivalence class of the object p∈ U relative to I_B, are called the B-lower and B-upper approximation of X in S and respectively, denoted by $\underline{B}(x)$, $\overline{B}(X)$.The objects in B(X) can be certainly classified as members of X on the basis of knowledge in B, while objects in $\overline{B}(X)$can only be classified as possible members of X on the basis of B.

b. Knowledge reduction and the core of rough set

Knowledge reduction is a significant concept of rough set used for data analysis. It is divided into attribute reduction and attributes value reduction. Under the premise that the indiscernibility relation, classification, and decision-making ability of decision table information system is unchanged, attribute reduction refers to eliminate irrelevant or unimportant redundant attributes, and get the minimum subset of condition attribute. Attribute reduction is based on the importance to reduce. The attributes of an information system in the decision table are not all important. Some attributes may be redundant, can delete after simplified; only part of the conditions property must be retained. For an attribute subset P ⊆ A, if there exists the relation Q= P- r, satisfying Q ⊂ P and making IND (Q) = IND (P) tenable, while Q is the best subset, then Q is called the reduction of P, denoted with red (P). The intersection of all reduction attribute sets in P is called the core of P, denoted with core (P).

5. APPLICATION OF NAIVE BAYESIAN CLASSIFIER FOR CLASSIFICATION

Rapid progress in data collection techniques and data storage has enabled an accumulation of huge amounts of experimental, observational and operational data. As a result, massive data sets containing gigantic bulk of information can be found almost everywhere. A well known example of massive data set is the data set containing the observed information about network traffic in wireless LAN. Tremendous need to quickly and correctly analyze or manipulate such enormous data sets facilitates the development of data mining. Data mining is the research aimed at discovery of various types of knowledge from the large data set [5]. It is an integral part of more general process of knowledge discovery in data bases. Data mining is utilizing a special case of Bayesian networks namely naïve bayes for performing effective classification. In a data mining context, a classification is a task of assigning objects to their relevant categories [24]. The incentive for performing classification of data is to attain a comprehensive understanding of differences and similarities between the objects in different classes.

6. EXPERIMENTAL ANALYSIS

The experiment used a Intel(R) Core i3 2310 M CPU @ 2.10GHz computer with 3GB RAM, and implemented on a Windows 7 Ultimate 32-bit Operating System. There has been quite some research done in the area of anomaly detection in networks. Most of this research has been conducted after the KDD99Cup data set that could be used for anomaly detection and knowledge discovery. But this data set cannot be used to test anomaly detection algorithms for wireless networks. This is because there exist a vast difference between the fixed network where current intrusion detection research are taking place and the wireless networks. This makes it very difficult to apply intrusion detection techniques developed for one environment to another. Compared with wired networks where traffic monitoring is usually done at switches, routers and gateways, the wireless environment does not have such traffic concentration points where the IDS can collect audit data for the entire network. This could be done at the access points but once again range is a problem.

a) Data collection

We experimented with two data sets. First, we collected three weeks of traces in the Wireless Lab with 10 access point and 100 clients using Wireshark. A module was then written to clean this data and extract the statistics that we desired. At this point it became essential to figure out what features should be stored in each feature vector (or data item). Every feature (listed in Table. 1) that we choose should provide some valuable information and reflect some important property of the network which an administrator would otherwise like to monitor.

b) Data pre-processing and feature extraction

A few simple and basic data pre-processing techniques like sampling and filtering are applied for the sake of easy and smooth operation of the experiments. Rough sets can't deal with continuous attributes directly, so we must discrete the continuous attributes.

We extracted randomly some non-repetitive connection records, and store into training database. Samples with known and unknown attacks are merged together so as to render two types of data only viz. Normal and Anomaly data. The features identified are listed below:

Table-1. Features identified for anomaly detection using Wireless LAN traffic.

S. No.	Features	Description
1	SrcMac	The MAC address of the source device.
2	DstMac	The MAC address of the destination.
3	NumFrames	The number of frames sent from the source to destination.
4	AvgFrmSize	The average size of frames in bytes, sent from source to destination.
5	NumDeauths	The ratio of number of De-authentication frames sent to the number of frames sent.
6	NumDisassocs	The ratio of number of Disassociation frames sent to the number of frames sent.
7	NumRetries	The ratio of number of retransmitted frames sent to the number of frames sent.
8	NumCRCErrs	The ratio of number of error frames sent to the number of frames sent.

9	AvgSignal	The average signal strength of frames sent from source to destination.
10	AvgNoise	The average noise (in terms of percentage of signal) in frames sent from source to destination.
11	Strength	Signal strength
12	SeqNo	Sequence Number
13	SourcePkts	Number of packets a station sourced
14	SourceErrPkts	Number of errors in source packets
15	DestErrPkts	Number of errors in destination packets

We might be also interested in reducing some of the condition attributes, i.e. to know whether all conditions are necessary to make decisions specified in a table. To this end we will employ the notion of a reduct (of condition attributes). By a reduct we understand a minimal subset of condition attributes which preserves the consistency factor. That means that in view of the network traffic data the most important condition attributes causing anomaly are identified and they cannot be eliminated from our considerations, whereas other attributes play a minor role and can be mutually exchanged as factors causing anomaly.

Algorithm for finding minimal set of the data set is described below:

```
Algorithm :Reduct Set Computation
Input:
            Information system A=(U,A)
Output:
            Set RED_A(A) of all reducts of A
Method:
            Compute indiscernibility matrix M(A)=C_ij
            Reduce M using absorption laws
            d-number of non-empty fields of reduced   M
            Build a families of sets R_0,R_1,……R_d in the following
            way:
            begin
            Compute R_1
            i=1
            while i > 0 do
            begin
if stop then return
if R_i= ϕ then
                        begin
                        i=i-1
                        continue
                        end
Remove from family R_i, the first element
Compute R_i+1
i=i+1
            If i=r then
            begin
                        Remove from R_d redundant elements
                        RED_A(A)= RED_A(A) U R_d
                        i=i-1
            end
end
end
```

Reduct set computation yields only 10 (minimal set) attributes out of 15. The features from 11 to 15 are not In order to verify the ability of classification of data sets after reduction; we have adopted the following test methods. Experiment was conducted using Weka (Waikato Environment for Knowledge Analysis) software. Weka is a collection of machine learning algorithms for data mining tasks [9]. The algorithms can either be applied directly to a dataset or called from our own Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. WEKA consists of Explorer, Experimenter, Knowledge flow, Simple Command Line Interface, Java interface. To analyse the Wireless network traffic two experiments are conducted. One is with real time wireless network and second is with Kyoto 2006+ dataset [11] with number of records are 25192 and 22544 respectively.

7. PERFORMANCE MEASURES

The effectiveness of IDS is evaluated by its ability to make correct predictions. According to the real nature of a given event compared to the prediction from the IDS, four possible outcomes are shown in Table-2, known as the confusion matrix. True negatives as well as true positives correspond to a correct operation of the IDS; that is, events are successfully labeled as normal and anomaly, respectively. False positives refer to normal events being predicted as attacks; false negatives are attack events incorrectly predicted as normal events. Based on the above confusion matrix, a numerical evaluation can apply the following measures to quantify the performance of IDSs:

True negative rate (TNR): $\frac{TN}{TN+FP}$ also known as specificity.

True positive rate (TPR): $\frac{TP}{TP+FN}$ also known as detection rate (DR)

The most popular performance metrics are detection rate (DR) together with false alarm rate (FAR). An IDS should have a high DR and a low FAR. Other commonly used combinations include precision and recall, or sensitivity and specificity.

Table-2. Results of Wireless Network traffic data for 25192 instances.

Wireless Network traffic Data set	**Correctly Classified Instances**	**%**	**Incorrectly Classified Instances**	**%**	**Training time to build a model(in Seconds)**
Classification with original attributes	22570	89.6	2622	10.4	0.61
Classification with Reduced data set	22500	89.3	2692	10.7	0.32

Table-3. Results of Kyoto 2006+data set for 22544 instances.

Kyoto 2006+ Data set	No.of Correctly Classified Instances	%	Incorrectly Classified Instances	%	Training time to build a model (in Seconds)
Classification with original attributes	18205	80.75	4339	19.25	0.56
Classification with Reduced data set	18105	80.31	4439	19.69	0.3

When comparing the results of two data sets the percentage of correctly classified instances of reduced data set are approximately nearer to the correctly classified instances of original data set. Due to the reduction in attributes, the training time of the classifier is improved.

8. CONCLUSIONS

In this paper, we have proposed an intrusion detection method using Naïve Bayesian classifier with Rough set theory. Rough set theory based feature reduction of network traffic data handles minimal set of attributes and vagueness that reduces complexity of the IDS. We have compared the performance of the Naïve bayes classifier with two different data sets. One is with real time wireless network traffic data set collected using Wireshark and the second one is using Kyoto2006+ data set. The comparison of the classification result in the two cases is demonstrated which yields the best average accuracy. This knowledge is used to generate alert message to network administrator to monitor the wireless network properly from various security threat.

REFERENCES

[1] P. Kavitha., M. Usha. Classifier Selection Model for Network Intrusion Detection using Data Minin,CiiT. International Journal of Data Mining and Knowledge Engineering.Vol 3,No.12 , 2011.

[2] P. Kavitha., Usha.M., Detecting Anomalies in WLAN using Discrimination Algorithm. 4th International Conference on Computing, Communication and Networking Technologies - ICCCNT 2013,July 2103.

[3] P. Kavitha., M. Usha. Anomaly Based Intrusion Detection in WLAN Using Discrimination Algorithm Combined With Naïve Baysian Classifier. AJTIT Vol. 61,March 2014.

[4] Tarek S. Sobh., Wired. and wireless intrusion detection system : Classification, good characteristics and state-of-the-art Elsevier. 2005

[5] Ertoz L., Eilertson E., Lazarevic A., Tan P., Srivastava J., Kumar V., Dokas P. The MINDS – Minnesota Intrusion Detection System, Next Generation Data Mining, MIT Press. 2004.

[6] Khoshgoftaar T.M., Nath S.V., Zhong S., Seliya N. Intrusion detection in wireless networks using clustering techniques with expert analysis, in proc. of the ICMLA 2005:Fourth International Conference on Machine Learning andApplications, pp. 120-125, 2005.

[7] Zhong S., Khoshgoftaar T.M., Nath S.V., A clustering approach to wireless network intrusion detection, in proc. Of the International Conference on Tools with Artificial Intelligence. ICTAI 2005, pp. 190-196, 2005.

[8] Zhong T. M. Khoshgoftaar. and N. Seliya. Clustering based network intrusion detection, International Journal of Reliability, Quality, and Safety Engineering,2007.

[9] M. Xue. and C. Zhu. Applied Research on Data Mining Algorithm in Network Intrusion Detection, International Joint Conference on Artificial Intelligence, 2009.

[10] An intrusion detection Model, Dorothy E. Denning, IEEE 1986.

[11] Traffic Data from Kyoto University's Honeypots http://www.takakura.com/Kyoto data.

[12] Zhang Wenxiu,Wu Weizhi.An Introduction to survey for Studies of Rough Set Theory, Fuzzy Systems and Mathematics[J], 2000, 14(4), PP:1-12

[13] J. Liu., F. Min., S. Liao. and W. Zhu. Test cost constraint attribute reduction through a genetic approach, Journal of Information & Computational Science 10: 3 (2013) 839–849.

[14] S. K. Pati. and A. K. Das. Constructing minimal spanning tree based on rough set theory for gene selection, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.4, No.1, January 2013.

[15] Z. Pawlak., Rough sets. The Tarragona University seminar on Formal Languages and Rough Sets in August 2003.

[16] J. Qian., D. Q. Miao., Z. H. Zhang. and W. Li. Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation, Elsevier, International Journal of Approximate Reasoning (2011) 212- 230.

[17] R Xu., J Li., F Zhang. R Levy., W Lee. Agent-based cooperative Anomaly detection for Wireless Ad hoc Networks, - Parallel and Distributed, 2006

[18] Anomaly Detection Approaches for Communication Networks MarinaThottan, Guanglei Liu, Chuanyi Ji.2009

[19] M. Balazinska. and P. Castro. Characterizing mobility and network usage in a corporate wireless local-area network, In The 1st Int. Conf. Mobile Systems, Applications, and Services, 2003

[20] T. Velmurugan. and T. Santhanam.Computational Complexity between k-Means and k- Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points, Journal of Computer Science, 6 (3): 363-368, 2010.

[21] Yang Su., Gwo-Jong Yu. Chun-Yuen Lin A real-time network intrusion detection system for large scale attacks based on an incremental mining approach ,Elsevier 2008

[22] Y. Li and L. Guo, An Active Learning Based on TCM-KNN Algorithm for Supervised Network Intrusion, Computer and Securtiy, 26: 459-467, 2007

[23] R. Luigi., T.E. Anderson.and N. McKeown. Traffic Classification using Clustering Algorithms. ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Pisa, Italy, ACM Press. pp. 281-286, Sep. 11-15,2011.

[24] B.A. Nahla., B. Salem. and E. Zied. Naïve Bayes vs Decision Trees in Intrusion Detection Systems, ACM Symposium on Applied Computing,Nicosia, Cyprus, 2004.

[25] Xiang C., M. Y. Chong. and H. L. Zhu. Design of Multiple-Level Tree Classifiers for Intrusion Detection System, IEEE Conference on Cybernetics and Intellligent Systems (CCIS 2004), Singapore, pp: 873-878, 2004.

[26] http://www.cs.waikato.ac.nz/~ml/weka/index.html.