International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 17665-17674

http://www.ripublication.com

Intrusion Detection and Classification of Attacks Using Fuzzy Logic

Sabna Sulaiman
M.Tech Computer Science and Engineering
Dept. of Computer Science
SRM University, Kattankulathur 603203
Chennai, Tamil Nadu, India
sabnasulaiman123@gmail.com
Dr. M.Pushpalatha
Professor, Dept. of Computer Science
SRM University, Kattankulathur 603203
Chennai, Tamil Nadu, India
pushpalatha.m@ktr.srmuniv.ac.in

Abstract

Attacks have increased with the increased use of wireless technologies; therefore it became difficult to provide security to wireless networks than wired networks. The attacks that are already there in the database can be easily detected and classified than the unknown and less frequently occurring attacks. Hence, more importance should be given to the need for efficient intrusion detection system. In our proposed work, we are implementing binary classifier to classify the data in to normal or attack. Types of attacks can be clearly classified using an ensemble of fuzzy classifiers with feature selection and multi boosting techniques. Use of Multi Boosting helps us to built decision tree which helps to classify the type of attack, in an efficient way which will take care of the instances which are misclassified by previous classifiers. For the unknown attacks, features of it are extracted for the accurate binary classification and also it improves the detection of attacks that occur less frequently in the training data.

Keywords: Intrusion detection, Classification of attacks, Fuzzy logic, Multi Boosting

Introduction

Attacks are increasing as the increase in the use of wireless technologies. Therefore, the need for effective intrusion detection system is necessary[1]. Intrusion detection

system monitors the network or system for the malicious activities. There are two types of network names as network based intrusion detection systems (NIDS) and host based intrusion detection systems (HIDS). To monitor the traffic to and from all devices on the network NIDS are placed on the points within the network[6] and It also performs an analysis of traffic passing on the entire subnet whereas, HIDS run on individual hosts or devices on the network and it monitors the inbound and outbound packets from the device inorder to alert the user or administrator if malicious activity is detected. All intrusion detection system (IDS) use one of the two techniques namely Statistical anomaly-based IDS or Signature based IDS

- Statistical anomaly based IDS: An IDS which is anomaly based that will monitor network traffic and compare it against an established baseline. The baseline will identify normal and attack packets. The bandwidth that is generally used protocols that are used, ports and devices that connect are examined. It also gives alarm to the administrator or user when anomalous traffic that is different than the baseline is detected.
- Signature based IDS: A signature based IDS will monitor packets on the network and comparison is done with the database of signatures or attributes from known malicious threats [2]. The issue is that there will be a delay between a new threat being discovered and the signature for detecting that threat being applied to our IDS. During that time delay our IDS would be unable to detect the new threat.

Intrusion detection systems are evolving products. However, this area continues to change as new research influences the design of products. The main problems that affect IDS are taken and tried to eliminate it as much as possible in our work such as false alarms, Traffic and attribute selection

The first limitation that said above is the intrusion detection system may give errors like false positive and false negative alarms. An efficient intrusion detection system should be free from false alarms.

- False positive: A false positive occurs when normal attack is mistakenly classified as malicious and treated accordingly.
- False negative: This is the case where an intrusion detection system does not generate an alert when an intrusion is actually taking place.

Another limitation is of the traffic of data in the network. We need to capture the packets from the network for the intrusion detection analysis. Packet capturing from heavy traffic is very difficult and it will be easy if we are able to select the types of network that we need to analyze like time based which will limit the analysis up to certain time, host based which will helps to analyze the data flow for certain systems.

The third limitation is about attribute selection which is very important. The captured packet will contain large number of attributes which are the combinations of both informative and uninformative attributes. Uninformative and unwanted attributes are termed as noise. We should select the only relevant and needed attribute as per our need which will increase the accuracy and computation power of the system.

Therefore we have to consider the above limitations before the system is being designed. In our work we are building an effective Intrusion detection system by

considering the selection of informative attribute, reducing false alarms and the detection and classification of less frequently occurred and unknown attacks.

In our work we are taking KDD CUP 99 dataset for evaluating the performance of six classification algorithm like Six algorithms namely Decision tree, Ripper, Neural Network, Navie Bayes, K-NN,SMO with respect to three parameters, correctly classified, incorrectly classified and accuracy rate. The best performing algorithm is used finally for the proposed intrusion detection system. The attacks that are not in the database are difficult to detect and classify therefore we are using a machine leaning technique that extract the feature of the network using feature extraction method in fuzzy logic is used for feature extraction[3]. For the feature extraction phase four types of network are taken namely intrinsic network, Host based network, Time based network and content based network. After extracting the features, feature classification is done to normal traffic or attacked traffic using binary classifier. Finally Multi Boosting technique is used to design the decision tree which will help to observe the sequence and frequency of the attacks. The search for certain sequence of attack type is also enabled in our proposed method.

The following section is like in session II Background, the related works of intrusion detection and its drawbacks are discussed. In session III Proposed method, the proposed intrusion detection system is explained in detail. In session IV Evaluation, the performance analysis of the six algorithms using appropriate data mining tool is explained and also the results of intrusion detection using feature extraction method of fuzzy logic are also explained. In session V, the conclusion and future work are explained.

Background

Snort is an intrusion detection system [4] which detects the attacks. Its performance will be depending up on the traffic that is there in the network. Snort is capable of performing analysis of real-time traffic and packet logging on IP networks. It can perform protocol analysis, pattern matching and can be used to detect a variety of attacks. EMERALD [5] is a software-based solution implementation that utilizes lightweight sensors distributed over a network for real-time detection of malicious attack. EMERALD sensors monitor activity both on host servers and network traffic streams. It has the limitations that sensors generally have like energy constrain, capacity. IDSs in the Open Market that employ data mining techniques have already been released as parts of commercial security packages.In the work that is done by D.Barbara, J.Couto, S.Jajodia, and N.Wu , ADAM (Audit Data Analysis and Mining) [6] is an intrusion detection system for intrusion detection with the help of data mining techniques. It has the database of training data sets which are said to be normal data set that is free of attacks. ADAM has several useful capabilities such as classifying an item as normal event, known and unknown attack, even it will matching of audit trial data and the rules it give rise to are done.

C4.5 decision tree is used to select features that have been described to improve Naives Bayesian learning in the work done by Chotirat “Ann” Ratanamahatana and Dimitrios Gunopulos[7]. C4.5 decision tree selects features for Naïve Bayesian

classifier to use because it does not use redundant attributes in the construction of decision trees and they cannot generate different splits of training data. C4.5 relay on the most appropriate feature it can find from the large set of data which leads to higher accuracy in both Bayesian classifier and C4.5 decision tree.

The work by Mrutyunjaya Panda and Manas Ranjan Patra [8] Proposed a framework of network based intrusion detection system based on Naïve Bayes Classification Algorithm. Here the proposed work is compared with neural network approach and it is found that the Naïve bayes frame work gives higher detection rate with less computation time and low cost factor but it generates more false positive alarms which is the drawback of the system. Naïve Bayseian network are restricted network that only have two layers and it assumes complete independence between the information nodes this can be count as the another limitation of the work. The work by Cristina Abad et al [9] argues for the need for correlating data among different logs to improve the accuracy of the intrusion detection system. It shows how different attacks are being logged in different network devices and argues that when only single log is analyzed some attacks cannot be detected in a proper way.

In [10], the authors propose a method of intrusion detection using a fuzzy neural network. It algorithm combines artificial neural network (ANN) and fuzzy Inference systems (FIS), as well as evolutionary algorithms. They create an algorithm that uses fuzzy rules and allow new neurons to be created in order to accomplish this. Usage of Snort is also there in order to gather real time data for training the algorithm and to compare their technique with that of an augmented neural network.

As some authors are only concentrating to the certain types of network and we can see that in most of intrusion detection systems false alarm rate is not considering. Some of the works are also not concentrating to the efficiency in the selection of attributes. We are coming with a system that will take four types of network like Intrinsic, Host based, Time based and Content based network so that we can limit the input data only within the network that we are selecting which will increase the computational efficiency. Our aim is also to increase the accuracy of the detection system with less error like reducing false positive and false negative. For this aggregation result of binary classifiers and MultiBoosting is applied. We are also concentrating in the detection of unknown and less frequently occurring attacks in an efficient way by feature extraction technique which is a fussy logic[11] method that uses the efficient selection of informative attribute for the less computation time.

Proposed Work

The proposed work is to classify the data in to normal and attack for this KDD cup 99 data sets are taken for the analysis. The first step is the selection of attribute which plays a key role in building effective IDS as time required for building classification models increases with increase in number of attributes. By eliminating redundant, irrelevant or uninformative attributes a subset of original attributes that are most helpful for classifications of attacks are selected. Selection of attributes and mining with less number of attributes can help in saving computing time, increase in the accuracy of classification models to build efficient models.

Attacks are of two types namely known and unknown attacks. The attacks that are already in the database and attacks that are occurring very frequently are called known attacks and the attacks that occur less frequently and those are not in the databases are called unknown attacks. The intrusion detection and classification of attacks are easier for these type attacks on the other hand intrusion detection and classification of unknown attacks are very difficult therefore, we are providing feature selection option so that we can select one of four types of network namely Intrinsic, Host based, Time based and Content based networks which are described as below. [12].

1. Intrinsic Network: Tcp trace utility is used to extract information from packets. The new intrinsic features such as data bytes are extracted.
2. Host based Network: Host-Based Intrusion operate on information collected within an individual computer system and attack on a particular system or host. The file data or the operating system alone is targeted by the attacker.
3. Time based Network: Within the given time, the numbers of connections to the same service, the error to the same service and to the different services are extracted.
4. Content based Network: The flow of message throughout the network is driven by the content of the messages rather than the explicit address assigned by the senders of the message.

Next step after feature selection is the feature extraction. After feature selection the data sets will be loaded in to the database for the further processing. Binary classification is also called as many value logic that deals with values that are more approximate rather than fixed and exact. It has truth value ranges from 0 and 1 that helps in classifying the incoming data as normal or attack. If the data is detected as attack then feature extraction method of fuzzy logic, which discovers relations between variables in large databases is used to extract the feature of the attack like its source, protocol, service, category that are the needed attributes and stored in .xml format that makes the classification as per their attack types become easier.

False alarms are reduced by checking the probability that the packet can be an attack and by extracting rules from the training dataset. After the classification decision tree are made with the help of Multi Boosting which is the extension of Adaboost that will take care of the instances which are misclassified by previous classifiers. From the decision tree it is possible to analyze the frequency of the attacks and a suitable intrusion prevention system can be designed as a future work. It is also possible to check the sequence of the attack that had happened in final commit phase. It will give search result if the corresponding searching sequences of attack are present. Fig 1 shows the architectural diagram of proposed intruder detection and classification of attack model.

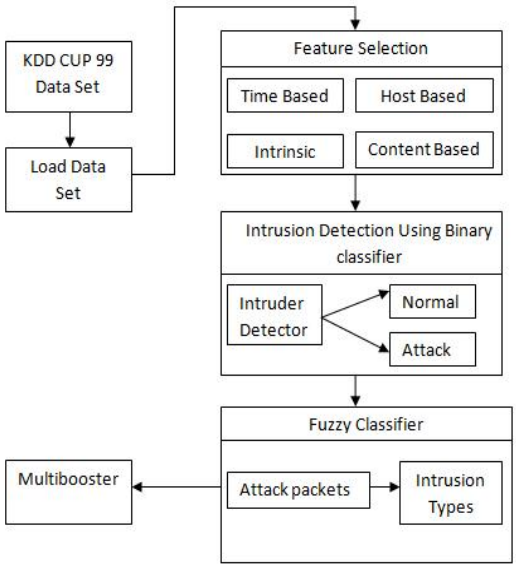

Figure 1: Proposed Classification Model

Results and Evaluation

In our proposed system an efficient intrusion detection system is designed with the help of binary classifier and fuzzy logic for feature extraction which will classify the packets in to attack or normal after feature selection from the one of four type of network namely Intrinsic, Host based, Time based and Content based networks. The type of attack is classified using feature extraction classifier.

Table I shows the classification of attacks after binary classification and after extracting the features using fuzzy logic. The instances are classified as Normal or as an attack. If it is an attack it is classified it in to which type of attack it belongs to.

Table 1: Classification of Attacks In Intrinsic Network

Duration	Protocol	Service	Total bytes	Attack name
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	icmp	ecr_i	1032	smurf
0	tcp	private	0	neptune
0	tcp	private	0	neptune
0	tcp	smtp	1703	Normal
1	tcp	smtp	1876	Normal
0	udp	domain_u	30	Normal
0	tcp	smtp	1577	Normal

Table II shows the attacks and the value of occurrence of the attacks in intrinsic network. Same type of table can be generated for times based network, host based network and content based network also. Twenty seven types of attacks are given to the database for the classification.

Table 2: Attack Types and Its Frequency of Occurrence In Intrinsic Network

ATTACK TYPES			
Attack	BACK	DICT	EJECT
	60	0	0
Attack	IMAP	IPSWEEP	LAND
	0	0	8
Attack	PERLMAGIC	PHF	POD
	0	0	8
Attack	SPY	SYSLOG	TEARDROP
	0	0	10
Attack	FFB	FORMAT	FTP-WRITE
	0	0	0
Attack	GUEST	LANDMODULE	MULTIHOP
	0	0	0
Attack	NEPTUNE	NMAP	PORTSWEEP
	24	27	0
Attack	ROOTKIT	SATAN	SMURF
	0	26	86
Attack	WAREZ	WAREZCLIENT	WAREZMASTER
	0	0	0

The prediction of attack can be done in an easy way when decision tree is constructed. Detection tree is selected from the performance evaluations of algorithms. We are taking three matrixes like correctly classification, incorrectly classification and accuracy rate for selecting the best performing algorithm.

1) Correctly Classification: It gives the percentage of correctly classified data out of total instances and it is also called as true positive (tp). It is calculated for decreasing false positive (fp).
2) Incorrectly Classification: It gives the percentage of incorrectly classified data out of total instances and it is also called as true negative (tn). It is calculated for decreasing false negative (fn).
3) Accuracy Rate: The accurate classifications of the instances are calculated with respect to false positive, false negative, true positive and true negative. Depending up on this value algorithm is selected for the further process.

$$Accuracy\ rate \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

The comparisons of algorithms are performed on WEKA TOOL.KDD CUP 99 data sets are taken and Six algorithms namely [13] Decision tree, Ripper, Neural Network, Navie Bayes, K-NN,SMO are evaluated for the correctly classification, Incorrectly classification and accuracy rate of the algorithm[14].

1) Decision Tree Classifier: It uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value.
2) Ripper: This implement propositional rule learner which will reduce the error rate. It is based in association rules with reduced error pruning

3) Neural Networks: They process records one at a time, and "learn" by comparing their classification of the record with the known actual classification of the record. The errors from the initial is fed back into the network, and used to modify the networks algorithm for iterations.
4) Navies Bayes: They are a family of probabilistic classifiers that are based on Bayes' theorem with strong (naive) independence assumptions between the features.
5) K-NN: It is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression
6) SMO: This implementation replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default.

The simulation experiment has been conducted and the result obtained is indicated by means of graphs in fig 2. By the experimentation it is found that decision tree is performing well in terms of accuracy rate of 98.67% which is calculated using the equation 1 compared to other six algorithms with KDD CUP 99 dataset as its input.

Since decision tree is having high accuracy rate as per the evaluation that we have done, decision tree is constructed with the frequency of occurrence of attack so that the further incoming packets can be easily predicted and classify as attack or normal and also to the types of attacks.

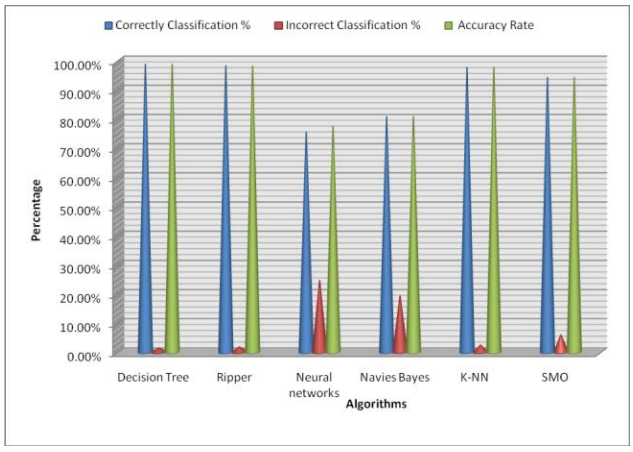

Figure 2: A graph showing the correctly classified, incorrectly classified and accuracy rate percentage and relay percentage comparison for Decision tree, Ripper, Neural Network, Navie Bayes, K-NN and SMO algorithm

Conclusion

Attacks are increasing as the increase in the use of wireless technologies. Therefore, the need for effective intrusion detection system becomes essential. In our work we are designing an efficient intrusion detection system using suitable data mining technique.

The KDD cup 99 data sets are taken for the analysis. The attributes that are most helpful for classification and detection decision making are selected. The intrusion detection and classification of attacks are easier for the known attacks on the other hand for unknown attacks; we are providing feature extraction method of fuzzy logic to extract the feature as per the needed attributes. Four types of network namely intrinsic network, Host based network, Time based network and content based networks are concentrated for feature selection are done and stored separately. Binary classification is done for classifying the packet in to normal and attack file. Feature extractions are done and stored separately in .xml format that are stored helps the classification in an easy way. False alarms are reduced by checking the probability that the packet can be an attack and by extracting rules from the training dataset. After the classification decision tree are made with the help of Multi Boosting which is the extension of Adaboost that will take care of the instances which are misclassified by previous classifiers.

From the decision tree it is possible to analyze the frequency of the attacks and a suitable intrusion prevention system can be designed as a future work. In our work we are taking the KDD CUP 99 data set, for future work we can take the real time traffic data using snort by analyzing packets from networking devices and even prediction of unknown attacks can be done by correlation and normalization rules which will also helps to reduce false alarms in more efficient way with suitable prediction methods..

Performance evaluation of six algorithms Decision tree, Ripper, Neural Network, Navie Bayes, K-NN and SMO algorithm are done based up on the three parameter correctly classifies, incorrectly classified and accuracy rate by taking the KDD CUP 99 data sets. It is examined that decision tree is performing well compared to other five algorithms in terms of accuracy rate of 98% than other five algorithms with KDD CUP 99 data set as its input.

References

[1] Christos Douligeris, Aikaterini Mitrokotsa, 2004, ”DDoS attacks and defense mechanisms: classification and state- of-theart”, ComputerNetworks:The International Journal of Computerand Telecommunications Networking, Vol. 44, Issue 5, pp: 643 - 666.

[2] J.Oberheide, E. Cooke, and F. Jahanian, 2008, “Cloudav: N-version antivirus in the network cloud,” in 2008 Proceedings. USENIX Security Symposium.

[3] Lee, W. and D. Xiang, 2001, ”Information-theoretic measures for anomaly detection”, IEEE Symposium on Security and Privacy, IEEE.

[4] “Snort Intrusion Detection System, Oct. 1, 2011, ” http://www.snort.org/ (last accessed.

[5] Porras, A. and Neumann, P. G., October 1997, ”EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances”, National Information Systems Security Conference.

[6] D.Barbara, J.Couto, S.Jajodia, and N.Wu, 2001, "ADAM: A test bed for exploring the use of data mining in intrusion detection", SIGMOD, vol30, no.4, pp 15-24.

[7] Chotirat "Ann" Ratanamahatana and Dimitrios Gunopulos, "Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection," Computer Science Department University of California Riverside, CA 92521 1-909-787-5190.

[8] Mrutyunjaya Panda and Manas Ranjan Patra, , 2007, Network intrusion detection using naive bayes. International journal of computer science and network security, 7(12): 258-263.

[9] Cristina Abad, Jed Taylor, Cigdem Sengul, William Yurcik, Yuanyuan Zhou, and Ken Rowe, 2003, Log correlation for intrusion detection: A proof of concept. In Computer Security Applications Conference. Proceedings. 19th Annual, pages 255-264. IEEE.

[10] S.chavan, K.Shah, N.Dave, S.Mukherjee, A.Abraham, and S.Sanyal, 2004, "Adaptive neuro-fuzzy Intrusion detection systems", ITCC, Vol 1.

[11] G. Florez, SM. Bridges, Vaughn RB, 2002, "An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection", Annual Meeting of The North American Fuzzy Information Processing Society Proceedings.

[12] Levent Ertoz and Eric Eilertson and Aleksandar Lazarevic and Pang-Ning Tan and Vipin Kumar and Jaideep Srivastava and Paul Dokas, 2004, "MINDS - Minnesota Intrusion Detection System", Next Generation Data Mining, IT Press.

[13] Luo, J, 1999, "Integrating fuzzy logic with data mining methods for intrusion detection", Master's thesis, Mississippi State Univ.

[14] Chotirat "Ann" Ratanamahatana and Dimitrios Gunopulos, "Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection," Computer Science Department University of California Riverside, CA 92521 1-909-787-519.