Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 6, Issue, 11, pp.9562-9568, November, 2014

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

REVIEW ARTICLE

REAL TIME INTRUSION DETECTION SYSTEM USING HYBRID CONCEPT

*Monika Kumari and Sanchika Bajpai

Herman's Park, 30 Bund Garden E-10 Flat No, Pune-03, India

ARTICLE INFO

Article History:
Received 22nd August, 2014
Received in revised form
08th September, 2014
Accepted 24th October, 2014
Published online 18th November, 2014

Key words:
Network Security,
Intrusion Detection,
ANN, Naive Bayes Classifier

ABSTRACT

With the expansion of computer networks, security has become a important issue for computer systems. The threat from attackers and criminal enterprises has grown with the expansion of Internet, therefore, intrusion detection systems have become a core component of computer network due to such threats. In this paper, we present framework integrated with neural network to build an effective intrusion detection system. In paper ANN and Naives Bayes strategies are used to detect the attacks in the system by using dynamic data or real time data. Naives Bayes is used to predict the attacks in the system. In prediction it has to specify that attacks is good or bad. After prediction the data is send to ANN which define the types of attacks are present in system. The systems are compared with existing approaches of intrusion detection which either uses neural network or based on layered framework.

Copyright © 2014 Monika Kumari and Sanchika Bajpai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

With the growth of network-based services and information on networks, network security is becoming more and more important. The highly connected computing world has intruders and hackers with new facilities for their destructive purposes. Intrusion detection techniques are used to protect against computer attacks to secure network architecture design, firewalls, and personal screening. Thus, intrusion detection systems play a important role in network security. The costs of temporary or permanent damages caused by unauthorized access of the attackers to computer systems have give different organizations to increasingly implement various systems to monitor data flow in their networks. Thus the security and continuty of the service can be protect by the intrusion detection system and it can handle large amount of data without affecting the performance.

Intrusion detection system has two main approaches:

Misuse based system

In a misuse detection based intrusion detection system, intrusions are detected by searching for activities that correspond to known signatures of intrusions and compare them against a database of signatures or patterns of known threats. This is similar to the way most software detects threats. The issue is that there will be a lag between a new threat being discovered in the wild and the signature for detecting that threat being applied to your IDS.

Anomaly based system

In anomaly based intrusion detection system, detection is performed by detecting changes in the patterns of utilization or behavior of the system. An IDS which is anomaly based will monitor network traffic and compare it against an established baseline. Neural networks are a form of artificial intelligence that uses multiple artificial neurons, networked together to process information. This type of network has the capability to learn from patterns, and extrapolate results from data that has been previously entered into the network's knowledge base. Neural network algorithms are emerging nowadays as a new artificial intelligence technique that can be applied to real-life problems.

This ability makes neural network applications extremely valuable in intrusion detection (Herv´e Debar, ?). Presents an approach of user behavior modeling that takes advantage of the properties of neural network algorithms coupled with expert system. The data presented to the systems consist of attack specific keyword-counts in network traffic. In (Ryan *et al.*, 1998 and Lippmann, 1999) neural networks have been applied to build keyword-count-based misuse detection systems. Authors in (Ghosh and Schwartzbard, 1999) also uses neural network to analyze program behavior profiles for both anomaly detection and misuse detection to identify the normal system behavior. The results of a study on intrusion detection on IIS (Internet information services) utilizing a hybrid intrusion detection system. Applies neural network approach to probing attacks that are the basis of other attacks in computer network systems show neural network pattern recognition back propagation algorithm to be effective in intrusion detection.

Corresponding author: Monika Kumari,
Herman's Park, 30 Bund Garden E-10 Flat No, Pune-03, India.

Comparing different neural network classifiers, the back-propagation neural network (BPN) is shown to be more efficient in developing IDS. However, the simulation time required to induce models from large datasets is long. The authors show that the output from these classifiers can be combined to generate a better classifier rather than selecting the individual best classifier. Similarly in combination of 'weak' classifiers are used where the individual classification power of weak classifiers is shown to be slightly better than that of random guessing. Authors in (Chuanyi and Sheng, 1997) and describe a data mining framework for adaptively building intrusion detection models. The proposed a distributed intrusion detection framework based on autonomous mobile agents. It uses a layered framework to build a network IDS which can detect a wide variety of attacks reliably and efficiently when compared to the traditional network IDS but the accuracy of less occurring attack is not good.

Therefore, an attempt is made in this paper to build an IDS by integrating layered framework with neural network so as to combine the advantages of both the approaches. Thus, an integrated IDS is proposed which can detect a wide variety of attacks with less false alarm rate and can operate efficiently in high speed network. Comparing different neural network classifiers, the back-propagation neural network (BPN) is shown to be more efficient in developing IDS. However, the simulation time required to induce models from large datasets is long. The attacks were classified, according to the actions of the attacker. Each attack type falls into one of the following four main categories:

a) Denials-of Service (DoS) attacks have the aim of denying services provided to the user, computer or network. A common trick is to severely overload the targeted system. (e.g. apache, smurf, Neptune, Ping of death, back, mailbomb, etc.).
b) Probing or Surveillance attacks have the aim of gaining knowledge of the existence of a computer system or network. Port Scans of a given IP-address range fall in this category.
c) User-to-Root (U2R) attacks have the aim of gaining root access on a particular computer or system on which the attacker previously had user level access.
d) Remote-to-Local(R2L) attack is an attack in which a user sends packets to a system over the internet, which the user does not have access to in order to expose the system vulnerabilities and exploit privileges which a local user would have on the computer (e.g. xclock, dictionary, guest_password, phf, sendmail, xsnoop, etc.).

Paper Organization: Section I has introduced the basic ideas in intrusion detection and the motivations for this study. Section II reviews some basic ideas in neural network theory and presents an overview of some of the previous studies that have applied neural networks in intrusion detection. Section III deals with proposed system. Section IV deals with the dataset, attack types, and the features used for classifying network connection records in this study describes the implementation procedure. Section V concludes the paper with a discussion and possibilities for future work.

LITERATURE SURVEY

In [2007] Mrutyunjaya Panda and Manas Ranjan Patra (Mrutyunjaya Panda and Manas Ranjan, 2007) implemented a system on Network intrusion detection using Navie Bayes. They built a framework of Network intrusion detection system using Naïve Bayes algorithm. The framework detects attacks in the given datasets using this navie bayes algorithm. Navie Bayes network is restricted network which has two layers. This gives the limitation to this research. Mehdi Moradi and Mohammad Zulkernine (Mehdi MORADI and Mohammad ZULKERNIE, ?) implemented a system on neural network system for intrusion detection and classification of attacks. The authors used the ANN algorithm to implement the system. Research off-line intrusion detection system is implemented using multi layer perception artificial neural network. Now, we are developing a system using hybrid approach. In this paper we are using two algorithms which gives better attack detection results. An approach for a real time intrusion detection system, intended to classify the normal and attack patterns and the type of the attack. In real time IDS hybrid approach is used and dynamic data is used. The packets are captured from real time data only. Here observed that all the packets which are captured from dynamic dataset are send to training dataset. After training dataset work the prediction of attacks is done by using Naïve Bayes. After prediction attacks are gone towards the ANN. ANN findout the attacks and final result is shown.

IMPLEMENTATION DETAILS

Artificial Neural Networks (ANN)

The ability of software computing techniques to dealing with untrusted data makes them attractive. However, ANNs are the most commonly used software computing technique in IDSs (Srinivas Mukkamala, 2002 and Sinclair *et al.,* 1999). Some applications used software computing techniques which not similar to ANNs in intrusion detection. For example, genetic algorithms have been used along with decision trees to automatically generate rules for classifying network connections. An ANN is an information processing system which work like biological nervous systems, such as the brain which work as process information. It is combination of a large number of interconnected processing neurons working with each other to solve problems. Each processing neuron is sum of element. The learning process is an optimization process. In this process parameters of the best set of connection coefficients (weighs) for solving a problem are found and includes the following basic steps:

-ANN has neural network with a number of inputs (vectors each representing a pattern)
-ANN will check how closely the actual output generated for a specific input matches the desired output.
-ANN will change the neural network parameters (weights) to better outputs.

The output of each neuron is used as the input to all of the neurons in the next layer. Some IDS designers used ANN as a recognition technique. Recognition can be implemented by using a feed-forward neural network which has been trained accordingly. Whenever the neural network is used, it identifies the input pattern and tries to output the corresponding class. When a connection record that has no output associated with it is given as an input, the neural network gives the output that corresponds to a taught input pattern that is least different from the given pattern. During training, the neural network parameters are optimized to associate outputs with corresponding input patterns (every input pattern is represented by a feature vector extracted from the characteristics of the network connection record).. The ability of software computing techniques for dealing with uncertain and partially true data makes them attractive to be applied in intrusion detection. Some studies have used soft computing techniques other than ANNs in intrusion detection. The most commonly reported application of neural networks in IDSs is to train the neural net on a sequence of information units, each of which may be an audit record or a sequence of commands For example, genetic algorithms have been used along with decision trees to automatically generate rules for classifying network connections. However, ANNs are the most commonly used software computing technique in IDSs (James Cannady, 1998; Fox *et al*., 1990 and Debar *et al*., 1992).

An approach for a neural network based intrusion detection system, intended to classify the normal and attack patterns and the type of the attack, has been presented in this paper. We applied the early stopping validation method which increased the generalization capability of the neural network and at the same time decreased the training time. It should be mentioned that the long training time of the neural network was mostly due to the huge number of training vectors of computation facilities. However, when the neural network parameters were determined by training, classification of a single record was done in a negligible time. Therefore, the neural network based IDS can operate as an *online* classifier for the attack types that it has been trained for. The only factor that makes the neural network off-line is the time used for gathering information necessary to compute the features.

A two layer neural network was also successfully used for the classification of connection records. Although the classification results were slightly better in the three layer network, application of a less complicated neural network was more computationally and memory wise efficient. From the practical point of view, the experimental results imply that there is more to do in the field of artificial neural network based intrusion detection systems. The implemented system solved a three class problem. However, its further development to several classes is straightforward. As a possible future development to the present study, one can include more attack scenarios in the dataset. Practical IDSs should include several attack types. In order to avoid unreasonable complexity in the neural network, an initial classification of the connection records to normal and general categories of attacks can be the first step. The records in each category of intrusions can then be further classified to the attack types.

Naive Bayes

The naive bayes model is a simplified Bayesian probability model. The probability of result is encoded in the model along with the probability of the variables occurring given that the result should occurs. The probability of proof variable gives the result occurs is assumed to be independent of the probability of other proof variables give the results occur. The knowledge of the attributes, they record whether or not a attack actually occurred. Consider the category of whether a attack occurred or not as the naive Bayes classifier. This is the knowledge that we are interested in. The attributes proof that the attack has occurred. Figure 1 below shows the framework for a Naive Bayes model to perform intrusion detection. The number of n attributes, the naive bayes classifier makes 2n! results. The results of the naive bayes classifier are often correct. The results examines the circumstances under which the naïve bayes classifier performs. The naive bayes classifier operates on a strong assumption. That means the probability of one attribute does not affect the probability of the other. Training data noise can only be decreased by choosing good training data. The training data must be divided into various groups by the machine learning algorithm. Bias is the error due to groupings in the training data being very large. Variance is the occur due to those groupings being too small.

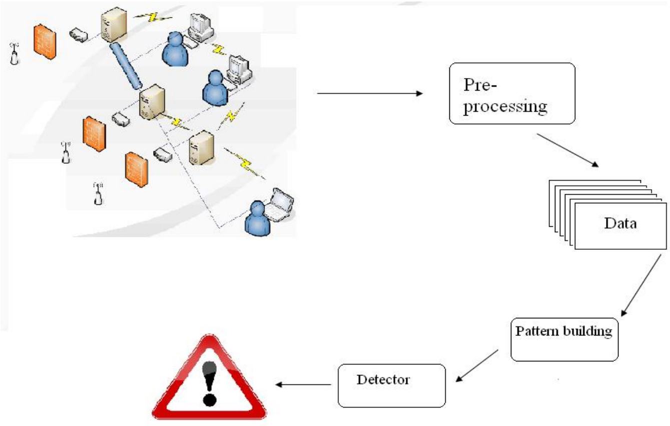


Fig.1. The framework of the Intrusion Detection Model

Fig.1. The framework of the Intrusion Detection Model. In the training phase, the naïve bayes algorithm calculates the probabilities of a theft given a particular attribute and then stores this probability. It is repeated for each attribute, and the amount of time taken to calculate the probabilities for each attribute. In the testing phase, the amount of time taken to calculate the probability of the given class for each case in the worst case is proportional to n, the number of attributes. In worst case, the time taken for testing is same as that for the training phase. In the Naïve Bayes, we have seen proposed framework of NIDS based on Naïve Bayes algorithm. The framework explain the patterns of the network services. With the patterns, the framework detects attacks in the datasets using the naïve Bayes Classifier algorithm. Compared to the Neural network approach, their approach achieve high detection rate, it is less time consuming and has low cost. However, it generates somewhat more false results. As a naïve Bayesian network is a restricted network that has only two layers and assumes complete independence between the information nodes. In order to solve this problem so as to reduce the false results, active platform or event based classification may be thought of using Bayesian network.

Architectural Details

IDS under consideration combine the advantages of both layered framework and neural network The proposed IDS is used to detect four common types of attacks like Denial of Service (DoS), Probe, Remote to Local (R2L),User to Root (U2R) and normal records also. In paper the hybrid approach is used. Hybrid approach is used because we are using two algorithm. In this architecture data enter as a packet then this packet is scanned after that packet analyzer do the analyzing of packet. After analyzing the packet go through the detailed signatures here some particular signatures are present and then the packet will send to the dataset labeling here final labeling is done decide that this data should send to ANN or Naïve Bayes. Figure 2 shows the proposed system.

Steps

1. First packets are entered then packets are analysis and scanned.
2. Then signatures of packets are searched.
3. After that labeling of dataset is done.
4. Then we have two algorithms ANN and Naive Bayes.
5. Here real time packets are used.
6. These packets are sending towards training dataset.
7. After training dataset data is send towards the Naïve Bayes algorithm.
8. Naives Bayes algorithm predict the attacks.
9. After prediction the searching of good or bad attacks are done.
10. If bad attacks are predicted then they are sending towards the ANN algorithm.
11. ANN algorithm finds out the attacks types and shows the results.

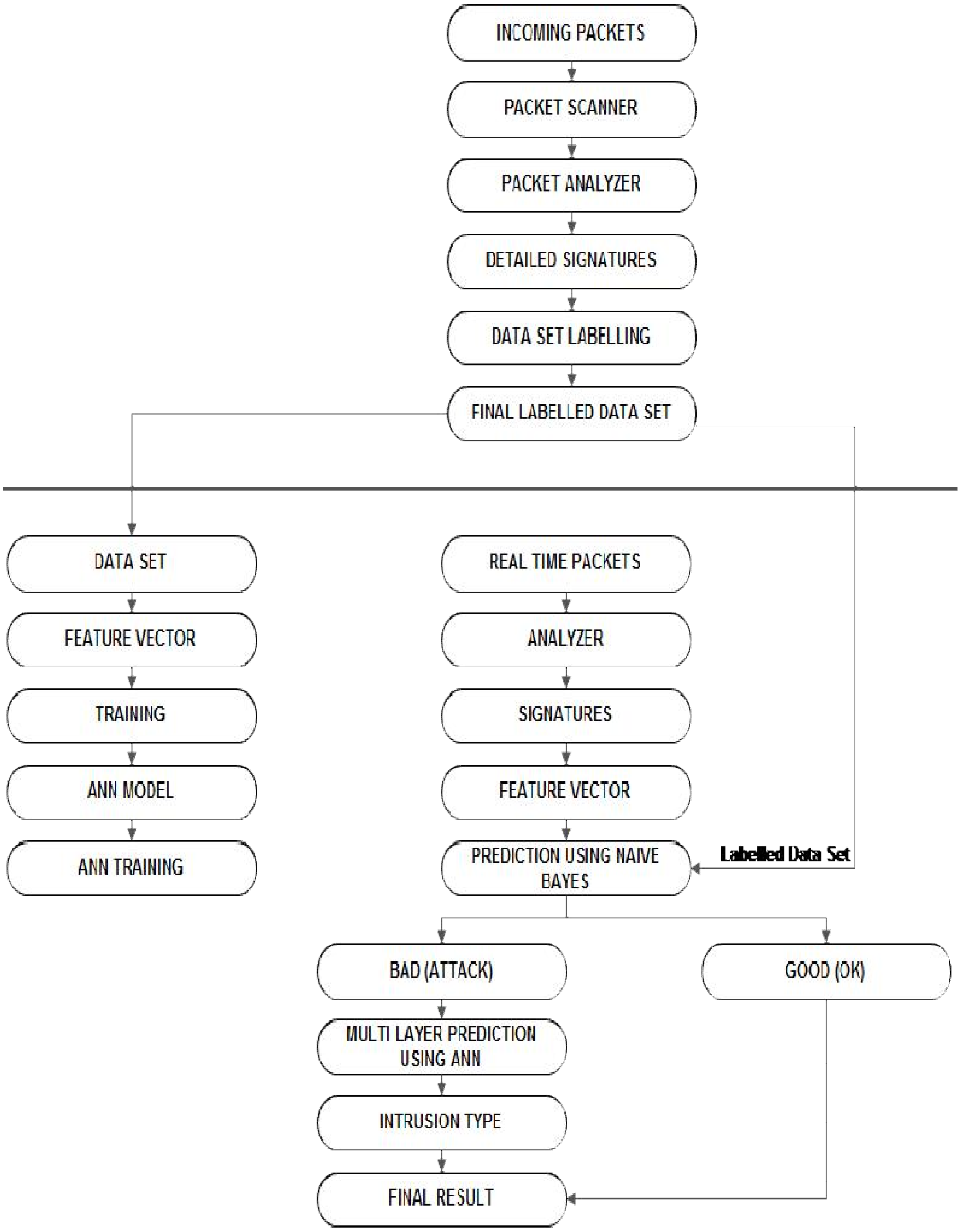


Fig.2. Operation flow for proposed system

DATASET AND ATTACK DESCRIPTION

In previous proposed architecture the static dataset is used in the system. The static dataset is limited data which not reliable. Now in this paper we are using real time dataset or dynamic dataset. To make the data reliable we are using real time dataset. In this paper we don't have to store the dataset in advance. In previous proposed architecture experiments are conducted using. KDD Cup 99 dataset which consist of a set of 41features derived from each connection and a label which specifies the connection records as either normal or specific attack type. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment. Generally the attacks fall into four main categories namely DoS, Probe, R2L and U2R. KDD-cup.data_10_percent.gzis used as training and validation dataset having exactly 494,021 instances with 22 attack types.

Probe attacks

The probe attacks are used to collect information about the target network from a source that is external to the network. Hence, basic connection level features such as the "duration of connection" and "source bytes" are significant while features like "number of files creations" and "number of files accessed" are not expected to provide information for detecting probes.

Precision Class 1	(Relevant Intersect Retrieved) / Retrieved	Correct Retrieved Object / Retrieved Objects	18
Precision Class 2	(Relevant Intersect Retrieved) / Retrieved	Correct Retrieved Object / Retrieved Objects	0.8583333
Recall Class 1	(Relevant Intersect Retrieved) / Relevant	Correct Retrieved Object / Actual Objects	0.85
Recall Class 2	(Relevant Intersect Retrieved) / Relevant	Correct Retrieved Object / Actual Objects	0.82

Data Set Name	Actual Objects	Retrieved Objects	Correct Retrieved Objects
Class 1	20	18	17
Class 2	25	24	23

DoS Attacks

The DoS attacks are used to force the target to stop the service(s) that is (are) provided by flooding it with probes illegitimate requests. To detect DoS attacks, it may not be important to know whether a user is "logged in or not."

R2L Attacks

The R2L attacks are one of the most difficult to detect as they involve the network level and the host level features. U2R

Attacks

The U2R attacks involve the semantic details that are very difficult to capture at an early stage. Such attacks are often content based and target an application. IDS the advantages of both layered framework and neural network The proposed IDS is used to detect following types of attacks like Denial of Service (DoS), Probe, Remote to Local (R2L),User to Root (U2R) and normal records also. In paper hybrid approach is used. Hybrid approach is used because we are using two algorithm. In this architecture data enter as a packet then this packet is scanned after that packet analyzer do the analyzing of packet. After analyzing the packet go through the detailed signatures here some particular signatures are present and then the packet will send to the dataset labeling here final labeling is done decide that this data should send to ANN or Naïve Bayes.

RESULTS

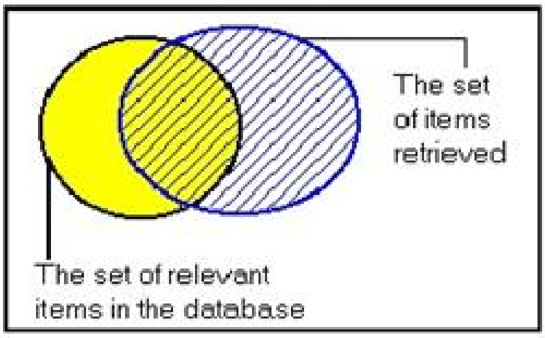
In the below table we are analyzing the result and performance of attacks. Here types of attacks are shown and training dataset, no. of iteration and minimum time to detect a attack is calculated after implementation of project. This all values can be calculated after implementation only.

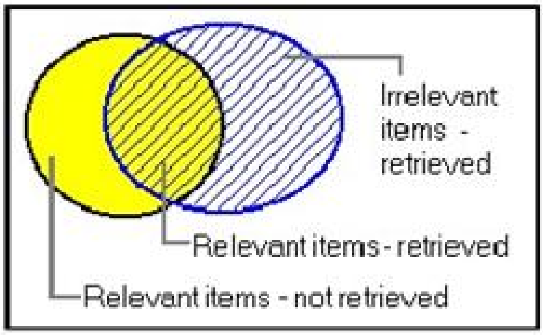
FAR and FRR Graph

This graph, sometimes called the Equal Error Graph, is the one most often used by researchers trying to understand the performance of their Verification system. It shows the False Accept and False Reject Rates at all thresholds. Minimizing the Crossover of the 2 plots is generally the goal of the researcher. Information on how well a system is handling Impostors can be understood from the steepness of the FAR plot. The user of a Verification System will also use this graph to calculate where to set their operating threshold. The graph will show the expected False Accept and False Reject rates at any chosen threshold.

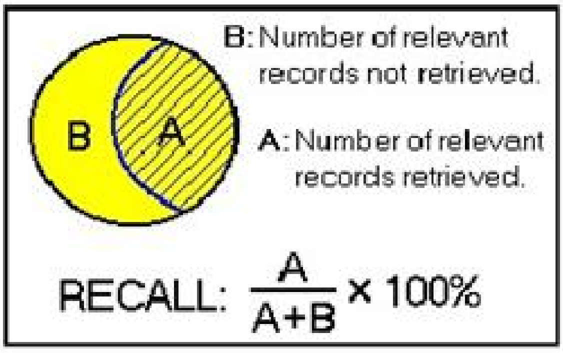
PRECISION and RECALL

Precision and recall are the basic measures used in evaluating search strategies. Records are assumed to be either relevant or irrelevant (these measures do not allow for degrees of relevancy).The actual retrieval set may not perfectly match the setoff relevant records.

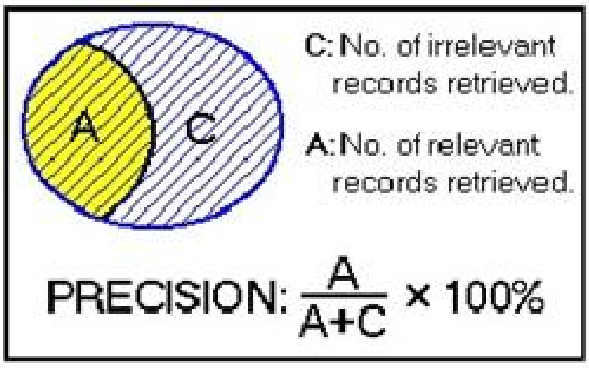

The set of items retrieved
The set of relevant items in the database


Irrelevant items - retrieved
Relevant items - retrieved
Relevant items - not retrieved

RECALL is the ratio of the number of relevant records Retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.


B: Number of relevant records not retrieved.
A: Number of relevant records retrieved.
B
A
RECALL: $\frac{A}{A+B} \times 100\%$

PRECISION is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

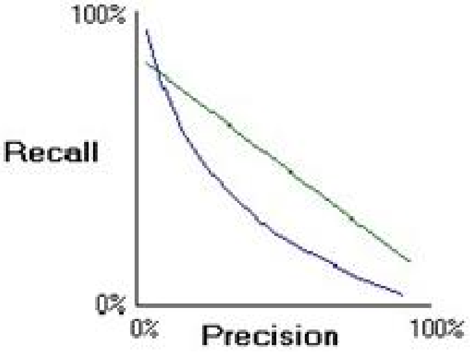

C: No. of irrelevant records retrieved.
A: No. of relevant records retrieved.
A
C
PRECISION: $\frac{A}{A+C} \times 100\%$

RECALL AND PRECISION ARE INVERSELY RELATED

As recall ↑ precision ↓

conversely:

As recall ↓ precision ↑



In the graph above, the two lines may represent the performance of different search systems. While the exact slope of the curve may vary between systems, the general inverse relationship between recall and precision remains.

Conclusion

An approach for a real time intrusion detection system, intended to classify the normal and attack patterns and the type of the attack. In real time IDS hybrid approach is used and dynamic data is used. The packets are captured from real time data only. Here observed that all the packets which are captured from dynamic dataset are send to training dataset. After training dataset work the prediction of attacks is done by using Naïve Bayes. After prediction attacks are gone towards the ANN. ANN find out the attacks and final result is shown.

REFERENCES

Chuanyi Ji, Sheng Ma, 1997. “Combinations of Weak Classifiers,” *IEEE Transactions on Neural Networks*, vol.8(1), pp. 32–42.

Debar, H., Becker, M. and Siboni, D. 1992. “A neural network component for an intrusion detection system,” *Proceedings of 1992 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, California*, pp. 240 – 250.

Fox, K., Henning, R., Reed, J. and Simonian, R. 1990. "A neural network approach towards intrusion detection," *Proceedings of 13th National Computer Security Conference, Baltimore, MD*, pp. 125-134.

Ghosh, A., Schwartzbard, A. 1999. "A study in using Neural Networks for Anomaly and Misuse Detection," *Proceedings of the 8th USENIX Security Symposium.*

Gopi K. Kuchimanchi, Vir V. Phoha, Kiran S. Balagani, Shekhar R. Gaddam, 2004. “Dimension Reduction Using Feature Extraction Methods for Real-time Misuse Detection Systems,” *Proceedings of the 2004 IEEE Workshop on Information Assurance and Security, June.*

Herv´e Debar, Monique Becke, Didier Siboni, “A Neural Network Component for an Intrusion detection system,” *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pp. 240–250.

Iftikhar Ahmad, Azween B Abdullah, Abdullah S Alghamdi, 2009. “Application of Artificial Neural Network in Detection of Probing Attacks,” *IEEE Symposium on Industrial Electronics and Applications*, pp. 557-562.

James Cannady, 1998. “Artificial neural networks for misuse detection,” *Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), Arlington*, VA.

Kapil Kumar Gupta, Baikunth Nath, Ramamohanarao Kotagiri, 2010. “Layered Approach Using Conditional Random Field for Intrusion Detection,” *IEEE Transactions on Dependable and Secure Computing*, vol. 7(1), pp. 35-49, March, 2010.

Lippmann, R., Cunningham, R. 1999. "Improving Intrusion Detection Performance using Keyword Selection and Neural Networks," RAID Proceedings, West Lafayette, Indiana, Sept 1999.

Mehdi MORADI and Mohammad ZULKERNIE,”A Neural Network based system or Intrusion Detection and Classification of Attacks.

Mrutyunjaya Panda and Manas Ranjan Patra, 2007. ”Network Intrusion Detection using Navie Bayes”, *International*

Journal of Computer Science and Network Security,Vol.7 No.12.

Mukhopadhyay, I., Chakraborty, M., Chakrabarti, S. and Chatterjee, T. 2011. “Back Propagation NeuralNetwork Approach to Intrusion Detection System,” *International Conference on Recent Trends in Information Systems.*

Nidhi Srivastav, “Novel intrusion detection system integrating layered framework with Neural Network”, *IEEE International Advance Computing Conference*.

Ryan, J., Lin, M. and Mikkulainen, R. 1998. "Intrusion Detection with Neural Networks,” Advances in Neural Information Processing Systems, vol. 10, MIT Press.

Sinclair, C., Pierce, L. and Matzner, S. 1999. “An application of machine learning to network intrusion detection,” *Proceedings of 15th Annual Computer Security Applications Conference (ACSAC '99), Phoenix, AZ*, pp. 371-377.

Srinivas Mukkamala, 2002. “Intrusion detection using neural networks and support vector machine,” *Proceedings of the 2002 IEEE International* Honolulu, HI.

****