ISBN 978-93-5156-328-0
International Conference of Advance Research and Innovation (ICARI-2014)

Intrusion Detection Based On Artificial Intelligence Techniques

Shefali Singh, Krati Saxena, Zubair Khan
Department of Computer Science and Engineering, Invertis University, Bareilly, Uttar Pradesh, India

Article Info

Article history:
Received 2 January 2014
Received in revised form
10 January 2014
Accepted 20 January 2014
Available online 1 February 2014

Keywords

Intrusion Detection System,
Decision Tree,
Self-Organizing Maps

Abstract

Information technology has become a main component to support critical infrastructure services in various sectors of our society. In effort to share information and streamline operations, organizations are creating complex networked systems and opening their networks to customers, suppliers, and other business partners. Whereas most users of these networks are legitimate users, an open network exposes the network to illegitimate access and use. Increased network complexity, greater access, and a growing emphasis on the internet have made network security a major concern for organizations. The number of computer security breaches has risen significantly in the last three years. While traditional approaches to network security have focused on prevention, network intrusion detection has become increasingly important in recent years to enable firms to reduce undetected intrusion. Intrusion Detection System is one of the most important security systems to detect intrusions in a variety of networks in a distributed environment. Here, we are doing a comparative study on Intrusion Detection System based on Artificial Intelligence techniques. The main techniques which are discussed here are Decision Trees, and Self-Organizing Maps (SOM). We are describing these techniques and determining how these techniques aid in detecting intrusions in a networking environment and which is more suitable for intrusion detection002E

1. Introduction

Intrusion Detection System is one of the most important security systems to detect intrusion in the distributed networking environment. It is a system for the identification of attacks in the network and takes corrective actions to prevent them [8]. In a network, there are so many suspicious activities that need to be concerned both at network level and host level. An intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality or availability of a system resource. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices. Intrusion detection technique is a proactive security technique which can provide real-time protection for information systems. It has been put forward and used for more than two decades in the field of information security [13]. It can detect intrusions from external network as well as intrusion attempts of unauthorized users and illegal operations of authorized users from internal network. Because of the increasing dependence which companies and government agencies have on their computer networks the importance of protecting these systems from attack is critical.

To design an Intrusion Detection System, there are two main approaches:-

1. Misuse Based Ids
2. Anomaly Based Ids

Misuse Based Ids deals with the knowledge of system vulnerabilities and known attacks patterns. Misuse detection finds intrusion who attempts to break into a system by exploiting some risks and vulnerabilities. It uses rules to describe events. It uses the rule to look for events that possibly fit an intrusion scenario. These events may be monitored by system call monitoring or using and it records [8].

Corresponding Author,
E-mail address: shef.7july@gmail.com

Anomaly Based Ids is also known as behavior-based detection. In this approach, all normal activities of the system are abstracted to establish a normal behavior model. Those activities that differ from the normal behavior model are considered to be intrusions [8]. The main advantages of anomaly based Ids is that it can recognize unknown attacks and high false positive rate is its disadvantage.

In the field of intrusion detection, the algorithm is used in two stages: tolerance and testing. Mature detectors are generated in tolerance phase, and then used to detect abnormal activities in testing phase [8].

The **tolerance phase** is described as follows:

(1) Define the "Self" data;
(2) Generate random candidate detectors; and
(3) Match each candidate detector with "Self" data. If it matches with any "Self" data, it is discarded, otherwise it is added to the collection of mature detectors.

A flow diagram of tolerance phase is presented in Fig. 1.

The **testing phase** is similar to tolerance phase. We match each system activity or behavior with mature detectors. If it matches with any mature detector, it is abnormal, otherwise it is normal. A flow diagram of testing phase is presented in Fig. 2.

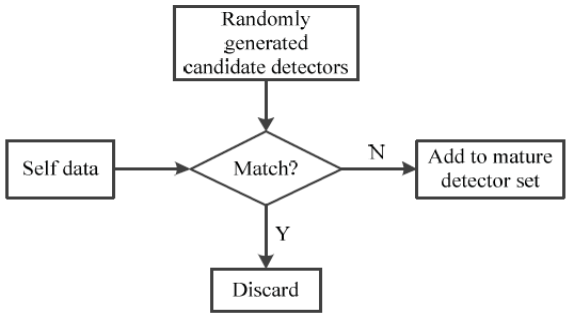


Figure 1. Tolerance phase using negative selection algorithm

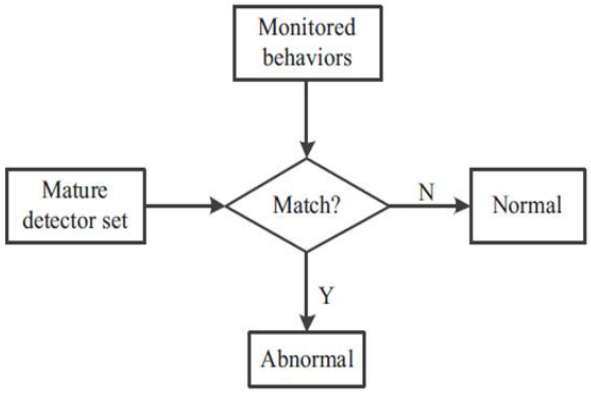


Figure 2. Testing phase using negative selection algorithm

There are mainly four types of networking attacks [12]:-

- **Denial of Service (DOS):-** It is a type of attacks that denies a user to access a machine such as Smurf, Ping, Back, Mail bomb, UDP storm etc. In this attack, a hacker makes memory resources too busy to serve legitimate networking request.
- **Remote to user Attacks (R2L):-** This attacks deals with sending packets to a machine over the internet and the user does not have access to those packets. For eg. Xlock, Xnsnoop, Phf, Guest, Sendmail dictionary etc.
- **User to Root Attacks (U2R):-**In this attacks, the hacker starts off on the system with the normal user account and mainly attempts to abuse vulnerabilities in the system in order to gain super user privileges. Eg. Xterm, Perl.
- **Probing: -** In this attack, a networking device or a machine is scanned by the hacker in order to determine vulnerabilities in the system that may be exploited later so as to compromise the system. eg. Postsweep, Nmap, Mscan, Satun, Saint etc.

Main Attack Classes	22 different attack types
DOS- Denial of Service	back, land, neptune, pod, smurf, teardrop
U2R- User to Root	buffer_overflow, loadmodule, perl, rootkit
R2L- Remote to User	ftp_write, guess_password, imap, multihop, phf, spy
Probe	ipsweep, nmap, portsweep, satan

Two Flavors of Ids:

- **Supervised Learning.** The correct answers are known and this information is used to train the network. This type of learning utilizes both input vectors and output vectors. The input vectors are used to provide the starting data, and the output vectors can be used to compare with the input vectors to determine some error. In a special type of supervised learning, *reinforcement learning*, the network is only told if it's output is right or wrong. Back-propagation algorithms make use of this style [18].
- **Unsupervised Learning.** The correct answers are not known (or just not told to the network). The network must try on its own to discover patterns in the input data. The input vectors are used solely. Output vectors generated will not be used to learn from. Also and possibly most importantly: no human interaction is needed for unsupervised learning. This can be an extremely important feature, especially when dealing with a large and/or complex data set that would be time-consuming or difficult to a human to compute [18].

ISBN 978-93-5156-328-0

2. Related Work

Tao Xu, PengYunfeng [8] presented an Intrusion detection approach inspired by biological memory cell. They have proposed a system that detects the intrusions in the system using biological memory cell. The result provides better performance than ordinary anomaly detection approaches with higher true positive rate and lower false positive rate.

Manish Kumar et.al [2] described intrusion detection system using Decision tree algorithm. In this paper they analyze a classification model for misuse and anomalyattack detection using decision tree algorithm.

A M Chandra shekhar et.al presented Intrusion detection technique by using k-means, Fuzzy neural network and SVM classifier for attaining high detection rate [19].

Cheng Leung Lui et.al [13] has described Agent based intrusion detection system using data mining approaches. Theresult shows that the frequent patterns mined from the audit data could be used as reliable agents, which outperformed from traditional signature-based NIDS.

Intrusion Detection Systems Using Decision Trees and Support Vector Machines presented by Johnson Thomas [15].The experimental result shows that Decision trees gives better overall performance than the SVM.

NarbikBashahIdris and Bharanidharan Shanmugam [14] described artificial intelligence techniques applied to intrusion detection. This paper proposes a dynamic model Intelligent

Intrusion Detection System, based on specific Al approach for intrusion detection. The techniques that are being investigated includes neural networks and fuzzy logic with network profiling,that uses simple data mining techniques to process the networkdata.

JuhaVesanto et.al presented Self-organizing map in Matlab: the SOM Toolbox.In this article, the SOM Toolbox and its usage are shortly presented. Also its performance in terms of computational load is evaluated and compared to a corresponding C program [17].

3. Data Set

The KDD99 dataset was used in the 3rd International Knowledge Discovery and Data Mining Tools Competition for building a network intrusion detector, a predictive model capable of distinguishing between intrusions and normal network connections [16]. In 1998, DARPA intrusion detection evaluation program, a simulated environment was set up to acquire raw TCP/IP dump data for a local-area network (LAN) by the MIT Lincoln Lab to compare the performance of various intrusion detection methods.

KDD Cup'99 of DARPA is being used in Intrusion Detection. They are having such attributes:-

No.	Features	No.	Features
1	duration	22	is_guest_login
2	protocol_type	23	count
3	service	24	srv_count
4	flag	25	serror_rate
5	src_bytes	26	srv_serror_rate
6	dst_bytes	27	rerror_rate
7	land	28	srv_rerror_rate
8	wrong_fragment	29	same_srv_rate
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	dst_host_srv_count
13	num_compromised	34	dst_host_same_srv_rate
14	root_shell	35	dst_host_diff_srv_rate
15	su_attempted	36	dst_host_same_src_port_rate
16	num_root	37	dst_host_srv_diff_host_rate
17	num_file_creations	38	dst_host_serror_rate
18	num_shells	39	dst_host_srv_serror_rate
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmds	41	dst_host_srv_rerror_rate
21	is_host_login		

These are the original values of KDD Cup'99 data [8]:

0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,9,9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,19,19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,235,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,29,29,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal.

4. Confusion Matrix

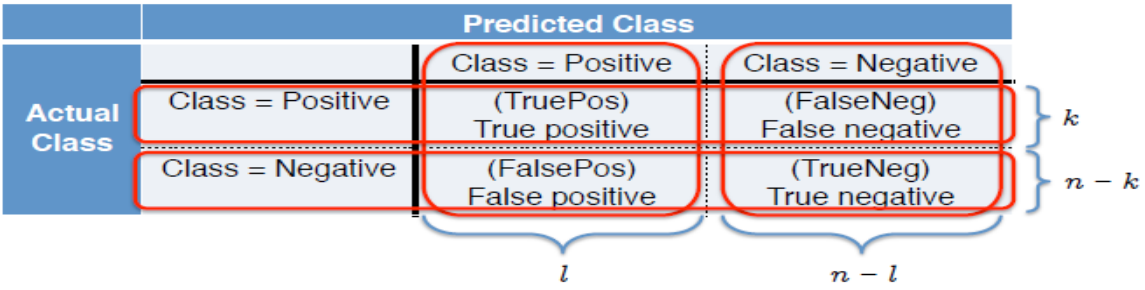
Suppose IDS gets n packets: k attack, n – k benign (where $0 \le k \le n$)
IDS thinks: l attack, n – l benign (where $0 \le l \le n$)
Confusion matrix helps us understand how IDS performed (correct/incorrect classification)
Context: positive means malicious, negative means benign [8].

True Negative Rate $= \frac{TN}{TN+FP}$ True Positive Rate $= \frac{TP}{TP+FN}$

False Negative Rate $= \frac{FN}{FN+TP}$

False Positive Rate $= \frac{FP}{FP+TN}$

ISBN 978-93-5156-328-0

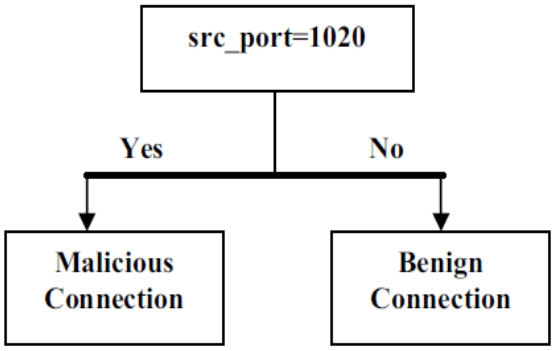


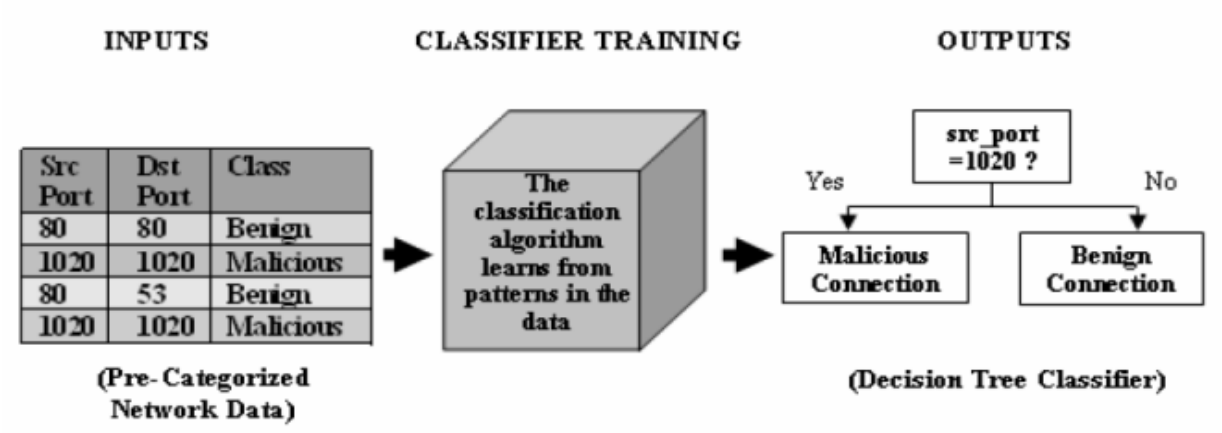
- **True positive:** IDS classified MALICIOUS packet as *malicious*
- **True negative:** IDS classified BENIGN packet as *benign*
- **False negative:** IDS classified MALICIOUS packet as *benign*
- **False positive:** IDS classified BENIGN packet as *malicious*

- True Positive (TP) is the number of attacks correctly classified
- True Negative (TN) is the number of normal records correctly classified
- False Positive (FP) is the number of normal records incorrectly classified
- False Negative (FN) is the number of attacks incorrectly classified

5. Decision Tree

Decision tree is one of the classification algorithms in data mining. The Classification algorithm is inductively learned to construct a model from the pre classified data set. Each data item is defined by values of the attributes. Classification may be viewed as mapping from a set of attributes to a particular class. The Decision tree classifies the given data item using the values of its attributes. Decision trees are one example of a classification algorithm [10]. Classification is a data mining technique that assigns objects to one of several predefined categories.





Algorithm [2]:-

1. For each attribute *a* Find the normalized information gain ratio from splitting on *a*
2. Let *a_best* be the attribute with the highest normalized information gain
3. Create a decision *node* that splits on *a_best*
4. Recurse on the sublists obtained by splitting on *a_best*, and add those nodes as children of *node*

Some advantages of the decision tree as a classification tool are:-

1. Decision trees are self-explanatory and when compacted they are also easy to follow. In other words if the decision tree has a reasonable number of leaves, it can be grasped by non–professional users. More over decision trees can be converted to a set of rules. Thus, this representation is considered as comprehensible.
2. Decision trees can handle both nominal and numeric input attributes.
3. Decision tree representation is rich enough to represent any discrete value classifier.
4. Decision trees are capable of handling datasets that may have errors.
5. Decision trees are capable of handling datasets that may have missing values.

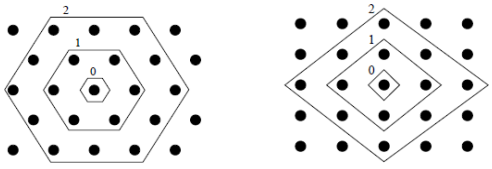
Limitations of Decision Tree [2]

1. Most of the algorithms (like ID3 and C4.5) require that the target attribute will have only discrete values.
2. As decision trees use the "divide and conquer" method, they tend to perform well if a few highly relevant attributes exist, but less so if many complex interactions are present.
3. The greedy characteristic of decision trees leads to another disadvantage that should be pointed out. This is its over–sensitivity to the training set, to irrelevant attributes and to noise.

6. Self Organizing Maps

Self-organizing maps learn to cluster data based on similarity, topology, with a preference of assigning the same number of instances to each class. Self-organizing maps are used both to cluster data and to reduce the dimensionality of data [4]. They are inspired by the sensory and motor mappings in the mammal brain, which also appear to automatically organizing information topologically. The Self Organizing Map is an extremely powerful mechanism for automatic mathematical characterization of acceptable system activity.

A SOM consists of neurons organized on a regular low dimensional grid. Each neuron is a d-dimensional weight vector (prototype vector, codebook vector) where d is equal to the dimension of the input vectors [5]. The neurons are connected to adjacent neurons by a neighborhood relation, which dictates the topology, or structure, of the map. In the Toolbox, topology is divided to two factors: local lattice structure (hexagonal or rectangular) and global map shape (sheet, cylinder or toroid) [17].

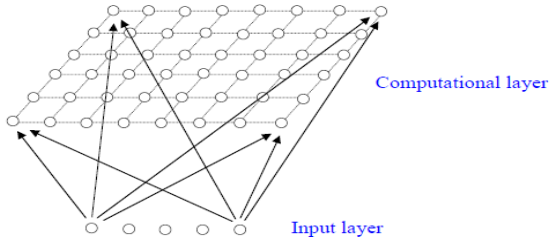


Neighborhoods (0, 1 and 2) of the centermost unit: hexagonal lattice on the left, rectangular on the right. The innermost polygon corresponds to 0-, next to the 1- and the outmost to the 2-neighborhood.

Advantages:

1. Simple and easy-to-understand algorithm that works.
2. Topological clustering
3. Unsupervised algorithm that works with nonlinear data set.
4. The excellent capability to visualize high-dimensional data onto 1 or 2 dimensional space makes it unique especially for dimensionality reduction.

The SOM system is known as a Kohonen Network. It is a type of neural network. They were developed in 1982 by TuevoKohonen, a professor emeritus of the Academy of Finland. Self- Organizing Maps are aptly named. "Self-Organizing" is because no supervision is required. "Maps" is because they attempt to *map* their weights to conform to the given input data. The nodes in a SOM network attempt to become like the inputs presented to them. In this sense, this is how they learn [9]. They can also be called "Feature Maps", as in Self-Organizing Feature Maps. Retaining principle 'features' of the input data is a fundamental principle of SOMs, and one of the things that makes them so valuable. Specifically, the topological relationships between input data are preserved when mapped to a SOM network. This has a feed-forward structure with a single computational layer of neurons arranged in rows and columns. Each neuron is fully connected to all the source units in the input layer:



A one dimensional map will just have a single row or column in the computational layer.

Applications of SOM:

- Color Classification
- Image Classification

Algorithm [9]:-

1. **Initialization**– Choose random values for the initial weight vectorsw_j.
2. **Sampling**– Draw a sample training input vector x from the input space.
3. 3.**Matching**– Find the winning neuron $I(x)$that has weight vector closest to the input vector, i.e. the minimum value of

$$d_j(x) = \sum_{i=1}^{D} (x_i - w_{ij})^2$$

4. **Updating**– Apply the weight update equation

$$\Delta w_{ij} = \eta(t) T_{j,I(x)}(t)(x_i - w_{ij})$$

Where $T_{j,I(x)}(t)$is a Gaussian neighborhood and $\eta(t)$is the learning rate.

1. **Continuation**– keep returning to step 2 until the feature map stops changing.

Limitations of Self Organizing Maps [9]

Time consuming algorithm, this is because as the number of neurons affects the performance of the algorithm. And as the number increases the computation increases which results in increasing computational time.

7. Experimental Result

Decision Tree algorithm, and Self-Organizing Maps are tested on KDD Cup'99 dataset and compared that which is better for intrusion detection. There are generally two types of attacks in network intrusion detection: the attacks that involve single connections and the attacks that involve multiple connections (bursts of connections). The standard metrics in Table 1 treat all types of attacks similarly thus failing to provide sufficiently generic and systematic evaluation for the attacks that involve many network connections.

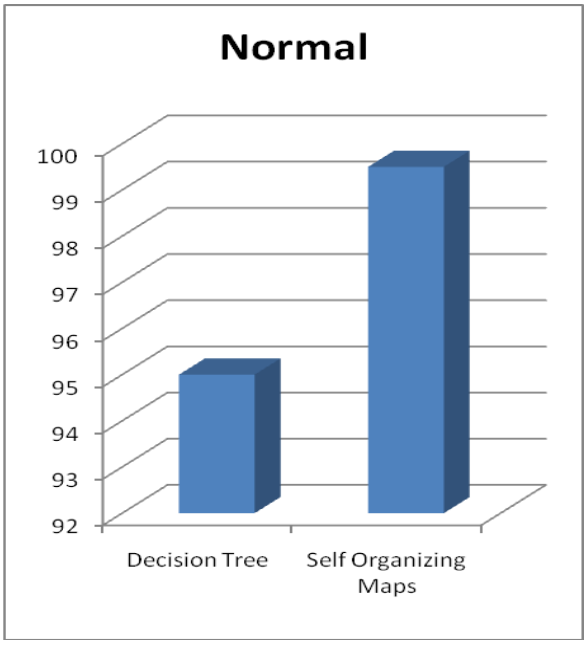
Table: 1. Confusion Matrix for Evaluation Of Intrusion Detection

Confusion Matrix		Predicted Class	
		Normal	Intrusion / Attack
Actual Class	Normal	True Negative	False Positive
	Intrusion / Attack	False Negative	Correctly Detected

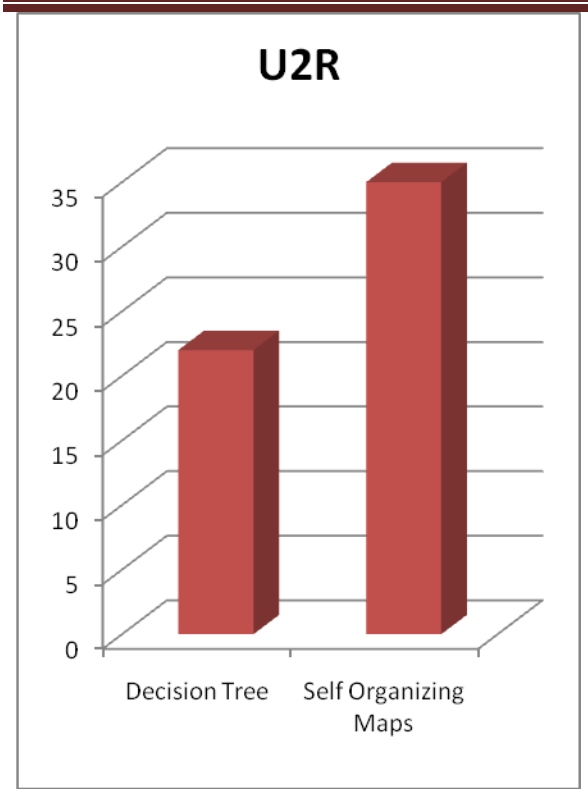
The experiment is carried out on an intrusion detection realdata stream which has been used in the Knowledge Discovery and Data Mining (KDD) 1999 Cup competition. In KDD99dataset the input data flow contains the details of the network connections, such as protocol type, connection duration, protocol type etc. Each data sample in KDD99 dataset represents attribute value of a class in the network data flow, and each class is labeled either as normal or as an attack with exactly one specific attack type. In total, 41 features have been used inKDD99 dataset and each connection can be categorized into five main classes as one normal class and four main intrusion classes as DOS, U2R, R2L and Probe. There are 22 different types of attacks that are grouped into the four main types of attacks DOS, U2R, R2L and Probe tabulated in Table 2.

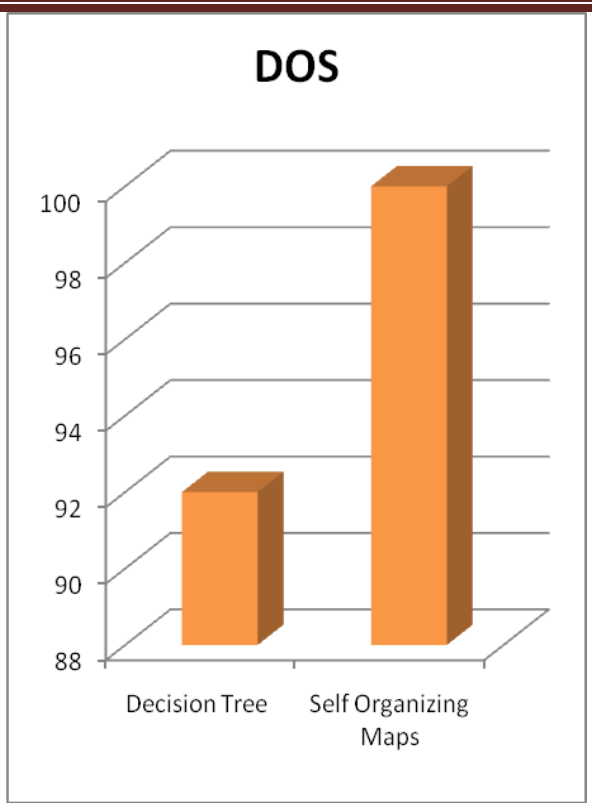
Table: 2. Ii. Different Types Of Attacks

Main Attack Classes	22 different attack types
DOS- Denial of Service	back, land, neptune, pod, smurf, teardrop
U2R- User to Root	buffer_overflow, loadmodule, perl, rootkit
R2L- Remote to User	ftp_write, guess_password, imap, multihop, phf, spy
Probe	ipsweep, nmap, portsweep, satan

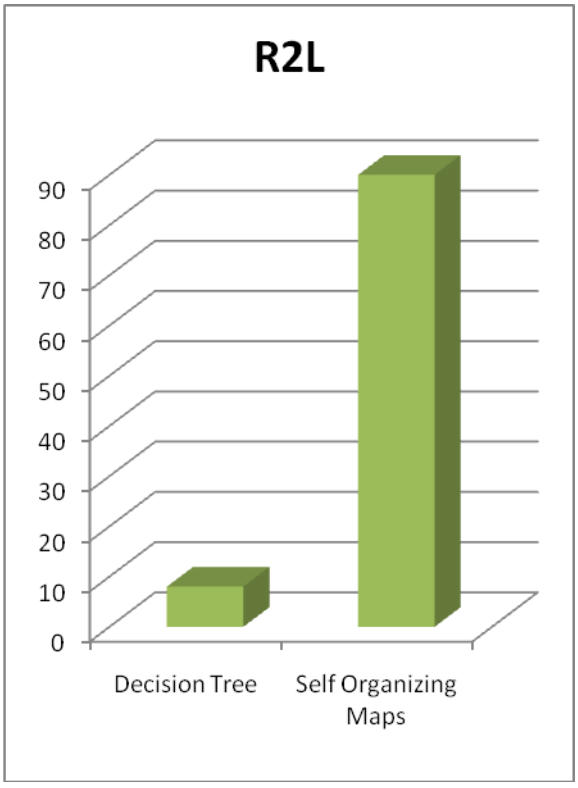


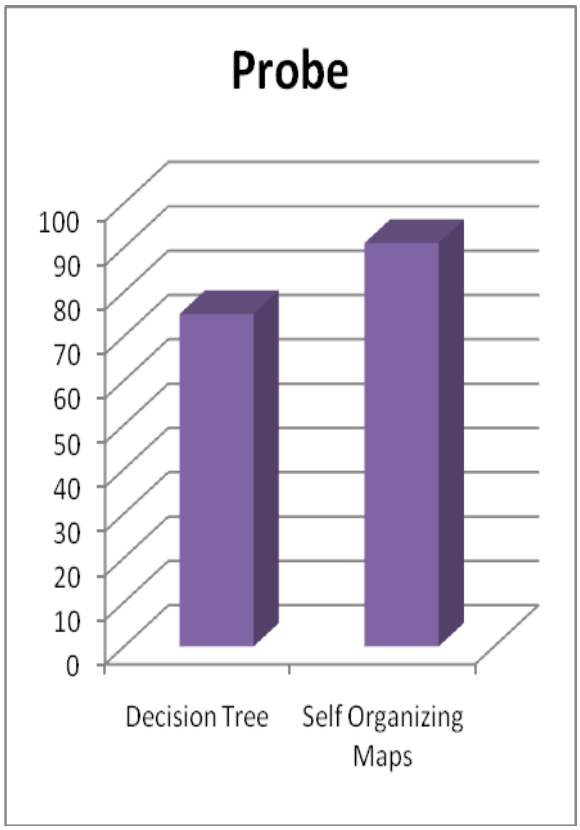
ISBN 978-93-5156-328-0





(a) Normal with 41 Attributes (b) U2R Attacks with 41 Attributes





(e) Probe Attacks with 41 attributes

8. Conclusion

Intrusion Detection System is one of the most important security systems to detect intrusions in a variety of networks in a distributed environment. Here, we are doing a comparative study on Intrusion Detection System based on Artificial Intelligence techniques. The main techniques which are discussed here are Decision Trees, and Self-Organizing Maps (SOM). We are describing these techniques and determining how these techniques aid in detecting intrusions in a networking environment and which is more suitable for intrusion detection. In Decision Tree, Most of the algorithms require that the target attribute will have only discrete values. As decision trees use the "divide and conquer" method, they tend to perform well if a few highly relevant attributes exist, but less so if many complex interactions are present. The greedy characteristic of decision trees leads to another disadvantage that should be pointed out. This is its over–sensitivity to the training set, to irrelevant attributes and to noise. But in Self-Organizing Maps, it is the time consuming algorithm. This is because as the number of neurons affects the performance of the algorithm. And as the number increases the computation increases which results in increasing computational time. As SOM is time consuming even though it is most suitable for the intrusion detection then decision tree.

References

[1] J. R. Quinlan, Induction of Decision Trees. Machine Learning, 1:81-106, 1986
[2] Manish Kumar, M. Hanumanthappa, Intrusion Detection System Using Decision Tree Algorithm, 2012
[3] S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica 31(2007) 249-268, 2007
[4] V. K. Pachghare, Intrusion Detection System Using Self Organizing Maps, 2009
[5] P. Lichodzijewski, A. Zincir-Heywood, M. Heywood, "Dynamic intrusion detection using self-organizing maps, 2002
[6] McHugh, J.: Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory. ACM Trans. on Information and System Security 3 (2000) 262-294.
[7] Aikaterini Mitrokotsa- Detection Denial of Service Attacks Using Emergent Self Organizing Maps, Greece, 2005
[8] Tao Xi, An Intrusion Detection Approach Inspired by biological memory cells, China, 2012
[9] A Behavior Based Approach to Host-Level Intrusion Detection using self-organizing maps, India, 2008
[10] Sandhya Peddabachigari, Ajith Abraham, Johnson Thomas, Intrusion Detection Systems Using Decision Trees and Support Vector Machines (2004), Vector Machines, International Journal of Applied Science and Computations
[11] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993
[12] Christine Dartigue, A New Data-Mining Based Approach for Network Intrusion Detection, USA, 2009
[13] Cheung-Leung Lui, Ting Yee Cheung, Agent based network intrusion detection System using data mining approaches, Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05) © 2005 IEEE
[14] Norbik Bashah Idrisand Bharanidlran Shanmugam, Artificial Intelligence Techniques Applied to Intrusion Detection
[15] L. Brieman, J. Friedman, R. Olshen, C. Stone, Classification of Regression Trees. Wadsworth Inc., 1984
[16] The KDD Archive. KDD99 cup dataset, 1999
[17] JuhaVesanto, Johan Himberg, EsaAlhoniemi, Juha Parhankangas, Self-organizing map in Matlab: the SOM Toolbox"Laboratory of Computer and Information Science, Helsinki University of Technology, Finland (1999).
[18] Shyam M. Guthikonda, Kohonen Self-Organizing Maps, 2005
[19] A M Chandrashekhar, Intrusion detection technique by using K-means, Fuzzy Neural Network and SMV classifier, India