iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** VII **Month of publication:** July 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Robust Intrusion Detection System by Utilizing Support Vector Machine and Error Back Propagation Neural Network

Gaurav Soni[1], Rachna Trivedi[2]
[1]*Assistant Professor, TIT Science, Bhopal,*
[2]*MTECH Scholar Computer Science & Engineering, TIT Science, Bhopal*

Abstract: ***The concept of Intrusion Detection System is used in the work. The data set is used for training and testing. Various numeric features of dataset are selected for better accuracy.SVM that is Support Vector Machine is trained for classifying normal and intruded sessions in the dataset. The work is tested in various parameters like Accuracy, Recall, precision and F measure. Once the Intruded sessions are found, EBPNN that is Error Back Propagation Neural network is Trained and Tested for the type of intrusion they are DOS, R2l, U2R and Probe. The Accuracy is tested in second module also on the Basis of Recall, Precision, and F measure***
Keywords Intrusion Detection System, EBPNN, SVM, Network Security, Supervised Learning, Neural Network

I. INTRODUCTION

An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces electronic reports to a management station.
Types of Intrusion Detection Systems (IDS)
Intrusion detection systems (IDS) can be classified into different ways.
The major classifications are
Active and passive IDS,
Network Intrusion detection systems (NIDS) and host Intrusion detection systems (HIDS)

A. Types of attack

Data packets are sent from the attacker nodes to the victim node or nodes. Attacks are generated randomly using a random function. The type of attack generated is classified to be a Probe, R2L, U2R or DOS attack.

1) *Denial-of-service (DOS)* attack is an attempt to make a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. Probe-Relative to computer security in a network, a probe is an attempt to gain access to a computer and its files through a known or probable weak point in the computer system.
2) *U2R Attack:* User to Root exploits are a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system
3) *R2L Attack:* A Remote to User attack occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.

B. Support Vector Machine (SVM)

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, and then the data are transformed in such a way that the separator could be drawn as a hyper plane. Following this, characteristics of new data can be used to predict the group to which a new record should belong.

Error back propagation neural network

Back propagation,

An abbreviation for "backward propagation of errors", is a common method of training artificial neural networks used in conjunction with an optimization method such as gradient descent..The method calculates the gradient of a loss function with respect to all the weights in the network.

The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function**.**

Back propagation requires a known, desired output for each input value in order to calculate the loss function gradient.

II. RELATED WORK

A. Machine learning techniques for intrusion detection system

Intrusion detection is considered as one of the foremost research areas in network security, the challenge is recognize unusual access that could lead to compromising the interconnected nodes. Anomaly-based intrusion detection system, that utilizes machine learning techniques such as single classifier and hybrid classifier have the capability to recognize unpredicted malevolent. In this paper, we examine different machine learning techniques that have been proposed for detecting intrusion by focusing on the hybrid classifier algorithms. The objective is to determine their strengths and weaknesses. From the comparison, we hope to identify the gap for developing an efficient intrusion detection system that is yet to be researched.

B. Artificial Neural Network based System for Intrusion Detection using Clustering on Different Feature Selection

Intrusion Detection System (IDS) is an example of exploitation Detection System that works for detecting malicious attacks. This can be described as software for security management. Many researchers have proposed the Intrusion Detection System with various techniques to achieve the best accuracy. In this paper it is projected that intrusion detection system with the combination of k-means clustering and artificial neural network to improve the system. To obtain a better answer benchmark dataset was divided into training and testing part and then cluster the dataset into five different divisions

After appliance these functions we have proposed a comparative analysis between them and choose the best accuracy rate among them. Here, it has been verified that, using the clustering technique a better accuracy rate can be found that improve the system with the perfect neural network functions which is the probabilistic neural network. It is also important to select efficient feature sets for better accuracy.

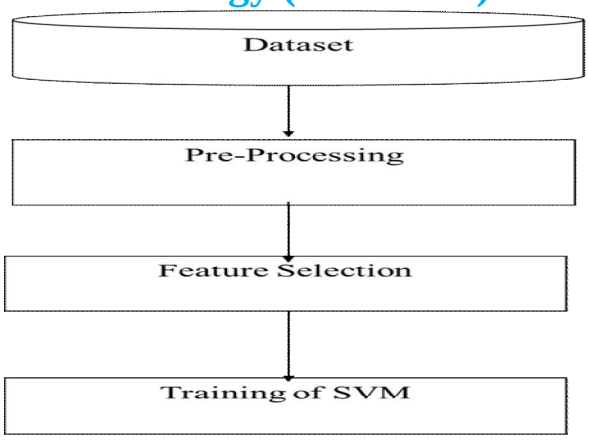
III. PROPOSED METHODOLOGY

In order to research cyber-attack techniques, data needs to be collected. It is important to collect data reliably in order to provide the best data analysis.

A. Proposed Work

Whole work is divide into two modules. First is separation input data into two class of safe and unsafe session from the dataset by using SVM (Support Vector Machine).

Then in second module identification of type of intrusion was done in unsafe network this identification process is done by Error Back Propagation Neural Network.



FLOW CHART- proposed work first module

Data Set

In this work intrusion detection was done on the very famous dataset known as KDD99

This dataset is used by different researcher for various purposes

In KDD99 dataset total 494,021 number of session present in which about 97,277 number of session are normal of condition where rest of sessions 391,458 are attack one.

In this percentage of various attacks are 79.24% DOS, 0.23% R2L, 0.01% U2R, 0.83% Probe.

Whole dataset consist of total 43 attribute which describe various features of the network packet transferring from sender to receiver.

Pre-Processing: Here information likes type of protocol, socket type, etc. are removed. As presence of these information increases confusion for the SVM training.

This can be understand as let raw data session is

{0,tcp,htp_data,,421,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,2,0.00,0.00,0.10,0.00,1.00,0.00,0.00,120,25,0.12,0.03,0.15,0.00,0.00,0.12,0.05,0.00,normal,20}

After applying pre-processing session Dataset will be

{1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,2,0.00,0.00,0.10,0.00,1.00,0.00,0.00,120,25,0.12,0.03,0.15,0.00,0.00,0.12,0.05,0.00,normal }

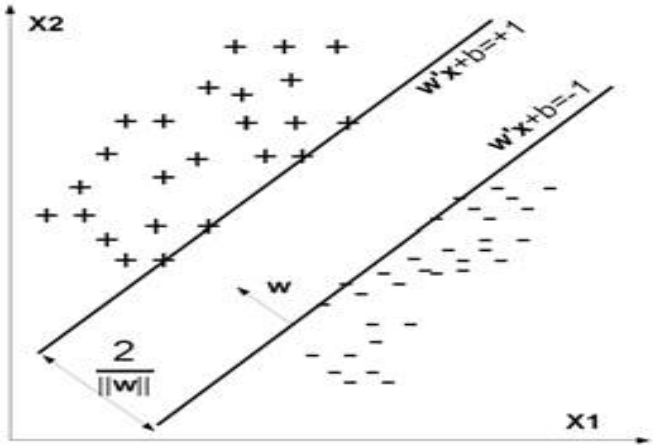
Feature Selection: In this step of first module features present in the session are identified and store in a matrix.

Separate matrix is prepared for safe and unsafe mode. This can be understood as the feature matrix of safe and unsafe have two features from each session:

F11:{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,150,25,0.17,0.03,0.17,0.00,0.00,0.00,0.05,0.00}

F12: {normal}

So each feature matrix has two columns, where first present numeric values and other present its type.



GRAPH-Training data samples for two different classes are denoted by + and - signs

Session Separation:

In this step of first module trained SVM is used for the separation of session in two category first is safe mode and other is unsafe mode. So session from the testing dataset is pass one by one in the trained SVM.

Output of the SVM is two class + or -. So first evaluation is required for the correct identification of session of their respected category.

Proposed Algorithm

Input: KDD

Output: SVM (Trained Support Vector Machine)

DS←Pre-Process(KDD) // DS (Filter Dataset

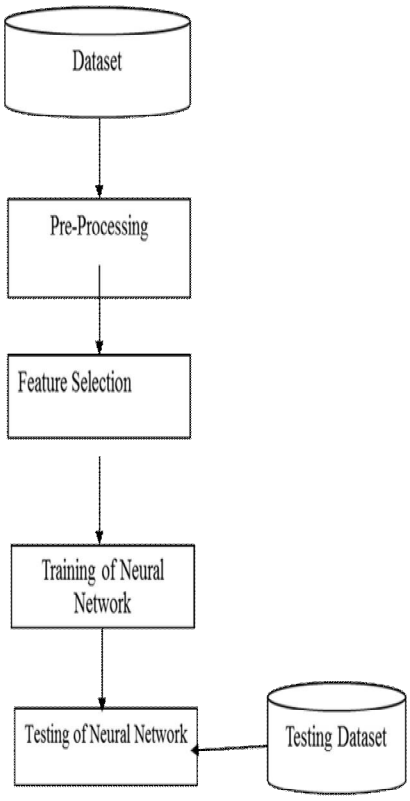
F[n, c]←Feature(DS) // F (feature), n (Total Session), c (safe and unsafe state of session)

Loop 1: n SVM←Train_SVM(F[n, c])

End Loop

Module 2. Error Back Propagation Neural Network

In this module dataset of attacked sessions are utilize to train the EBPNN. Then trained dataset is utilized for the testing of unknown attack session.



FLOW CHART-Proposed work second module.

Pre-Processing: Here removal of useless information in dataset is done in this step. This is same as filtering of pre-processing done in first module.

Feature Collection: Here as dataset contain only attack sessions, but each contain variety of attacks so separation of each type of

attack session in there group is done in this step.

Training of Neural Network:

In this step features gained after pre-processing is put in Error Back Propagation neural network. As each input vector is already present in segmented form, where position of the values presents various network feature. Let us assume an input vector V = [f1, f2, f3……..fn] now for each input vector its corresponding class is identified by a unique number set of output vector consist of those unique number values. So output vector is like [1, 2, 3, ……. .5] and its segment is Xi. So this vector act as the input to the neural network while Xi act as the desired output

Proposed Algorithm

Input: KDD, I (Number of Iteration)

Output:EBPNN(Error Back Propagation Neural Network)

DS←Pre-Process(KDD) // DS (Filter Dataset

2 F[n, c]←Feature(DS) // F (feature), n (Total Session) c(class of intrusion)

3 Loop 1:I

4 Loop 1:n

5 EBPNN←Train_EBPNN(F[n c])

6 End Loop

7 End Loop

IV. EXPERIMENTS AND RESULTS

A. *MATLAB Toolbox*

In order to implement above algorithm for intrusion detection system MATLAB is used, where dataset is used of different size.

Neural Network Toolbox includes command-line functions and apps for creating, training, and simulating neural networks.

This makes it easy to develop neural networks for tasks such as data-fitting, pattern recognition, and clustering. After creating networks in these tools, it can automatically generate MATLAB code to capture work and automate tasks.

3 Evaluation Parameter

To test our result this work uses following measures the accuracy , Precision, Recall and F-score. These parameters are depending on the TP, TN, FP and FN.

$$Precision = \frac{True\_Positive}{True\_Positive + False\_Positive}$$

$$Recall = \frac{True\_Positive}{True\_Positive + False\_Negative}$$

$$F\_Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

In order to make the better evaluation for this work one more parameter has introduced that is accuracy of the class of the intrusion. Accuracy of the work is calculated by:

Accuracy = (true positives +false negatives)/(Total_Normal + Total_Intrusion)

Results

Data-Set Size	SVM
	Accuracy
3000	99.967
5000	99.86
8000	98.662
9000	99.567

From above table it is obtained that with the increase in dataset size accuracy rate remain above 99% which is quite a big achievement. This is because SVM can perfectly make two class partitions. Slight decrease in rate was done because of types of data increases so confusion for SVM also increases. As number of patterns is more in the dataset so results are more accurate after a consistent data type.

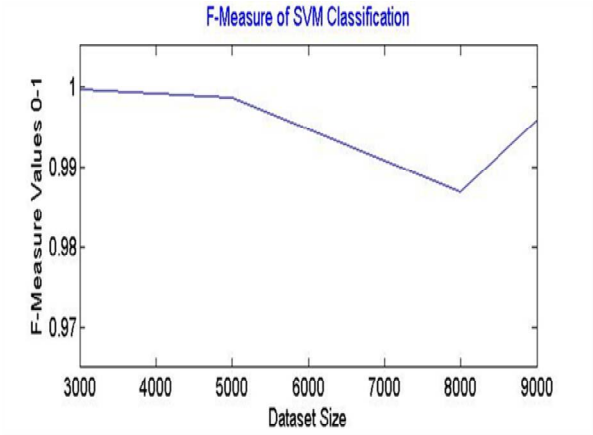
Data-Set Size	SVM		
	Precision	Recall	F-Measure
3000	0.99932	1	0.99966
6000	0.99722	1	0.99861
8000	0.97598	0.99801	0.98687
9000	0.99316	0.99849	0.99582

SVM Precision, Recall, F-measure values at different Dataset Sizes.

Volume 4 Issue VII, July 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

SVM Based Two Class Partition		
Data-Set Size	Actual	Proposed Work
3000	1468	1532
6000	2513	2487
8000	4120	3880
9000	4675	4325

SVM based two class results.



. F-Measure comparison of SVM at different dataset size.

Results of 8000 Dataset size			
Attacks	Actual	Proposed Work	Accuracy
DOS	2826	2826	100
Probe	786	786	100
R2L	214	213	99.5327
U2R	53	53	100

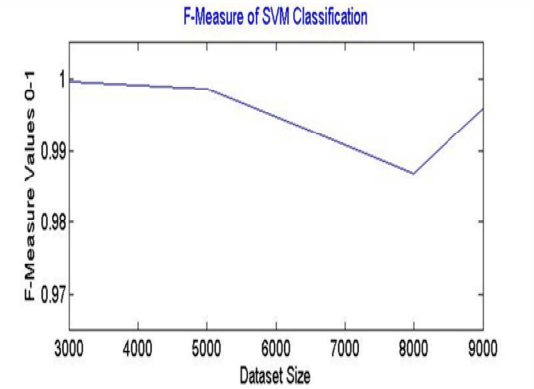
Represent detection of different attack from the proposed work.

It is seen that proposed work has achieved 100 % accuracy at different dataset size which is highly acceptable. While proposed work well in case of R2L and U2R attack are also giving good result .So overall accuracy for identification of type of attack is high and appreciable.

Results Obtained From EBPNN

Results of 9000 Dataset size			
Attacks	Actual	Proposed Work	Accuracy
DOS	3159	3159	100
Probe	890	890	100
R2L	222	221	99.5495
U2R	53	53	100

Represent detection of different attack from the proposed work.



. F-Measure comparison of SVM at different dataset size.

Results Obtained From EBPNN

Results of 9000 Dataset size			
Attacks	Actual	Proposed Work	Accuracy
DOS	3159	3159	100
Probe	890	890	100
R2L	222	221	99.5495
U2R	53	53	100

Represent detection of different attack from the proposed work.

V. CONCLUSIONS AND FUTURE SCOPE

A. Conclusion

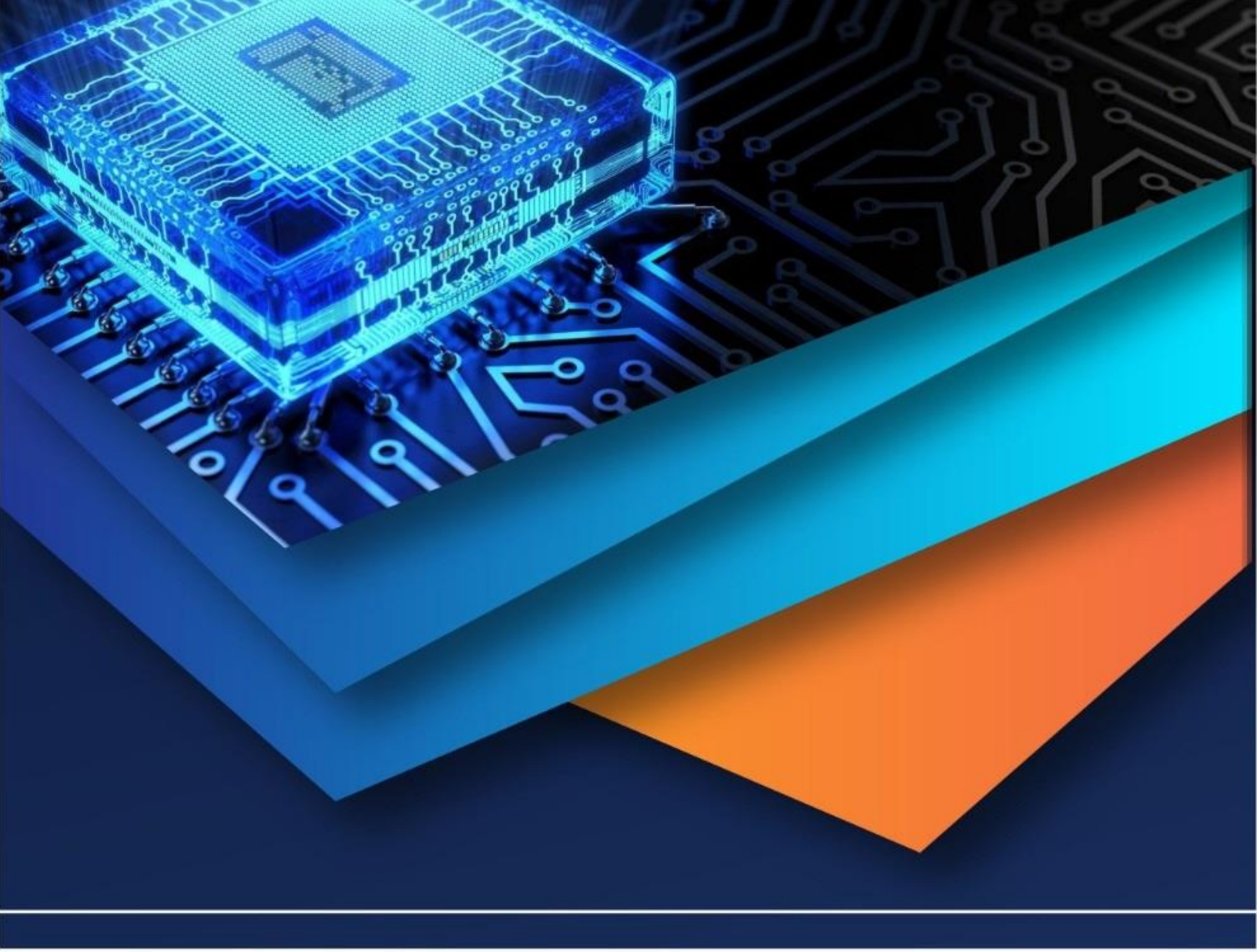
In this work solving the computer networking problem to give the specific way to prevent the security problems intrusion detection system (IPS) has a main techniques to secure the computer networks problems. Many way to check my types of attracters problems on the network concepts such as DOS (Denial of service), (R2L) Remote to local, (U2R) User to remote, Probe etc. In this work a combination of SVM and Feed Forward neural network is done for the detection of intrusion in network. Results obtain are highly appreciable as trained networks identify all type of intrusions more than 99.9% of accuracy. Here probe, Dos, U2R type of attacks are 100% detected, while R2L attack is 99.8% detected

B. Future Scope

As In future it need to be improved by putting data on the unsupervised network, so it automatically update the new behavior of the intruder. In future as this work utilizes only KDD'99 dataset, while there are other dataset as well for learning the feature and detect different intrusion. In the future, attempt can be made to add implementation of some more algorithms and techniques. A numeric comparison also needs to be done between the existing method and proposed method to illustrate the advantages of proposed system over existing system.

REFERENCES

[1] https://en.wikipedia.org/wiki/Support_vector_machine
[2] https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
[3] BBC Analysis reveals popular adobe password. [http://www.bbc.co.uk/ news/technology-24821528; accessed Nov-2013].
[4] Better Host Review Distributed denial-of-service attack image. [Http: //www.betterhostreview.com/tag/DDOS-attack-protected-hosting; accessed 10-Oct-2013].
[5] Callegati, F., Cerroni, W., and Ramilli, M. Man-in-the-Middle Attack to the HTTPS Protocol. Security & Privacy, IEEE, 7(1):78
[6] Cavallaro, L. (2013). Malicious Software and its Underground Economy: Two Sides to Every Story.
[7] [https://www.coursera.org/course/malsoftware; accessed April-2013]. 44, 47
[8] Chamales, G. The honeywall CD-ROM. Security & Privacy, IEEE, 2(2):77
[9] Corporation, A. Jpgraph. [http://jpgraph.net/; accessed 13-Feb-2013].
[10] DiMino,A.M. Another look at a cross-platform DDOS botnet. [http://sempersecurus.blogspot.co.uk/2013/12/ another-look-at-cross-platform-DDOS.html; accessed Feb-2014]. (2014)
[11] Dittrich, D. The ethics of social honey pots. Available at SSRN 2184997. (2012)
[12] F-Secure[http://www.f-secure.com/en/web/home_gb/home; accessed Nov-2013].
[13] Fettig, A. and Lefkowitz, G. Twisted network programming essentials. O'Reilly Media, Inc. 6 (2005)
[14] Fu, X., Yu, W., Cheng, D., Tan, X., Stre , K., and Graham, S. (On recognizing virtual honey pots and countermeasures. In Dependable, Autonomic and Secure Computing, 2nd IEEE International Symposium on, pages 211{218. IEEE. 6 (2006).
[15] Futex Analysis of a Linux malware (2014).
[16] Fyodor), G. L. [http://nmap.org/; accessed Nov-2013]. 28
[17] GoDaddy Operating Company,
[18] Gooneys, J. M. Top 100 adobe passwords with count. [http:// stricture-group.com/files/adobe-top100.txt; accessed Nov-2013]. vii,
[19] Halliday, J. samplesshd-tty.c. [https://github.com/substack/libssh/ blob/master/examples/samplesshd-tty.c; accessed Oct-2013].
[20] Zhang, V. (2013). Trojanized appy bird comes on the heels of takedown by app creator. Accessed Nov-2013].
[21] Y.K.Jain, S. Singh "Honey pot based Secure
[22] R.Baumann, C.Plattner "honey pots" Diploma Thesis in Computer Science, 2002



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY