Roshna R.S,Vinodh Ewards / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 2, March -April 2013, pp.1440-1445

Botnet Detection Using Adaptive Neuro Fuzzy Inference System

Roshna R.S,Vinodh Ewards
Department of computer science ,karunya university ,Coimbatore

ABSTRACT

Botnets are the most serious threat against cyber-security. They provide several services of illegal activities such as denial of service attacks, malware dissemination, phishing, and click fraud without the permission of computer users. Nowadays botnets are the interesting research topic related to cyber threat and cyber crime prevention system. Botnet is collection of compromised computers named as bot,each bot can be considered as malicious software application which can be controlled by remote system from outside network through command and control channel.Most of existing behavior based techniques are not able to detect and predict the botnet as they change their structure and pattern.Here presenting new technique named as botnet detection using Adaptive Neuro Fuzzy Inference System (ANFIS), which can train the system for future prediction .Here we are discussed about different behavior based botnet detection and their draw backs along with new behavior based detection using Anfis algorithm.

Keywords Botnet; bot; P2P; HTTP,patterns; malicious activities,Anfis,Fuzzy

I. INTRODUCTION

Computer networks are critical to modern society. An extensive range of business, infrastructure, and human needs, such as communications, utilities, banks, and leisure services are provided by systems that rely on the secure and efficient operation of networks. Now day's botnet is popular techniques for spreading internet crimes. Botnet is a compromised network of computer called as bots which is remotely controlled by commands send by attacker with the intention of spreading spam's, generating distributed denial of service, information hacking such as making malicious activities. Bots or zombies are software applications which can be easily attached to victim computer through social engineering activities, email attachment, etc. Bot can be categorized in to three main types such as IRC typed bot, HTTP based bot, PEER –PEER based bots.

A botnet, or the army of bots (zombies), is comprised of more than thousands or tens of thousands of compromised computers. Although statistics show that the number of botnets is increasing most Internet users are still unaware of what is going on and how serious the problem is. Many of these users' computers are easily compromised by bot malware and then become members of botnets. Since bot malware usually does not affect regular uses of compromised computers, bot masters or bot herders can control these compromised computers remotely and ask them to carry out malicious activities, such as sending SPAMs, launching distributed denial of service (DDoS) attacks, and stealing personal private information. Here botnet detection is mainly concentrated in Behaviour based detection. Dataset is collected from University of Victoria about 11.15 gb and analysed the packets using wireshark ,and extracted the features for all detection techniques.

II. RELATED WORKS

1.1 Botnet detection using machine learning

Machine learning technique for identifying IRC based botnet traffic. Machine learning technique divides the task in to two sections. (1) Distinguishing between IRC and NON IRC traffic, (2) distinguishing between botnet and real IRC traffic in [3].

Stage 1: identifying chat traffic

Author explore machine learning based classification in identifying IRC chat traffic in three dimension that is the classification scheme, the subset of characteristics features used to describe the flows and the size of training set. For identifying the IRC based chat flows from other traffic author suggest three classification scheme namely such as J48, naive bayes, and Bayesian networks. In J48 classification is based on decision trees and here each internal node represent test on one or more attribute, and each leaf node corresponds to decision outcomes .The bayesian network is used directed acyclic graph, which helps to capture the dependences among samples. Classification of samples is carried out based on this graphical representation of the conditional probability distribution of sample features. In J48 scheme author explored selecting attributes based on how IRC traffic different from other non Chat flow, that is packet involves chat has only small packet size, flow duration compared to other services such as long ftp transfers .

Stage 2: identifying botnet traffic

Here author explores approach to distinguish botnet from legitimate IRC chat flows. They were trying to train machine learning classifiers such as J48, naïve bayes with their tasted traffic traces which collect botnet IRC flows and real IRC flows.

1.2 Fuzzy pattern based filtering algorithm.

Fuzzy pattern based filtering algorithm is based on behavior based botnet detection for all type of botnet. With the help of fuzzy membership function, intended to identify malicious domain names and IP address. In this techniques define membership function for generated failed DNS queries, have similar DNS query interval, generate failed network connections an similar payload size for network connections. The main advantage of fuzzy membership function is that, it can easily altered and modified in order to improve the performance. Fuzzy pattern based filtering algorithm is help to detect human like behavior of botnet in [8].

Algorithm of fuzzy pattern recognition filtering algorithm:

Step 1: Traffic Reduction

Step2: Feature extracted from DNS packets.

Step3: Feature extracted from network flow

Step4: Initialize fuzzy pattern recognition.

In fuzzy pattern recognition ,we classify in to two phases ,named as DNS phase and TCP phase.

For DNS phase,define a feature vector x= (β, r) where α as fixed set that contain n counters is the total number of DNS responses. Where r is the number of failed DNS responses. Define membership function for inactive malicious DNS query. Define membership function for malicious DNS query. Define membership function for normal DNS query.

For The network flow phase, define a feature vector x=(β,r) where β is the total number of DNS responses. Where r is the number of failed DNS responses. Define membership function inactive malicious IP address. Define membership function of malicious IP address. Define membership function of Normal IP address.

Step5: Analyze the result and detecting botnet.

III. THE PROPOSED ANFIS PATTERN RECOGNITION FILTERING ALGORITHM FOR BOTNET DETECTION

At a first look ,fuzzy logic can be considered to be worst. There are a lot of reasons of getting such results. First of all, the rule size was limited and corresponds to membership functions given by the developer. A better result is based up on both number of membership functions and rules. The restriction of fuzzy rules and fuzzy sets is due to the ANFIS constraint. The problem was wantto choose the same FIS in both Fuzzy and in ANFIS methods to be able to compare with another. Time taken to learn anfis is very short compare to neural network. so that ANFIS reaches to the target faster than neural network. Anfis is more preferable than neural network for handling complex problem

1.3 Botnet detection using adaptive neuro fuzzy inference system

ANFIS(Adaptive neuro fuzzy inference system)is a kind of neural network, which incorporating the techniques of fuzzy inference system. Both artificial model and fuzzy logic are used in ANFIS. The usage of artificial intelligence has been applied widely in most of the fields of computation studies. Main feature of this concept is the ability of self learning and self-predicting some desired outputs. The learning may be done with a supervised or an unsupervised way. Neural Network study and Fuzzy Logic are the basic areas of artificial intelligence concept. Adaptive Neuro-Fuzzy study combines these two methods and uses the advantages of both methods.

1.4 Anfis (adaptive neuro fuzzy inference system)

ANFIS is an adaptive neuro fuzzy network which allows the usage of neural network topology along with fuzzy logic. It not only includes the characteristics of both methods, but also avoids disadvantages of both fuzzy logic and artificial neural network. ANFIS combines both neural network and fuzzy logic, it is capable of handling complex problems. Even if the targets are not given, ANFIS may reach the optimum result rapidly.

1.5 Problem statement

Goal of the proposed project is to identify malicious DNS and IP address. Identified domains have to classify active and inactive bot. Active domain name and IP address can be easily identified because they started to contact C& C server and download commands, and send feedback etc. But in the case of inactive domain names, they are previously contacted C & C server but they currently inactive for some time.

To achieve this goal, the proposed project target on three sub problem.

1.5.1 Traffic reduction

To reduce the input traffic to the proposed system, have to filter irrelevant traffic by all eliminating packet except DNS packet. With the help of traffic reduction algorithm, able to increase the speed of the detection process.

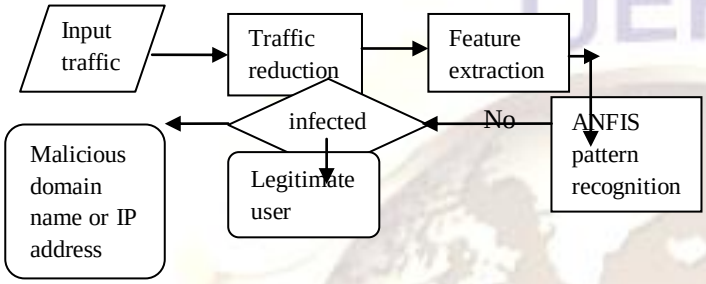
1.5.2 Feature extraction

In this section, identifying the behavior of the botnet from legitimate user has to identify ideal features. This feature is going to use the detection of botnet.

1.5.3 Pattern recognition using ANFIS

Once feature is identified, have to select pattern recognition technique. Pattern recognition technique able to find botnet, based up on the extracted features.

IV. FLOW CHART OF ANFIS PATTERN RECOGNITION ALGORITHM



V. ALGORITHM OF ANFIS PATTERN RECOGNITION FILTER

Step 1: Traffic Reduction
Step2: Feature extracted from DNS packets.
Step3: Feature extracted from network flow
Step4: Initialize anfis pattern recognition.

VI. EXPERIMENTAL SETUP

Implementation detail contains the environment setup, packet capturing, packet filtering, feature extraction, fuzzy pattern recognition.

1.6 Software requirements

1.6.1 Honey pot

In computer terminology, a honeypot is a trap set to detect, deflect, or in some manner counteract attempts at unauthorized use of information systems. Generally it consists of a computer, data, or a network site that appears to be part of a network, but is actually isolated and monitored, and which seems to contain information or a resource of value to attackers.

VII. IRC server

Internet Relay Chat (IRC) is a protocol for real-time Internet text messaging (chat) or synchronous conferencing. It is mainly designed for group communication in discussion forums, called channels, but also allows one-to-one communication via private message as well as chat and data transfer, including file sharing.

VIII. Wireshark

Wireshark allows the user to put network interface controllers that support promiscuous mode into that mode, in order to see all traffic visible on that interface, not just traffic addressed to one of the interface's configured addresses and broadcast/multicast traffic. However, when capturing with a packet analyzer in promiscuous mode on a port on a network switch, not all of the traffic traveling through the switch will necessarily be sent to the port on which the capture is being done, so capturing in promiscuous mode will not necessarily be sufficient to see all traffic on the network. Port mirroring or various network taps extend capture to any point on net; simple passive taps are extremely resistant to malware tampering.

IX. Matlab

MATLAB is an on-line system providing machine aid for the mechanical symbolic processes encountered in analysis. It is capable of performing, automatically and symbolically, such common procedures as simplification, substitution, differentiation, polynomial factorization, indefinite integration, direct and inverse Laplace transforms, the solution of linear differential equations with constant coefficients, the solution of simultaneous linear equations, and the inversion of matrices. It also supplies fairly elaborate bookkeeping facilities appropriate to its on-line operation

X. MySQL

MySQL is primarily an RDBMS and ships with no GUI tools to administer MySQL databases or manage data contained within the databases. Users may use the included command linetools, or use MySQL "front-ends", desktop software and web applications that create and manage MySQL databases, build database structures, back up data, inspect status, and work with data records

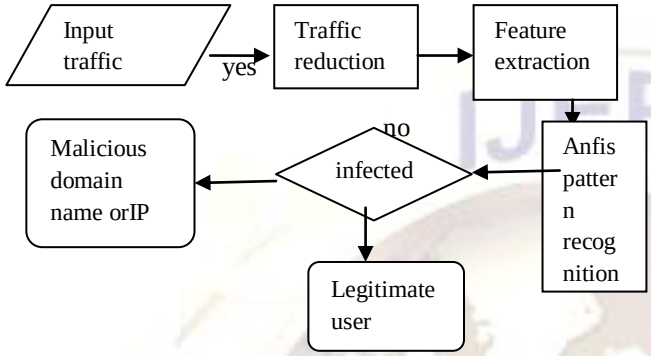
1.7 Hardware requirements

1. Operating system
   Ubuntu 12.04 desktop
   Ubuntu 11.04 desktop
   Ubuntu 12.04 server
   Windows XP
2. Processor: 2.56 ghz dual core
3. Ram:4GB

1.8 Environment setup

In workstation installed with ubuntu 12.04 and created kernel based virtual machine environment. After installing Kernal based virtual machine in workstation, installed virt-manager. with the help of virt manager and created three virtual box .After creating three virtual box ,tried to install operating systems such as windows xp, and two ubuntu 11.04.In one ubuntu kernel based virtual

machine installed with honeypot to attract attacker. Through honeypot collected the kaiten.c executable code. To execute the kaiten bot from windows XP, installed IRC (internet relay chat) server in one ubuntu12.04 server, and created channel named karunya. And executed kaiten bot with IRC server for one minute. And Captured the bot packet using wireshark ,then stored in MySql database in the ubuntu kernel based virtual machine for further analysis.



Implementation steps

1.9 Traffic reduction

It is common that input raw packet traces contain many different types of packet traces contain many different types of packets. Since most of them are not relevant to botnet detection. They should be filtered out with an accurate and efficient traffic reduction algorithm. It enables a botnet detection system to run in a more efficient way

With the help of program "wireshark" able to filter DNS packets. And stored in a file .pcap format.

1.10 Feature extraction

Bots activities often start with DNS queries. If the domain name of a Command &Control server cannot be resolved or the resolved IP addresses are unreachable (offline hosts or invalid IP addresses), the bot is inactive.If one of the contacting IP addresses is valid and a bot is able to comunicate to the Command &Control server, it is an active bot. Bots can be classified in to two types, active and inactive bots. It is not difficult to distinguish active and inactive bots. An active bot is always able to establish connections with one Command &Control server. On the other hand, an inactive bot could receive a number of DNS failure messages and it is not able to reach a Command &Control server. Therefore, extract features from DNS queries and network flows and then the extracted features are used to detect Command &Control server addresses.

1.11 Feature extracted from DNS packets

Bot usually operate with particular behavior. Some of the behavior is distinguishable from normal behavior and hence features of the behavior can be extracted to detect bots.In the DNS phase, for each identified domain name, define a feature vector x = (β,γ) for the domain name,Where.β is the total number of DNS responses,γ is the number of failed DNS responses.

In this phase, define the following three states and each state has its own membership function:Inactive malicious DNS query assume that a DNS query about an inactive malicious domain name usually gets a failed DNS response. Therefore, more failed DNS responses should lead to a higher membership value. Based on the observation, define a membership function X1 which is used to calculate the probability of being an inactive malicious DNS query. The function X1 is defined as

$$X1(x) = 1 - (\beta - \gamma)/\beta \quad (1)$$

- Since malicious DNS queries usually have similar time interval. If most DNS queries for an identified domain name have similar time intervals, it could be a malicious domain name. Define a membership function X_2 to calculate probability of contacting a malicious domain name

$$X2(x = \begin{cases} \frac{\max(\alpha)}{\sum\alpha}, & \sum\alpha > \rho \\ 0, & otherwise \end{cases} \quad 2)$$

- Normal DNS query define a membership function X3 to calculate the probability of being a normal DNS query. If an identified domain name has no failed DNS response, low query frequency, and diverse time intervals, it would be a benign domain name.

$X_3(x) = 1-\max\{X_1(x), X_2(x)\}$ (3)

10.6 Feature extracted from TCP packets

In network flow phase, define a feature vector x = (α, β, r) for each destination IP address identified in network flows, where the maximum time interval between a request and its response is less than n seconds. Define a as a fixed size set that contains n counters a = {a1, a2, . . . ,an}. Each counter in a has an initial value of zero. Given a segment of a network trace containing m request-response pairs, the time intervals between a request and the corresponding response can be measured and then form a sequence S= {s1, s2, . . . , sm}. b is the total number of network requests.

The maximum payload size is less than b bytes. We define c as a fixed size set that contains b + 1 counters , c = {r0, r1, r2, . . . , rb}. Each counter in c has an initial value of zero. Given a segment of a network trace containing t network flows, the payload size of each network flow is extracted and form a sequence P = {p1, p2, . . . , pt}.

In this phase, define the following three states and their corresponding membership functions:

- Inactive malicious IP address , assume that if an IP address receives many requests but does not respond, it is highly probable that the destination IP address is an inactive malicious IP address. Define a membership function X1 to calculate the probability of being an inactive malicious IP address: In the equation, ρ is a threshold for the number of retries. When a destination IP address has been reconnected for more than ρ times, the destination IP address is treated as malicious.

$$X_1 = \begin{cases} 1, & \sum\alpha = 0 \; and \; \beta \geq \rho \\ 0, & otherwise \end{cases} \quad (4)$$

- Malicious IP address since computers with malicious IP addresses, provide the some commands to bots, it can be observed that connections to these malicious IP addresses would have similar payload sizes. Without counting a payload size of zero, define a membership function X2 to calculate the probability of being a malicious IP address: X2(x) =

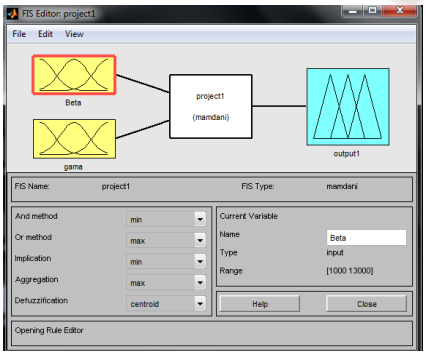
$$\begin{cases} \frac{\max(\gamma)}{\beta - r_0}, & \beta - r_0 \geq \rho \\ 0, & otherwise \end{cases} \quad (5)$$

- Normal IP address defines a membership function X3 to calculate the probability of being a normal IP address. If a destination IP address has no failed network flows and the payload sizes are diverse, it would be a benign address. Therefore, the function X3 is defined as

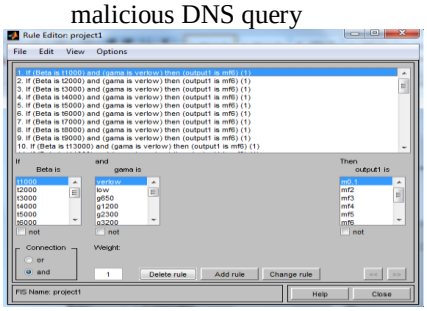
$$X_3(x) = 1\text{-}\max\{X_1(x), X_2(x) \quad (6)$$

XI. RESULTS

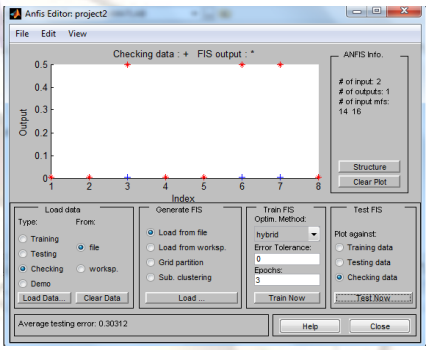
Experimental results show that the proposed Anfis has high detection rates of 95.29% and 95.24% for malicious domain names and malicious IP addresses, respectively. In addition, the Anfis algorithm can detect inactive botnets, which can be used to identify potential vulnerable hosts.



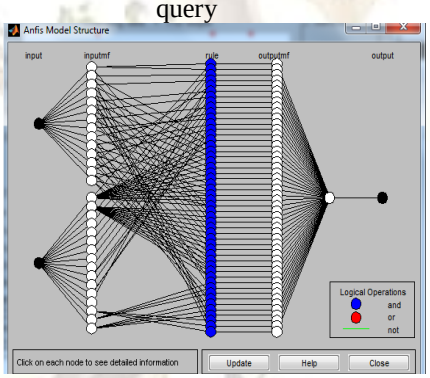
Fuzzy system generated for inactive malicious DNS query



Fuzzy ruleset generated for inactive malicious DNS query

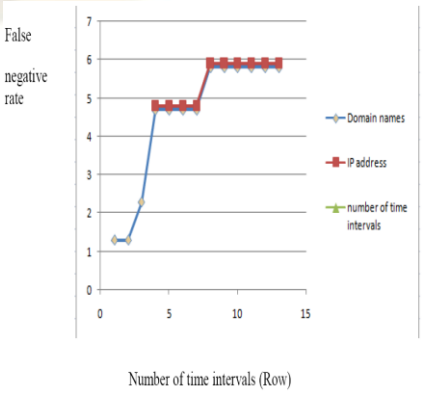


Anfis window for inactive malicious DNS query



Anfis structure for Inactive malicious DNS query

Graph representing detection rate os Anfis pattern recognition algorithm



Malicious Domain names and IP address: false negative rate vs. Row

XII. CONCLUSIONS

Botnets pose a significant and growing threat against cyber-security. Botnet is a collection of internet-connected computers whose security defenses have been breached and control ceded to a malicious party. Each compromised device, known as a bot, is created when a computer is penetrated by software from malicious distribution. The controller of the botnet is able to direct the activities of these compromised computers through communication channels formed by standards-based network protocol such as irc and http, peer-peer. In this paper, we propose an extensible Anfis pattern recognition-based filtering algorithm for botnet detection. Based on common bot host behavior observed from DNS and TCP traffic, our Anfis algorithm is divided into three stages (1) traffic reduction: reduce input raw packet traces and speed up the processing of bots specific activities (2) feature extraction: extract features from the reduced input packet traces and (3) Anfis pattern recognition: with extracted features, detect bot-relevant malicious domain names and IP addresses based on the maximum membership principle. Here use real bots to generate botnet traces to evaluate the proposed Anfis algorithm. Experimental results show that the proposed Anfis has high detection rates of 95.29% and 95.24% for malicious domain names and malicious IP addresses, respectively. In addition, the Anfis algorithm can detect inactive botnets, which can be used to identify potential vulnerable hosts.

REFERENCES

[1] J. Goebel and T. Holz, "Rishi: Identify bot contaminated hosts by irc nickname evaluation," in Proc. 1st Workshop on Hot Topics in Understanding Botnets, 2007.

[2] Sunny Behal, Amanpreet Singh Brar, Krishan Kumar "Signature-based Botnet Detection and Prevention", http: //www.rimtengg.com/ iscet/proceedings /pdfs/advcom p/148.pdf

[3] C. Livadas, R. Walsh, D. Lapsley, W.T. Strayer, Usilng machine learning technliques to identify botnet traffic, in: Proceedings of the 31st *IEEE Conference on Local Computer Networks*, IEEE, pp. 967–974, 2006.

[4] H. Choi, H. Lee, H. Lee, H. Kim, Botnet detection by monitoring group activities in DNS traffic, in: Proceedings of the *7th IEEE International Conference on Computer and Information Technology*, , pp. 715–720, 2007.

[5] S.S.Garasia, D.P.Rana, R.G.Mehta ,"Http Botnet Detection Using Frequent Patternset Mining" in :proceedings of [Ijesat*] International Journal Of Engineering Science & Advanced Technology* Volume-2, Issue-3, 619 – 624

[6] Zhiyong Huang* and Xiaoping Zeng"Detecting and blocking P2P botnets through contact tracing chains"in proceedings of Int. J. *Internet Protocol Technology*, Vol. 5, Nos. 1/2, 2010

[9] Snort IDS web page. http://www.snort.org, March 2006.

[7] Hossein Rouhani Zeidanloo" New Approach for Detection of IRC and P2P Botnets"in proceedings of *International Journal of Computer and Electrical Engineering*, Vol.2, No.6, December,1793-8163,2010

[8] Kuochen Wang, Chun-Ying Huang" A fuzzy pattern-base filtering algorithm for botnet detectionin proceedings of *Comput. Netw.* (2011), doi: 10.1016 /j.comnet. 2011.0 .026

[9] Snort IDS web page. http://www.snort.org, March 2006.

[10] J.R.Binkley and S.Singh,"An algorithm for anomaly-based botnet detection," in Proc. *USENIX Steps to Reducing Unwanted Traffic on theInternet Workshop* (SRUTI'06), , 2006, pp 43–48.

[11] G. Gu, J. Zhang, and W. Lee, "Botsniffer: Detecting botnet command and control channels in network traffic," in Proc. *15th Annual Network And distributed System Security Symposium* (NDSS'08), 2008.

[12] A. Karasaridis, B. Rexroad, and D. Hoeflin, "Wide-scale botnet detection and characterization," in Proc. 1st Workshop on *Hot Topics in Understanding Botnets*, 2007

[13] D. Dagon, "Botnet Detection and Response, *The Network is the Infection*," in OARC Workshop, 2005

[14] J. Kristoff, "Botnets," in 32nd Meeting of the North American Network Operators Group, 2004.

[15] A.Schonewille and D.J.van Helmond. "The Domain Name Service as an IDS,"Master's Project, University of Amsterdam, Netherlands, Feb 2006, http: //staff.science.uva.nl /~delaat/snb-2005-2006/p12/report.pdf

[16] N. F. A. Ramachandran and D. Dagon, "Revealing botnet membership using dnsbl counter-intelligence," in Proc. 2nd *Workshop on Steps to Reducing Unwanted Traffic on the Internet* (SRUTI '06), 2006.

[17] W. Strayer, D. Lapsley, B. Walsh, and C. Livadas, Botnet Detection Based on Network Behavior, ser. *Advances in Information Security.Springer*, 2008, PP. 1-24.

[18] M. M. Masud, T. Al-khateeb, L. Khan, B. Thuraisingham, K. W. Hamlen, " Flow-based identification of botnet traffic by mining multiple log file," in Proc. *International Conference on Distributed Frameworks & Applications (DFMA)*, Penang, Malaysia, 2008.

[19] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol- and structure independent botnet detection," in Proc. 17th *USENIX Security Symposium*, 2008.

First Author: Roshna R.S ,perusing Mtech in karunya university ,Thamil Nadu

Second Author: Mr.Vinodh Ewards ,Assistant professor ,Karunya university,Thamil Nadu