Scientific Journal of Impact Factor(SJIF): 3.134

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 5, May -2015

Frequent Pattern Reorganization Approach to HTTP Botnet Detection

Binal Panchal[1], Mitesh Thakore[2], Hardik Upadhyay[3]

[1]Computer Department, MEC, Basna
[2]Computer Department, MEC, Basna
[3]Computer Department, GPERI, Mehsana

Abstract- Botnet is the most frightful and malicious threats that usually occurs in the current cyber security. It can be defined as collection of compromised computers that are controlled remotely by hackers or botmaster. Malicious activities of botnet such as launching DDoS (distributed denial of service)attacks, sending spam, Trojan, Phishing emails, information harvesting and click fraud. Recently malicious botnets progresses into HTTP botnets out of typical IRC and P2P botnets. The defining characteristics of botnet are the use of command and control channel through which they can be updated and directed. Data mining techniques and algorithms allow us to automate detecting characteristics from large amount of data, which the conventional heuristics and signature based methods could not apply. One major data mining technique for extracting valuable pattern of the botnet attack is FP-growth algorithm. It aims to design and implement mechanism to detect bot activity. After perform proposed approach we can discovers regularities and irregularities in large amount of data set.

Keywords- Botnet, Botnet Detection, HTTP Botnet, Hping3, Frequent Pattern Mining

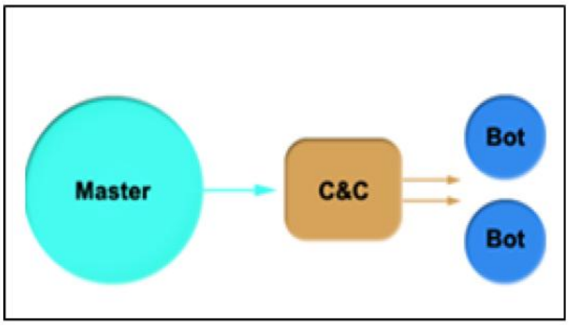
1 INTRODUCTION

Parallel and Distributed computing are widely accepted, due to the improvement and advancement in network bandwidth and computing, so they have been the obviously targeted by attacker [1]. The groups of compromised computers are controlled by one or group of attacker known as "Botmaster" [2]. Botnet operators can use the aggregated power of many bots to exponentially raise the impact of those dangerous activities. A single bot might not be a danger for the Internet, but a network of bots certainly is able to create huge malfunctioning. A study shows that, on a typical day, about 40% of the 800 million computers connected to the internet in a botnet in year 2008 [3]. Communication, resource sharing and curiosity have been great motivators for underground research and hacking. The major attacks under Botnet are DDoS, Scanning, Phishing, Click fraud, spamming [4].

2 THEORETICAL BACKGROUNDS

Once an attack is initiated by a group of computer nodes having different locations controlled by a malicious individual or controller, it may be very hard to trace back to the origin due to the complexity of the Internet [9]. Because of these reasons it has become very serious problem nowadays.

2.1 Command and Control Centre



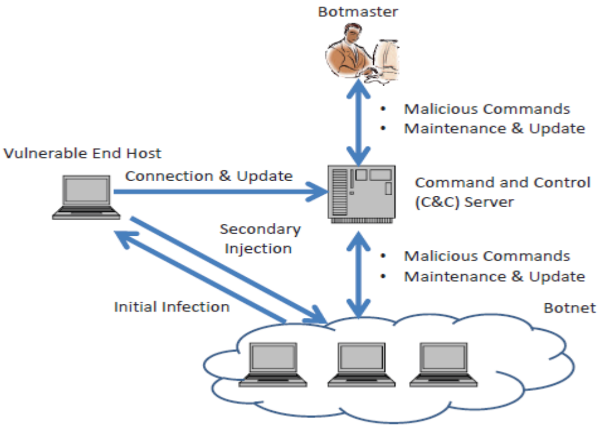
(Fig. 2.1- Botnet Command & Control Architecture)

The backbone of botnet is command and control(C&C) channel that is responsible for setting up the botnet, controlling the activities of the bots, issuing commands, once a C&C channel is detected, and the whole botnet is exposed [10]. The Botmaster computer communicates with its bots by a command and control (C&C) channel, which passes commands from the botmaster to bots, and transmits stolen information from infected machines to their master. Basically botnet activities can be classified as three parts:

1) Searching- Searching for vulnerable and unprotected computers.
2) Dissemination (Infection) - the Bot code is distributed to the computers (targets), so the targets become Bots.
3) sign-on - the Bots connect to BotMaster and become ready to receive command and control traffic.

2.2 Botnet Life Cycle

A typical botnet can be created and maintained in five phases including: 1) Initial infection, 2) Secondary injection, 3) connection, 4) Malicious command and control, 5) Update and maintenance. This life-cycle is depicted in Fig 2.



(Fig.2.2 Botnet Life Cycle [11])

3. LITERATURE REVIEW

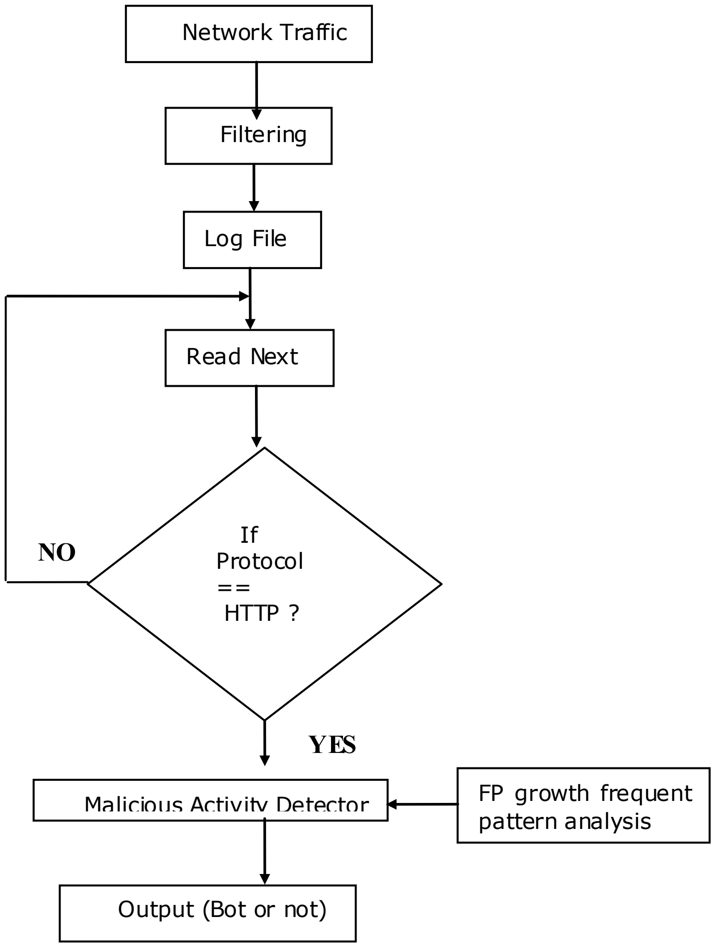
Detection Technique	Proposed Method	Advantage and working of method	Disadvantage
Sajjad Arshad [14]	Cluster Based Data Mining Tech	Better with normal data flow	Hard to work with large database and if large number of white list domains are included
Dae-il jang [21]	Analysis of HTTP2P Botnet: case study Waledac, IEEE-2009	Efficiently detect with normal data traffic	Can’t handle encrypted traffic
Jae-Seo Lee[20]	The Activity Analysis of Malicious HTTP-based botnets using Degree of Periodic Repeatability, IEEE-2008	Overcome existing DNS filtering method.	High false positive ratio as no separation of normal traffic and bot traffic
Geobl and Holz Rishi [16]	n-gram analysis and a scoring system to detect bots that use uncommon communication channels (REGEX)	IRC detection	easily misguided by disguised nicknames, port numbers and cannot detect encrypted as well as non IRC botnet
Masayuki Ohrui [23]	Apriori-prefix span hybrid approach for automated detection of Botnet coordinated attack, IEEE-	Apriori deals with subset of events without considering the order of events	High false positive ratio

	2011		
Roberto Perdiscia et al [19]	network-level behavioral malware clustering system, structural similarities among malicious HTTP traffic	HTTP method like, GET, POST analyzed	HTTP request and response encryption is major limitation
S. S. Garasiya [22]	Botnet Detection with use of Time stamp and frequent pattern set	No need for prior knowledge of botnet such as botnet detection	candidate key generation and false positive ratio

After study and research work in this area we can conclude that Data mining techniques are easily applicable in network flow data.

4. PROPOSED SOLUTION

To detect HTTP Botnet using by analyzing frequent pattern from network logs. We will be using FP growth frequent pattern analysis approach to detect the malicious activity.



(Fig 4.1 Proposed Botnet detection flow chart)

Proposed system contains main four components:

1) Network Traffic: Traffic monitoring is responsible to detect the group of hosts that have similar behavior and communication pattern by inspecting all network traffic. Each flow record has following information: Source IP address, Destination IP address, Source Port, Destination Port, Number of packets transferred in both directions, Time of packet received and transferred to particular IP address.

2) Filtering: Filtering is responsible to filter out irrelevant traffic flows. The main objective of this part is to reduce the traffic workload and makes the rest of the system perform more efficiently. This step will reduce bot irrelevant traffic. If this is not performed well then it can raise false positive rates.
Filter out based on black lists and white lists: If the source or the destination address of a packet is well-known, it is often not necessary to check it. Hence, the packet can be safely ignored.

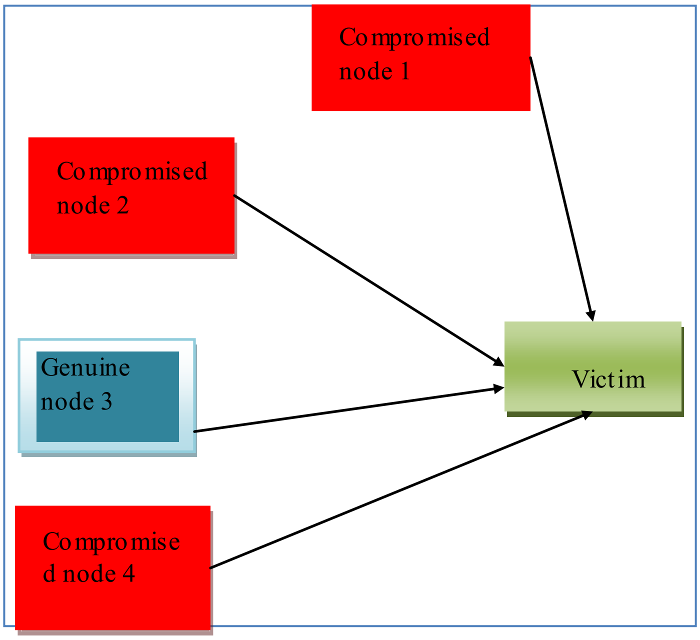
3) Network Traffic Separator: After Filtering network traffic contains many different types of packets, like, TCP, DHCP, Broadcast, Local Network packets etc. Some packets from this traffic are not part of Botnet so they should be taken off from the network traffic. Network Traffic Separator is responsible to separate HTTP traffic from the rest of traffic and sends them to centralized part. Like most network protocols, HTTP uses the client-server model and HTTP protocol.
In the format of HTTP request message, HTTP methods are to be focused. Three common HTTP methods are "GET", "HEAD" or "POST". HTTP bots connects to their C&C periodically. Therefore, the traffic is inspected and if the first few bytes of an HTTP request contain "GET', "POST" or "HEAD"; it's the indication of HTTP protocol and separating those flows and redirect them to Centralized part.

4) Malicious Activity Detector: This part will detect malicious activity generated by Botmaster. Data mining technique is used for extracting suspicious activity. A common approach to detect frequent itemset from the dataset is Apriori. The research proved that Apriori will be time consuming as it scans whole dataset to generate candidate set. So,by using FP – growth frequent pattern analyzing algorithm data as well. By using latest FP –growth approach we can provide good throughput of system in terms of high detection rate with low false positive alarm and in precise time constraints.

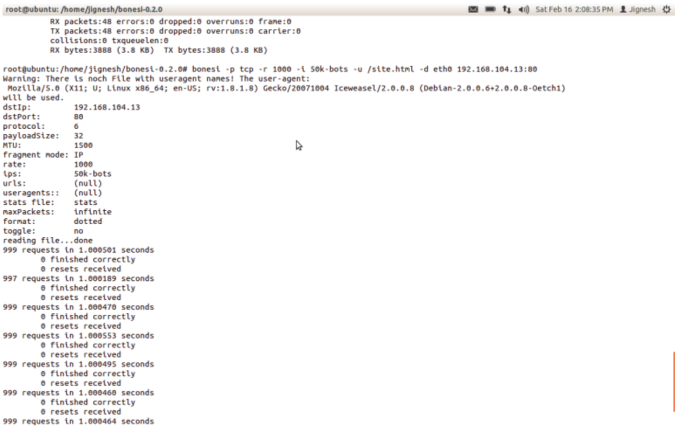
5. IMPLEMENTATION OF PROBLEM AND RESULT ANALYSIS

Implementation is done in two phases:
Phase 1: Attack scenario generation and generation of actual dataset using simulating DoS attack
Phase 2: Filtering and Implementing algorithm to identifying malicious nodes

Our implementation is passive methodology for Botnet Detection. Let us assume that attacker has compromised multiple systems. As main motive of Botnet is to launch a DoS/DDoS attack, Let’s consider following scenario for implementation.



(Fig.5.1- Attack Scenario)

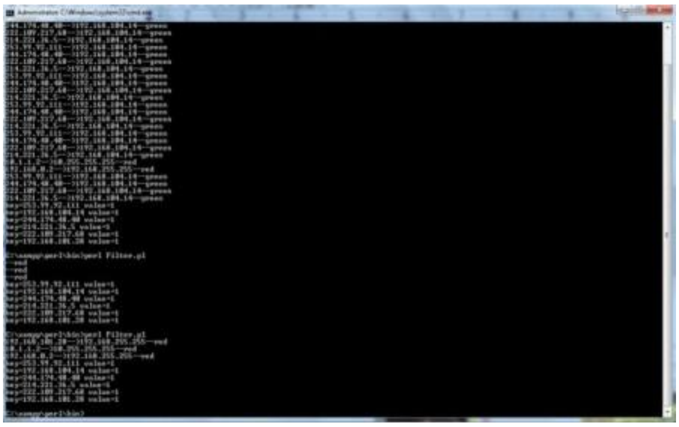


```
root@ubuntu: /home/jignesh/bonesi-0.2.0                    Sat Feb 16 2:08:35 PM  Jignesh
          RX packets:48 errors:0 dropped:0 overruns:0 frame:0
          TX packets:48 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:3888 (3.8 KB)  TX bytes:3888 (3.8 KB)

root@ubuntu:/home/jignesh/bonesi-0.2.0# bonesi -p tcp -r 1000 -i 50k-bots -u /site.html -d eth0 192.168.104.13:80
Warning: There is noch File with useragent names! The user-agent:
 Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.8.1.8) Gecko/20071004 Iceweasel/2.0.0.8 (Debian-2.0.0.6+2.0.0.8-Oetch1)
will be used.
dstIp:          192.168.104.13
dstPort:        80
protocol:       6
payloadSize:    32
MTU:            1500
fragment mode:  IP
rate:           1000
ips:            50k-bots
urls:           (null)
useragents::    (null)
stats file:     stats
maxPackets:     infinite
format:         dotted
toggle:         no
reading file...done
999 requests in 1.000501 seconds
        0 finished correctly
        0 resets received
997 requests in 1.000189 seconds
        0 finished correctly
        0 resets received
999 requests in 1.000470 seconds
        0 finished correctly
        0 resets received
999 requests in 1.000553 seconds
        0 finished correctly
        0 resets received
999 requests in 1.000495 seconds
        0 finished correctly
        0 resets received
999 requests in 1.000460 seconds
        0 finished correctly
        0 resets received
999 requests in 1.000464 seconds
```

(Fig.5.2- Botmaster launching DDoS attack using bonesi 0.2.0)

Now, on victim node wireshark is installed for capturing network traffic. These packet data are irregular like some are broadcast data, other network traffic, genuine traffic which are not our target data. so, thos traffic are removed by filtering process.

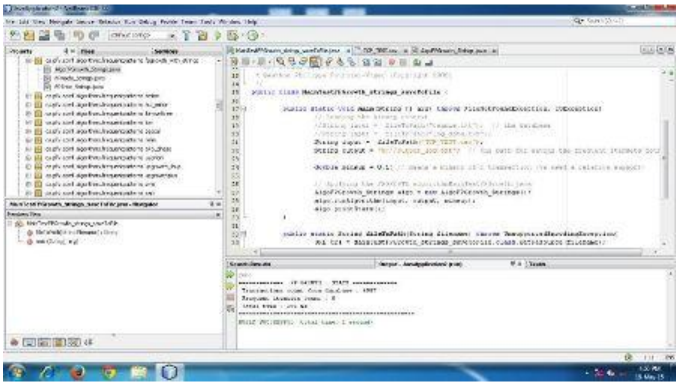


(Fig.5.3 Process of filtering)

After applying Filtering process, we get Red_file including legitimate data which are the input to FP growth algorithm.

FP-growth algorithm execute in NetBeans IDE 7.0. FP-growth algorithm detects frequently occurred items with count and min_sup value.

This min_sup value is change with the no. of records. For large amount of data, we can take min_sup 20%, 25%, 30% as per we required for efficient result. In our approach we take the minimum support is 10% and also takes Packet data up to 4,987 then we can get the result as below:



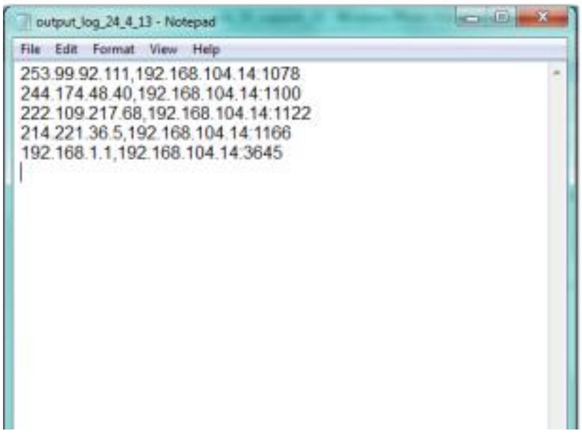
(Fig 5.4 Executing FP- growth with min_sup 10%)

After the execution of algorithm it display the output information like,

……………….FP Growth stats……………….

Transactions count from database: 4987
Frequent Item sets count: 5
Total time: 296 ms

It generates the output file like



```
253.99.92.111,192.168.104.14:1078
244.174.48.40,192.168.104.14:1100
222.109.217.68,192.168.104.14:1122
214.221.36.5,192.168.104.14:1166
192.168.1.1,192.168.104.14:3645
```

(Fig. 5.5 output of Frequent IP with min_sup 10%)

Result Analysis:

Analysis of Frequent occurred IP with different min_sup value,

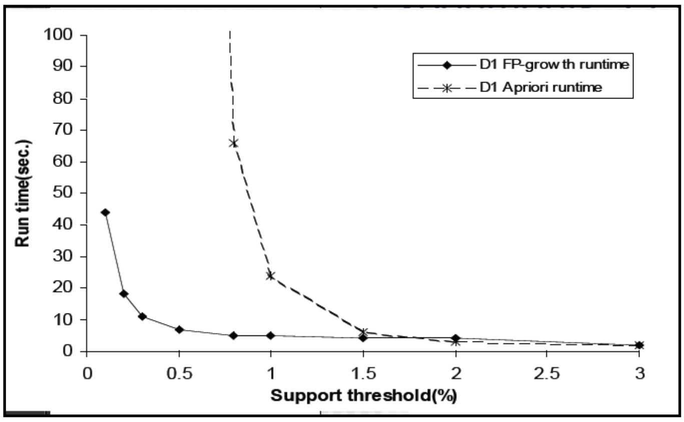
Source_IP	Destination_IP	Count	Min_Sup
253.99.92.111	192.168.104.14	1078	0.1
244.174.48.40	192.168.104.14	1100	
222.109.217.68	192.168.104.14	1122	
214.221.36.5	192.168.104.14	1166	
192.168.1.1	192.168.104.14	3645	
192.168.1.1	192.168.104.14	3645	0.2

After the analysis of result with different min_sup value, we can conclude that 253.99.92.111, 244.174.48.40, 222.109.217.68, 214.221.36.5 and 192.168.1.1 are the malicious IP address.
So, to protect our system from this malicious server we can bypass these traffic or IP via Firewall. So we can protect our system from attacker.

Comparison between Ariori and FP-Growth:

FP-growth work faster than Apriori as number of transaction will increase. In test data numbers of transactions are 10000 max so large significance of time complexity cannot be measured. But as proved in large transaction it will help to reduce time complexity. In Apriori if you increase database you may not receive all frequent items as FP Growth as it also affect on confidence which is not in FP Growth algorithm.

Comparison Chart:



6. CONCLUSION

Web-based botnets are difficult to detect, because the C&C messages of web botnet are spread over HTTP protocol hiding behind normal flows. On web traffic, data mining techniques are easily applicable. In proposed framework, Incoming and outgoing network traffic is monitored then filtering and separation on packet data or network traffic is done. In our proposed detection framework, we monitor the group of hosts that show frequent communication pattern, and finding malicious host on them. There is no need for any prior knowledge about botnet signature.

Data mining techniques are easily applicable on network flow information. It will require less time and memory compared to Apriori.

FUTURE WORK

In future work, for blacklist and whitelist filtering process can be implemented dynamically so overhead of manual task can be eliminated. This algorithm can be modified by using more improving parameter like timestamp, destination IP.

REFERENCES

[1]Chung-Huang Yang, Kuang-Li Ting, “Fast Deployment of Botnet Detection with Traffic Monitoring”, Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 856-860, 2009.

[2]Haritha S. Nair, Vinodh Ewards S E “A Study on Botnet Detection Techniques”, International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012

[3]Botnet scams are exploding, 2008 http://usatoday30.usatoday.com/tech/news/computersecurity/2008-03-16-computer-botnets_n.htm.

[4]Nicholas Ianelli, Aaron Hackworth. Botnets as a Vehicle for Online Crime - CERT and CERT Coordination Center are registeredin the U.S. Patent and Trademark Office. 2005

[5]Oregon Man Cops Plea in eBay DDOS Attack, http://www.internetnews.com/security/article.php/3574101

[6]Worm strikes down Windows 2000 systems,http://www.cnn.com/2005/TECH/internet/08/16/computer.worm/index.html

[7]Kraken botnet Wikipedia, http://en.wikipedia.orglwikilKraken_botnet, 2008.

[8]Zeus botnet steals $47M from European bank customers,2012. http://news.cnet.com/8301-1009_3-57557434-83/zeus-botnet-steals-$47m-from-european-bank-customers/

[9]Erdem Alparslan, Adem Karahoca and Dilek Karahoca. BotNet Detection: Enhancing Analysis by Using Data Mining Techniques, Downloaded from http://dx.doi.org/10.5772/48804 (BOOK)

[10]Sonal P.Patil, Swatantra Kumar. Botnet-A Network Threat, International Conference on Recent Trends in Information Technology and Computer Science (IRCTITCS), Pages 29-35, 2011.

[11] Xiaonan Zang, Athichart Tangpong, George Kesidis and David J. Miller. Botnet Detection Through Fine Flow Classification. Departments of CS&E and EE, The Pennsylvania State University, University Park, PA, 16802. CSE Dept Technical Report No. CSE11-001, Jan. 31, 2011

[12]Alireza Shahrestani, Maryam Feily, Rodina Ahmad, Sureswaran Ramadass. ARCHITECTURE FOR APPLYING DATA MINING AND VISUALIZATION ON NETWORK FLOW FOR BOTNET TRAFFIC DETECTION, International Conference on Computer Technology and Development,IEEE, Pages 33-37 2009

[13]Zhang yanyan,Yao Yuan, Study of Database Intrusion Detection Based on Improved Association Rule Algorithm, IEEE. Pages 673-676, 2010

[14]Sajjad Arshad, Maghsoud Abbaspour, Mehdi Kharrazi, Hooman Sanatkar. An Anomaly-based Botnet Detection Approach for Identifying Stealthy Botnets, Presented in International Conference on Computer Application Industrial Electronics, IEEE, Pages 564-569, 2011

[15]Wang Zilong, Wang Jinsong, Huang Wenyi, Xia Chengyi. "The Detection of IRC Botnet Based on Abnormal Behavior" Second International Conference on MultiMedia and Information Technology, IEEE, Pages 146-149, 2010

[16]J. Goebel and T. Holz. Rishi: Identify bot contaminated hosts by irc nickname evaluation In USENIX Workshop on Hot Topics in Understanding Botnets (HotBots 07), 2007

[17]Claudio Mazzariello. IRC traffic analysis for botnet detection, The Fourth International Conference on Information Assurance and Security, IEEE, Pages 318-323, 2008

[18]Hossein Rouhani Zeidanloo, Mohammad Jorjor Zadeh shooshtari, Payam Vahdani Amoli, M. Safari, Mazdak Zamani. A Taxonomy of Botnet Detection Techniques, IEEE, Pages 158-162, 2010

[19]Roberto Perdiscia, Wenke Leea, and Nick Feamstera, Behavioral Clustering of HTTP-Based Malware and Signature Generation Using Malicious Network Traces, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA, USENIX, 2010

[20]Jae-Seo Lee, HyunCheol Jeong, Jun-Hyung Park, Minsoo Kim, Bong-Nam Noh. "The Activity Analysis of Malicious HTTP-based Botnets using Degree of Periodic Repeatability" published at International Conference on Security Technology, IEEE, Pages 83-86, 2008.

[21]Dae-il Jang, Minsoo Kim, Hyun-chul Jung, Bong-Nam Noh. "Analysis of HTTP2P Botnet :Case Study Waledac", Published in IEEE 9th Malaysia International Conference on communications,15 -17 December 2009 Kuala Lumpur Malaysia, Pages 409-412, 2009.

[22]S.S. Garasiya, D. P.Rana and R.G.Mehta, 202 "HTTP Botnet Detection using Frequent pattern set mining"- International Journal of Engineering Science and Advance Technology vol.2, Issue.3, pp.619-624 ,2014

[23] Masayuki Ohrui and Hiroaki Kikuchi, "Apriori–Prefix Span Hybrid Approach for Automated Detection of Botnet Coordinated Attcks" Published in International Conference on Network-Based Information Systemsacks, pages 92-97, 2011.