International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-4, April 2013

Spam Mail Visualization through Open Relay On Firewall Gateway

Shashank Shekhar, Gurpreet Singh

Abstract- A picture is able to tell one thousand words. Pictorial representation of any matter is clearer than text. Present time security is more challengeable task in the computer field. No one can say any model/algorithms/idea can't be crack. This paper is basically an idea to visualise the spam mail send by spammer through open relay. Email log file has many information. A lot information can be extracted from a log file. In this paper, two types of log files are described. These are email and firewall logs. This paper shows how a log file is able to visualize the information, attack, and protection against spammer/attacker.

This paper described different method for visualizing the attack through open relay via firewall gateway. Spam mail is a great problem on the email server. Many attack are done on this server via open relays. Firewall may be a great idea to protect the sever/ system by the block or pass the spam mail. This paper also describes the forensic analysis of different attacks via log file on the email server. It is able to visualize worm, virus, dictionary attack, man in the middle attack from log files. Basically this paper represents an idea about the answers of the "wh" words on the email server. Those are "who"(Source and destination), "when"(time), "where"(port and address), "what"(visualization of different activity).

Keywords- Log files, MTA, SMTP, Email, Spam, Directory harvest, Firewall, Virus, Worm, Firewall Ruleset,

I. INTRODUCTION

A spam mail, also known as junk mail or unsolicited bulk email, is involving identical message to numerous recipients. Spam mail may contains malware as scripts or other executable file attachments, phishing link or any other harmful programs like Trojan or virus. There is no clear definition of Spam. We can't categorize clearly between spam and email. But an email is not Spam itself. If the matter or attached document/ file are offensive and the mail is no an attack on the ability of the user to receive or email or use of internet. Generally, a commercial or bulk message is classify as Spam. If we sign up for particular message for a company or organization, and found mail everyday then it is not in category of spam. An open relay (also known as an *insecure relay* or a *third-party relay*) is an SMTP e-mail server. It allows third-party relay of e-mail messages.

Manuscript published on 30 April 2013.
* Correspondence Author (s)
Shashank Shekhar,born in Bihar, and received B. tech degree in Information Technology, from Sikkim Manipal Institute of Technology, Majitar, India.
Gurpreet Singh, born in Punjab and received B.E. degree from Sant Longowal Institute of Technology in 2009,.M.E. degree from Thapar University, Punjab, India.

An open relay makes it possible for an unscrupulous sender to route large volumes of). Due to this the server, who is typically unaware of this problem, donates network and computer resources to the sender's purpose. when a spammer hijacks a server or infect the server of great an organization or network may suffer system crashes, equipment damage, and loss of business with great damage.[14,15,16]
Email works with Simple mail transfer Protocol. Its working commands[5]: MAIL FROM: <reverse-path>;RCPT TO: <forward-path>;RCPT TO: <forward-path>……(for each reciepient);If unknown recipient: response "550 Failure reply"; DATA email headers and contents VRFY username (Often disabled);250 (user exists) or 550 (no such user)

1.1 Problem definitions and objectives:

Email server suffers a lots of anomalous and malicious activities, attacks on server, problems of open relay, identify large emails delay, different types of attack and infections on the networks The firewall run on the restricted environment and can use application gateways.There are a lots of drawback in the base mechanisms and not possible to detect unknown attacks and anomalies detection for high rates of false vulnerabilities.
This article's motive is to visualize the firewall gateway for spam mail with different attacks, anomalies anomalous activities on email server through open relay. Also visualize the all activities done on the firewall gateway with spam mail and activities after infection of network. i.e. blocked, unblocked port, on the email server, worm detection in the email network ,traffic flow analysis by email log file in the pictorial view. Through open relay a spammer can send mail or access any system, from any domain or subdomain networks.

II. A STATISTICAL ANALYSIS OF EMAIL LOG FILE AND FIREWALL

2.1. Email server log:

Email server has two types of log files. First email log is MTA (mail transfer agents) and second log file are POP and IMAP. Email MTA log contains: [12,13]

1. Timestamp: Timestamp contains that time when a mail starts processing.
2. Sender and Recipient: It contains the sender and receiver email addresses.
3. Size: It gives the total size of email message.
4. Message ID: It contains the unique id of message.
5. Status of the mail: it gives the status of the mail which is delivered or not.
6. Time delay: the total time taken receiving and delivered.
7. Total no of recipients: the no of recipients in this mail.

8. Relay: there are two mail server. It gives information about both mail server for receiving and handling.

Jan 22 11:43:27 linux sm-mta[17309]: k28F25Rux017397: from=<gurpreet.16523 @lpu.co.in >, size= 2558, class=0, nrcpts= 1, sgid=<000001c68cf8$2bc42250$e7d4a8c0@mvu82>, proto=SMTP, daemon=MTA, relay= 4 host62-225.pool8711.interbusiness.it [213.33.117.26] Jan 22 11:39:09 linux2 sm-mta[17301]: k28F25Rux017397: to=<mail2shashank@in.com>, ctladdr=<mail2shashank@in.com> (0/0), delay= 00:00:08, xdelay= 00:00:00, mailer=local, pri=294718, dsn=2.0.0, stat= Sent

2.2 Email firewall log:

Firewall generate log entries like traffic flows. The firewall shows that blocked and unblocked packet also. The comman field of a firewall contains the following:[9,10]

1. Time stamp: This gives that time when the log file is recorded.
2. Packet Size: The size of the capture packet.
3. Action: An indication that packet is blocked or not.
4. Ports: It gives information which port is used in the communication for email.
5. IP address: It gives the IP address of the both end points.
6. Ethernet address: address of the previous of next and previous communication points.
7. Rule no: It helps for recognize for passing and blocking port.
8. Direction: It gives the flow of direction about in & out.

Jan 22 20:00:05.60794 rule 57/0(match): pass in on xl1: 231.145.39.55. 1030 > 117.19.2.250. 53: 8010 [1au] A? mx1.mail.in.com. (74) (CF)

III. LOG VISUALIZATION

Defining the Problem→data assess→parsing→graph and coloring the nodes→filtering and aggregation of nodes→ governing [9]

3.1 On Email Server:

For open relay, which accept email not only in the domain also accept in sub domain. The other party can do unwanted activity. It must be prevented. There are some way to find this spam mail and stop it. Sendmail log, which is used to detect to someone, who is abusing email server fig:6.

1. Collect the raw (log file) data and find out more suitable log to extract the information. Generally two types of log file are at email server. MTA and IMAP, POP. We need MTA log file for finding all the information.
2. Parse the email logs by Afterglow, i.e. provide send mail parser. We can also use sed, awk, perl for scripting. The output will be in the .CSV format.
3. Now we specify the domains. If a mail is not in the domain, it means that these are abusing spam mails. It was legitimately sent by mail server. Now colouring process has done and abusing mail is colouring with specific colour which are identify easily. We can use afterglow for identify the source and destination recipient.
4. Spam mails are send to lots of address point. It means that a large no of email send by one source point. Also in these spam mails contain a lots of attached file. These may contains a lot of virus, Trojan, or picture which harms the networks and the system. Delay, flags of the header, average packet size and IP headers can be an arm to visualize the spam mail. In this case it can be visualize by box plots graph and GraphViz.
5. Sometimes multiple email are sent and queued for a large periods of time. It may be possible that all emails are sent from multiple or single invalid mail address. Due to this the mail network made busy and mail doesn't delivered and network become busy. It can be visualized by link graph.

3.2 Firewall visualization:

Generally A Spam mail, which is sent by open relay, will be them who are sending by invalid addresses. Sometimes a lot of same mail send to invalid address to make busy the network for a long amount of time. This is also attack on the server. Visualize this invalid address.

From log file 1st we have to identify the ruleset and extract all the problems and misconfigurations. For this process we need historical data.

Step 1: parse the log file and find out the source address with sequence of the rule in the firewall.

Step 2: find out the blocked and passed sources for investigating it.

Step 3: find out destination port with sequence rule no and extract the all information Action i.e. pass and block from it.

Step 4: colouring the nodes and find where the network is blocked or passed.

The email traffic allowed outbound for a bunch of system/machines. i.e. port no 25. In case of spam mails, those are send by one address, are blocked by a sequence no(rule no). valid mails are passed by another rule. If two email server access the internal machine which doesn't follow rule. It means that it may be affected with malware.[13,14]

IV. PROPOSED ALGORITHM

Algorithm : Proposed algorithm to extract the required fields of the firewall.

Inputs: rules sequences

Output: Desired output field by firewall

for each field in rule.fields/{action}
if E.field = dest_ip or E.field = src_ip
E.field = *.*.*.*;
end if
if E.field = dest_port or E.field = src_port
E.field =*
end if
for each E.ruleset Ri in rules { Extract the E.field in autonomous system}
E.domain ← extract domain of E.field
If E.domain = E.field then Internal machine
else
External machine.
end if
end for
for each E.ruleset Rj in rules { extract overlapping in the system}
if E.field(src_ip)= E.field then overlapping
If complete overlapping then block
elseif partial overlapping then correlation
elseif weak overlapping then pass

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

end if
end for
Xfield =0
for each rule i in rules { count the number of rules having field value equal to the common subset value}
if Ri. field = E.field
Xfield =Xfield +1
end if
end for
end for
X = max(Xdest_ip,Xscr_ip,Xdest_port,Xscr_port) { choose the most common field }
if X=Xsrc_port
return src_port
end if
if X=Xdest_port
return dest_port
end if
if X=Xsrc_ip
return src_ip
end if
if X=Xdest_ip
return dest_ip
end if

V. ANALYSIS

5.1 Firewall Analysis:

This picture (Fig:1)is firewall link graph based on the capture log. The blocked and passed traffic has been shown with port no, source and destination address. This graph also shows the internal(211.153.35.41) and external machines (212.134.34.93)presents in the graph. Port no 25 which is running on the email server. Machine 212.134.34.93 got access with the port 137,132, 449. The mail server (port no. 25) passed(137,25) and block (132,198) some of the traffic which is concern with valid mail or spam mail/redundancy.

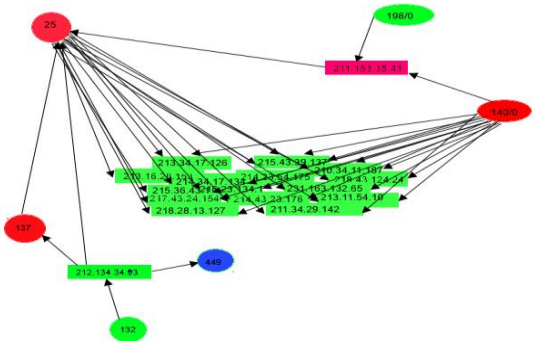

Fig: 1 Firewall link graph

5.2 Check SMTP Open relay via graph: Open relay can be check very easily via telnet. But it can be check by using graph. For showing this, any processing and visualization tool (e.g. AfterGlow), can be used. [13]

1. To check open relay we create a domain for email server.
2. After parsing email log file, create a link graph.
3. All nodes in the domain are coloured by specific colour.
4. We got some nodes which are not in the domain and these are not coloured and These are abusing the server. Unknown, window, guest are undefined node who are abusing server.

5.3 Extraction of information:

5.3.1 Clique and bridge visualization of email :

The clique ,shown in fig 1, and bridge in the graph, gives the information about the user. Clique & bridge[1] are a way to find the following information:

(a) who send email to whom. i.e. source and destination address of email sender and receiver.
(b)Mails are send by any internal recipients or external.i.e. IP address of sender for domain and sub domain or outside the domain in the network.
(c)Who is more communicating? i.e large no of link in the clique. The degree of nodes (out degree) give the no of mails sent by sender. (d) spam mail generally send one to many mapping. It means that clique can be a way to find suspicious recipient.
(e) A recipient can be connected with a bridge from two clique. It means that recipient is communicating with multiple network

5.3.2. Correlation and connections between nodes in multinetwork graph.

Correlation between the nodes will give the information about internal and external machines in the multi-network graph. Correlation is a mathematical relationship between two or more variables. If a graph G = {V, E}, a subset S ⊆ V of the vertices is connected if there exists a path in G between every pair of vertices in S. Correlation has three types: [14]
(a) Perfect correlation: In a autonomous network graph G, A node is connected to all the node then the probability of relationship r = 1, known as perfect correlation.
(b) Strong correlation: In a autonomous network graph G, a node is not connected all the node and the probability of relationship 0 .25 ≤ r < 0.75, known as strong correlation.
(c) Weak correlation: In a autonomous network graph G, a node is connected to few nodes and the probability of relationship 0 ≤ r < 0.25, is known as weak correlation.

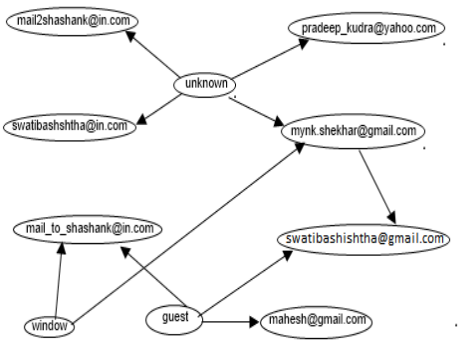

Fig no:2 open relay visualization

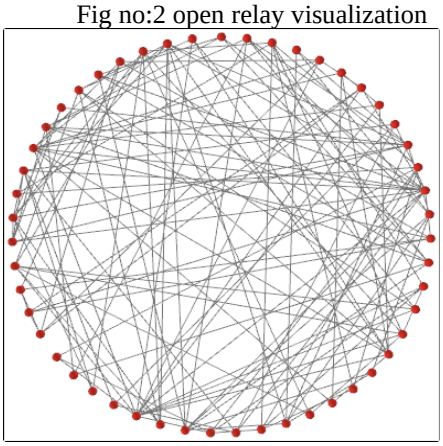

Fig 3 : Clique of small email network.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

In a graph G(V,E) has : (a) A subset S ⊆ V of the vertices, connected to nodes in domain D_i, is termed as internally connected nodes. It means that the nodes are communicating in the same domain. (b) A subset S ⊆ V of the vertices are connected of nodes which belongs to different domains D_i, D_j..... Externally connected nodes, it means the nodes are communicating in the different domains.(c) A subset S ⊆ V of the vertices are not connected with the nodes of any domain termed as disconnected nodes. It means that they are not communicating.[7].A multinetwork email connected preferential attached shown in the fig no:4.

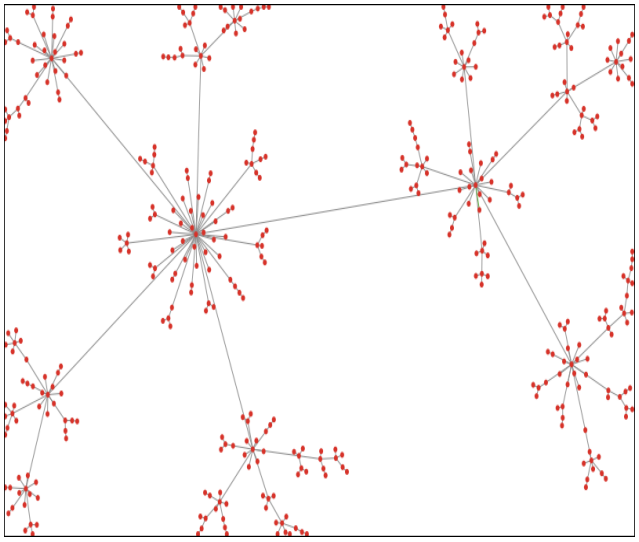

Fig no: 4 A multinetwork connected email nodes

5.*3.3 Overlapping and size of email:*

If overlapping [17] has been found in the graph then it can be clear from the given figure no : 5

Size of mail: If mail is in large size and there is no valid address. It means that it is an attack [9].It shows in the link graph figure no:6. It is abusing the mail with invalid IP address generated by AfterGlow. Sometimes large mail may take more response time due to the busy traffic from a address. It is not an attack. For this, compare two delays time from previous/unaffected and present/affected. If time delay more than threshold time then it is affected. It may be a way to find out man in the middle attack. For finding this attack. (a) Monitor the Source and destination address for specify the victims and attacker. Also verify the port no by firewall log for identifies the targeted service of an attack. (b) Compare the size of mails, bridge and clique in the mail. (d) Compare roundtrip time delay to deliver a mail.(e) compare with threshold value of the time delay

5.3.4 Directory Harvest attack by the spammar:

In Directory Harvest attack, spammer write a program and find the sequence of the alphanumeric/alphabetic/numeric sequence for email address for sending spam. There are many tools to find out "@" symbol for scanning the email for spam mail. Many of the results are incorrect, but after filter the mail address, it will work as targets for sending spam mail. The attacker selects a number of destination domains.[11]

1. The selection can be based on public information avail-able about the domains. (number of expected users on the system)
2. The attacker gains control over innocent victim computers and turns them into zombies. This can be done using a Trojan program or e-mail worm, etc.
3. The attacker carries out the attack by controlling the zombies.
4. The attacker gathers the results from the zombies and analyzes it.

(a) Find DNS and IP addresses. (b) Find the destination port no and source port no with valid/invalid ip address of mail. (c) firewall can block it as mention in the rule set of firewall. (d) If strong relationship it means that no need to block. If weak relationship then block it. (d) Use obfuscation for email address to disguising the spammers [6].(e) Create an disposal email address and use encoding tool to publish email address[6].(f) avoid confirming the unsubscribe request. Delete this mail permanently

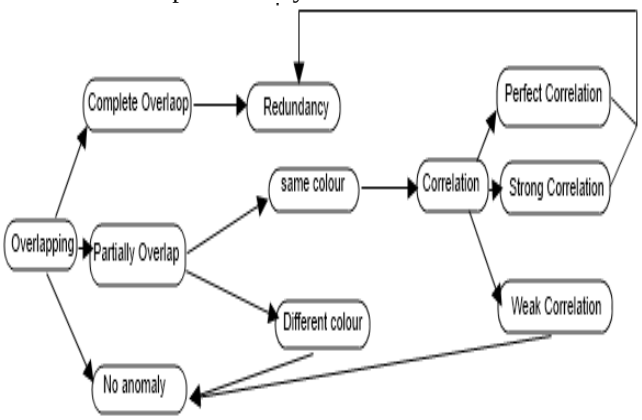

Fig no: 5 overlapping

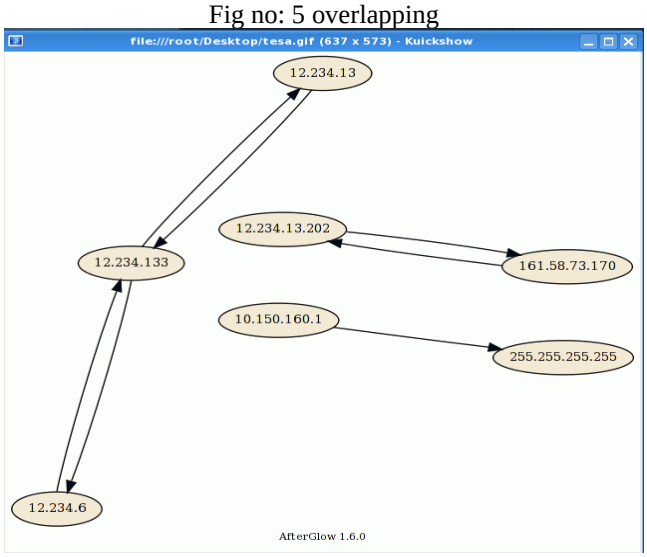

Fig:6 An email abusing

5.3.5 Infection of virus and worm on the network and server:.

If virus has been infected in the system after opening spam mail then connected node can't be visualize. The infected node became disabled for the all neighbors as shown in the graph. Sometimes virus infected node work as a server. It is shown in the figure no 9.It can find by the degree of centrality. Degree of centrality is a measure of the importance of a node embedded in the email network analysis fig no.8.For visualization of the email network centrality analysis of the email network, Hierarchical layout is used. It involves higher placement of a node with a high centrality value and a node with lower centrality value can be interpreted with the height of a node position. [3,4]

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

At first, the node set is divided into different layers L = { L_1 , L_2, . . . , L_h }, so that if u ∈ L_i and v ∈ L_j for edge(u , v), then i < j . These are ordered set of collection of layers. For parallel lines, all edges points in the same direction.

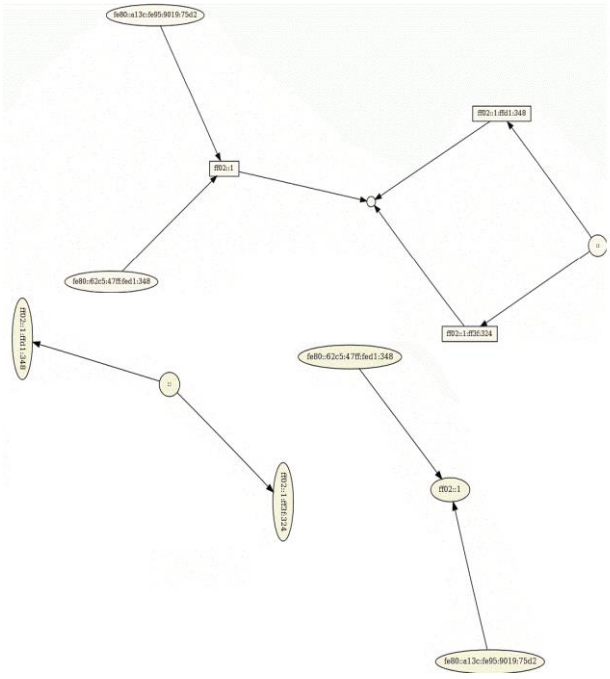

Figure no 7: Blocked by firewall, access is denied.

These directions of the edges plays a great role for partitioning the nodes in to the different layers. Consider the undirected edge {u,v} and d_u and d_v be the degree of centrality for nodes u and v, as a source and target respectively.

In case of hierarchical layout, the degree of centrality of the node will be weakly connected. Lower centrality values nodes are placed above and higher centrality value placed below for each nodes. In the same layer, without edges between the nodes, a weak relation between the centrality values and vertical position of the nodes.[3]

1. Partition the node set into layers. And split each layer into n parts where n ≥ 2;
2. Find Order the nodes in each layer and wall;
3. Assign x -, y -, and z -coordinates to all nodes.
4. Calculate CG(v) = 1/(Max$_{t€V}$ d_G(v; t)) where CG denotes graph centrality (Hage and Harary, 1995)[8]

Some activity by infected node after infection:[9]

(a) Jump: When an infected node choose location randomly on the permutation to perform scan along the network. i.e. infected location (Z) chooses new location (Z+h), termed as jump.

(b) Old infection: When an infected node jumps on previously infected node, termed as old infection.

(c) New Infection: when an infected node hits another node, which is not infected termed as new inefection.

(d) kth jump: when an infected node used a particular sequence to performing infection or made a list to perform the activity of infection termed as kth jump. In this case if location Zth perform infection at Z+k location. After this another node infected list or priority.

(e) 0- jump infection: When infected node hits another node, but it is already infected and retired or loosing connection in the network this jump is 0- jump worm infection.

If worm has been infected in the email network then it generates an enormous amount of traffic. Sometimes edges have been disabled for connected network. Worm can be seen very easily and also find out by long chain of the message and spanning tree for a graph. Most worm propagate with constant payload. Also average packet size of a worm can be specified with fixed length and size. Due to the enormous traffic the node become invisible. It can also specified by header and protocol of the headers. Another characteristics is the scanning behavior. The infected node sends the message follow 'one to many' property. Worm is basically detect by the comparison of graphs ideal and infected.

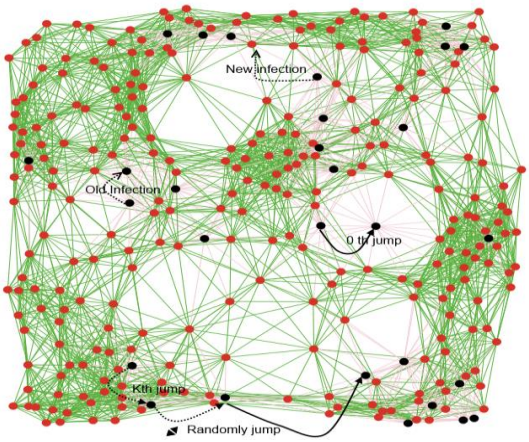

Fig: 8 Virus infected node in the network

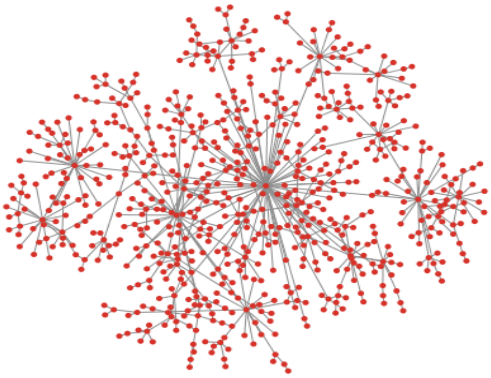

Fig:9 Virus infected Server and nodes

Fig no:10 Infected server after applying degree of centrality (step wise).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

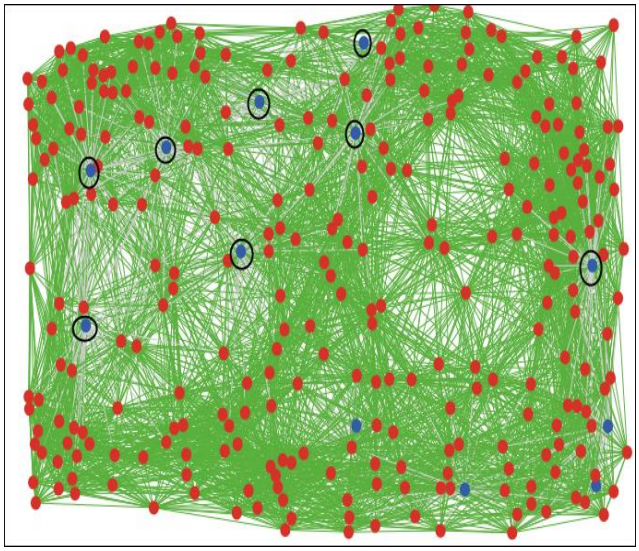

Fig no: 11 Infected Nodes act as a server.[3]

VI. CONCLUSION

Open relay is the best way to send the spam mail. As mention in the paper, it can be visualize and blocked by a firewall. In the domain of the mail server, nodes might be suffer from different attacks which can be detect from above methods. These methods can be disabled mail relay functions, avoid hijack of users, man in the middle attack, disallow third party relay, and avoid email bombing, blocked the spam, worm in the network, virus infection, and different attack as mentions. These methods are not perfect because it can't be use online. It will work but couldn't show more efficiency as like on-line tools. On line tools gives more information. But consider a situation that we are not able to go online to find the redundancy of the subject. In this case it will be work like a forensic analysis of the subjects. Security might be cracked by a different idea but it leaves some spot to find out all information regarding the subjects.

A spammer can send Spammail through open relay using proxy server and perform attack like DDos attack.. It can also visualize by the graph. It contains a great work to find out the spammer.

REFERENCES

[1] C. Bron, and J. Kerbosch. 1973. Finding all cliques of an undirected graph. Comm. ACM 16(9), pp. 575-577.

[2] Wei-Jen Li,Shlolo Hershkop,Salvatore j. Stolfo, VizSec/DMSEC'04, October 29, 2004, Washington, DC. USA.ACM 1-58113-974-8/04/0010

[3] Xloyan Fu, Xlaobin Shen,Seok-Hee Hong,Yingcin Wu, Nikola S. Nikolov,Kal Xu,Visualization and analysis of Email Networks, IEEE Asia Pacific Symposium on Visualization 2007(APVIS2007), Sydney, Australia, 2007.

[4] Weidong Huang, Colin Murray, Xiaobin Shen, Le Song, Yingxin Wu, Lanbo Zheng. Visualisation and Analysis of Network Motifs.,IEEE 9th International Conference Information Visualisation, London England, 6-8 July 2005.

[5] *Dan Boneh ,Unwanted Traffic*: *Denial of Service and Spam email*. CS 155. Spring 2009.

[6] Tobias Eggendorfer Jörg Keller, Preventing Spam By Dynamically Obfuscating Email-Addresses, IEEE, 200

[7] J. Baumes, M. Goldberg , M. Magdon- Ismail, A. Wallace. Discovery of hidden group in the communication networks, Rensselaer Polytechnic Institute, 21st Feb 2004.

[8] Ulrik Brandes , A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2):163-177, (2001).

[9] Parbati kumar Manna, Shigang Chen, Sanjay Ranka, Inside the permutation-scanning worms: propagation modeling and analysis.

[10] Wei-Jen Li , Shlomo Hershkop, Salvatore J. Stolfo, Email Archive Analysis Through Graphical Visualization , VizSEC/DMSEC'04, October 29, 2004, Washington, DC. USA.

[11] Boldizs´ar Bencs´ath Istv´an Vajda, Efficient Directory Harvest Attacks

[12] Olu Akindeinde,Security Analysis and data Visualization, October 16,2009

[13] Raffael Marty, Applied Security Visualization, Pearson Education,Inc 2009,

[14] Richard Blum, Open source Email security, Sam publication, 2002

[15] Mark Ciampa,Security+guide to network security fundamental,3rd edition, 2009

[16] George M. Marakas Introduction to DataMining, Wirehousing and Visualization, Core Concepts, Eearson Education,Inc, 2009

[17] Nidhi Sharma FireViz: A Personal Firewall Visualizing Tool, Massachusetts Institute Of Technology, Masters of Engineering in Computer Science and Engineering,June2005.

Shashank Shekhar,born in Bihar,and received B.tech degree in Information Technology, from Sikkim Manipal Institute of Technology,Majitar,India. Currently Pursuing, M.tech,in Computer Science and Engineering from Lovely Professional University, Phagwara, Punjab,India.

Gurpreet Singh, born in Punjab and received B.E. degree from Sant Longowal Institute of Technology in 2009,.M.E. degree fromThapar University, Punjab. Currently working in Lovely Professional University, Punjab as Asst. Professor of Computer Sc. And Engg. From 1.5 year.