Detection of Botnet using Spamming Behaviour of the System

Shivanna D, Dr. K Raghuveer
Student of Computer Network Engineering, Professor & Head of the Department
Information Science and Engineering,
The National Institute of Engineering, Mysuru, India

***Abstract**:* Malicious software would normally infect the network or machines by exploiting the security threats and problems present in the browsers to filtrate a system when the customer or user visits a threatening or potentially dangerous websites. Now compromised systems can be used to form the Bot-nets under the external control (controlling outside the machine network), which are then used for spamming or to disable a website by sending bogus requests. In this system we are going to use a network forensic technology to find the concealment techniques and the characteristics of the Bot-net affected systems. And here we are proposing the computational intelligence techniques to detect Bot-nets.

Bot-net has lot of characteristics (developed by many skillful developers, dynamic nature and of high flexibility) such as DoS, DDoS, flooding, spamming etc... Of many characteristics of Bot-nets in our project we have considered the spamming as the one way of detecting whether the system got really affected by the Bot-net. The protocols (such as SMTP, POP3 and IMAP) which are necessary for emails operations will be considered for finding or detecting the affected system. The system will have certain bandwidth and even the network will have certain bandwidth, when any system got affected by the Bot-net then there might be a chance of using the bandwidth for potentially dangerous activity. So the bandwidth will be wasted. So the spamming is one of those kinds and in this project we specially look into the system is affected by the spam by checking the email transactions. More the number of outgoing emails more the chance of system got affected by the Bot-nets.

***Index Terms* – Botnet, Network Forensics, Fuzzy Logic, Artificial Intelligence**

I. Introduction

Spam is an Unsolicited bulk email (huge number of emails), sent to a large or many number of user's email addresses, where the users of those emails, who have not given consent for or agreed to receive such kind of an emails. Spam is used for an advertising a service (any kind of modern world service) or a product, more like getting the junk mail through the post. The real example of spam is an email from an unknown or forged address advertising Viagra kind of things, even here the user has not agreed to such content delivery. It is a big potential threat to the internet, as it consumes around 60% of all the network traffic, it includes email traffic too. Spam costs consumers and Internet Service Providers (ISPs) huge amount of cost in bandwidth charges only, this will be really expensive it continues for a long time. Even we find many numbers of methods for fighting spam which are technically strong; the spammers will somehow manage to work on sending the spams in anyway.

Spammer normally does not pay much to the spam operations. They achieve this through the method of exploiting open email servers, which are very prone to spamming attack. The spammer would want to send just one email message to an improperly configured server to reach huge number of email users, where the bulk transfer is handled through the mis-configured server. Here recipients do also need to pay cost simply, even for receiving the emails. As the ISPs no need to bear for these mails, finally the consumers have to pay for spamming.

Various legislations have been implemented to handle the spammers, but still it is very deterrent to handle spammers.

II. Literature Review

2.1 Network Forensics

Network forensics is the process in which capture, recording, and analysis every network events or transactions so as to make sure that there might be a problem where the security holes are present in the system.

2.2 Bot-Net Identification

The plan is to identify systems that might have been infected by the Bot-net software. The purpose here is not to find such Bot-net controllers, known as spammers. Email traffic analysis focuses on getting the information about the behaviour of email users based on the sender, receiver, and date/time information and other details derived from the email messages. The initial step here is to build a profile of each static IP machine or system to test network's behaviour, or so called "normal".

A number of the metrics is chosen build such behaviour for each machine which contains static IP:

1. Volume (in other words number) of weekly outgoing email traffic or the frequency of incoming connections in to the SMTP. Network traffic through SMTP ports.
2. Consistency of weekly outgoing email volume (here measured in number of packets).
3. Sending delays between emails.

4. Median of sending delays of emails, I.e., between the emails.
5. Volume of weekly incoming email traffic (Number of packets received). Traffic through IMAP/POP ports.
6. Ratio of Outgoing to the incoming mail traffic (number of outgoing packets/incoming emails packets).

A high ratio of outgoing (metric 1) to in coming (metric 5) email traffic may also indicate a spamming Bot-net.

Using the above metrics, usually calculate change in the email traffic behaviour which ranges from -1 to 1.

- 0 indicates the little change in email sending behaviour of a machine.
- 1 indicates the large change in email sending behaviour of a machine

2.3 Fuzzy Logic in artificial Intelligence

Fuzzy logic is more or less the binary logic that is it may be either 1 or 0. It is a continuous of values which will always between 0 and 1. This can also be said in other way as 0% to 100%. For an example consider of the variable YOUNG, say age of 5 years is 100% young, age of 18years is 15% young and exactly 18 years is considered as 30% of young and age 30 is 0% young, as per the fuzzy age below 18 years is 100% young and above 18 years is 0% young. So finally using fuzzy the conclusion can be drawn that it is 0 or 1(0% or 100%).

When there are N observations given from x_1, x_2 ……x_n in a discrete time series, so we can form n-1 forms like (x_1, x_2), (x_2, x_3)…. (x_{n-1}, x_n), the time series between 2 successive time series like x_t and x_{t+1} can be calculated as

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\left[\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2\right]}}$$

Where $\bar{x}_{(1)} - \sum_{t=1}^{N-1} x_t/(N-1)$ is the mean of first n-1 observations.

&

$$\bar{x}_{(2)} = \sum_{t=2}^{N} x_t/(N-1)$$

is the mean of last n-1 observations.

Here r_1 is called as an autocorrelation coefficient or serial correlation coefficient.

The r_1 is usually lies between -1 to 1 where near 0 means the normal behaviour and near the 1 is the value for abnormal behaviour and the coefficient is computed for each metric or observation, so coefficient is calculated for current behaviour and normal behaviour and difference between them is calculated to find the behaviour which may be normal or abnormal. Value close to zero tells there is a more of normal behaviour and close to 1 indicates the abnormal behaviour.

III. Implementation Of The Proposed System

3.1 First part of the implementation is Network Forensic

Capturing the network events where it collects only needed information, as there is no need to have bilk information to serve the purpose.

Port numbers of protocols given.

SMTP: Standard ports: 25, 587, and 465.

POP3/POP: Standard ports: 995 and 110.

IMAP: Standard ports: 143 and 993.

Once the required data is extracted next step is to use that data to check the behaviour of the system.

3.2 How the behaviour is actually calculated?

The details of mail transactions or events of a week are considered in computing the behaviour of the system. The details for the normal behaviour are taken when system is not affected from any threats. In this stage the system is full of zero problems (assumption). So the details of mail events will be stored into the XML file system. next at each successive week the details regarding the mails are added embedded in to the files. So the system might get affected in any of the weeks. So details of the normal behaviour of the machine, which was stored in files when the system was not affected by any kind of threat will be checked against the present week data.

Serial auto correlation coefficient method is used to find the behaviour of the system or machine, for both the initial week and the present week. And the difference between them will be computed, if the difference is towards zero says there is a less chance of system got affected from the bot-net. If the value is towards 1 then there is a more chance that system got affected from the bot-net.

3.3 Explanation With Example

For the current behaviour the last 7 days data will be used so that it shows how the behaviour of the system at present is.

Example:

	Normal ()	Current	Difference
Metric 1	0.2	0.2	0
Metric 2	0.3	0.1	0.2
Metric 3	0.5	0.25	0.25
Metric 4	0.25	0.10	0.15
Metric 5	0.33	0.33	0
Metric 6	0.4	0.2	0.2

Once the entire values calculated then next thing is to compare, the above table is an example to that where Normal and current depicts the behaviour at initial and present week respectively.

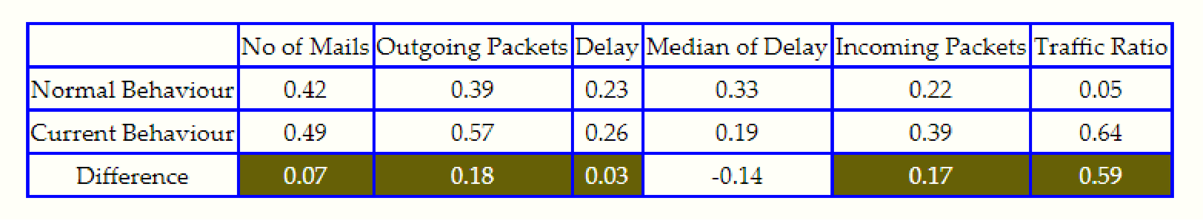
Now the behaviour is calculated in a following manner.

$$\mathbf{z= (0, 0.2, 0.25, 0.15, 0.0, 0.2) = 0.2 \text{ for } z \text{ in } [0, 1].}$$

The value calculated above "z" decides the behaviour.

IV. RESULTS AND DISCUSSION

The results of the developed system to compute the behaviour.



	No of Mails	Outgoing Packets	Delay	Median of Delay	Incoming Packets	Traffic Ratio
Normal Behaviour	0.42	0.39	0.23	0.33	0.22	0.05
Current Behaviour	0.49	0.57	0.26	0.19	0.39	0.64
Difference	0.07	0.18	0.03	-0.14	0.17	0.59

Table 4.1 Descriptive Statistics

Graph representation of behavioral differences

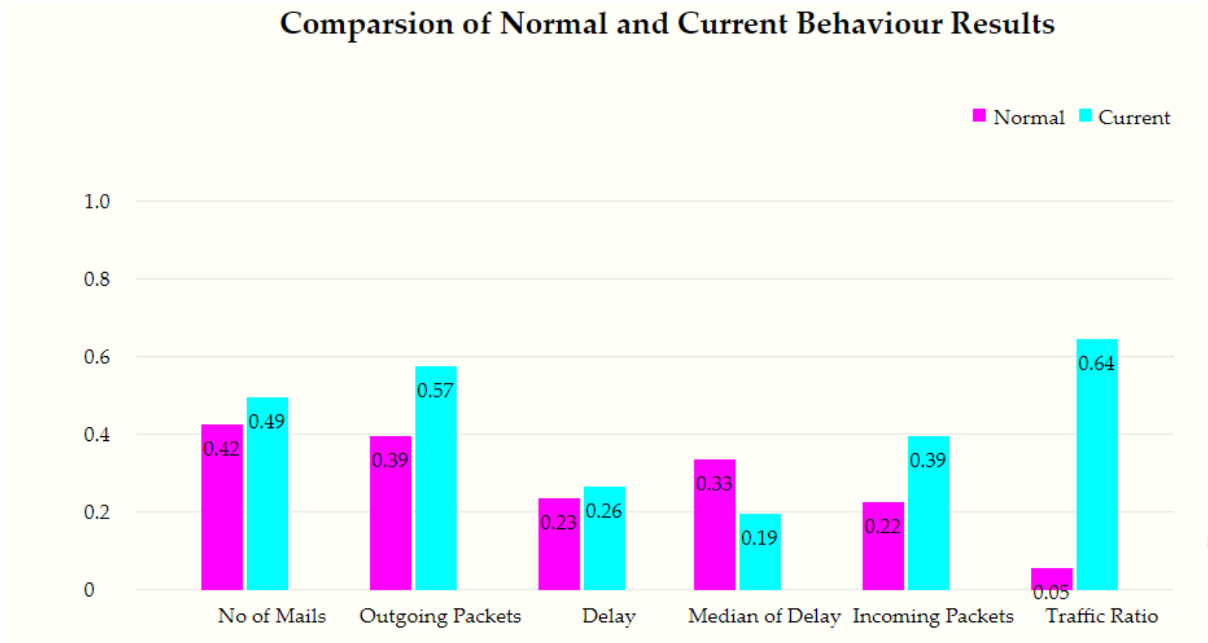


Figure 4.2 Comparison Results

According to the above table after running the system developed, value calculated for Behaviour changes is **0.208.**

And to know where the system stands for the present behaviour is pictorially represented in the following graph.

The value 0.208 is in Some What Normal Behaviour

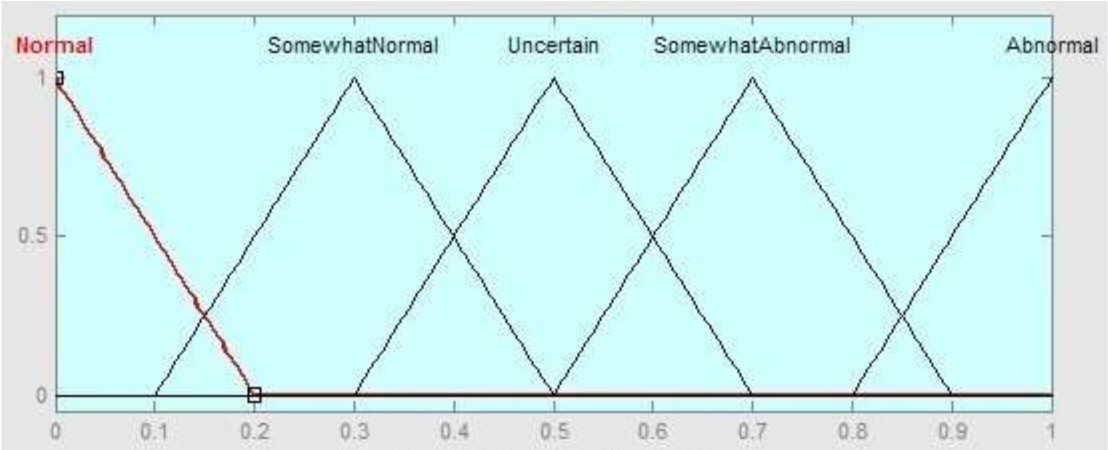


Figure 4.3 Pictorial representation of the Graph

V. CONCLUSION

This software will implement a network forensic tool that can be used by any organisation to detect the spamming Bot-nets on their network. This will enable the users to stop their computers being used to send unsolicited bulk spam emails. It will also save the cost in terms of data storage costs for spam emails that are being sent out by their network. We use here metrics and how these metrics would get filtered for "noise" so as not to corrupt the data that was captured by the network forensic tool.

REFERENCES

[1]. Chatfield, C. (1996) The Analysis of Time Series: An Introduction, 5th edn, Chapman and Hall, London.

[2]. M. J.-H. Lim, M. Negnevitsky, and J. Hartnett, "A fuzzy approach for detecting anomalous behaviour in e-mail traffic," in 4th Australian Digital Forensics Conference, C. Valli and A. Woodward, Eds. Perth, Western Australia: School of Computer and Information Science, Edith Cowan University, 2006, pp. 36 - 49.

[3]. Using network forensics and artificial intelligence techniques to detect Bot-nets on an organizational network, I. Vural, H.S. Venter University of Pretoria, 2010 Seventh International Conference on Information Technology.

[4]. Patrick F. Dunn, Measurement and Data Analysis for Engineering and Science, New York: McGraw--Hill, 2005 ISBN 0-07-282538-3

[5]. J. E. Dickerson, J. Juslin, 0. Koukousoula, and J. A. Dickerson, "Fuzzy intrusion detection," Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference, vol. 3, pp. 1506-1510, 2001.

[6]. Internet Service Providers' Association, 2008. 'What is Spam?' Available: http://www.ispa.org.za/spam/whatisspam.shtml. [April 2009]

[7]. https://www.tutorialspoint.com/software_testing/software_testing_quick_guide.html

[8]. https://www.techopedia.com/definition/29998/system-design

[9]. https://www.techopedia.com/definition/29998/system-design