ISSN 22773061

Efficient Detection of SPAM messages and SPAM zombies in the Internet using Naïve-Bayesian and Sequential Probability Ratio Test (SPRT)

K.Munivara Prasad[1] A.Rama Mohan Reddy[2] K Venugopal Rao[3]

[1]Department of Computer Science and Engineering, JNTUH, Hyderabad

[2]Professor and Head, Department of CSE, SVUCE,SV University, Tirupati

[3]Professor and Head , Department CSE, GNITS, Hyderabad

Abstract-The Internet is a global system of interconnected computer networks that provides the communication to serve billions of users worldwide. Compromised machines in the internet allows the attackers to launch various security attacks such as DDoS, spamming, and identity theft. Compromised machines are the one of the major security threat on the internet. In this paper we address this issue by using Naïve-Bayesian and SPRT to automatically identify compromised machines in a network. Spamming allows the attackers to recruit the large number of compromised machines to generate the SPAM messages by hiding the identity, these compromised machines commonly known as spam zombies. We used Naïve-Bayesian and manual methods to detect the SPAM messages and used SPRT technique to identify the spam zombies from the SPAM messages. We proved that the Naïve-Bayesian approach minimizes the error rate, false positives and false negatives compared to the manual approach in the process of detecting SPAM message. Our evaluation studies based on one day email trace collected in our organization network that shows Naïve-Bayesian and SPRT are the effective and efficient systems in automatically detecting SPAM messages and compromised machines in a network.

Keywords: SPAM messages, DDoS Attack, false positives false negatives and Naïve-Bayesian approach

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 7 , No 1

editor@cirworld.com
www.cirworld.com, member.cirworld.com

ISSN 22773061

I.INTRODUCTION

In today's computing world, internet plays an important role in our daily lives (in almost every aspect).It is the place where we do lot of things just sitting at one place. Internet not only influences the people to do positive works but also influences the people to trouble others by posing many attacks. These attacks are posed by the attackers directly or indirectly.

Attacks are broadly classified into two types, one is automatic attacks and other type is manual attacks. Most of the successful attacks are from the automated generated code injected by the attackers. These are very dangerous which includes Denial of Service (DoS), Distributed denial of Service (DDoS)[18], E-mail Worms, Viruses, Worms, Trojan horses, phishing attacks etc.

Internet e-mail worms are very popular because they are very hard to track. After creating a worm, attacker uses one of the many anonymous e-mail services to launch it. Most of them are in huge size and the user is enticed to execute the worm. The worm first load into the machines main memory and it looks for additional email addresses to send itself to.

Attackers get the control over the machines, to launch the attacks on targeted machine, which are formally known as drones, bots, zombies or compromised machines. In E-mail applications these are called as spam zombies because these zombies generate huge number of spam messages to launch the attack on the target machine. It is given that spamming is the major security challenge in the email communication. According to the Report of 2012 march more than 75% of all email traffic is occupied by the spam [2]. The detection of these spam zombies is cumbersome for the system administrators.

Spamming [2][1] is an important threat plaguing the internet from the past decades. More than 75% of traffic is spam and in that 0.4% was malicious [2]. The attacker floods these spam mails by using the collection of compromised machines known as botnet or zombie army to the target machines. These compromised machines are called spam zombies. Normally spam is given as UBE/UCE i.e., Unsolicited Bulk or Commercial E-mail. Spam message is an unwanted message to the users because, they occupy the network bandwidth, disk space, connection time and money, could hide viruses and pornography information and can tempt the users to send their money and the confidential details.

E-mail spamming [2][17] turn out to be the major platform for the attackers because of its unique behavior, low cost and high speed. Spamming is the major resource for the attackers to get the incentives and they are earning around $200 billion dollars per year. Spamming attracts the attackers day by day because it is the easiest and cheapest communication available today.

In general E-mail applications and providers uses spam filters to filter the spam messages. Spam filtering is a technique for discriminating the genuine message from the spam messages. The attackers send the spam messages to the targeted machine by exalting the filters, which causes the increase in false positives and false negatives. False positive is the misclassification of good message as a spam message and false negative is the misclassification of spam message as a good message. Efficient spam filter aims to minimize the false positive and false negatives.

A Spam filter [14][15] detects the spam messages at three levels likely at Network level, user level and policy based technique.

- Network level technique: Detection of spam messages at network level is very difficult. The approaches used in this technique are, Domain verification and the Challenge Response Systems. Domain verification technique uses sender name, domain, and route information obtained by the SMTP to filter the messages. Example is one cannot send the message to the incoming route in a network path. Whereas the Challenge Response Systems technique probes the sender by asking questions to ensure that the sending side is not an automated bot. This technique minimizes the huge number of messages sent from the automated bots.
- User level technique: the two popular methods used at this level are content based and parameter based (white and blacklists). Content based approach discriminates the genuine message (ham) from spam based on the number of spam words exists in the mail content and weightage of the spam words. Parameter based approach uses the basic parameters of the mail for discrimination.
- Policy based technique: This is not a technical approach. Through pricing e-mail, accusing the spammers by law, some spamming can be reduced.

The efficient detection of spam message improves the performance of the E-mail application and providers, by minimizing the false positives and false negatives. But spam message detection alone does not provides the efficient solution for this problem. Detection of spam zombies along with the spam messages improves the performance and which restricts the zombies from being send the spam messages in future.

In this paper we define an efficient approach to detect the spam messages and spam zombies. We divided the entire process into two phases, in the first phase we are detecting the spam messages using manual and Naïve Bayesian spam filter and in the second phase SPRT is used to detect the spam zombies.

II. RELATED WORK

In spam classification, the existing technologies related to spam categorization are very complicated and gives poor results when compared to Naive-bayesian spam filter.

Choi et.al proposed a technique to detect the bots based on the DNS queries generated. Based on the similarity in the group activity of the DNS traffic the bots are detected in this paper. In [6] the botnets are detected based on the passive analysis on flow data.

Xie et al. developed an effective tool named DBSpam to detect proxy-based spamming activities in a network relying on the packet symmetry property of such activities [8]. We intend to identify all types of compromised machines involved in spamming, not only the spam proxies that translate and forward upstream non-SMTP packets (for example, HTTP) into SMTP commands to downstream mail servers as in [5].

BotHunter uses the IDS trace [3] to detect the bots by comparing the inbound intrusion alarms with the outbound communication patterns. SPRT algorithm focuses on any spamming activity unlike BotHunter which depends on specifics of malware infection process.

An anomaly-based detection system named BotSniffer [4] identifies botnets by exploring the spatial-temporal behavioral similarity commonly observed in botnets. It focuses on IRC-based and HTTP-based botnets. In BotSniffer, flowsare classified into groups based on the common server thatthey connect to. If the flows within a group exhibit behavioral similarity, the corresponding hosts involved are detected as being compromised. BotMiner [4] is one of the first botnet detection systems that are both protocol and structure independent. In BotMiner, flows are classified into groups based on similar communication patterns and similar malicious activity patterns, respectively. The intersection of the two groups is considered to be compromised machines. Compared to general botnet detection systemssuch as BotHunter, BotSniffer, and BotMiner, SPOT is a light weight compromised machine detection scheme, by exploring the economic incentives for attackers to recruit the large number of compromised machines.

A powerful statistical method, Sequential Probability Ratio Test has been successfully applied in many areas of networking security, such as portscan activities [9], proxy-based spamming activities [10], MAC protocol misbehavior in wireless networks.

III. PROPOSED WORK

3.1 Zombies or Bots

The term bot or zombie [11][12] is originated from the word robot. Bot is a compromised system that is controlled by a botmaster or attacker. The attacker identifies the compromised systems in the internet to generate huge traffic or spam messages towards the target machine. The compromised machine refers the machine running with no antivirus software or older versions of antivirus software. Bot may inject or send traffic to the target machine but in turn it may be a part of the substantial traffic flow attacking to the target machine. Generally botmaster gains the control over an individual bot if the security levels of the bot are very low. Likewise bot groups also searches for the poor security entry points (hosts or victims) to destruct the other hosts.

What Bots do

- Uses to propagate the malware,
- Consumes the network traffic
- Harvest the usernames and the passwords
- Uses p2p networks to propagate the malware
- Uses spam messages to get the incentives

Types of the Bots:

AgoBot:It is the most commonly used bot, where more than 500 versions are available. It was developed in C++ and installs the source code directlyin GPL. The advantage of this bot is ,it provides easy commands and scanners addition by extending the CCommandHandler and CScanner class and allows the users to add their own methods to them .It is very hard to do reverse engineering. This acts as a protocol that uses other than the IRC and used for packet sniffing and sorting the traffic.

SDBot:It was developed in c language. Since it is written with poor coding many attackers attracted to use it. The advantage of this is very easy to understand and easy to create the new bots.

3.2 Botnet or zombie armies

Botnet [11][12][13] is a collection of compromised computers that are controlled by a botmaster through commands to forward malicious code, viruses, or spam.Botmaster sends the commands through IRC channels or using other private tools to the bots to execute their commands in the bots automatically. The compromised machines or bots do not know that some other computer is controlling that system. Normally home based, less secured systems will be hosted for these purposes. Bot master gets the complete control over the botnet so that he consumes the complete bandwidth of the network for generating and flooding the spam messages towards the target machine to earn good incentives. The simple commands can scan for the other bots to populate the botnet or can pose threats to the victim machines to fulfill its desire.

3.3 Spam Messages

Spam messages are the special purpose messages [2] to attract the users and to make the user to listen to their fraud words. Like normal message, it also contains the structure of header and body. Where header contains the details of the message and the body contains the content of the message including the subject. Each spam message have some unique properties that distinguishes with the other messages, which includes Empty To: filed, missingTo: field, more number of recipients in the Cc: filed, no or suspected message ID's,BCc: filed exists, same but fake addresses usage in From: and To: fields etc.

3.4 Spam Filtering

The process of discriminating genuine messages from the spam messages is known as Spam filtering [14]. Spam Filtering methods are broadly classified into two types namely content based methods and parameter based methods.

Content based filtering:

This method classifies the genuine messages and spam messages by considering the body of the message.Now a days the spammers are very intelligent and they are using fake identities to send spam messages instead of using their original identity. This Content based algorithms classifies the messages efficiently even the spammer uses the fake identities. Some of the content based spam filtering algorithms is Bayesian, SVM, KNN, Naïve-Bayes etc.

The advantage of this method is it provides efficient results even the spammer uses ip spoofing. Drawback of this method is it takes more time compared to parameter based methods.

Parameter based filtering:

This method discriminates the genuine messages from spam messages by considering the parameters of the message. Parameters of a message include From, To, Received by, Subject, IP address, port addresses etc.

The advantage of this method is very easy to implement and no need to open a message for getting parameters. It takes very less time for processing compared to content based methods. It fails when the spammer uses the fake identity or uses ip spoofing. Some of the parameter based filtering techniques are Blacklists, White lists, Challenge/Response methods etc.

Blacklists are the ip addresses/domain names/e-mail addresses of the real time spammers. In real world many black lists are available. These are called Real time Black Lists(RBL). By comparing the RBL one can filter out the spammer messages undoubtedly. In the same way white lists are the ip addresses/domain names/e-mail addresses of the trusted parties. To maintain these two techniques by an independent user is a tough job while most of the organizations follow them.

3.5 Content Based Spam filtering Methods

Content based spam filtering [15] uses the body of the message for discriminating genuine messages from the spam messages. This method extracts the spam words from the body of the message and based on the threshold calculated for the spam words the messages are discriminated. The words are defined as the spam words by training the system with spam messages and weight is calculated for each word.

Naïve-Bayesian Algorithm:

It is a probability based algorithm to discriminate genuine message from spam message. Naive bayes [16] approach is slight modification to the bayesian algorithm. It is the most effective spam filtering content based algorithm. More detailed description is given in 3.6.

K-Nearest Neighbors:

This method of classification is based on the distance measure among the messages. The distance is measured based on the features between the messages like Euclidean distance measurement. This method doesn't need any training phase, so the incoming messages will be directly measured with the available sample messages.

Bayesian Classifier:

One of the most popular spam filters available today is bayes classifier, which is based on the probabilistic method of classification. The discrimination process is carried out based on the features extracted from the collection of previous spam messages of same kind. It means that, the previously classified spam words of spam message occur in the present spam message more frequently than in the normal message. In this method each word probabilities are calculated in the training phase. After calculating the word probabilities, the message will be classified based on the combination of the words that are commonly presented in the training and in the given message.

We define two classes of messages namely genuine message (HAM) and spam messages (SPAM). Probability distribution function is used to define the classification of these messages.

$$P\left(\frac{c}{x}\right) = \frac{p\left(\frac{x}{c}\right) . p(c)}{p(x)}$$

Where$P(x/c)$ is the probability of the message with feature x from the class c. $p(c)$ is the a-priori probability of class c and $p(x)$ is the a-priori probability of message (x).

The probability of the messages is calculated using the above equation. Normally SPAM message contains more probability (>0.8) than a normal message.

3.6 Naïve Bayesian spam filter

This spam filter [16] uses the principles of Bayesian spam filter for discriminating HAM from the SPAM. Normal Bayesian with the independent features assumption in the calculation of conditional probability is called the Naïve Bayesian spam filter. Here the features in the text is considered independent among each other even though there is an inter dependency. For example, in a phrase "HURRY UP" the two words HURRY and UP is interdependent. That is after the occurrence of the word HURRY there are more chances to come the word UP than any other word. Fortunately the practical results are shown that the dependency will not play a major role in the decision of the SPAM classification.

3.7 SPRT (Sequential Probability Ratio Test)

This is a statistical approach [17] of testing between two hypothesis one is null and the other is alternative hypothesis (H_0, H_1).Based on the observations, a one dimensional random walk will be moved between two boundaries(A,

B).Whenever the variable value touches either of the boundaries the test is stopped and the corresponding result is considered according to the boundary. This can be illustrated as following

$\Lambda_n \leq A \longrightarrow$ Accept H_1 and stop the test

$\Lambda_n \geq B \longrightarrow$ Accept H_0 and stop the test,

$A < \Lambda_n < B \longrightarrow$ Take additional observation and continue the test.

The boundaries are calculated by the user-desired false positives (α) and the false negatives (β). False positive is the case where the algorithm accepts H_1 when H_0 is true. False negative is the case where the algorithm accepts $H_0$when H_1 is true.

$$\Lambda_n = \ln \frac{\Pr(X_1, X_2, X_3, \ldots, X_n | H_1)}{\Pr(X_1, X_2, X_3, \ldots, X_n | H_0)}$$

$\Lambda_n$is the nth observation which is calculated in the below form:

$$\Lambda_n = \ln \frac{\prod_1^n \Pr(X_i|H_1)}{\prod_1^n \Pr(X_i|H_0)} = \sum_{i=1}^{n} \ln \frac{\Pr(X_i|H_1)}{\Pr(X_i|H_0)} = \sum_{i=1}^{n} Z_i$$

Here X_i is the Bernoulli variable, which are independent and identically distributed. We can denote the probability of an observation coming from the H_0 as θ_0 and H_1 as θ_1

$$\Pr\langle X_i = 1|H_0\rangle = 1 - \Pr\langle X_i = 0|H_0\rangle = \theta_0$$

$$\Pr\langle X_i = 1|H_1\rangle = 1 - \Pr\langle X_i = 0|H_1\rangle = \theta_1$$

$\sum_{i=1}^{n} Z_i$ Means if the observation is $X_i = 1$, $\ln \frac{\theta_1}{\theta_0}$ is added to the variable Λ otherwise $\ln \frac{1-\theta_1}{1-\theta_0}$ is added to the variable Λ. Now after adding each observation boundaries are checked. Whenever it crosses the boundaries, the test will be stopped.

According to the Wald's the boundaries calculated by the user-desired false positives and the false negatives as

$$A \geq \ln \frac{\beta}{1-\alpha}, \; B \leq \ln \frac{1-\beta}{\alpha}$$

For practical purpose we consider the above equations as

$$A = \ln \frac{\beta}{1-\alpha}, \; B = \ln \frac{1-\beta}{\alpha}$$

This consideration will be almost negligible because the with the true false positives and false negatives we can write the equation as

$$\alpha_1 \leq \frac{\alpha}{1-\beta}, \beta_1 \leq \frac{\beta}{1-\alpha}$$

And

$$\alpha_1 + \beta_1 \leq \alpha + \beta$$

So true and user desired false positives and negatives are almost similar.

Another important computation that can be obtained from the SPRT test is the number of observations to reach a machine compromised or normal. We can compute the average number of observations by

$$E[N|H_1] = \frac{\beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}}{\theta_1 \ln \frac{\theta_1}{\theta_0} + (1-\theta_1) \ln \frac{1-\theta_1}{1-\theta_0}}$$

,

$$E[N|H_0] = \frac{(1-\alpha).\ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}}{\theta_1.\ln \frac{\theta_1}{\theta_0} + (1-\theta_1).\ln \frac{1-\theta_1}{1-\theta_0}}$$

The number of observations will depend on the user defined values of the four parametersθ_0,θ_1, α, β. System administrators has to be very careful when providing these 4 parameter values.

IV. SPAM MESSAGES &SPAM ZOMBIE DETECTION

We have divided the entire detection process into two phases. The first phase defines the detection of SPAM messages and the second phase defines the detection of attack source or SPAMS zombies.

4.1 Phase I: SPAM Message Detection

We propose two methods for detecting SPAM messages namely Manual method and Naïve-Bayesian method.

4.1.1 Manual Method:

In manual method the system is first trained with the SPAM messages. In the training phase the system extracts the tokens or words from the SPAM message and calculates the weight for each word. The weight for the frequently occurred words usually has the maximum weight compared to the normal words and these words are considered as the SPAM words. The SPAM count or weight for each word is stored in the database.

In the detection phase, the manual method extracts the tokens or words from the incoming message and checks SPAM weight from the database for each word. The average weight of all the tokens or words from the incoming message exceeds the defined threshold value then the message is classified as the SPAM message otherwise it is HAM or genuine message.

Algorithm

Training phase:

- Take the collection of SPAM messages.
- Extract the words or tokenize each and every message.
- Calculate the weight for each word or token.
- Store the results in the database.

Detection phase:

- Extract the words or tokenize each and every message.
- Extract the weights of the words or tokens from the database.
- Calculate the average weights of the words extracted from the message.
- Define the threshold value.
- Check the average weight with the threshold value.

4.1.2 Naïve-Bayesian classifier

The naïve bayes filtering contains two phases of processing which includes training phase and the classification phase. In training phase, the equal number of SPAM and normal messages is trained to get the probabilities of the each and every occurred word in the message and these are stored as a reference for lateral retrieval purpose. In the classification phase, the words are extracted from each message to calculate spammicity and hammicity based on the previously calculated word probabilities. Spammicity defines the probability of SPAM and hammicity defines the probability of HAM or genuine message.

In training phase, Probability of a word or token is calculated by the formula

$$S_{token} = \frac{S\_Count_{token}}{S\_Count_{token} + H\_Count_{token}}$$

Where S_{token} the spammicity of a token is, S_Count_{token} is the number of spam messages that contain this token. H_Count_{token} is the number of ham messages that contain this token.

In classification phase, the message total spammicity and hammicity are calculated by this formula

$$S_{message} = \prod_{i=1}^{n} S_{token_i}$$

Where $S_{message}$ is the spammicity of a message.hammicity can be calculated by the product of the hammicity of the each and every token.

$$H_{message} = \prod_{i=1}^{n} (1 - S_{token_i})$$

Here $H_{message}$ is the hammicity of a message. Hammicity can be directly calculated by using the spammicity of a message. Since spammicity and hammicity of a message are opposite in nature, the total probability will be 1(because of two classes).

Algorithm

Training phase:

- Take equal number of spam and HAM or legitimate messages.
- Extract the words or tokenize each and every message.
- Calculate the spammicity (S_{token}) and hammicity (1-S_{token}) of each token.
- Store the results in the database.

Classification phase:

- Extract the words or tokenize each and every message.
- Retrieve the probabilities of the words or tokens.
- Calculate the total spammicity ($S_{message}$) of the message.

4.2 Phase-II (ZOMBIE DETECTION)

In the first phase of detection, we are using Manual and Naïve-Bayesian methods for detecting SPAM messages. Detection of SPAM messages increases the performance of the system, but it is not the final solution for the problem because the attacker again uses the same machine for transmitting the SPAM messages. Instead of detecting the SPAM message alone, it is better to detect the source of the SPAM messages, so that the machine can block the accepting of messages in future. Detection of the SPAM zombies blocks the compromised systems from being transmit the SPAM messages .In this section we proposed an approach for detecting the zombies based on the IP addresses.

The first phase of detection discriminates the SPAM and HAM messages. The HAM messages are directly accepted and SPAM messages are stored in the spam folder. Now the detection of SPAM zombies is done based on the source IP addresses of the SPAM messages identified. Threshold is defined based on the type of network for identifying the SPAM zombies, if the network is busy network then large threshold is used otherwise threshold is maintained based on the traffic volume.

In spam zombie detection, the above explained SPRT algorithm is used. Here the hypothesis is either compromised machine (H_0) or normal machine (H_1). The observations are the messages generated by the machines. The random walk of a variable (Λ) is based on the messages generated by machines (M).

In a network, a compromised machine will generate more number of SPAM messages than a normal machine. That is the probability of a SPAM message coming from a compromised machine is more than (the probability a spam message coming from) a normal machine.

$$Pr\langle X_i = 1|H_1\rangle > Pr\langle X_i = 1|H_0\rangle$$

Algorithm:

1: Record the IP address of a message sending machine.

2:Get the all 4 parameters (θ_0, θ_1, α, β)

3:Let n be the new observation

4: Assume spam message as $X_n = 1$ and the normal message as $\theta_0 = 0$

5:if($\theta_0 = 1$) //Spam msg

6: $\Lambda_n = \Lambda_n + \ln\frac{\theta_1}{\theta_0}$

7: else //Non-spam msg

8: $\Lambda_n = \Lambda_n + \ln\frac{1-\theta_1}{1-\theta_0}$

9: end if

10: if($\Lambda_n \geq B$) then

11: machine is compromised

12: else if($\Lambda_n \leq A$) then

13: machine is normal, take new observations

14: $\Lambda_n = 0$.

15:else

16: Test continues with new observations.

17:end if

The explanation for the above algorithm is as follows. First IP address of a sending message machine is recorded, then the system administrator sets the 4 parameters according to his network conditions. Each message coming from the machine is an observation in the random walk over two boundaries A and B.

In that, if observation is a spam message ($X_i = 1$), $\ln\frac{\theta_1}{\theta_0}$ is added to the variable Λ otherwise $\ln\frac{1-\theta_1}{1-\theta_0}$ is added to the variable Λ.

$\Lambda_n \leq A \longrightarrow$ Accept the machine as Normal machine and terminate the test,

$\Lambda_n \geq B \longrightarrow$ Accept the machine as Compromised machine and terminate the test,

$A < \Lambda_n < B \longrightarrow$ Take additional observation and continue the test.

Boundaries (A and B) are calculated with the user-desired false positives and false negatives and these are not affect the true false positives and false negatives. Because of the fractional values (ranges from 0.01 to 0.05) of the false positives and the false negatives error rate is minimized.

V. EXPERIMENTAL RESULTS

In our organization, we are maintaining three mail servers and 25000 users utilize the services from that servers. Every day 2, 00,000 to 3, 00,000 mails are transmitting in the network. Because of the SPAM messages the performance of the network is decreased, to overcome the limitation of the network a spam filter is deployed that is based on the content based Naïve-Bayesian in each and every machine. We tested the spam filter accuracy for messages passed through the server and got the results around 99% accuracy for false positives and 95% accuracy for false negatives.

In this process, first every message is send through the spam filter to categorize the message as SPAM or HAM, then the results were passed to the SPRT algorithm for zombie detection. This algorithm classified the machines as compromised or normal based on the given false positives and false negatives parameters. We fixed the false positives and false negatives as 0.01 and 0.01for our network and the threshold for detecting the zombies is maintained as three. That is when a machine sends three spam messages is identified as a compromised machine.

ISSN 22773061

First we tested the accuracy of spam filter that is based on the naïve Bayesian algorithm for 3, 00,000 messages. Above 2, 95,000 were successfully categorized by the filter. In that more than 55,000 messages were SPAM messages.

In naïve Bayesian algorithm, we have created the database with 2, 00,000 tokens. To create a token database we trained the system with 5, 00,000 SPAM and HAM messages. Each and every token is stemmed and got the spammicity and hammicity of the token. In the classification phase the message is tokenized and checked with the database for spammicity and hammicity probabilities. For example the token "free" has the spammicity probability as 0.996 and hammicity probability as 0.003 (approximated to 4 decimalfraction). For the tokens that are not available in the database are assigned as 0.5 (neutral) probabilities. By identifying the message as SPAM or HAM, we have updated the database for lateral retrievals. Finally by calculating the average probabilities of spammicity and hammicity of a message we are classifying the message as SPAM or HAM.The classification of messages results are shown in the figure 5.1.

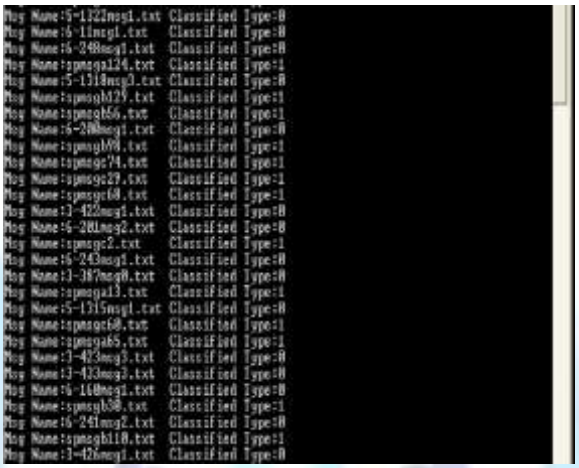

Fig 1: Naïve-Bayesian classification results

Here classified type 0 means HAM or normal message and classified type 1 means SPAM message. In the figure statistics of the classified messages are shown. To test the messages we considered the LingSpam corpus of messages. We got 98% of accuracy, false positives as 0.01% and false negatives as 0.02% for our approach.

To detect a compromised machine we (System administrator) have defined four parameters which includes normal machine SPAM messages sending probability (θ_0), compromised machine SPAM messages sending probability (θ_1), false positive and false negative rates. Normally the range of the 4 parameters is like this:θ_0, is 0.1 - 0.2, θ_1, is 0.8 - 0.9, false positives are around 0.05 - 0.001. θ_0, Value is from 0.1-0.2 which means that the chance of getting SPAM messages is 10-20 percent from a normal machine.

We applied Manual and Naïve-Bayesian approaches separately for 50,000 messages, 1, 00,000 messages and 3, 00,000 messages respectively for our organization network. The messages were passed through SPRT algorithm to record the IP addresses of each sending machine. We have calculated the error rate for each input message group and also calculated the false positives and false negatives for the same.

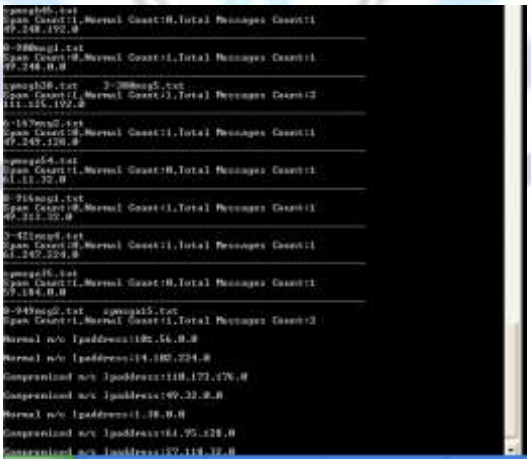

Figure 2: SPRT spam zombie detection algorithm results

Number of messages/Method	Error Rate (%)		
	50,000	1,00,000	3,00,000
Mathematical	9.6	11.9	10.86
Naïve-Bayesian	1.6	1.5	0.93

ISSN 22773061

Table 1: Wrongly classified messages or Error rate of the messages using Manual and Naïve-Bayesian methods.

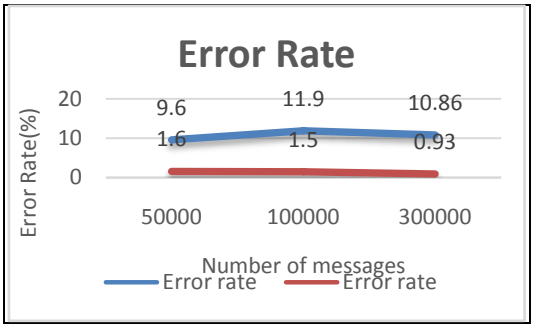

Figure 3: Graphical representation of wrongly classified messages or Error rate of the messages using Manual and Naïve-Bayesian methods.

Number of Packets	False Negatives (%)	
	Manual	Naïve-Bayesian
50000	2.14	2
100000	2.35	2.3
300000	1.48	1.33

Table 2: Calculation of False Negatives for the messages

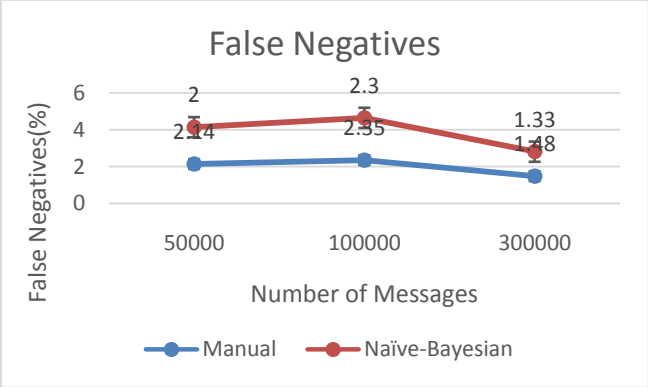

Figure 4::Graphical representation of False Negatives for messages

Number of Messages	False Positives (%)	
	Manual	Naïve-Bayesian
50000	18.54	1.1
100000	21.79	0.62
300000	20.17	0.53

Table3: Calculation of False Positives for the messages

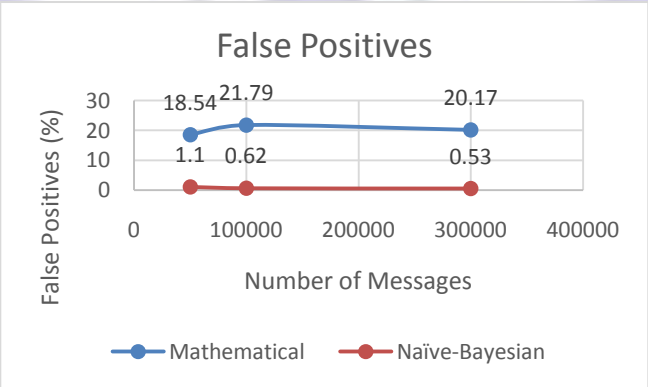

Figure 5::Graphical representation of False Positives for the messages

VI. CONCLUSION & FUTURE WORK

SPAM messages are the major problem for the internet users. In this paper we proposed naïve-bayesian approach to detect the SPAM messages in the internet and we extended our research to detect the source of the SPAM

messages called as SPAM Zombies using sequential probability ratio test (SPRT). We proved that the combination of naïve-bayesian and SPRT improves the efficiency of the system by minimizing the error rate, false positives and false negatives in the detection process of SPAM messages and the SPRT improves the accuracy of detecting the SPAM Zombies from the SPAM messages defined by the naïve-bayesian approach.

We proposed content based approach for detecting the SPAM messages and this provides an efficient and accurate detection results, but the processing time is proportionate to the number of messages. Parameter based approaches takes relatively less processing time. In future we are planning to implement the same with parameter based approach.

VII. REFERENCES

1. BRODER, A.Z., GLASSMAN, S.C., MANASSE, M.S., AND ZWENG, G. Syntactic clustering of the web.In WWW'97.
2. Z. Duan, K. Gopalan, and X. Yuan, "Behavioral Characteristics of Spammers and Their Network Reachability Properties," TechnicalReport TR-060602, Dept. of Computer Science, Florida State Univ.,June 2006.
3. G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "BotHunter: Detecting Malware Infection through Ids-Driven Dialog Correlation," Proc. 16th USENIX Security Symp., Aug. 2007.
4. G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic," Proc. 15th Ann. Network and Distributed System Security Symp. (NDSS '08),Feb. 2008.
5. Y. Xie, F. Xu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming Botnets: Signatures and Characteristics," Proc. ACM SIGCOMM, Aug. 2008.
6. Botnet Detection by Monitoring Group Activities in DNS Traffic Hyunsang Choi, Hanwoo Lee, Heejo Lee, Hyogon KimKorea University.
7. L. Zhuang, J. Dunagan, D.R. Simon, H.J. Wang, I. Osipkov, G. Hulten, and J.D. Tygar, "Characterizing Botnets from Email SpamRecords," Proc. First Usenix Workshop Large-Scale Exploits and Emergent Threats, Apr. 2008.
8. M. Xie, H. Yin, and H. Wang, "An effective defense against email spamlaundering," in ACM Conference on Computer and CommunicationsSecurity, Alexandria, VA, October 30 - November 3 2006.
9. J. Jung, V. Paxson, A. Berger, and H. Balakrishnan, "Fast portscan detection using sequential hypothesis testing," in Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, May 2004.
10.] M. Xie, H. Yin, and H. Wang, "An effective defense against email spam laundering," in ACM Conference on Computer and Communications Security, Alexandria, VA, October 30 - November 3 2006.
11. F.C. Freiling, T. Holz, and G. Wicherski, "Botnet Tracking: Exploring a Root-Cause Methodology to Prevent Distributed Denial-of-Service Attacks," Proc. 10th European Symp. Research in Computer Security (ESORICS), Sept. 2005.
12. E. Cooke, F. Jahanian, and D. McPherson, "The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets," Proc. Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI), July 2005.
13. B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski,R. Kemmerer, C. Kruegel, and G. Vigna, "Your Botnet Is My Botnet: Analysis of a Botnet Takeover," Proc. ACM Conf. Computer Comm. Security, 2009.
14. M. Tariq Banday, Tariq R. Jan Effectiveness and Limitations of Statistical Spam Filters International Conference on "New Trends in Statistics and Optimization" Department of Statistics, University of Kashmir, Srinagar, India, from 20th to 23rd October, 2008.
15. Xiao Mang Li ; Ung Mo Kim "A hierarchical framework for content-based image spam filtering" 8th International Conference on Information Science and Digital Content Technology (ICIDT),IEEE, 2012 Volume: 1 , 2012 , Page(s): 149 – 155.
16. Taninpong, P. ; Ngamsuriyaroj, S. Incremental Adaptive Spam Mail Filtering Using Naïve Bayesian Classification ,10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD '09. IEEE.
17. Duan, Z. Dept. of Comput. Sci., Florida State Univ., Tallahassee, FL, USA Peng Chen; Sanchez, F.; Yingfei Dong; Stephenson, M.;" Detecting Spam Zombies by Monitoring Outgoing Messages ", IEEE Transactions on Dependable and Secure Computing March-April 2012,Volume**:** 9 , Issue: 2 Page(s): 198 - 210 .
18. K. Munivara Prasad ,A. Rama Mohan Reddy, K.venugopal Rao,An efficient detection of flooding attacks to Internet Threat Monitors (ITM) using entropy variations under low traffic,IEEE Third International Conference on Computing Communication & Networking Technologies (ICCCNT), 2012 , 26-28 July 2012,pages-1-11.