www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 11 Nov. 2016, Page No. 19119-19123

Spam Zombie Detection and Blocking Mechanism

Mr. Rajan S. Moravkar#[1], Prof. Ujwala.V. Gaikwad#[2]
[1]kingrajan28@gmail.com
[2]ujwalagaikwad@ternaengg.ac.in

Department of Computer Engineering,
Terna Engineering College, Nerul Navi Mumbai, Maharashtra, INDIA

ABSTRACT -- **One of the key security threats on the Internet are called Compromised machines. The compromised machines in the network are recognize or distinguish using SPOT algorithm. SPOT algorithm is worked on The basis of powerful statistical tool called as Sequential Probability Ratio Test (SPRT) [1]. The Proposed system Spam Zombie Detection and Blocking Mechanism is an online spam zombie detection system in network which designed with Iterative Dichotomiser 3 (ID3) type logic which will increase the existing system performance and efficiency along with the detection it will also blocks the zombie system detected within the network. Proposed system is designed for the private mailing system. It also try to provide the enhanced security mechanism by blocking hacked machines.**

I. INTRODUCTION

In today's computing world, almost in every aspect internet plays an important role in our daily lives. Internet not only influences the people to do positive works but also influences the people to trouble others by posing many attacks like doing spam sending viruses in email. These attacks are posed by the attackers indirectly or directly. There are two types of Attacks, one of them is manual attacks and the other type is automatic attacks. Most of the successful attacks are from the automated generated code injected by the attackers. These are very dangerous some of them are DDos, Dos, E-mail Worms, Worms, Viruses. These machines are called zombies, drones or compromised machines. The Report of 2012 march says that more than 75% of traffic is spam and in that 0.4% was malicious these spam zombies is tough job for the system administrators [2].

The Spam message is an unwanted message to the users because it occupy the network bandwidth, disk space, connection time, money and hide viruses inside spam message .The Spam filtering is a technique that classifies a message into two categories (good and spammed message).

Effective spam filter aims to minimize the false positive percentage. There are many methods available to filter out the spam.

II. RELATED WORK

A. CI Anti spamtechniques:

CI Anti-spam techniques is designed on basis of behavioral characteristics. This anti-spam techniques need to understand the behavioral characteristics of spammers that distinguishes it from senders of non spam messages. The effectiveness and feasibility of CI anti-spam techniques[3] affected by the behavioral characteristics of the spammers such as distributions of non-spam and spam messages by spam statistics, ratios of spam messages from different spammers etc.

Advantage:
1. Botsniffer has very promising detection accuracy with very low false positive rate.
Disadvantage:
1. Botnet CC traffic is difficult to detect.

B. DMTP filtering models:

Receiver-pull and Sender-push. In Receiver-pull, it allow receiver to control over system and when they want any data from sender. In Sender-push, delivery of traffic is controlled by a sender and receivers just accept whatever sender had sent. DMTP (Differentiated mail transfer protocol) [5] which is a pull based model as a counterpart to the spam problem, which grants the control over the message delivery to the receiver. In the push-based white list and black list are defined along with receiver to determine whether to accept the message.
Advantages:
1. DMTP can easily deploy on internet.
2. Spam retrieval behaviors of receivers determine the delivery rates of spam.
Disadvantage:
1. New patterns are hard to detect.

C. Spam Zombie Blocking:

This system not only detects the spam zombies, but also has the facility of blocking the zombie system. This system mainly gives results on the basis of the false negative and false positive error rates. These false negative and false

positive probabilities can be bounded by user defined threshold value. If the number of the messages send by the particular machine exceeds the threshold limit then that machine is considered to be compromised [1].

D. Parameter based filtering:

Parameter based filtering is used to classify a message as either non spam or spam by considering the parameters of the message. Parameters of a message include To, From, Received by, IP address, Subject etc. By using parameter based filtering most of the spam messages will be filtered or detected. Most of the messages will be filtered by checking the parameters of the message [6]. Some of the parameter based filtering techniques are Whitelists, Blacklists, Challenge/Response methods etc.Blacklists are the IP addresses/e-mail addresses/domain names of the real time spammers. In real world many black lists are available. These are called Realtime Black Lists (RBL).

E. Content Based Spam filtering Methods

Content based spam filtering works on the content of the message ie, the body of the message to decide it's a spam or not spam. These methods depend on the training of the previous messages [7].

F. K-Nearest Neighbors:

This method of classification is based on the distance measure among the messages. The distance can be measured based on the features between the messages by equations like Euclidean distance measurement. This method doesn't need training phase where the incoming messages will be directly measured with the available sample messages. So the time complexity of each message is of O (n) [9].

Major security challenge on the Internet is the existence of the large number of compromised machines. Such machines have been increasingly used to launch various security attacks including spamming and spreading malware, DDoS, and identity theft. They are often used to launch various security attacks such as spamming and spreading malware, DDoS, and identity theft [1].

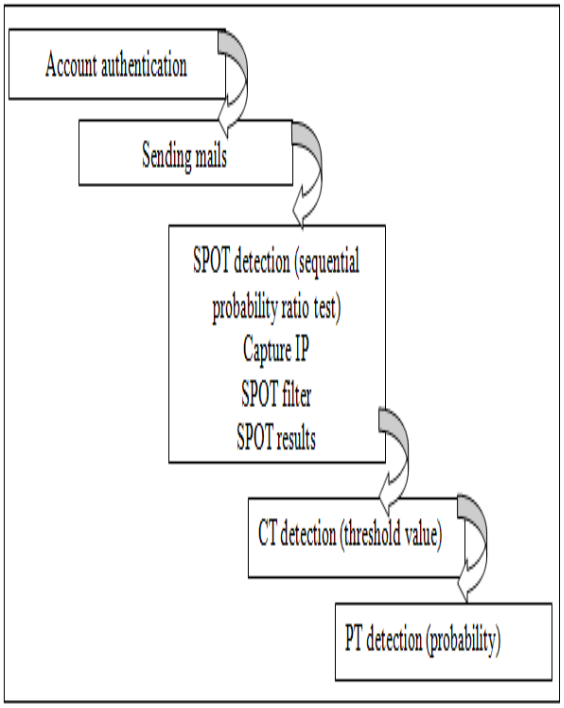

Fig.1 Detecting Spam Zombies by Monitoring Outgoing Messages

There is need to control the existing compromised systems over the network that perform the various security attacks. This paper focuses mainly on the detection of the compromised machines that send the spam messages which are also known as spam zombies. This system does not require the spamming global characteristics such as the spamming patterns of the botnets and the size of the botnets. This system has tool with the help of which an administrator can detect the compromised machines automatically. Thus this system is known as an online botnet detection system. Here the name given to this spam zombie detection system is SPOT system which monitors the outgoing messages.

The statistical method called Sequential Probability Ratio Test is used to design the SPOT system. The SPRT method is used to test the two hypotheses Spam Zombie Detection and Blocking Mechanism which the machine is compromised and the machine is not compromised. This tool helps to minimize the expected number of observations used to take the decision. Here the user define the threshold limit for the false positive and false negative probabilities required by the SPRT method. Thus the SPOT system can quickly identify the spam zombies within the network

DOI: 10.18535/ijecs/v5i11.72

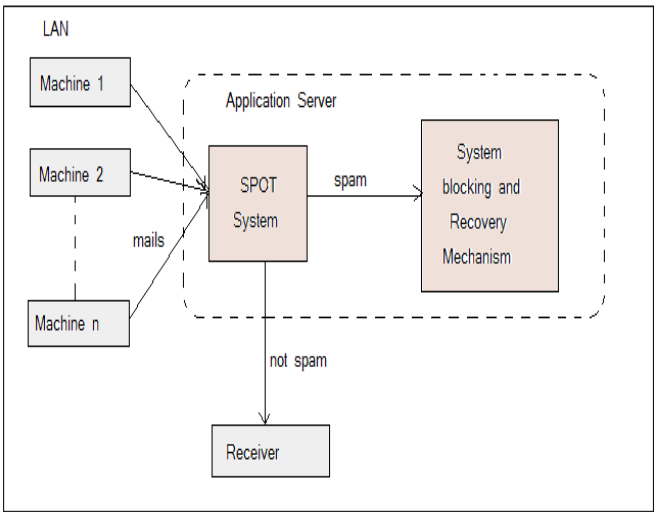

Fig. 2 System Architecture

The Existing system has some basic steps like Account authentication, sending mails, SOPT detection, CT detection, and PT detection as shown in fig 1 to identify the message is spam or not.

The SPOT System or existing system will be connected with more than two machines.

So when machines send message it will be 1st pass to the SPOT system as shown in fig2.

III PROPOSED WORK

The Proposed system use ID 3 algorithm logic [10] where it will get called instead of SPOT decision making logic for filtering purpose. Decision tree is one of widely used filtering or classification method in Data Mining field whose core problem is the choice of splitting attributes [4]. In ID3 algorithm, information theory is applied to choose the attribute that has the biggest information gain as the splitting attribute in each step and a recursive way is used to generate decision tree until threshold condition is reached.

The proposed improved ID3 based on weighted modified information gain called ωID3 is introduced in the proposed system [10]. Also proposed system will store newly introduced email spam behavior in the database which will also not only increase but also help to filter the input messages without any calculation and it will increase the existing system performance.

In the existing system ID3 algorithm will try to improve the performance and accuracy of the system. As shown in fig 3 the ID3 algorithm will do the filtering process just before the message is sent to SMTP server. So the steps will be Account authentication (Gmail authentication), sending mails, ID3 algorithm, CT detection, and PT detection. Additional to this database also help to try to improve the performance of existing system.

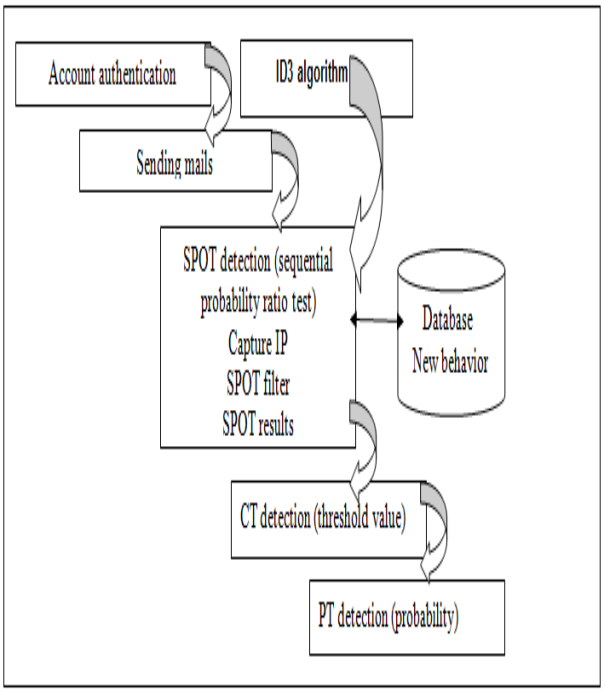

Fig. 3 Spam Zombie Detection and Blocking Mechanism

ALGORITHM FOR BLOCKING THE SYSTEM

The blocking functionalities of the system works as follows:

1: System is a Zombie.
2: Let n be the number of the important mails.
3: 'GId' be Gmail ID and 'PW' Gmail password,
4: declare total and threshold value
4: if (entered Gmail ID==GId) and (Pass word== PW) then
5: start sending the mail for compose mail screen.
6: for loop i=0 to i<5
7: choose the correct mail sender
8: If(choose correct sender) then
9: total++
10: endIf
11: endfor
12: if (total>=threshold) then
13: continue with account.
14: change your password.
15: else
16: block the machine temporarily.
17: senders 'Physical address' block.
18: endelse
19: else
20: enter correct question and password.
21: endelse.

If the system is found as a Zombie system it has blocked temporarily and the user of that system when tries to login then he is informed that the system has been blocked. If the user wants to recover the system then it works as per the above algorithm.

DOI: 10.18535/ijecs/v5i11.72

Login ()
{
1. First take user system 'Physical Address' as its unique.
2. Check whether the user 'Physical Address' address is blocked or not.
3. If (Physical Address' is blocked by system) {
4. Acknowledge user that the machine has been compromised or unblocked it.
}
5. Else {
6. Continue with the account.
}}

IV Performance Evaluation

Finding out the accuracy of the spam detection system is to find out how accurate it detects the spam from test dataset. Proposed system used the down emails from different training data set.

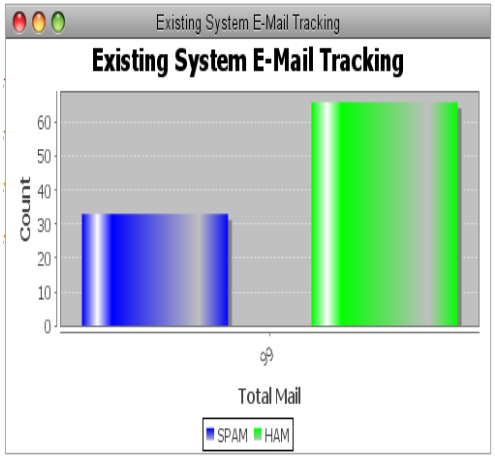

Fig. 4 Existing system E-Mail Tracking
Existing system spam: 33
Existing system ham: 66

True Positive (TP), states the number of spam mails correctly classified as spam. True Negative (TN) states the number of non spam mails correctly classified as non spam. False Positive (FP) states the number spam mails classified as non spam. False Negative (FN) states the number of non spam mails classified as spam. Accuracy gives the performance of the system. The following tables show that increase in the dataset the accuracy of system is improved.

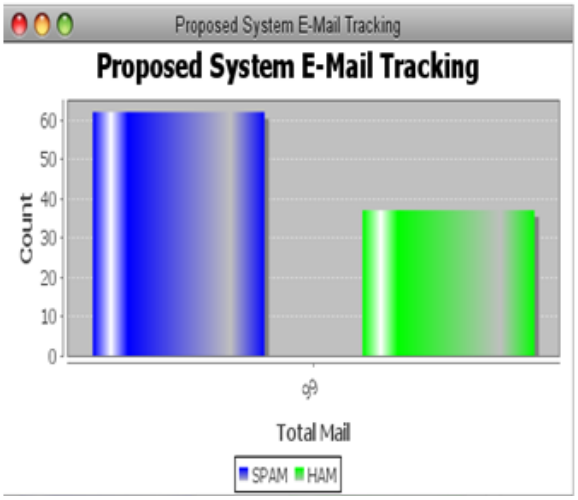

Fig. 4 Proposed System E-Mail Tracking
Proposed system spam: 62
Proposed system ham: 37

	Number of email	True Positive	True Negative	False Positive	False Negative
Existing system	99	26	52	13	7
Proposed system	99	59	33	4	3

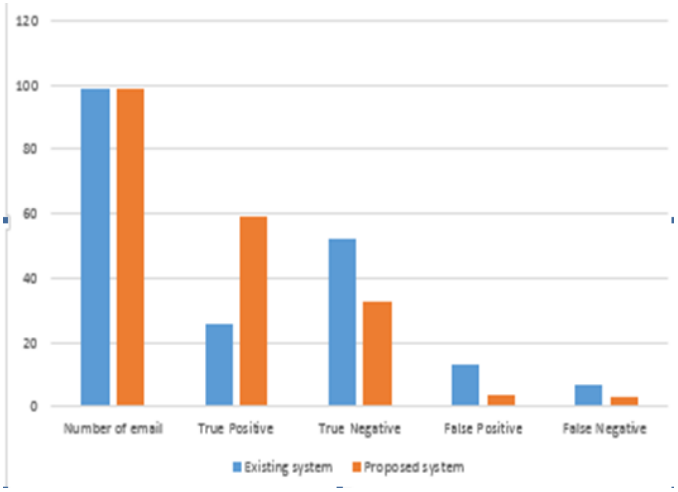

Fig. 6 Comparative graph of existing and proposed system

	Number of email	True Positive	True Negative	False Positive	False Negative	Accuracy
Existing system	99	26	52	13	7	79.5918367
Proposed system	99	59	33	4	3	92.9292929

Table I performance evolution of existing and proposed system

	Number of email	Number of error	TN (%)	TP (%)	FN (%)	FP (%)
Existing system	99	20	53.06%	26.53%	7.14%	13.27%
Proposed system	99	7	33.33%	59.60%	3.03%	4.04%

Table II comparison of existing and proposed system in percentage

The comparative study on this dataset also shows that proposed system for spam detection performs better than existing spam detection system.

V CONCLUSION

The proposed system is using the ID3 algorithm in order to detect the spam zombies. Depending upon the threshold limit system design to minimize the number of the required observation for detecting the spam zombies. The proposed system gives a complete spam detection system which can efficiently process the matching of spam emails. It also provide the temporarily blocking mechanism in which if the system is identified as the spam zombie then the system physical address gets blocked so that it cannot send the spam messages further. The proposed system will try to help to recover the blocked or compromised system in case if the system was hacked by an attacker and was used as a spam zombie. The systems Performance was evaluated on basis its accuracy. The comparative study of existing and proposed system concludes that proposed system is more efficient and accurate.

REFERENCES

[1] Zhenhai Duan, Peng Chen, Fernando Sanchez Florida State University, Yingfei Dong "Detecting Spam Zombies by Monitoring Outgoing Messages"University of Hawaii, Mary Stephenson, James Barker Florida State University,IEEE Transaction on dependable and secure Computing Vol.9 No2 March/April 2012

[2] http://securelist.com, Maria Namestnikova on June 14, 2012.

[3] Alex Brodsky University of Winnipeg, Winnipeg, MB, Canada, R3B 2E9, "Dmitry BrodskyMicrosoftCorporation,Redmond",WA,USA,98033,US ENIX Association Berkeley, CA, USA ©2007

[4] Chen Jin, Luo De-lin ,Mu Fen-xiang. "An Improved ID3 Decision Tree Algorithm," Proceedings of 2009 4th International Conference on Computer Science & Education, 2009, pp.127-130.

[5] Bass T Center for Inf. Protection ,Mclean VA,US, Watt G , "A simple framework for filtering queued SMTP mail (cyberwar countermeasures) " MILCOM 97 Proceedings (Volume:3)

[6] Menna, F. Dip. Ing. e Scienza dell"Inf. (DISI), Univ. of Trento, Trento, Italy "Simulation of SPIT Filtering: Quantitative Evaluation of Parameter Tuning" June 2009

[7] Lian Lin ; Comput. & Inf. Eng. Coll., Xiamen Univ., Xiamen ; Zhongwen Li ; Liang Shi" Spam Mail Filtering Based on Network Processor" Oct. 2008

[8] The SpamAssassin data set is publicly available on website. The URL for web site is http://wiki.apache.org /spamassassin /SpamAssassin.

[9] Viswanath, P. ; Dept. of Comput. Sci. & Eng., Rajeev Gandhi Memorial Coll. of Eng. & Technol., Nandyal, India ; Sarma, T.H. Recent Advances in Intelligent Computational Systems (RAICS) "An improvement to k-nearest neighbor classifier", 2011 IEEE

[10] Jun-Hui Liu ; Dept. of Inf. Eng., Zhengzhou Coll. of Animal Husbandry Eng., Zhengzhou, China ; Na Li" Optimized ID3 algorithm based on attribute importance and convex function", Dec. 2011