Spam Zombie Detection and Blocking Mechanism

Mr. Rajan S. Moravkar#1
Prof. Ujwala Gaikwad#2
Department of Computer Engineering,
Terna Engineering College, Nerul Navi Mumbai, Maharashtra, INDIA
[1]kingrajan28@gmail.com
[2]ujwalavg@gmail.com

Abstract --Compromised machines are one of the key security threats on the Internet. The compromised machines in the network are identified using SPOT algorithm. SPOT algorithm is designed on a powerful statistical tool called as Sequential Probability Ratio Test (SPRT) [1]. Proposed Spam Zombie Detection and Blocking Mechanism is an online spam zombie detection system in network which has Iterative Dichotomiser 3 (ID3) type logic which will increase the existing system efficiency and performance 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 and providing authentication. It will also provide the strong authentication having question answer mechanism.

*****

I. INTRODUCTION

In today's figuring world, web assumes an imperative part in our day by day lives in verging on each viewpoint. Web not just impacts the general population to do constructive works additionally impacts the general population to inconvenience others by posturing numerous assaults. These assaults are postured by the aggressors specifically or in a roundabout way. Assaults are for the most part of two sorts, one of them is programmed assaults and the other sort is manual assaults. The majority of the fruitful assaults are from the robotized created code infused by the aggressors. These are exceptionally unsafe some of them are Dos, DDos, E-mail Worms, Viruses, Worms .These machines are called automatons, zombies or traded off machines. The Report of 2012 walk says that more than 75% of activity is spam and in that 0.4% was pernicious these spam zombies is extreme employment for the framework directors [2].

The Spam message is an undesirable message to the clients since it possess the system data transfer capacity, circle space, association time, cash and cover up infections inside spam message .The Spam sifting is a strategy that characterizes a message into two classifications (great and spammed message). Powerful spam channel intends to minimize the false positive percentage. There are numerous strategies accessible to sift through the spam.

II. RELATED WORK

A. CI Anti spam techniques:

To develop antispam techniques, need the understanding of behavioral characteristics of spammers that distinguishes it from senders of non spam messages. The feasibility and effectiveness of CI anti spam techniques[3] affected by the behavioral characteristics of the spammers such as distributions of spam and non-spam messages by spam ratios, statistics 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:

Sender-push and Receiver-pull. In Sender-push, conveyance of activity is controlled by a sender and recipients simply acknowledge whatever sender had sent. In Receiver-pull, it permit collector to control over framework and when they need any information from sender. DMTP (Differentiated mail exchange convention) [5] which is a force based model as a partner to the spam issue, which gives the control over the message conveyance to the recipient. In the push-based white rundown and boycott are characterized alongside collector to figure out if to acknowledge the message.

Advantages:

1. Spam retrieval behaviors of receivers determine the delivery rates of spam.
2. DMTP can easily deploy on internet.

Disadvantage:

1. New patterns are hard to detect.

C. Spam Zombie Blocking:

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

These are the strategies that are utilized to characterize a message as either spam or non-spam by considering the parameters of the message. Parameters of a message incorporate From, To, Received by, Subject, IP address and so forth. By utilizing parameter based separating the vast majority of the spam messages will be sifted. The vast majority of the messages will be sifted by checking the parameters of the message [6]. A portion of the parameter based sifting systems are Blacklists, Whitelists, Challenge/Response strategies etc. Blacklists are the IP addresses/area names/email locations of the continuous spammers. In certifiable numerous boycotts are accessible. These are called Real time Black Lists (RBL).

E. Content Based Spam filtering Methods

Content based spam filtering uses the content ie, the body of the message to decide a message as spam. These methods depend on the training of the previous messages [7].

F. K-Nearest Neighbors:

This technique for characterization depends on the separation measure among the messages. The separation can be measured in light of the elements between the messages by comparisons like Euclidean separation estimation. This strategy needn't bother with preparing stage where the approaching messages will be specifically measured with the accessible specimen messages. So the time many-sided quality of every message is of O (n) [9].

Real security challenge on the Internet is the presence of the expansive number of traded off machines. Such machines have been progressively used to dispatch different security assaults including spamming and spreading malware, DDoS, and data fraud. They are frequently used to dispatch different security assaults, for example, spamming and spreading malware, DDoS, and wholesale fraud [1].

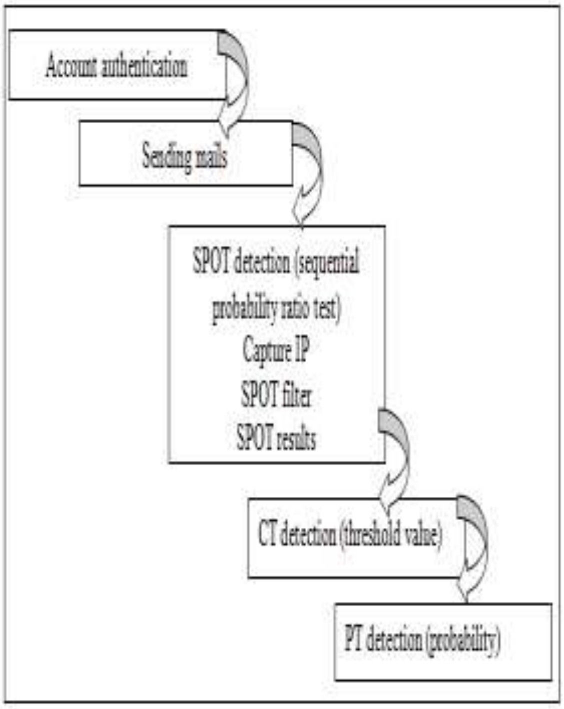
Account authentication
Sending mails
SPOT detection (sequential probability ratio test)
Capture IP
SPOT filter
SPOT results
CT detection (threshold value)
PT detection (probability)

Fig 1.Detecting Spam Zombies by Monitoring Outgoing Messages

There is have to control the current bargained frameworks over the system that perform the different security assaults. This paper chiefly concentrates on the recognition of the bargained machines that send the spam messages which are otherwise called spam zombies. This framework does not require the spamming worldwide attributes, for example, the span of the botnets and the spamming examples of the botnets. This framework has apparatus with the assistance of which a chairman can identify the bargained machines naturally. Along these lines this framework is known as an online botnet identification framework. Here the name given to this spam zombie discovery framework is SPOT framework which screens the active messages.

The factual technique called Sequential Probability Ratio Test (SPRT) is utilized to plan the SPOT framework. The SPRT strategy is utilized to test the two theories Spam Zombie Detection And Blocking Mechanism which the machine is traded off and the machine is not bargained. This instrument minimizes the normal number of perceptions used to take the choice. Here the client can characterize as far as possible for the false positive and false negative probabilities required by the SPRT strategy. Hence the SPOT framework can rapidly recognize the spam zombies inside the system

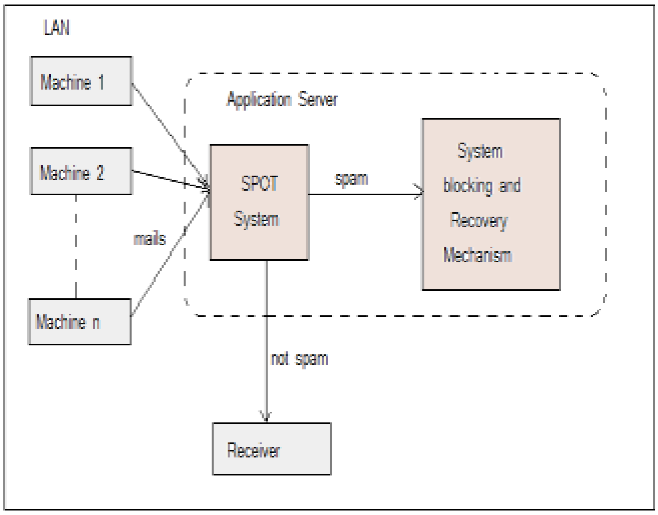

Fig 2.System Architecture

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 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

Proposed system will use ID 3 algorithm logic [10] where it will get called all before SPOT decision making logic for filtering purpose. Decision tree is one of widely used classification or filtering 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.

An improved ID3 based on weighted modified information gain called ωID3 is proposed in the system [10]. Also system will store newly introduced email spam behaviour in data base which will also increase or help to filter the input messages which will increase the existing system performance.

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

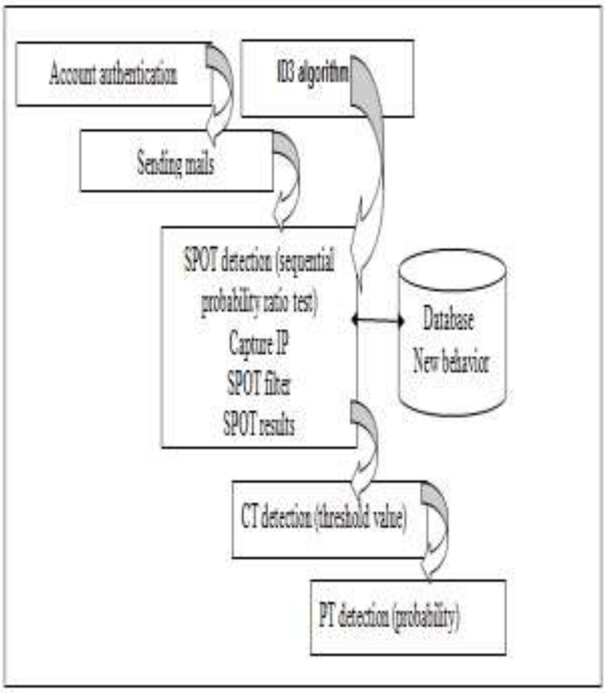
Account authentication
ID3 algorithm
Sending mails
SPOT detection (sequential probability ratio test)
Capture IP
SPOT filter
SPOT results
Database
New behavior
CT detection (threshold value)
PT detection (probability)

Fig 3.Spam Zombie Detection and Blocking Mechanism

ALGORITHM FOR BLOCKING THE SYSTEM

The blocking functionalities of the system works as follows:

```
System is a Zombie.
Let n be the number of the important mails.
'que' be security question and 'ans' be given answer,
declare total and threshold value
if (selected Question==que) and (ans== answer) then
    take first 5 important mails subject.
    for loop i=0 to i<5
        choose the correct mail sender
        If(choose correct sender) then
            total++
        endIf
    endfor
    if (total>=threshold) then
        continue with account.
        change your password.
    else
        block the account permanently.
        senders 'mac' block.
    endelse
else
    enter correct question and password.
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.

```
login()
{
1. First take user account name its password and user system
'MAC' address as its unique.
2. Check whether the user account or 'MAC' address is
blocked or not.
3. If (account is blocked by system) {
4. Acknowledge user that the account is permanently blocked
}
5. Else{
6. Continue with the account.
}}
```

IV CONCLUSION

The proposed system will use the Sequential Probability Ratio Test algorithm in order to detect the spam zombies along with ID3 algorithm. Depending upon the threshold limit given by the user this system intends to minimize the number of the required observation for detecting the spam zombies. The proposed system will also provide the blocking mechanism in which if the system is identified as the spam zombie then the system gets blocked so that it cannot send the spam messages further. The proposed system will try to help to recover the blocked system in case if the system was hacked by an attacker and was used as a spam zombie.

The system Performance will be evaluated on basis of Precision, Recall and Accuracy.

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,USENIX 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