J. Comp. Sci. Comm., 2(2)1-7, 2016

© 2016, TextRoad Publication

ISSN 2356-8844
Journal of Computer Sciences and Communication
www.textroad.com

DSPOT: System to Detecting Spam Bots Based on the Network Edge

Sahar Bahramzadeh, Mehdi Hosseinzadeh, Majid Askarzadeh

Science and Research Branch, Azad Islamic university, Saveh, Iran

ABSTRACT

In recent years, most emails received by users were spams and unfortunately they allocated high percentage of bandwidth of email. What since made spam emails as a security thread for computer networks is high percentage of spam emails sent by Botnet networks. Therefore it is possible for an email spam receiving system turning to a bot and finally this new bot will participate in any malicious operations of this Botnet such as DDos attacks, data theft and spam sending. Consequently, if we detect these compromised machines sending spam, called spam bots in our internal network, it will both prevent spam emails and resulting in Botnet detection. This paper aimed to design and implementation network edge-based DSPOT system for detecting the spam bot systems and detects spam bots on online based by monitoring the output messages. This system is a developed from SPOT system. The performance of this system has been investigated on email traffic of an internet service provider (ISP). Results of this test indicate that comparing to previous systems, the speed and accuracy of this system is higher and it can detect 99 spam bot out of 312 IP Address with only 5 spam bot couldn't be detected by this system.

KEYWORDS— Detecting Spam Bot, Detecting Spammer, Algorithm for Botnets Detection

I. INTRODUCTION

Today, email is considered as a main part of human living, but by development of internet network and using of email for increased speed of data sharing to users in different subjects and despites the usefulness of this technology, we can see mass emails from some individuals and organization on daily based to some extent that they comprise high percentage of total email traffics. On 2010, the rate of spam emails included 88.5% of total emails sent in any day, i.e. up to 61.6 billion emails in each day, included spam emails [1].

Spam emails comprise the biggest problem of emails; because they allocate high size of bandwidth to themselves. In recent years, the security risk and threat of spams has been increased as well. Spams risks are because of high percentage of spam emails sent by Botnets. On 2011 for example, in some months, Botnets were responsible for sending about 95% of spam emails [2].

One of the objectives of Botnets is using the bots for sending spam emails. These spam emails are either for advertisement purposes or resulting in downloading malwares

on user's system followed by making a bot by such system and finally this new bot will participate in any malicious operation of Botnet such as DDos, data theft and spam sending. A type of Botnet using bots for sending the spam emails called "Spam Botnet" and any spam sending bot called "Spam Bot".

There have been considered different approach in spammer detecting systems designed since, for detecting the spam bots. Some of such systems detect spammer system relying on content-based methods in which there has been used an anti-spam filter like spam assassin. Some other ones are being detected relying on traffic specifications belonged to spammers that there is no analysis applied on the text of email. In the third approach, called "Hybrid Approach", one can use both previous approaches as two efficient and complementary factors to detect the spammers [3].

This paper introduces DSPOT system. This is a developed version of SPOT system [4] performing based on network edge and detects spam bots on online based, by analysis of output email. Network edge means connecting point of a network or organization with other networks. This system has used on hybrid approach; meaning besides considering the content of email, we mainly analyze traffic specifications of email packets. We believe that this system is an evolved version of SPOT because it lacks the limitations observed in SPOT. Benefiting from combined approaches, i.e. using anti-spam filter as a content-based method and analyzing the traffic specifications of email as a non-content-based method, we tried to increase the rate of detecting the spam bot.

The remainder of the paper is organized as follows. Part 2 of this paper is allocated to related work in Botnet detection area. Part 3 of this paper introduces designed system and pre-requirements needed for developing DSPOT. Part 4 determines algorithm used in DSPOT of different viewpoints. Part 5 indicates the results from DSPOT evaluation and its performance result in a real network; and part 6 allocated to the limitations of study and finally, the conclusion is presented in part 7.

II. LITERATURES REVIEW

This part reviews the similar studies previously conducted and focused on some studies dealt with detecting compromised machines based on spammers' activity:

[*] **Corresponding Author:** Sahar Bahramzadeh, Science and Research Branch, Azad Islamic university, Saveh, Iran
Email: Sahar.bahramzade@gmail.com

On 2006, Xie et al, William University provided DBspam for detecting the spam proxies. This tool functioned on-line and based on network edge. Spam washing was conducted by open proxies was an extended trick for hiding the actual source sending email spam [5]. Comparing with designed tools, this study aims to not only investigate spam proxies but also all spam bots in the network.

On 2007, Gu et al. developed applied program, BotHunter based on malware infection scenario. This tool is to some extent an IDS. It compares sequential steps of communication and executive operation of a machine with models previously detected based on full successful infection process of a malware and detects potential infected machines in the network [6]. Comparing to BotHunter stressing on detecting whole life cycle of a bot, system designed in this paper only focuses on detecting the spam activities of bots in the network; therefore, it has less processing complexities than BotHunter.

Two recent studies [7] and [8] conducted on 2008 aimed to detect general specifications of Botnets such as their size and members belonged to them by analyzing emails received from a big mail server such as Hotmail. As indicated in a study [7], the methodology included introduction of AutoRE framework in order to extract and analysis polluted URLs embedded in the context of email and consequently clustering the emails according to such URLs. And methodology indicated in study [8] included tracing the similar spam emails received from campaigns belonging to existing Botnets. Generally, such analysis has been conducted on recorded traffic of mail server. Therefore, such designed tools may not be effective in online determination of spam bot in the network. On the other side, such approaches focus on better perception of general specifications of spam Botnets than determining the spammer in an internal network. While this study aims to design a system for determining the spammers in an internal network on online based.

SPOT system has been designed on 2012 to detect spam bot in the internal network [4]. This system functions based on online email traffic analysis and based on network edge. The main drawback of this system is that being a spammer is only detected by relying on contents of emails sent by sender, while today's spammers arrange the content of their emails such not being detected by anti-spam filters. Therefore, under such conditions it is better to study the traffic specifications of email packets besides considering the contents of email using spam assassin to such extent that in some cases, we can detect the presence of spammer without benefiting from email content analysis and only by using such properties. In SPOT system also the messages sent are analyzed randomly and independent from each other, but in this system, we concentrated on abundant and similar messages sent by a machine as well. Comparing to SPOT that considered any sender as an end user, we normally consider email sender as an end user or an internal mail server. Along with development of SPOT system we decide to detect spammer bots in DSPOT, to apply the parameters as mentioned above besides analyzing the email content.

III. A Review on Work Scenario In DSPOT

This designed system analyzes output emails of network on port 25. In some networks, the traffic of port 25 is blocked by edge router [9], but it is unlike in most networks. Anyway, because our system is located before edge router, according to Fig. 1, it can analyze the traffic of port 25 by receiving a copy of network traffic using TAP device and detect the spammers in that network.

Normally, email senders in a network are either final user or internal mail servers. As indicated in Fig. 1, there are also email senders or end user (c) and or internal mail servers (S) in our network. After our internal mail servers being detected, their ip is considered as white list. Even though such mail servers send many spam emails, because their actual sender machines couldn't be traced, therefore, such emails will not be analyzed and they will be considered as secure. DSPOT doesn't know ip address of mail servers before, but by running the application program it will periodically detect them and add to its database.

As mentioned above, detected spammer systems are among end users systems in the network. According to our assumptions, an end user machine in this network is either normal or compromised and this paper focused on detecting the compromised machines called spam bot.

As in introduction, anti-spam filter, spam assassin, running on DSPOT is used as a content-based method and we considered the analysis of traffic properties of senders as a non-content method (based on behavior).

A. Detecting Spam Bots In Network

A system is spam bot in DSPOT according to content-based approach if:

1) *Sending 3 emails with spam content:* According to achievements of SPOT study, if a system sends three emails that according to spam assassin filter the content of those emails are detected as spam, then the sender is spam bot [4].

A system is spam bot in DSPOT according to non content-based approach (behavior-based) if:

2) *Sending email not in allowed time:* In DSPOT, we defined a text file called in which we determined the permissible time for sending email based on any ip.

Therefore, if ip belonging to a subnet sends email not in

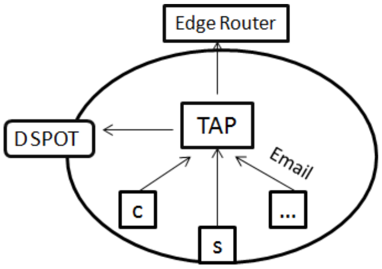


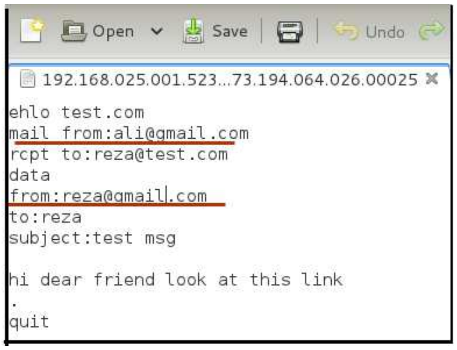
Fig. 1. Situation of DSPOT system in the allowed time is called a spammer.

The reason for this idea is that some spam bots are systems belonging to a specific organization or company with a normal behavior during the day, but in some hours during the night with no user under such machines, they force by Botnet control centers to send emails. Hence, if we permit the ips belonging to such subnets to send email only in their normal period, by such way we can detect some spam Botnets in different subnets. Some users, of course like home users are allowed to send email 24 hours a day.

3) Sending email under conditions that content of sender field is different in the header and body of email: Under normal conditions, the content of sender field in the header must be similar to its content in the body of email [10]. Otherwise, such anomaly in the behavior of sender indicates it is a spammer (Fig. 2).

4) Sending N emails from a machine under conditions with 70% different senders: In SPOT systems, the frequency of emails sent by a machine couldn’t be considered. We know that spammers mostly intend to send similar and many emails. Results of our studies indicate that under normal state, emails sent by a machine are sent in an interval not very long maximum with 3 or 4 different IDs. But if for example 7 out of 10 emails sent by different IDs, it is considered as an anomaly and sender is called as a spammer.

5) Email receiver isn’t a permitted mail server belonging to the sender’s domain: According to Fig. 3, in the first phase of communication in SMTP protocol, the receiver of email packet is legal mail server of sender domain [11]. But some spammers, mail relay server, send the message directly to destination mail server and because this is illegal and against protocol standards, under such conditions this sender is considered as a spam bot.



192.168.025.001.523...73.194.064.026.00025

```
ehlo test.com
mail from:ali@gmail.com
rcpt to:reza@test.com
data
from:reza@gmail.com
to:reza
subject:test msg

hi dear friend look at this link
.
quit
```

Fig. 2.Different sender fields in header and body

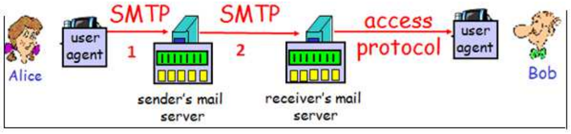


Fig. 3. Stages for comunication in SMTP protocol

B. Detecting Email Sending Internal Mail Servers in the Network

Because email senders in the network might be our permissible internal mail servers, so it is necessary to make some arrangements to detect them and because in this study, we put the internal mail servers in white list, therefore, we will not analyze emails sent by them and consider them as secure.

Generally, in a domain, some machines are only permitted to send email from that domain. Ip of such machines has been recorded in SPF record belonged to that domain in DNS server. On the other side, in any domain, only some machines are allowed to receive email from that domain. Ip of these machines in the MX record belonging to that domain has been recorded in DNS server. Therefore, in such domains, MX recording machines are responsible for receiving email and SPF recording machines are responsible for sending email. Therefore, if email sender IP is among SPF machines, mail sender is a legal server. In some other domains with no SPF record, MX record machines are responsible both for sending and receiving the email. Therefore, in such domains, if Sender's System IP belongs to MX record set, it is a legal mail server [12].

IV. Spam Bot Detection Algorithm

According to stages for implementing the algorithm, when an output message attains to our system, different parameters like IP Address of sender system, sender domain and other parameters as indicated in TABLE I are saved by which we can conduct some necessary studies. Initially, by reviewing the situation of ports in sender system, ensuring that they are client (line 4), we will study next conditions.

- If sender IP takes action to send email not in the time allowed in file, it is a spam bot (line 5, 6).
- Otherwise, should the content of sender field in the header and body of email are different in the email sent by this Ip, it will be a spam bot (line 7,8).
- If sender IP performs normally in both above cases, but destination of email isn't a server from email sender's domain, sender IP will be detected as a spam bot, because according to what mentioned in part 3, client must initially send email for the server of sender's domain (line 9, 10).
- Upon violating all above conditions, it is necessary to study the content of emails by an anti-spam filter like spam assassin by which if sender has sent three spam
- emails, it is considered as a spam bot (lines 11 to 27) [4].
- Upon violating above mentioned conditions, if 70% of sender ID field of emails is different among N emails sent of an IP, the sender is a spam bot (line 28, 29).
- If sender machine isn't a client, it is necessary to study if it is a mail server. For this reason, we will compare sender IP with a set of IPs present in MX and SPF record belonging to the sender's domain. Thus, if it belonged to that set, it is a legal mail server; otherwise it is a Mail Reley Server (line 30, 31).

TABLE I. Parameters required By An Email

Field	Description
S_time	email sending time
Src_domain	domain of email sender
Dst_domain	domain of email receiver
sip	IP of email sender
dip	IP of email receiver
m_size	Size of email (byte)
body_id	sender Id in the header of email
header_id	sender id in the body of email

Algorithm SPOT [4]

A. A General Review On Algorithm Function

As discussed in the algorithm stages, after receiving email and recording all required specifications (parameters indicated in TABLE I), DSPOT system determines if the sender machine is a spam bot or not.

First, Relying on non-content analysis (analysis of email traffic specifications), such as time of sending the message, content of sender fields in the header and body of message and IP of message receiver, determines the situation of sender. Upon correct parameters of email under reviewed conditions, sender will be determined as a spam bot. one of the benefits of our system than previous systems is that we can determine a spam bot even by an email.

Next, by using SPOT system algorithm, i.e. relying on content- based analysis approach and using an anti-spam filter called spam assassin, we will study the conditions of a spammer in sender system. Proportional with a score allocated to emails sent by spam assassin, it can indeed show if it is a spam email. If a system sends three spam emails, it will be a spam bot. Because parameters θ1, θ0, β and α are defined by user, we also determined some values for parameters similar to what mentioned in SPOT. $\alpha = 0.01$, $\beta = 0.01$, $\theta_0 = 0.2$, $\theta_1 = 0.9$ [4].

Should emails sent by client couldn't be considered in any above conditions, we will deal with the header of emails consecutively sent by it (if it has a sequence of emails sent).

Finally, if the situation of ports in sender system is similar to a main server, we will study its conditions for being a mail server relying on traffic properties of email. These conditions, as indicated at the end of algorithm are also similar to primitive conditions of algorithm based on traffic properties of email.

Algorithm DSPOT spam bot detection

```
An outgoing message arrives at DSPOT
Get IP address of sending machine m
// all following parameters specific to machine m
if sender is End-User {
if ( time of sending email is illegal ) then
    Machine m is compromised.
else if (header's from != body's from)then
    Machine m is compromised.
else if (dip != MX_dns (source domain)) then
    Machine m is compromised.
    Let n be the message index
    Let Xn = 1 if message is spam, Xn = 0  otherwise
    if (Xn == 1) then
    // spam, Eq. 3
       + = ln
    ....
    Test continues with new observations
    else
    Test continues with an additional observation

else if (n == N & (70% senders'IDs is different))  then
    Machine m is compromised.}
else if (sip= =MX_dns or SPF_dns(source  domain )) then
    Machine m is Legitimate Mail Server
end if
```

V. Evaluation Of DSPOT Sysytem Efficiency

In this part, we will evaluate the efficiency of DSPOT System. Such evaluation is conducted by studying 2 month of email traffics belonging to an ISP in Iran. DSPOT system performs based on network edge and by running an applied program; it will detect spam bots in the network. An email sending IP in our system is either a spam bot or normal. Upon any conformity of behavior of this sender with parameters as mentioned in algorithm, it is a spam bot, otherwise it is normal.

As indicated in Fig. 4, for evaluating the performance of DSPOT and comparing it with similar systems designed before, we send a copy of output traffic of network by TAP simultaneously to both systems SPOT and DSPOT. Only by studying the traffic of output email on port 25, SPOT and DSPOT will detect spam bots.

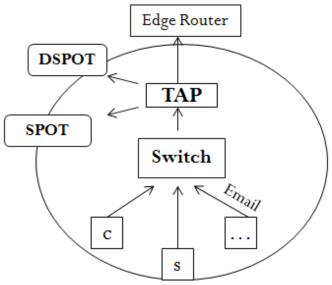


Fig. 4.Comparing the Performance of DSPOT with other Systems

Results from traffic analysis for 21,547,982 emails by DSPOT indicated in TABLE II. As indicated in TABLE II, among 312 IP addresses sending the emails in related network, there were 104 addresses recognized as compromised IP addresses (number of spam bots actually present in the network) and DSPOT could recognize about 100 IP addresses among 104 compromised machines as spam bots. While SPOT has only recognized 93 addresses as spam bot machine (under SMTP communications). Therefore under such conditions and according to this traffic, the detection rate of SPOT system was only 89.4% while the detection rate of our system was 96.1%. Thus, DSPOT system has higher capability for detecting the spam bot machines than SPOT system. As indicated in last column of TABLE II, our system only failed to recognize 4 compromised machines this is while SPOT system wasn't able to recognize 11 compromised machines. Such inability to detection is related to error rate present in both systems, i.e. error rate of SPOT was 10.6% and error rate of DSPOT was only 3.9%. the detection rate is indeed the same true positive and error rate is the same as false negative.

According to above results obtained from comparing the performance of both systems, DSPOT and SPOT, we conclude that logically due to its usage of Hybrid Approach, DSPOT has higher accuracy (higher detection rate and lower error rate) for detecting the spam bots in the network.

VI. Limitations and future work

To facilitate conducting this study, we didn't consider emails sent by dynamic IPs and only analyzed emails belonging to static IPs. But as you know, some part of spam emails are sent by devices with dynamic address. Therefore, detecting spam zombies using dynamic IPs will be studied in a future work.

Because email packets are sent under different protocols and naturally on different ports as well, we aimed to investigate only the traffic of email under SMTP protocol on port 25. Then, there are certainly other spammers sent their emails under other protocols on ports 465 and or 578 and development of DSPOT for activity on such ports will be studied in a future work.

VII. Conclusion

In this study, we developed DSPOT system to detect spam bots in internal network. DSPOT is a developed version of SPOT and by analyzing the output emails of network edge, online, it detects spam zombies. In order for reducing the error rate in this approach, DSPOT analysis has been conducted based on both content of email and their traffic properties.

Performance of DSPOT was evaluated by analyzing the traffic of emails entering to an ISP. Studies indicate that this system has a proper and more accurate performance for detecting the spam bots in the network and can be used as a complementary tool in Unified Threats Management Systems(UTM).

TABLE II. A SUMMARY OF SITUATION OF EMAIL SENDER IPS

System Type	*Total # IP*	*Detected*	*Confirmed (%)*	*Missed (%)*	
DSPOT	312	104	100 (96.1)	4(3.9)	
SPOT	312	104	93 (89.4)	11(10.6)	

REFERENCES

[1] Analysis of Spam Activity Trends, 2011, Available:

http://www.symantec.com/threatreport/topic.jsp>?id=spam_fraud_activity_trends&aid=analysis_of_spam_activity_trends [Accessed: March. 01, 2012].

[2] Percentage of spam sent from Botnets, 2011. Available:

http://www.symantec.com/threatreport/topic.jsp?id=spam_fraud_activity_trends&aid=analysis_of_spam_delivered_by_Botnets [Accessed: March. 02, 2012].

[3] A. Ramachandran, and N. Feamster, "Understanding the Network-Level Behavior of Spammers", in *Proc. ACM SIGCOMM, NY*, USA. 2006, pp 291-302.

[4] Zhenhai Duan, Peng Chen, Fernando Sanchez, Yingfei Dong, Mary Stephenson, James Barker, "Detecting Spam Zombies by Monitoring Outgoing Messages" ,2012, *IEEE Trans. Inform. Theory*, vol 9, pp 198-210.

[5] 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 -November3 2006.

[6] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, “Bothunter: Detecting malware infection through ids-driven dialog correlation,” *in Proc. 16th USENIX Security Symposium*, Boston, MA, Aug. 2007.

[7] Y. Xie, F. Xu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, “Spamming Botnets: Signatures and characteristics,” in *Proc. ACM SIGCOMM*, Seattle, WA, Aug. 2008.

[8] L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang, I. Osipkov, G. Hulten, and J. D. Tygar, “Characterizing Botnets from email spam records” in *Proc. of 1st Usenix Workshop on Large Scale Exploits and Emergent Threats*, San Francisco, CA, Apr. 2008.

[9] S. Linford, “Increasing spam threat from proxy hijacking,” http://www.spamhaus.org/news.lasso?article=156.

[10] J. Klensin, “Simple Mail Transfer Protocol,” RFC 5321, Oct. 2008.

[11] Jim Kurose and Keith Ross, “Application Layer,” in *Computer Networking: A Top Down Approach*, 5th edition, Addison-Wesley , 2010, pp.123-125.

[12] Fernando Sanchez and Zhenhai Duan and Yingfei Dong, “Blocking Spam By Separating End-User Machines from Legitimate Mail Server Machines”, *Proceedings of 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS),* Redmond, Perth, Australia, 2011.