QUERY BASED APPROACH TOWARDS SPAM ATTACKS USING ARTIFICIAL NEURAL NETWORK

Gaurav Kumar Tak and Shashikala Tapaswi

ABV- Indian Institute of Information Technology and Management
Gwalior (M.P.), INDIA
gauravtakswm@gmail.com , stapaswi@hotmail.com

ABSTRACT

Currently, spam and scams are passive attack over the inbox which can initiated to steal some confidential information, to spread Worms, Viruses, Trojans, cookies and Sometimes they are used for phishing attacks. Spam mails are the major issue over mail boxes as well as over the internet. Spam mails can be the cause of phishing attack, hacking of banking accounts, attacks on confidential data. Spamming is growing at a rapid rate since sending a flood of mails is easy and very cheap. Spam mails disturb the mind-peace, waste time and consume various resources e.g., memory space and network bandwidth, so filtering of spam mails is a big issue in cyber security.

This paper presents an novel approach of spam filtering which is based on some query generated approach on the knowledge base and also use some artificial neural network methods to detect the spam mails based on their behavior. analysis of the mail header, cross validation. Proposed methodology includes the 7 several steps which are well defined and achieve the higher accuracy. It works well with all kinds of spam mails (text based spam as well as image spam). Our tested data and experiments results shows promising results, and spam's are detected out at least 98.17 % with 0.12% false positive.

KEYWORDS

Artificial neural network, Spam, Scam, Cross Validation, Virus, Worms & Trojan

1. INTRODUCTION

Now days, Email (Electronic mail) communication plays a great role in the human life due to its fast and free availability, lower or free cost. It is more useful for many corporate because of some features like newsletters, business correspondence, Email marketing, Advertisements etc.

Like Freelancer.com Support use email service for business correspondence to send the emails and messages to its authorized members. Google news alerts use it for the news letter.Naukri.com, DevNetworkIndia.org and etc. use email service for the new jobs advertisements massively. Inkfruit , ZoomIn, Fashnvia.com (India) and etc. use email service for their product marketing and their advertisements. Many times, these mails like Product advertisements, job advertisements, news alerts are meaningful for the email users but sometimes, they generate spam mails over the mail-inbox.

Today, Email and chat services are the most common, instantaneous and successful Internet applications, which are threatened by spam mails and spam chats. These Service can be accessed using mobile internet or low speed internet. Spam mails can be an advertisement or notification of porn website, porn video, phishing website, Nigerian scam, medicines advertisements, adult content etc.

DOI : 10.5121/ijaia.2010.1407

Spammers collect e-mail addresses from chartrooms, public networking websites, customer lists, newsgroups, and worms, viruses which harvest users' address books, and are sold to other spammers. They also use a practice known as "e-mail appending" or "epending" in which they use known information about their target (such as a postal address) to search for the target's e-mail address. Much of spam is sent to invalid e-mail addresses. Spam averages 78% of all e-mail sent [14].

The spam detection problem seems more serious over mailboxes today. Without a spam filter, one email user might receive over hundreds of mails daily and find that most of them are of spam category. Spam mails consume unnecessary traffic over the internet as well as email service provider. Moreover, receiving spam mails are with no use for email users.

In the employed system, a highly simplified architecture of artificial neural networks is used to detect the misbehaviour of incoming mails.

An artificial neural network is a mathematical model which works on the principles of biological neural networks. Generally it is referred as neural network (NN).Using neural network model; we can easily map the complex inputs with the complex outputs.

Some of the silent features of ANN are as follows,

- ✓ They represent a highly connected network of neurons - the basic processing unit.
- ✓ They operate in a highly parallel manner.
- ✓ Each neuron does some amount of information processing.
- ✓ It derives inputs from some other neuron and in return gives its output to other neuron for further processing.
- ✓ This layer-by-layer processing of the information results in great computational capability.
- ✓ As a result of this parallel processing, ANNs are able to achieve great results when applied to real-life problems.

A typical architecture of neural network is depicted in figure1.

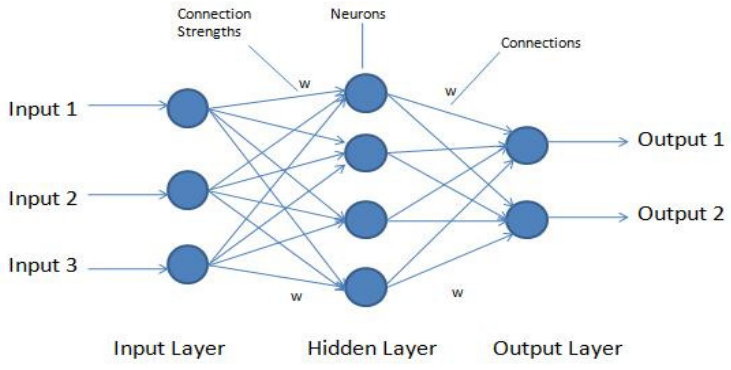


Figure 1. Architecture of Neural Network

Neural network performs its operations in two phases: learning phase and testing phase.

Learning Phase: In the proposed methodology, we have taught several SQL attacks to the network in a supervised manner. We entrust the system with several variants of any attack and assign it a particular *label.* Thus we can see that system learns by feeding various patterns of the same attack.

During the *training process of neural network*, matrix of inbox mails and spam mails is used as input matrix to the neural network. In the proposed methodology, the input matrix is updated after defined time interval.

Any neural network adjusts the weights of attacks in order to learn in a supervised or unsupervised manner.

In our method of learning, each candidate attack taught to the network is associated with a weight matrix. Weight matrix associated with the k^{th} spam is assigned the label $W_{k.}$. Weight matrix is updated with the progress of the learning of the spam mail. This matrix is initialized to zero when learning phase starts. An input pattern corresponding to the spam is taught to the submitted to the network.

According to information compiled by Commtouch Software Ltd., E-mail spam for the first quarter of 2010 can be broken down as follows[15].

Table 1. E-mail Spam by Topic

Pharmacy	81%
Replica	5.40%
Enhancers	2.30%
Degrees	1.30%
Casino	1%
Phishing	2.30%
Weight Loss	0.40%
Other	6.30%

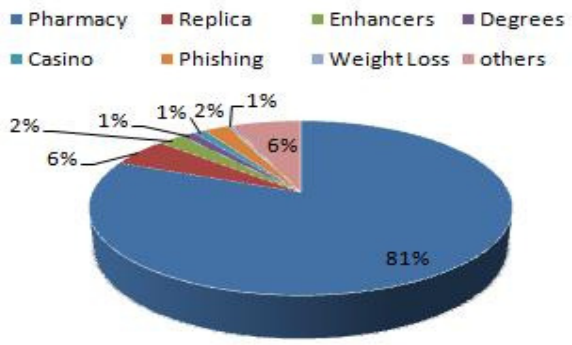


Figure 2**.** Spam e-mails distribution by topic

Due to following characteristics, currently the identification process of spam mails is a difficult problem [3].

[1] Spam heterogeneity
[2] Spam definition

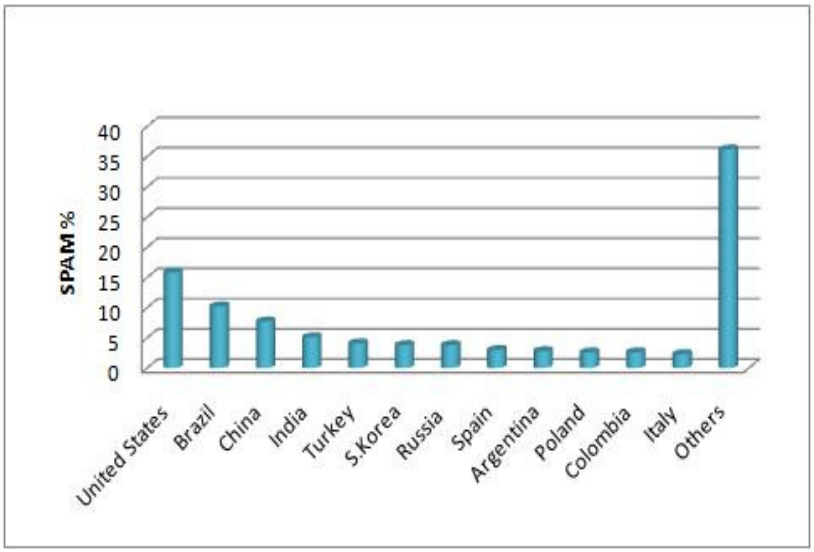


Figure 3**.** Represents the spam distribution over various countries.

By continent, Asia continues to dominate in spam, with more than a third of the world's unsolicited junk email relayed by the region. Asia covers 34.8% spam mails over all the spam mails. The breakdown of spam relaying by continent is as follows [6]:

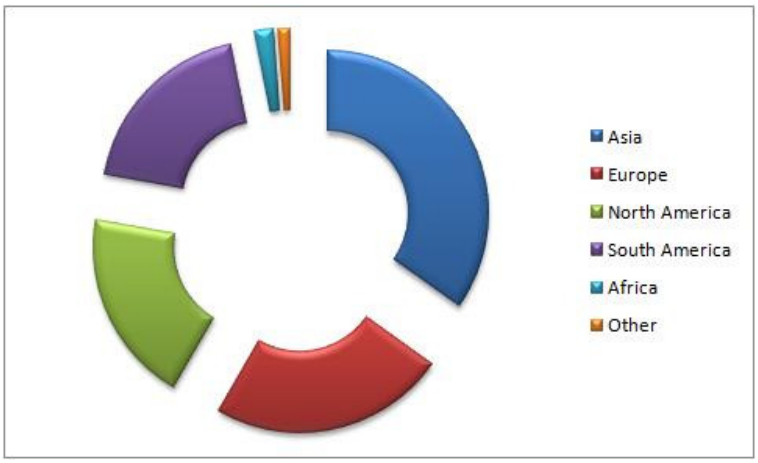


Figure 4. Spam distribution over various regions.

2. SPAMMER APPROACHES AND THEIR ATTACK

There are many techniques adopted by the spammer or attackers to collect and store the email addresses or personal information etc. Some of those approaches are from posts to UseNet with

email address, from mailing lists, from web pages, from various web and paper forms, via an Ident daemon, from a web browser, from IRC and chat rooms, from finger daemons, from AOL profiles, from domain contact points, by guessing & cleaning, from white & yellow pages, from a previous owner of the email address, by having access to the same computer, using social engineering, from the address book and emails on other people's computers, buying lists from others, by hacking into sites and etc.

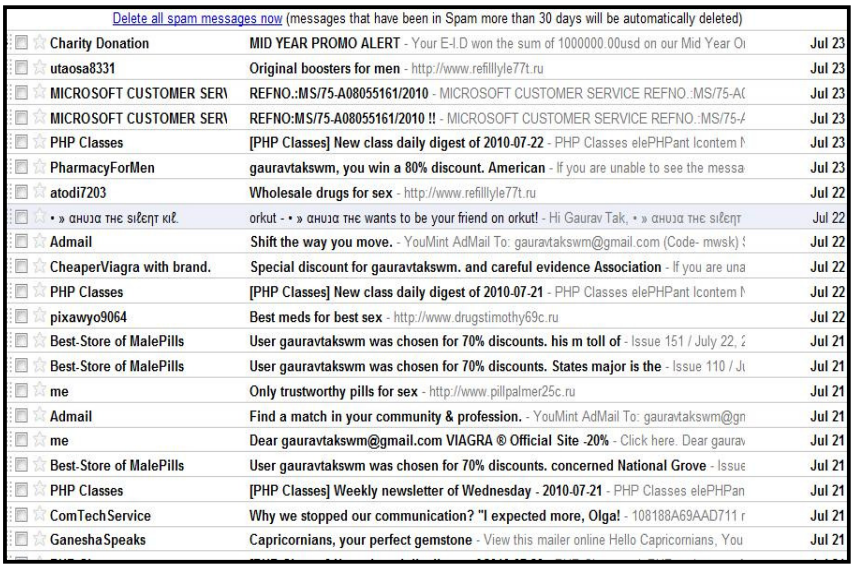


Figure 5**.** A spam box folder over the mailbox.

With a marketing service, a person can arrange his contacts by certain demographics so that he can create custom mailing lists. This means that he can have some newsletters that go to all customers while also having some that only go to women or men or people with a history of shopping in a particular category. These tailored mailing lists ensure that your messages are only received by customers who may be interested in the subject matter, keeping those who likely would not be from feeling as though they are being spammed and unsubscribing.

Currently, a lot of social networking sites exits over WWW. Some sites are really useful but some creates spam mails over the mailbox. With social networking sites ,when a person joins some social networking website (like shtyle.fm , yaari.com, indiarocks.com, mycantos.com, facebook.com, tagged.com etc.), then these social site use some script to approach contacts (contact mail list) of that person and send invitation to his contacts to join the same social site. Many times they fill spam mails in peoples' inbox using this approach. There are also many several attacks over the mailbox by the spammers. Some spammers generate spam mails over the mailbox using the manual script but some use machine generated scripts to generate the spam mails.

3. RELATED WORK

In literature, there are many techniques described for the detection of spam and mail filtering. Some of the techniques are described as follows:

In [16], A Rule approach has been proposed for the detection of spam mails. The discussed approach uses the training and testing phases of data. Moreover, the stale and obsolete spam rules suspend during the training. This action is used for improving the spam filtering

efficiency. However, the time complexity is higher due to the rules generation and their execution. E. Damiani et al. discussed some basic properties of the spam mails. They focused on the reasons of the popularity of spam mails. The uses of the digests in the proposed approach to identify spam mails in a privacy-preserving way is a fundamental technique for collaborative ltering[3].A social network is constructed based on email exchanges between various users in [11][12]. Spammers are identified by observing abnormalities in the structural properties of the network. Many times spammer uses the public social sites for increasing their mail list database. However, it is a reactive mechanism since spammers are identified after they have already sent spam. In [13] a novel approach has been discussed, which creates a Bayesian network out of email exchanges to detect spam. Though Bayesian classifiers can be used for detecting spam emails, they inherently need to scan the contents of the email to compute the probability distributions for every node in the network. Since many times it is not possible, to detect spam mails for the particular inbox and its requirement for filtering the spam mails [4].

Nitin Jindal et al. discussed an approach of review spam. Review spam is quite

different from Web page spam and email spam, and thus requires different detection techniques[17]. There is an effective technique to detect the spam mail that is 'Fast Effective Botnet Spam Detection'. It uses the header information of mails to detect the spam mails. It is useful for both 'Text based spam' as well as 'image based spam'. It analyzes the sender IP address, sender email address, MX records and MX hosts [1].

One approach is also described to detect the spam mails, it use the Bayesian calculation for single keyword sets and multiple keywords sets, along with its keyword contexts to improve the spam detection [5].

4. PROPOSED METHODOLOGY

Before proposing a new methodology for spam detection, we are aware of this fact that most of time spam mails and scams are spread out using the machine generated script. In this paper, we are proposing a new query based cross layer approach for the above that is based on the above facts and some other spam features.

Our system uses some knowledge base and query generation using the history of previous mails and spam mails which is specific for the each user or its mailbox. Using the knowledge base, detection of spam mails is performed. It also maintains some keywords list, which can easily be pointed out as some words or content in the incoming mail, then perform the detection operation.

Many times when a person clicks a URL which is present in his mailbox, (that URL has been provided by the spammers) then mail address of the person is captured by the spammer and is easily inserted in spammer's database.

Proposed spam detection approach, follow the few steps to indentify the spam mails which are as follows:

1) **Analyze the mail content**: Firstly, proposed approach analyze the mail content and sender mail address of the mail, then cross analyze and compare the content and sender address of the previous spam mails if content and sender address both are already present in any of the previous spam mails then it directly declares the mail as ***"a spam"*** (a spam is already present with the same sender and same mail content)**.**

If the some fraction of incoming mail content matches with the any previous spam mail then mail is filtered using the spam threshold value (S_t). The spam threshold value can be defined as the mathematical value which decides the performance and accuracy of spam detection system. It can be different for various systems. It is used to indentify the spam mails with the partially matching case.

If S_t =0.7 and matching fraction of the content of mail matches with the previous declared spam mails is greater than equal to 0.7, then the mail is declared as ***"a spam"***.

Matching fraction of the content= max. $(NM_1/N_1, NM_2/N_2, \ldots, NM_p/N_p)$

NM_p: Total number of exactly matched words of incoming mail with the p-th spam mail.

N_p: Total number of words in p-th spam mail.

P: The total no. of recent mails which are available in the spam mail list corresponding to that user.

Using the analysis step ,following mail from PHP-classes is detected as spam mail because it was already present in the spam folder and user never communicated with the sender mail id.

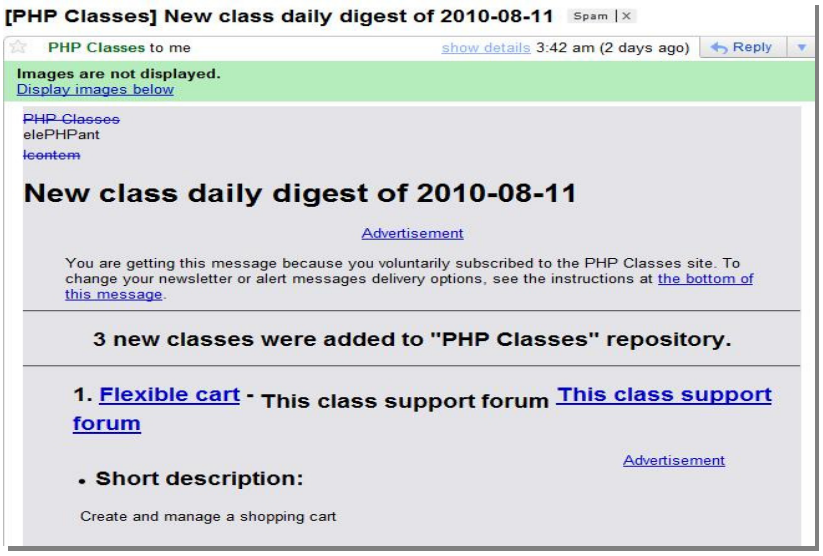


Figure 6**.** A Spam mail from the PHP Classes

2) **Trusted Knowledge Base**: Knowledge Base is always a good, efficient and faster approach to give the results based on historical data. It is used some queries to execute the results. It also follow some update operation to make the result efficient based on the system requirements.

In the Trusted Knowledge Base, database of trusted sender is stored over the inbox based on the frequency of the communication of mails. The Knowledge Base is also updated upon the requirement of inbox or threshold count of incoming mails.

This Knowledge Base is responsible to the detection of spam mails when sender of incoming is already kept in the trusted zone.

If the sender is not the trusted sender then next steps would be executed to indentify the spam mails.

3) **Keywords knowledge Base**: To execute this step, A knowledge base is maintained at mail server for each user which stores the spam keywords (already defined by the specific user).During this step, proposed approach analyzes the keywords of mails with the keywords knowledge base of spam which is prepared by the particular user for detection of spam. Using the result it decides that incoming mail belongs to the spam category or not. If incoming mail has not been declared as "***spam***" then execute the other steps to indentify the spam mails.

4) **Sender mail address**: Our proposed methodology extract the sender mail address using the mail header (check the *from* field or *reply-to* field to get the sender email address) and analyze it to indentify the spam. Using the sender email address, system finds that have any communication been done previously between receiver and sender or not? If receiver has already communicated with that mail address, then mail is declared as "***not a spam***". But if receiver has never communicated, then system explores the contact list of the receiver.

If the sender mail address already present in the contact list then the mail is declared as "***not a spam***"**.**

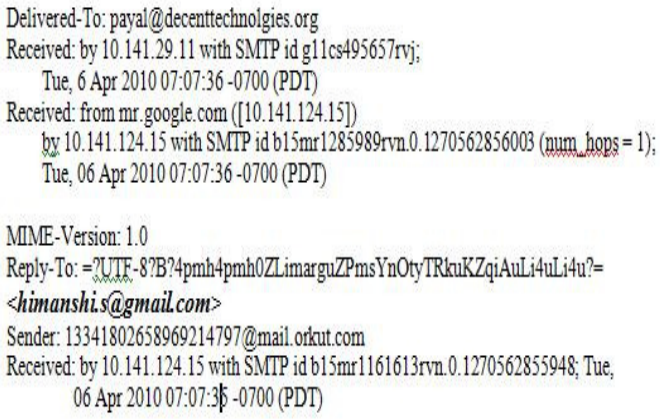
This step is very useful with the public networking site because many times networking sites send invitation using someone contacts.

In the given figure, It is shown that we have received the spam mail in the inbox of vikas@decenttechnologies.org from Skoot.com server and our proposed approach is able to detect the spam easily.

```
Delivered-To: vikas@decenttechnolgies.org
Return-Path: <skoost@skoost.com>
Received: from skoismta10.skoost.com (mx198.skoost.com
[80.248.17.198])
by mx.google.com with ESMTP id t10si1017262rvl.81.2010.04.14.19.41.23;
Wed, 14 Apr 2010 19:41:23 -0700 (PDT)
Received: from skoissql01 (unknown [80.248.18.51])
by skoismta10.skoost.com (Postfix) with ESMTP id AFA3B2ACDBC
for < vikas@decenttechnolgies.org >; Thu, 12 Mar 2010 02:41:22 +0000
(GMT)
MIME-Version: 1.0
From: Skoost <skoost@skoost.com>
Sender: Skoost <skoost@skoost.com>
To: "vikas@decenttechnolgies.org " < vikas@decenttechnolgies.org >
Reply-To: Skoost <skoost@skoost.com>
Date: 12 Mar 2010 02:41:22 +0000
Subject: A gift box - Skoost
Content-Type: multipart/alternative;
boundary=--boundary_6430361_eb7ef93d-fac4-4e36-bbca-0ea779a0dfbf
Message-Id: <20100415024122.AFA3B2ACDBC@skoismta10.skoost.com>


----boundary_6430361_eb7ef93d-fac4-4e36-bbca-0ea779a0dfbf
Content-Type: text/plain; charset=utf-8
Content-Transfer-Encoding: quoted-printable
```

Figure 7. Extracted mail header of the inbox "vikas@decenttechnoloies.org"



Delivered-To: payal@decenttechnolgies.org
Received: by 10.141.29.11 with SMTP id g11cs495657rvj;
Tue, 6 Apr 2010 07:07:36 -0700 (PDT)
Received: from mr.google.com ([10.141.124.15])
by 10.141.124.15 with SMTP id b15mr1285989rvn.0.1270562856003 (num_hops = 1);
Tue, 06 Apr 2010 07:07:36 -0700 (PDT)

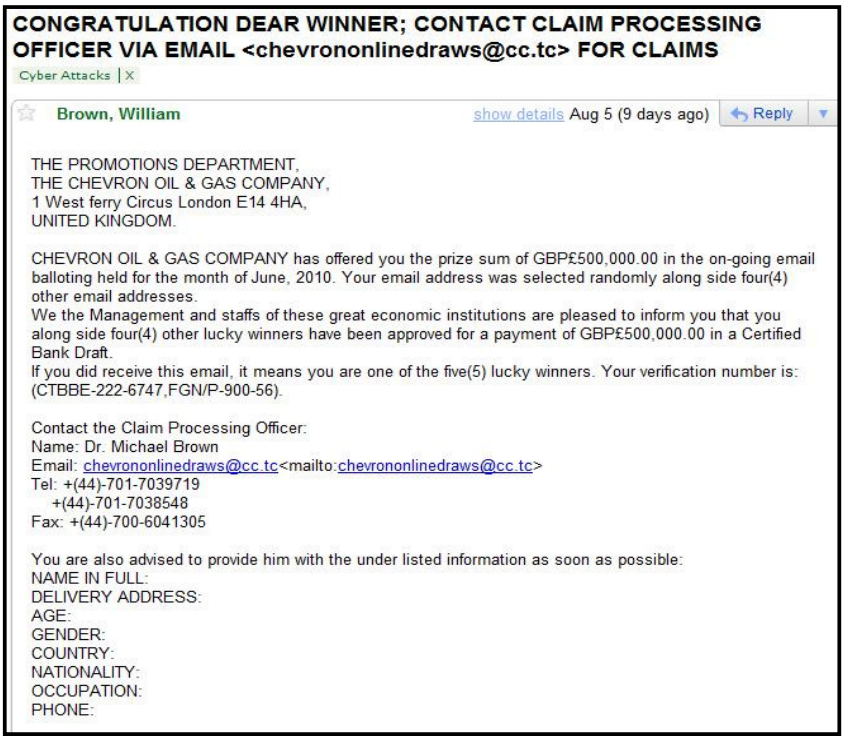
MIME-Version: 1.0
Reply-To: =?UTF-8?B?4pmh4pmh0ZLimarguZPmsYnOtyTRkuKZqiAuLi4uLi4u?=
<himanshi.s@gmail.com>
Sender: 13341802658969214797@mail.orkut.com
Received: by 10.141.124.15 with SMTP id b15mr1161613rvn.0.1270562855948; Tue,
06 Apr 2010 07:07:35 -0700 (PDT)

Figure 8**.** Extracted mail header of the inbox "payal@decenttechnoloies.org"

In the above example, user payal (receipt email address: payal@decenttechnologies.org) has sender user himanshi.s@gmail.com in her contact list. So the received mail will be declared as "***not a spam***".

5) Sender Location: This step is useful when mail user receive a mail from the another country which already belongs to the spam mail country. Our approach finds the sender mail server location and then compares the location with the spam mails location. Using this step, we are able to filter out some lottery spam and some Nigerian scams too.

Using this step following mail is easily detected as spam mail because nation of mail inbox is INDIA and incoming mail server exists in US and receiver has never communicated with the US mail sender so it can be detected as spam mail.



CONGRATULATION DEAR WINNER; CONTACT CLAIM PROCESSING OFFICER VIA EMAIL <chevrononlinedraws@cc.tc> FOR CLAIMS

Cyber Attacks | X

Brown, William show details Aug 5 (9 days ago) Reply

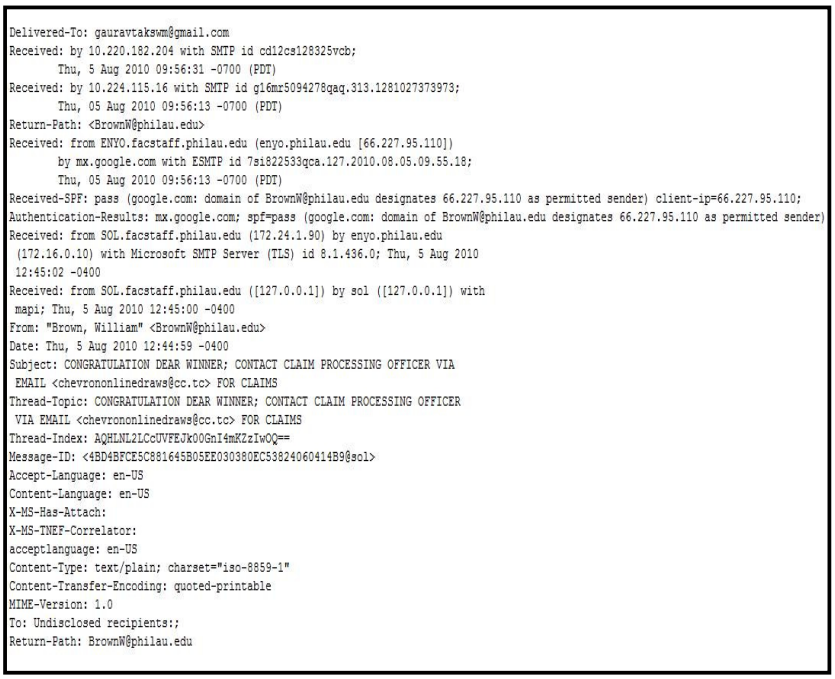
THE PROMOTIONS DEPARTMENT,
THE CHEVRON OIL & GAS COMPANY,
1 West ferry Circus London E14 4HA,
UNITED KINGDOM.

CHEVRON OIL & GAS COMPANY has offered you the prize sum of GBP£500,000.00 in the on-going email balloting held for the month of June, 2010. Your email address was selected randomly along side four(4) other email addresses.
We the Management and staffs of these great economic institutions are pleased to inform you that you along side four(4) other lucky winners have been approved for a payment of GBP£500,000.00 in a Certified Bank Draft.
If you did receive this email, it means you are one of the five(5) lucky winners. Your verification number is: (CTBBE-222-6747,FGN/P-900-56).

Contact the Claim Processing Officer:
Name: Dr. Michael Brown
Email: chevrononlinedraws@cc.tc<mailto:chevrononlinedraws@cc.tc>
Tel: +(44)-701-7039719
+(44)-701-7038548
Fax: +(44)-700-6041305

You are also advised to provide him with the under listed information as soon as possible:
NAME IN FULL:
DELIVERY ADDRESS:
AGE:
GENDER:
COUNTRY:
NATIONALITY:
OCCUPATION:
PHONE:

Figure 9. A Lottery Spam to capture personal information of mail user.



```
Delivered-To: gauravtakswm@gmail.com
Received: by 10.220.182.204 with SMTP id cd12cs128325vcb;
        Thu, 5 Aug 2010 09:56:31 -0700 (PDT)
Received: by 10.224.115.16 with SMTP id g16mr5094278qaq.313.1281027373973;
        Thu, 05 Aug 2010 09:56:13 -0700 (PDT)
Return-Path: <BrownW@philau.edu>
Received: from ENYO.facstaff.philau.edu (enyo.philau.edu [66.227.95.110])
        by mx.google.com with ESMTP id 7si822533qca.127.2010.08.05.09.55.18;
        Thu, 05 Aug 2010 09:56:13 -0700 (PDT)
Received-SPF: pass (google.com: domain of BrownW@philau.edu designates 66.227.95.110 as permitted sender) client-ip=66.227.95.110;
Authentication-Results: mx.google.com; spf=pass (google.com: domain of BrownW@philau.edu designates 66.227.95.110 as permitted sender)
Received: from SOL.facstaff.philau.edu (172.24.1.90) by enyo.philau.edu
 (172.16.0.10) with Microsoft SMTP Server (TLS) id 8.1.436.0; Thu, 5 Aug 2010
 12:45:02 -0400
Received: from SOL.facstaff.philau.edu ([127.0.0.1]) by sol ([127.0.0.1]) with
 mapi; Thu, 5 Aug 2010 12:45:00 -0400
From: "Brown, William" <BrownW@philau.edu>
Date: Thu, 5 Aug 2010 12:44:59 -0400
Subject: CONGRATULATION DEAR WINNER; CONTACT CLAIM PROCESSING OFFICER VIA
 EMAIL <chevrononlinedraws@cc.tc> FOR CLAIMS
Thread-Topic: CONGRATULATION DEAR WINNER; CONTACT CLAIM PROCESSING OFFICER
 VIA EMAIL <chevrononlinedraws@cc.tc> FOR CLAIMS
Thread-Index: AQHLNL2LCcUVFEJk00GnI4mKZzIwOQ==
Message-ID: <4BD4BFCE5C881645B05EE030380EC53824060414B9@sol>
Accept-Language: en-US
Content-Language: en-US
X-MS-Has-Attach:
X-MS-TNEF-Correlator:
acceptlanguage: en-US
Content-Type: text/plain; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
MIME-Version: 1.0
To: Undisclosed recipients:;
Return-Path: BrownW@philau.edu
```

Figure 10. Lottery Spam header to find out the sender mail location.

Many mail server use the sender location approach to indentify the spam so they ask to the users country and location at the time of mail registration.

6) **Misbehaviour of incoming mail**: This step is executed using the artificial neural network. Artificial Neural Network (ANN) is a scientific discipline that is concerned with the design and development of algorithms that allow computers to adapt their behaviour based on data. ANN automatically learns to recognize complex patterns and makes intelligent and efficient decisions based on data.

   In the spam filtering ANN learns the complex pattern of mails and makes intelligent, efficient decisions based on the incoming mail.

   Proposed methodology executes training phase testing phase using sample set of the mailbox to complete this step.

   During this step, we are able to predict any misbehaviour event of incoming mails; Machines generated mails, flood of mails over inbox. Misbehaviour can be predicted using the time factor, some sender mail address, some attacks.

   To detect the Misbehaviour, training phase is executed after each threshold value of incoming mail over inbox.

7) **Cross Validation**: During this step, system will verify the sender that sender is a genuine human user or machine generated user using some cross request.

   If the incoming mail is machine generated email, it implies that sender is not human user. So the machine generated mails are not able to validate their identity. Most of the spam mails are detected during this step.

5. IMPLEMENTATION AND ANALYSIS

We have conducted the analysis of spam mails using the proposed methodology on some inboxes of different peoples We have created the environment using some web technologies HTML, script languages, AJAX, XML and MySql tools for implementing the methodology. We also applied some basic concepts of PHP, AJAX, MySQL and JavaScript from the references [7] [8]. **Figure**8 represents the diagrammatic representation of the proposed methodology.

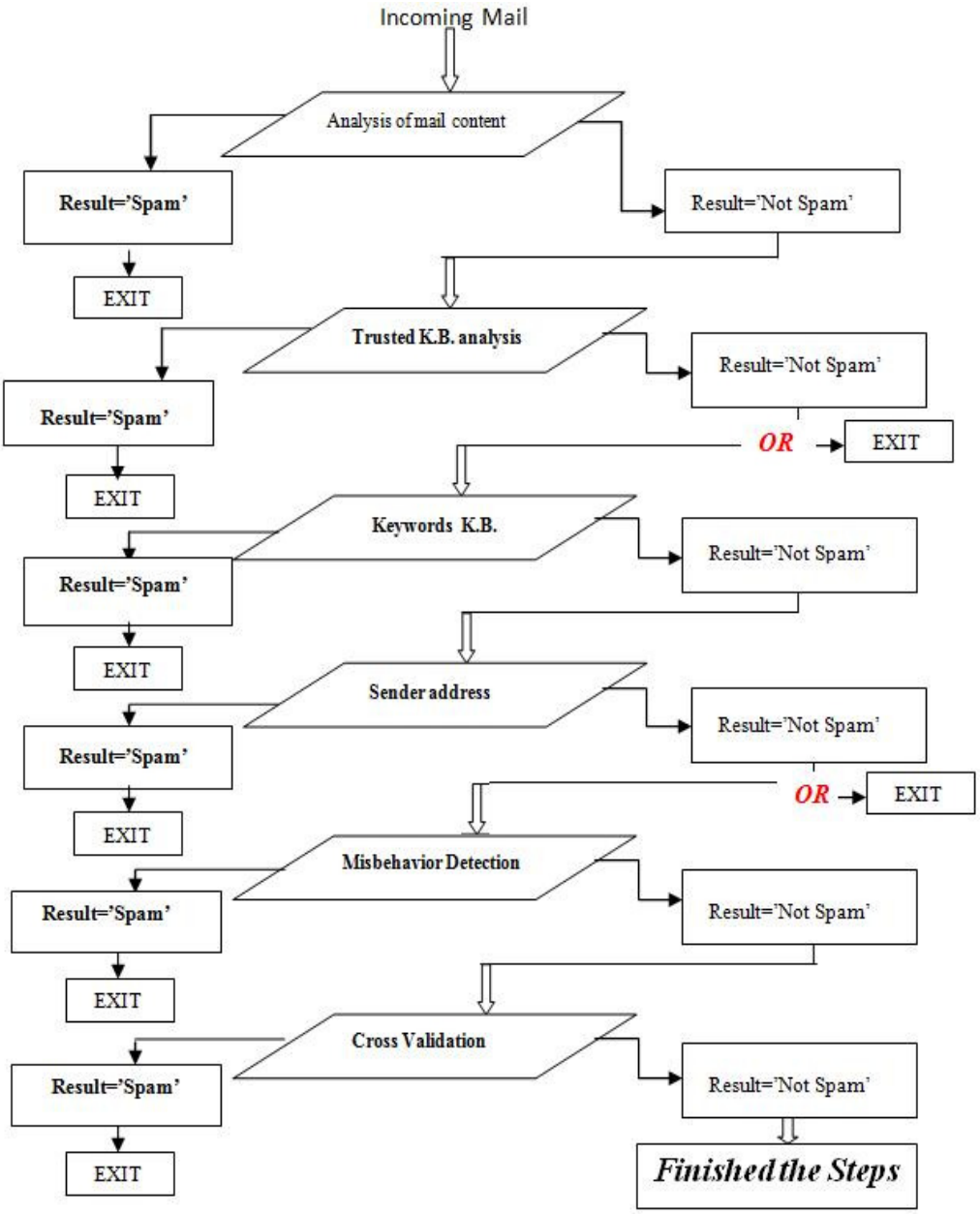


Figure 11. Diagrammatic representation of the proposed methodology.

1) Extract Mail Content

- Analyze the matching pattern and calculate the matching fraction with the previous spam mail and then compare the matching fraction with the spam threshold value.

If (matching fraction > = S_t)

then mail = 'spam';

Exit;

else

Go to step2;

2) To find the sender belongs to the trusted zone of the specific user then it performs some query operations. The Trusted Knowledge Base is responsible to maintain the status of the sender user. This Knowledge base is created using some frequent and recent received and sent mails

If (sender exists in trusted knowledge base)

then mail='not a spam';

exit;

else

Go to step3;

3) Analysis the mail content using 'spam keywords knowledge base (already declared by the user).

If (mail content matches with the spam keywords knowledge base)

then mail= 'spam';

exit;

else

Go to step4;

4) Analysis the sender mail addresses using the contact list and previous received mails.

-Extract mail header then Separate sender mail address.

If (sender mail address is available in (contact list or previous communicated mails)

then mail = 'not a spam';

else

Go to step5;

5) Detect the spam mail using the Sender location step.

Sender_Location ()

{

S_location=find_location();

/* Using some crawling operations over the internet*/

/*Find the location of Sender mail server using the mail header*/

If(S_location not belongs to the receiver Location/nation)

Then mail = 'a spam';

Else

then mail = 'not a spam';

/ (sender belongs to the receiver location) */*

Go to step6;

}

6) It is the complex step of artificial network; we are not able to map the step using the functions. The step is executed using some artificial tools and API.

7) Detect the spam mail using the cross validation approach.

Cross validation ()

{

Send (simple equation / puzzle, sender mail address)

If (validation=true)

Then mail = 'not a spam';

Else

then mail = 'a spam';

/ (sender is a machine user) */*

}

We have recorded the incoming mail activities and sender mail addresses over 4 months (Apr,2010 to july,2010) at mailbox of an organization. We have not implemented our proposed methodology for the detection of spam mails in Apr,2010 but during May,2010 and July,2010 , we have implemented it and recorded the activities of incoming mails and also analyzed the behavior of incoming for the artificial neural network step. The following table data represents the recorded activities over the various mailboxes.

Table 2. Represents the data of recorded activities over mailboxes.

Month	Apr,2010	May, 2010	Jun , 2010	July, 2010
Inbox	15870	17961	18460	17123
Spam	4692	7234	7494	7031
False Match	83	43	23	29
Total mail	19562	25195	25954	23157
% Spam Caught	24.8 %	28.7%	28.9%	30.4%
% False Match	0.42 %	0.17%	0.089%	0.099%

We can get the performance information of the proposed methodology using the experimented results which are shown in table2.We can easily compare these results and performance with the previously described approaches of spam filtering.

Fig 12 and 13 represents all the complete scenario of experiments results.

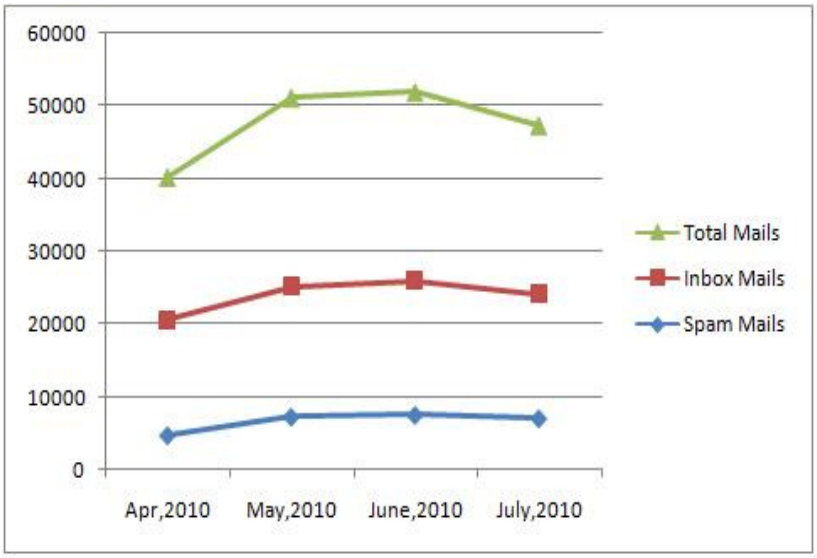


Figure 12. Analysis of Total mails, Inbox Mails, Spam Mails over mailboxes.

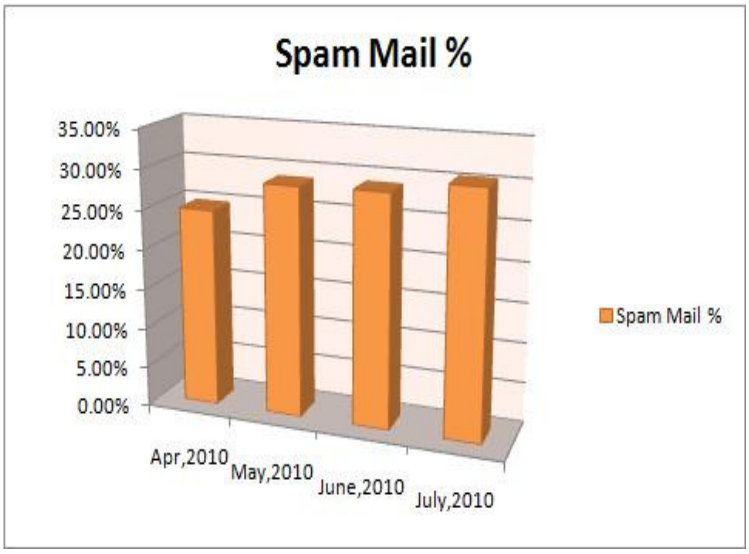


Figure 13. Analysis of Spam mails % over the Apr,2010-July,2010 months.

6. CONCLUSIONS AND LIMITATION

Our work is inspired by a situation of large number of spam mails over the mailbox, those we have easily encountered. We have recorded the incoming mail activities of various mail boxes of an university server over 4 months and analyzed those mails to get the better results and better performance of spam filtering. From table data, we can all results of spam mails, inbox mails, false match easily for the given time period. The experiment results provide the complete scenario of the problem and accuracy of spam detection. Our system indicated that the spam

was filtered out with 98.17 % with 0.12% false positive. Table 2 represents the recorded data over the 4 months time period.

Limitation of the proposed method is that it needs more hardware for the execution and higher memory space. So many times, it increases the workload of the mail server. So to implement the proposed methodology for large mail servers, we need intelligent mail servers which are can be reduced the time complexity and provide better performance of spam filtering, So that we can easily manage higher computation load. Due to more hardware specification and higher computation load, the cost of implementation of proposed methodology is much higher.

ACKNOWLEDGEMENT

The authors would like to thank ABV-Indian Institute of Information Technology and Management, Gwalior for the support provided for this work.

REFERENCES

[1] Kobkiat Saraubon, Benchaphon Limthanmaphon, "Fast Effective Botnet Spam Detection," iccit, pp.1066-1070, 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, 2009.

[2] *Chun-Chao Yeh, Chia-Hui Lin, "Near-Duplicate Mail Detection Based on URL Information for Spam Filtering", ,pp. 842-851, Volume 3961/2006,Information Networking. Advances in Data Communications and Wireless Networks, Book Series: Lecture Notes in Computer Science.*

[3] DAMIANI DE CAPITANI , E. DAMIANI , S. DE , CAPITANI VIMERCATI , S. PARABOSCHI , P. SAMARATI, "AN OPEN DIGEST-BASED TECHNIQUE FOR SPAM DETECTION",IN PROCEEDINGS OF INTERNATIONAL WORKSHOP ON SECURITY IN PARALLEL AND DISTRIBUTED SYSTEMS ,2004.

[4] G. Kesidis , A. Tangpong , C. Griffin, A sybil-proof referral system based on multiplicative reputation chains, IEEE Communications Letters, v.13 n.11, p.862-864, November 2009.

[5] *Biju Issac, Wendy Japutra Jap, Jofry Hadi Sutanto, "Improved Bayesian Anti-Spam Filter ," iccet, vol. 2, pp.326-330, 2009 International Conference on Computer Engineering and Technology, 2009.*

[6] http://www.sophos.com/pressoffice/news/articles/2009/04/dirtydozen.html.

[7] PHP, AJAX, MySql and JavaScript Tutorials, http://www.w3schools.com/

[8] Luis von Ahn, Manuel Blum, Nicholas Hopper, and John Langford. CAPTCHA: Using Hard AI Problems for Security. In Eurocrypt.

[9] Weinstein, L.: Inside risks: Spam wars. Communication of ACM, Vol. 46, No. 8,(2003) 136–136.

[10] Corbato, F.J.: On computer system challenges. Journal of ACM, vol. 50, No. 1,(2003) 30–31.

[11] A. J. O'Donnell,W. Mankowski, and J. Abrahamson. Using e-mail social network analysis for detecting nauthorized accounts. In Third Conference on Email and Anti-Spam,Mountain View, CA, July 2006.

[12] P. O. Boykin and V. P. Roychowdhury. Leveraging social networks to fight spam. Computer, 38(4):61–68, Apr. 2005..

[13] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. A Bayesian approach to filtering junk E-mail. In Learning for Text Categorization: Papers from the 1998 Workshop,Madison, Wisconsin, 1998.

[14] "A Brief History of Spam - Time". 2009-11-02. http://www.time.com/time/business/article/0,8599,1933796,00.html. Retrieved 2010-05-01.

[15] "Q1 2010 Internet Threats Trend Report". http://www.commtouch.com/download/1679. Retrieved 2010-05-18.

[16] Yu-Fen Chiu , Chia-Mei Chen, Bingchiang Jeng, Hsiao-Chung Lin, "An Alliance-based Anti-Spam Approach",Third International Conference on Natural Computation(ICNC2007),2007 IEEE.

[17] Nitin Jindal,Bing Liu, "Analyzing and Detecting Review Spam", Seventh IEEE International Conference on Data Mining(ICDM),2007 IEEE.

Authors

Gaurav Kumar Tak is a student of 5th Year Integrated Post Graduate Course (B.Tech. + M.Tech. in Information and Communication Technology) in ABV-Indian Institute of Information Technology and Management Gwalior, India. His primary research areas of interest are Cyber Crime and Security, Wireless Ad-Hoc Network, Web Technologies.

S. Tapaswi is Professor in IT Dept., ABV-IIITM, Gwalior, India. She earned her Ph.D. (Computer Engineering) from Indian Institute of Technology, Roorkee, India in 2002, M.Tech (Computer Science) from University of Delhi, India in 1993 and B.E. from MITS, Gwalior, India in 1986. Her primary research areas of interest are AI, ANNs, Fuzzy Logic, Digital Image Processing, Computer Networks, Mobile Networks, Adhoc networks, Information Security etc.