

A REVIEW PAPER ON IMAGE SPAM FILTERING

Sneha Nikam¹ and Rujata Choudhari²

Abstract- Now-a-days, use of internet has become an inseparable part of our day-to-day life. Wide usage of internet such as, social network, Emails, Marketing, Advertising etc., results in online communication and information exchange. This attracts spiteful users to social network as well, leading towards sending unrequested messages in bulk quantity, with obtaining unambiguous permission of receiver i.e. "Spam". Spam is nothing but unwanted messages. Over a last decade, unwanted messages, has become a major problem for users e.g., large amount of spam E-mails into user's mailboxes. Not necessary Spam will be in textual form always, it can be in image form as well, e.g. advertisement text in images attached to emails. There are various methodologies have been developed to fight with this problem. The main purpose to write this paper is to take review of what is image spam, working of spam filters, methods to identify spam along with Spam filtering techniques for an image based spam.

Keywords – *Spam; Email spam; Image spam; spam filter.*

I. INTRODUCTION

Today, the Internet is one of the most effective and efficient way of communication. Whether it is through Facebook, MySpace, Yahoo, or another website, the internet gives us the opportunity to connect with all people to give information about all over the world. While the Internet has many good qualities and can be very useful, it can also be a source of trouble. We can't imagine living without e-mails and video calls. When we talk about electronic spam, we are talking about electronic junk mail or junk newsgroup postings. The fact that an e-mail is classified as spam on two basic criteria. A) It is unwanted. B) This is the deciding factor. These emails are sent in large quantities to numerous recipients. A major negative effect of electronic spam is that it uses a lot of network bandwidth. Moreover, it may include malware in the form of scripts or other executable attachments. After opening or downloading such e-mails, we might mistakenly invite viruses to enter into our system.

Technically, the term "spam" is Internet slang that refers to unsolicited commercial email (UCE) or unsolicited bulk email (UBE). People generally refer to this kind of communication as junk email to equate it with the paper junk mail that comes through the US Mail. Unwanted email most often contains advertisements for services or products, but few well-known marketers use UCE for advertisement, e.g. offers of software for collecting email addresses and sending UCE, Other "Get Rich Quick" or "Make Money Fast" (MMF) schemes etc. Generally Spam found in various categories like text messages. Spammers are constantly expanding the variety of their offer and inventing new ways to attract victims. Now-a-days spammers have started use of images to hide the fake messages, e.g. hiding the spam message into images which are sent as email attachments. Many a times filtering programs can't identify such spam emails, but the hidden message becomes visible to the recipients when email will get open by them. Successful design of spam filtering software will be helpful to avoid receiving such type of spam messages.

Hence the aim of this paper is to survey the current literature in the field of spam with focus on specific spam filtering techniques used in image spam. The views presented here is to show to the best of our knowledge how much work has been done in each of the spam domains.

¹ Department of Computer Engineering St. Francis Institute of Engineering, Mumbai, Maharashtra, India

² Department of Computer Engineering St. Francis Institute of Engineering, Mumbai, Maharashtra, India

Rest of the paper is organized as following. Section 2 and Section 3 summarizes the description about image spam and how exactly the spam filter works. Section 4 gives details about spam filtering Techniques. Section 5 provides conclusion and future research directions.

II. WHAT IS IMAGE SPAM?

Image spam is a type of spam, or rather, a spamming technique, in which a spam message is delivered in the form of an image. It is junk email that replaces text with images as a means of fooling spam filters. This is done in an attempt to by-pass spam filters that scan for particular keywords. In image spam, Image delivery works by embedding code in an HTML message that links to an image file on the Web. Image spam takes larger no of network resources than text spam because image files are larger than ASCII character strings. Larger files require more bandwidth and, as a consequence, cause greater degradation of transfer rates.

Image-Based Spam contains its un-wanted content inside an embedded graphic file (typically appears in GIF format, but can also appear as JPG, PNG, BMP etc.), making it difficult for some spam filters to identify. These unsolicited e-mails contain no relevant text or hyperlinks. The message may appear to be a text message (see example fig 1(a) & fig 1(b)); however, it is merely an image of text.

An image spam email is formatted in HTML, which usually has only non- doubtful text with an embedded image which later sent either as an attachment or via the use of self-referencing HTML with image data in its consignment. The embedded image carries the target message and most email clients display the message



Figure 1. . (a) Image containing only text



Figure 1. (b) Image spam with photographic elements

Recently, spammers have been experimenting with new techniques such as “broken images,” i.e. splitting a single image into smaller images that it together like puzzle pieces. This technique makes it even more difficult for text based anti-spam engines to detect and block.

Another visible technique is to send animated GIFs, with multiple frames of random “noise”. These random pixels act similarly to the randomized images that are not animated, simply with another level of complexity. In some cases, the animated GIFs contain subconscious messages (e.g. “buy... buy... buy”) embedded into frames that

bind by very quickly. Animated GIF spam is much heavier, on average, than static Image-Based Spam. There are many techniques spammers used to create spam images, some types of spam images are discuss as follows:

A. Types of Spam Images

- *Text Only Images*- Some images contain only text



Figure.2.1 Text only image spam

- *Randomization* – To prevent signature based anti-spam devices, spammers add random color stripes, random colored pixels, shades of colors.



Figure. 2.2 Image with random color pixel

- *Wild backgrounds*- It is difficult for OCR to detect Text in the images



Figure. 2.3 Image with wild background

- *Animated gif and multipart Images*- The images is split into multiple parts, some containing the message and others containing some animation. The frames in the image rotate fast enough to display only the final result to the user.
- *Standard Images*- These are neat looking images, none of the above tricks are employed and that gives it a genuine look. The entire message is contained in the image and hence scanners cannot detect it. In fact many of the images that come in as spam today have a professional look making them look just like photographs. Figure 2.4 below give two such examples.

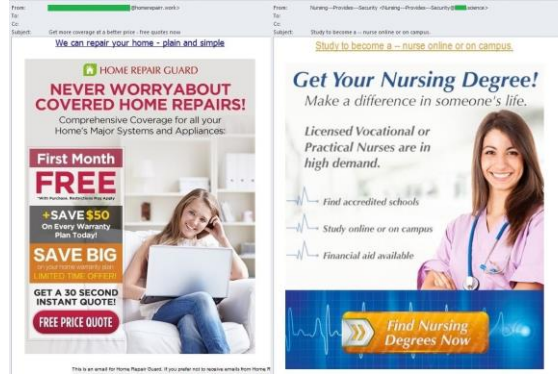


Figure. 2.4 Standard Images

III. SPAM FILTER

A spam filter is a kind of program that is used to detect unwelcome and unwanted email and stop those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for certain criteria on which it bases judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not particularly effective, too often omitting perfectly legal messages basically called as false positives and hire actual spam through. More stylish programs, such as Bayesian filters or other heuristic filters, attempt to identify spam through doubtful word patterns or word frequency.

A. Structure of Spam Filter:

Spam filtering is designed to distinguish between real mail and spam mail. Spam filters are a specialized technical fix against spam which helps end-users to keep their mailboxes clean. Spam filters can be operated on Internet Service Providers (ISPs), email servers, or users' email clients [6]. Several modules which analyze different characteristics of input emails, like address of the sender and the recipient, textual content, header format and mail attachments are involve in spam filter. The filter may implement on the user's system or on a server serving the same purpose for multiple users at a single time. The basic structure for Spam filter is as shown in figure 2.

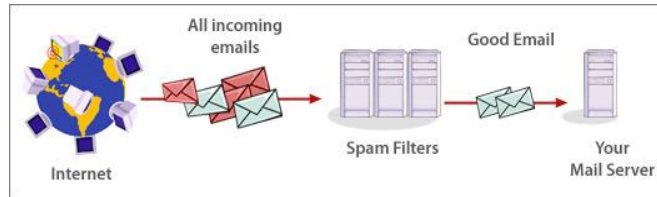


Figure 3. Basic structure of Spam filter

As shown in figure, all incoming mails will be filtered by Spam filter one at a time. Spam filter then distinguish the received mails is Spam or Real message. Once the identification is done, the real message will send to receiver's inbox and spam message will be separate out by filter. If any misclassification is found i.e. either spam in the inbox or real message block by filter, the errors may be reported to the filter to improve its performance [6].

In general, a spam filter can be defined as an application which implements a function [2]:

$$\left\{ \begin{array}{ll} f(m, \theta) = C_{spam}, & \text{if decision is "spam"} \\ C_{real}, & \text{otherwise} \end{array} \right. \quad (1)$$

Where,

m = message to be classified,

θ = a vector of parameters

C_{spam} and C_{real} = labels to be assigned to the messages.

Usually, spam filters are designed on the basis of machine learning classification techniques. In such a technique the vector of parameters θ is the result of training the classifier on a precollected dataset [8].

IV. IMAGE SPAM FILTERING TECHNIQUES

Day-by-day the methods of spam filtering are improving gradually. Spammers always try to decrease filtering efficiency by developing new spamming techniques. The reverse response i.e. reactivity of spammers asks for corrective measures from filter developers, which in the field of spam filtering may be termed as opposing reactivity [8]. Out of all such a spam filtering techniques, few are discussed as follows.

A. *Distributed adaptive blacklist :*

By definition, Spam is always delivering to large number of recipients. There is little if any customization of spam messages to individual recipients is to be done. Each recipient of a spam, however, in the absence of prior filtering, must press his own "Delete" button to get rid of the message. Distributed blacklist filters let one user's Delete button warns millions of other users as to the spamminess of the message [6].

Tools such as Razor and Pyzor operate around servers that store digests of known spams. When a message is received by an MTA, a distributed blacklist filter is called to determine whether the message is a known spam. These tools use clever statistical techniques for creating digests, so that spams with minor or just different headers resulting from transport routes do not prevent recognition of message identity. In addition, maintainers of distributed blacklist servers frequently create "honey-pot" addresses specifically for the purpose of attracting spam (but never for any legitimate correspondences).

B. *Rule Based Filtering :*

Content filtering techniques relied on the specification of lists of words or regular expressions disallowed in mail messages. Thus, if a site receives spam advertising "herbal Viagra", the administrator might place these words in the filter configuration [8]. The mail server would hence reject any message containing the phrase.

There are three types of possible disadvantages of filtering : First, it can be time-consuming for maintenance perspective. Second, it is prone to false positives. Third, these false positives are not equally distributed since content filtering is prone to reject legal messages on topics related to products frequently advertised in spam [8]. A system administrator who attempts to reject spam messages which advertise mortgage refinancing, credit or debit may inadvertently block legitimate e-mail on the same subject.

Spammers used to change the phrases and spellings frequently which they used. This may increase more work for the administrator. However, it also has some advantages for the spam fighter. If the spammer starts spelling "Tiara" as "Tlara" or "Tia_ra", it makes it harder for the spammer's intended audience to read their messages.

Content filtering can also be implemented to examine the URLs present (i.e. spam vertising) in an email message. This form of content filtering is much harder to disguise as the URLs must resolve to a valid domain name. Extracting a list of such links and comparing them to published sources of spam vertised domains is a simple and reliable way to eliminate a large percentage of spam via content analysis.

C. *Support Vector machine:*

SVM (Support Vector Machine) is an algorithm to classify the data set expressed by the vector into two classes [9]. In the spam filter by SVM, mail is classified into spam mail and regular mail by using SVM. SVM uses the data set expressed by the vector as an input. To classify mail by SVM, the mail data necessary expressed a vector. For separating the tokens, the vector conversion of the text is used same as Bayesian spam filter, and then appearance frequency with the token code which is corresponding to appeared token is calculated. To define the token code, all tokens that appear to mail are extracted beforehand. As for the appearance frequency, the definitions by the occurrence count or TF-IDF are thought.

The SVM filter procedure for classifying whether a mail is spam mail or regular mail is as follows. The procedure separates into pre-processing (filter learning) and the classify processing (filtering) [9].

- *pre-processing(filter learning):*
 - 1) Collect spam mails and regular mails.
 - 2) Separate all mail to tokens.
 - 3) Calculate appearance frequency of each token.
 - 4) Define token code.

- 5) Make vector sets with the token code and its appearance frequency
- 6) Construct filter (classification rules) by learning filter
By SVM with vector sets and label (the mail is spam Mail of regular mail).

- *Classify Processing(Filtering):*
 - 1) Separate a mail to tokens.
 - 2) Make vector with token code and its appearance frequency.
 - 3) Using SVM filter and this vector, classify the mail.

C. Bayesian Spam Filter:

Bayesian spam filter is a spam filter which is based on Bayesian theory. [9] In Bayesian theory, the probability of a certain cause when a certain event occurs can be calculated by the probability of all cause of event and the conditional probability that the event occurs by a certain cause. The filter separates by the probability whether the spam mail or not from the appearance probability of the character string (token) used with mail based on Bayesian theory. The word (or, the stem) and the character string that n character are consecutive are used as a token.

D. K-Nearest Neighbors :

The most popular one in this category is k-Nearest Neighbors (k-NN). The value of k designates the number of neighbors used for classification. A significant step of this method is the choice of similarity function between messages. The method frequently used to compute the similarity measure between messages is the “cosine distance”, where cosine is defined as the angle between the vectors representing the compared messages [9]. This distance function normalizes the length of the messages, and hence considered effective.

E. Technique of search engines:

When it acts on text e-mails, classification techniques of text seem to be efficient. However, spammers do not stop to invent tricks to avoid filters. One of these tricks is to include the hyperlink of a Web page which contains the advertising text/image in the body of the message only. The problem become then a web content classification. Public search engines have been anticipated to overcome this kind of spams which offers a mean to classify the websites. The principle of this technique is to analyze automatically the contents of the pages referred by the links sent in the messages likely to be spams.

Bing outlines some ways of discovering and then filtering such spam within the algorithm. It includes:

- *Content Quality*

This method is useful just to Access the quality of content. At a high level, spammer’s overall goal is to drive ad and affiliate clicks, the content of the page is important only to the extent that it helps to facilitate said goal. To put it another way, spammers generate content targeted at search engines and their algorithms, whereas legitimate SEOs generate content for their customers. The result is that, in most cases, spam pages have inadequate content with limited value to the user. We use this fact to facilitate detection. There are literally hundreds, if not thousands, of signals used to make this assessment, ranging from simple things like number of words on the page to more complex concepts of content uniqueness and utility.

- *Ad Location & Quantity*

With this we can take a look at the presence of ads on a page. About every page on the web contain ads. Presence of ads doesn’t make the page bad, let alone spam. What to be needed is, care about things like a) how many ads appear on the page, b) what type of ads (e.g. banner, grey-over’s, pop-ups), and c) how intrusive/disruptive they are.

- *Page Layout*

It is also important to look at the position & layout of the information on the page. E.g. where is the main content located? Where are the ads located? Do the ads take up the prime real estate or are they neatly separated away from the main content (e.g. in the header/ footer or side pane)? Is it easy for users to mentally separate content from ads?

- *Spammers Use Content Generation Techniques:*

Many times spammers use content generation techniques to quickly “maximize web presence” through mass content production via (a) copying other’s content (either entirely or with minor tweaks), b) using programs to

automatically generate page content, c) using external APIs to populate their pages with non-unique content. This can be avoided by using “creative clustering algorithms” to detect these attempts.

- *Spammers Use Other Techniques To Boost Rankings:*

Spammers may use other methods such as a) stuffing page body/ url/ anchors with keywords, b) performing link manipulation via link farms, link networks, forum post abuse and c) including hidden content on the page not meant for human consumption. To deal with all these, searching algorithms can be used to look for content outliers across the web and if things look unnatural, it can be detected. For link manipulation, web graph (page/ site in links and out links) can be used to identify possible link manipulation.

- *Action Taken On Spam/Spammers*

There are different levels of action to be taken on spam, like (a) demoting the page, (b) neutralizing the effect of specific spam techniques or (c) removing the page/ site out of the index all-together. The level of action depends on a) the extent/ egregiousness of the spam techniques involved and b) the potential value the page presents to the users

V.CONCLUSION

We have referred various papers to take the review regarding image spam and seen different types of image spam and their content. Along with that, we have discussed basic idea about image spam and how spam filter works. As spammers have countless techniques for creating a spam image, the research for a perfect spam filter is always fruitful. Several works have been proposed and almost all of these methods have the common objectives of high processing speed and high accuracy, to make it applicable in time critical environment like the Internet. Future work includes analysis and comparison of these some techniques reviewed in terms of computation and time complexity along with accuracy.

REFERENCES

- [1] Christina, V, S Karpagavalli, and G Suganya. "A Study On Email Spam Filtering Techniques". *International Journal of Computer Applications* 12.1 (2010): 7-9. Web.
- [2] B Mehta, S Nangia and M Gupta. "Detecting Image Spam using Visual Features and Near Duplicate Detection" *International World Wide Web conference committee* ACM 978-1-60558-085-2/08/04, April 21–25, 2008, Beijing, China.
- [3] R Sharma, G Kaur. "Spam Detection Techniques: A Review" *International Journal of Science and Research (IJSR)*, ISSN:2319-7064, Vol.4 Issue 5, May 2015.
- [4] A *whitepaper* on "Mail-Secure Image based spam treatment".
- [5] M Kamble, C Dule. "Image Spam Detection: A Review" *International Conference on Advances in Computer science and electronics engineering*, ISBN: 978-981-07-1403-1, doi:10.3850/978-981-07-1403-1 624
- [6] M Das, V Prasad "Analysis of Image Spam in Email based on content analysis" *International Journal on Natural Language Computing (IJNLC)*, Vol.3, No.3, June 2014.
- [7] Gordon V. Cormack, "Email Spam Filtering: A Systematic Review", *Foundations and Trends in Information Retrieval* Vol. 1, No. 4 335–455, 2008.
- [8] Blanzieri Enrico, Bryl Anton, "A Survey Of Learning-Based Techniques Of Email Spam Filtering", *Artificial Intelligence Review*, Volume 29, Issue 1, Springer, pp 63-92, 2008.
- [9] Taira, H., Haruno, M.: Feature Selection in SVM Text Categorization, *Journal of Information Processing Society of Japan*, Vol.41, No.4, pp.1113-1123 (2000). (In Japanese)