Image Spam Detection : A Review

Ms. M. Kamble
Computer Science & Engineering
GHRCE
Nagpur, India
minalkamble327@gmail.com

Ms. Chhaya Dule
Asst. Prof. Computer Science & Engineering
GHRCE,
Nagpur, India
chhaya06@rediffmail.com

***Abstract*— Today, the internet is the most powerful tools throughout the world. However the explosive growth of unsolicited emails has prompted the development of numerous spam filtering techniques It needlessly obstruct the entire system. Spammers are creating new ways against anti-spam technology. By the end of 2006, the nature of spam had totally shifted. The newest of which is the image-based spam. They are difficult to detect using traditional techniques. How to deal with the image spam is a difficult problem for the entire world. Image spam has become the main form of spam, it is a problem crying out for solutions to effectively filter such spam nowadays. This paper considers various systems that has been proposed for the detection of image spam by considering text, content and image features.**

Keywords—Spam, Image Spam, OCR

I. INTRODUCTION

As the use of email for the communication is increasing, the number of unwanted 'spam' is also increasing[1]. For example, there's the occasional joke sent in mass from friend to friends and back again, or that all-important virus alert, or the occasional inspiration, etc[2].

Fig 1 : Natural images

Fig 2 : Spam images

Most readers spend large amount of time detecting such messages[3].There is also cost related to server which manages the large amount of emails related to the system. When large number of messages are sent in bulk by the spammers, it adversely affects the performance of the system. Users have to pay long distance connection charges which increases with the unwanted messages. Most of the spammers send the emails in fraudulent way, by using the software which hides the identity[2]. There are various ways which spammers uses to get the email addresses. They buy it from various companies and acquire email addresses and sometimes they also hack the account.

Image Spam is an e-mail solicitation that uses graphical images of text to avoid filters[4]. Recently, though, it reached an unprecedented level of sophistication and took off. A year ago, fewer than five out of 100 e-mails were image spam. Today, up to 40 percent are. Meanwhile, image spam is the reason spam traffic overall doubled in 2006, according to antispam company . It is expected to keep rising.

There are various aims of using images in email, from simply making the email more attractive, or adding a look of professionalism, to attempting to evade text based spam filters and signatures. The use of remote images in particular has been steadily increasing over the last 16 months.

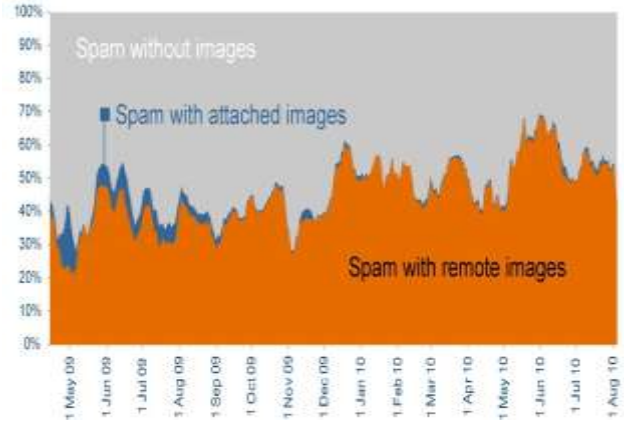

Fig3: Spam Image Survey

The basic idea behind image spam is that it is difficult to detect using spam filtering software designed to detect patterns

in text in the plain-text E-mail body[5]. Attempts to filter text in image spam are easily defeated because optical character recognition of text in image spam can be prevented using a variety of obfuscation techniques which will not prevent the spam image from being read by human beings[3]. Obfuscation techniques can include , blurring of text outlines , construction of the image from multiple image layers assembled within an HTML e-mail use of animated image formats, random noise added to the image (also known as confetti) to prevent the detection of multiple similar images using hash algorithms.

Currently, the surest known countermeasure for image spam is to discard all messages containing images which do not appear to come from an already white listed E-mail address. However, this has the disadvantage that valid messages containing images from new correspondents must either be silently discarded, or that bogus "backscatter" bounce messages must necessarily be generated to the reply-to addresses in junk mail messages, enabling denial-of-service attacks by spammers, as well as a directory harvesting attack. Another common technique for image spam detection is to analyze what percentage of the email is actually an image, as image spam often contains very little text content.

II. IMAGE SPAM

Following are the techniques which spammers use in order to prevent the detection of the spam image.

A. GIF Layering :

Like the word splitting divides words into multiple images to prevent spam filters, an image spam can be divided into pieces. Pieces of a message are layered to create a complete, legible message [6].

Fig 4. GIF Layering

B. Optical Character Recognition Duping :

OCR works by measuring the geometry and structure in images, searching for shapes that match the shapes of letters, then converting a matched geometric shape into real text. To defeat OCR, spammers upset the geometry of letters by changing the colors, so that OCR can't recognize a letter even as the human eye easily recognizes it. The effect is something like blurred characters in an eye test.[7]

C. Word Splitting and Ransom Notes :

If OCR catches up to the color tricks in image spam, a spammer's next defense is word splitting. By dividing the image and leaving space in between the pieces, any image the OCR engine is examining is only a piece of a letter with its own distinct geometry.

Instead of word splitting, some spammers have employed a ransom note technique in which each letter in the spam message is its own image, and each letter image includes background noise and other baffling techniques. A program cobbles together randomized letter images to make words. The effect looks like a classic ransom note with a mishmash of letters cut out from magazines.[13]

D. Geometric Variance :

Many filters can intercept mass mailings based on their sameness. Images, though, can be altered easily without disturbing the message inside them. Thus one spam message will arrive as dozens of differently shaped images, and each time the colors of the text images will have changed, as will the randomly generated speckling and pixel and word salads. No two images are alike despite the fact that they carry similar messages.

II. IMAGE SPAM DETECTION

Various anti spam technologies are proposed in filtering text based spam emails which usually compare the contents of emails against specific keywords. The latest image spam is not possible to detect by the most anti spam software. Consequently, variety of methodology has been implemented in current anti spam system to filter the image spam.

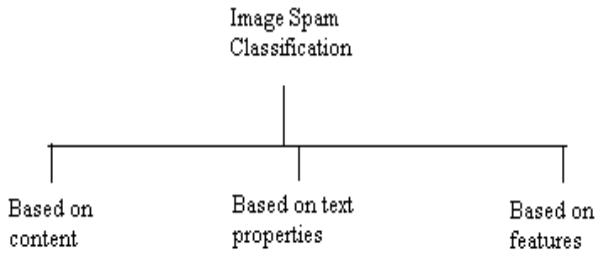

Fig 5: Classification of image spam

A. Classification Based on contents

In [8] author used probabilistic tress to detect the spam images in email. It has made use of global image features to train classifier to distinguish the spam images.

[9] used a fast and low-cost feature extraction and classification framework to find the large amounts of image spam using decision tress and support vector machines. The authors also compare SVM classifier with the decision tree method. It has been proved that SVM classifier is better.

B. Classification Based on text properties

[10] used low-level image processing techniques in order to detect one of the characteristics of most image spam. It proposed method to detect the presence of noisy text into an image. The output was in the form of crisp value in the range or some real value. The presence of noisy text was considered as key to distinguish the spam.

C. Classification Based on image features

Many researchers used images features to filter the image-spam mails. In [11] features of images are divided into two major classes viz, the high-level features and the low-level features.

i. The high level features refers to the file name, file size, file format, and so on.[11] used these feature to identify image spam.
ii. The low level features includes color, shape, etc.

In image processing and photography, a color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image's color space, the set of all possible colors.The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. For monochromatic images, the term intensity histogram may be used instead. For multi-spectral images, where each pixel is represented by an arbitrary number of measurements (for example, beyond the three measurements in RGB).

Advantages of color histogram

i. It is a trouble-free feature
ii. It can be calculated by a one simple pass of whole image
iii. The range of the colors in an image spam is limited

[3] used this color histogram to distinguish image spam.

IV . CONCLUSION

In this paper we have discussed about the nature of image spam. Spammers use different techniques to make the traditional software unable to detect this new kind of spam. Various authors have proposed methods for detecting the image spam. Some have used content based method , some have used text based methods. But feature based methods proved to be more accurate and efficient among them.

V . FUTURE SCOPE

The aim is to investigate the features of image spam and select an optimum feature to classify the same using maximum likelihood classifier. The idea is to develop a method to filter spam based on image content, rather than text content Performance of the system will be then measured in terms of the accuracy and precision.

References

[1] Hrishikesh B. Ardhye, Gregory K. Myers, James A. Herson , (2005), "Image Analysis for Efficient Categorization of Image-based Spam E-mail".
[2] How do I stop spam?
[3] M. Soranamageswari, Dr. C. Meena, (2010), "An Efficient Feature Extraction Method for Classification of Image Spam Using Artificial Neural Networks" ,pp. 169-172
[4] Nobuo Kumagai Masayoshi Aritsugi, (2005) , "On Applying an Image Processing Technique to Detecting Spams" .
[5] Congfu Xu, Yafang Chen, Kevin Chiew, (2010) "An Approach to Image Spam Filtering Based on Base64 Encoding and N-Gram Feature Extraction", pp. 171-177
[6] www.csoonline.com/article/221254/image-spam-by-the-numbers
[7] Giorgio Fumera, Ignazio Pillai, Fabio Roli. , (2006), "Spam Filtering Based On The Analysis Of Text Information Embedded Into Images", Journal of Machine Learning Research, pp .2699-2720
[8] Yan Gao, Ming Yang, Xiaonan Zhao, (2008), " Image Spam Hunter"
[9] Sven Krasser et al,(2007), " Identifying Image Spam based on Header and File Properties using C4.5 Decision Trees and Support Vector Machine Learning "
[10] Battista Biggio, Giorgio Fumera, Ignazio Pillai, Fabio Roli, (2007) "Image Spam Filtering by Content Obscuring Detection," Fourth Conference on Email and Anti-Spam.
[11] Masahiro Uemura, Toshihiro Tabata (2008), "Design and Evaluation of a Bayesian-filter-based Image Spam Filtering Method".