Innovative Science and Technology Publications

International Journal of Future Innovative Science and Technology
ISSN: 2454- 194X Volume - 2, Issue - 2

Manuscript Title

Efficient System for Detecting Image Spam and Identifying New Child Abuse Media in a Network

C. Sreebanupriya
PG Student, Information Technology,
Sri Venkateswara College of Engineering
Sriperambudhur, Tamil Nadu, India.
banuu23@gmail.com

A.Indumathi
Assistant Professor, Information Technology,
Sri Venkateswara College of Engineering
Sriperambudhur, Tamil Nadu, India.
aindumathi@svce.ac.in

May – 2016

www.istpublications.com

International Journal of Future Innovative Science and Technology, ISSN: 2454- 194X
Volume-2, Issue-2, May - 2016 editor@istpublications.com

Efficient System For Detecting Image Spam And Identifying New Child Abuse Media In A Network

C. Sreebanupriya
PG Student, Information Technology,
Sri Venkateswara College of Engineering
Sriperambudhur, Tamil Nadu, India.
banuu23@gmail.com

A.Indumathi
Assistant Professor, Information Technology,
Sri Venkateswara College of Engineering
Sriperambudhur, Tamil Nadu, India.
aindumathi@svce.ac.in

ABSTRACT

The increasing levels of child sex abuse and spam words are being shared in Websites pose a significant challenge to permanently block their activities. Although a number of steps taken to stop such offender activity, as they cannot detect the on-going child abuse and image spam. In this, Email spam is a subset of electronic spam involving nearly identical messages and images sent to numerous recipients by email. In real time browsers such as yahoo, Rediff mail or Gmail haven't aware of injected spam, abuse images and abuse videos. In the survey of 2015, 90% of the email users has affected with image spam, including features like child pornography and child abuse. Here introduced OCR scanner in to website server which scan the image and separate the spam data in the images, while comparing with global spam database. Also the proposed nudity algorithm for the abuse images that shares virally into the website today, can able to detect the skin color of the image that typically rely on hash value being stored and those detected images are tagged with an IP address, username, Mac address to block the user permanently from being accessed from the server.

Keywords-Email service Provider; Computer Crime; Image Classification; Text Analysis; Tagging; Child protection; paedophilia.

I. INTRODUCTION

When an image spam is distributed over a network, it isa larger drain in the network resource than the literal spam because an image file is larger than the text file and it requires a higher bandwidth. As a consequent, it causes a grater degradation of transfer rates. The image spam has the different formats, such as .gif, .bmp, .jpg, .png, .gif etc., most images are in the format of .gif and .jpg. In fact, the process of classifying the image spam groups [2] because an image contains many properties, for example brightness, contrast, and radian. Presently, there are tremendous methods in the image spam detection process[2]. Unfortunately, these methods can detect only detect only images of texts, or humans or bodies appearances.

The proliferation of the Internet file sharing systems has transformed the distribution of child sexual abuse media into a crime without geographical boundaries. While there is scientific debate on whether the online pedophile is a new type of offender [15] or if those with a pre-disposition to offend are responding to the opportunities afforded by the new forms of social media [18], empirical evidence points to the problem of Internet-based paedophilia as endemic. Number of searches and transfer of abuse images are increasing day-to-day.

The severity of the problem has resulted in a number of solutions that can monitor such activity. Tools such as the Child Protection System (CPS) are able to capture data about paedophile activity on networks. However, these systems rely on matching the files shared on a network against a hash based value. As a result, they are not able to identify new child abuse media that may be released on to the network. Furthermore, Originators of the such media on hands-can be abusers and their early detection and apprehension can safeguard their victims from further abuse.

Therefore, here proposes a new method that enhances the image spam detection. So, the image spam, not only the images of text or human pictures, but also other image spam such as

images of advertisements, can be detected. Further it can also detect the offender activity of sharing abuse images in the website by introducing Nudity algorithm.

II. PROBLEM DEFINITION

The primary focus of this paper is on identifying and avoiding an image spam e-mail across the web. Security has always been an important aspect of quality of the service provided. Our aim is to develop a novel framework using OCR which helps to ensure that we won't have spam in our e-mail inbox. By identifying the spam it relives the user from worrying about the spam email. There are a number of providers who provide the e-mail filtering software to whom the clients will use like Google, Yahoo, and Rediffmail etc. When these software provides inbuilt filtering software, there is a major threat of mix image spam for a number of reasons like e-mail message from unknown source, link leads to an unknown web pages. The messages that are sent by any unknown source must be prevented. Before going to the rest of this paper, all necessary definitions must be stated.

1) Image spam e-mail

An image is classified as an image spam if and only if the following conditions are true more than one.

1. The image is delivered by an unknown source.

2. The image is attached with the main body of e-mail.

3. The image is a hyperlink to an unknown web Page.

2) Format of an Image Spam

The format of an image spam is the same as standard format of image over the Internet like Subject, Message Body, and Receiver Address.

3) Characteristics of an image spam e-mail

An image spam can be classified in two categories: pure image, or mixed image. The pure image spam is the spam contains only image(s); the mix image spam consists of images and a text message attached to an email.

In the year 2008, jordan [6] proposed a detection technique for an image spam using near-duplicate detection technique. This technique produces a non-spam image repository. Then, when users receive an image, it will be compared with images in the image database for the spam filtering process. The received image will be eliminated when the feature vector of it is different from the feature of the images in the image repository. Similar to this research, Battista et al. [2] proposed a method to classify an image spam by comparing the received image with the original image in the database. The received image will be terminated when there is a difference between these two images. These two techniques have limitation in detecting image spam because some spam consists of both texts and images. Thus, the proposed techniques cannot be applied. Therefore, [12] proposed a method to detect the image spam that consists of both texts and images. This technique uses one-class Support Vector Machines (SVM) to classify a spam from the received e-mail. Moreover, the researchers separated three sets of features to detect the image spam; these features are embedded text features, Banner and graphic features and image location features. The filtering process [1] tries to recognize an image spam using the OCR tool. This technique is similar to [12] because it detects spam images , except that these spam flow via e-mail and the detection uses a non-uniform background, pixels of different colors for each

character, and distortion of text lines or single characters.

Francesco et al. [4] proposed two different image processing techniques which are used to detect the image spam that composed of both text and image. The components-based method on [5] SIFT method to detect the image spam. This method detects image spam that was modified from the converted of content text to an image and the embedded spam message through an email. Some image spam was identified using a boosting tree which is learning based prototype system announced by [13]. The detection system is called as the Image Spam Hunter. The other method that applied a tree structure to classify the spam is proposed by Sven et al. [11] uses a decision tree and a support vector machine as the classifiers. The advantage is that it can detect a large amount of image spam. A matching mechanism proposed by [7] uses using user specified image content focuses on text message hidden in spam images using SIFT algorithm that led many researches to use these properties as attributes of an object to detect as the image spam. A new method of Jordan [6] will converts JPEG images to ASCII using JP2A. Peizhou et al. [10] identifies the image spam by using properties of an image as attributes. He applied file properties and a Histograms algorithm for image spam detection. This method is called as the FH algorithm. This algorithm is the first part of 2-step image spam classification while the second part is the comparison of the histogram, both gray and color histograms, models are used for image testing. Although the image spam is rapidly grow over the Internet, another unwanted image message called ham also causes problem to the Internet users.

From the researches mentioned above, all the detection mechanisms require 100% accuracy of image mapping between the incoming message and the standard image in the database or corpus. This constrain limits the efficiency of detecting spam because the spam can modify itself from the original pattern to varieties patterns. Thus, the method proposed will unlock this limitation.

III. APPROACH

A method to detect a spam from the body of an e-mail, called using OCR and Nudity algorithm. Thus, the certified e-mail can be distinguished from the spam e-mail. Whenever an e-mail arrives at the e-mail server, it will be sent for the OCR scanner to separate the content according to its characteristics which can be either text or images. On conversion it is notified whether to be verified with global spam database if it contains text on further it will be sent to the nudity detection. we applied a more comprehensive approach during the selection process by combining paedophile keyword information. More specifically, we first created a dictionary-based filter containing a manually extended version of the paedophile keyword with global database. Secondly From images, we extract: colour-correlograms, skin features, having a resultant percentage about the nudity in the image. If the percentage exceeds the given average from our algorithm, it is considered as nudity and further the image cannot be uploaded in to the website. The result from this module will be ended up at the evaluation module where the e-mail will be determined as a certified e-mail or an image spam e-mail. Figure 1 shows the system architecture of detecting spam text and image spam in to a website.

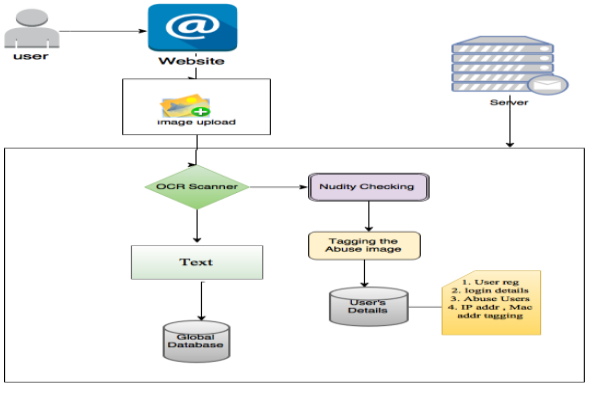

Fig. 1 System Architecture of detecting spam text and image spam in to a website.

A. OCR

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic translation of scanned images of handwritten, typewritten or printed text into machine-encoded text. It is widely used to convert books and documents into electronic files, to computerize a record-keeping system in an office, or to publish the text on a website. OCR makes it possible to edit the text, search for a word or phrase, store it more compactly, display or print a copy free of scanning artifacts, and apply techniques such as machine translation, text-to-speech and text mining to it. . Figure 2 illustrates the example of an image spam e-mail.. This way, the latest OCR technology is quite useful when we need coding our legal papers. When people scan any document it's stored in form of an image, which can't be edited. Whilst OCR has made it possible to scan any printed document and relocate it into word-processing software, such as MS word where we can easily edit it as per our need. OCR is an advanced technology that allows people to alter different sorts of documents including scanned paper documents, PDF files or pictures captured by a digital camera into editable and searchable data as well. Using OCR the system recognizes the static shape of the character. It is necessary to understand that OCR technology is a basic technology also used in advanced scanning applications. Figure 3 shows how when Figure 2 is split into text and image using OCR.

imagename	images
make money.TIF	0x494924009CC0020080341E2FF673A1FEDC73BA5F4FF733...

Fig. 2 Shows the separate text and image when OCR applied.

B. E-mail process

This module includes sending and receiving a mail system. Mail composing page have e-mail address of recipient, subject and the content. All the mail received in the corresponding mail inbox.

C. Image Separation Module

In this module, whenever an e-mail arrives at the e-mail server, the OCR tool converts or extracts the received e-mail according to the content based on its characteristics. Since there are 2 types of outcomes from the previous module, texts and images, then details of each type must be separately defined. When considering the text message obtained from the original image, Examples for keywords are such as prices, win, trust, and join now.

1.Keywords in an image

All keywords of the advertised spam are recorded. A popular resource to collect these keywords is the Internet, such as trash box of web mails, the spam box of web mails, shopping web site, including the Internet site of spam list. Figure 4 shows sample of an image spam.

Fig. 4 Sample of an image spam

2. Pure Image spam

This repository stores only images that are counted as spam. Thus, this corpus consists of

small images, and its properties such as Colour-Correlograms, Skin-Feature, and Nudity value.

IV.ALGORITHM DEVELOPMENT AND IMPLEMENTATION

A. NUDITY ALGORITHM

Nudity algorithm is used for detecting the spam images with abuse content while sharing into the website. It depends on few factors like Colour-correlograms, HSV color space, Skin features. Depending upon the value in the hash , it will be compared with the optimum range of percentage, then then image decided to be sent over the network or not.

Colour-Correlograms describe the occurrence probability of a color in a pixel's neighborhood (see e.g. [9], [10], [12]). Hence, they represent the local spatial correlation of colors in images. Here, we apply a special variant of the colour-correlogram ,namely the auto colour-correlogram, which describes the probability of the identical color *c* reoccurring within a distance *d* of the current pixel in image *I*.

(1)

$$\alpha d\ c\ (I) = \gamma dc,c(I)$$

For an optimal performance, this feature is computed in HSV color space ([14]).The Skin-Feature is based on a RGB skin color model, which was generated using manually segmented images from the COMPAQ database. The presence of skin is indicated via a *skin-probability-map* (SPM) (see also [17]).

$$P(skin|c) = Pskin(c)Pskin(c) + Pnon{-}skin(c) \quad (2)$$

For computing the skin feature, first the SPM is transferred into a Skin-Segmentation-Mask(SSM) via morphological operations and adaptive thresholding. Next, the mean intensities of SPM and SSM are calculated, as well as their center and variance of skin mass. This yields 14 dimensional descriptor representing image skin properties.

V.ADVANTAGES

We conducted a live testing workshop on websites; our web application toolkit is useful in detecting the abuse images and nudity value for all the images. It can also be able to detect the image spam efficiently, when the E-mail browsers receives a mail. It will efficiently scan type of image containing the word to verify whether it is spam or not. This will reduce overall occurrence of the offender attack taking place in real time websites.

VI. EVALUATION

The evaluation in the detection architecture where all texts and images obtained from the e-mail will be identified as a certified message or an image spam message. In doing so, the technique defined previously will be used, all converted texts and images will compare with data stored in this database, also checks the nudity value for the abuse images.

1) Keywords in an image

Since there are large amount of spam in the website sharing, the searching speed is a significant issue that must be concerned; otherwise, the mailing system will have a huge problem in managing the arrival queue. OCR scanner will be used the Computer graphic or computer vision to the image retrieval problem that is the problem of searching for images, similar means that the search will be analyzed the true contents of the image. On comparing the words retrieved from is verified with global spam database and evaluated.

2) Pure image spam

Nudity algorithm on which we used helps to recognize the content of the image with Colour-correlograms, skin feature by making Skin-Segmentation-Mask(SSM) via morphological operations and adaptive thresholding. Next, the mean intensities of SPM and SSM are calculated, as well as their center and variance of skin mass. Parameters were experimentally determined on a development set of each training partition during cross validation. The scores we report are average precision, recall and F-score. These are standard evaluation metrics that can be computed based on the number of true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn) in a confusion matrix. The recall score for each class provides information on the number of filenames that were successfully retrieved, while the precision score takes into account all retrieved filenames for each class and evaluates how many of them were actually relevant. The F-score is then the harmonic mean of precision as given in equation (3) - (5).

Precision = tp/(tp+ fp) (3)
Recall =tp/(tp+ fn) (4)
Fscore= 2.Precision.Recall/(Precision + Recall) (5)

The results of the experiments are shown in Table I and table II.

TABLE I: PERFORMANCE PARAMETER

Classifier	Natural Images	Spam Images
Natural images	Number of images correctly classified as natural images	Number of spam images is classified as natural image
Spam Images	Number of natural images is classified as spam image	Number of images correctly classified as spam images

TABLE II: CLASSIFICATION RESULTS

Approach	F-Score		Precision (P)		Recall (R)	
	Ham	Spam	Ham	Spam	Ham	Spam
SVM machines	90.5%	86.6%	84.5%	80.6%	88.3%	85.7%
FH Histogram	94.6%	92.1%	88.7%	84.1%	90.5%	89.6%
OCR and Nudity	96.5%	95.4%	90.5%	88.7%	92.0%	91.4%

VII. DISCUSSION

Today spam is unavoidable on the Internet and spams are evolving from text to image over the years each having different attributes. Even though many researchers have given different approaches to detect the uncertified e-mail spam, the problem is the time taken to detect the image spam and also requires 100% accuracy while detecting our accuracy of the spam images including child abuse media containing 20,000 images which have been collected from various web sources like flickr.com, youtube.com, pichunter.com, redtube.com having numerical feature representations are parameterized under the algorithm.

VIII. CONCLUSION

Spam is crucial problem across the Internet because it is evolved from text to image. Some of the E-mail spam filtering software could not identify the partial image spam. So this paper proposed a new technique to identify and avoid the received image spam across the web also by detecting the abuse images shared over website. This paper deals with detecting the offender activity in the website. As a result the uncertified image spam and sharing abuse images is sent to the spam folder. Finally it relieves the user from worrying about the received image spam.

VII. FUTURE WORK

The users sharing the abuse image against child Protection will be permanently blocked using IP address, user behavior, MAC address. In future, visual words and videos can be detected considering the number of frames, acoustic features, Visual Words and Pyramids to provide a texture based content representation. Therefore a number of websites can be blocked against such illegal offender activities that virally go in the internet.

VIII. REFERENCES

[1] B.Battista, F.Giorgio, P.Ignazio, and R.Fabio, "Image Spam Filtering Using Visual Information", in Proc.14th International Conference on Image Analysis and Processing (ICIAP 2007), Department of Electrical and Electronic Engineering, University. Of Cagliari, Italy, 2007.

[2] B.Battista, F.Giorgio, P.Ignazio, and R.Fabio,"Image Spam Filtering by Content Obscuring Detection " in Proc. 4th international Conference on E-mail and Anti- Spam(CEAS 2007), California, USA, April 2008.

[3] B.Battista, F.Giorgio, P.Ignazio, and R.Fabio,"Image Spam Filtering by Content Obscuring Detection " in Proc. 4th international Conference on E-mail and Anti- Spam(CEAS 2007), California, USA, April 2008.

[4] G.Francesco, P.Antonio, PI.Antonio, and S.Carlo, "Using heterogeneous features for anti-spam filters", in Proc. 19th International Conference on Database and Expert Systems Application, pp.670-674, September 2008.

[5] H.Hailing, G.Weiqiang, and Z.Yu, "A Novel Method for Image Spam Filtering", in Proc. 9th International Conference for Young Computer Scientists, Zhang JiaJie, Hunan China, pp.826-830, November 2008.

[6] N.Jordan, M.Daniel, C.D.Nunes, and A.John, "Image Spam-ASCII to the Rescue!" in Proc. 3rd International Conference on Malicious and Unwanted Software (MALWARE), pp.65-68, 2008.

[7] C.Junwei, Z.Lichun, and L.Yueu, "Application of Scale Invariant Feature Transform to Image Spam Filter", in Proc. 2nd International Conference on Future Generation Communication and Networking Symposia,IEEE Computer Society, vol. 3, pp.55-58, 2008.

[8] D.Mark, G.Reuven, and B.E.Ari, "Learning Fast Classifiers for Image Spam", in Proc.4th International Conference on Email and Anti-Spam(CEAS 2007), Microsoft Research Silicon Valley, Mountain View, California, USA, 2007.

[9] K.Sven, T.Yuchung, G.Jeremy, A.Dmitri, and J.Paul, " Identifying Image Spam based on Header and File Properties using C4.5 Decision Trees and Support Vector Machine Learning", in Proc. IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, June 2007.

[10] W.C. Wu, C.T. Kwang, Z.Qiang, and W.L.Yi, "Using Visual Features For Anti-Spam Filtering", in Proc. IEEE

International Conference on Image Processing (ICIP 2005), Genoa, Italy, Vol. 3, pp. III- 509-12, September 11-14, 2005.

[11] G.Yan, Y.Ming, Z.Xiaonan, P.Bryan, W.Ying , N.Thrasyvoulos, Pappas, and C.Alok, "Image Spam Hunter", in Proc. Acoustics, Speech and Signal(ICASSP2008), pp. 1765-1768, March 31 2008- April 4 2008.

[12] W. Zhe , J.William, L.Qin, C.Moses, and L.Kai, "Filtering Image Spam with Near-Duplicate Detection", in Proc. 4th International Conference on E-mail Anti-Spam(CEAS 2007), California, USA, April 2008.

[13] FIVES: Forensic Image and Video Examination Support. EC Safer Internet project. More information can be found at http://fives.kau.se

[14] I-Dash: The Investigator's Dashboard. EC Safer Internet project. More information can be found at http://www.i-dash.eu/.

[15] D. Middleton, "Internet Sex Offenders", Ch. 12 in Assessment and Treatment of Sex Offenders: A Handbook, 2009.

[16] T. Deselaers, L. Pimenidis, H. Ney, "Bag-of-Visual-Words Models for Adult Image Classification and Filtering", *ICPR*, pp. 1-4, 2010.

[17] D. Zorth, R. Ji, T. Chen, T. Breuel, S. Chang, "Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs", *ACMMultimedia*, 2013.

[18] H. Yu, C. Ho, Y. Juan, C. Lin, "LibShortText: A Library for Short-text Classification and Analysis", Technical Report,http://www.csie.ntu.edu.tw/cjlin/papers/libshorttext.pdf, 2014.