Spam Image Clustering: a Novel Approach in Revealing Common Sources of Spam

Chengcui Zhang, PhD, University of Alabama at Birmingham, Department of Computer and Information Sciences, CH 127 1530 3rd Ave S, Birmingham, AL 35294; Wei-Bang Chen, MS, University of Alabama at Birmingham, Department of Computer and Information Sciences, CH 128 1530 3rd Ave S, Birmingham, AL 35294; Richa Tiwari, MS. University of Alabama at Birmingham, Department of Computer and Information Sciences, CH 128 1530 3rd Ave S, Birmingham, AL 35294; Xin Chen, PhD, University of Alabama at Birmingham, Department of Computer and Information Sciences, CH 128 1530 3rd Ave S, Birmingham, AL 35294; and Gary Warner, BS, University of Alabama at Birmingham, Department of Computer and Information Sciences, CH 127 1530 3rd Ave S, Birmingham, AL 35294*

Abstract

Spam, unsolicited bulk messages, grows exponentially each year and according to a recent study, comprises almost 85% of all the email in the world. Interestingly, nearly 80% of spam is sent by approximately 200 spammers. Since spammers continually discover newer and better techniques to get around the Internet security filter, a major challenge in this field is to identify the methods spammers use and determine how to impede them. Even law enforcement is involved in tracing these criminals and trying to bring down the servers sending bulk spam emails.

The main objective of this project is to help forensic science researchers find those spammers by using data mining and image processing techniques. There are many existing research works that target the detection and filtering of spam emails from regular emails, but little work has been done in identifying the common sources of these emails. The spammers often attach images, usually advertising a particular product, in their spam emails to escape the detection of text-based spam filters. This study is based on the hypothesis that, by clustering these images on the basis of their textual content, foreground illustrations and overall visual effects, it can be concluded that the images in the same cluster originated from a common source and hence might come from the same spammer.

In this paper, 1190 spam images were used for analysis and clustering to identify common sources, i.e., the spammers. Each cluster contains images with similar background texture, foreground illustration and/or textual content, indicating that a common template has been used to produce them. Spam images can be typically classified into two categories- those that are mainly text based and those that contain mainly sub-images often referred as foreground illustrations. To extract the text from the image, the OCR technique is used. For identifying the illustrations in the foreground, the color-code histogram method is employed. In addition, since the layout of foreground illustrations is an important feature of a specific spam template, it is also included in the clustering process by measuring the layout similarities between spam images.

It is often the case that a spammer may create two spam images using the same template for selling different products, and thus, the two spam images will likely to agree in their text layouts but are different in terms of the actual embedded text. Hence the text layouts of all the images are also examined and compared in the clustering process. In addition, the background texture of each image is also analyzed by extracting the Tamura directionality feature.

The clustering of images that contain mainly text is performed separately from that of those that contain mainly foreground illustrations. The spam images with mainly text are grouped together according to their similar background texture and text layout. A cutoff value of 65 percent, for thresholding the similarity value between images, is applied. For those images that contain mainly illustrations, their foreground illustration layout and color code histogram features are used in clustering. These two features have their own advantages and complement each other in clustering similar spam images. A two-stage clustering method is used, in which the first stage uses features like shape, layout, and color histogram of the illustrations, while the second stage uses color code histogram to further refine the clusters. In the performance evaluation of the proposed algorithm, F-measure is used, which combines both precision and recall measures. The ground truth for this work was obtained manually by visual verification based on the visual appearance of those images and consists of 61 clusters. The result evaluation is performed at both stages in clustering as well as on the collective clustering result. The proposed algorithm produces an average of 69 clusters and an average F-measure value of 0.779, which is very close to the ground truth. The evaluation results show that the hypothesis, which states that the spam images that are visually similar can provide a substantial clue in detecting the commonality in the source of the spam and hence detecting the spammers, is quite likely to be true.

The proposed approach adopts the data mining and image processing techniques in the field of computer forensics, which brings an innovative interdisciplinary perspective to this line of research. Our study on spam image clustering goes beyond traditional means of spam filtering and is among a few recent efforts in identifying the common source of spam. Just like the other approaches in this field, e.g., spam clustering according to common subjects and/or common IP addresses, the clustering of spam images is only one key piece of a complex jigsaw puzzle. Those techniques, when put together properly, can reveal the common source of spam and thus contribute to the capturing and impeding of those elusive cyber criminals.

Key Terms: Spam Images, Clustering, Computer Forensics.