



Enhancing Enterprise Email

Security

Accountability

Productivity

Image Spam - The New Face of Email Threats

Email solicitations that use graphical images of text, are the latest and most dreadful form of spam. Image spam was almost unheard of back in 2005, but today nearly 40% un-wanted email is image spam. Also, image spam is the main reason spam traffic doubled in 2006, and it is getting worse. Unfortunately, current generation of spam filters are unable to keep up with this new threat.

Indeed, email messages with images (typically including html text and links) have been around for almost a decade now. So, how is image spam any different from other marketing messages which include images? And why is it so difficult for spam filters to catch image spam?

Current generation of spam filters rely heavily on keyword matching techniques. They inspect text content - such as words, phrases, web-site addresses (URL), domains, IP addresses, and other textual patterns in an email. Those patterns are then matched against black-lists, or dictionaries of objectionable things that make any given message "spammy". If a message seems spammy enough, the filter blocks it. For example, a spam for pornography would most likely include a picture with a link pointing to their web-site. The filters would detect that web-site address as spammy, while ignoring the actual picture itself.

On the other hand, recent genre of image spam do not have any underlying link, or any relevant text that would give it away as spam. Indeed, majority of these image spams are pitching penny-stocks - showing the stock symbol in an image, with no links or other "spammy" text. To make matters worse, many of these image spams also trick the filters, by including lots of irrelevant but "good" text - such as news-stories from NY-Times, or excerpts from Bible. As a result, the spam filters do not find any objectionable text in such messages, and let them thru.

Ironically, the deceptive trick behind image spam is graceful in its simplicity - computers can't see the images.

In contrast, PUREmail, a second generation email security engine, is using a different approach to successfully identify image spam. **PUREmail uses image processing techniques to "visualize" those images, and then "interpret" them using artificial intelligence - thus closely simulating human perception.**

While these image processing techniques have been around for quite some time (and are commonly found in image scanning / image editing software) - their usefulness is only limited to presenting visual data to a human, who then responds with certain action. For example, a graphic designer may "fix" an image during an iterative process of several edits. On the other hand, email filters must make an intelligent interpretation "on-the-fly" without human intervention or visual feedback. That is why, many popular commercial email filters which have lately experimented with certain image processing techniques, have not been able to achieve consistent results and accuracy. **However, a cocktail of these techniques, mixed in a specific order and calibration - a complex algorithm perfected by PUREmail - produced nearly 98% accuracy in detecting image spam.**



Enhancing Enterprise Email

Security

Accountability

Productivity

Here is a brief overview of some of these techniques, their individual deficiencies, and how PUREmail combines them effectively to combat image spam:

- **OCR** - Optical Character Recognition works by searching images for shapes that match the alpha-numeric characters (letters), then translating a matched geometric shape into real text. As such, OCR would seem to be the most obvious choice for detecting image spam, since it is the closest thing to "vision" computers can get.

However, to defeat OCR, spammers upset the geometry of letters and use varying colors. As a result, OCR can't "see" a letter, but the human eye easily recognizes it. Spammers also sprinkle the image with confetti-like speckles ("random noise"). This doesn't affect the legibility of the necessary information, but makes every message unique - so that filter can not find patterns or high volumes of identical images.

- **Noise Reduction** - To overcome the limitation of OCR, PUREmail first removes the background noise by using "pixel-blending" technique. It smoothens the image by taking the average value of adjacent pixels. While "pixel-blending" largely succeeds in removing speckles of noise, it also has a negative side-effect. It makes the image lose its sharpness and contrast - making it harder for OCR to detect geometric shapes of letters.
- **Edge Detection** - After removing the background noise, PUREmail restores the sharpness of the characters by using "edge-detection" technique. An aggregate sampling of color gradient and statistical variance provides a "Histogram" which helps identify the continuous boundaries or "edges" of geometrical shapes. Once the edges are accurately established, the character recognition becomes easy.
- **Skin Detection** - PUREmail has developed a unique way of flagging messages containing pornography images. It works on a simple premise that pornography images expose excessive amount of skin - i.e. nearly 80% of the color on such images is of the same hue (varying shades of skin color) - while remaining 20% colors on the palette represent ALL other objects in the picture. Based upon this 80-20 rule, and some other statistical modeling, PUREmail is able to detect porn spam with nearly 99% accuracy.

Although, a combination of these techniques have produced great results, it comes at a cost. Analyzing an image is not so easy. There is so much data in an image — millions of pixels made of 0s and 1s, which must be sampled, analyzed, adjusted and re-analyzed, and then interpreted (using artificial intelligence) - all in real time. This requires intense calculations and massive computing power. PUREmail has developed a unique architecture of "clustered" Linux servers (inexpensive PC compatible computers), which provide parallel processing, load balancing, fail-over and high-availability - at a fraction of cost of high end servers.

To be sure, the technology developed by PUREmail is not perfect (at only 98% accuracy). However, it has enormous potential - even beyond image spam. PUREmail applies artificial intelligence techniques and natural language analysis to determine the overall context of the entire email message. This context can be further matched with "corporate policies" and "individual preferences" to determine the relevance of each email to its recipient. After all, the definition of spam is highly subjective. What is junk for one person, may be useful for another.