Mail-SeCure™

Inappropriate Content Control™

Whitepaper

Image Spam in a nutshell

Spammers are consistently creating sophisticated new weapons in their armed race against Anti-Spam technology, the latest of which is known as Image-based Spam. Spammers have used images in their messages for years, in most cases to offer a peek at a pornographic Web site, or to illustrate the effectiveness of their miracle drugs. But as more of their text-based messages started being blocked, Spammers searched for new methods and realized that putting their words inside the image could frustrate text filtering. The use of other people's computers to send their bandwidth-hogging e-mail made the tactic practical. The number of unsolicited messages containing images has grown significantly throughout 2006 and now represents 25 to 45 percent of all junk e-mail (Spam Doubles, Finding New Ways to Deliver Itself, NYT, December 6th, 2006).

Image Spam is expected to continue its growth and spread process. Through constant monitoring, PineApp has identified that Image-based Spam tends to be distributed in massive waves; at one of the distribution peaks, PineApp measured Image-based Spam as 30% of all global Spam. Image-based Spam creates bandwidth and storage problems, since a typical Image-based Spam message weighs more than three times that of a regular Spam message. During Image-Spam distribution peaks, bandwidth and storage requirements increase by 70% in average. Therefore, Image-based Spam leads to loss of productivity and IT resources' heavy burden.

PineApp has implemented a unique solution which enables decoding images and treats them with RPD (Recurrent Pattern Detection) module, similarly to other types of Spam. This solution improves PineApp's already superior Spam detection rate and maintains its low false positive rate.

Questions & Answers

What is Image-based Spam?

Image-based Spam is Spam which contains its unwanted content inside of graphics (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; however, in reality it is merely an image of text. Often the content of such messages are penny stock "Pump & Dump" schemes and other malicious types of Spam. Since creating Image-based Spam requires more technical know-how than basic textual Spam does, it originates in areas such as Asia and Eastern Europe, which have technically advanced Spammers.

What are the most common trends in Image-based Spam?

Spammers have been experimenting with new techniques such as "broken images," i.e. splitting a single image into smaller images that fit together like puzzle pieces. This technique makes it even more difficult for anti-Spam engines to detect and block.

Another highly used technique is to send animated GIFs, with several frames of random noise. These random pixels act similarly to randomized images which are not animated, simply with another level of complexity. In some cases, the animated GIFs contain subliminal messages (e.g. "buy… buy… buy") embedded into frames that flash very quickly. Animated GIF Spam is much heavier, in average, than static Image-based Spam.

Why is it so Difficult for Most Anti-Spam Engines to detect and Block Image-based Spam?

These unsolicited emails contain no text or hyperlinks, so most Anti-Spam Engines cannot detect this type of Spam. Often the message will contain a text copied from legitimate books, in order to fool Bayesian (context based) filters.

Spammers have figured out a way to fool even those engines that try to analyze the image data itself: they slightly vary the images in each message. They do this easily by changing the shade of the border or background, changing line spacing or margins, or even adding tiny specks to the background; these types of changes are invisible to the eye (or irrelevant to the reader), but **completely change the way the data appears to most Anti-Spam Engines**. The result is a huge quantity of Image-based Spam, which contains random patterns with almost no repetitions.

Currently, none of the traditional Anti-Spam technologies – content-based, Bayesian, Heuristic, URL Filtering etc. – are able to prevent this type of Spam on a consistently accurate basis. Thus, the Image-based Spam defense engine is an essential tool in the fight against Spam.

Inappropriate Content Control

Inappropriate content, such as pornography, has become a major concern in email based communication, as it leads to exposure to improper content, possible sexual harassment and even exposure to Viruses and other malware, generated in email-linked pornographic websites.

PineApp has added its new-generation Inappropriate Content Control (ICC) technology into Mail-SeCure appliance as an optional module.

The preliminary ICC package includes two central features: IWF (Internet Watch Foundation) database lookup and an image analyzer engine.

Using IWF database lookup, Mail-SeCure is able to block email messages which contain links to websites or URLs of child pornography, child abuse, etc.

The image analyzer engine identifies inappropriate or pornographic images in digital data transmission and other multimedia files (such as videos and PowerPoint files). The system can either block or delete the suspicious mail and notify the administrator, according to the customer's preferences and configured severity level.

Questions & Answers

What is Inappropriate Content?

Inappropriate content ranges from soft pornography to hard pornography, and on to indecent photographs and indecent pseudo-photographs of children.

What is the importance of Inappropriate Content Control?

Corporations are being sued by employees who have been sexually harassed or bullied by their co-workers,

through exposure to inappropriate content in the workplace. Under the Protection form Harassment Act (1997), corporations are vicariously liable to these employees complaining of harassment. A corporation's **only** defense in such cases is interdiction (preventing the harassment in the first place) and taking all reasonably, practical measures to avoid the act.

It is a known fact that the prevention of an event, which would otherwise give rise to a cause of action in law, is far better than defending the action later.

The new generation of Inappropriate Content Control (ICC) technology, is the way in which corporations can prevent such action from occurring.

How does the IWF Database work?

The new IWF-Compliance feature offers greater email compliance by tagging messages which contain references to child abuse image sites (i.e. child pornography sites). PineApp has integrated a URL blacklist provided by the Internet Watch Foundation (www.iwf.org.uk) into the ICC module. As a result, PineApp now also provides a new anti-child abuse feature to their customers.

The IWF works in collaboration with the police, governmental organizations, the wider online industry and the public to combat the availability of illegal online content. The URL list is updated daily with new sites hosting child abuse images.

How does the Image-Analyzer technology work?

The Image Analyzer engine scans the composition of images, in order to determine attributes that indicate whether the image may be of a pornographic nature. It uses sophisticated analytical processes consisting of thousands of algorithms. These include 11 different detection methods and hundreds of individual scanning parameters, set to provide sufficient information to reliably distinguish between pornographic and non-pornographic images.