

Emerging Trends in Fighting Spam

An Osterman Research White Paper

Published June 2007

SPONSORED BY



Why You Should Read This White Paper

Spam has been a serious problem for email administrators and users alike for more than five years, growing from one in six emails in 2002 to approximately three out of four emails today. In fact, after five years of dealing with spam and throwing significant resources at the problem, 54% of messaging decision-makers in mid-sized and large organizations still view spam as a serious or very serious problem according to a recent Osterman Research survey.

The fundamental difficulty in solving the spam problem is that spammers are becoming more sophisticated in their approach to defeating conventional spam-blocking technologies. Among these techniques is the use of image spam that is more difficult to detect and block than text-based spam. In addition, image spam messages are typically much larger than their text-based counterparts and so consume dramatically more storage and bandwidth.

This white paper focuses on the problems caused by image spam, as well as other spammer techniques for delivering content through existing spam-blocking defenses. The paper also discusses Symantec's approach to solving the problem of image spam, botnets and other threats.

The fundamental difficulty in solving the spam problem is that spammers are becoming more sophisticated in their approach to defeating conventional spam-blocking technologies.

The Growing Problem with Image Spam

Image spam is a growing and serious problem facing email administrators and end users alike. In a March 2007 survey conducted by Osterman Research, more than 60% of messaging decision makers cited image spam as a problem for their organizations – nearly one in five decision makers cited it as a serious or 'huge' problem.

What Exactly is Image Spam?

The concept of image spam is really quite simple: instead of using text characters as in a normal spam message, text characters are presented in one or more images. These messages typically contain very little, if any, text, instead relying on the presentation of content in one or more image files.

Why is Image Spam Such a Problem?

Unlike text-based spam, image spam presents a number of unique challenges to IT administrators trying to rid their organizations of this content, as well as end users who receive these messages in their inbox:

- Conventional spam-blocking tools that rely on textual analysis of incoming messages are designed to look for keywords and other content within the body of messages. While these tools generally work well at detecting suspicious content, they do not work well against image spam, since there is usually very little text within these messages – the image of the text displayed within the image cannot be detected using conventional technologies.

It is important to note that content-scanning technologies are just one tool of the several that should be used to block spam, including reputation analysis and others, as discussed later in this paper.

Unlike simple misspellings of words that are used in text-based to fool spam-blocking content filters, image spam uses a variety of techniques to fool signature-based and other content filtering tools.

- The average image spam is typically five to ten times larger than a text-based spam message, resulting in much greater impacts on bandwidth and storage. For example, if image spam represents only 10% of the total number of spam messages received by an organization, it may consume 50% or more of the total bandwidth and storage required to receive and process it.
- Image spam volumes rose dramatically during 2006 and are continuing in 2007. For example, in 2005 image spam represented less than four percent of all spam messages received – as of the first half of 2007, that figure is more than 40%.

Image Spam Soaks Up Corporate Resources

In short, the fundamental problem with image spam is that it consumes significant quantities of system resources, including CPU cycles, bandwidth and storage. As Symantec correctly notes, the amount of image spam that an organization receives can be closely correlated with the use of system resources.

New Approaches Are Needed

Image Spammers Use Innovative Techniques

Unlike simple misspellings of words that are used in text-based to fool spam-blocking content filters, image spam uses a variety of techniques to fool signature-based and other content filtering tools, including:

- Use of different background colors with the same message content.
- Randomly-placed pixels to make each message look unique to a spam filter.
- The use of 'snow' in a message.
- The use of odd, unusual or slanted fonts that make detection of content more difficult.
- The use of multiple images to make up a single image.

Some of these techniques are shown in the image spam example below.

Not unlike text-based spam, image spam uses a variety of appealing messages to induce recipients to open the emails and take action.

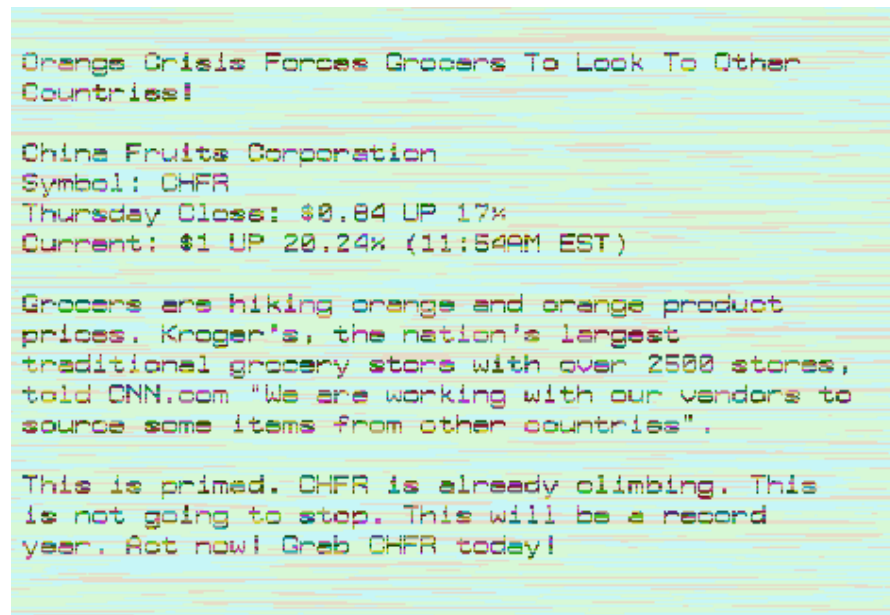


Image Spam Content is Appealing

Not unlike text-based spam, image spam uses a variety of appealing messages to induce recipients to open the emails and take action. For example, many spam messages contain stock 'pump-and-dump' schemes designed to drive up the price of questionable or marginal stocks. Many image spams contain subject lines with timely information on upcoming storms or other events. Lots of image spam contains advertisements for cheap software or weight loss breakthroughs or dating sites.

In short, the messages contained in image spam are largely the same as those sent in text-based spam, but the use of images as the delivery vehicle means that these messages have a higher likelihood of reaching their intended audience by defeating existing spam-blocking defenses.

Obfuscation Techniques Will Become More Onerous

Image spammers will continue to make increased use of randomization in order to make detection of their content more difficult. These techniques will include greater use of the randomization techniques already used, such as changing colors and inserting random pixels in messages, as well as displaying text at different angles in an attempt to render image spam-detection systems useless.

Other Spam Problems are Worsening

Image spam is by no means the only problem faced by organizations that are trying to protect their users from the growing onslaught of undesired content infiltrating their networks. For example:

- A growing proportion of spam is generated by zombies that are part of enormous botnets of infected computers. Symantec reported in March 2007 that it had discovered more than six million zombies worldwide during the second half of 2006 – a 29% increase compared to the first half of 2006. More than 80% of spam is today generated by zombies.

Spam coming from botnets is very difficult to detect and stop. Zombies typically send relatively small amounts of spam – perhaps only 100 or fewer messages per day – which makes differentiating between botnet and valid traffic difficult. For example, Symantec found that 42% of the spam received by one of its customers came from IP addresses that sent five or fewer messages.

Even if a significant percentage of botnet traffic could be blocked, the speed with which spam campaigns are started and stopped, the fact that the actual machines used to send spam change often, and the enormous numbers of zombies that can be used for a particular spam campaign make blocking zombies' IP addresses relatively ineffective.

Spam coming from botnets is very difficult to detect and stop. Zombies typically send relatively small amounts of spam – perhaps only 100 or fewer messages per day – which makes differentiating between botnet and valid traffic difficult.

- During the second half of 2006, spam volumes increased dramatically, growing by roughly 100% in just six months. This has inundated many networks that are now receiving quantities of spam that tax their storage and bandwidth.
- Phishing attempts are becoming more sophisticated and still fool a significant percentage of even sophisticated email users.
- Spam is now morphing into a more sophisticated form of attack in which spam and phishing attempts are integrated into a much greater threat. Consequently, organizations should be evaluating technologies that take an integrated approach to spam and phishing.

For example, Symantec's Response Network provides the ability to identify spam and phishing in an integrated fashion by pushing rules to gateway devices for filtering out specific spam attributes and URL lists that are known phishing sites. This provides for more unified protection at the gateway.

- It is also important to note that spam is also a vector for other forms of attacks that can include viruses, spyware and other forms of malware.

The 'Real' Problems Caused by Spam

Many reports have cited the negative effect that spam has on user productivity, focusing on the amount of time that users spend dealing with spam. However, productivity is really not a critical issue in the context of spam management, since most corporate users receive relatively little spam in their inbox because of the effectiveness of spam filtering tools. Even if a user spends a few minutes each day deleting spam from their mailbox, they will typically spend less time involved in this activity than they will talking around the water cooler or taking a long lunch.

Instead, the real problems with spam focus much more on the impact that spam has on straining bandwidth and dramatically increasing messaging-related storage requirements. For example, Osterman Research has consistently found that growth in messaging storage requirements is the leading problem faced by messaging administrators, a problem caused in part by increasing quantities of spam entering corporate networks. This makes it increasingly important to deploy gateway solutions that are not tied directly to email servers, since these solutions

The real problems with spam focus much more on the impact that spam has on reducing email server performance, taxing bandwidth and dramatically increasing messaging-related storage requirements.

can stop spam before it impacts email servers and networks downstream.

Further, users have become very sensitive to spam, so much so that an increase of just one or two spam messages received per day will cause many users to complain to their IT administrators, a problem that many administrators would like to eliminate. This is further evidence that current spam filtering technologies actually work quite well.

Spammers Will Become More Difficult to Stop

Spammers are becoming more sophisticated and more difficult to combat. They will use more image spam and more sophisticated botnets that will randomize not only messages, but also the patterns in which these messages are sent. Spammers will use a variety of additional binary content formats, including video and audio, in an attempt to break through spam defenses. The net result will be either more IT resources spent on combatting spam or smarter ways of approaching the problem. It will become increasingly important to have available a worldwide network to discover threats as quickly as possible so that rules can be created as soon as possible after the discovery of new threats. It will also be important to use vendors that have a consistent track record of detecting and resolving threats early so that problems can be averted quickly and efficiently.

Reputation scores can be derived from multiple sources, including analysis of spam verdicts in local mail flow and analysis of broader traffic volumes from a variety of sources across the Internet, such as spam traps, user feedback, network topology, mail volumes and other sources.

Symantec's Solution to Image Spam

Stopping image spam and the growing onslaught of spam in general requires a multi-faceted approach:

- **Reputation analysis**

The reputation of the sender of spam, as defined by the domain or the IP address from which spam is being sent, is an important component in filtering image spam and other unwanted content. Reputation scores can be derived from multiple sources, including analysis of spam verdicts in local mail flow and analysis of broader traffic volumes from a variety of sources across the Internet, such as spam traps, user feedback, network topology, mail volumes and other sources.

Using this information, dynamic block lists can be developed to simply block incoming email from suspected spamming sources, or the more sophisticated technique of traffic shaping can be employed, in which

traffic from suspect sources is simply throttled. Symantec has found that up to 80% of incoming image spam, for example, can be stopped simply by properly evaluating the reputation of the sender. It is important to use a vendor that scans enormous of Internet traffic so that as many potential threats as possible are examined.

- **Image structure**

Symantec has found that a continuous analysis of image spam attacks reveals certain characteristics in the structure of image spam. These characteristics can then be used to distinguish the structure of image spam which can act as a fingerprint to detect and block these messages. It is important to note that this process does not require local image analysis which would require enormous amounts of processing power on the customer site.

- **Image content**

Despite the challenges presented by image obfuscation techniques, analysis of image content can be successful in identifying image spam for some attacks. Symantec's approaches include using a 'fuzzy matching' technique that can be used to compare a known spam image to a potential spam image, and identifying the 'spammy' characteristics of an image.

Symantec's approach is to view spam management in a layered manner, running incoming content through more than 20 different technologies in order to minimize the amount of spam that reaches the end user. For example, the initial layer through which incoming content is run focuses on reputation analysis. These systems can eliminate most spam messages quickly and with a minimum of CPU cycles. The next step is to run the remaining content through systems that will purify the mail stream by performing deeper analysis of the structure and content of each message, eliminating most of the remaining spam content and thereby minimizing the negative impacts that spam has on email server performance, storage and bandwidth.

Symantec's approach is to view spam management in a layered manner, running incoming content through more than 20 different technologies in order to minimize the amount of spam that reaches the end user.

Summary

Image spam is a serious and growing threat for organizations of all sizes and is part of a growing trend toward more sophistication in the way that spammers are attempting to defeat spam-blocking systems. Because image spam can defeat traditional anti-spam technologies, new approaches are needed to identify and block image spam effectively. These tools must be part of a holistic and multi-layered approach to spam management that uses reputation analysis and deep content inspection in order to minimize the amount of spam that reaches the end user.

Image spam is a serious and growing threat for organizations of all sizes and is part of a growing trend toward more sophistication in the way that spammers are attempting to defeat spam-blocking systems.

© 2007 Osterman Research, Inc. All rights reserved.

No part of this document may be reproduced in any form by any means, nor may it be distributed without the permission of Osterman Research, Inc., nor may it be resold by any entity other than Osterman Research, Inc., without prior written authorization of Osterman Research, Inc.

THIS DOCUMENT IS PROVIDED "AS IS". ALL EXPRESS OR IMPLIED REPRESENTATIONS, CONDITIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OR FITNESS FOR A PARTICULAR PURPOSE, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE DETERMINED TO BE ILLEGAL.