SECURE COMPUTING

White Paper

www.securecomputing.com

Image spam: The latest attack on the enterprise inbox

Secure Computing® has been solving the most difficult network and application security challenges for over 20 years. We help our customers create trusted environments both inside and outside their organizations.

Table of contents

Secure Computing Corporation

Corporate Headquarters
4810 Harwood Road
San Jose, CA 95124 USA
Tel +1.800.379.4944
Tel +1.408.979.6100
Fax +1.408.979.6501

European Headquarters
No 1 The Arena
Downshire Way
Bracknell
Berkshire, RG12 1PU UK
Tel +44.0.870.460.4766
Fax +44.0.870.460.4767

Asia/Pac Headquarters
1604-5 MLC Tower
248 Queen's East Road
Wan Chai Hong Kong
Tel +852.2520.2422
Fax +852.2587.1333

Japan Headquarters
Shinjuku Mitsui Bldg. 2, 7F
Nishi-Shinjuku 3-2-11
Shinjuku-ku Tokyo, 160-0023
Japan
Tel +81.3.5339.6310
Fax +81.3.4496.4537

For a complete listing of all our global offices, see *www.securecomputing.com/goto/globaloffices*

Introduction

Spammers have long attempted to bypass anti-spam software by incorporating their sales pitch into an image, rather than sending it as plain text. When they first adopted this practice, they were able to evade simple content recognition tools. As image spamming grew in popularity, anti-spam vendors developed signatures designed to detect specific image spam messages. In doing so, the anti-spam software was able to reference these signatures and reject identical or nearly identical messages. However, spammers have now fired a new barrage of image spam using randomized images that appear identical to the human eye, yet appear to be entirely unique to most anti-spam software. Many of the changes to the images contained within spam messages are so subtle that they require a pixel-by-pixel examination of the image in order to detect the differences.

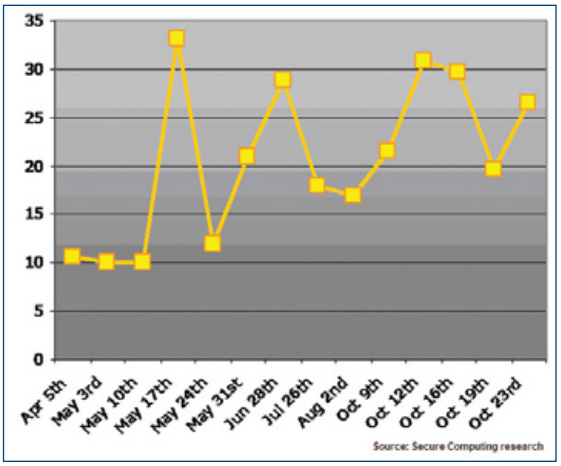


Figure 1: Percentage of image spam in total spam volume.

In recent months, the level of image spam seen by Secure Computing® appliances has increased by nearly 200 percent (Figure 1). This sudden spike in image spam volume can be attributed to the fact that the majority of anti-spam software struggles to detect this new method, making it more appealing and profitable to spammers. Traditional techniques used for detecting and blocking spam have thus far been unable to provide equal effectiveness when dealing with the new image spamming methods.

To combat the growing epidemic of image spam, Secure Computing has taken dramatic steps to quickly and accurately detect these messages, stopping them before they reach enterprise inboxes. TrustedSource™, Secure Computing's revolutionary reputation system, identifies unwelcome messages on multiple levels:

- Sender reputation: Is the sender known to have sent spam in the past?
- Message reputation and fingerprinting: Does the message contain elements of spam encountered previously?
- Image fingerprinting: Does a comparison of the image contained in the message contain similarities to known spam images?

Tricks image spammers use

The following sections contain actual image spam messages analyzed by Secure Computing, and an examination of some of the techniques used by spammers to fool anti-spam software.

Sliced images

Very often, image spam messages are not composed of a single image, but of multiple images pieced together to appear as one. The red lines in Figure 2 indicate "cuts" in the image, similar to the creation of a jigsaw puzzle. This technique is effective against many anti-spam solutions because it bypasses the signature files that have been designed to detect individual images. Spammers send out multiple versions of the same message by slicing it randomly and then reassembling it within the email.

Figure 2: Sliced image.

Random pixel modification

Here, the spammer randomly changes individual pixels within the image that would otherwise likely go unnoticed by the reader (see Figure 3). As a result, each separate iteration of this image will appear completely unique to most anti-spam software. By using random pixel modification, spammers can create virtually unlimited

versions of the exact same message and fool anti-spam software into identifying each as different from the last.

Color modification

Spammers have unlimited flexibility in the number of colors and fonts they can use in image spam messages; each change results in new pixel locations and identifiers, further modifying the image's properties and

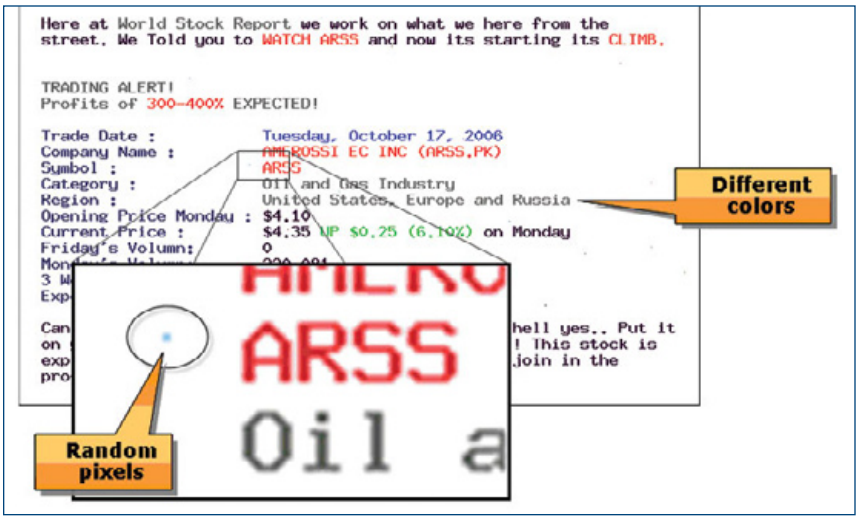


Figure 3: Random pixel and color modification.

distancing it from any signatures that may have been developed using previous versions of the graphic.

Multi-frame animated images

Figure 4 represents the latest modification to image spam messages. This method is completely unique because, instead of sending a single image containing the message, the spammer has created an animated .gif file with multiple frames. The image shown below consists of two separate frames, each containing seemingly random

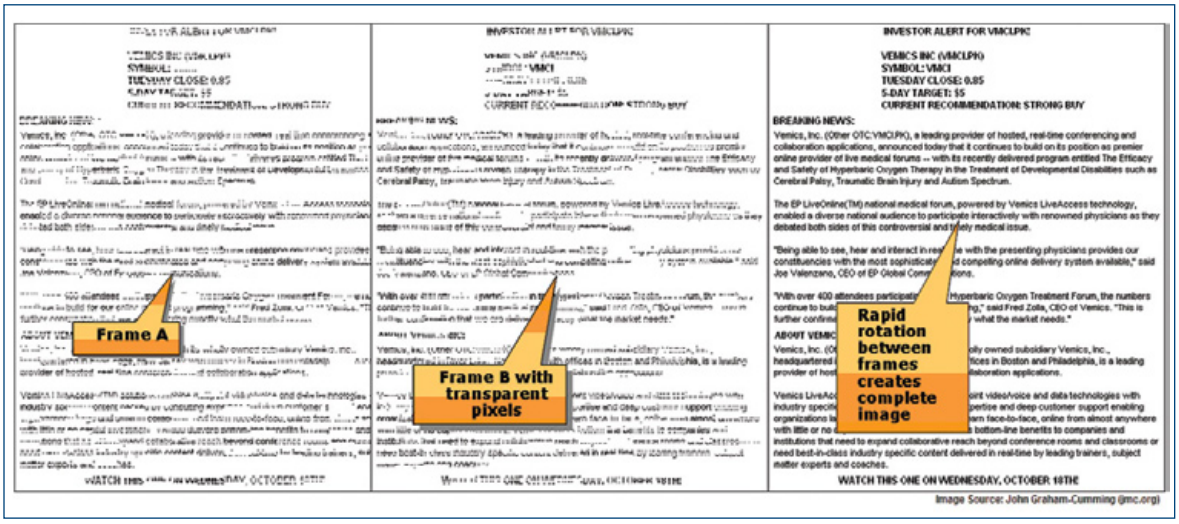


Figure 4: Multi-frame animated image spam.

lines. When superimposed upon each other, the two frames create a complete image. The frames rotate at a rate so fast that the human eye is unable to detect the animation and sees only the final image.

The future of image spam

Spammers rarely rely on one technique for very long, so Secure Computing's research team remains hard at work developing techniques to identify and block spam messages that have not yet been encountered. We anticipate that future iterations of image spam messages will incorporate photographs with hand-written text

Figure 5: Photograph-based image spam with hand-written characters

superimposed over them. By using hand-written characters, spammers can effectively hide their messages from Optical Character Recognition (OCR) technology, which requires known fonts in order to be effective.

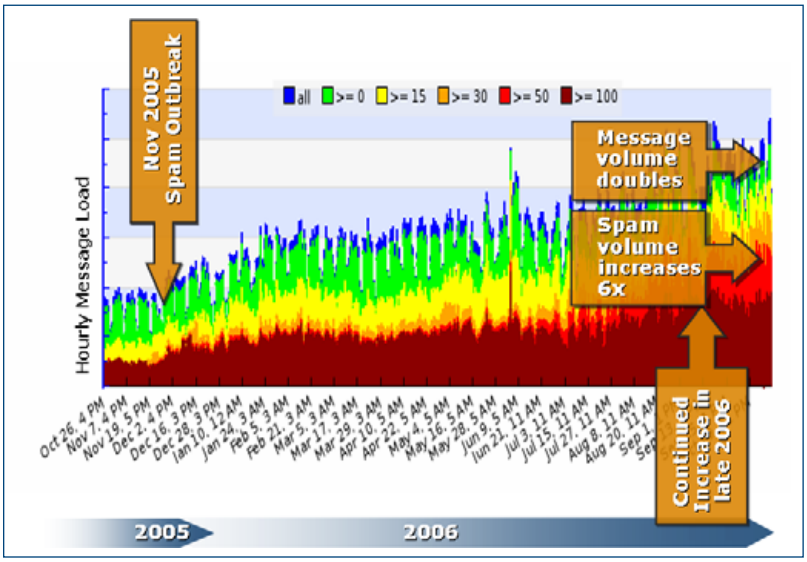
OCR: Why doesn't it stop image spam?

Faced with increasing frustration from customers inundated by image spam, many vendors have turned to old technology to combat a new problem. Optical Character Recognition (OCR) is a technique that attempts to translate images with characters (such as image spam) into text that can be "read" by software. OCR is, in theory, a solid concept—however, it is slow, processor-intensive and relatively easy to fool.

- **Resource depletion:** Reliance on OCR as a method of detecting image spam is untenable due to multiple hardware limitations. Because OCR must open each image and compare it to multiple sets of known characters, it necessitates heavy processor loads and causes unacceptable delays in message throughput.
- **Known technology:** Spammers are aware of OCR's techniques, and have long circumvented it by rotating fonts, characters and pictures. Now, the new instances of animated images such as the ones shown in the Multi-Frame Animated Image or Hand-Written Image examples, render OCR completely useless for detecting image spam. Because the OCR technology is designed to inspect only the one frame of an image, the use of multiple frames allows the spammers to completely bypass full inspection. OCR's reliance on character sets makes it an easy target for spammers using hand-written images as opposed to images that include known fonts.

The decision by anti-spam vendors to rely on OCR to detect image spam has proven to be a poor one, as spammers continuously take steps to stay ahead of known technologies. Multi-frame animated images and similarly deceptive image spam are merely another step in rendering OCR irrelevant and ineffective.

TrustedSource: Stopping image spam to reclaim the inbox



The TrustedSource reputation system keeps enterprises ahead of the spammers in the ongoing battle for the inbox by leveraging global intelligence on email senders and the types of messages they send. With more than 7000 sensors located in 48 countries worldwide, Secure Computing sees more email sent

Figure 6: The recent trend of rising spam volumes has been identified and analyzed in deatil by Secure Computing due to increased spammer activity and the continuously improving effectiveness of TrustedSource.

to enterprises and governments than any other messaging security vendor. As a result of Secure Computing's unique view into email traffic, TrustedSource is fed more intelligence than any other reputation system, resulting in superior accuracy when developing a reputation score.

How does TrustedSource work?

Like a virtual credit agency, TrustedSource assigns a reputation score and further classifies senders as good, bad or suspicious based on an in-depth analysis by processing more than a dozen behavior attributes of each sender. TrustedSource is the first and only reputation system to combine traffic data, whitelists, blacklists and network characteristics with the unparalleled strength of global enterprise data, enabling the system to define a reputation for every sender, not just those that have been encountered in the past. As opposed to other offerings that do not integrate reputation into the spam scoring, TrustedSource data provides the most accurate and effective protection against spam, viruses and other unwanted traffic.

TrustedSource analyzes billions of messages per month from Secure Computing's global enterprise network sensors located in enterprises and government institutions. This allows TrustedSource to create a virtual continuum of IP scores, eliminating the estimation and guesswork required in less advanced reputation systems that rely on ranges of only a few dozen score possibilities. Sender reputation scores in TrustedSource are based on both sender history and message characteristics. When formulating a sender's reputation score TrustedSource considers such factors as:

- When was this sender seen for the first time?
- How much email is this sender responsible for?
- Does the sender both send and receive email, or only send emails? Senders who have an abnormally high ratio of sent mail to received mail usually turn out to be spammers.
- Is the sender's behavior sporadic or continuous? Spikes in sending behavior, as opposed to a steady stream of inbound and outbound mail, is a key indicator of spamming.

TrustedSource then utilizes this profile to watch for deviations from expected behavior for any given sender. Secure Computing's IronMail® email gateway appliances positioned around the globe report back to TrustedSource on all mail flow they are seeing, giving TrustedSource a real-time view of worldwide mail traffic. Any deviations from predicted behavior are picked up by TrustedSource and if a new reputation score is derived for a given sender, that new score is immediately made available to all IronMail units in the field.

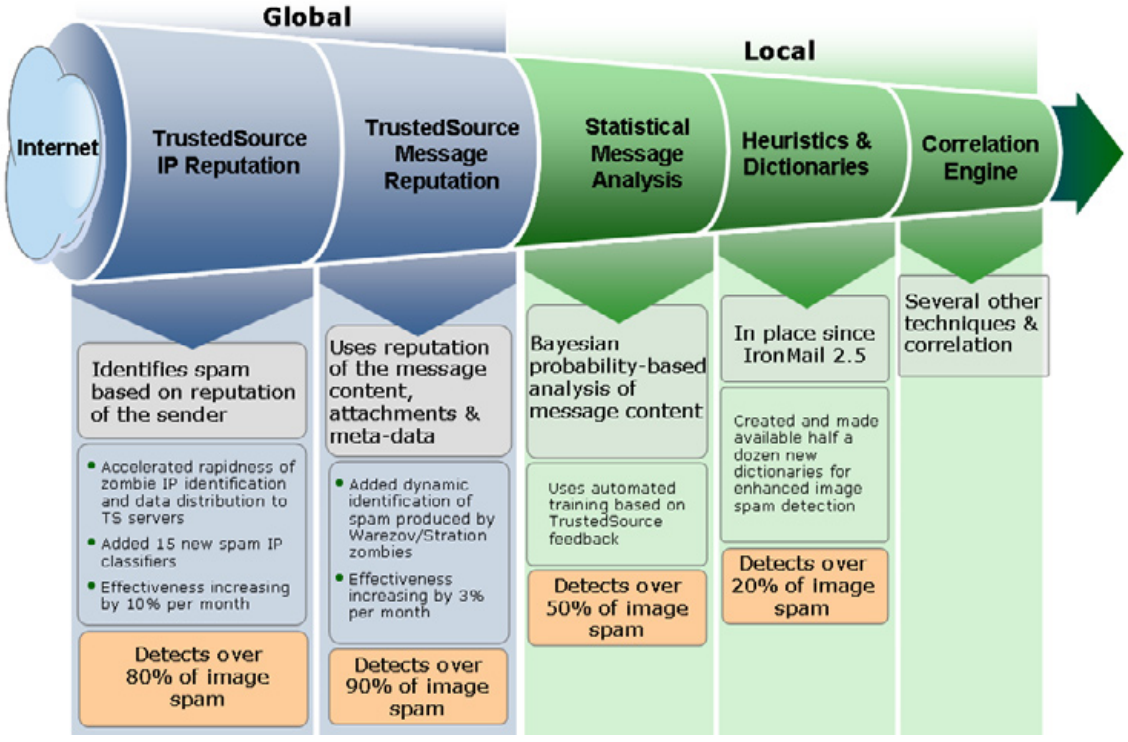


Figure 7: Secure Computing's IronMail appliances receive real-time global intelligence from TrustedSource, while simultaneously applying local processes built into the box to detect image spam. This combination of multiple reputation and analysis technologies provides the highest image spam detection rate in the industry, and reults in continuous improvement as more data is accumulated.

Message reputation and fingerprinting

Sender reputation data from TrustedSource has been incorporated into the IronMail appliance since early 2002, providing real-time behavior analysis on more than one-third of the world's enterprise messaging traffic. In many environments, Secure Computing has been able to block 80 percent of connections based purely on reputation data, increasing security levels while maintaining a false positive rate of less than one in one million. Now, TrustedSource is more effective than ever at fighting nebulous messages such as image spam, thanks to the development of Message Reputation scoring.

The Message Reputation function within TrustedSource is particularly effective in identifying spammers who are using image proliferation and manipulation to evade detection. Leveraging Secure Computing's network of IronMail customers, TrustedSource:

- Analyzes thousands of fingerprints per message, including embedded images and attachments.
- Exchanges more than 20 fingerprints for each message with the TrustedSource network.
- Correlates reputations assigned to each identity by intelligently aggregating the global behavioral and sending pattern knowledge available for each.
- Provides real-time data exchange to allow for instantaneous detection of new zombies (PCs compromised by hackers and used to send spam) and new spam outbreaks such as image-based spam.

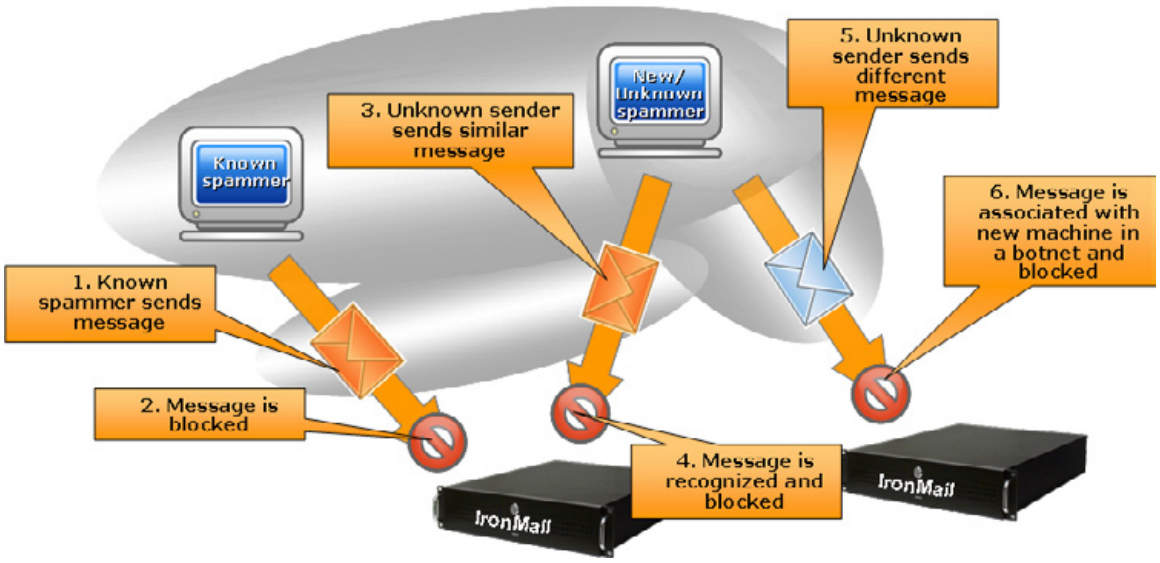


Figure 8: Despite the rise in the global spam volumes and the relatively high success rate of image spam, TrustedSource's effectiveness continues to improve, thanks to technologies such as Message Reputation scoring.

The TrustedSource Message Reputation technology includes a sophisticated and highly accurate message fingerprinting function, and intelligent correlation of that message information with sender-behavior based reputation. No other reputation system offers these capabilities.

The Message Reputation fingerprinting component of TrustedSource tracks specific message compositional elements, and is a highly advanced bulk message detection mechanism. TrustedSource creates thousands of hashes for each message, which are then intelligently processed to identify varying degrees of similarity among messages being sent from all over the world. This fingerprinting engine is very effective at identifying even randomization and obfuscation in image spam messages. This component alone contributes significantly to TrustedSource's effectiveness, but when correlated with sender reputation intelligence, it provides unprecedented accuracy and protection.

Image fingerprinting

To combat the next wave of photograph-based image spam, Secure Computing uses a process known as *normalization* to remove "noise" and obfuscation attempts from an image and strip it down to its basic components. Once the unnecessary content has been stripped from the image, "fingerprints" of the image are created, and then used to identify future photograph-based image spam with similar characteristics (Figure 9). this technique will defeat the attempts of image spammers to hide their messages within a photograph, and will allow Secure Computing IronMail appliances to analyze photograph-based image spam messages and identify them with the same efficiency and reliability as with current types of image spam.

Figure 9: Using normalization to detect photograph-based image spam.

Summary

The recent surge in spam volumes is due in large part to the advancement in image spam technology. In keeping with their standard operating tactics, spammers are constantly creating new techniques either in response to, or in advance of, anti-spam software solutions. By avoiding reliance on OCR and applying TrustedSource reputation to each sender and message encountered, IronMail is able to block more image spam, at a higher rate and with lower false positives than any other gateway security provider. By combining years of industry-leading research with the power of Secure Computing's global network of sensors, TrustedSource continues its tradition of keeping enterprises ahead of spammers and other lurking threats.