

# The Rise and Fall of Image Spam

March, 2008

## Contents

What is Image Spam?	2
Why Image Spam?	2
Image Spam Techniques	2
Why did Image Spam Rise and Fall?	8
Drawbacks of OCR Technology	8
Effectiveness of MailMarshal Against Image Spam	9
Why is MailMarshal So Effective?	9
Conclusion	10

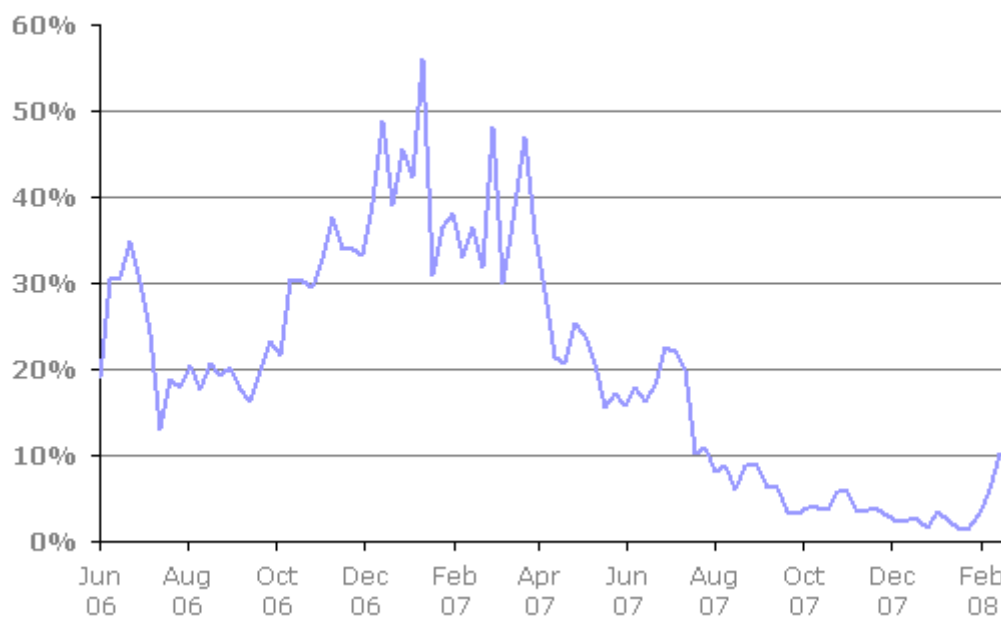
This document discusses the spammer's use of images in spam email, and documents MailMarshal's success in combating this spam type.

## What is Image Spam?

Image spam is simply spam with an attached image. Spam with pictures is not new but in the past almost all spam images were actively downloaded from the web by the email client upon viewing.

Spam with an attached image started to appear in numbers in the second half of 2005. Image spam exploded in mid 2006 and by early 2007 it had reached a peak of over 50% of total spam received. However, by the end of 2007, image spam had dropped back to under 10%. Spammers largely abandoned the technique in the face of better anti-spam defenses against the technique.

Image Spam: Proportion of All Spam



Source: Marshal Spam Archives

## Why Image Spam?

Why did spammers turn to image spam? To avoid anti-spam devices it seems. Image spam has several advantages for spammers:

- It gets the message across, by-passing anti-spam techniques that scan the message body for spam-like text.
- Pretty graphics can allow for a colorful and “professional looking” message.
- It allows for several techniques for spammers to randomize each message, again by-passing anti-spam techniques based on signatures of the message or image, or technologies based on Optical Character Recognition (OCR) i.e. extracting text from images.

## Image Spam Techniques

Various image spam techniques have been used by spammers. This section highlights some of the tricks employed.

## WHITEPAPER – Rise and Fall of Image Spam

### It's only "text"

Some images consist only of words; the image below is from a recent spam email:

Hello,

You have been chosen to participate in an invitation only limited time event!

Are you currently paying over 9% for your mortgage? STOP! We can help you lower that today!

Answer only a few questions and we can give you an approval in under 30 seconds it is that simple!

And stop fighting for lenders, let them fight for you!  
Make them work for your business by giving you the lowest rates around!  
Two hundred and thirty thousand dollar loans are available for only three hundred and forty dollars/month!

**WE ARE PRACTICALLY GIVING AWAY MONEY!**

Think your credit is too bad to get a deal like this? THINK AGAIN!  
We will have you saving your money in no time!

Are you ready to save your money?

Regards.

### Randomization Techniques

To avoid signature-based anti-spam devices, spammers have acquired the ability to randomize each image file on the fly, by subtly altering size, color and inserting random pixels in each file. Here are some examples:

### Stock Spots

Note the insertion of colored pixels at random:



Forecast for July, 2006

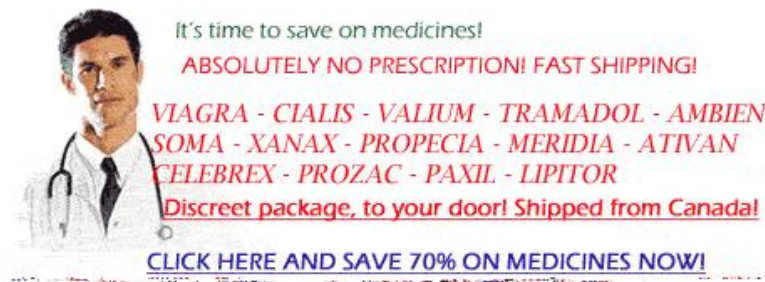
Price: \$5.50

Price Target: \$12.00

## WHITEPAPER – Rise and Fall of Image Spam

### Color Streak

Note the multicolored lines of pixels at the bottom of the image:



### Shades of Gray

Can you see the subtle changes in color between these two images?

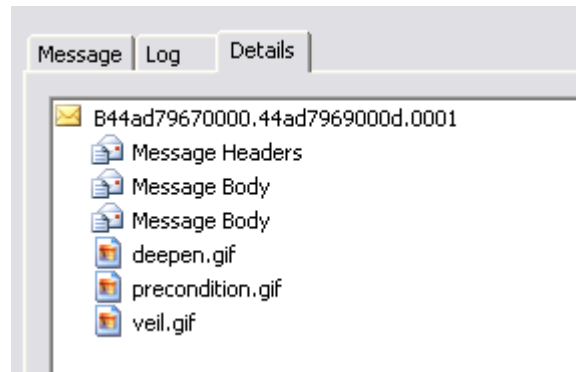


### Stock Splits

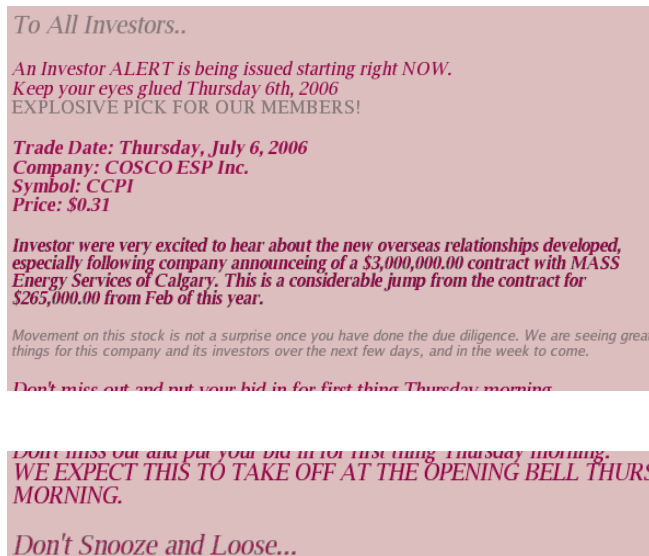
This trick, associated with a certain type of stock spam, splits the original image into multiple images, again to avoid signature filters and possibly image scanning software as well.

## WHITEPAPER – Rise and Fall of Image Spam

Note the three attached images, viewed in a MailMarshal Console:



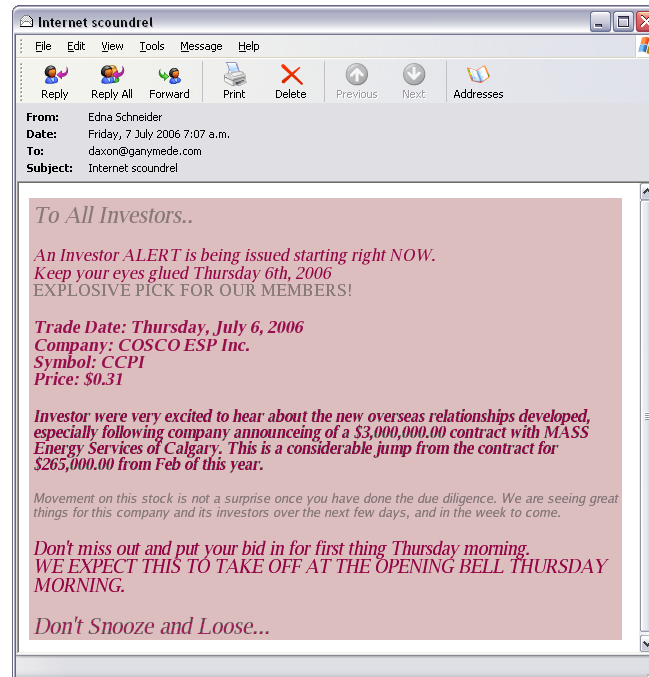
The images are these:



DAY

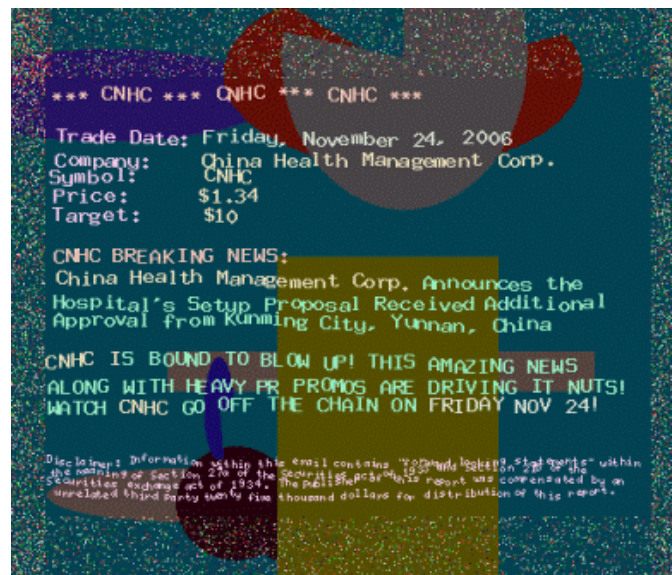
## WHITEPAPER – Rise and Fall of Image Spam

The email client re-assembles the pieces when the message is opened:



## Wild Backgrounds

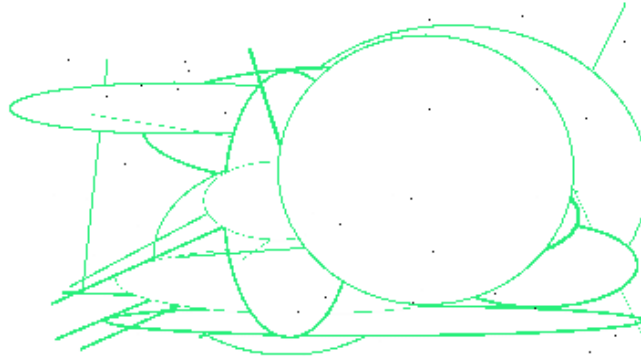
In their quest to bypass anti-spam filters, spammers have sometimes taken their 'art' to extremes, especially to fool OCR (Optical Character Recognition) technology that aims to scan images to extract text. Some images use highly colored and patterned backgrounds, uneven letters, and randomly inserted pixels around the border. Each image is unique and hard to read by any software attempting to use OCR.



## WHITEPAPER – Rise and Fall of Image Spam

### Animated Gifs

Another trick employed is the use of animated gifs. These are created by combining multiple gif images in one file, which when displayed one after another, gives the appearance of movement. Below are two frames from an animated gif spam. The first frame is random stuff, designed to make each image unique and to fool OCR technologies. The second frame is the actual spam message.



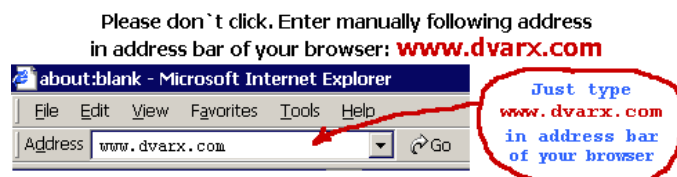
### Want the degree but can't find the time?

WIKI ABOUT IPSA  
The people who we talked to for this article said that work is the best way to get a degree.  
Patterson Herbert is now a graduate.  
It is difficult to find a job these days so you need to be a college graduate.  
Have people check the same website they are all on the internet to find the degree and the work  
that has the degree to get it all in one go.  
Don't waste time trying to find a job and then trying to find a degree to get it all in one go.  
It is a good idea to find a job and then try to find a degree to get it all in one go.  
The way to get a degree is to find a job and then try to find a degree to get it all in one go.  
Just type the URL in the address bar of your browser and you will be able to access it.  
CALL US TODAY AND GET YOUR WORK  
EXPERIENCE THE CHANCE TO EARN YOU  
THE HIGHER COMPENSATION YOU DESERVE!

CALL NOW:

**1-213-596-5768**

Another animated gif example gets around the issue of providing a URL in the message body for a user to click on. Instead the URL is provided in the image. The image is also animated which draws attention to the URL and the user must manually enter it in a browser. The aim here is to evade filters that utilize URL blacklists.



## WHITEPAPER – Rise and Fall of Image Spam

### Standard Images

Let's not forget straightforward "standard" images. No obvious tricks are employed, but the mere fact the essential message is contained in the image renders text scanners useless. Moreover, as opposed to the wild tricks described above, standard images can lend an air of professionalism and authenticity to the email:



The screenshot shows a website for "Canadian Pharmacy" with the tagline "#1 Internet Online Drugstore". It features a grid of nine medication listings, each with a small image of the product, the name, and the price. Below the grid is a promotional message and a "Click here" button.

Medication	Our price
Viagra	\$1.53
Viagra Soft Tabs	\$1.78
Cialis Soft Tabs	\$3.61
Cialis	\$2.23
Viagra Jelly	\$4.25
Levitra	\$3.63
Soma	\$0.67
Kamagra	\$5.56
Herbal Phentermine	\$2.3

We reduced prices and added new products!

[Click here](#)

### Why did Image Spam Rise and Fall?

At its peak, image spam was a huge a problem. The mere use of images renders simple text-based scanners useless. The ability of spammers to randomize and split images means each image is unique. Systems relying on "fingerprinting", signatures or OCR become less effective. Moreover, image spam is larger and the explosion of image spam in late 2006 caused headaches for many companies and ISPs as their infrastructure struggled to cope with the extra burden.

However, during 2007, spammers increasingly moved away from image spam, likely in response to an improved range of anti-spam defenses specifically targeting image spam. Also, as image spam is larger, spambots send less spam per hour than smaller, non-image spam. These factors saw spammers swing back to sending greater numbers of 'ordinary' plain text spam in 2007.

### Drawbacks of OCR Technology

Many vendors have turned to OCR (Optical Character Recognition) in an effort to improve their detection rates against image spam. This technology aims to scan images to extract text.

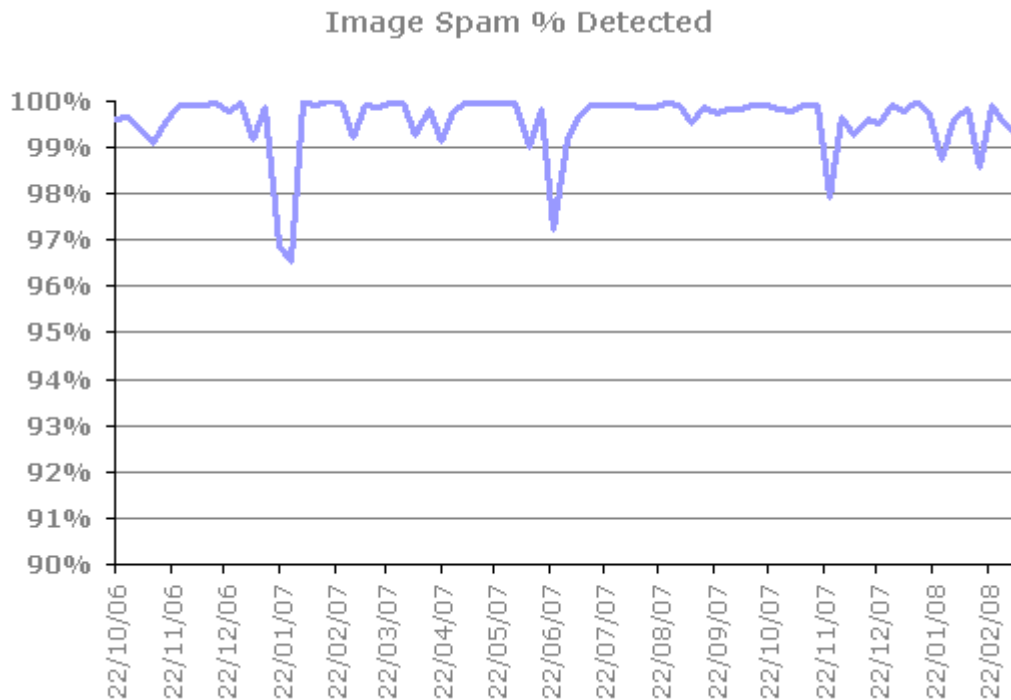
While this idea sounds good, in practice there are a number of problems. It is very resource intensive - scanning all images would cripple a high volume mail server. And, even if you do successfully extract text from an image, you are then left with the problem of weighting or scoring that text – hitting some of the limitations of traditional key word scanning. Furthermore, as we have seen in this paper, the spammers have been quick to adjust to OCR technology by introducing irregular and wavy text, and wild backgrounds in their images.

## WHITEPAPER – Rise and Fall of Image Spam

In sum, OCR is not suitable as a frontline defense against image spam. There is plenty of other information about a typical image spam message to identify it as spam.

### Effectiveness of MailMarshal Against Image Spam

MailMarshal performs exceptionally well against image spam. The graph below shows MailMarshal performance against image Spam since the image spam 'flood' in latter half of 2006. Performance typically remained well above 99% with an average over the period of 99.4%. MailMarshal's lowest point in the whole period still remained above 96.5%.



Source: Marshal live "honeypot" streams

### Why is MailMarshal So Effective?

MailMarshal utilizes the "Defense-in-Depth" concept, where multiple technologies are used to lessen the risk from one component being circumvented. Solutions relying on only one or two technologies are extremely vulnerable to changing spammer tactics. MailMarshal utilizes defense-in-depth with a range of proven anti-spam and content filtering technologies to provide a true layered anti-spam system.

In particular, MailMarshal incorporates Marshal's own SpamCensor technology, which analyzes message content, size and composition, and attachments. The SpamCensor incorporates unique scanning technology that targets the footprints left by spambots - devices installed on compromised computers which are responsible for the overwhelming majority of spam sent today. By targeting these 'botprints' and other repeatable characteristics, the SpamCensor detects new waves of spam, over and over again. The fact that spam may include an image is of little significance. Rather, it is the other information in the email around the image that is important. MailMarshal seeks to target those distinctive patterns.

## WHITEPAPER – Rise and Fall of Image Spam

### **Conclusion**

The image spam phenomenon of late 2006 showed the inventiveness and technological prowess of spammers. Image spam was initially successful at evading anti-spam devices based on simple text scanning, signatures and even newer technologies such as OCR.

Enterprises need a multi-faceted solution that goes beyond traditional spam detection technologies. MailMarshal provides such a solution, and it has proven itself very effective against image spam.

THIS DOCUMENT AND THE SOFTWARE DESCRIBED IN THIS DOCUMENT ARE FURNISHED UNDER AND ARE SUBJECT TO THE TERMS OF A LICENSE AGREEMENT OR A NON-DISCLOSURE AGREEMENT. EXCEPT AS EXPRESSLY SET FORTH IN SUCH LICENSE AGREEMENT OR NON-DISCLOSURE AGREEMENT, MARSHAL LIMITED PROVIDES THIS DOCUMENT AND THE SOFTWARE DESCRIBED IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. SOME JURISDICTIONS DO NOT ALLOW DISCLAIMERS OF EXPRESS OR IMPLIED WARRANTIES IN CERTAIN TRANSACTIONS; THEREFORE, THIS STATEMENT MAY NOT APPLY TO YOU.

This document and the software described in this document may not be lent, sold, or given away without the prior written permission of Marshal, except as otherwise permitted by law. Except as expressly set forth in such license agreement or non-disclosure agreement, no part of this document or the software described in this document may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, or otherwise, without the prior written consent of Marshal. Some companies, names, and data in this document are used for illustration purposes and may not represent real companies, individuals, or data. This document could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein. These changes may be incorporated in new editions of this document. Marshal may make improvements in or changes to the software described in this document at any time.

© 2008 Marshal Limited, all rights reserved.

U.S. Government Restricted Rights: The software and the documentation are commercial computer software and documentation developed at private expense. Use, duplication, or disclosure by the U.S. Government is subject to the terms of the Marshal standard commercial license for the software, and where applicable, the restrictions set forth in the Rights in Technical Data and Computer Software clauses and any successor rules or regulations.

Marshal, MailMarshal, the Marshal logo, WebMarshal, Security Reporting Center and Firewall Suite are trademarks or registered trademarks of Marshal Limited or its subsidiaries in the United Kingdom and other jurisdictions. All other company and product names mentioned are used only for identification purposes and may be trademarks or registered trademarks of their respective companies.



Marshal's Worldwide and EMEA HQ  
Marshal Limited,  
Renaissance 2200,  
Basing View,  
Basingstoke,  
Hampshire RG21 4EQ  
United Kingdom

Phone: +44 (0) 1256 848080  
Fax: +44 (0) 1256 848060

Email: [emea.sales@marshal.com](mailto:emea.sales@marshal.com)

Americas  
Marshal Inc.  
5909 Peachtree Dunwoody Road NE,  
Suite 770,  
Atlanta,  
GA 30328  
USA

Phone: +1 404 564-5800  
Fax: +1 404 564-5801

Email: [americas.sales@marshal.com](mailto:americas.sales@marshal.com)  
[info@marshal.com](mailto:info@marshal.com) | [www.marshal.com](http://www.marshal.com)

Asia-Pacific  
Marshal Software (NZ) Ltd  
Suite 1, Level 1, Building C  
Millennium Centre  
600 Great South Road  
Greenlane, Auckland  
New Zealand

Phone: +64 9 984 5700  
Fax: +64 9 984 5720

Email: [apac.sales@marshal.com](mailto:apac.sales@marshal.com)