MessageLabs®

Be certain

PDF Spam: Spam Evolves, PDF becomes the Latest Threat

A MessageLabs Whitepaper: August 2007

Nick Johnston, Anti-Spam Development at MessageLabs

Table of Contents

Spammers are known to be creative and versatile in their attempts to bypass spam filters.

Introduction

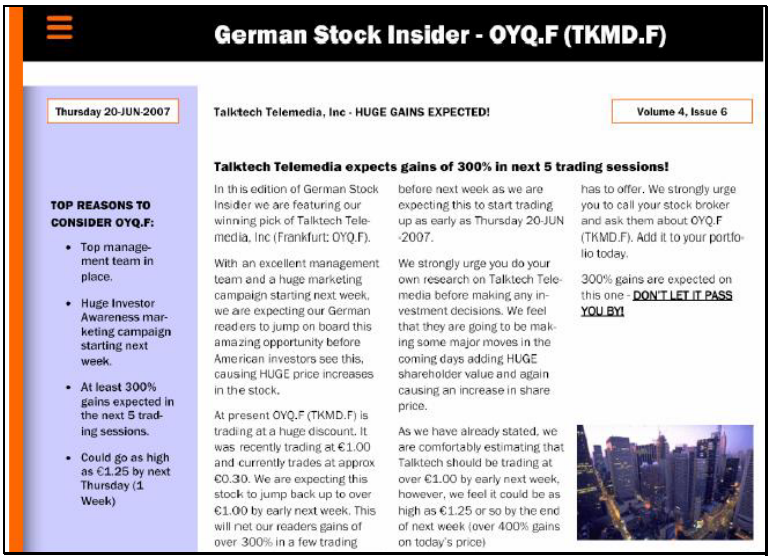
Spammers are known to be highly creative and versatile in their attempts to bypass spam filters. For years, image spam has been very popular, with spammers using a variety of different techniques to randomise their images, making detection more difficult. As both MessageLabs and the wider anti-spam community have improved their image processing techniques, spammers are increasingly switching to a new format: PDF

This MessageLabs White Paper explains what PDF spam is and how spammers are trying to stay a step ahead of techniques designed to tackle it. The information and analysis presented here is based on MessageLabs' hands-on experience of providing messaging and web security management services for over 15,000 clients worldwide, with around 1.5 billion emails processed each week on their behalf.

The Beginning of PDF Spam

PDF (Portable Document Format) is a popular document format invented by Adobe Systems, and is widely used for document exchange in the business world. As such, it is a "trusted" format, and many naïve anti-spam solutions automatically whitelist all messages containing a PDF file. Such is the importance and general acceptance of PDF in the business world that practically all computers in a corporate environment will have a PDF viewer installed. This makes PDF an excellent "vector" for spam messages.

MessageLabs first saw large-scale PDF spam in the middle of June 2007. This "spam run" or "campaign" was a "pump and dump" scam promoting a German stock. Many new types of spam start primitively, and PDF spam was no exception. This first spam run included exactly the same document in each message, making it easy to stop the messages using hashes or "fingerprints" (like MD5 for example).



German Stock Insider - OYQ.F (TKMD.F)

Thursday 20-JUN-2007

Talktech Telemedia, Inc - HUGE GAINS EXPECTED!

Volume 4, Issue 6

Talktech Telemedia expects gains of 300% in next 5 trading sessions!

TOP REASONS TO CONSIDER OYQ.F:

- Top management team in place.
- Huge Investor Awareness marketing campaign starting next week.
- At least 300% gains expected in the next 5 trading sessions.
- Could go as high as €1.25 by next Thursday (1 Week)

In th is edition of German Stock Insider we are featuring our winning pick of Talktech Telemedia, Inc (Frankfurt: OYQ.F).

With an excellent management team and a huge marketing campaign starting next week, we are expecting our German readers to jump on board this amazing opportunity before American investors see this, causing HUGE price increases in the stock.

At present OYQ.F (TKMD.F) is trading at a huge discount. It was recently trading at €1.00 and currently trades at approx €0.30. We are expecting this stock to jump back up to over €1.00 by early next week. This will net our readers gains of over 300% in a few trading

before next week as we are expecting this to start trading up as early as Thursday 20-JUN -2007.

We strongly urge you do your own research on Talktech Telemedia before making any investment decisions. We feel that they are going to be making some major moves in the coming days adding HUGE shareholder value and again causing an increase in share price.

As we have already stated, we are comfortably estimating that Talktech should be trading at over €1.00 by early next week, however, we feel it could be as high as €1.25 or so by the end of next week (over 400% gains on today's price)

has to offer. We strongly urge you to call your stock broker and ask them about OYQ.F (TKMD.F). Add it to your portfolio today.

300% gains are expected on this one - **DON'T LET IT PASS YOU BY!**

PDF Spam example 1

PDF Spam Evolves Quickly

Soon after seeing the first major PDF spam run, MessageLabs began seeing more. But this time, each message had a different PDF file attached. Spammers have long had the ability to randomise images, and have now updated their botnet software to simply insert these random images into PDF documents. This technique means that each PDF that a spammer sends out will be different, and will be more difficult to stop.

Randomised PDFs

The images below (taken from randomised PDFs) illustrate this concept well. Each image includes exactly the same text, but the shape of the image is different, as are the colours:

Spammers have long had the ability to randomise images, and have now updated their botnet software to simply insert these random images into PDF documents.

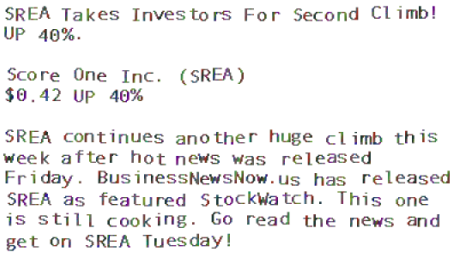


Image 1: Randomised PDF Spam 1

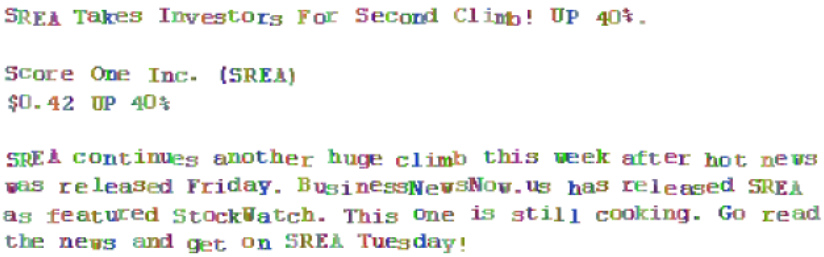


Image 2: Randomised PDF Spam 2

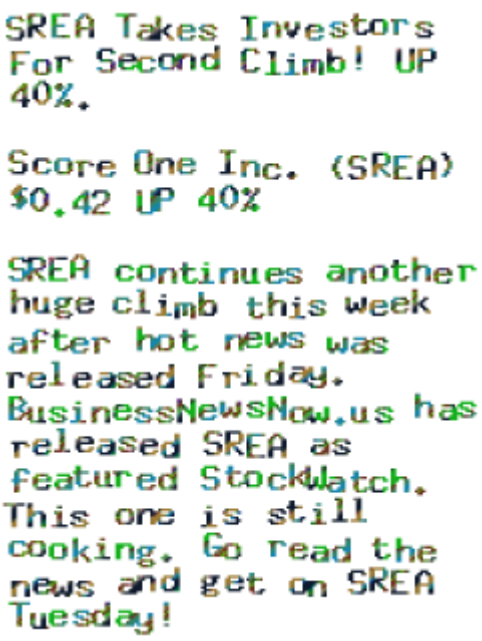


Image 3: Randomised PDF Spam 3

In contrast to legitimate business PDF files, PDF files from this randomised spam run do not use standard paper sizes such as A4 or Letter. For example, one document might be 74.4 × 96 mm, and another might be 168.3 × 54.7 mm.

Corrupted Files to Avoid Detection

Many images in image spam were deliberately corrupted – in other words, the images were constructed without complying with the appropriate specification or standard. By corrupting files, spammers make it more difficult for the analysis tools used by the anti-spam companies to open and analyse the images. Some computer programs would fail to process such images, and indeed these images could cause some programs to become much slower, use more resources or crash. However, spammers rely on the fact that other applications (like many common email clients) are more forgiving and display the images without problems.

MessageLabs has seen similar tactics employed with PDF spam, detecting many corrupted PDF documents. It's unclear if this corruption is accidental or deliberate, but as with corrupt images, strict processing programs tend to fail on these PDF files and so analysis and identification becomes more difficult. The messages can still be viewed by the recipients though because Adobe's Acrobat® Reader displays the PDF correctly (by rebuilding part of the PDF document's internal structure). Some older versions of Acrobat Reader® briefly display a dialog box telling the user that the file is damaged and is being repaired, but this requires no interaction from the user.

Variable Length PDF Files

A more recent tactic seen by MessageLabs is the use of variable length PDF documents.

A more recent tactic seen by MessageLabs is the use of variable length PDF documents. Until recently, most PDF documents sent in spam were simple, single page documents. In contrast, with variable length PDF spam, the first half or so of the first page includes the spam message, and the rest of the page and a random number of subsequent pages contain text "poison".

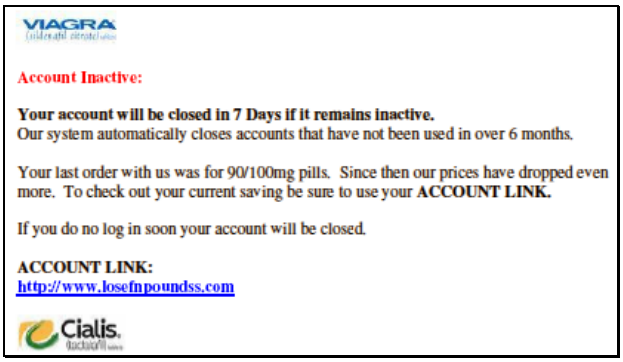
This poison is designed to foil statistical anti-spam techniques such as Bayes. We have seen spam PDF documents containing up to 14 pages of poison. The poison can be random words, programmatically-generated "nonsense text" or legitimate text "scraped" off popular web sites. Some examples of this text include:

```
But in light of their back-stabbing, Artificial
Intelligence-inspired offenses and their sinister,
temptation-ridden environment this response is
degenerate.

Ships from and sold by Amazon.

I also had my tripod and took several amazing long
exposure shots of  the interior.
```

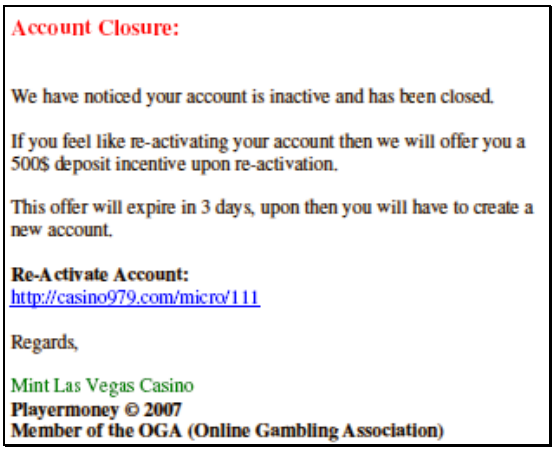
It is likely that spammers think longer PDF documents are more likely to be considered legitimate business documents like reports, manuals and so on.

PDF Spam Diversity

Most press coverage around PDF spam has solely concentrated on "pump-and-dump" stock spam. MessageLabs has seen PDF documents used in other types of spam, such as pharmaceutical spam and online casino spam. Recent examples include:



PDF Spam example 2



PDF Spam example 3

PDF Spam Construction

Spammers are using a wide variety of tools to produce their PDF documents. Many tools include their name as the document “producer” or “creator” in the PDF file itself. Some spammers are using common office applications such as Microsoft Word and OpenOffice:

```
/Producer(GNU Ghostscript 7.07)
/Creator(OpenOffice.org 1.1.4)

/Title(Microsoft Word - sancashtemplate.doc)
/Creator(PScript5.dll Version 5.2.2)
```

Some spammers have also used tools like PowerPDF, text2pdf and so on to produce their PDF documents. More recently, spammers have written their own tools to produce PDF documents. This gives them maximum flexibility, and lets them specify random “producer” names and titles which are difficult to detect by anti-spam software, for example:

```
Title: One of the most interesting things about the
present development of the automobile is the trend to
give cars  a retro look.
Producer: For pure and simple ugly no one has been able
to beat them
```

Title: , has a new promotion that puts its money where its mouth is.
Producer: The flights will be convenient for travellers coming from the U

Although many people are familiar with PDF documents, there are also some related formats which are comparatively unknown. Recently MessageLabs has seen spam claiming to have FDF (Forms Data Format) attachments, which also open with Adobe's Acrobat® Reader. The attachments are actually PDF files merely labelled with a '.fdf' extension. This is likely to be another attempt by the spammers to bypass anti-spam software that only looks at the file extension ('.fdf' in this case), rather than doing reliable checking of the actual file.

Meet the Threat

PDF spam is an increasing problem and now accounts for around 20% of spam.

PDF spam is an increasing problem and now accounts for around 20% of spam. The damage that spam can cause any business should never be underestimated. Efficiency, productivity and profitability can all take a serious hit if electronic junk email gains access to inboxes, with valuable time and effort eaten up in identifying and deleting unwanted messages. MessageLabs stops PDF spam using several different broad techniques:

- Skeptic® heuristics updated around the clock to ensure the highest protection possible from PDF spam.
- Automatic fingerprint-based blocking of known spam PDF files.
- Honeypot monitoring systems for identifying new PDF spam runs.
- Tools to detect corrupted PDF files.
- Generic approaches such as IP blacklisting.

As MessageLabs offers a managed anti-spam service, our customers benefit from seamless, continual system improvement. Combined with our 24 hours per day, 7 days a week, 365 days a year operations and development teams, this ensures that MessageLabs customers are always protected against PDF spam and other emerging spam threats.

www.messagelabs.com
info@messagelabs.com

Freephone UK
0800 917 7733

Toll free US
1-866-460-0000

Europe
HEADQUARTERS
1270 Lansdowne Court
Gloucester Business Park
Gloucester, GL3 4AB
United Kingdom

T +44 (0) 1452 627 627
F +44 (0) 1452 627 628

LONDON
3rd Floor
40 Whitfield Street
London, W1T 2RH
United Kingdom

T +44 (0) 207 291 1960
F +44 (0) 207 291 1937

NETHERLANDS
Teleport Towers
Kingsfordweg 151
1043 GR
Amsterdam
Netherlands

T +31 (0) 20 491 9600
F +31 (0) 20 491 7354

BELGIUM / LUXEMBOURG
Cullinganlaan 1B
B-1831 Diegem
Belgium

T +32 (0) 2 403 12 61
F +32 (0) 2 403 12 12

DACH
Feringastraße 9
85774 Unterföhring
Munich
Germany

T +49 (0) 89 189 43 990
F +49 (0) 89 189 43 999

Americas
AMERICAS HEADQUARTERS
512 Seventh Avenue
6th Floor
New York, NY 10018
USA

T +1 646 519 8100
F +1 646 452 6570

CENTRAL REGION
7760 France Avenue South
Suite 1100
Bloomington, MN 55435
USA

T +1 952 886 7541
F +1 952 886 7498

Asia Pacific
HONG KONG
1601
Tower II
89 Queensway
Admiralty
Hong Kong

T +852 2111 3650
F +852 2111 9061

AUSTRALIA
Level 6
107 Mount Street,
North Sydney
NSW 2060
Australia

T +61 2 8208 7100
F +61 2 9954 9500

SINGAPORE
Level 14
Prudential Tower
30 Cecil Street
Singapore 049712

T +65 62 32 2855
F +65 6232 2300