Blue Coat Malware Search Analysis Project
Michelle Ashton, Aaron Smith, Josh Sudweeks
Malware Ecosystem Analysis

Purpose
• Use Blue Coat's WebPulse request logs to perform malware ecosystem analysis starting with pages overrun by link spam. Develop one or more automated techniques to perform this same analysis and identify malware relays or malware sources.

Key Tasks
• Identify blogs or other sites with link spam
• Harvest links from these sites
• Correlate data with WebPulse request logs
• Understand their role in the malware ecosystem
• Develop automated tools
Link Spam

• Links between pages that are present for reasons other than merit.

• Take advantage of link-based algorithms
  – This gives websites higher rankings the more other highly ranked websites link to it.

• Also aims at influencing other link-based ranking techniques such as the Hyperlink-Induced Topic Search (HITS) algorithm
  – Determines two values for a page: its authority, which estimates the value for a page, and its hub value, which estimates the value of its links to other pages
Link Farms

- Involves creating tightly-knit communities of pages referencing each other, also aim at influencing each other.
Link Farm Process

• Select 20 different sites (that are known to contain Link Spam) from list

• For each of the 20 Link Spam pages, capture and compare the HTML regularly for about a week
  o This is to see if the site's links change much
  o Used DRTR as a link-scraper
DRTR Server

Dynamic Real-Time Rating Server

• System that has been trained to differentiate between different categories using Statistical Analysis.
Comparing HTML with the ‘diff’ command in Linux, no changes shown from comparing the two files.
Findings in Link Farm Analysis

• Links were not being changed, the sites we looked at were being infected and abandoned.
• With Google changing the way they present web pages, link farms were not as successful in making it to Google Search Results first page.
• Google’s restructured algorithm puts less weight on Link Farms/Link Spam.
Malware Search Analysis

Purpose
• Identify malware links in top search results in order to protect Bluecoat customers from malware/spyware

Key Tasks
• Develop an automated way to find keywords
• Use these keywords to find links to malware
• Find additional ways to identify malware links
• Conceptualize process of malware link generation.
Process Used

Gather Data
- Google Trends/Search Results
- Current Events List
- Celebrity + Qualifiers

Setup – Database/Perl Scripts
- MySQL
- Perl Scripts to Extract, Transform, & Load Data
- Generate Malware Search Rules

Analyze Data
- MySQL Workbench
- Perl
- Excel

Results
- Patterns
- Trends
- Other Findings
Gather Keywords/Search Results
Categorize URLs based on Severity Level

Mark suspicion level for each URL. When you are finished click "Submit Query" at the bottom of the page to update the db.

It is not necessary to mark all of the URLs. If you feel like quitting after labeling just a few URLs, click "Submit Query" and only the URLs you have labeled will be updated.

<table>
<thead>
<tr>
<th>Comments</th>
<th>Severity</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>404</td>
<td>2 (Very Suspicious)</td>
<td><a href="http://www.mikeyboy.com/tmnyj.php?m=wildflower%20triathlon%202010">www.mikeyboy.com/tmnyj.php?m=wildflower%20triathlon%202010</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>w3sitesdesign.net/usaj.php?page=hot+in+cleveland+sitcom</td>
</tr>
<tr>
<td></td>
<td></td>
<td>secretimpact.com/gssol.php?off=chemistry%20regents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tempmusic.net/opchc.php?go=leon%20rose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>introchem.chem.okstate.edu/conference/paper.php?go=albie+manzo+law+school</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gloriaandrosendo.com/fbmhu.php?ad=dorell%20wright%20girlfriend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>eiseman.us/kfrq.php?t=afrkala</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bluehillsmotor/oymfl.php?a=amazon%20amanda%20website</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.the-numbers.com/movies/2000/KTTN.php">www.the-numbers.com/movies/2000/KTTN.php</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>teamhinga.net/kxuwq.php?do=schnep%20farm%20coupons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mr2guru.com/etnx.php?a=gail+lumet+buckley+photo</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.guitar-addicts.com/tabs/id/29651.php">www.guitar-addicts.com/tabs/id/29651.php</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.partiyat.com/ols/931th/sonya.php">www.partiyat.com/ols/931th/sonya.php</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>jclutilla.com/jcrf.php?page=pennsylvania%20primary%20election%202010%20date</td>
</tr>
<tr>
<td></td>
<td></td>
<td>apadc.com/trxyp.php?ad=gray%30s+anatomy+season+finale+2010+date</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unswapbook.com/mahwm/doxrq.php?qt=730999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lagarch.net/gxwca.php?page=kris%20oversteeg%20wiki</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hhaircraft.com/kgqoy.php?p=frank+luca+biography</td>
</tr>
<tr>
<td></td>
<td></td>
<td>im3mods.com/ouiper.php?topic=american%20idol%20top%203%202008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wings.buffalo.edu/publications/spectrum/print.php?id=30817</td>
</tr>
<tr>
<td></td>
<td></td>
<td>willowridgefamily.com/oidxr.php?page=mike%20brown%20fired</td>
</tr>
<tr>
<td></td>
<td></td>
<td>themanton.com/madcpi.php?ola=mywebboo</td>
</tr>
</tbody>
</table>
Query Database for Patterns/Trends
Malware Rules

• Bluecoat gave us rules to help find malware
  – Shady tld
  – Yellow flag term
  – Contains .php and keyword in url
  – Contains red flag domain
  – .PHP
Shady Top Level Domains

- .co.cc, .biz, .pl, .cn, .cz, .mx,

<table>
<thead>
<tr>
<th></th>
<th>URLs Searched</th>
<th>Suspicious Pages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Trends</td>
<td>194</td>
<td>83</td>
<td>43%</td>
</tr>
<tr>
<td>Events</td>
<td>76</td>
<td>7</td>
<td>9%</td>
</tr>
<tr>
<td>Celebrities</td>
<td>68</td>
<td>29</td>
<td>43%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>338</strong></td>
<td><strong>119</strong></td>
<td><strong>35%</strong></td>
</tr>
</tbody>
</table>
Yellow Search Term

- Wiki., Blog., /forums/, /members/

<table>
<thead>
<tr>
<th></th>
<th>URLs Searched</th>
<th>Suspicious Pages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Trends</td>
<td>19</td>
<td>11</td>
<td>58%</td>
</tr>
<tr>
<td>Events</td>
<td>110</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Celebrities</td>
<td>132</td>
<td>11</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>261</td>
<td>22</td>
<td>8%</td>
</tr>
</tbody>
</table>
## Php and keyword in URL

<table>
<thead>
<tr>
<th></th>
<th>URLs Searched</th>
<th>Suspicious Pages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Trends</td>
<td>115</td>
<td>9</td>
<td>8%</td>
</tr>
<tr>
<td>Events</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Celebrities</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>116</strong></td>
<td><strong>9</strong></td>
<td><strong>8%</strong></td>
</tr>
</tbody>
</table>
## Red Flag Domain

- Domains that commonly host malware

<table>
<thead>
<tr>
<th></th>
<th>URLs Searched</th>
<th>Suspicious Pages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Trends</td>
<td>91</td>
<td>21</td>
<td>23%</td>
</tr>
<tr>
<td>Events</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Celebrities</td>
<td>44</td>
<td>38</td>
<td>86%</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
<td>59</td>
<td>44%</td>
</tr>
</tbody>
</table>
New Rule

• By analyzing URLs that are known by bluecoat, google and us to contain malware, we found a new rule
  – URLs that contain 5 random letters followed by .php, then followed by more characters.

'www.excelinteriorny.com/ddvjd.php?topic=amy+schumer+30+rock'
[a-zA-Z]{5}.php

- 5 random letters followed by .php
  – (9 common dictionary words excluded)

<table>
<thead>
<tr>
<th></th>
<th>URLs Searched</th>
<th>Suspicious Pages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Trends</td>
<td>1443</td>
<td>906</td>
<td>63%</td>
</tr>
<tr>
<td>Events</td>
<td>4</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>Celebrities</td>
<td>23</td>
<td>5</td>
<td>22%</td>
</tr>
<tr>
<td>Total</td>
<td>1470</td>
<td>912</td>
<td>62%</td>
</tr>
</tbody>
</table>
[a-zA-Z]{5}.php tuning

• With 9 dictionary words removed
  – 2072 pages, 1065 suspicious, 51% suspicious

• With 41 dictionary words removed
  – 1114 pages, 992 suspicious 89% suspicious

• Including malware Google spotted
  – 1656 pages, 1524 suspicious, 93% suspicious
Malware by Date

![Graph showing malware by date with lines for googe_trends, Events, and Celebs]
Google Found Malware By Date

- google_trends
- celebs
- events

Graph showing trends from May 1, 2010 to July 15, 2010.
Findings

• Out of 75 links – 17 would be from sye628.xorg.pl
• Other domains include:
  – systellex.com
  – hhaircraft.com 3
  – www.excelinteriorny.com
  – bluehillsmoto.com
  – rrcrc.com
  – familytiesweddings.com
  – systellex.com
  – helix-x.com
  – kristaremmersmorgan.com
Vs Google

• Google marked 745 pages as suspicious
  – ‘This site may harm your computer’
• Our filters marked 526 of those pages (70%)

• We found a total of 2333 suspicious pages
• Google also found 526 of those pages (22%)
Vs Bluecoat

• 198 sites flagged suspicious or malware by Bluecoat
  – 190 caught by shady tld filter
    • Only 18 confirmed by us, other were not looked at

• No overlap between Bluecoat and google
Perl Script

- Searches the database for xxxxx.php links within 1 day
- Returns URL, keyword and rank
- Verify and block daily
- Average of 13 links per day
Recommendations

• Implement and run similar process daily.
• Expand Search to find/include more trends & patterns.
• Refine rules to easily identify Malware.
• Purge data weekly/monthly after results are found (1.6 Million URLs over 2 months).
Questions?
Firefox security page

Reported Attack Page!

This web page at nadadenepoimetnormalnoblahamuha.xorg.pl has been reported as an attack page and has been blocked based on your security preferences.

Attack pages try to install programs that steal private information, use your computer to attack others, or damage your system.

Some attack pages intentionally distribute harmful software, but many are compromised without the knowledge or permission of their owners.

Get me out of here!   Why was this page blocked?

Ignore this warning
Findings

Windows Security Center

Security essentials
- Get the latest security and virus information from Microsoft
- Check for the best updates from Windows Update
- Get support for security-related issues
- Get help about Security Center
- Change the way Security Center alerts me

Firewall
- ON

Automatic Updates
- ON

Virus Protection
- NOT FOUND

Windows did not find antivirus software on this computer. Antivirus software helps protect your computer against viruses and other security threats. Click Recommendations for suggested actions you can take. This does antivirus software help protect my computer?

Note: Windows does not detect all antivirus programs.

Manage security settings for:
- Internet Options
- Windows Firewall
- Automatic Updates

At Microsoft, we care about your privacy. Please read our privacy statement.
Example

• mrfeature.com/media.php?name=3981-michael-jackson-final-rehearsal
Downloaded .exe

Support Center

Downloading Registry Repair Software...

If your download does not begin automatically, please click here.

©2010

RegCureSetup_RW.exe
1.9 MB