Data Mining the eCriminals: Interesting things lurking in APWG statistics

Patrick Cain
APWG

APWG
Committed to Wiping Out Internet Scams and Fraud
We Publish Statistics

New Phishing Sites by Month June '06 - June '07

<table>
<thead>
<tr>
<th>RANK</th>
<th>TLD</th>
<th>TLD Location</th>
<th># Unique Phishing attacks 2H2010</th>
<th>Unique Domain Names used for phishing 2H2010</th>
<th>Domains in registry Oct 2010</th>
<th>Score: Phish per 10,000 domains 2H2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.th</td>
<td>Thailand</td>
<td>125</td>
<td>66</td>
<td>51,433</td>
<td>12.6</td>
</tr>
<tr>
<td>2</td>
<td>.ir</td>
<td>Iran</td>
<td>204</td>
<td>169</td>
<td>175,430</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>.ma</td>
<td>Morocco</td>
<td>73</td>
<td>34</td>
<td>30,009</td>
<td>9.3</td>
</tr>
<tr>
<td>4</td>
<td>.ie</td>
<td>Ireland</td>
<td>112</td>
<td>96</td>
<td>151,023</td>
<td>6.4</td>
</tr>
<tr>
<td>5</td>
<td>.tk</td>
<td>Tokelau</td>
<td>2,899</td>
<td>2,420</td>
<td>4,030,739</td>
<td>8.3</td>
</tr>
<tr>
<td>5 (tie)</td>
<td>.ke</td>
<td>Kazakhstan</td>
<td>49</td>
<td>28</td>
<td>50,054</td>
<td>5.5</td>
</tr>
<tr>
<td>5 (tie)</td>
<td>.cc</td>
<td>Cocos Islands</td>
<td>4,093</td>
<td>55</td>
<td>100,000</td>
<td>5.5</td>
</tr>
<tr>
<td>7</td>
<td>.in</td>
<td>India</td>
<td>523</td>
<td>421</td>
<td>721,185</td>
<td>5.3</td>
</tr>
<tr>
<td>8</td>
<td>.my</td>
<td>Malaysia</td>
<td>68</td>
<td>56</td>
<td>108,211</td>
<td>5.1</td>
</tr>
<tr>
<td>9</td>
<td>.hu</td>
<td>Hungary</td>
<td>385</td>
<td>255</td>
<td>542,000</td>
<td>4.7</td>
</tr>
</tbody>
</table>
Why Publish Stats?

• To gauge how bad (or good) things are
• And, we’re not trying to sell you something
  – Vendor neutral
• We’re not trying to be alarmist
• It does allow for trending
• Can identify obvious areas for improvement

• [Everybody has a problem with them]
Phishing Terminology

• Phishing – Using social engineering to extract personal data or credentials from a victim.

• A phishing campaign is composed of:
  – Lures – A message used to entice a victim to respond.
    • “I am your bank. Give me your password.”
  – Collector - System used to collect and hold personal data and credentials
  – Credentials
    • Bank or system passwords
    • Tax numbers, birth dates, etc
  – Takedown – Disable collector
Total Number of Lures Seen

• Counting the number of (unique) lures and brands and collectors was fun…
  … for a little while 😞

• The goal was to educate banks that phishing was real
  – It worked. Then the stats lost their luster

• Now, the stats are based on domains and TLDs
  – A twice-yearly global phishing domains report is published
  – Use the stats to let registries compare themselves
  – .com & .net account for about 50% of all phish
## Attacks and Domains for 3 Years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phishing Domain Names</strong></td>
<td>-</td>
<td>47,342</td>
<td>56,959</td>
<td>55,698</td>
<td>126,697</td>
<td>48,244</td>
<td>67,677</td>
</tr>
<tr>
<td><strong>Unique campaigns</strong></td>
<td>28,818</td>
<td>26,678</td>
<td>30,454</td>
<td>30,131</td>
<td>28,775</td>
<td>28,646</td>
<td>42,624</td>
</tr>
<tr>
<td><strong>TLDs used</strong></td>
<td>145</td>
<td>155</td>
<td>170</td>
<td>171</td>
<td>173</td>
<td>177</td>
<td>183</td>
</tr>
<tr>
<td><strong>IP-based phish</strong></td>
<td>5,217</td>
<td>3,389</td>
<td>2,809</td>
<td>3,563</td>
<td>2,031</td>
<td>2,018</td>
<td>2,318</td>
</tr>
<tr>
<td><strong>Malicious reg domains</strong></td>
<td>-</td>
<td>-</td>
<td>5,561</td>
<td>4,382</td>
<td>6,372</td>
<td>4,755</td>
<td>11,769</td>
</tr>
<tr>
<td><strong>IDN domains</strong></td>
<td>10</td>
<td>52</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
## Detail from the 2H2010 Report

<table>
<thead>
<tr>
<th>Rank</th>
<th>TLD</th>
<th>TLD Location</th>
<th># Unique Phishing Attacks 2H2010</th>
<th>Unique Domain Names used for Phishing 2H2010</th>
<th>Domains in Registry 2010</th>
<th>Score: Phish per 10,000 domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.th</td>
<td>Thailand</td>
<td>125</td>
<td>65</td>
<td>51,438</td>
<td>12.6</td>
</tr>
<tr>
<td>2</td>
<td>.ir</td>
<td>Iran</td>
<td>295</td>
<td>169</td>
<td>175,600</td>
<td>9.6</td>
</tr>
<tr>
<td>3</td>
<td>.ma</td>
<td>Morocco</td>
<td>73</td>
<td>34</td>
<td>36,669</td>
<td>9.3</td>
</tr>
<tr>
<td>4</td>
<td>.ie</td>
<td>Ireland</td>
<td>112</td>
<td>96</td>
<td>151,023</td>
<td>6.4</td>
</tr>
<tr>
<td>5</td>
<td>.tk</td>
<td>Tokelau</td>
<td>2,533</td>
<td>2,429</td>
<td>4,030,709</td>
<td>6.0</td>
</tr>
<tr>
<td>6 (tie)</td>
<td>.kz</td>
<td>Kazakhstan</td>
<td>49</td>
<td>28</td>
<td>50,534</td>
<td>5.5</td>
</tr>
<tr>
<td>6 (tie)</td>
<td>.cc</td>
<td>Cocos Islands</td>
<td>4,963</td>
<td>55</td>
<td>100,000</td>
<td>5.5</td>
</tr>
<tr>
<td>7</td>
<td>.in</td>
<td>India</td>
<td>523</td>
<td>421</td>
<td>791,165</td>
<td>5.3</td>
</tr>
<tr>
<td>8</td>
<td>.my</td>
<td>Malaysia</td>
<td>68</td>
<td>55</td>
<td>108,21</td>
<td>5.1</td>
</tr>
<tr>
<td>9</td>
<td>.hu</td>
<td>Hungary</td>
<td>365</td>
<td>255</td>
<td>542,000</td>
<td>4.7</td>
</tr>
</tbody>
</table>
## Many Years as a Trend

<table>
<thead>
<tr>
<th>Year</th>
<th>1H2008</th>
<th>2H2008</th>
<th>1H2009</th>
<th>2H2009</th>
<th>1H2010</th>
<th>2H2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Hong Kong</em></td>
<td>Venezuela</td>
<td>Peru</td>
<td><em>Thailand</em></td>
<td>Thailand</td>
<td>Thailand</td>
</tr>
<tr>
<td>2</td>
<td><em>Thailand</em></td>
<td><em>Thailand</em></td>
<td><em>Thailand</em></td>
<td>Korea</td>
<td>Korea</td>
<td>Iran</td>
</tr>
<tr>
<td>3</td>
<td>Belize</td>
<td>Belize</td>
<td>Belize</td>
<td>Ireland</td>
<td>Ireland</td>
<td>Morocco</td>
</tr>
<tr>
<td>4</td>
<td>Venezuela</td>
<td><em>Soviet Union</em></td>
<td>Belgium</td>
<td>Belgium</td>
<td>Poland</td>
<td>Ireland</td>
</tr>
<tr>
<td>5</td>
<td><em>Chile</em></td>
<td>Romania</td>
<td>Romania</td>
<td>Romania</td>
<td>Chile</td>
<td>Tokelau</td>
</tr>
<tr>
<td>6</td>
<td>Romania</td>
<td><em>Chile</em></td>
<td>Taiwan</td>
<td>Malaysia</td>
<td>Malaysia</td>
<td>Korea</td>
</tr>
<tr>
<td>7</td>
<td>Liechtenstein</td>
<td>Korea</td>
<td>Korea</td>
<td>.eu</td>
<td>Greece</td>
<td>Cocos Islands</td>
</tr>
<tr>
<td>8</td>
<td>.name</td>
<td>Vietnam</td>
<td><em>Chile</em></td>
<td>Iran</td>
<td>Romania</td>
<td>India</td>
</tr>
<tr>
<td>9</td>
<td>Taiwan</td>
<td>Russia</td>
<td>Ireland</td>
<td>Poland</td>
<td>Vietnam</td>
<td>Malaysia</td>
</tr>
<tr>
<td>10</td>
<td>Korea</td>
<td>Taiwan</td>
<td>Malaysia</td>
<td>Mexico</td>
<td>Czech Rep</td>
<td>Hungary</td>
</tr>
</tbody>
</table>
Collector Site Uptimes

![Bar chart showing collector site uptimes for different periods from 1H2008 to 2H2010. The chart compares average and median values. The x-axis represents the periods, and the y-axis shows the uptimes in percentages.](chart.png)
The future of Statistics

• The numbers and pictures are nice…. 

…but what are we REALLY trying to do?
Adventures in Statistics

• One use of the stats is to convince the banks, governments, polizei, etc, that there is a problem
  – … and to calm down the media hounds

• Phishing, spam, CC fraud, etc used to be distinct
  – Now, organized crime is involved
  – Even minor groups have turned into cooperatives
  – It’s now lumped up as Electronic crime (eCrime)

• Everybody knows the numbers are increasing
  – But they’re only our numbers
  – How do we get to see a bigger picture?
The real purpose of stats… 😊

• The goal is to catch the bad guy
• How do we get countries to devote resources to eCrime?
• How do we get LEA’s attention?
  – We need justice’s attention
• How do we get Justice’s attention?
  – Define risks; education
  – Sounds like a paper.. 😊 (Has it been done before?)
What got into Pat?

• We hang out internationally
  – We try and get countries to take eCrime seriously
    • How do we get cops/gov’ts actionable?

• Lots of people use our stats as a driver for change
  – But get/give different conclusions
    are the current stats meeting the ‘mission’?
  – I wondered if we were looking at the stats ‘big picture’ wrong
A Diversion

• Interaction with the UN eCrime Commission convinced us that some organizations, companies, and member-states will never report any type of specific eCrime statistics.

• This is bad
  – Stats help countries prioritize response
  – Stats help plan response actions
  – Our stats won’t help (non-country specific) you!

• It will get worse
  – APT, night dragon, cheese slider, etc

• What’s a crime fighter to do?
Modify Our Current Stats?

- Define the risks to an organization from the internet
  - Kind of like what ISO/IEC 27032 may do
- Refine some (general) threats from those risks
- Identify threat-specific malicious behaviour

- Report stats as ‘threats and risks’ based
  - We’ll need new types of reporting
  - And more people to report things
  - Or not. Use it ‘internally’, too
So how could this be useful?

• I volunteered to lead an effort to write an “Internet Threat Assessment” to help our friends and us come up with useable stats, understand the risks, and educate justice ministries.

• This is live research; views welcome
  – ‘Live’ as in still changing
The Top-Level Risks

- Financial Loss
- Data Misuse
  - Proprietary
  - Personal
- Content Controls
  - Content Restrictions
  - Access to Prohibited Content
- Business Interference
- Loss of Network Control
- Distribution of Prohibited Speech
- Loss of Privacy
- (Reputation)
- (People/Knowledge)
Digging into the Risks/Threats

- Financial Loss
  - Fraudulent transactions
  - Improper Credential Use
  - Laundering Activities
  - Extortion
- Proprietary Data Misuse
  - Possession
  - Corruption, Deletion
  - Misuse
  - Cyber Stalking
- Personal Data Misuse
  - Possession
  - Alteration
  - Misuse/Trafficking?
  - Falsification
- (Controlling Content)
- Access to Prohibited Content
  - Illegal porn
  - Pirated artistic works
- Distribution of Prohibited Speech
  - Hate speech
  - Death threats
  - Cyber-bullying
- Business Interference
  - DOS
- Loss of Network Control
  - Network Service Unavail – (DOS)
  - Network Compromised
- Loss of Privacy
  - Data Aggregation
Down to the Details

• Map the Risks to likely attacks
  – Using CAPEC mappings (initially)
• Describe how to determine, collect, report those attacks
  – Let people do it themselves
  – Maybe convince some collusion to get area statistics
<table>
<thead>
<tr>
<th>Risk</th>
<th>Company</th>
<th>Government</th>
<th>Person</th>
<th>Alien</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Loss</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Data Misuse</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proprietary</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Controlling Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to Prohibited Content</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Restrictions</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Distribution of Prohibited Speech</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Business Interference</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss of Network Control</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Data Misuse</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Loss of Privacy</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
The Path Forward

• Flush out a document
  – Humorously called: Internet Risk Assessment
  – Why do a doc? Set the tone; define vocabulary
  – Use it as a tool to educate our ‘friends’

• Longer-term
  – Get more data (from others) into the stats
  – Provide our squishy-stats in a more general form so we track evolution.
Our overall next steps

• Run an eCrime IODEF Pilot this fall to see if this all works
  – Multi-country, multi-language, multi-grief
  – Can we report and understand set scenarios
  – See if we can collect the new types of stats

• (unrelated) Figure out how to measure eCrime
Other Event Info

• CrimeFighters want more data in our stats
  • Collect more data items
• As we slop data around, there’s more to agree on…
  – Data Sharing Restrictions
  – The attack ‘method’
  – The ‘impact’ of the attack
• LEO guidance on data to put in a report
• Watch ITU-related and other efforts
Additional Information

• Special thanks to
  – Greg Aaron of Afilias
  – Rod Rasmussen of Internet Identity

• For the Global Phishing Report

• All reports are available on
  – http://apwg.org/resources.html
Thank you

Pat Cain
Resident Research Fellow
APWG
pcain@antiphishing.org