Using Early Results from the ‘spamHINTS’ Project to Estimate an ISP Abuse Team’s Task

Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

CEAS, Mountain View
28th July 2006
ISP Email Handling

- Smarthost
- MX host

The Internet
Email Log Processing @ Demon

Detection of spam (black) and viruses (red)
## Detection Ratios (spam)

<table>
<thead>
<tr>
<th>Count</th>
<th>AS</th>
<th>Description</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>80319</td>
<td>AS4134</td>
<td>CHINANET (CN)</td>
<td>55%</td>
</tr>
<tr>
<td>75980</td>
<td>AS4766</td>
<td>Korea Telecom (KR)</td>
<td>62%</td>
</tr>
<tr>
<td>47578</td>
<td>AS4812</td>
<td>China Telecom (CN)</td>
<td>59%</td>
</tr>
<tr>
<td>18683</td>
<td>AS9318</td>
<td>Hanaro Telecom (KR)</td>
<td>53%</td>
</tr>
<tr>
<td>12609</td>
<td>AS4837</td>
<td>CNC (CN)</td>
<td>38%</td>
</tr>
<tr>
<td>5792</td>
<td>AS12322</td>
<td>Proxad (FR)</td>
<td>38%</td>
</tr>
<tr>
<td>4941</td>
<td>AS3786</td>
<td>Dacom Corporation (KR)</td>
<td>67%</td>
</tr>
<tr>
<td>4779</td>
<td>AS7738</td>
<td>TeleBahia (BR)</td>
<td>21%</td>
</tr>
<tr>
<td>3929</td>
<td>AS9277</td>
<td>Thrunet (KR)</td>
<td>62%</td>
</tr>
<tr>
<td>3911</td>
<td>AS3320</td>
<td>Deutsche Telecom (DE)</td>
<td>16%</td>
</tr>
</tbody>
</table>
Ratios

• Significantly better than previously reported
  – was circa 1% at time of CEAS 2005
  – now 25% - 50%
  – Main reason is better heuristics (esp HELOs)

• Applying to Demon… we are detecting 20 customers a day, so maybe 40-80 are actually infected at maximum
Suppose All Email is Bad 😞

• Typically, remote ASs are sending email from 0.25% … 1.25% of their address space
• Applying this ratio to Demon address space (~200K customers) means 500 to 2500 customers will have a problem and so far we have detected almost none of them 😞
LINX samples 1 in 2000 packets (using sFlow) and makes the port 25 traffic available for analysis…
sFlow Results : Demon Customers

8639 servers

6927 clients

3847 on "AOL whitelist"
Conclusions

• Log processing spots ~20 problems a day
• Remote data suggests 40–80 Demon customers actually have a problem
• Worst case analysis on customer counts suggests 500–2500 with a problem
• sFlow data suggests 3500+ problems
  – but much more work is needed!
Using Early Results from the ‘spamHINTS’ Project to Estimate an ISP Abuse Team’s Task

Richard Clayton

http://www.cl.cam.ac.uk/~rnc1/earlyResult.pdf