Analysis of Web-based Malware

Keio University (Japan)
Masayoshi MIZUTANI (mizutani@sfc.wide.ad.jp)

WIDE Project (http://www.wide.ad.jp)
Internet Research Lab. (http://www.sfc.wide.ad.jp)
Internet Security Center (http://isc.sfc.wide.ad.jp)

2009/10/6 Internet2
Summary

• Recently web-based malware have been increasing[1]
  • drive-by-download attack malware
  • Traditional malware: sending exploit code by itself (e.g. NetBIOS)
• Traffic data analysis by executing captured malware
  • Communication methods
  • Comparison on traffic volume
  • Comparison on types of TCP packets
  • Destination IP Addresses
  • HTTP Request of passive attack
• Comparison between traditional and web-based malware
  • from viewpoint of network security monitor

Analysis Environment

• Dynamic analysis on virtual machines
  • Virtual Machine: VirtualBox 2.0
  • Guest OS: Windows XP (Version 5.1 Build 2600)
  • Internet Explorer (Version: 6.0.2600.0000)
• Capturing traffic data generated by executing malware
  • Traditional malware
  • Web-based malware (by [2])

Data set

• $D_p$: Traffic data of web-based malware from the analysis environment
• $D_a$: Traffic data of traditional malware from the analysis environment
• $D_c$: Traffic data of traditional malware from Cyber Clean Center (Telecom ISAC in Japan http://www.ccc.go.jp/)
• $D_m$: Normal traffic data from MAWI traffic archive (http://tracer.csl.sony.co.jp/mawi/)

<table>
<thead>
<tr>
<th></th>
<th>Date</th>
<th>Duration</th>
<th>Number of data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_p$</td>
<td>2008.7-2008.11</td>
<td>20 min</td>
<td>255</td>
<td>Web-based malware</td>
</tr>
<tr>
<td>$D_a$</td>
<td>2008.5-2008.10</td>
<td>60 min</td>
<td>484</td>
<td>Traditional malware</td>
</tr>
<tr>
<td>$D_c$</td>
<td>2008.4</td>
<td>–</td>
<td>144</td>
<td>CCC DATASET 2008</td>
</tr>
<tr>
<td>$D_m$</td>
<td>2009.1</td>
<td>15 min</td>
<td>250</td>
<td>MAWI Traffic Archive</td>
</tr>
</tbody>
</table>
### Analysis (1/5)

**Communication methods**

- **Web-based malware**
  - Majorities are HTTP traffics
  - Nor IRC and attacking traffic

- **Traditional malware**
  - Majorities IRC traffic
  - Attacking other hosts

![Diagram showing communication methods and protocol of downloading malware.](image)
Analysis (2/5)

Comparison of traffic volume

Red: (Dp) Web-based malware (drive-by-download attack)
Blue: (Da) Traditional malware
Comparison of TCP packet type

- **Red**: (Da) Web-based malware (drive-by-download attack)
- **Blue**: (Dp) Traditional malware
- **Green**: (Dm) MAWI backbone traffic archive (2009/1/8 14:00-14:15)

Over 80% web-based malware communicate with under 10 IP addresses

About 60% normal host communicate with over 10 IP addresses
HTTP Request of web-based malware

- Most of HTTP Request is similar to requests generated by general website

**Request Path that are appeared on HTTP Communication (Top 5)**

<table>
<thead>
<tr>
<th>PATH of HTTP requests</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>86</td>
<td>33.72%</td>
</tr>
<tr>
<td>/cgi-bin/index.cgi</td>
<td>61</td>
<td>23.92%</td>
</tr>
<tr>
<td>/forum.php</td>
<td>47</td>
<td>18.43%</td>
</tr>
<tr>
<td>/in.cgi</td>
<td>29</td>
<td>11.37%</td>
</tr>
<tr>
<td>/soft3/common/16.gif</td>
<td>22</td>
<td>8.63%</td>
</tr>
</tbody>
</table>

**Request Variable Name that are appeared on HTTP GET Request (Top 5)**

<table>
<thead>
<tr>
<th>Variable names</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>109</td>
<td>42.75%</td>
</tr>
<tr>
<td>v</td>
<td>39</td>
<td>15.29%</td>
</tr>
<tr>
<td>t</td>
<td>37</td>
<td>14.51%</td>
</tr>
<tr>
<td>ad</td>
<td>35</td>
<td>13.73%</td>
</tr>
<tr>
<td>ver</td>
<td>28</td>
<td>10.98%</td>
</tr>
</tbody>
</table>
Conclusion

• Conclusion
  • There are differences between traditional/web-based malware
  • Web-based malware is likely generic user’s actions
  • Obsoleted traditional intrusion detection system

• Discussion
  • Monitoring methods by state transition and correlation analysis (e.g. BotHunter[3]) are efficient
  • We have developed detection method based-on the results
    • Behavior rule based intrusion detection system[4]

Thank you for your attention

• Acknowledgement
  • Prof. Dr. Keiji Takeda (Keio university)
  • Prof. Dr. Osamu Nakamura (Keio university)
  • Prof. Dr. Jun Murai (Keio university)

• More reference:
  • ROOK (implementation): http://rook.sourceforge.net