Flow Visualization Using MS-Excel
Visualization for the Common Man

Presented by Lee Rock and Jay Brown
US-CERT Analysts
Einstein Program
Background

- US-CERT Mission
- Einstein Program
  > Large volumes of traffic
  > Architecture limitations
- Proactive vs. Reactive analysis
- Slow application certification process
Pro’s and Con’s

• Pro’s:
  – Visualization allows for rapid analysis
  – Patterns are easy to identify
  – Flexibility in analysis
  – Most enterprises have MS Office (Excel)

• Con’s:
  – Excel plotting engine is limited
  – Max of 65K records (recommend <= 50K)
  – Data must be imported and formatted
  – Memory management is an issue
Data Preparation Steps

- Data Pull
- Data Reduction
- Importing Data
- Data Formatting
- Sample analysis slides
Data Pull

Analysts have several options when trying to pull interesting datasets. Several methods we find useful are:

- Collecting data during non-business hours
  - Reduces traffic from users; helps expose automated sessions
- Search for outbound traffic only
  - Reduces noise from scanning, etc.
- Filtering for packets with the PSH/ACK flags set in the initial flags field
  - Focuses the traffic on sessions where data is actually transferred
- Filtering for packets with the SYN flag set in the initial flags field
  - Focuses on sessions initiated by your organization
- Limit traffic to records under 5K bytes
  - Most cyclical sessions (beaconing) happen in this range

Traffic should be refined to provide the best possible dataset for analysts to work with.
Data Reduction

To further enhance the concentration of suspicious data, analysts should:

• Remove replies from servers (responses to inbound server requests)
  – Looking for genuine outbound traffic
• Remove loud, common talkers (instant messenger, web crawlers, etc)
  – Reduces the noise, especially in web traffic
• “Whitelists” and “blacklists” are helpful for filtering

This is an iterative approach – Analyze, Research, Remove.
## Importing Data

Data is imported from a pipe delimited text file

<table>
<thead>
<tr>
<th>sIP</th>
<th>dIP</th>
<th>sPort</th>
<th>dPort</th>
<th>proto</th>
<th>packets</th>
<th>bytes</th>
<th>flags</th>
<th>sTime</th>
<th>dur</th>
<th>eT</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.147.82.96</td>
<td>10.130.166.158</td>
<td>80</td>
<td>5515</td>
<td></td>
<td>11</td>
<td>8258</td>
<td>FS FA</td>
<td></td>
<td>0.290</td>
<td>2007/10/29 15:07:39</td>
</tr>
<tr>
<td>10.150.165.218</td>
<td>10.54.98.176</td>
<td>80</td>
<td>5705</td>
<td></td>
<td>86</td>
<td>12032</td>
<td>FS FA</td>
<td></td>
<td>0.552</td>
<td>2007/10/29 15:07:41</td>
</tr>
<tr>
<td>10.95.46.146</td>
<td>10.34.134.191</td>
<td>80</td>
<td>5705</td>
<td></td>
<td>1</td>
<td>40</td>
<td>A</td>
<td></td>
<td>0.000</td>
<td>2007/10/29 15:07:42</td>
</tr>
<tr>
<td>10.94.132.147</td>
<td>10.168.141.231</td>
<td>80</td>
<td>49094</td>
<td></td>
<td>10</td>
<td>7348</td>
<td>FS FA</td>
<td></td>
<td>0.825</td>
<td>2007/10/29 15:16:20</td>
</tr>
<tr>
<td>10.226.143.219</td>
<td>10.162.254.83</td>
<td>80</td>
<td>49297</td>
<td></td>
<td>33</td>
<td>43588</td>
<td>FS FA</td>
<td></td>
<td>0.490</td>
<td>2007/10/29 15:16:23</td>
</tr>
<tr>
<td>172.25.4.165</td>
<td>10.161.142.75</td>
<td>80</td>
<td>47356</td>
<td></td>
<td>10</td>
<td>7047</td>
<td>FS FA</td>
<td></td>
<td>0.859</td>
<td>2007/10/29 15:21:42</td>
</tr>
<tr>
<td>10.120.9.241</td>
<td>10.36.140.83</td>
<td>80</td>
<td>47409</td>
<td></td>
<td>7</td>
<td>1839</td>
<td>FS FA</td>
<td></td>
<td>0.326</td>
<td>2007/10/29 15:23:28</td>
</tr>
<tr>
<td>10.8.58.141</td>
<td>10.254.147.27</td>
<td>5705</td>
<td>80</td>
<td></td>
<td>55</td>
<td>3589</td>
<td>FS FA</td>
<td></td>
<td>1.591</td>
<td>2007/10/29 15:23:42</td>
</tr>
<tr>
<td>10.215.49.170</td>
<td>10.30.5.168</td>
<td>49094</td>
<td>80</td>
<td></td>
<td>8</td>
<td>969</td>
<td>FS FA</td>
<td></td>
<td>0.825</td>
<td>2007/10/29 15:23:50</td>
</tr>
<tr>
<td>10.207.158.173</td>
<td>10.15.150.60</td>
<td>49297</td>
<td>80</td>
<td></td>
<td>22</td>
<td>1666</td>
<td>FS FA</td>
<td></td>
<td>0.480</td>
<td>2007/10/29 15:23:56</td>
</tr>
<tr>
<td>192.168.45.69</td>
<td>10.88.159.210</td>
<td>47356</td>
<td>80</td>
<td></td>
<td>10</td>
<td>1796</td>
<td>FS FA</td>
<td></td>
<td>0.874</td>
<td>2007/10/29 15:24:22</td>
</tr>
<tr>
<td>10.115.234.230</td>
<td>10.144.241.122</td>
<td>24503</td>
<td>6</td>
<td></td>
<td>11</td>
<td>6735</td>
<td>FS FA</td>
<td></td>
<td>0.220</td>
<td>2007/10/29 16:01:28</td>
</tr>
<tr>
<td>10.52.224.171</td>
<td>10.532.170.176</td>
<td>24601</td>
<td>6</td>
<td></td>
<td>10</td>
<td>1475</td>
<td>FS FA</td>
<td></td>
<td>0.236</td>
<td>2007/10/29 16:01:29</td>
</tr>
<tr>
<td>10.144.199.78</td>
<td>10.208.138.229</td>
<td>64021</td>
<td>6</td>
<td></td>
<td>10</td>
<td>6437</td>
<td>FS FA</td>
<td></td>
<td>0.161</td>
<td>2007/10/29 16:09:08</td>
</tr>
<tr>
<td>10.9.152.19</td>
<td>10.233.152.178</td>
<td>64124</td>
<td>6</td>
<td></td>
<td>7</td>
<td>1310</td>
<td>FS FA</td>
<td></td>
<td>0.247</td>
<td>2007/10/29 16:09:10</td>
</tr>
<tr>
<td>10.116.235.116</td>
<td>10.127.192.234</td>
<td>64021</td>
<td>6</td>
<td></td>
<td>7</td>
<td>40</td>
<td>A</td>
<td></td>
<td>0.000</td>
<td>2007/10/29 16:09:09</td>
</tr>
<tr>
<td>10.0.158.212</td>
<td>10.131.10.198</td>
<td>40079</td>
<td>9</td>
<td></td>
<td>11</td>
<td>6378</td>
<td>FS FA</td>
<td></td>
<td>0.247</td>
<td>2007/10/29 16:40:58</td>
</tr>
<tr>
<td>10.40.145.167</td>
<td>10.229.195.82</td>
<td>40167</td>
<td>9</td>
<td></td>
<td>15</td>
<td>15095</td>
<td>FS FA</td>
<td></td>
<td>0.317</td>
<td>2007/10/29 16:41:78</td>
</tr>
<tr>
<td>10.33.232.60</td>
<td>10.224.241.122</td>
<td>83835</td>
<td>9</td>
<td></td>
<td>21</td>
<td>21877</td>
<td>FS FA</td>
<td></td>
<td>0.409</td>
<td>2007/10/29 16:42:34</td>
</tr>
<tr>
<td>10.76.204.191</td>
<td>10.52.57.127</td>
<td>2403</td>
<td>80</td>
<td></td>
<td>6</td>
<td>1439</td>
<td>FS FA</td>
<td></td>
<td>0.232</td>
<td>2007/10/29 16:42:38</td>
</tr>
<tr>
<td>10.6.3.30</td>
<td>10.218.84.41</td>
<td>24601</td>
<td>80</td>
<td></td>
<td>7</td>
<td>808</td>
<td>FS FA</td>
<td></td>
<td>0.225</td>
<td>2007/10/29 16:42:39</td>
</tr>
<tr>
<td>10.123.207.187</td>
<td>10.211.245.126</td>
<td>69021</td>
<td>80</td>
<td></td>
<td>8</td>
<td>2544</td>
<td>FS FA</td>
<td></td>
<td>0.153</td>
<td>2007/10/29 16:09:08</td>
</tr>
<tr>
<td>10.239.204.27</td>
<td>10.140.242.83</td>
<td>54021</td>
<td>80</td>
<td></td>
<td>7</td>
<td>40</td>
<td>R</td>
<td></td>
<td>0.000</td>
<td>2007/10/29 16:09:08</td>
</tr>
</tbody>
</table>
Columns within the spreadsheet should be aligned to each field of the flows, Einstein data is formatted to encompass:

- Source IP
- Destination IP
- Source Port
- Destination Port
- Protocol
- Packets
- Bytes
- Flags
- Start Time
- Duration
- End Time
- Sensor
- Type
- Initial Flags

<table>
<thead>
<tr>
<th></th>
<th>dIP</th>
<th>dIP</th>
<th>sPort</th>
<th>dPort</th>
<th>protocol</th>
<th>packets</th>
<th>bytes</th>
<th>flags</th>
<th>sTime</th>
<th>dur</th>
<th>eTime</th>
<th>sensor</th>
<th>type</th>
<th>initialFlag</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>52611</td>
<td>22</td>
<td>6</td>
<td>2569</td>
<td>3385741</td>
<td>FS PA</td>
<td>2007/10/17</td>
<td>10:40:722</td>
<td>9:5</td>
<td>2007/10/17</td>
<td>10:50:222</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>7774</td>
<td>22</td>
<td>6</td>
<td>136</td>
<td>10750</td>
<td>PA</td>
<td>2007/10/17</td>
<td>08:28:293</td>
<td>1795.961</td>
<td>2007/10/17</td>
<td>08:23:964</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>7774</td>
<td>22</td>
<td>6</td>
<td>106</td>
<td>9046</td>
<td>PA</td>
<td>2007/10/17</td>
<td>08:36:714</td>
<td>1600.05</td>
<td>2007/10/17</td>
<td>08:36:764</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>01:18:787</td>
<td>0</td>
<td>2007/10/17</td>
<td>01:18:787</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>06:18:690</td>
<td>0</td>
<td>2007/10/17</td>
<td>06:18:690</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>11:18:690</td>
<td>0</td>
<td>2007/10/17</td>
<td>11:18:690</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>16:18:514</td>
<td>0</td>
<td>2007/10/17</td>
<td>16:18:514</td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>21:18:433</td>
<td>0</td>
<td>2007/10/17</td>
<td>21:18:433</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>26:18:349</td>
<td>0</td>
<td>2007/10/17</td>
<td>26:18:349</td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>31:18:257</td>
<td>0</td>
<td>2007/10/17</td>
<td>31:18:257</td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>36:18:164</td>
<td>0</td>
<td>2007/10/17</td>
<td>36:18:164</td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1321</td>
<td>22</td>
<td>6</td>
<td>956</td>
<td>40630</td>
<td>PA</td>
<td>2007/10/17</td>
<td>09:32:261</td>
<td>1798.994</td>
<td>2007/10/17</td>
<td>09:31:275</td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>41:18:088</td>
<td>0</td>
<td>2007/10/17</td>
<td>41:18:088</td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>x.x.x.x</td>
<td>Y.Y.Y.Y</td>
<td>1688</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>41</td>
<td>A</td>
<td>2007/10/17</td>
<td>46:17:971</td>
<td>0</td>
<td>2007/10/17</td>
<td>46:17:971</td>
<td>X</td>
</tr>
</tbody>
</table>
US-CERT analysts use two methods to format the Einstein time fields into a format that is able to be plotted:

A: Use the `- - legacy-timestamps` switch to place the time in a `MM/DD/YYYY HH:MM:SS` format from the default `MM/DD/YYYYTHH:MM:SS.MMM`

B: Utilize the replace function in excel to remove the milliseconds from the time and replace the `T` placeholder with a space:
Plot

Creating charts from the selected data, allows for quick pattern identification.
You can “zoom” in to specific data points, by changing the scale of the axis

- Right click on the axis
- Select “Format Axis”
- Click on the “Scale” tab
- Adjust scale as desired
- Works for both axis
- Remember to remove
Highlight

By hovering over a data point in the series an analyst can locate the point in the rest of the records by filtering for the displayed information.
AutoFilter

Method A – Drop down list:
Select the desired value from the drop down list

Method B – Custom Filter:
Select data by using Excel’s built in boolean logic search functions
Sample Analysis Slides

- Scatter Plot Analysis
  - Byte Based Patterns
  - Duration Based Patterns
  - sPort vs. dPort Patterns
  - IP Based Patterns
  - Application Pattern
Byte Based Patterns
Duration Based Patterns
IP Integer Patterns


ARIN
Comprehensive View
Case Study
Multi-day View
Case Study Conclusion

After notifying the agency in question, the machines that were generating this traffic were found and forensically examined. The malware turned out to be a keystroke logger that posted data to a specific website and retrieved commands embedded on the same site. Prior to this incident, there was no malware associated with this site.
Additional Analysis

Determining application patterns
  – Identifying specific applications

Working with gateway traffic
  – Structured gateway
  – Proxy gateway
  – Gateway mannerisms
Gateway Mannerisms
Future Directions

• Split view analysis
• Coloring data
• Application coloring
• sPort colored by app
• Gateway coloring to IP
Coloring Example
Green = HTTP, Dark Green = HTTPS, Blue = DNS, Red = Other
Application Coloring

Green = HTTP, Blue = DNS, Red = Other
Color sPort vs Application
Contact Info

• **Technical comments or questions**
  – US-CERT Security Operations Center
  – Email: soc@us-cert.gov
  – Phone: +1 888-282-0870

• **Media inquiries**
  – US-CERT Public Affairs
  – Email: media@us-cert.gov
  – Phone: +1 202-282-8010

• **General questions or suggestions**
  – US-CERT Information Request
  – Email: info@us-cert.gov
  – Phone: +1 703-235-5111

• **For more information, visit** [http://www.us-cert.gov](http://www.us-cert.gov)
Questions?