Security Visualizations: A Case Study

Carrie Gates
cgates@cert.org
Overview

1. Our first foray into visualization
   • Along with some striking results! 😊
2. Which led into graphing data …
   • Along with some not-so-striking results 😞
3. Which resulted in some lessons about the usefulness of visualizing data for security
What were we doing?

Trying to detect scans

A scan is a reconnaissance technique aimed at *multiple* targets.
Hypothesis

There is an obvious delineation between normal activity and scans when examining the number of destination IP addresses contacted per source IP.

**Scans:** many destinations contacted

**Normal:** few destinations contacted

(e.g. Williamson in 2003)
What did we expect?

- Two clusters: one with only a few destinations and lots of sources, and one with lots of destinations and only a few sources
The results...
A step back

We should note at this point that:

1. This network contains > 16 million hosts

2. It has multiple border routers that are geographically dispersed as well as multiple administrative domains

3. The subnets are also dispersed in IP space

So this represents something fairly large, that might be observed on the other networks.
Was this normal?

Still present, 7 months later.
Except....

Disappears abruptly on August 11, 2003

This corresponds with the release of Blaster
Random sources?

- 3 particular /8’s
- SYN only flows to port 80
- Unique src-dest pairs
- But targets not randomly distributed
- Mostly 1 specific /8 targeted
- Lots to particular /16s inside
Just to reappear…

Reappeared slowly from 9-13 February 2004

Disappeared abruptly on June 1, 2004

SYN-only flows to port 80 (10 times normal volume!)
Welchia/Nachi

So how can we use this?

- Can see large scale events
- Good for forensics but not quick recognition / response
End result?

Started looking at ways to visualize information

Two main events:
1. Web portal for scan db
2. SC|Net
Eye Candy

Scan Report for Aug 17, 2005

World Scans

Top ten source countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Scans</th>
</tr>
</thead>
<tbody>
<tr>
<td>cn</td>
<td>9669</td>
</tr>
<tr>
<td>us</td>
<td>8129</td>
</tr>
<tr>
<td>De</td>
<td>6157</td>
</tr>
<tr>
<td>de</td>
<td>3346</td>
</tr>
<tr>
<td>fr</td>
<td>3275</td>
</tr>
<tr>
<td>kr</td>
<td>2781</td>
</tr>
<tr>
<td>tw</td>
<td>2731</td>
</tr>
<tr>
<td>es</td>
<td>1626</td>
</tr>
<tr>
<td>gb</td>
<td>784</td>
</tr>
<tr>
<td>pl</td>
<td>771</td>
</tr>
</tbody>
</table>

Top ten scanned ports

<table>
<thead>
<tr>
<th>Port</th>
<th>Scans</th>
</tr>
</thead>
<tbody>
<tr>
<td>1025</td>
<td>11156</td>
</tr>
<tr>
<td>80</td>
<td>10949</td>
</tr>
<tr>
<td>6129</td>
<td>5415</td>
</tr>
<tr>
<td>4662</td>
<td>4989</td>
</tr>
<tr>
<td>4899</td>
<td>4575</td>
</tr>
<tr>
<td>25</td>
<td>4444</td>
</tr>
<tr>
<td>5900</td>
<td>3716</td>
</tr>
<tr>
<td>3410</td>
<td>3605</td>
</tr>
<tr>
<td>5000</td>
<td>3294</td>
</tr>
<tr>
<td>62452</td>
<td>2394</td>
</tr>
</tbody>
</table>
Trending of Scan Info

We would like to animate this to show changes over time.
Frustrations

1. Can’t tell why something happened (e.g., spike in scans in late August)

2. Some graphs don’t really tell you anything useful/new (e.g., eye candy graph)

3. Not really sure what to graph!
   • Need something that provides information
ScNet (Supercomputing 2005)

Helped perform operational security

Used web portal (Zero), which can have new graphs added easily and taps into R (stats package) for graphs.
Connection Information

Most useful to view (e.g., internal scanning could indicate worms)
### More useful visualization

<table>
<thead>
<tr>
<th>Time</th>
<th>IP</th>
<th>Hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-09-22 14:00:00</td>
<td>10.1.76.163</td>
<td>122304</td>
</tr>
<tr>
<td>2005-09-22 15:00:00</td>
<td>10.1.76.163</td>
<td>111394</td>
</tr>
<tr>
<td>2005-09-22 16:00:00</td>
<td>10.1.76.163</td>
<td>38915</td>
</tr>
<tr>
<td>2005-09-22 13:00:00</td>
<td>10.1.76.163</td>
<td>24184</td>
</tr>
<tr>
<td>2005-09-22 13:00:00</td>
<td>10.1.76.204</td>
<td>84</td>
</tr>
<tr>
<td>2005-09-22 14:00:00</td>
<td>10.1.76.204</td>
<td>79</td>
</tr>
<tr>
<td>2005-09-22 13:00:00</td>
<td>10.1.76.190</td>
<td>70</td>
</tr>
<tr>
<td>2005-09-22 13:00:00</td>
<td>10.1.76.71</td>
<td>57</td>
</tr>
</tbody>
</table>

Especially if we have lots of /24 subnets!
Tables are not always best

Blue dot: center
Green dot: outlier
Lessons Learned

1. Visualizations are useful for:
   - Trending (e.g., time series)
   - Places where outliers are not obvious
   - Observing clusters
   - Large-scale changes in baseline (e.g., contact surface)

2. Visualizations are **not** useful for:
   - Places where you just want a min/max value - use a sorted table of values!
   - When there is a *LOT* of data - some events occluded
Lessons Learned

3. Most important part: figuring out what to visualize in the first place and why visualization is the best approach to solving a particular problem (versus, for example, an automated detection system, or a table!)

4. The value of eye candy
Thanks!

• John Prevost (Zero, eye candy)
• Josh McNutt (clustering graph)
• Brian Trammell (Sc|Net graph and table)
• SiLK development team
• John McHugh
• Bill Yurcik