A Different Approach to Logs

Vlad Grigorescu
Carnegie Mellon University
Information Security Office
> finger vladg@cmu.edu

Senior Information Security Engineer
Networking Monitoring/IDS Design

github.com/grigorescu

@0f010d

speakerdeck.com/vladg
System and service logs are designed for developers, engineers, and admins.

While they’re not designed for security, we rely heavily on them because many times they’re the best that we can do.

speakerdeck.com/vladg
Example: DNS

Feb 25 12:50:34.110 queries: info: client 10.0.0.3#1035: query: 22.example.com IN A -

- Was it recursive? Any other flags set?
- What was the response? Where did the response come from?
- Are malformed requests logged?
- What if someone doesn’t use your DNS servers?
If We Need to Analyze a Protocol...

- Let’s analyze all instances of it.
- Let’s log anything we might need to know in the future
  ...without logging too much.
- Take advantage of our tools understanding a protocol.
Our Solution:

Bro
How We Use Bro

- Inspects all traffic between VLANs
- Analyzes the protocols it sees to create forensically-sound logs.
- Automatically blocks IPs, notifies users
- Integrates with CIF
- Generates inventory data (certificates, services) and summary statistics.
<table>
<thead>
<tr>
<th><strong>Timestamp</strong></th>
<th>Fri Nov 13 12:25:22.211 EDT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UID</strong></td>
<td>lluYKisMFvh</td>
</tr>
<tr>
<td><strong>Originator</strong></td>
<td>192.168.1.2:64122</td>
</tr>
<tr>
<td><strong>Responder</strong></td>
<td>4.2.2.2:53</td>
</tr>
<tr>
<td><strong>Protocol</strong></td>
<td>UDP</td>
</tr>
<tr>
<td><strong>Transaction ID</strong></td>
<td>61551</td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td>addons.mozilla.org</td>
</tr>
<tr>
<td><strong>Query Class</strong></td>
<td>CINTERNET</td>
</tr>
<tr>
<td><strong>Query Type</strong></td>
<td>A</td>
</tr>
<tr>
<td><strong>Response Code</strong></td>
<td>NOERROR</td>
</tr>
<tr>
<td><strong>Authoritative Answer</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Truncation</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Recursion Desired</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Recursion Available</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Answers</strong></td>
<td>[amo.glb.mozilla.net, 63.245.209.91]</td>
</tr>
<tr>
<td><strong>TTLs</strong></td>
<td>[1, 20]</td>
</tr>
</tbody>
</table>
Bro-Supported Protocols

• ~37 protocols:
  • HTTP
  • SMTP
  • DHCP
  • SSL
  • SSH
  • SIP
  • Modbus

• Coming soon!
  • RADIUS
  • MySQL
  • Netflow v9
  • Win32 PE
Other Log Types

- Bro reads syslog!
- Input framework
  - Files
  - Plugin-based
Results

17,000 users, 2 Gbps to the Internet
Results

17,000 users, 2 Gbps to the Internet

- Average of 12,000 logs/second
- Peaks of 50,000 logs/second
- 1 billion logs/day

...now what?
Apache Lucene

- Full text indexing and searching
- Supports complex queries
- Actively developed from 1999
- Used by: Akamai, Apple, Comcast, IBM, LinkedIn.
### Syntax Cheatsheet

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(host:&quot;google.com&quot; OR host:&quot;yahoo.com&quot;) AND id.resp_p:80</td>
<td>Matches google.com or yahoo.com on port 80</td>
</tr>
<tr>
<td>host:&quot;ns?.*.cn&quot;</td>
<td>Matches ns1.mail.cn, ns2.mail.cn, but not ns01.mail.cn</td>
</tr>
<tr>
<td>subject:&quot;viagra~&quot;</td>
<td>Fuzzy search: vi4gra, viagr4, etc.</td>
</tr>
<tr>
<td>subject:&quot;buy viagra&quot;~5</td>
<td>Distance search: buy and viagra, no more than 5 words away.</td>
</tr>
<tr>
<td>ts:134066078821 TO * AND from:[<a href="mailto:alice@cmu.edu">alice@cmu.edu</a> TO <a href="mailto:bob@cmu.edu">bob@cmu.edu</a>]</td>
<td>Search from a given timestamp to now, where the from address is alphabetically between alice and bob.</td>
</tr>
</tbody>
</table>
ElasticSearch

- “Wraps” Lucene
- Distributed, highly available
- Per-operation persistence
- RESTful API, JSON
- Native Bro support
Lucene is Fast...

- A single Dell R720 server
- 3.8 billion logs
- 272 Lucene indexes
Lucene is Fast...

<table>
<thead>
<tr>
<th>Query</th>
<th>ms</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>method:SUPERGET</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>method:PUT</td>
<td>469</td>
<td>83,739</td>
</tr>
<tr>
<td>query:.pk AND qtype_name:AAAA</td>
<td>66</td>
<td>6,150</td>
</tr>
</tbody>
</table>

3.8 billion logs
Introducing Brownian

- An interactive web interface to Bro logs in ElasticSearch.
- Adds context by understanding Bro logs.
- Designed to quickly eliminate noise.
- Use the displayed data to help you build advanced queries.
- Leverage plugins to query other data sources, or to streamline the IR workflow.
http://brownian.bro.org

DNS Lookup: support.dell.com is an alias for support.ins.dell.com.
support.ins.dell.com has address 143.166.224.210

DNS Lookup: 210.224.166.143.in-addr.arpa domain name pointer s3b-support.us.dell.com.

<table>
<thead>
<tr>
<th>ts</th>
<th>uid</th>
<th>id.orig_h</th>
<th>id.orig_p</th>
<th>id.resp_h</th>
<th>id.resp_p</th>
<th>trans_depth</th>
<th>method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fri Nov 20 13:31:55.891</td>
<td>jo0dndG6Sve</td>
<td>192.168.1.106</td>
<td>1052</td>
<td>143.166.224.210</td>
<td>80</td>
<td>9</td>
<td>POST</td>
</tr>
</tbody>
</table>
Questions?

- bro.org  (brownian.bro.org)
- elasticsearch.org
- github.com/grigorescu/Brownian
- speakerdeck.com/vladg