REAL-TIME THREAT HUNTING

TIM CROTHERS
<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Source</th>
<th>Destination</th>
<th>Protocol</th>
<th>Length</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000000</td>
<td>10.3.14.102</td>
<td>10.3.14.1</td>
<td>DNS</td>
<td>80</td>
<td>Standard query 0x246b A totyaaclumatic.top</td>
</tr>
<tr>
<td>2</td>
<td>0.148493</td>
<td>10.3.14.1</td>
<td>10.3.14.102</td>
<td>DNS</td>
<td>96</td>
<td>Standard query response 0x246b A totyaaclumatic.top A 104.199.9.203</td>
</tr>
<tr>
<td>3</td>
<td>0.156603</td>
<td>10.3.14.102</td>
<td>104.199.9.203</td>
<td>TCP</td>
<td>66</td>
<td>49158 - 80 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=256 SACK_PERM=1</td>
</tr>
<tr>
<td>4</td>
<td>0.585613</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>60</td>
<td>0 - 49158 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460</td>
</tr>
<tr>
<td>5</td>
<td>0.585709</td>
<td>10.3.14.102</td>
<td>104.199.9.203</td>
<td>TCP</td>
<td>60</td>
<td>49158 - 80 [ACK] Seq=1 Ack=1 Win=64240 Len=0</td>
</tr>
<tr>
<td>6</td>
<td>0.587433</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>60</td>
<td>0 - 49158 [ACK] Seq=1 Ack=1 Win=64240 Len=0</td>
</tr>
<tr>
<td>7</td>
<td>0.587483</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>60</td>
<td>0 - 49158 [ACK] Seq=1 Ack=1 Win=64240 Len=0</td>
</tr>
<tr>
<td>8</td>
<td>2.427121</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>1514</td>
<td>[TCP segment of a reassembled PDU]</td>
</tr>
<tr>
<td>9</td>
<td>2.427141</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>1410</td>
<td>[TCP segment of a reassembled PDU]</td>
</tr>
<tr>
<td>10</td>
<td>2.427473</td>
<td>10.3.14.102</td>
<td>104.199.9.203</td>
<td>TCP</td>
<td>60</td>
<td>49158 - 80 [ACK] Seq=1 Ack=2817 Win=64240 Len=0</td>
</tr>
<tr>
<td>11</td>
<td>2.427481</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>1514</td>
<td>[TCP segment of a reassembled PDU]</td>
</tr>
<tr>
<td>12</td>
<td>2.427484</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>1514</td>
<td>[TCP segment of a reassembled PDU]</td>
</tr>
<tr>
<td>13</td>
<td>2.427486</td>
<td>104.199.9.203</td>
<td>10.3.14.102</td>
<td>TCP</td>
<td>1514</td>
<td>[TCP segment of a reassembled PDU]</td>
</tr>
</tbody>
</table>

Frame 6: 134 bytes on wire (1072 bits), 134 bytes captured (1072 bits)
Ethernet II, Src: HewlettP 1c:47:ae (00:00:00:1c:47:ae), Dst: Netgear b6:93:f1 (20:e5:2a:b6:93:f1)
Internet Protocol Version 4, Src: 10.3.14.102, Dst: 104.199.9.203
Transmission Control Protocol, Src Port: 49158, Dst Port: 80, Seq: 1, Ack: 1, Len: 80

HTTP Transfer Protocol
GET /search.php HTTP/1.1\r\nHost: totyaaclumatic.top\r\nConnection: Keep-Alive\r\n\r\n[Full request URI: http://totyaaclumatic.top/search.php]
[HTTP request 1/1]
[Response in frame: 239]
ANYONE RECOGNIZE THIS?
ANYONE RECOGNIZE THIS?
• Too few experienced practitioners
• Takes too long to develop experienced practitioners
• Too much data to look through
• Hunts are periodic
CYBER HUNTING BENEFITS

• Find unknown malicious activity
• Program to drive detection improvement
• Fantastic mentoring vehicle
MACHINE LEARNING TO THE RESCUE...
ENOUGH TALK... DEMO TIME!
SO WHAT DID YOU JUST SEE?

• Python script using a trained Naïve Bayes algorithm based model against 37,440 HTTP headers
• Found 46 things that looked suspicious
• 0.12% suspicious
### SUSPICIOUS ENTRIES

<table>
<thead>
<tr>
<th>Line</th>
<th>Looks Suspicious</th>
<th>X-Requested-With</th>
<th>Cookie</th>
<th>Connection</th>
<th>Keep-Alive</th>
<th>DNT</th>
<th>If-Modified-Since</th>
<th>Host</th>
<th>Referer</th>
<th>Accept-Language</th>
<th>Accept-Charset</th>
<th>Accept-Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>371</td>
<td>Looks Suspicious</td>
<td>X-Requested-With</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Host</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
</tr>
<tr>
<td>8074</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>16100</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>16191</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>16762</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>16976</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17050</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17113</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17121</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17203</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17344</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17347</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17427</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17442</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17456</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17617</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>17745</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18073</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18190</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18344</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18521</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18606</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18700</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>18968</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>19055</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>20632</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>24155</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
<tr>
<td>24699</td>
<td>Looks Suspicious</td>
<td>Cookie</td>
<td>Connection</td>
<td>Keep-Alive</td>
<td>DNT</td>
<td>If-Modified-Since</td>
<td>Wed</td>
<td>Referer</td>
<td>Accept-Language</td>
<td>Accept-Charset</td>
<td>Accept-Encoding</td>
<td></td>
</tr>
</tbody>
</table>
WHAT'S NEEDED TO DO THIS?

• Python
• Sci-kit Learn & Pandas (python modules)
• Packet captures of non-malicious activity
• Packet captures of malicious activity
• Bro
• Bro HTTP_Header script
• Assimilate python scripts

Customized code at: https://github.com/Soinull/assimilate
STEP BY STEP

1. Collect and process training data
2. Train model
3. Assess actual data files
4. Validate suspicious entries
5. Retrain as needed to improve accuracy
6. 🍺
TRAINING DATA

MALWARE-TRAFFIC-ANALYSIS.NET

Internal Malicious Traffic

Wireshark

Malicious Data

Labeled Data

Normal Data

Test Data

Training Data
PROCESSING PACKET CAPTURES

• Install customized HTTP_Headers Bro module
• Process all packet captures with “Bro –r”
event http_all_headers(c: connection, is_orig: bool, hlist: mime_header_list)
{
    local my_log: Info;
    local origin: string;
    local identifier: string;
    # local event_json_string: string;
    local event_kv_string: string;

    # Is the header from a client request or server response
    if ( is_orig )
        origin = "client";
    else
        origin = "server";

    # If we don't have a header_info_vector than punt
    if ( ! c$http || ! c$http$header_info_vector )
        return;

    print c$http$header_info_vector;
# Example script to iterate over pcap files to get corresponding http.log and httpheader.log files

```bash
for file in ../*.*.pcap
do
    name=${file##*/}
    echo $name
    base=${name%.*.pcap}
    echo $base
    cp ../"$file" .
    bro -r "$file" custom/BrowserFingerprinting/http-headers.bro
    mv http.log ../"$base"_http.log
    mv httpheaders.log ../"$base"_httpheaders.log
    rm -f *.log *.pcap
done
```
PROCESSING PCAPS
ASSIMILATE-TRAIN.PY

```
timcrothers@wodeergediannao:/Desktop/assimilate$ ls
```

Terminal window showing a directory listing with files and folders named `assimilate`, `assimilate_test`, and `assimilate_test_am`. The current directory is `/Desktop/assimilate`.
BUILDING ML MODELS FOR HUNTING

• More data == More accuracy
• More data == Slower speed
• Bro Header Normalization == Lower Accuracy
• Tighter Scoping == More Accuracy
ASSIMILATE-ASSESS.PY
data = DataFrame({'header': [], 'class': []})
blr = BroLogReader()
print('Reading normal data...')
data = data.append(blr.dataFrameFromDirectory(opts.normaldata, 'good'))
print('Reading malicious data...')
data = data.append(blr.dataFrameFromDirectory(opts.maliciousdata, 'bad'))
print('Vectorizing data...')
vectorizer = CountVectorizer()
counts = vectorizer.fit_transform(data['header'].values)
classifier = MultinomialNB()
targets = data['class'].values
classifier.fit(counts, targets)
print('Writing out models...')
joblib.dump(vectorizer, opts.vectorizerfile)
joblib.dump(classifier, opts.bayesianfile)
TAKEAWAYS

• Pandas & Sci-kit Learn make Data Science & Machine Learning available to everyone

• ML tools have progressed to the point that cyber hunters can use them as black boxes

• ClearCut & Assimilate are starter tools that are easily modified to adding serious ML capabilities to your hunting efforts
NEXT STEPS

• Bro fix for header normalization
• Integration with additional validation
• Additional data models
• Different Features/Use Cases
• Streaming support
ADDITIONAL RESOURCES

• ClearCut – github.com/DavidJBianco/Clearcut
• Assimilate – github.com/Soinull/assimilate
• Bro Scripts & Log Reader – github.com/ClickSecurity/data_hacking
Tim Crothers
@soinull
badsecurity@gmail.com