Some Lessons Learned From OSINT Tool Development & Operations

Mike Geide
PUNCH Cyber Analytics Group
About

- Former life:
  - USG SOC/CSIRC Analyst (US-CERT and others)
  - Security researcher at Zscaler – data-rich secure web gateway cloud service provider
  - Author of Poortego and some other tools

- Present:
  - Co-founder and CTO at PUNCH Cyber Analytics Group

Mike Geide | mike@punchcyber.com | punchcyber.com
Poortego

- Why:
  - Intelligence tools largely commercial and some short-comings
  - Doing threat intelligence on budget
  - Sensed need/want for a FOSS equivalent to Maltego
- Developed a prototype
  - Ruby code - run stand-alone or as Metasploit plugin
  - Presented: SecTor, RSA Europe
  - Mentioned in Team Cymru Dragon Newsbytes, on CIF mailing list, etc.
- https://github.com/mgeide/poortego
Some things I’ve learned

- Getting contributors is hard… people are busy
  - Maybe more luck if was Python and/or web-based

- I got busy (and didn’t want to maintain code)
  - Start company, have 2\textsuperscript{nd} kid, train for marathon, etc.

- Not going to compete with a commercial entity
  - Developer resources, QA, GUI/features, Documentation, Support
  - Maltego Tungsten is cool (see BlackHat 2013 preso)

- Intelligence components got a budget all the sudden
  - Threat intel, big data, APT, etc. buzz words
Poortego Future - TBD

- Death?
  - Instead purchase a Maltego license for $650
  - Write local transforms and leverage Python frameworks like Sploitego (https://github.com/allfro/sploitego)

- Or lots more coding?
  - Team-up with other developers if there is interest
  - Update code-base to support Python-based API/transforms
  - Make client/server web-based (nodejs?)
  - Leverage graph database? (neo4j)
  - Add in a smattering of features and cool graphics
  - Integrate, integrate, integrate …
Integration versus “yet another tool”

- Maltego machines will refresh/link off of what you know, but there is other data out there that you don’t know about.

- More and more data feeds and tools that provide levels of intelligence are available – but they are not integrated.

- Case study of how we handle in-house…
“The Hub” Overview

- Intended to solve our “integration problem”
  - Leverage any/all intelligence projects (FOSS or commercial)
    - Let them do what they’re good at / intended to do
  - Integrate into workflow quickly and easily
    - Versus modifying each underlying project code
  - One screen for querying or adding data quickly
    - Query/caching service providing actionable output
“The Hub”
Architecture

- CouchDB – result data broker (JSON)
  - http://couchdb.apache.org/

- ElasticSearch / CouchDB Elastic River – index documents
  - http://www.elasticsearch.org/
  - https://github.com/elasticsearch/elasticsearch-river-couchdb

- Python and Bottle web framework
“The Hub” Integration

- In-house we leverage existing FOSS projects in the Hub:
  - MITRE’s CRITS – threat campaign tracking
    - https://github.com/MITRECND/
  - CIF instances – pulls of data feeds (e.g., ZeusTracker)
    - https://github.com/collectiveintel/cif-v1
  - Cuckoo – malware sandbox
    - https://github.com/cuckoobox/cuckoo
  - Moloch – pcap analysis and repository
    - https://github.com/aol/moloch
  - IOCextractor
    - https://github.com/stephenbrannon/IOCextractor
  - News-pet – RSS open-source feeds
  - Internal projects
    - “Ntropi” – internal domain and resolution tracking project
    - Norman Sandbox (commercial)
    - And Public/Private/Internal data sources (our “Stream”)
“The Hub” Simplicity

- **Ask**
  - Single query interface to all sources
    - Tags, email, hashes, domains, ips

- **Give** – directly provide data or a source of data
  - Allows importing of data from a URL, comma separated list, or plaintext file
    - If URL, and not already a source known to hub, it is added as “tell” source too

- **Tell** – tell the system where to go look on its own
  - Review/Add scheduled open source information gathering
    - Known sources, blogs, tweets, email lists, etc.
“The Hub” – Ask

### “The Hub” – Ask / Pivot

#### ntropi

<table>
<thead>
<tr>
<th>Domain</th>
<th>IP Address</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>yahooeast.net</td>
<td>184.168.221.37</td>
<td>2013-09-20T19:00:44</td>
</tr>
</tbody>
</table>

#### c/f

0.65


<table>
<thead>
<tr>
<th>Domain</th>
<th>IP Address</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>eabblankchair.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rtblankchair.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ali.blankchair.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dlfreshdns.org</td>
<td></td>
<td></td>
</tr>
<tr>
<td>downloadmp3server.servemp3.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.yahoeast.net">www.yahoeast.net</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t.co/7njbo81a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t.co/8umlq6xe</td>
<td>180.150.228.102</td>
<td>163.17.117.90</td>
</tr>
<tr>
<td>110.45.156.5</td>
<td>66.153.86.14</td>
<td>6546123.com</td>
</tr>
<tr>
<td>8bab4b518400725ba60cb50afde3b28701</td>
<td>5d6c0511808bb11d0b55d9a872fd</td>
<td>4d2573e68539973ab0bb3a8fb87582</td>
</tr>
<tr>
<td>dbd6b1032d7fb475f260111bf1d77856e645e2987c3d3a682b6a34737</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### ntropi

<table>
<thead>
<tr>
<th>Domain</th>
<th>IP Address</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.yahooeast.net">www.yahooeast.net</a></td>
<td>184.168.221.37</td>
<td>2013-09-20T19:00:51</td>
</tr>
<tr>
<td>yahooeast.net</td>
<td>184.168.221.37</td>
<td>2013-09-20T19:00:44</td>
</tr>
<tr>
<td>boeing-job.com</td>
<td>184.168.221.37</td>
<td>2013-04-12T20:15:10</td>
</tr>
</tbody>
</table>
“The Hub” – Ask / Pivot (2)
“The Hub” - Give

give the hub

/09/operation-deputydog-zero-day-cve-2013-3893-attack-against-japanese-targets.html

apt

{
  "docType": "report",
  "docMeta": {
    "published_date": "2013-09-24 00:45:21",
    "tags": [
      "apt"
    ]
  },
  "docContents": {
    "extracted_info": [
      {
        "type": "name",
        "value": "img20130823.jpg"
      },
      {
        "type": "md5",
        "value": "8aba4b5184072f2a50c5ecfe326701"
      }
    ]
  }
}
“The Hub” – Tell

tell the hub\{sources\}

<table>
<thead>
<tr>
<th>source uri</th>
<th>path</th>
<th>frequency</th>
<th>last checked</th>
</tr>
</thead>
<tbody>
<tr>
<td>http://</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://blog.malwaretracker.com">http://blog.malwaretracker.com</a></td>
<td></td>
<td>daily</td>
<td>2013-09-24 10:00:23</td>
</tr>
<tr>
<td><a href="http://blog.shadowserver.org">http://blog.shadowserver.org</a></td>
<td></td>
<td>daily</td>
<td>2013-09-24 10:00:30</td>
</tr>
</tbody>
</table>
“The Hub” – Tell (History)
Conclusions

- Lots of projects, data, etc.
  - “Big Data” is hot right now

- Spend more time integrating and doing your analysis than developing
  - Unless you enjoy maintaining code ;)

- “The Hub” style integration is effective and simple to use
  - Let the complexities reside in any of the other integrated projects

And yes, PUNCH can help set this up for you
(shameless plug)

Mike Geide  |  mike@punchcyber.com  |  punchcyber.com
Sidebar – GraphDB (neo4j)

[Image: Neo4j database with a query and a graph visualization]