Enabling GEOINT and Cyber Security

Travis Pinney
Berico Technologies
@tlpinney
Mil-OSS LANT2 – 2013-Aug-08
What is CLAVIN is NOT?

• A cyber security solution
What is CLAVIN?

• Cartographic Location And Vicinity INdexer
• Geoparser
• Extracts location names
• Resolves geospatial entities
• Open source
• Runs on Hadoop
Under the hood

1. UNSTRUCTURED TEXT DOCUMENT
2. 3RD PARTY NAMED ENTITY RECOGNIZER
3. LIST OF LOCATION NAMES
4. GEONAMES.org GAZETTEER
5. LUCENE INDEX
6. CLAVIN LOCATION RESOLVER
7. STRUCTURED GEO DATA
CLAVIN handles

- Ambiguous references
- Typos & phonetic spellings
- "Ivory Coast" == "Côte d'Ivoire"
Demo Time

Locations Parsed and Resolved From Text

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Lat, Lon</th>
<th>Country Code</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>6446345</td>
<td>US</td>
<td>49, 1, 1.96667</td>
<td>FR</td>
<td>9</td>
</tr>
<tr>
<td>298795</td>
<td>Turkey</td>
<td>39.05901, 34.91155</td>
<td>TR</td>
<td>9</td>
</tr>
<tr>
<td>2215636</td>
<td>Libya</td>
<td>28, 17</td>
<td>LY</td>
<td>8</td>
</tr>
<tr>
<td>88319</td>
<td>Benghazi</td>
<td>32.11667, 20.06667</td>
<td>LY</td>
<td>7</td>
</tr>
<tr>
<td>2215636</td>
<td>Libyan</td>
<td>28, 17</td>
<td>LY</td>
<td>5</td>
</tr>
<tr>
<td>3199389</td>
<td>Herzegovina</td>
<td>43, 17.83333</td>
<td>BA</td>
<td>1</td>
</tr>
<tr>
<td>1149361</td>
<td>Afghanistan</td>
<td>33, 66</td>
<td>AF</td>
<td>1</td>
</tr>
<tr>
<td>529468</td>
<td>Southeast Europe</td>
<td>43.53333, 3.98333</td>
<td>FR</td>
<td>1</td>
</tr>
<tr>
<td>1319</td>
<td>Bengazi</td>
<td>32.11667, 20.06667</td>
<td>LY</td>
<td>1</td>
</tr>
<tr>
<td>10630</td>
<td>Cairo</td>
<td>30.06263, 31.24967</td>
<td>EG</td>
<td>1</td>
</tr>
<tr>
<td>5455014</td>
<td>American</td>
<td>33.52813, -105.74332</td>
<td>US</td>
<td>1</td>
</tr>
<tr>
<td>831053</td>
<td>Kosovo</td>
<td>42.58333, 21</td>
<td>XK</td>
<td>1</td>
</tr>
<tr>
<td>3277605</td>
<td>Bosnia</td>
<td>44.25, 17.83333</td>
<td>BA</td>
<td>1</td>
</tr>
<tr>
<td>685608</td>
<td>Balkans</td>
<td>44, 23</td>
<td>RO</td>
<td>1</td>
</tr>
<tr>
<td>6783140</td>
<td>Gadafi</td>
<td>12.43269, 14.24894</td>
<td>CK</td>
<td>1</td>
</tr>
<tr>
<td>783754</td>
<td>Albania</td>
<td>41, 20</td>
<td>AL</td>
<td>1</td>
</tr>
<tr>
<td>718075</td>
<td>Macedonia</td>
<td>41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CLAVIN stats

- **Accurate**: 0.75
- **Fast**: 100 locations per second per CPU
- **Scalable**: processes 1 million documents in under 1 hour on a 9-node Hadoop cluster
- **Open Source**: Apache License on GitHub
CLAVIN Extensions

- Custom Geo Gazetteers (OpenStreetMap)
- IP Address Gazetteer Integration
- GeoJson Support
- Log Extractors
OpenStreetMap
OpenStreetMap

• Node
• Ways
• Relations
OpenStreetMap

• Node
  place=*  
  name=*
IP Address Gazetteer

MaxMind GeoLite2 Database
Parsing Strategies

GeoHash

IpAddress

LatLon

Dms (Degrees Minutes Seconds)
Parsing Strategies

GeoHash

IpAddress

LatLon

Dms (Degrees Minutes Seconds)
Parsing Strategies

GeoParserFactory.DefaultCoordinateParsingStrategies
.add(new IpAddressParsingStrategy());

// Get a parser instance.
GeoParser parser =
GeoParserFactory.getDefault("IndexDirectory/");

// Extract and resolve IP Addresses and coordinates from the text below.
ResolutionContext results =
parser.parse("This is a test  41.222.183.44");
Parsing Strategies

41.222.183.44 (ipaddress) was extracted as -6.8, 39.2833.
41.222.183.44 (ipaddress) was resolved as Golden Tulip Dar Es Salaam, Tanzania.
GeoJSON

geojson.org

1.0
https://github.com/Leaflet/Leaflet
Github
# Netstat

## Active Connections

<table>
<thead>
<tr>
<th>Proto</th>
<th>Local Address</th>
<th>Foreign Address</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>10.8.30.56:54082</td>
<td>74.125.228.118:443</td>
<td>ESTABLISHED</td>
</tr>
</tbody>
</table>
Netstat Feature

```javascript
var feature = {
    "type": "Feature",
    "geometry": {
        "type": "Point",
        "coordinates": [-122.0813, 37.4139]
    },
    "properties": {
        "src-ip": "10.8.30.56",
        "src-port": "54082",
        "dst-ip": "74.125.228.118",
        "dst-port": "443",
        "state": "ESTABLISHED"
    }
};
```
Netstat Feature
Tomcat Logs

50.78.x.x - - [04/Oct/2012:17:57:17 +0000] "GET /clavin-web/ HTTP/1.1" 200 5329
```
var feature = {
    "type": "Feature",
    "geometry": {
        "type": "Point",
        "coordinates": [122.0828, 37.3861]
    },
    "properties": {
        "ip": "50.78.X.X",
        "datetime": "2012-10-04T17:57Z",
        "action": "GET",
        "path": "/clavin-web/",
        "code": "200",
        "bytes": "5329"
    }
};
```
Tomcat Log Feature
Styling

• https://help.github.com/articles/mapping-geojson-files-on-github#styling-markers
var feature = {
    "type": "Feature",
    "properties": {
        "ip": "50.78.x.x",
        "datetime": "2012-10-04T17:57Z",
        "action": "GET",
        "path": "/clavin-web/",
        "state": "ESTABLISHED",
        "marker-color": "#00FF00"
    },
    "geometry": {
        "type": "Point",
        "coordinates": [122.0828, 37.3861]
    }
};
# Feature Color

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ip</strong></td>
<td>50.78.x.x</td>
</tr>
<tr>
<td><strong>datetime</strong></td>
<td>2012-10-04T17:57Z</td>
</tr>
<tr>
<td><strong>action</strong></td>
<td>GET</td>
</tr>
<tr>
<td><strong>path</strong></td>
<td>/clavin-web/</td>
</tr>
<tr>
<td><strong>code</strong></td>
<td>200</td>
</tr>
<tr>
<td><strong>bytes</strong></td>
<td>5329</td>
</tr>
</tbody>
</table>
Tools

GEOJSONLINT.COM
Network

Firewall Logs
HTTP Logs
Pcap
Netflow
Social Media

Twitter
Facebook
Panoramio
FourSquare
Social Media

heatmap.js
Social Media
Conclusion

• One small but important part of a network security analytic workflow
Links

https://github.com/Berico-Technologies/CLAVIN

https://github.com/Berico-Technologies/CLAVIN-contrib
clavin.bericotechnologies.com
@CLAVIN__ (two underscores)
@greenbacker
@tlpinney
@BericoTech (we’re hiring!)