SKYNET:
Courier Detection via Machine Learning

June 5, 2012
Given a handful of courier selectors, can we find others that “behave similarly” by analyzing GSM metadata?

It’s worth noting that:

• we are looking for different people using phones in similar ways
• without using any call chaining techniques from known selectors
• by scanning through all selectors seen in Pakistan that have not left Af/Pak (~55M)
From GSM metadata, we can measure aspects of each selector’s pattern-of-life, social network, and travel behavior.
This presentation describes our search for AQSL couriers using behavioral profiling

Behavioral Feature Extraction

Cross Validation Experiment on AQSL Couriers

Preliminary SIGINT Findings
Counting unique UCELLIDs shows that couriers travel more often than typical Pakistani selectors
By examining multiple features at once, we can see some indicative behaviors of our courier selectors.
Looking at a hierarchical clustering derived from all 80 features, the AQSL groups mostly stay together.

4 days of collect
Now, we’ll describe a cross validation experiment on the AQSL selectors that we were provided.
Our initial detector uses the centroid of the AQSL couriers to “find other selectors like these”

AQSL Cross-Validation Experiment

• 7 MSISDN/IMSI pairs

• Hold each pair out and score them when training the centroid on the rest

• Assume that random draws of Pakistani selectors are nontargets

• How well do we do?
Our initial detector uses the centroid of the AQSL couriers to “find other selectors like these”

AQSL Cross-Validation Experiment

• Initial experiments showed EER in 10-20% range

• Here, performance is much worse against these nontargets:
  • Seen in Pakistan
  • Not seen outside of Af/Pak
  • Not FVEY selectors
Statistical algorithms are able to find the couriers at very low false alarm rates, if we’re allowed to miss half of them.

**Random Forest Classifier**

- 7 MSISDN/IMSI pairs
- Hold each pair out and then try to find them after learning how to distinguish remaining couriers from other Pakistanis (using 100k random selectors here)
- Assume that random draws of Pakistani selectors are nontargets
- 0.18% False Alarm Rate at 50% Miss Rate
We’ve been experimenting with several error metrics on both small and large test sets

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Classifier</th>
<th>Features</th>
<th>100k Test Selectors</th>
<th>55M Test Selectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>False Alarm Rate at 50% Miss Rate</td>
<td>Mean Reciprocal Rank</td>
</tr>
<tr>
<td>None</td>
<td>Random</td>
<td>None</td>
<td>50%</td>
<td>1/23k (simulated)</td>
</tr>
<tr>
<td>Known Couriers</td>
<td>Centroid</td>
<td>All</td>
<td>20%</td>
<td>1/18k</td>
</tr>
<tr>
<td>+ Anchory Selectors</td>
<td>Random Forest</td>
<td>Outgoing</td>
<td>43%</td>
<td>1/27k</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.18%</td>
<td>1/9.9</td>
</tr>
</tbody>
</table>

Random Forest:
- 0.18% false alarm rate at 50% miss rate
- 7x improvement over random performance when evaluating its tasked precision at 100
To get more training data we scraped selectors from S2I11 Anchory reports containing keyword “courier”

Anchory Selectors

- Searched for reports containing “S2I11” AND “courier”
- Filtered out non-mobile numbers and kept selectors with “interesting” travel patterns seen in SmartTracker
Adding selectors from Anchory reports to the training data reduced the false alarm rates even further.

**Anchory Selectors**

- Searched for reports containing “S2I11” AND “courier”
- Filtered out non-mobile numbers and kept selectors with “interesting” travel patterns seen in SmartTracker
We’ve been experimenting with several error metrics on both small and large test sets.

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Classifier</th>
<th>Features</th>
<th>100k Test Selectors</th>
<th>55M Test Selectors</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>False Alarm Rate at 50% Miss Rate</td>
<td>Mean Reciprocal Rank</td>
<td>Tasked Selectors in Top 500</td>
</tr>
<tr>
<td>None</td>
<td>Random</td>
<td>None</td>
<td>50%</td>
<td>1/23k (simulated)</td>
<td>0.64 (active/Pak)</td>
</tr>
<tr>
<td>Known Couriers</td>
<td>Centroid</td>
<td>All</td>
<td>20%</td>
<td>1/18k</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random Forest</td>
<td>Outgoing</td>
<td>43%</td>
<td>1/27k</td>
<td></td>
</tr>
<tr>
<td>+ Anchory Selectors</td>
<td></td>
<td></td>
<td>0.18%</td>
<td>1/9.9</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.008%</td>
<td>1/14</td>
<td>21</td>
</tr>
</tbody>
</table>

Random Forest trained on Known Couriers + Anchory Selectors:
- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100
Now, we’ll investigate some findings after running these classifiers on +55M Pakistani selectors via MapReduce

Behavioral Feature Extraction

Cross Validation Experiment on AQSL Couriers

Preliminary SIGINT Findings
The highest scoring selector that traveled to Peshawar and Lahore is PROB AHMED ZAIDAN.
In the top 500 scoring selectors, 21 are tasked leading us to believe that we’re on the right track.
We have also discovered many untasked selectors with interesting travel patterns.
Preliminary results indicate that we’re on the right track, but much remains to be done.

**Cross Validation Experiment:**
- Random Forest classifier operating at 0.18% false alarm rate at 50% miss
- Enhancing training data with Anchory selectors reduced that to 0.008%
- Mean Reciprocal Rank is ~1/10

**Preliminary SIGINT Findings:**
- Behavioral features helped discover similar selectors with “courier-like” travel patterns
- High number of tasked selectors at the top is hopefully indicative of the detector performing well “in the wild”