LIFECYCLE OF A PHONE FRAUDSTER: FROM ACCOUNT RECONNAISSANCE TO TAKEOVER

Vijay Balasubramaniyan, Raj Bandyopadhyay, Telvis Calhoun
7 August, 2014
ONE IN EVERY 2,901 CALLS IS FRAUD

Number of Calls

- > 10B (85) 1,932
- > 100B (17) 3,311
- > 1,000B (4) 4,635

Number of Calls
PHONE FRAUD BY THE NUMBERS

12% – 94%
Awareness of fraud on the phone channel

5
Avg. # of calls needed to compromise an account

$700,000
Largest financial transaction loss stopped

$0.57
Dollars lost for every call into your call center
PHONEPRINTING™

Call Audio
- Requires 15 seconds of call audio

147 audio features in each fingerprint

- Loss:
  - Packet loss
  - Robotization
  - Dropped frames

- Spectrum:
  - Quantization
  - Frequency filters
  - Codec artifacts

- Noise:
  - Clarity
  - Correlation
  - Signal-to-noise ratio

Phoneprint™

Phone Type
Geo-Location
Risk Score

Requires 15 seconds of call audio
FRAUD DETECTION SYSTEM

Phone Traffic

- ANI
- Audio

Pindrop Consortium

- Reputation
- Geography
- Device Type
- Phoneprint
- Voiceprint

Analysis

Scoring

Notification

Risky

81

Fraud Analyst

API

Pindrop Security

Fraud Analyst
ANALYSIS AT SCALE

100 Million calls
18 Million Originating ANIs
12 Million Accounts
At score threshold of 60, stop 91% of fraud while passing 99% of genuine
FRAUD CALL DISTRIBUTION

Percentage of fraud calls

<table>
<thead>
<tr>
<th>Domestic</th>
<th>International</th>
</tr>
</thead>
<tbody>
<tr>
<td>48%</td>
<td>52%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell</th>
<th>Land</th>
<th>VoIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>29%</td>
<td>27%</td>
<td>44%</td>
</tr>
</tbody>
</table>
WEST AFRICA “ONE”
VOICE DISTORTION
CHARACTERISTICS OF CDR DATA

● Basic metadata about calls
  o Source and destination ANIs
  o Start and end timestamps of call

● Advanced application-specific metadata
  o IVR call flow information
  o Account numbers and other user information
  o Tower & base station information (for cell networks)
GRAPH REPRESENTATION OF CDRS

- Directed graph: ANIs as nodes, edges from source to target
- Edges annotated with timing and other information
- Can include other elements as nodes e.g. account numbers
FEATURES FOR CDR ANALYSIS

- **Reputation features**
  - Carrier, device type, prepaid status of source ANI
  - Complaints against source ANI

- **Velocity (graph) features**
  - Number of ANIs & accounts targeted by source ANI
  - Frequency and duration of calls from ANI
  - Application-specific features e.g. ANI scanning, number of authentication attempts
FEATURES FOR CDR ANALYSIS

- Behavior features from IVR
  - Call flow sequence in the IVR e.g. [Account Entry, PIN Entry, Balance Check]
  - Use call flow sequences from single or multiple calls
  - Break up sequences into short chunks
  - Represent chunks as fixed-length, numeric vectors
  - Select top K features using feature selection techniques e.g. chi-squared
CASE STUDY 1: CALLING CARD TELCO

● Premium rate services fraud
  o Fraudsters using stolen calling cards to call fake ‘premium’ numbers abroad
  o Use of automated robots to discover valid customer ANIs (ANI scanning) and dial out using those ANIs

● Our CDR analysis approach
  o Create features based on graph analysis, duration of calls, and interval between subsequent calls
  o Create a custom feature to identify scanning
WE DETECT ANI SCANNING

- Detect over 80% of premium rate fraud, up to 10 days before actual fraud calls
- ANI scanning feature detects 50% of fraud
CASE STUDY 2: BENEFITS PROVIDER

● Fraudulent claims in state benefits
  ○ Fraudsters suspected of performing reconnaissance over IVR to find valid info
  ○ Use of valid account info for account takeovers

● Our approach
  ○ Combine reputation, velocity and behavior features
  ○ Train model over labeled set of calls
  ○ Use model to score incoming calls
WE FIND IVR RECONNAISSANCE

- Suspicious activity on 46% of accounts up to 2 months before fraud
- Specific instances of reconnaissance in IVR

<table>
<thead>
<tr>
<th>ACEStart</th>
<th>ACEStart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Menu</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Balance PAN - Begin</td>
<td></td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8888</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Error</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Retry Menu</td>
<td>TryAgain</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8839</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Retry Menu</td>
<td>TryAgain</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************3994</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Business Logic</td>
</tr>
<tr>
<td>ACEStart</td>
<td>ACEStart</td>
</tr>
<tr>
<td>Language Menu</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Balance PAN - Begin</td>
<td></td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8883</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Retry Menu</td>
<td>Error</td>
</tr>
<tr>
<td>PAN Retry Menu</td>
<td>TryAgain</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8883</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>****</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>Balance Playback</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACEStart</th>
<th>ACEStart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Menu</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Balance PAN - Begin</td>
<td></td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8839</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Error</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************8399</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Entry</td>
<td>Timeout</td>
</tr>
<tr>
<td>PAN Retry Menu</td>
<td>Hang Up</td>
</tr>
<tr>
<td>ACEStart</td>
<td>ACEStart</td>
</tr>
<tr>
<td>Language Menu</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Balance PAN - Begin</td>
<td></td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************9491</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>****</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>Balance Playback</td>
</tr>
<tr>
<td>ACEStart</td>
<td>ACEStart</td>
</tr>
<tr>
<td>Language Menu</td>
<td>Error</td>
</tr>
<tr>
<td>ACEStart</td>
<td>ACEStart</td>
</tr>
<tr>
<td>Language Menu</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Balance PAN - Begin</td>
<td></td>
</tr>
<tr>
<td>PAN Entry</td>
<td>**************5629</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>****</td>
</tr>
<tr>
<td>PIN Entry</td>
<td>Balance Playback</td>
</tr>
</tbody>
</table>

- ANI: 209-532-XXXX
- 7 consecutive calls in 1 hour
- Sequence of invalid Account number (PAN) entry attempts, followed by successful PAN entry and PIN entry.
ML TRAINING AT SCALE
REAL-TIME PREDICTION ARCHITECTURE
MONGODB: LESSONS LEARNED

● **Bulk Ingest**
  - Use Journaled Write-Concern for Inserts. Acknowledged Write-Concern for Updates.

● **Query**
  - Use Aggregations API + Indexes to generate IVR and CDR features.

● **Prediction**
  - Store feature vectors as Binary BSON objects.
CONCLUSION

- Account takeover – acoustical anomalies
  - > 80% TDR, < 2% FPR
  - 52% coming from international locations
- Account reconnaissance – CDR analysis
  - 46% detected 2 months before attack
- Detect pre-crime, zero day attacks and repeat attacks on the phone channel by complete understanding of lifecycle of a fraudster
PINDROP SECURITY

vijay@pindropsecurity.com
raj@pindropsecurity.com
tcalhoun@pindropsecurity.com