Catching Bandits and Only Bandits: Privacy-Preserving Intersection Warrants for Lawful Surveillance

Aaron Segal, Bryan Ford, and Joan Feigenbaum
Yale University
FOCI 2014

“...an unspeakable blasphemy.” - @Dymaxion
Overview

• Mass Surveillance and Privacy – Introduction

• Privacy Principles for Open Surveillance Processes

• Case Study – High Country Bandits and Lawful Intersection Protocol

• Implementation & Evaluation
Motivation & Goals

“State of the art” discussion on surveillance and privacy:

- **Secret** processes for data collection

- Public is asked to **trust** the government

- Presumed **tradeoff** between *national security* and *personal privacy*

- Ideal world: **No surveillance**
Motivation & Goals

“State of the art” discussion on surveillance and privacy:

- **Secret** processes for data collection

- Public is asked to **trust** the government

- Presumed **tradeoff** between *national security* and *personal privacy*

- Ideal world: **No surveillance**
  - Realistic goal: **Surveillance with privacy protection**
Motivation & Goals

“State of the art” discussion on surveillance and privacy:

- Secret processes for data collection

- Public is asked to trust the government

- Presumed tradeoff between national security and personal privacy
  - No need to abandon personal privacy to ensure national security

- Ideal world: No surveillance
  - Realistic goal: Surveillance with privacy protection
Motivation & Goals

“State of the art” discussion on surveillance and privacy:

- **Secret** processes for data collection

- Public is asked to **trust** the government
  - Accountability guaranteed by existing **cryptographic technology**

- Presumed **tradeoff** between **national security** and **personal privacy**
  - **No need** to abandon **personal privacy** to ensure **national security**

- Ideal world: **No surveillance**
  - Realistic goal: **Surveillance with privacy protection**
Motivation & Goals

“State of the art” discussion on surveillance and privacy:

- Secret processes for data collection
  - Open processes for data collection
- Public is asked to trust the government
  - Accountability guaranteed by existing cryptographic technology
- Presumed tradeoff between national security and personal privacy
  - No need to abandon personal privacy to ensure national security
- Ideal world: No surveillance
  - Realistic goal: Surveillance with privacy protection
Privacy Principles for Surveillance

Open processes
- Must follow rules and procedures of public law
- Need not disclose targets and details of investigations

Two types of users:

- **Targeted users**
  - Under **suspicion**
  - Subject of a **warrant**
  - Can be **known** or **unknown**

- **Untargeted users**
  - No probable cause
  - Not targets of investigation
  - The vast majority of internet users
Open Privacy Firewall

I. Any surveillance or law-enforcement process that obtains or uses private information about *untargeted users* shall be an **open**, public, unclassified process.

II. Any **secret** surveillance or law-enforcement process shall use only:
   a. public information, and
   b. private information about *targeted users* obtained under authorized warrants via open surveillance processes.
Surveillance Privacy Principles

• Division of trust
  - No one agency can compromise privacy

• Enforced scope limiting
  - Overly broad group of users’ data is not captured

• Sealing time and notification
  - Finite, reasonable time before users are notified

• Accountability
  - Statistics presented on use of surveillance
Case Study – High Country Bandits

2010 case – string of bank robberies in Arizona, Colorado

FBI Intersection attack compared 3 cell tower dumps totaling 150,000 users

• 1 number found in all 3 cell dumps – led to arrest
• 149,999 innocent users’ information acquired
Intersecting Cell-Tower Dumps

• Law enforcement goal: Find *targeted, unknown* user whose phone number will appear in the intersection of cell-tower dumps

• Used in: High Country Bandits case, CO-TRAVELER program
  - Same principle for any collection of metadata

<table>
<thead>
<tr>
<th>Cell Tower A</th>
<th>Time $t_1$</th>
<th>Cell Tower B</th>
<th>Time $t_2$</th>
<th>Cell Tower C</th>
<th>Time $t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>203-555-4430</td>
<td></td>
<td>203-555-3222</td>
<td></td>
<td>203-555-7928</td>
<td></td>
</tr>
<tr>
<td>203-555-3435</td>
<td></td>
<td>203-555-3813</td>
<td></td>
<td>203-555-0599</td>
<td></td>
</tr>
<tr>
<td>203-555-2840</td>
<td></td>
<td>203-555-2786</td>
<td></td>
<td>203-555-6445</td>
<td></td>
</tr>
<tr>
<td>203-555-7691</td>
<td></td>
<td>203-555-7976</td>
<td></td>
<td>203-555-7511</td>
<td></td>
</tr>
<tr>
<td>203-555-1505</td>
<td></td>
<td>203-555-0392</td>
<td></td>
<td>203-555-2277</td>
<td></td>
</tr>
<tr>
<td>203-555-9589</td>
<td></td>
<td>203-555-5872</td>
<td></td>
<td>203-555-7976</td>
<td></td>
</tr>
<tr>
<td>203-555-7976</td>
<td></td>
<td>203-555-4891</td>
<td></td>
<td>203-555-2840</td>
<td></td>
</tr>
<tr>
<td>203-555-9266</td>
<td></td>
<td>203-555-9709</td>
<td></td>
<td>203-555-3222</td>
<td></td>
</tr>
</tbody>
</table>
Intersecting Cell-Tower Dumps

• Law enforcement goal: Find *targeted, unknown* user whose phone number will appear in the intersection of cell-tower dumps

• Used in: High Country Bandits case, CO-TRAVELER program
  - Same principle for any collection of metadata

<table>
<thead>
<tr>
<th>Cell Tower A</th>
<th>Cell Tower B</th>
<th>Cell Tower C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time $t_1$</td>
<td>Time $t_2$</td>
<td>Time $t_3$</td>
</tr>
<tr>
<td>• 203-555-4430</td>
<td>• 203-555-3222</td>
<td>• 203-555-7928</td>
</tr>
<tr>
<td>• 203-555-3435</td>
<td>• 203-555-3813</td>
<td>• 203-555-0599</td>
</tr>
<tr>
<td>• 203-555-2840</td>
<td>• 203-555-2786</td>
<td>• 203-555-6445</td>
</tr>
<tr>
<td>• 203-555-7691</td>
<td>• 203-555-7976</td>
<td>• 203-555-7511</td>
</tr>
<tr>
<td>• 203-555-1505</td>
<td>• 203-555-0392</td>
<td>• 203-555-2277</td>
</tr>
<tr>
<td>• 203-555-9589</td>
<td>• 203-555-5872</td>
<td>• 203-555-7976</td>
</tr>
<tr>
<td>• <strong>203-555-7976</strong></td>
<td>• 203-555-4891</td>
<td>• 203-555-2840</td>
</tr>
<tr>
<td>• 203-555-9266</td>
<td>• 203-555-9709</td>
<td>• 203-555-3222</td>
</tr>
</tbody>
</table>
Privacy-Protecting Solution

Based on Vaidya, Clifton (2005)

• A **private set intersection protocol** built to satisfy surveillance privacy principles
• Relies on **multiple, independent agencies** to execute protocol, providing division of trust, accountability
• Example:
  • Executive agency (FBI, NSA)
  • Judicial agency (warrant-issuing court)
  • Legislative agency (oversight committee established by law)
Each agency provides encryption key based on commutative, public-key, randomized encryption scheme

- Commutative encryption: \( \text{Dec}_A(\text{Dec}_B(c)) = \text{Dec}_B(\text{Dec}_A(c)) \)

Sources of phone metadata (telecoms) encrypt each data item using all agencies’ keys and give encrypted sets to repositories

When agencies agree on a warrant for intersection, repositories distribute encrypted data sets to agencies

- Agencies individually select temporary keys for a commutative, deterministic encryption scheme to be used for this intersection, then thrown away
Private Set Intersection Protocol – Phase 1

- An agency starts with data sets under *randomized* encryption by all agencies’ keys
- Each agency strips off its layer of *randomized* encryption, adds a layer of *deterministic* encryption using its temporary key, permutes the data sets, and sends them to next agency
Private Set Intersection Protocol – Phase 1

• An agency starts with data sets under *randomized* encryption by all agencies’ keys
• Each agency strips off its layer of *randomized* encryption, adds a layer of *deterministic* encryption using its temporary key, permutes the data sets, and sends them to next agency
Private Set Intersection Protocol – Phase 1

• An agency starts with data sets under *randomized* encryption by all agencies’ keys.

• Each agency strips off its layer of *randomized* encryption, adds a layer of *deterministic* encryption using its temporary key, permutes the data sets, and sends them to next agency.

<table>
<thead>
<tr>
<th>(203) 555-2469</th>
<th>(203) 555-7976</th>
<th>(203) 555-7976</th>
</tr>
</thead>
<tbody>
<tr>
<td>(203) 555-7976</td>
<td>(203) 555-8770</td>
<td>(203) 555-3179</td>
</tr>
</tbody>
</table>
Private Set Intersection Protocol – Phase 1

• An agency starts with data sets under *randomized* encryption by all agencies’ keys
• Each agency strips off its layer of *randomized* encryption, adds a layer of *deterministic* encryption using its temporary key, permutes the data sets, and sends them to next agency

(203) 555-2469  (203) 555-7976  (203) 555-7976
(203) 555-7976  (203) 555-8770  (203) 555-3179
Private Set Intersection Protocol – Phase 1

• An agency starts with data sets under randomized encryption by all agencies’ keys

• Each agency strips off its layer of randomized encryption, adds a layer of deterministic encryption using its temporary key, permutes the data sets, and sends them to next agency

(203) 555-2469
(203) 555-7976
(203) 555-7976
(203) 555-3179
(203) 555-8770
Private Set Intersection Protocol – Phase 2

• When phase I is done, each item has encrypted with deterministic encryption using temporary keys
• Matching ciphertexts = matching plaintexts = targeted users – keep
• Non-matching ciphertexts = untargeted users – discard

(203) 555-2469
(203) 555-7976
(203) 555-7976
(203) 555-8770
(203) 555-3179
Private Set Intersection Protocol – Phase 2

- When phase I is done, each item has encrypted with *deterministic* encryption using *temporary keys*
- Matching ciphertexts = matching plaintexts = targeted users – **keep**
- Non-matching ciphertexts = untargeted users – **discard**
Private Set Intersection Protocol – Phase 2

• When phase I is done, each item has encrypted with deterministic encryption using temporary keys
• Matching ciphertexts = matching plaintexts = targeted users – keep
• Non-matching ciphertexts = untargeted users – discard
Private Set Intersection Protocol – Phase 2

- After phase II, size of intersection revealed
- If intersection cardinality above pre-defined threshold, any agency can stop protocol
  - Prevents accidental compromise of privacy, e.g. “concert scenario”

(203) 555-7976
Private Set Intersection Protocol – Phase 2

- After phase II, size of intersection revealed
- If intersection cardinality above pre-defined threshold, any agency can stop protocol
  - Prevents accidental compromise of privacy, e.g. “concert scenario”
Private Set Intersection Protocol – Phase 2

- After phase II, size of intersection revealed
- If intersection cardinality above pre-defined threshold, any agency can stop protocol
  - Prevents accidental compromise of privacy, e.g. “concert scenario”
Private Set Intersection Protocol – Phase 3

- Once intersection is determined, each agency uses temporary key to remove its layer of encryption
- Set is permuted and passed on as in phase I
- Final results sent to all participants
- Temporary keys securely deleted

(203) 555-7976
Private Set Intersection Protocol – Phase 3

- Once intersection is determined, each agency uses temporary key to remove its layer of encryption
- Set is permuted and passed on as in phase I
- Final results sent to all participants
- Temporary keys securely deleted
Private Set Intersection Protocol – Phase 3

- Once intersection is determined, each agency uses *temporary* key to remove its layer of encryption
- Set is permuted and passed on as in phase I
- Final results sent to all participants
- Temporary keys securely deleted

(203) 555-7976
Protocol Satisfies Privacy Principles

• Satisfies principle of **Open Process**
  - Can openly standardize protocol, crypto *without* compromising investigative power

• Division of trust
  - No one agency can decrypt or perform intersection

• Enforced scope limiting
  - Any agency can stop protocol if sets or intersection are too large

• Sealing time and notification
  - Implementable by policy – all agencies get final data set

• Accountability
  - Because every agency must participate, no agencies can perform attack without other agencies learning and getting statistics
Implementation of Protocol

• We implemented our lawful set intersection protocol in Java

• Tested with three “agencies”, run on PlanetLab nodes distributed across the US (CT, TX, CA)

• Proof-of-concept
  - Unoptimized crypto library
  - One single-threaded worker per “agency”

Evaluation of Implementation

• Running time increases linearly with size of data sets

• Roughly 130-150 milliseconds per item of metadata

• High Country Bandits example with 50,000 items per set takes just under 2 hours to complete (43 minutes of CPU time per node)

<table>
<thead>
<tr>
<th>Items per node (KB)</th>
<th>Data sent per node (KB)</th>
<th>CPU time per node (s)</th>
<th>End-to-End runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>21</td>
<td>0.6</td>
<td>4.1</td>
</tr>
<tr>
<td>25</td>
<td>46</td>
<td>1.3</td>
<td>6.0</td>
</tr>
<tr>
<td>50</td>
<td>86</td>
<td>2.6</td>
<td>9.6</td>
</tr>
<tr>
<td>75</td>
<td>127</td>
<td>3.8</td>
<td>12.6</td>
</tr>
<tr>
<td>100</td>
<td>167</td>
<td>5.0</td>
<td>15.5</td>
</tr>
<tr>
<td>250</td>
<td>410</td>
<td>12.4</td>
<td>38.2</td>
</tr>
<tr>
<td>500</td>
<td>815</td>
<td>24.7</td>
<td>69.1</td>
</tr>
<tr>
<td>750</td>
<td>1220</td>
<td>36.9</td>
<td>103.0</td>
</tr>
<tr>
<td>1000</td>
<td>1625</td>
<td>49.3</td>
<td>137.2</td>
</tr>
<tr>
<td>2500</td>
<td>4055</td>
<td>123.0</td>
<td>369.9</td>
</tr>
<tr>
<td>5000</td>
<td>8106</td>
<td>245.6</td>
<td>724.9</td>
</tr>
<tr>
<td>7500</td>
<td>12156</td>
<td>369.4</td>
<td>1034.9</td>
</tr>
<tr>
<td>10000</td>
<td>16206</td>
<td>493.8</td>
<td>1402.3</td>
</tr>
<tr>
<td>50000</td>
<td>81009</td>
<td>2560.5</td>
<td>6971.2</td>
</tr>
</tbody>
</table>

Table 1: Experimental Results
Conclusions

- **Open** surveillance processes *can* and *should* be designed to meet law enforcement needs while protecting privacy

- Privacy-protecting surveillance is feasible using **existing** technology

- Directions for future work:
  - testing our protocol with optimized, multi-threaded implementation
  - creating privacy-protecting protocols to replace other forms of surveillance
  - testing with general-purpose Secure Multi-party Computation (SMPC) platforms such as FairPlay, Sharemind to automatically compile surveillance queries into privacy-protecting protocols
Thank you!