# 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

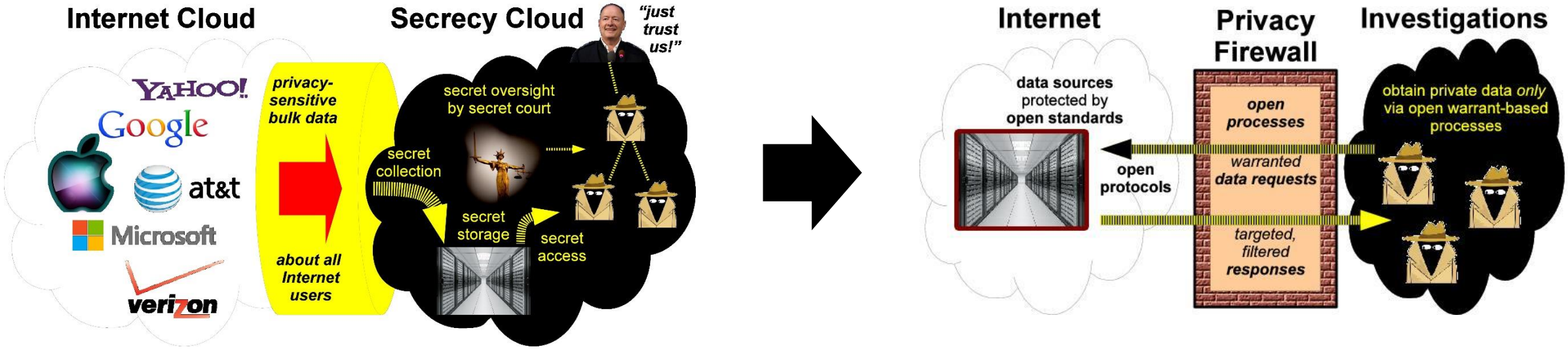
# Open Privacy Firewall

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

| Cell Tower A<br>Time $t_1$ | Cell Tower B<br>Time $t_2$ | Cell Tower C<br>Time $t_3$ |
|---|---|---|
| • 203-555-4430 | • 203-555-3222 | • 203-555-7928 |
| • 203-555-3435 | • 203-555-3813 | • 203-555-0599 |
| • 203-555-2840 | • 203-555-2786 | • 203-555-6445 |
| • 203-555-7691 | • 203-555-7976 | • 203-555-7511 |
| • 203-555-1505 | • 203-555-0392 | • 203-555-2277 |
| • 203-555-9589 | • 203-555-5872 | • 203-555-7976 |
| • 203-555-7976 | • 203-555-4891 | • 203-555-2840 |
| • 203-555-9266 | • 203-555-9709 | • 203-555-3222 |

# 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

| Cell Tower A<br>Time $t_1$ | Cell Tower B<br>Time $t_2$ | Cell Tower C<br>Time $t_3$ |
|---|---|---|
| 203-555-4430 | 203-555-3222 | 203-555-7928 |
| 203-555-3435 | 203-555-3813 | 203-555-0599 |
| 203-555-2840 | 203-555-2786 | 203-555-6445 |
| 203-555-7691 | **203-555-7976** | 203-555-7511 |
| 203-555-1505 | 203-555-0392 | 203-555-2277 |
| 203-555-9589 | 203-555-5872 | **203-555-7976** |
| **203-555-7976** | 203-555-4891 | 203-555-2840 |
| 203-555-9266 | 203-555-9709 | 203-555-3222 |

# 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)

# Private Set Intersection Protocol – Preparation

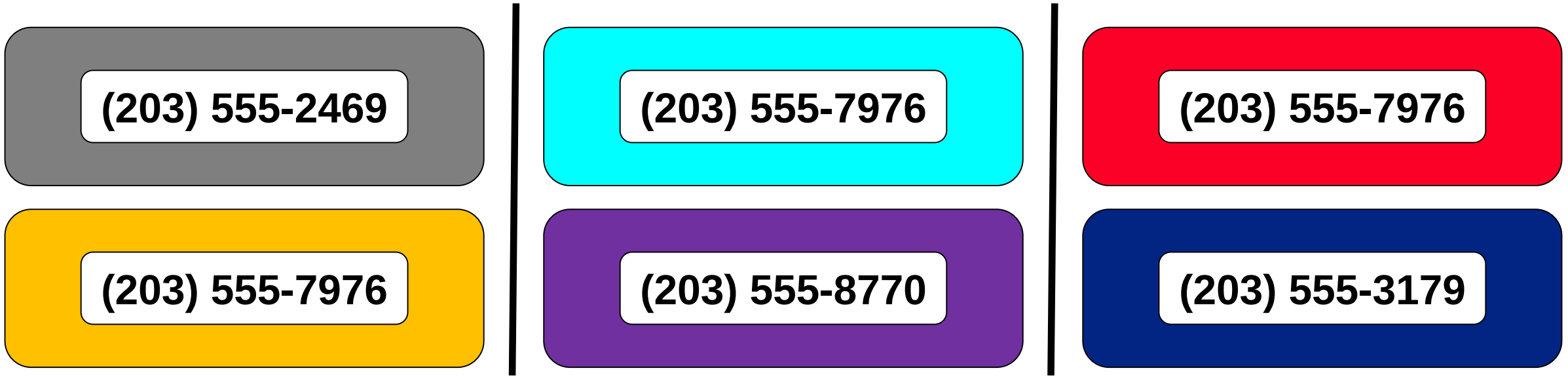
- Each **agency** provides encryption key based on *commutative, public-key*, *randomized* encryption scheme
  - **Commutative** encryption: $Dec_A(Dec_B(c)) = Dec_B(Dec_A(c))$
- **Sources** of phone metadata (telecoms) encrypt each data item using all agencies' keys and give encrypted sets to **repositories**
- 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

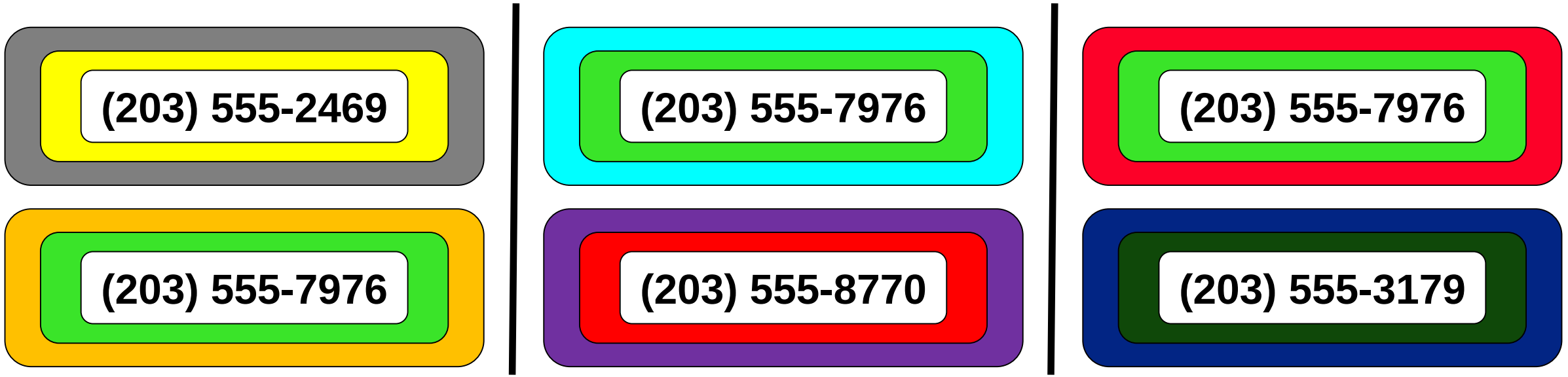
# Private Set Intersection Protocol – Phase 1

- 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



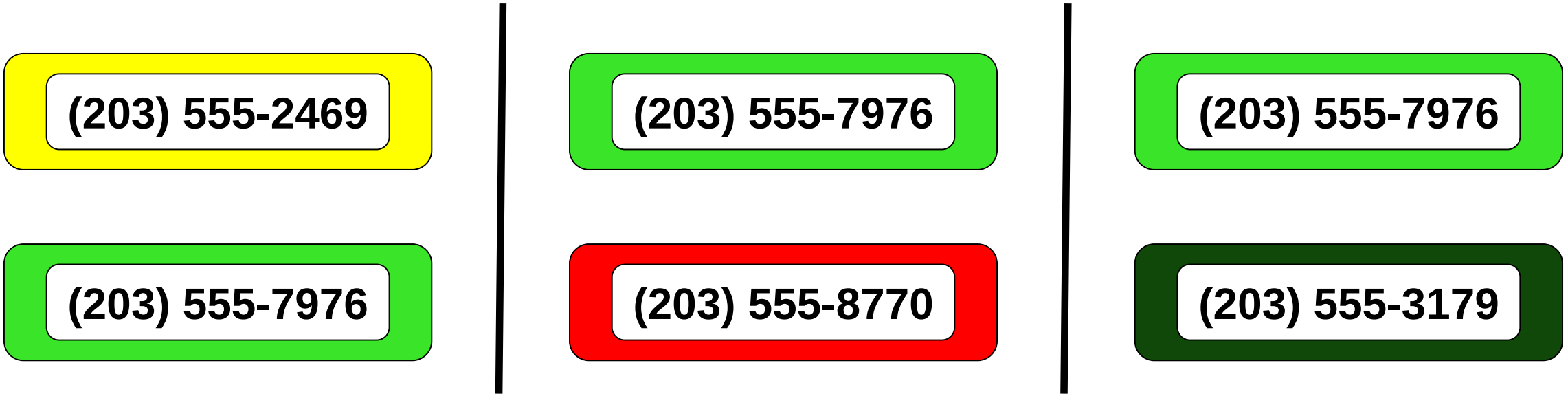
# Private Set Intersection Protocol – Phase 1

- 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



# 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

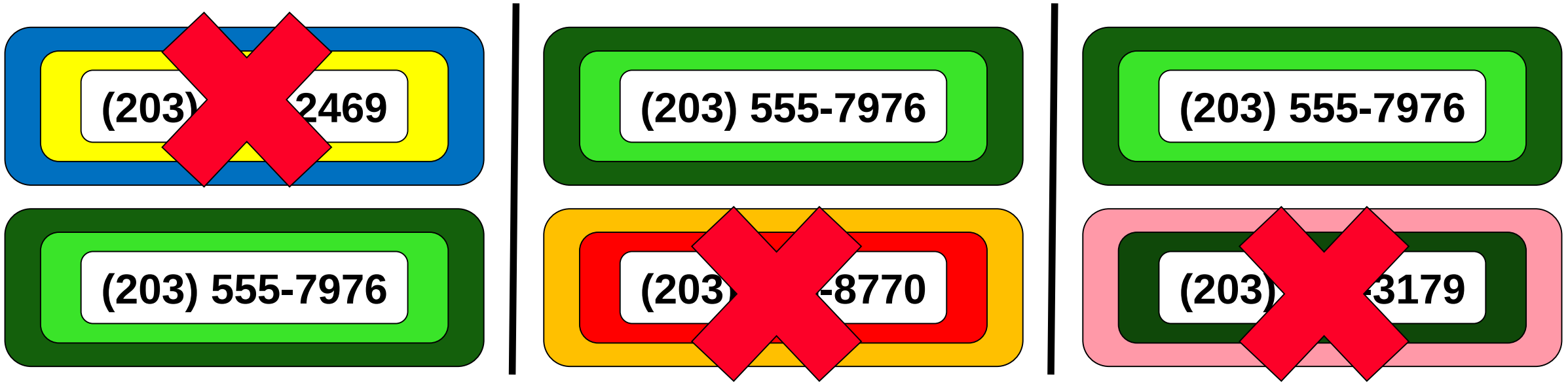
- An agency starts with data sets under *randomized* encryption by all agencies' keys
- Each agency strips off its layer of *randomized* encryption, adds a layer of *deterministic* encryption using its temporary key, permutes the data sets, and sends them to next agency

# Private Set Intersection Protocol – Phase 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

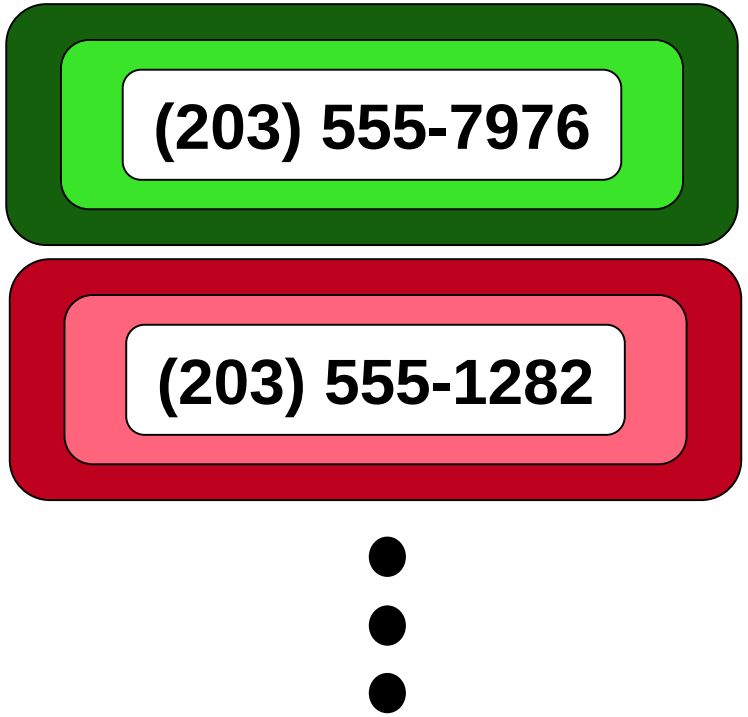
- When phase I is done, each item has encrypted with *deterministic* encryption using *temporary keys*
- Matching ciphertexts = matching plaintexts = targeted users – **keep**
- Non-matching ciphertexts = untargeted users – **discard**

# 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”

# 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

# 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”

https://github.com/DeDiS/Surveillance

# 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)

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

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!