Privacy-Preserving Surveillance

Joan Feigenbaum
Yale University
http://www.cs.yale.edu/homes/jf/

SUNY Global Center; October 13, 2017

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

Privacy-Preserving, Accountable Surveillance

- Identify an unknown target but preserve privacy of untargeted users
  - Collect a large set of encrypted data records (on both targeted and untargeted users), use a cryptographic protocol to winnow it down to just the records of the targets, and then decrypt only those records.

Privacy-Preserving, Accountable Surveillance

- Identify an unknown target but preserve privacy of untargeted users
  - Collect a large set of encrypted data records (on both targeted and untargeted users), use a cryptographic protocol to winnow it down to just the records of the targets, and then decrypt only those records.
- Distributed trust
  - No one agency can compromise privacy.
- Enforced scope limiting
  - No overly broad group of users' data are captured.
- Sealing time and notification
  - After a finite, reasonable time, surveilled users are notified.
- Accountability
  - Surveillance statistics are maintained and audited.

Segal, Ford, & F. Solution in FOCI 2014

- **Privacy-preserving set intersection**
  - Implemented protocol is a variation of Vaidya and Clifton's "secure set-intersection cardinality" protocol [J. Computer Security, 2005].
  - One key technical ingredient is the *mutual commutativity* of the **ElGamal** and **Pohlig-Hellman** encryption schemes:

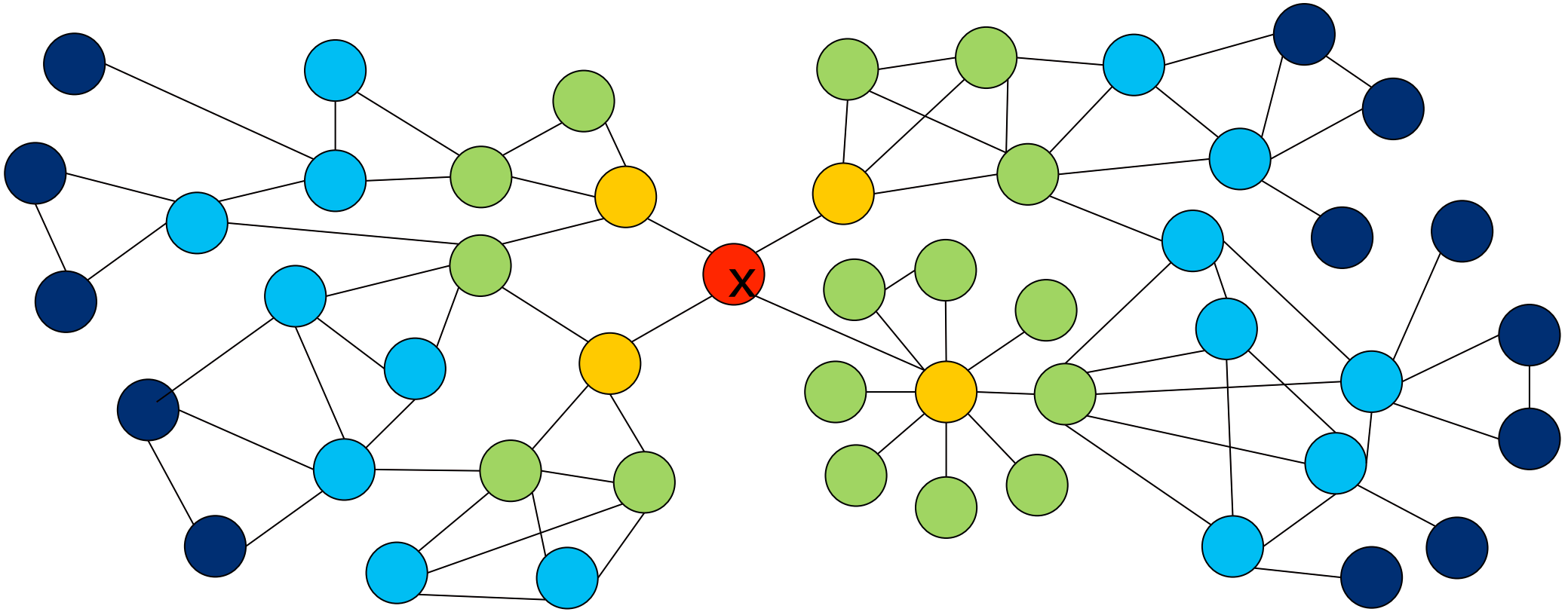
$$D_2(D_3(D_1(E_3(E_2(E_1(x)))))) = x$$

$$D_3(D_2(E_3(D_1(E_2(E_1(x)))))) = x$$

- **Efficient (offline) operation:** Completes 150,000-record instances in 10 minutes.

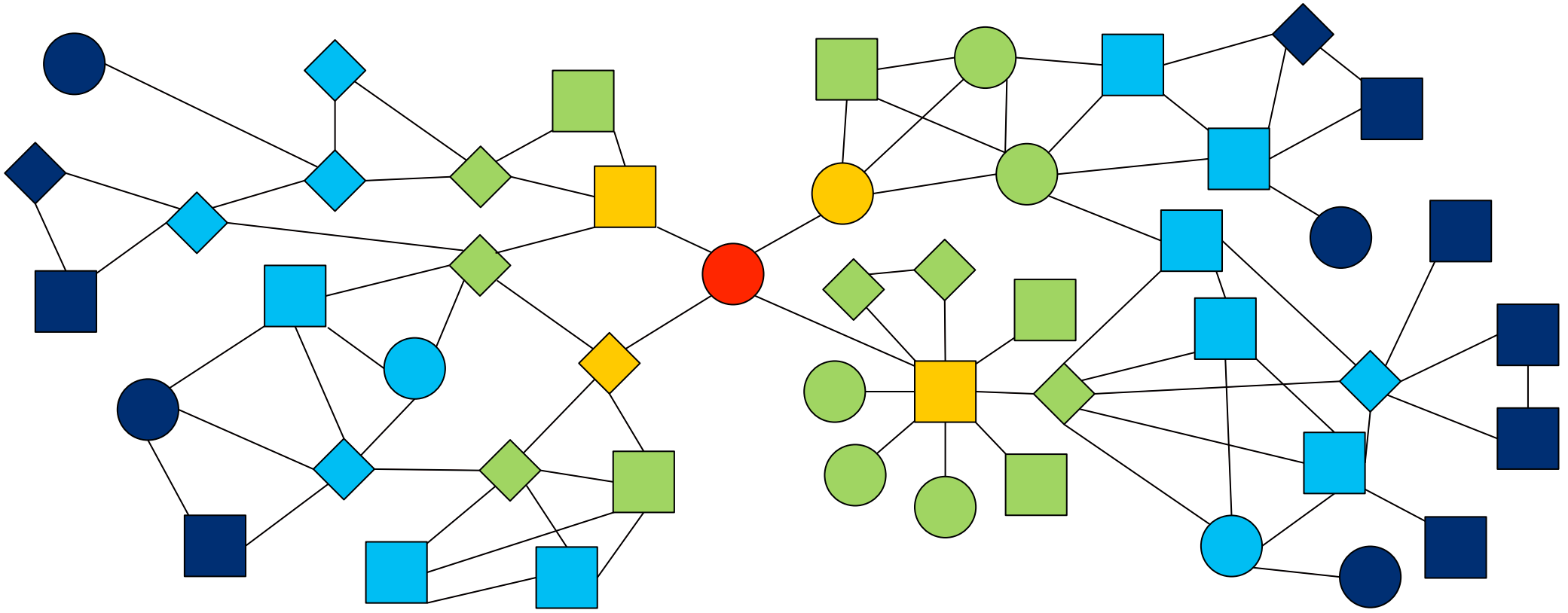
Contact Chaining

- Government knows phone number of target X.
- Goal: Consider the “k-contacts” of X (nodes within distance k).



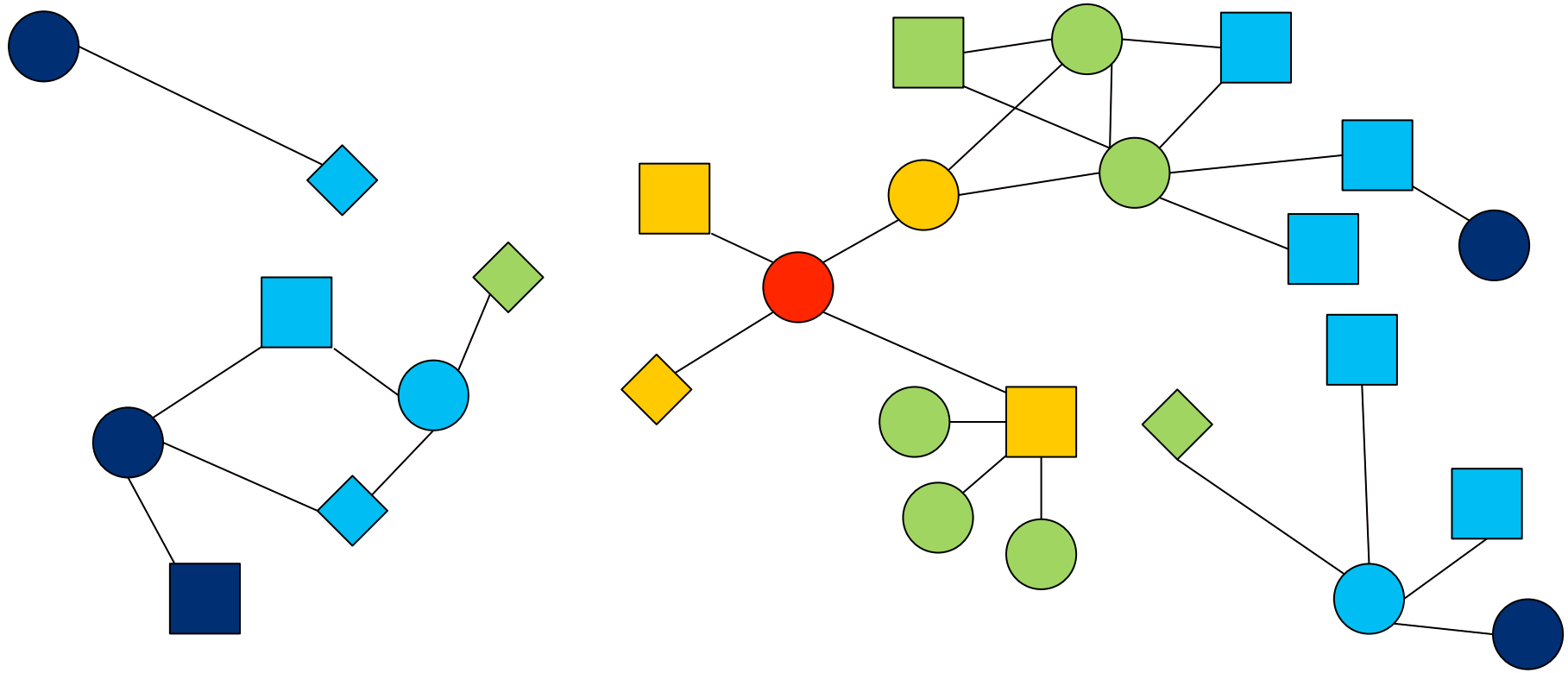
Privacy-Preserving Contact Chaining Goals

- Government learns actionable, relevant intelligence
- Telecommunications companies learn nothing more about other companies' clients



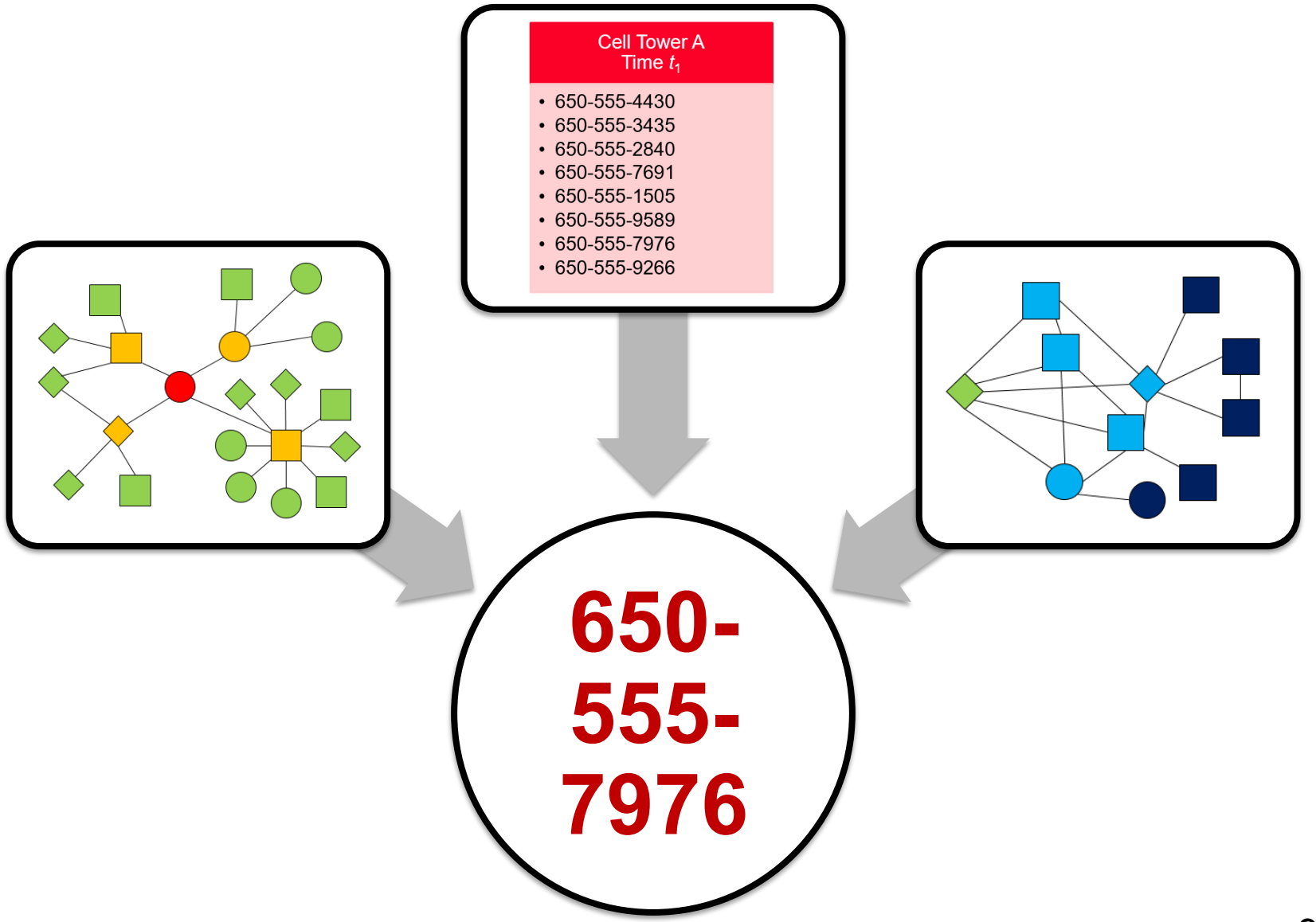
Privacy-Preserving Contact Chaining Goals

- Government learns actionable, relevant intelligence
- Telecommunications companies learn nothing more about other companies' clients



Using Contact Chaining - Main Idea

- Use privacy-preserving contact chaining protocol to get encryptions of *k*-contacts of target
- Use privacy-preserving set intersection to filter *k*-contacts and decrypt only new targets



Segal, F., & Ford Solution in WPES 2016

- Java implementation of a distributed-BFS-based protocol run on the Yale CS Cloud, pipelined into the set-intersection protocol
- Tested on real-network data (**http://snap.stanford.edu/data**)
- Varied
  - the target (starting node) X
  - the chain length k
  - the large-degree cutoff d
- Measured
  - end-to-end running time
  - total CPU time used by all telecoms
  - total amount of data exchanged
- All grew linearly in the number of ciphertexts in the reach set.

Related Work

- Kamara (2014) and Kroll, Felten, Boneh (2014)
  - Cryptographic protocols for privacy-preserving, accountable surveillance of **known** targets
- Kearns, Roth, Wu, Yaroslavtsev (2016)
  - Differentially private, graph-search algorithms for **distinguishing targeted users from untargeted users**
- Ongoing and future work at Yale
  - Multi-layer, attribute-based encryption
  - Privacy-preserving, accountable surveillance of social-network data
  - Privacy-preserving, accountable video surveillance
- Support from funding agencies (since ~ 2011)
  - SPAR (IARPA – PIR)
  - PROCEED and Brandeis (DARPA – PIR, SMC, HE, etc.)
  - HECTOR (IARPA BAA 17-05, Proposals due December 1, 2017)

Wide Range of Negative Reactions

- "Don't be evil": Crypto researchers should aim for "no surveillance."
- "Political infeasibility"
  - LE and IC won't accept distributed trust, scope limits, etc.
  - FISA courts (and other "rubber stamps") won't set meaningful limits or allow notification of targets or statistical reporting.
- "Technical infeasibility"
  - People who seek warrants won't know when these techniques are applicable, won't set appropriate parameters, and won't interpret results correctly.
  - SMPC and similar protocols are too hard to implement and deploy.
- "Lack of generality": Not worth the fixed costs (e.g., data infrastructure)
- "Don't give aid and comfort to the enemy"
  - Justification for bulk collection of encrypted data might be morphed into a justification to backdoor all crypto protocols (because of malice or ignorance).

QUESTIONS?

Back-up Slides

For more information, see:

- A. Segal, J. Feigenbaum, and B. Ford, “Open, privacy-preserving protocols for lawful surveillance,” **http://arxiv.org/abs/1607.03659**.
- A. Segal, B. Ford, and J. Feigenbaum, “Catching Bandits and *Only* Bandits: Privacy-Preserving Intersection Warrants for Lawful Surveillance,” in *Proceedings of the 2014 USENIX Workshop on Free and Open Communications on the Internet* (FOCI).
- A. Segal, J. Feigenbaum, and B. Ford, “Privacy-Preserving Lawful Contact Chaining (Preliminary Report),” in *Proceedings of the 2016 ACM Workshop on Privacy in the Electronic Society* (WPES).
- J. Feigenbaum and B. Ford, “Multiple Objectives of Lawful-Surveillance Protocols,” to appear in *Proceedings of the 2017 International Workshop on Security Protocols* (Cambridge SPW), **http://www.cs.yale.edu/homes/jf/01-feigenbaum-paper.pdf**.

Segal, Ford, & F. Solution in FOCI 2014

- Java implementation of protocol run on Yale CS Cloud
- High Country Bandits example with 50,000 items per set takes less than 11 minutes to complete.
- Note that this is an *offline* process.

Items	Data sent per node (KB)	End-to-End runtime (s)
10	21	1.0
25	46	1.1
50	86	1.3
75	127	1.6
100	167	1.7
250	410	2.9
500	815	4.9
750	1220	6.8
1000	1625	8.2
2500	4055	18.5
5000	8106	36.7
7500	12156	53.6
10000	16206	71.8
25000	40507	229.4
50000	81009	629.4

Table 1: Experimental Results

Implementation of Contact-Chaining Protocol

- Java implementation of protocol run in parallel on Yale CS Cloud
- Used actual network data from a Slovakian social network as “realistic” stand-in for a telephone network
- Created 4 “telecoms” owning 44%, 24%, 17%, and 15% of the network to simulate proportional sizes of largest 4 telecoms

Contact Chaining Experimental Setup

- Java implementation of protocol run in parallel on Yale CS Cloud
- Used actual network data from a Slovakian social network as “realistic” stand-in for a telephone network

Ciphertexts in result	Degree of Target x	Maximum Path Length k	Large Vertex Degree Cutoff d
582	40	2	50
1061	47	2	75
5301	128	2	150
10188	123	2	500
27338	32	3	200
49446	40	3	150
102899	230	3	100
149535	159	3	150
194231	128	3	500
297474	123	3	500

Contact Chaining Experimental Results

- Varied starting position, *k*, and *d* to examine a variety of neighborhood sizes
- Measured
  - End-to-end running time
  - CPU time used by all telecoms
  - Total bandwidth sent over network

Ciphertexts in result	End-to-end runtime MM:SS	Telecom CPU Time H:MM:SS	Bytes transferred MB
582	00:05	0:00:32	18
1061	00:06	0:00:57	6
5301	00:23	0:04:43	22
10188	00:37	0:08:41	36
27338	01:50	0:28:23	132
49446	03:15	0:46:28	222
102899	07:43	1:58:16	804
149535	10:25	2:42:49	896
194231	13:57	3:34:48	978
297474	21:51	5:41:43	1570

Contact Chaining Experimental Results