CSE 484 / CSE M 584: Computer Security and Privacy

Physical Security (recap)
Anonymity

Autumn 2018

Tadayoshi (Yoshi) Kohno
yoshi@cs.Washington.edu

Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, Ada Lerner, John Manferdelli, John Mitchell, Franziska Roesner, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Admin

- Lab 2 out Nov 5, due Nov 20, 4:30pm

- Looking ahead:
- HW 3 out Nov 19, due Nov 30
- Lab 3 out ~Nov 26, due Dec 7 (Quiz Section on Nov 29)

- No class Nov 21; video review assignment instead
  - Counts for class participation that day

Office Hours

- TA Office Hours this week:
  - Monday, 12-1pm, 5th floor breakout
  - Monday, 2:30-3:30pm, 4th floor breakout
  - Tuesday, 3-4pm, 4th floor breakout
- I still have office hours after class, but might be ~10 mins late

Admin

- Final Project Proposals: We are looking at them this week
- Final Project Checkpoint: Nov 30 – preliminary outline and references
- Final Project Presentation: Dec 10 – 12-15-minute video – **must** be on time
- Explore something of interest to you, that could hopefully benefit you or your career in some way – technical topics, current events, etc

Earlence’s Research

Voice Assistant Security Project

- Build and evaluate a system to defend against voice-confusion attacks
  - Hey Alexa, “ask capital won to …”
  - Hey Alexa, “ask capital one to … ”
- 1 position available for an undergrad (with funding starting next quarter)
  - Should have experience in working with programming languages
  - Preferable: taken courses on systems, and security
  - Self-motivated
- This can be a valuable experience for students who want to enter grad school, or those who want to evaluate whether research is for them
- Contact earlence@cs.washington.edu with transcript, CV
- earlence.com

General Link for Security & Privacy Research: http://goo.gl/forms/sD40kxIXM6

Physical Security and Digital Security

Connecting Ideas…

- Defense in Depth
  - Layers (safes in banks, etc.)
- Deterrents:
  - Home alarm systems
  - Video cameras (forensic trails)

Snake Oil

- Appearance of security may not equal security
- Many computer systems claim to provide a high level of security, when in fact they do not
- Similarly, some locks advertise themselves as being very secure, when in fact they are easy to circumvent

Denial of Service

- Door locks also subject to denial of service attacks
  - Break a (wrong) key in someone's door
  - Or gum
  - Or super glue
- Double-sided locks

One Size Doesn’t Fit All

- Different locks suitable for different purposes
  - Gym locker
  - Car
  - Bank vault
  - Nuclear missiles
  - …

There Exist Different Adversaries

- An outsider
- An (ex-)employee or previous tenant (who had a key)
- An insider (someone who makes the locks, keys the locks, or has a master key)

Electronic World

- Physical world:
  - Not a high degree of connectedness
  - (Yes, there’s exceptions, but generally …)
- Digital world:
  - Everyone can be everyone else’s “next door” neighbor
  - More potential for anonymity

Anonymity

Privacy on Public Networks

- Internet is designed as a public network
  - Machines on your LAN may see your traffic, network routers see all traffic that passes through them
- Routing information is public
  - IP packet headers identify source and destination
  - Even a passive observer can easily figure out who is talking to whom
- Encryption does not hide identities
  - Encryption hides payload, but not routing information
  - Even IP-level encryption (tunnel-mode IPSec/ESP) reveals IP addresses of IPSec gateways

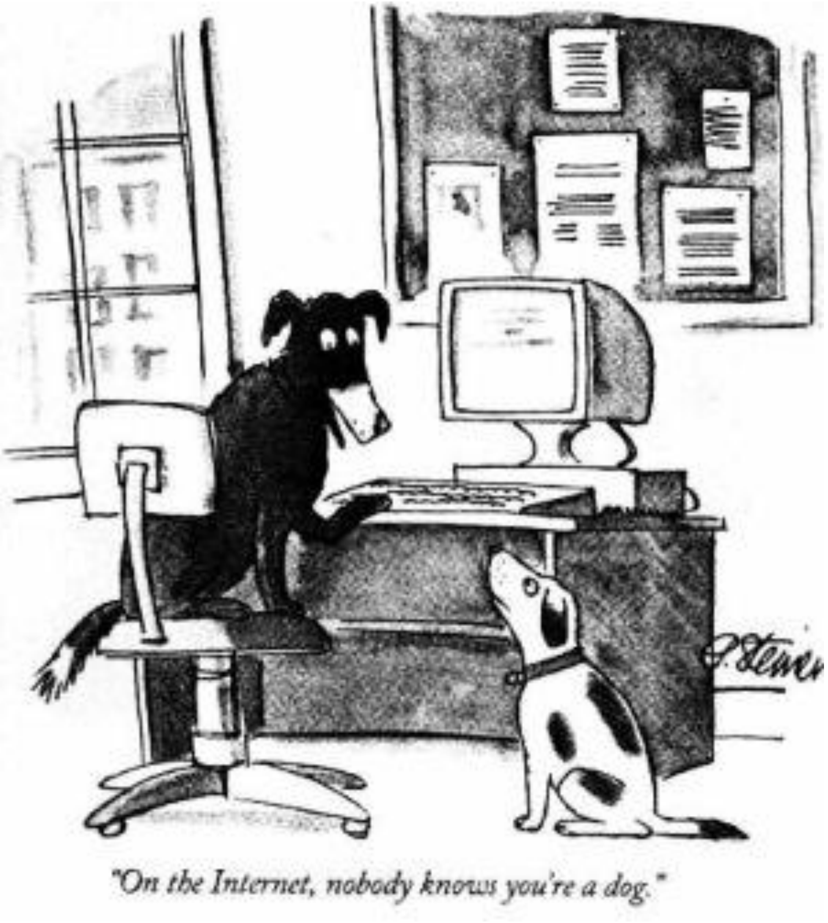
Questions

Q1: What is anonymity?

Q2: Why might people want anonymity on the Internet?

Q3: Why might people **not** want anonymity on the Internet?

Famous Cartoon – Is it True?



Applications of Anonymity (I)

- Privacy
  - Hide online transactions, Web browsing, etc. from intrusive governments, marketers, parents
- Untraceable electronic mail
  - Corporate whistle-blowers
  - Political dissidents
  - Socially sensitive communications (e.g., support groups)
  - Confidential business negotiations
- Law enforcement and intelligence
  - Sting operations and honeypots
  - Secret communications on a public network

Applications of Anonymity (II)

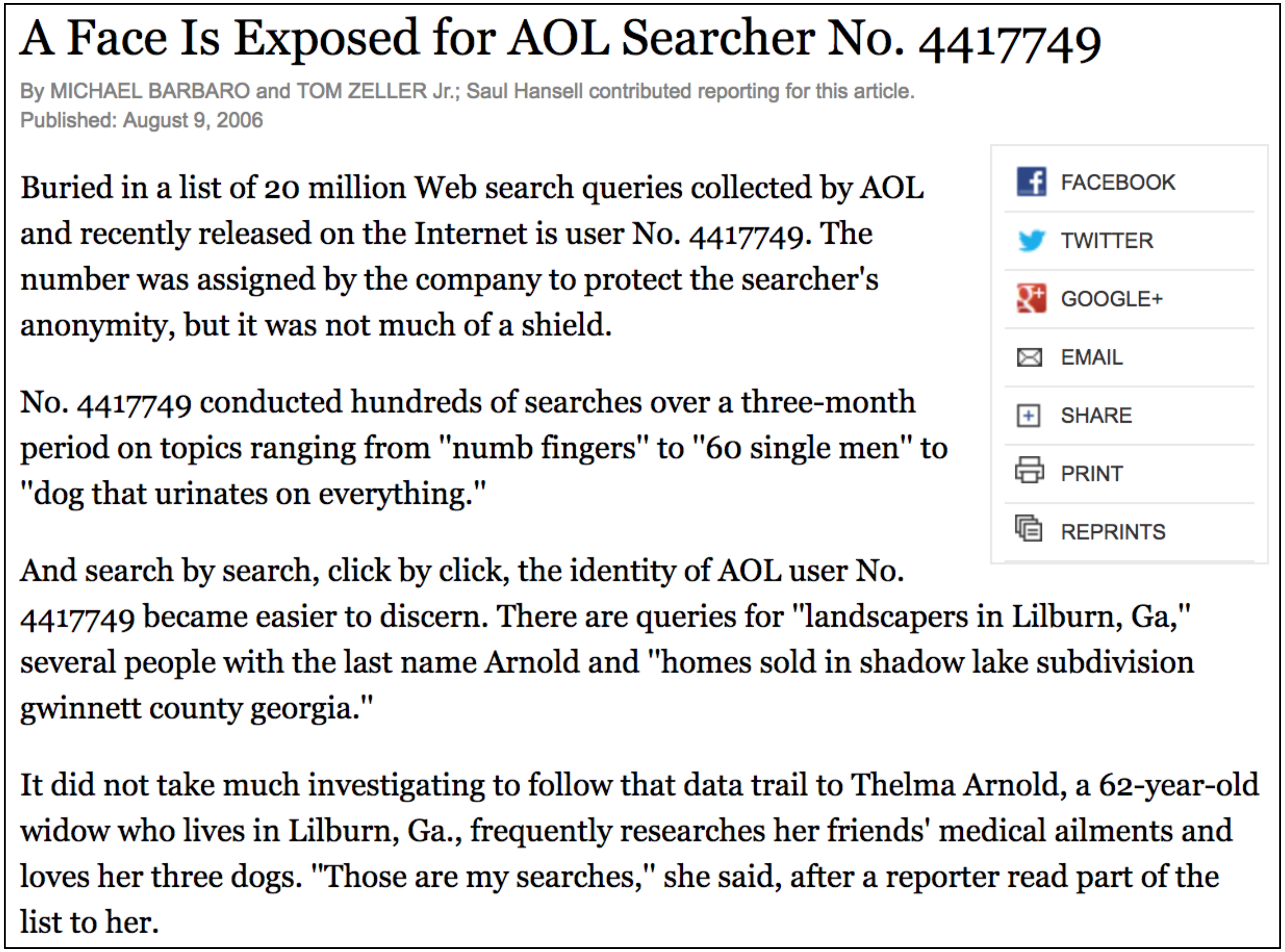
- Digital cash (from 1980s, but also modern crypto currencies like Zcash)
  - Electronic currency with properties of paper money (online purchases unlinkable to buyer's identity)
- Anonymous votes for electronic voting
- Censorship-resistant publishing

What is Anonymity?

- Anonymity is the state of being not identifiable within a set of subjects
  - You cannot be anonymous by yourself!
    - Big difference between anonymity and confidentiality
  - Hide your activities among others' similar activities
- Unlinkability of action and identity
  - For example, sender and email he/she sends are no more related after observing communication than before
- Unobservability (hard to achieve)
  - Observer cannot even tell whether a certain action took place or not

Part 1: Anonymity in Datasets

How to release an anonymous dataset?



A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.; Saul Hansell contributed reporting for this article.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

How to release an anonymous dataset?

- Possible approach: remove identifying information from datasets?

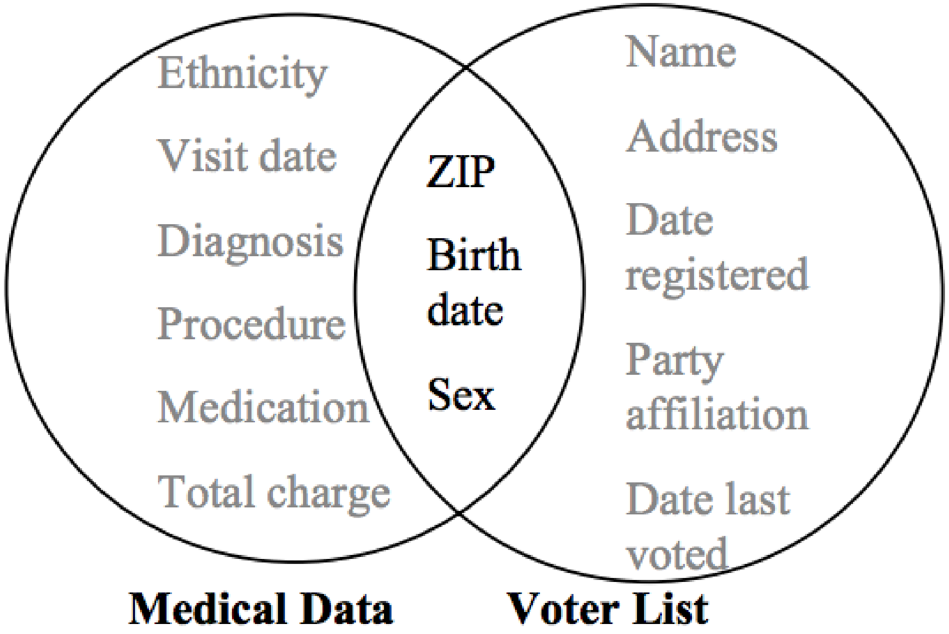
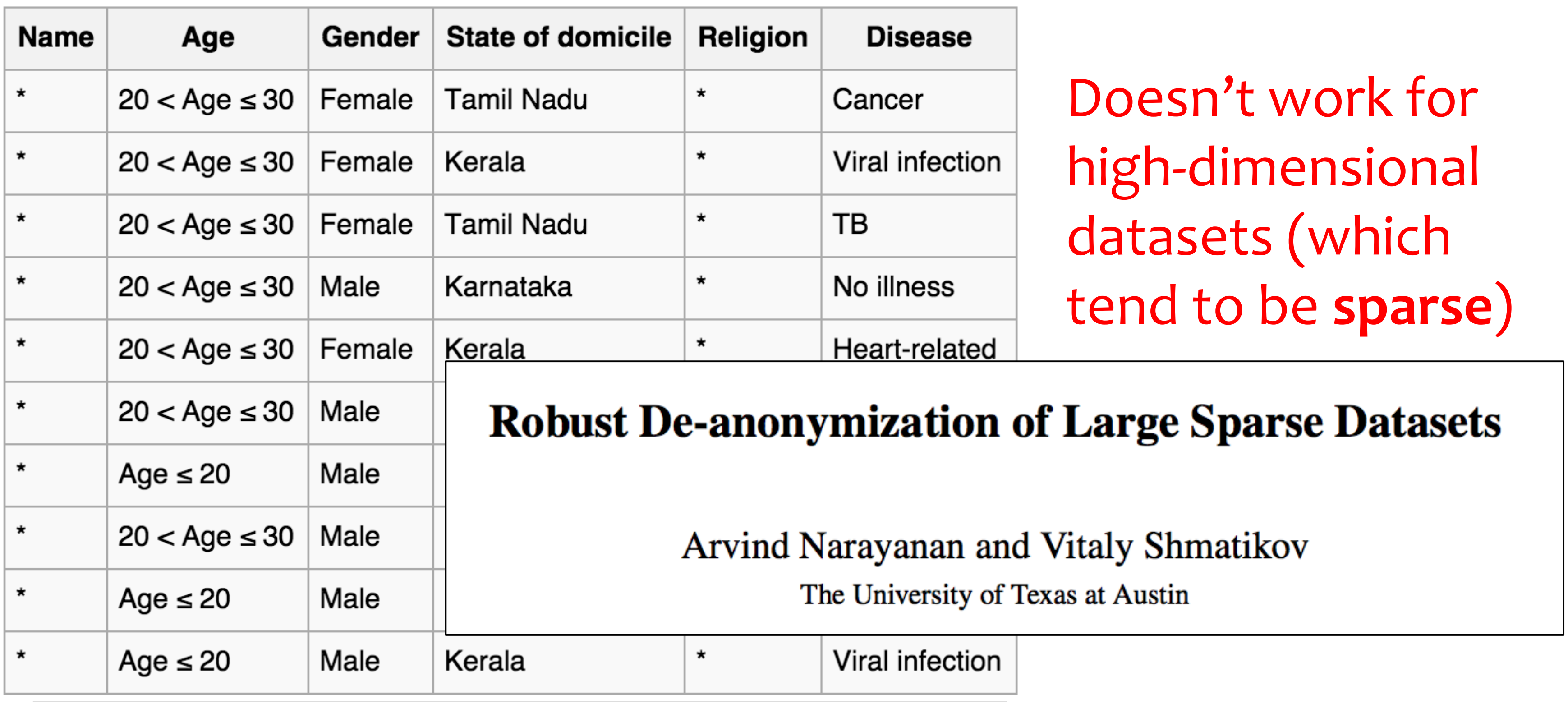


Figure 1 Linking to re-identify data

Massachusetts medical+voter data [Sweeney 1997]

k-Anonymity

- Each person contained in the dataset cannot be distinguished from at least k-1 others in the data.



Name	Age	Gender	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	Tamil Nadu	*	TB
*	20 < Age ≤ 30	Male	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male			
*	Age ≤ 20	Male			
*	20 < Age ≤ 30	Male			
*	Age ≤ 20	Male			
*	Age ≤ 20	Male	Kerala	*	Viral infection

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Doesn't work for high-dimensional datasets (which tend to be **sparse**)

[Dwork et al.]

Differential Privacy

- **Setting:** Trusted party has a database
- **Goal:** allow queries on the database that are useful but preserve the privacy of individual records
- **Differential privacy intuition:** add noise so that an output is produced with similar probability whether any single input is included or not
- Privacy of the computation, not of the dataset

Part 2: Anonymity in Communication

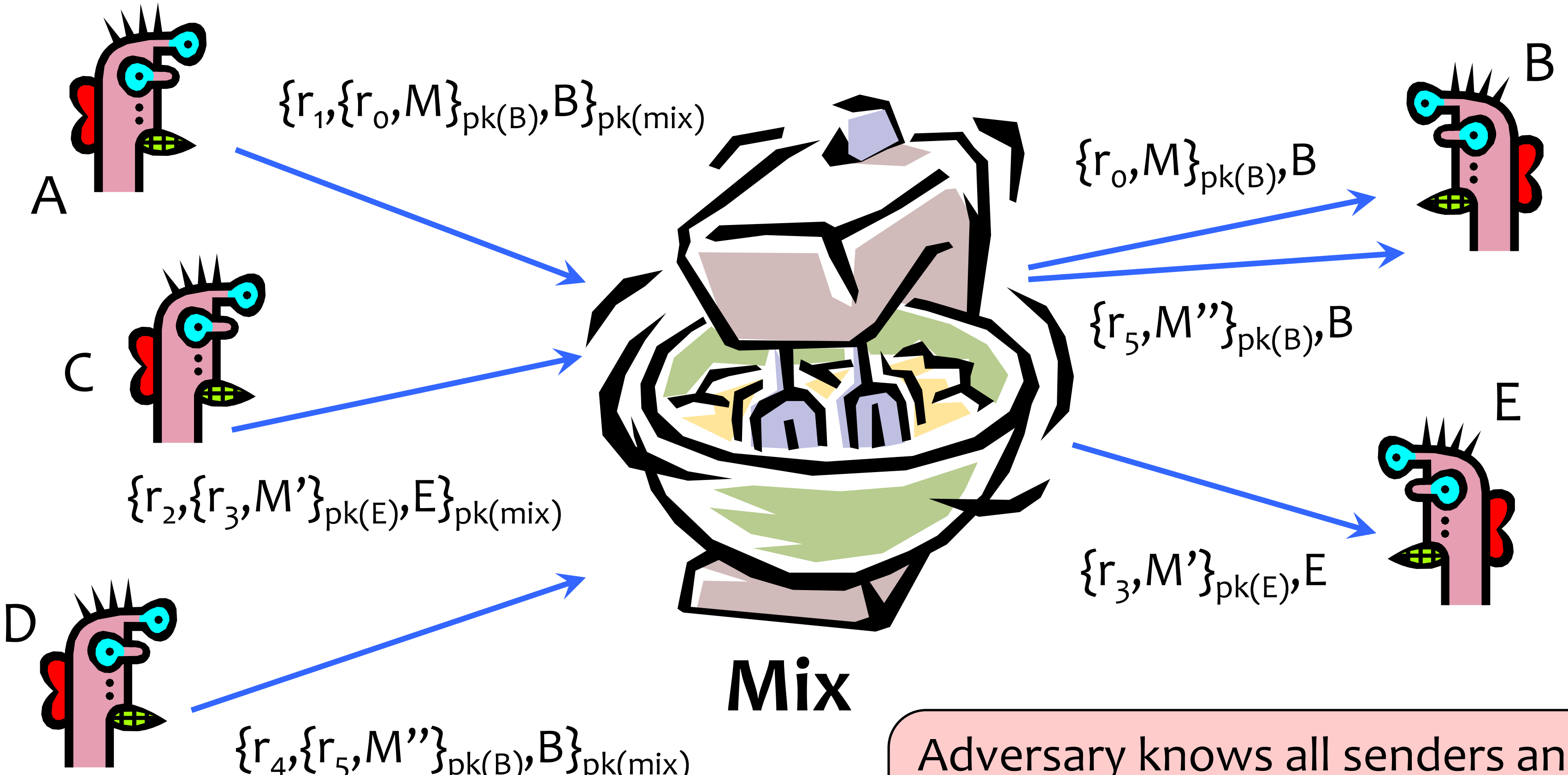
Chaum’s Mix

- Early proposal for anonymous email
  - David Chaum. “Untraceable electronic mail, return addresses, and digital pseudonyms”. Communications of the ACM, February 1981.

Before spam, people thought anonymous email was a good idea ☺

- Public key crypto + trusted re-mailer (Mix)
  - Untrusted communication medium
  - Public keys used as persistent pseudonyms
- Modern anonymity systems use Mix as the basic building block

Basic Mix Design



A

$\{r_1,\{r_0,M\}_{pk(B)},B\}_{pk(mix)}$

C

$\{r_2,\{r_3,M'\}_{pk(E)},E\}_{pk(mix)}$

D

$\{r_4,\{r_5,M''\}_{pk(B)},B\}_{pk(mix)}$

Mix

$\{r_0,M\}_{pk(B)},B$

B

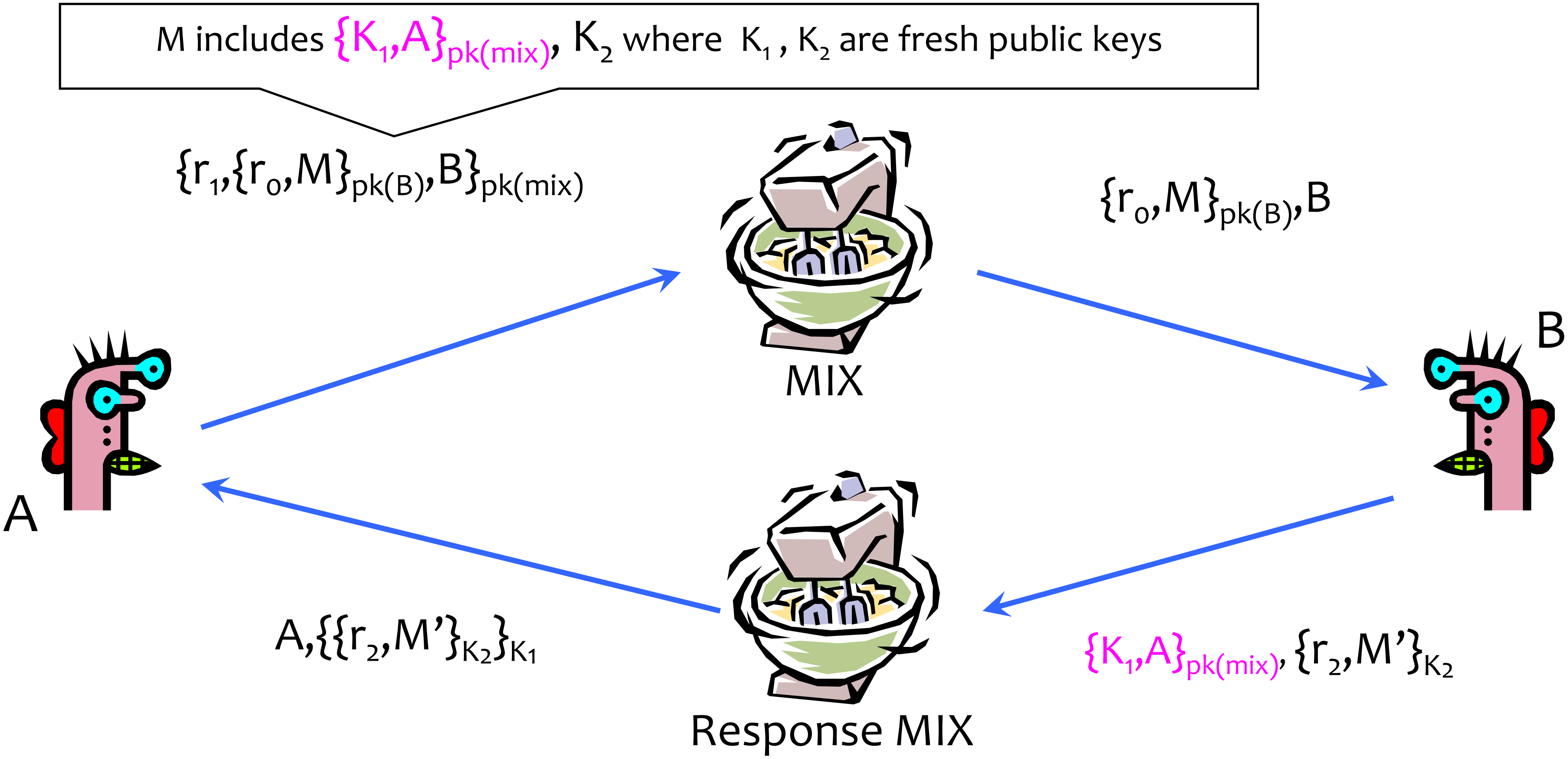
$\{r_5,M''\}_{pk(B)},B$

E

$\{r_3,M'\}_{pk(E)},E$

Adversary knows all senders and all receivers, but cannot link a sent message with a received message

Anonymous Return Addresses



M includes $\{K_1,A\}_{pk(mix)}$, K_2 where K_1, K_2 are fresh public keys

$\{r_1,\{r_0,M\}_{pk(B)},B\}_{pk(mix)}$

MIX

$\{r_0,M\}_{pk(B)},B$

A

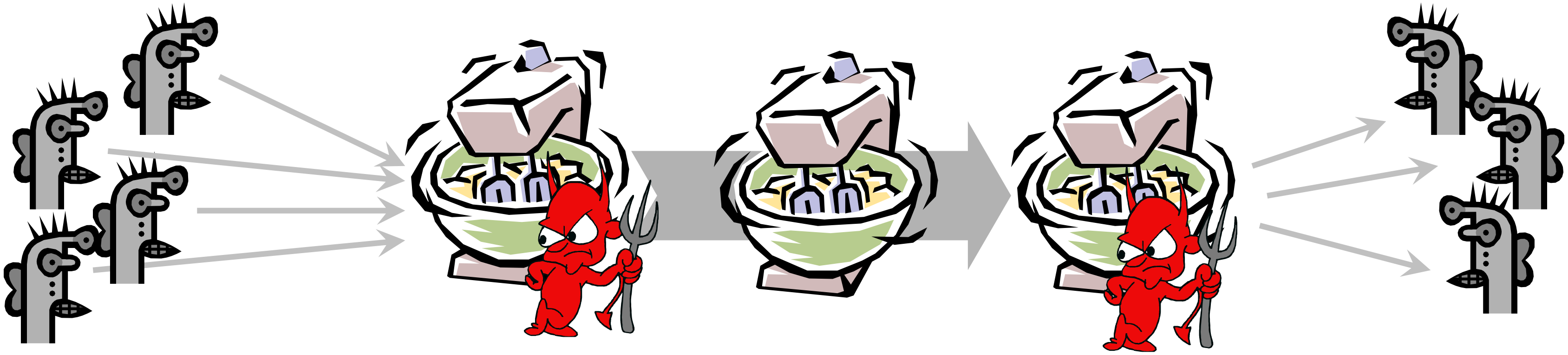
B

$A,\{\{r_2,M'\}_{K_2}\}_{K_1}$

Response MIX

$\{K_1,A\}_{pk(mix)}, \{r_2,M'\}_{K_2}$

Mix Cascades and Mixnets

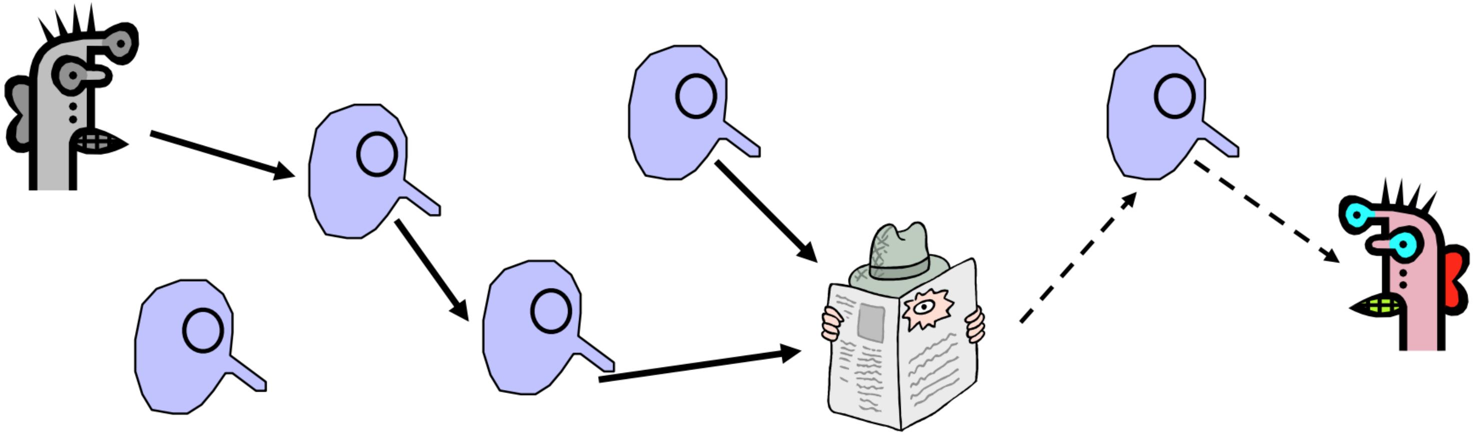


- Messages are sent through a sequence of mixes
  - Can also form an arbitrary network of mixes ("mixnet")
- Some of the mixes may be controlled by attacker, but even a single good mix ensures anonymity
- Pad and buffer traffic to foil correlation attacks

Disadvantages of Basic Mixnets

- Public-key encryption and decryption at each mix are computationally expensive
- Basic mixnets have high latency
  - OK for email, not OK for anonymous Web browsing
- Challenge: low-latency anonymity network

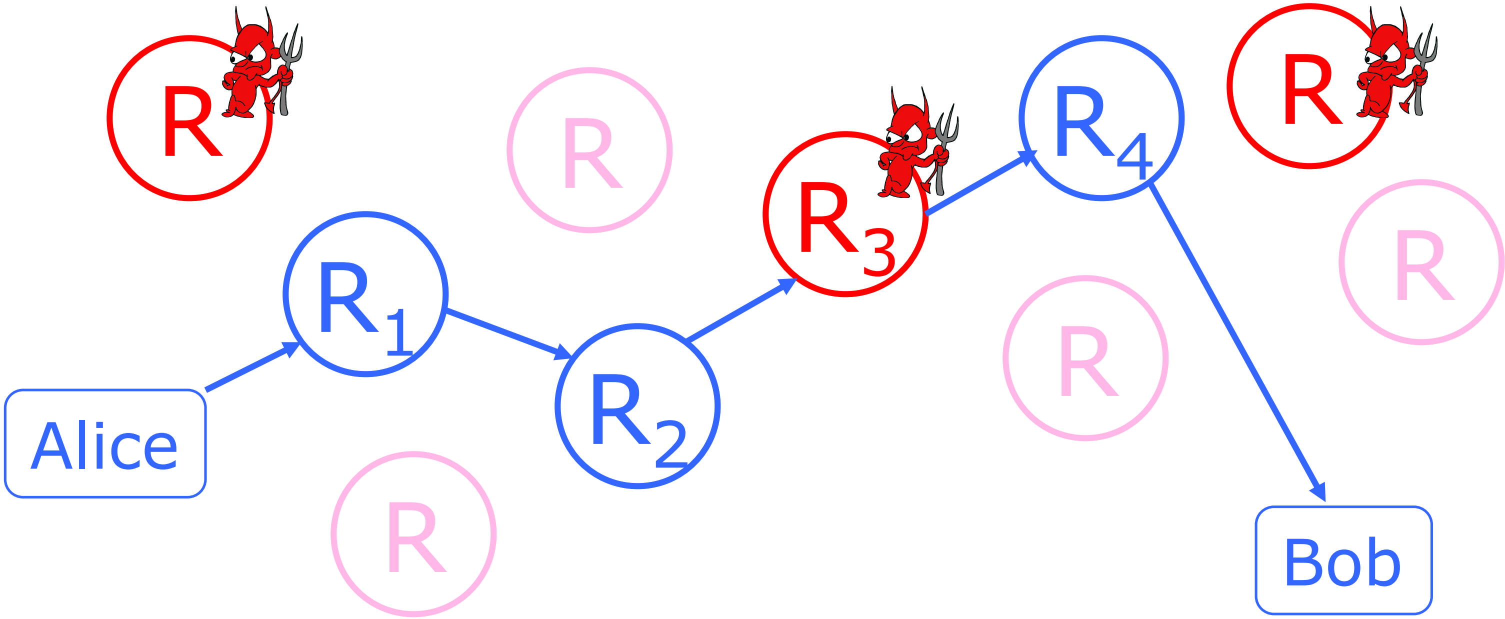
Another Idea: Randomized Routing



- Hide message source by routing it randomly
  - Popular technique: Crowds, Freenet, Onion routing
- Routers don't know for sure if the apparent source of a message is the true sender or another router

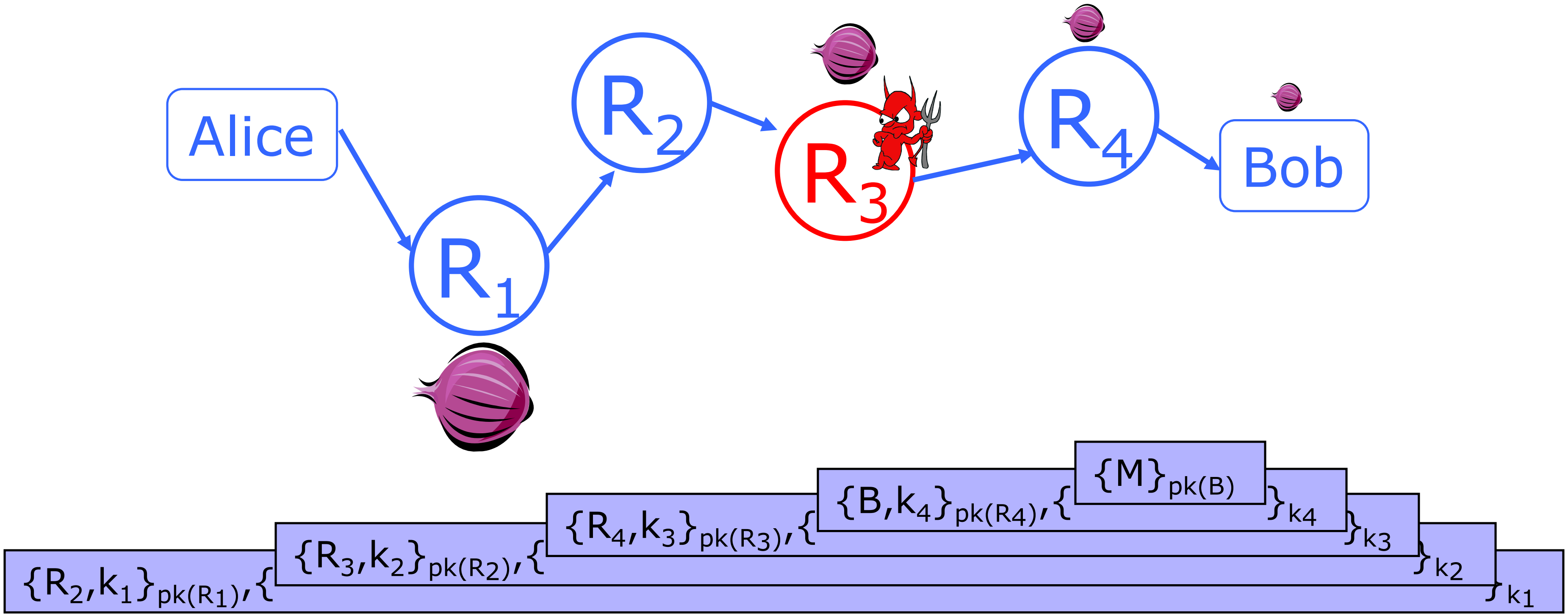
[Reed, Syverson, Goldschlag 1997]

Onion Routing



- Sender chooses a random sequence of routers
  - Some routers are honest, some controlled by attacker
  - Sender controls the length of the path

Route Establishment



$\{R_2,k_1\}_{pk(R_1)},\{\{R_3,k_2\}_{pk(R_2)},\{\{R_4,k_3\}_{pk(R_3)},\{\{B,k_4\}_{pk(R_4)},\{\{M\}_{pk(B)}\}_{k_4}\}_{k_3}\}_{k_2}\}_{k_1}$

- Routing info for each link encrypted with router’s public key
- Each router learns only the identity of the next router