

I8734: Foundations of Privacy

Course Overview

Anupam Datta
CMU
Fall 2014

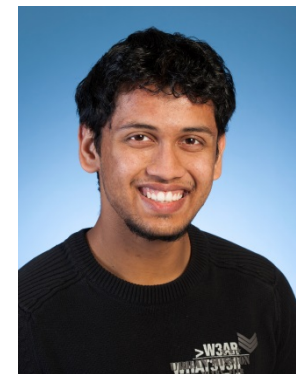
Logistics



Course Staff

- ▶ **Instructor: Anupam Datta**
 - ▶ Office hours: CIC 2118
 - ▶ Email: danupam@cmu.edu
 - ▶ Office hours: Mon 2-3PM EDT
 - ▶ Skype for SV students: anupam.datta2

- ▶ **TA: Amit Datta**
 - ▶ Office hours: CIC 2214
 - ▶ Email: amitdatta@cmu.edu
 - ▶ Office hours: Fri 1:30-2:30PM EDT
 - ▶ Skype for SV students: datta.amit



Extra office hours on demand

Logistics

- ▶ Lectures: Monday & Wednesday, 4:30-6:20 PM EDT (usually 80 minutes)
- ▶ Recitation: Friday 12:30-1:20pm EDT (attend!)
- ▶ Web page: <http://www.ece.cmu.edu/~ece734/>
- ▶ Course blackboard (for grades)
- ▶ Piazza (for all other communication)
 - ▶ Please enroll; you should have received invitation
- ▶ Course work and grading:
 - ▶ Homework (60%) – 4 x 15%
 - ▶ Best 4 of 5 homeworks
 - ▶ Course project (30%)
 - ▶ Class participation (10%)

Logistics (2)

- ▶ **Course Project:**
 - ▶ Teams of 2 (form team by end of week)
 - ▶ Project proposal: 1-2 pages (Due: Oct 1)
 - ▶ Deliverable Part I (Due: November 7)
 - ▶ Deliverable Part II (Due: Dec 1)
 - ▶ Written report: 5-10 pages (Due: Dec 1)
 - ▶ Final presentation in class (Dec 1)

Logistics (3)

Collaboration policy:

- ▶ You are allowed to discuss homework problems and approaches for their solution with other students in the class, but are required to figure out and write out detailed solutions independently and to acknowledge any collaboration or other source

[CMU Computing Policy](#)

[CMU Academic Integrity Policy](#)

Logistics (4)

Example Violations:

- ▶ Submission of work completed or edited in whole or in part by another person.
- ▶ Supplying or communicating unauthorized information or materials, including graded work and answer keys from previous course offerings, in any way to another student.
- ▶ Use of unauthorized information or materials, including graded work and answer keys from previous course offerings.
- ▶ ...not exhaustive list

If in doubt, ask me!

Logistics (5)

- ▶ **Pittsburgh class in HH D210 on**
 - ▶ Sept 10 (this Wed!)
 - ▶ Oct 15
 - ▶ Nov 12
- (Noted on class schedule online)

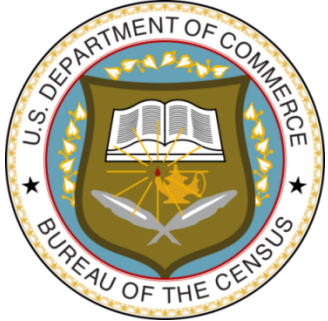
Logistics (6)

- ▶ Consent form for video recording lectures

Prerequisites

- ▶ An undergraduate course equivalent to 15-251 is required or permission of instructor
- ▶ An introductory course in computer security such as 18-487, 18-630, or 18-730 is recommended, but not required
- ▶ If in doubt, please talk to me after class
- ▶ Quick class poll

Personal Information is Everywhere



Google

facebook



amazon.com



flickr® from YAHOO!

Organizing Questions


- ▶ **What is privacy?**
 - ▶ From philosophical and legal conceptions to computer science and engineering
 - ▶ Inspiration from conceptions, but greater precision often through greater specificity

- ▶ **How can we protect privacy?**
 - ▶ Beyond creating laws and institutions
 - ▶ Computational mechanisms for privacy protection

Privacy Problems

Module I: Privacy through Accountability

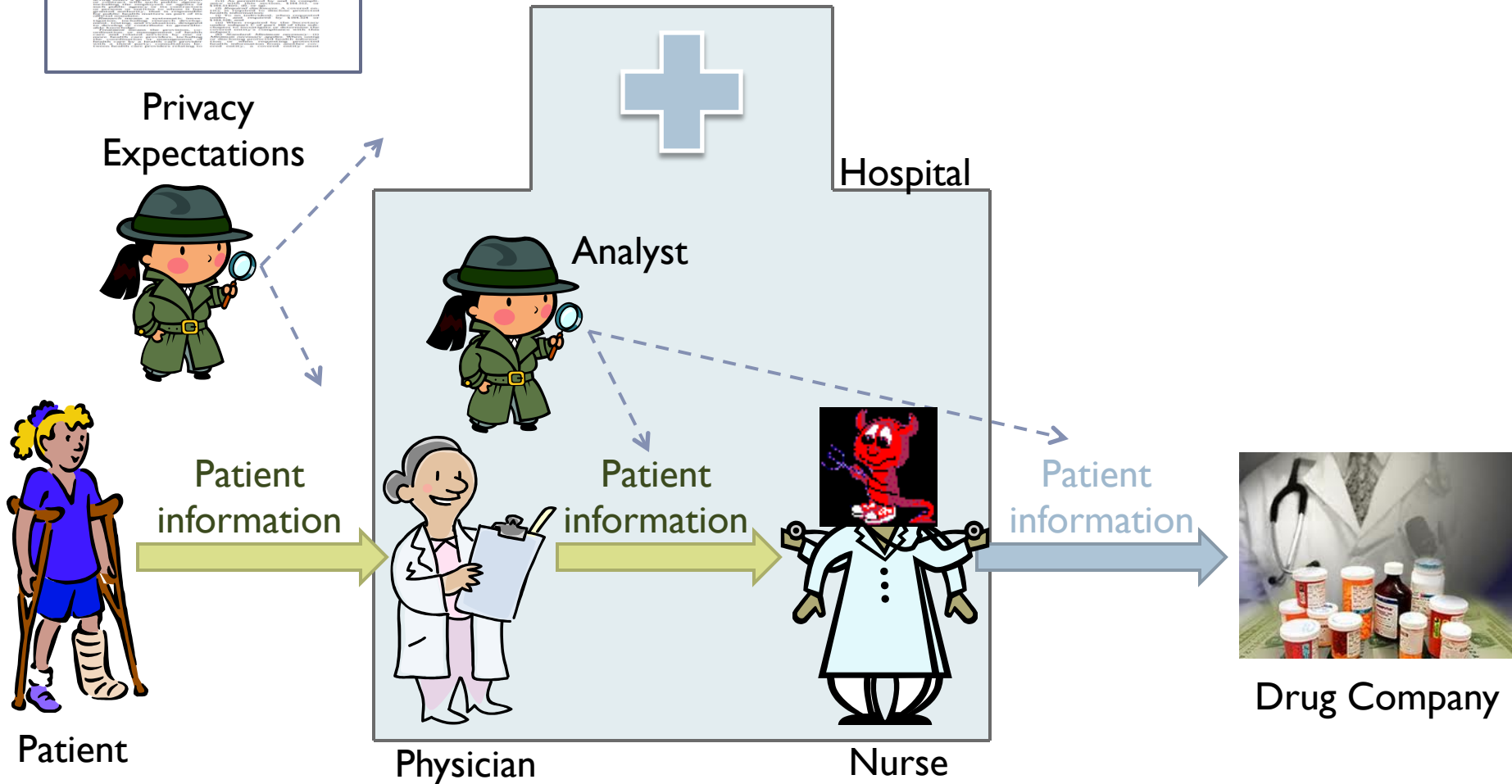
Research Challenge



Programs and People

Ensure organizations respect privacy expectations in the collection, use, and disclosure of personal information

Healthcare Privacy



Example from HIPAA Privacy Rule

A covered entity may disclose an individual's protected health information (phi) to law-enforcement officials for the purpose of identifying an individual if the individual made a statement admitting participating in a violent crime that the covered entity believes may have caused serious physical harm to the victim

▶ Concepts in privacy policies

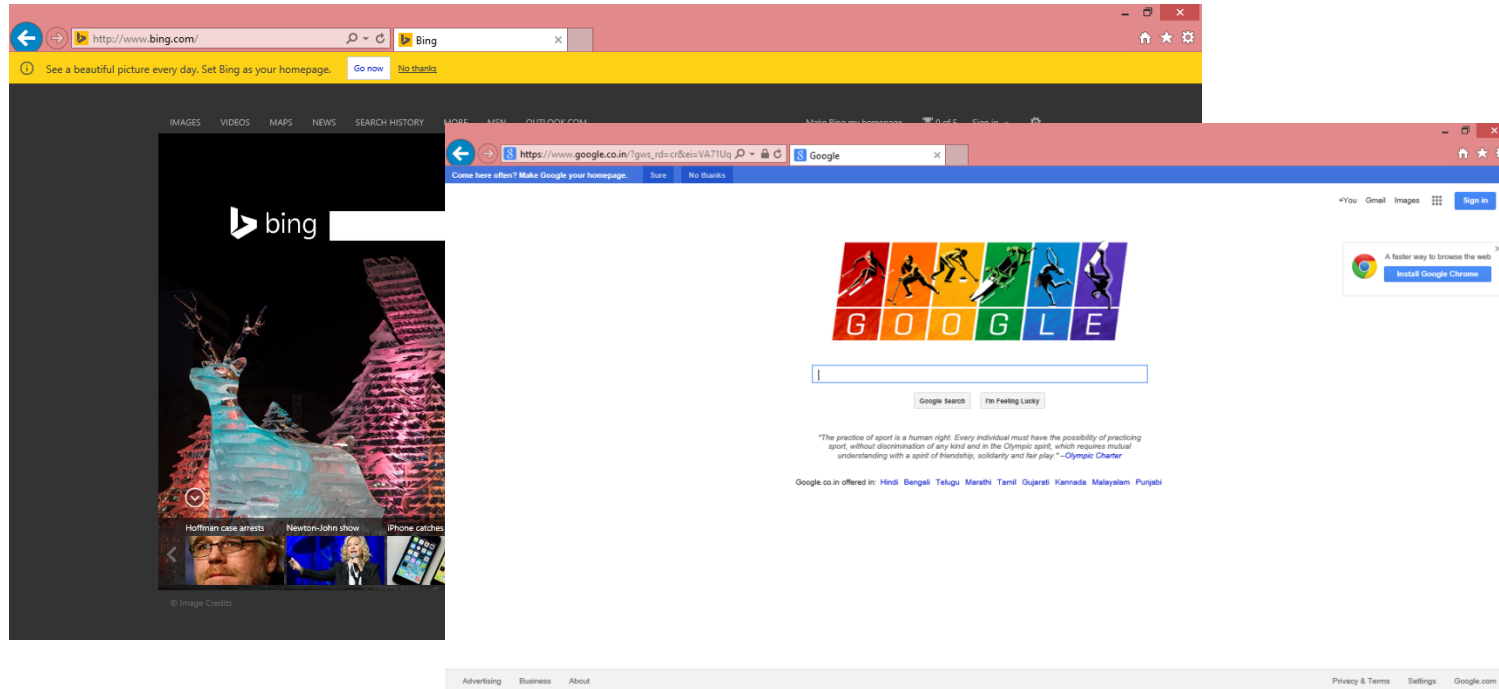
- ▶ **Actions:** send(p1, p2, m)
- ▶ **Roles:** inrole(p2, law-enforcement)
- ▶ **Data attributes:** attr_in(prescription, phi)
- ▶ **Temporal constraints:** in-the-past(state(q, m))

- ▶ **Purposes:** purp_in(u, id-criminal))
- ▶ **Beliefs:** believes-crime-caused-serious-harm(p, q, m)

Black-and-white concepts

Grey concepts

Web Advertising



Example privacy policies:

- ▶ Not use detailed location (full IP address) for advertising
- ▶ Not use health information for advertising

Privacy Compliance for Bing

The screenshot shows a browser window with two tabs. The first tab is the Bing homepage (http://www.bing.com/), featuring a search bar, navigation links (IMAGES, VIDEOS, MAPS, NEWS, SEARCH HISTORY, MORE), and a cartoon detective character. The second tab is the Bing Privacy Statement page (http://www.microsoft.com/privacystatement/en-us/b...), which contains the following content:

Bing Privacy Statement

This privacy statement applies to Bing websites, services, products and applications that collect data and display these terms. It does not apply to other Microsoft products and services that do not link to the Bing Privacy Statement.

Cookies & Similar Technologies

When you use Bing services with a web browser, we will place one or more "cookies" on your machine. For example, Bing uses a cookie with a unique identifier known as the Search ID to operate the service and enable certain search features. If you sign into Bing or other Microsoft services using a Microsoft account, we will set or read one or more additional cookies. We use these cookies to operate Bing services and provide you a more relevant search experience. You can use your browser settings to remove or block cookies on your computer.

We also use web beacons to help deliver cookies and compile analytics. These may include third-party web beacons, which are prohibited from collecting your personal information.

[Learn More](#)

Collecting Your Information

When you use Bing services, Microsoft may collect many kinds of information in order to operate effectively and provide you the best products, services and experiences we can. We collect information when you register, sign in and use our sites and services. We also may get information from other companies. We collect this information in a variety of ways, including from web forms, technologies like cookies, web logging and software on your computer or other device.

When you conduct a search, Microsoft collects the following:

- Search term and time and date of your search
- IP address, browser configuration and approximate location
- Any unique identifiers contained in the cookies

We store search terms (and the cookie IDs associated with search terms) separately from any account information that directly identifies the user, such as name, e-mail address, or phone numbers. We have technological safeguards in place designed to prevent the unauthorized correlation of this data and we remove the entirety of the IP address after 6 months, cookies and other cross session identifiers, after 18 months.

Bing provides search services to select partners and its users. Some examples include Yahoo! and Nokia. In order to provide these services, Bing services receive certain search related information from these partners that may include date, time, IP address, a unique identifier and other search related data.

[*Top of page](#)

[Learn More](#)

How We Use Your Personal Information

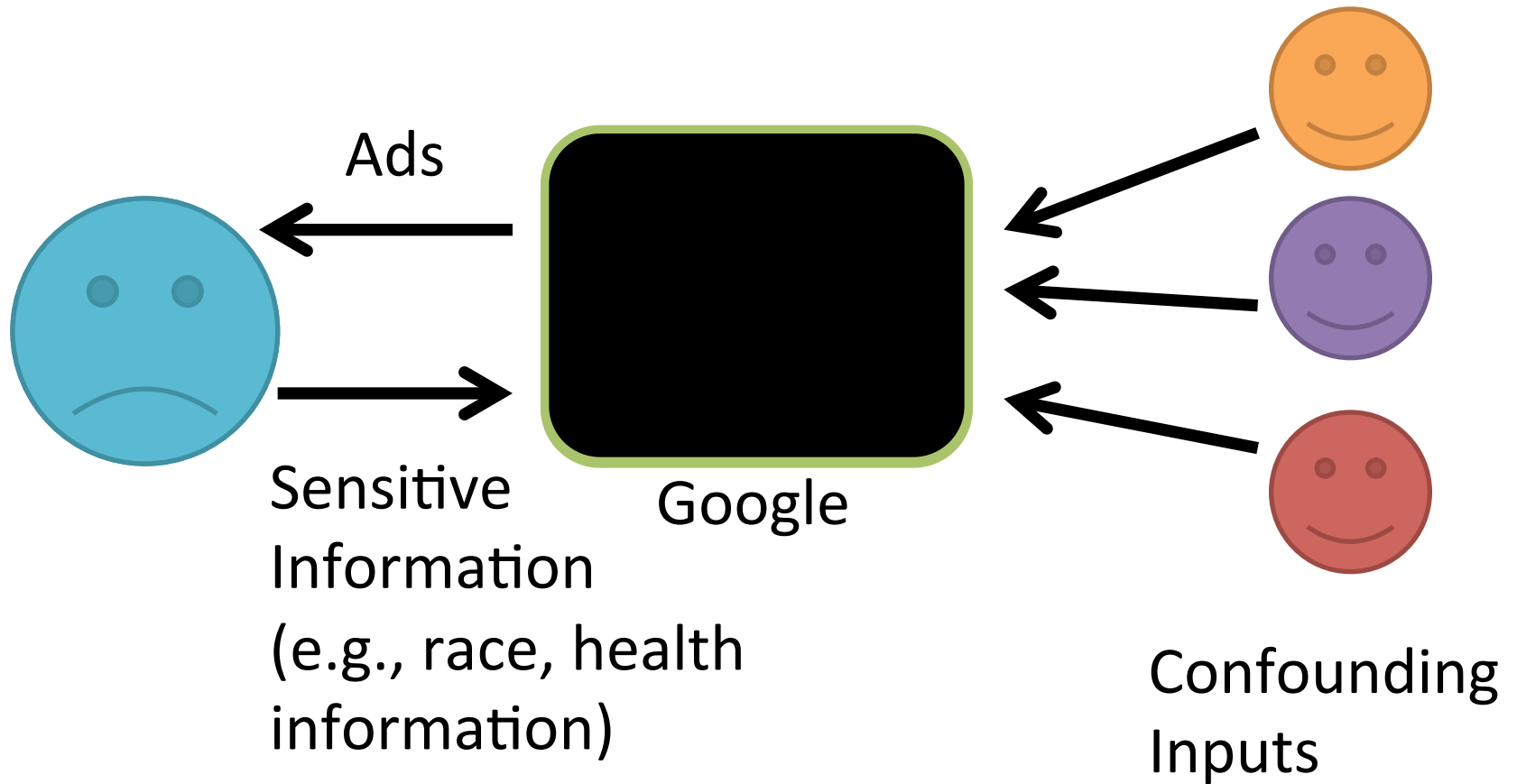
Cookies

Most Microsoft Sites use "cookies," small text files that can be read by a web server in the domain that put the cookie on your hard drive. We may use cookies to store your preferences and settings, help with sign-in, provide targeted ads and analyze site

Setting:

- ▶ Auditor has access to source code

Web Privacy: Advertising



Web Privacy: Online Tracking

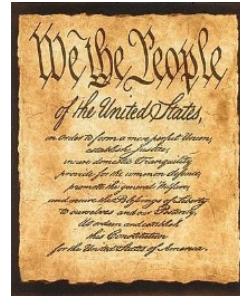


64

Independent tracking mechanisms on average on top-50 sites

Module I: Privacy through Accountability

- ▶ **Formalize Privacy Policies**
 - ▶ Precise semantics of privacy concepts (restrictions on personal information flow)
- ▶ **Enforce Privacy Policies**
 - ▶ Accountability
 - ▶ Detect
 - ▶ Intervene (blame, punish, fix code)
 - ▶ Explain



<http://www.andrew.cmu.edu/user/danupam/privacy.html>

Module I: Learning Outcomes

- ▶ Understanding of real-world privacy policies and laws
- ▶ Methods for detecting privacy violations

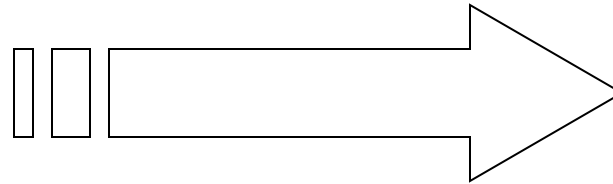
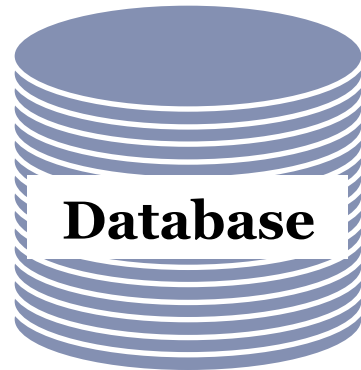
- ▶ Experience with audit tools for healthcare privacy
- ▶ Experience with web tracking investigation tool

Module I: Possible Projects

- ▶ Study privacy law enforcement in healthcare, education, banking with an audit tool
- ▶ Study web data use by Google, Bing, Facebook with an audit tool

Module II: Database Privacy

Database Privacy Goals



Government,
marketers,
researchers, ...

- Health records
- Census data
- Web search records

Conflicting goals:

- Provide useful **information**
- Protect **individual privacy**



[CNET](#) › [News](#) › [Corporate & legal](#)

August 7, 2006 9:59 AM PDT

AOL apologizes for release of user search data

By [Dawn Kawamoto](#) and [Elinor Mills](#)

Staff Writers, CNET News

Last modified: August 7, 2006 2:30 PM PDT

Related Stories

[Should Google be forced to hand over data?](#)

March 14, 2006

[Judge to help feds against Google](#)

March 14, 2006

AOL apologized on Monday for releasing search log data on subscribers that had been intended for use with the company's newly launched research site.

The randomly selected data, which focused on 658,000 subscribers and posted 10 days ago, was among the tools intended for use on the recently launched AOL Research site. But the Internet giant has since removed the search logs from public view.

stories

submissions

popular

blog

all stories

ask slashdot

book reviews

games

idle

yro

Anonymity of Netflix Prize Dataset Broken

Posted by **Zonk** on Tuesday November 27, 2007 @10:23AM
from the there-are-degrees-of-anonymity dept.



[KentuckyFC](#) writes

"The [anonymity of the Netflix Prize dataset has been broken](#) by a pair of computer scientists from the University of Texas, according to a report from the physics arXivblog. It turns out that an individual's set of ratings and the dates on which they were made are pretty unique, particularly if the ratings involve films outside the most popular 100 movies. So it's straightforward to find a match by comparing the anonymized data against publicly available ratings on the Internet Movie Database (IMDb) ([abstract on the physics arxiv](#)). The researchers used this method to find how individuals on the IMDb privately rated films on Netflix, in the process

Module II: Learning Outcomes

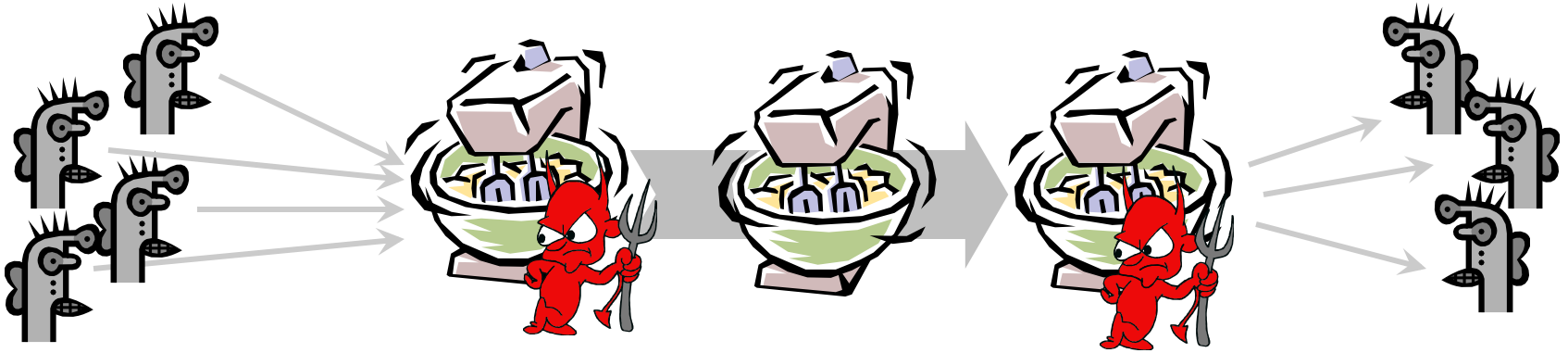
- ▶ Understanding of pitfalls in anonymizing databases
- ▶ Understanding of methods for releasing privacy-preserving statistics and their limitations

Module II: Possible Projects

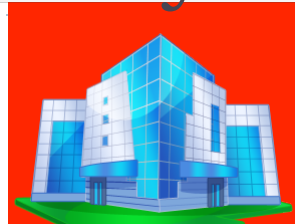
- ▶ Study privacy/utility tradeoffs with tool for releasing privacy-preserving statistics
- ▶ Study re-identification and other privacy risks of genomic data and their defenses

Module III: Cryptographic Mechanisms for Privacy Protection

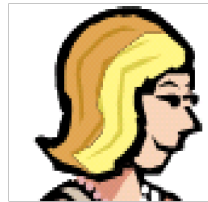
Anonymous Communication



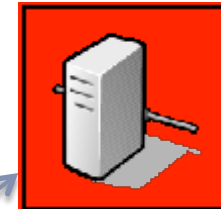
Anonymous Credentials



Organization



Alice



Service

“I have a cred from
Org saying
WA resident
Age >18”

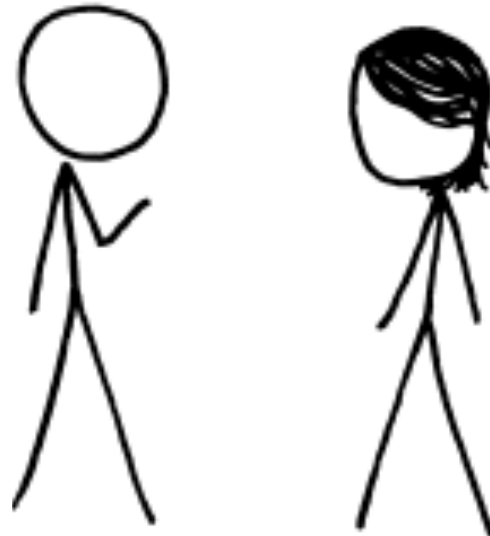
Cred from Org
Name Alice
Address
Birthdate
Birthplace
Citizenship
...

- Cannot
 - Identify Alice
 - Learn anything beyond the info she gives
 - Distinguish two users with the same attributes
 - Link multiple uses of the same credentials

Secure Two-Party Computation

Bob's Genome: ACTG...
Markers (~1000): [0,1, ..., 0]

Bob



Alice's Genome: ACTG...
Markers (~1000): [0, 0, ..., 1]

Alice



$$x = f(g_A, g_B)$$

Can Alice and Bob compute a function of their private data, without exposing anything about their data besides the result?

Module III: Learning Outcomes

- ▶ Understanding of cryptography behind
 - ▶ Anonymous communication
 - ▶ Anonymous credentials (zero-knowledge)
 - ▶ Biometric identification (secure computation)

Module III: Possible Projects

- ▶ Study improved cryptographic mechanisms for privacy protection, e.g. containing damage from large-scale data breaches
- ▶ Study privacy guarantees of and threats to using anonymous communication tools like Tor

An Organizing Viewpoint

Privacy as a right to *restrictions on personal information flow*

Student Introductions

- ▶ Who are you?
- ▶ Why are you here?

Homework for Next Class

- ▶ Read the Fair Information Practices Principles

[http://www.oecd.org/internet/ieconomy/
oecdguidelinesontheProtectionofPrivacyandTransborderFlow
sofpersonaldata.htm](http://www.oecd.org/internet/ieconomy/oecdguidelinesontheProtectionofPrivacyandTransborderFlowsofPersonalData.htm)

- ▶ Critically read the entire privacy policy of a Web services company of your choice

Homework Continued

Discussion questions:

- ▶ Try to find one example of a piece of the policy that maps to each principle.
- ▶ Can you find examples of principles that are not reflected in the policy?
- ▶ Can you find examples of policy clauses that reflect a principle that is not included in these principles?
- ▶ Are there policy clauses that could be more restrictive or less restrictive with respect to information use in order to better adhere to the principles?
- ▶ Are there parts of the policy that are too vague? If so, suggest alternatives.

Thanks! Questions?