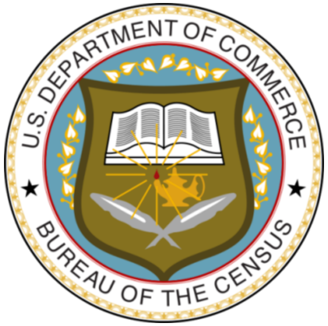
18734: Foundations of Privacy

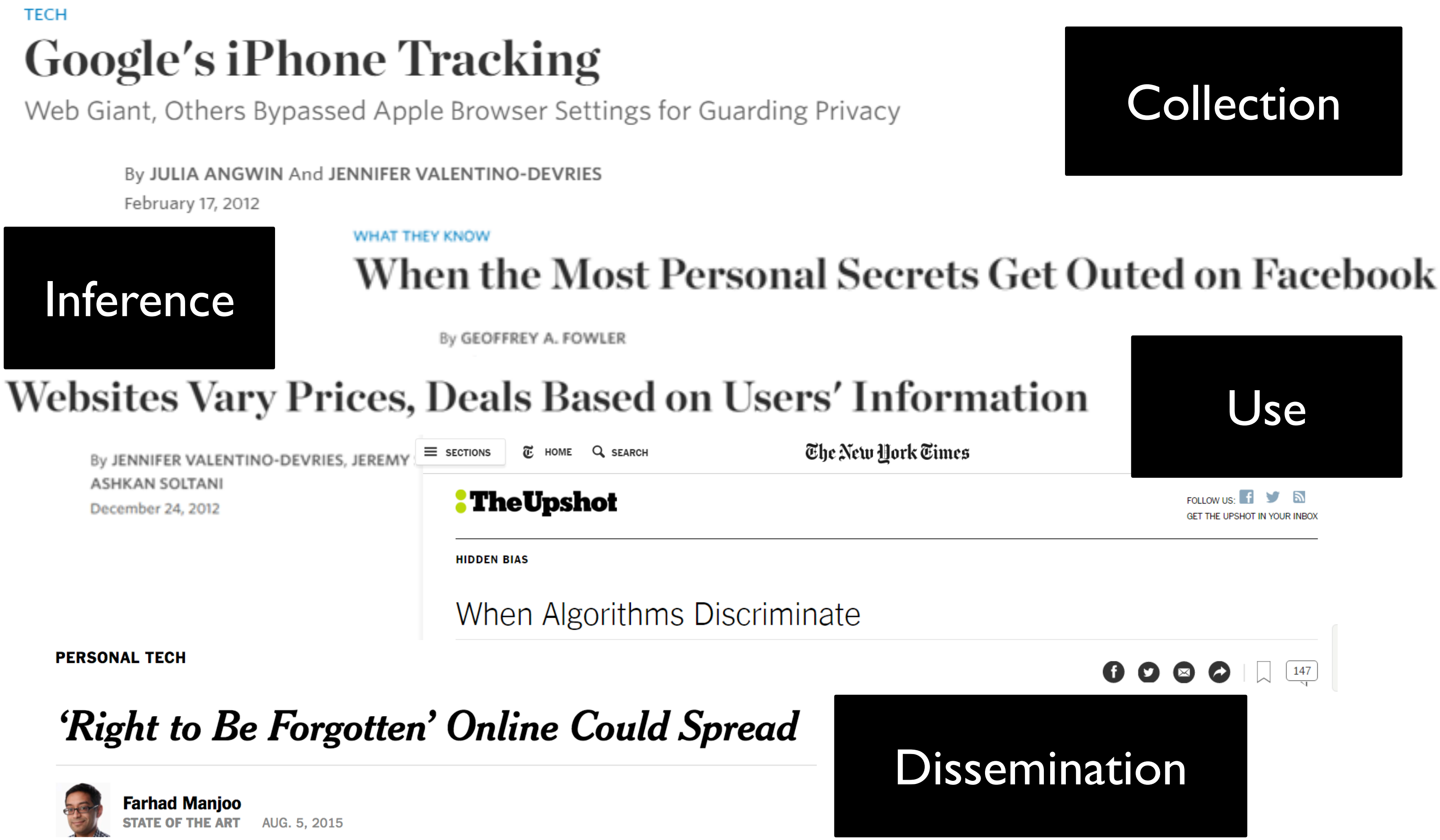
Course Overview

Giulia Fanti
CMU
Fall 2019
(Slides by Anupam Datta)

Personal Information is Everywhere



Privacy and Fairness Problems



TECH

Google's iPhone Tracking

Web Giant, Others Bypassed Apple Browser Settings for Guarding Privacy

By JULIA ANGWIN And JENNIFER VALENTINO-DEVRIES

February 17, 2012

Collection

Inference

WHAT THEY KNOW

When the Most Personal Secrets Get Outed on Facebook

By GEOFFREY A. FOWLER

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES, JEREMY

ASHKAN SOLTANI

December 24, 2012

Use

HIDDEN BIAS

When Algorithms Discriminate

PERSONAL TECH

'Right to Be Forgotten' Online Could Spread

Farhad Manjoo

STATE OF THE ART AUG. 5, 2015

Dissemination

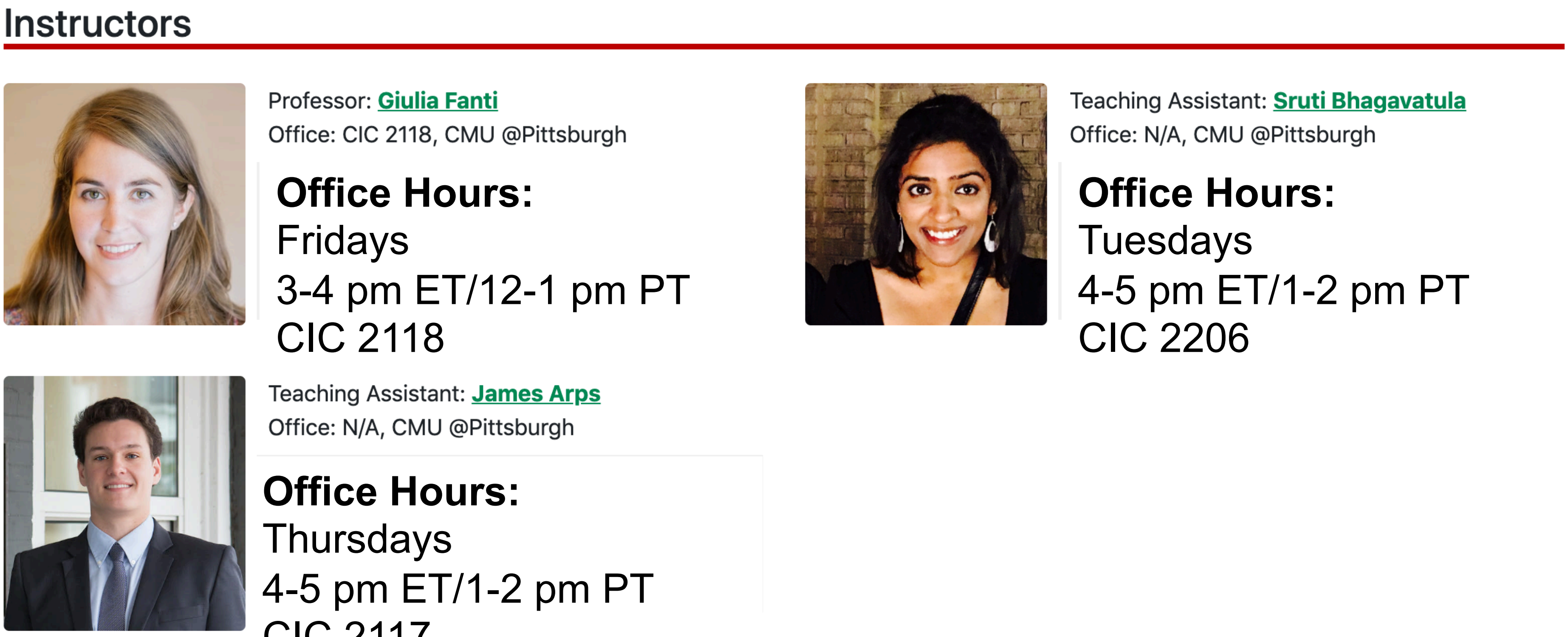
Organizing Questions

- What is privacy? What is fairness?
  - 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 and fairness?
  - Beyond creating laws and institutions
  - Computational mechanisms

Logistics

Course Staff



Instructors

Professor: Giulia Fanti
Office: CIC 2118, CMU @Pittsburgh

Office Hours:
Fridays
3-4 pm ET/12-1 pm PT
CIC 2118

Teaching Assistant: Sruti Bhagavatula
Office: N/A, CMU @Pittsburgh

Office Hours:
Tuesdays
4-5 pm ET/1-2 pm PT
CIC 2206

Teaching Assistant: James Arps
Office: N/A, CMU @Pittsburgh

Office Hours:
Thursdays
4-5 pm ET/1-2 pm PT
CIC 2117

Logistics

- Lectures: Monday & Wednesday, 12:30-2:20 PM EST WEH 4623 (B23 212)
- Recitation: Friday 12:30-1:20pm EST HH1107 (B23 211)
- Web page:
  - http://course.ece.cmu.edu/~ece734/index.html
- Canvas (for grades, homeworks, etc.) and Piazza (for all other communication)
  - Please enroll in Piazza; you will receive invitation shortly
- Course work and grading:
  - Homework (60%)
  - Course project (30%)
  - Class participation (10%)

Logistics

- Homework
  - ~1 week long
  - Due 10 min before the start of class
  - $n \approx 5$ total assignments
  - Will count best $(n-1)/n$ homeworks

- Late days
  - You have **eight**
  - You can use up to 3 per assignment
  - Each late day gives you 24 more hours
  - Do **not** ask me for more

Logistics

- Course Project:
  - Teams of 2 (no groups of 3; singletons come talk to me)
  - Project proposal: 1-2 pages (Oct 4)
  - Deliverable Part I: In-class presentation (Oct 30 – Nov 4)
  - Deliverable Part II: Written report: 7-10 pages (Dec 6)
  - In-class presentations (Dec 2, 4, 6)
  - If you come talk to instructors >=2 times about project progress (e.g., in OH) and have a borderline score, we will bump you up

Fall 2014 Course Projects

- Studies of personal information usage by Web services
  - Study on Facebook ads
  - Price Discrimination
  - Recommendations for news articles
  - Effect of cookies on Google ads
- Analytics to discover information usage by Web services
  - Abstaining Machine Learning
  - Ensemble Machine Learning
- Privacy Protecting the New York Taxicab Dataset
- Defense against Canvas Fingerprinting on the Web
- Privacy and Security issues of Android ads
- ML (Lasso Regression) over Encrypted Big Data

Fall 2015 Course Projects

- Secure Modular Embedding: Comparing Signals without revealing them
- Robust Ad Collection
- Inversion Attack on Machine Learning Models
- Privacy in Election Campaigns
- Improving Usability of Private Browsing Mode
- Investigating gender discrimination in popular employment websites
- Comparing Privacy Tools
- Google Advertising Platform Case study
- The Unexpected Danger of Multiple Social Media Accounts: Instagram and Twitter Reveal More than You Think
- Effects of Browser-Type on Internet Results

Logistics

- Class participation
  - In-class quizzes, first 10 minutes of class
  - Will discuss quizzes afterwards
  - Use participation on quizzes to determine class participation grade

  - Correct/Incorrect answers do not affect your grade
  - BUT if you are borderline and did well on these quizzes, I will bump you up

Logistics

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

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!

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, or18-730 is recommended, but not required
- If in doubt, please talk to me after class

Quick Class Poll

Privacy Problems

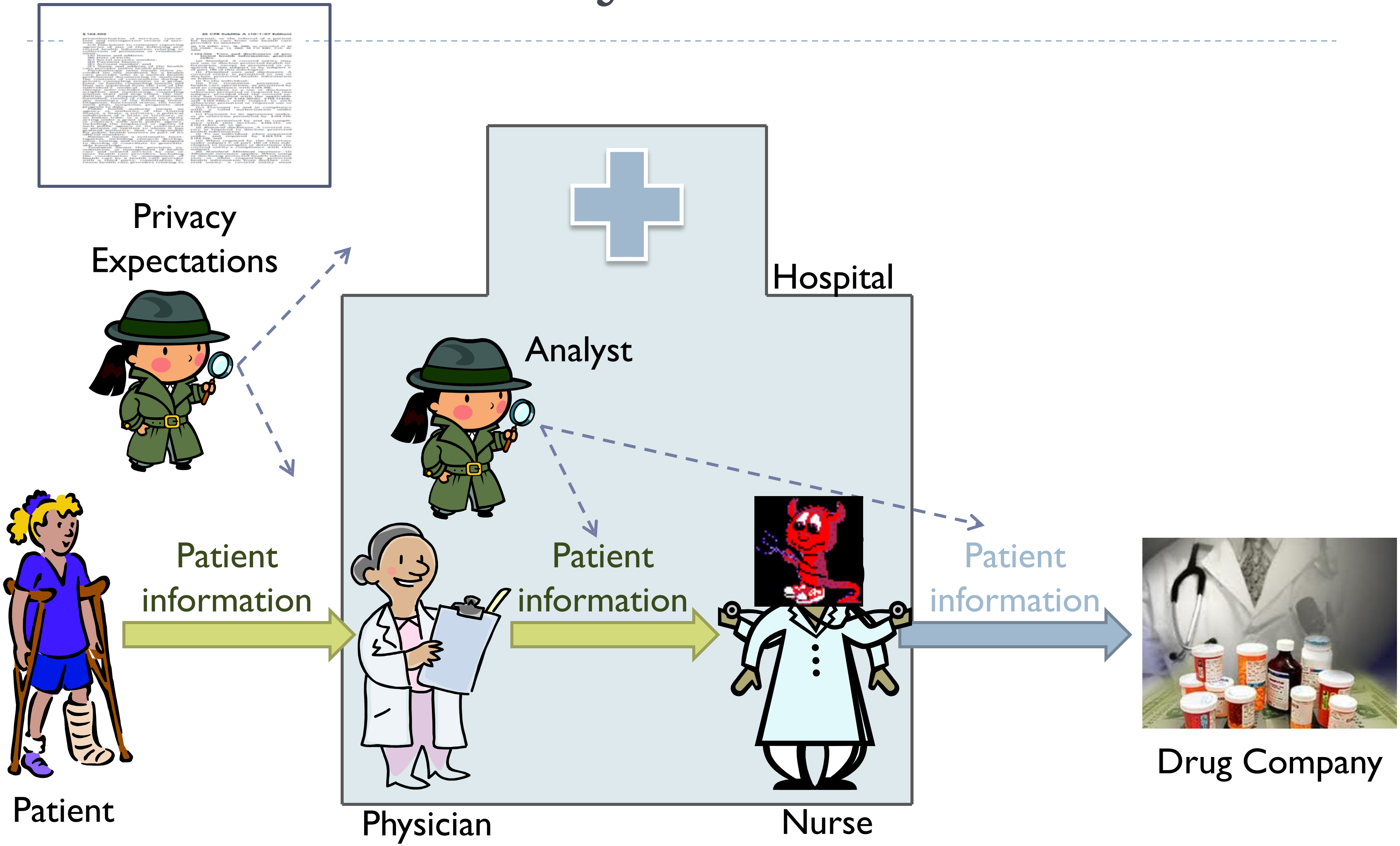
Module I: Privacy through Accountability

Collection

Use

Dissemination

Healthcare Privacy



Privacy Expectations

Hospital

Analyst

Patient information

Patient information

Patient information

Patient

Physician

Nurse

Drug Company

HIPAA Privacy Rule

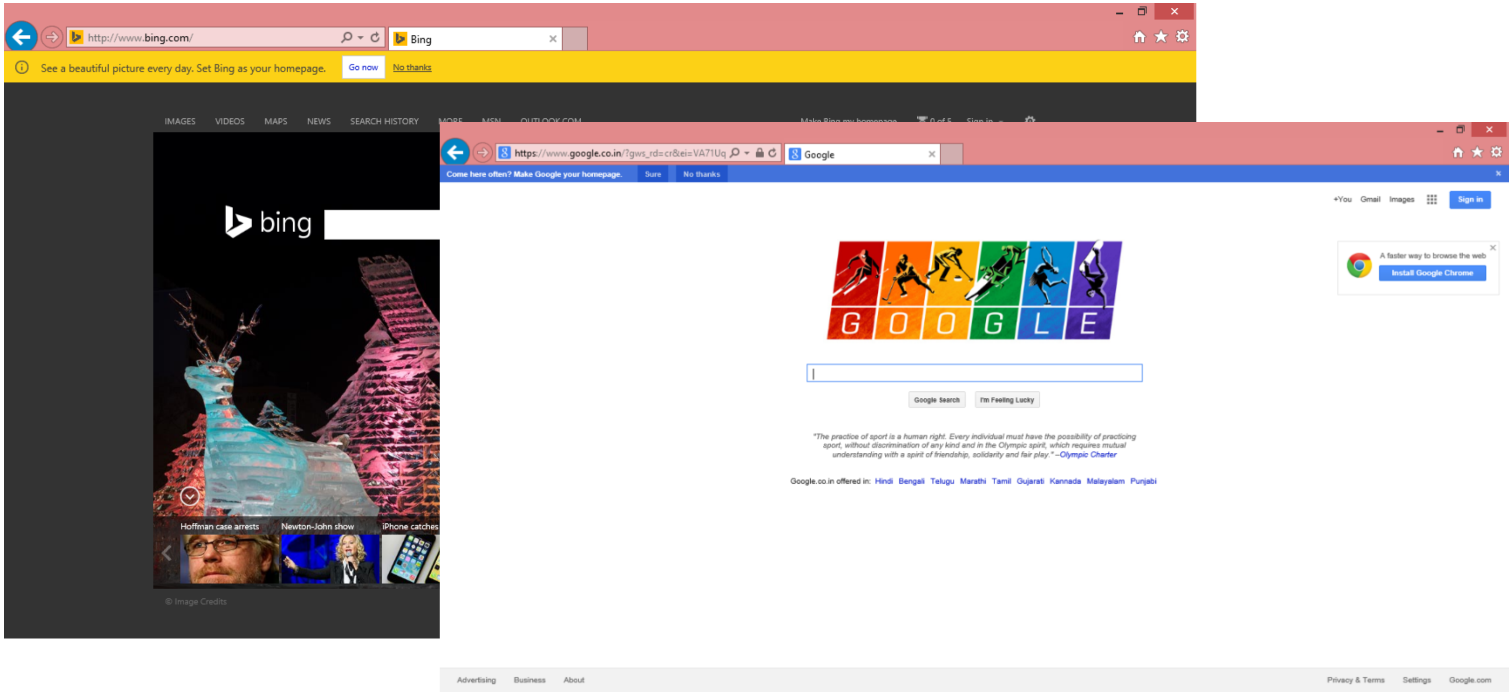
Use

Dissemination

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

Web Advertising

Use



http://www.bing.com/

See a beautiful picture every day. Set Bing as your homepage.

https://www.google.co.in/?gws_rd=cr&ei=VA71Uq

Come here often? Make Google your homepage.

A faster way to browse the web

"The practice of sport is a human right. Every individual must have the possibility of practicing sport, without discrimination of any kind and in the Olympic spirit, which requires mutual understanding with a spirit of friendship, solidarity and fair play." –Olympic Charter

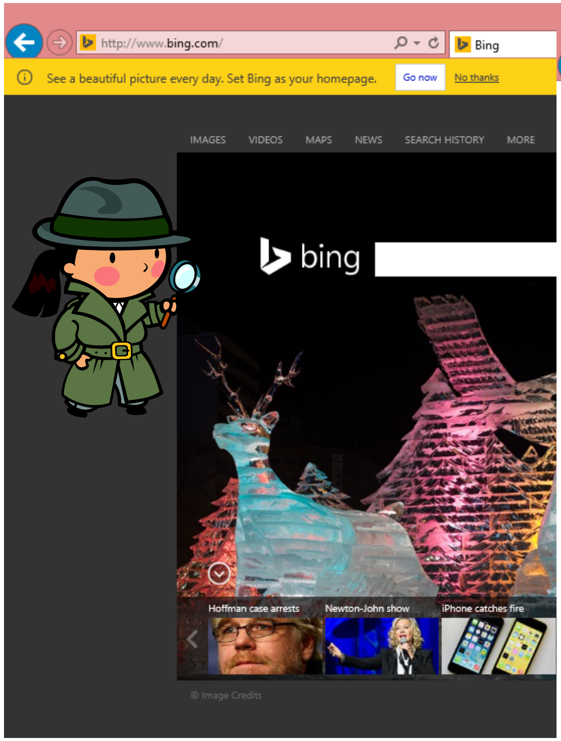
Google.co.in offered in: Hindi Bengali Telugu Marathi Tamil Gujarati Kannada Malayalam Punjabi

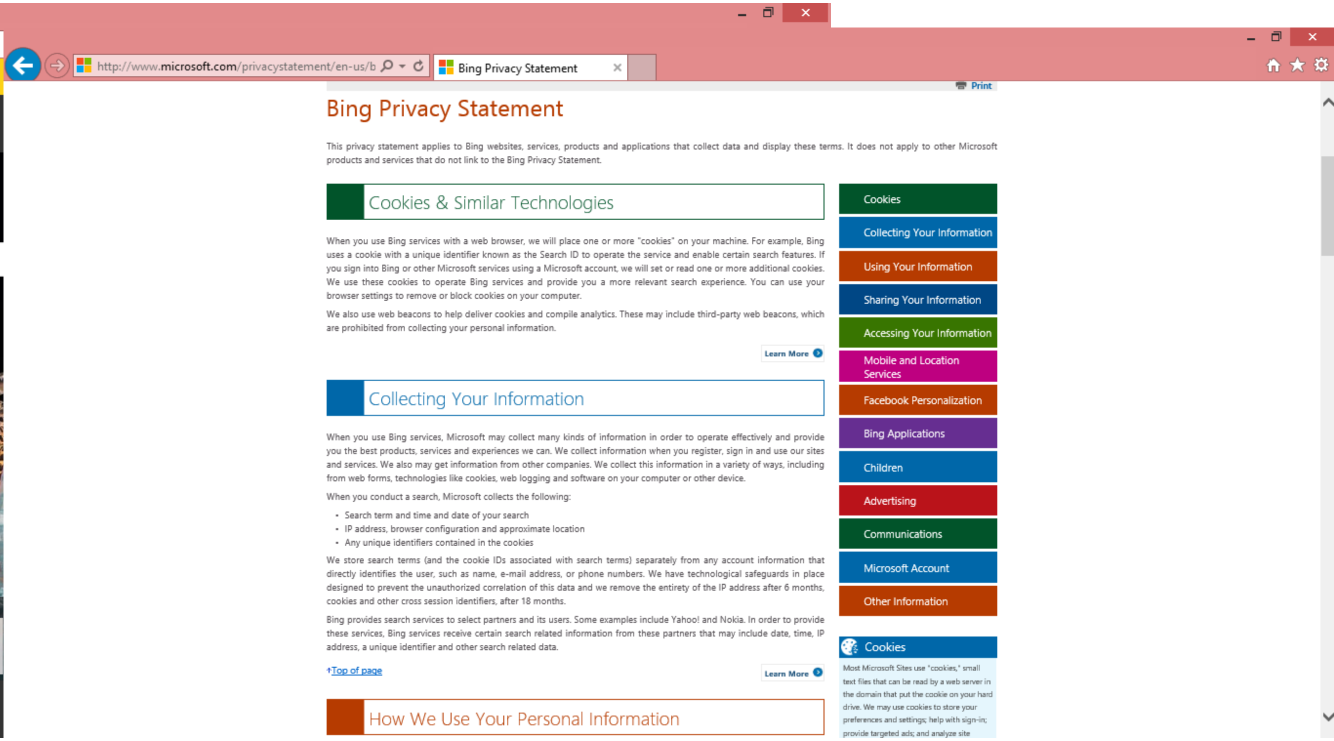
<u>Example privacy policies:</u>

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

Privacy Compliance for Bing

Use





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

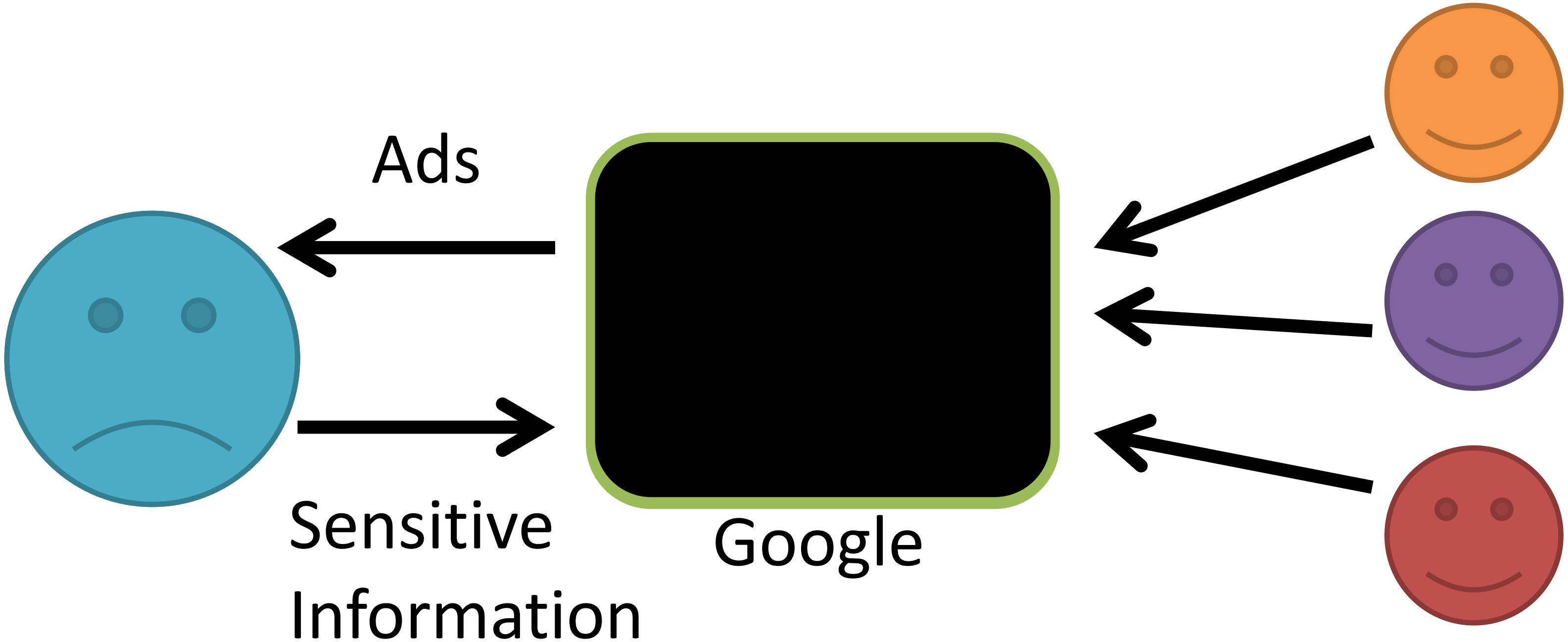
Most Microsoft Sites use "cookies," small text files that can be read by a web server in the domain that put the cookie on your hard drive. We may use cookies to store your preferences and settings; help with sign-in; provide targeted ads; and analyze site

Setting:

- Auditor has access to source code

Web Privacy: Advertising

Use



Ads

Sensitive Information (e.g., race, health information)

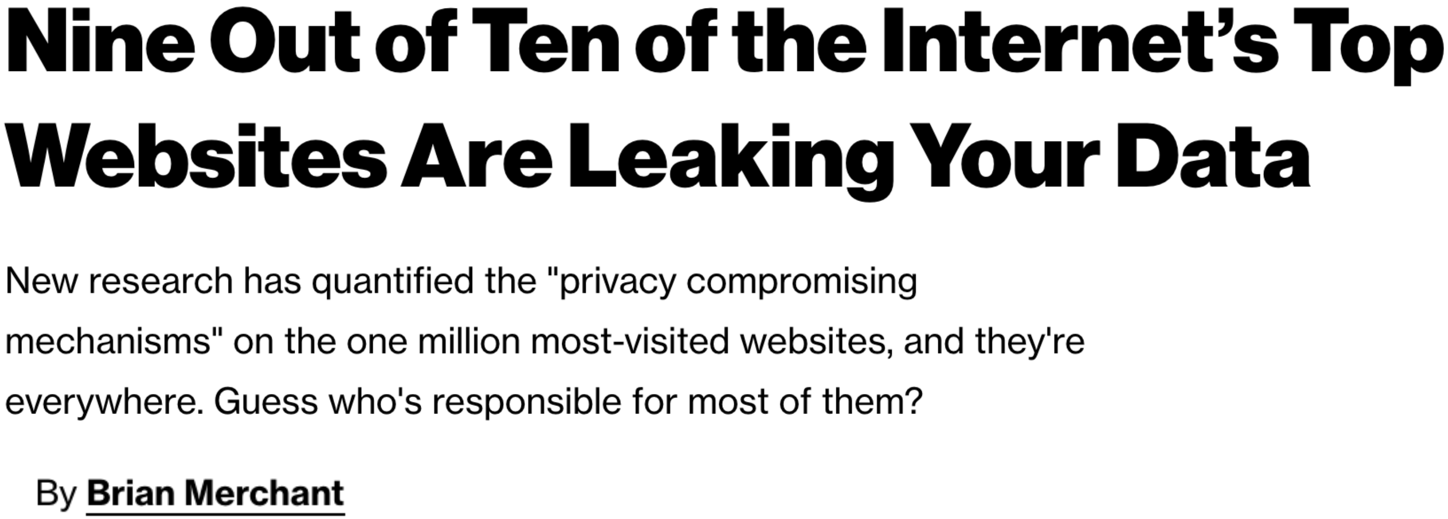
Google

Confounding Inputs

Web Privacy: Online Tracking

Collection

VICE



Nine Out of Ten of the Internet's Top Websites Are Leaking Your Data

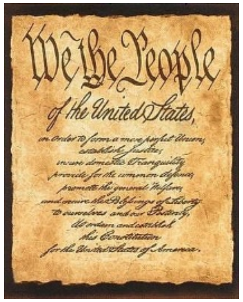
New research has quantified the "privacy compromising mechanisms" on the one million most-visited websites, and they're everywhere. Guess who's responsible for most of them?

By **Brian Merchant**

Module I: Privacy through Accountability

- Formalize Privacy Policies
  - Precise semantics of privacy concepts

    (restrictions on personal information flow)



- Enforce Privacy Policies
  - Accountability
    - Detect
    - Explain
    - Correct

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

- Practical experience
  - Use web tracking investigation tools
  - Interact with companies' privacy policies

Module II: Protecting Privacy and Fairness in Big Data Analytics

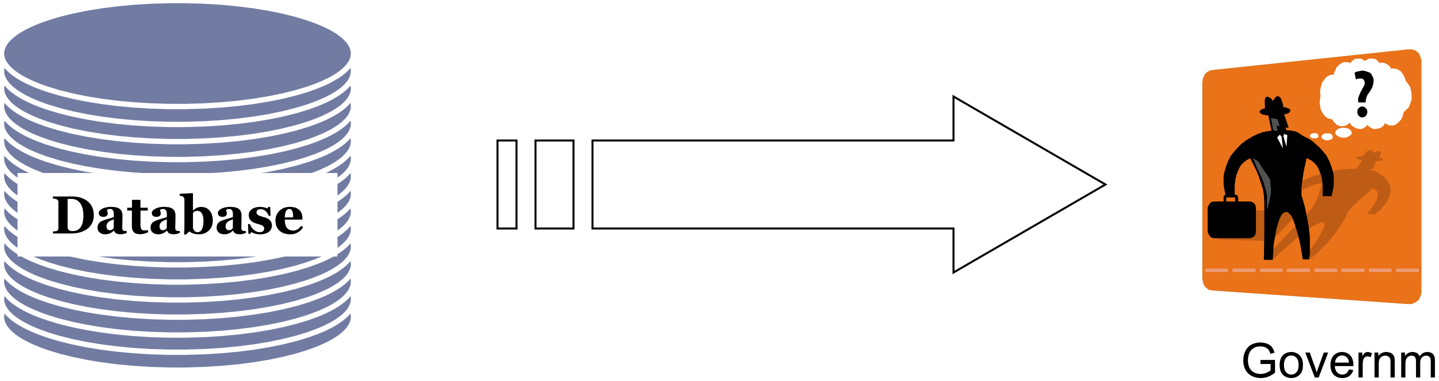
Collection

Inference

Use

Dissemination

Database Privacy Goals

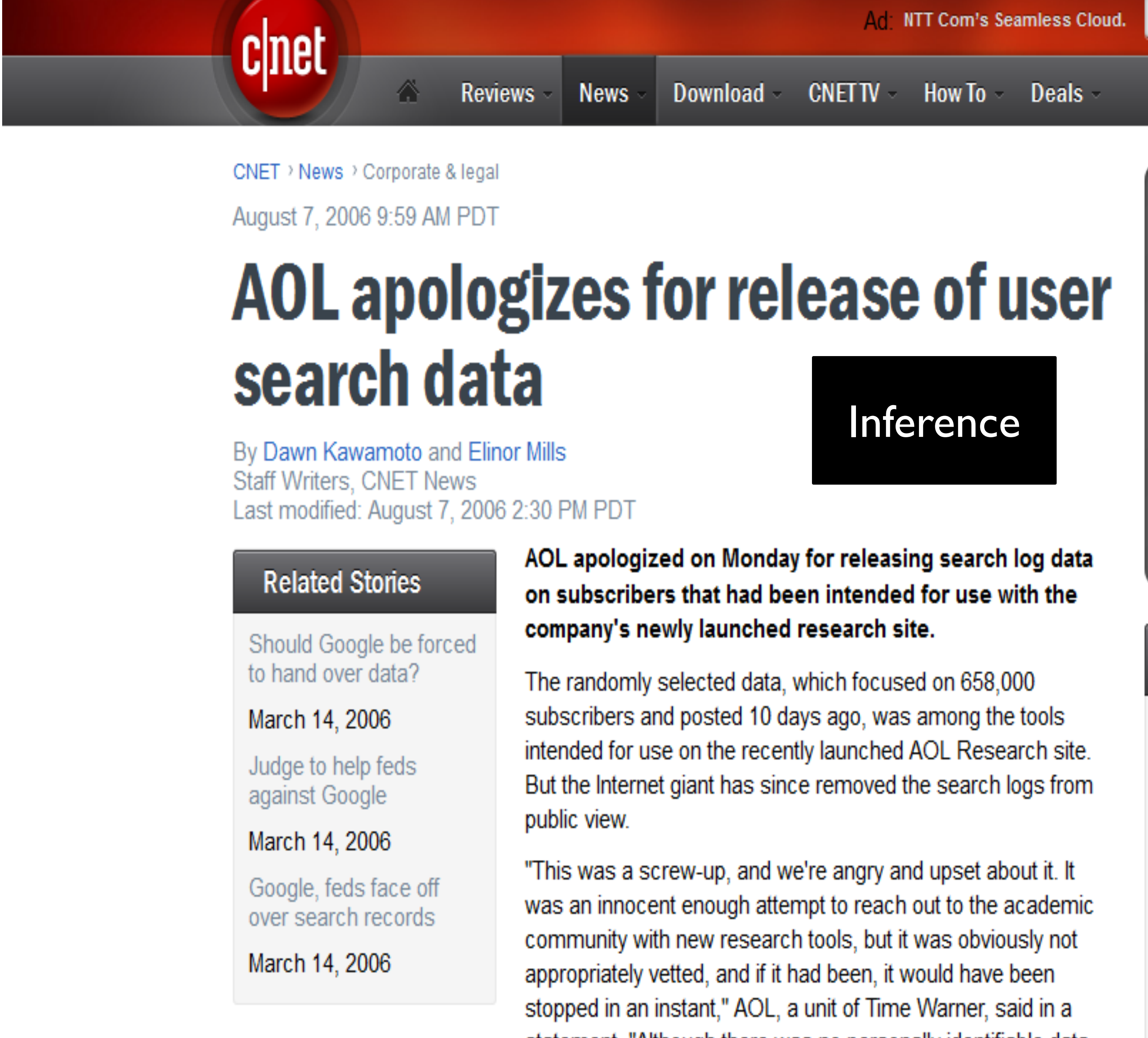


Government, marketers, researchers, …

- 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

Inference

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

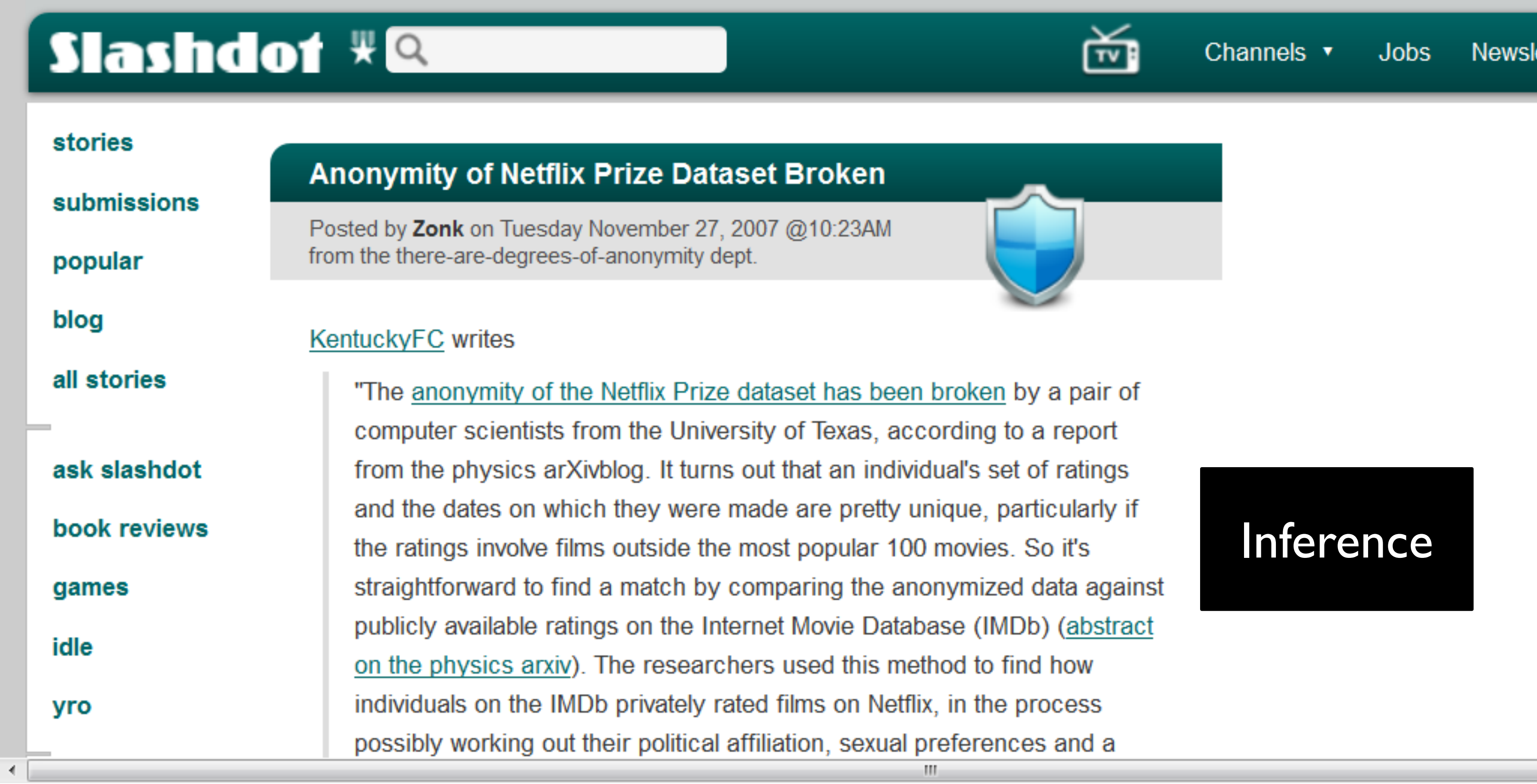
Judge to help feds against Google
March 14, 2006

Google, feds face off over search records
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.

"This was a screw-up, and we're angry and upset about it. It was an innocent enough attempt to reach out to the academic community with new research tools, but it was obviously not appropriately vetted, and if it had been, it would have been stopped in an instant," AOL, a unit of Time Warner, said in a statement. "Although there was no personally identifiable data



Anonymity of Netflix Prize Dataset Broken

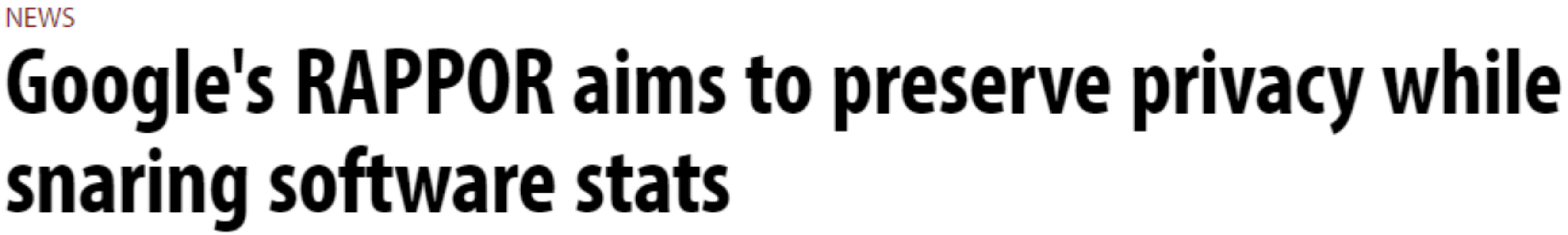
Posted by **Zonk** on Tuesday November 27, 2007 @10:23AM
from the there-are-degrees-of-anonymity dept.

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 possibly working out their political affiliation, sexual preferences and a

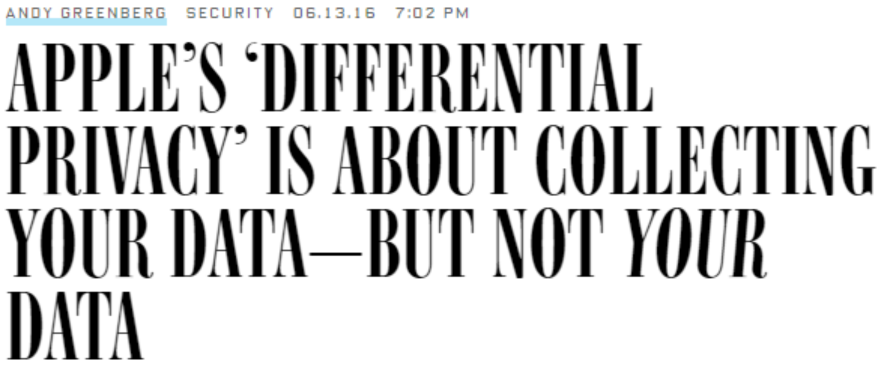
Inference

Privacy Solutions



NEWS

Google's RAPPOR aims to preserve privacy while snaring software stats



ANDY GREENBERG SECURITY 06.13.16 7:02 PM

APPLE'S 'DIFFERENTIAL PRIVACY' IS ABOUT COLLECTING YOUR DATA—BUT NOT *YOUR* DATA

Collection

Inference

Dissemination

Module II: Learning Outcomes

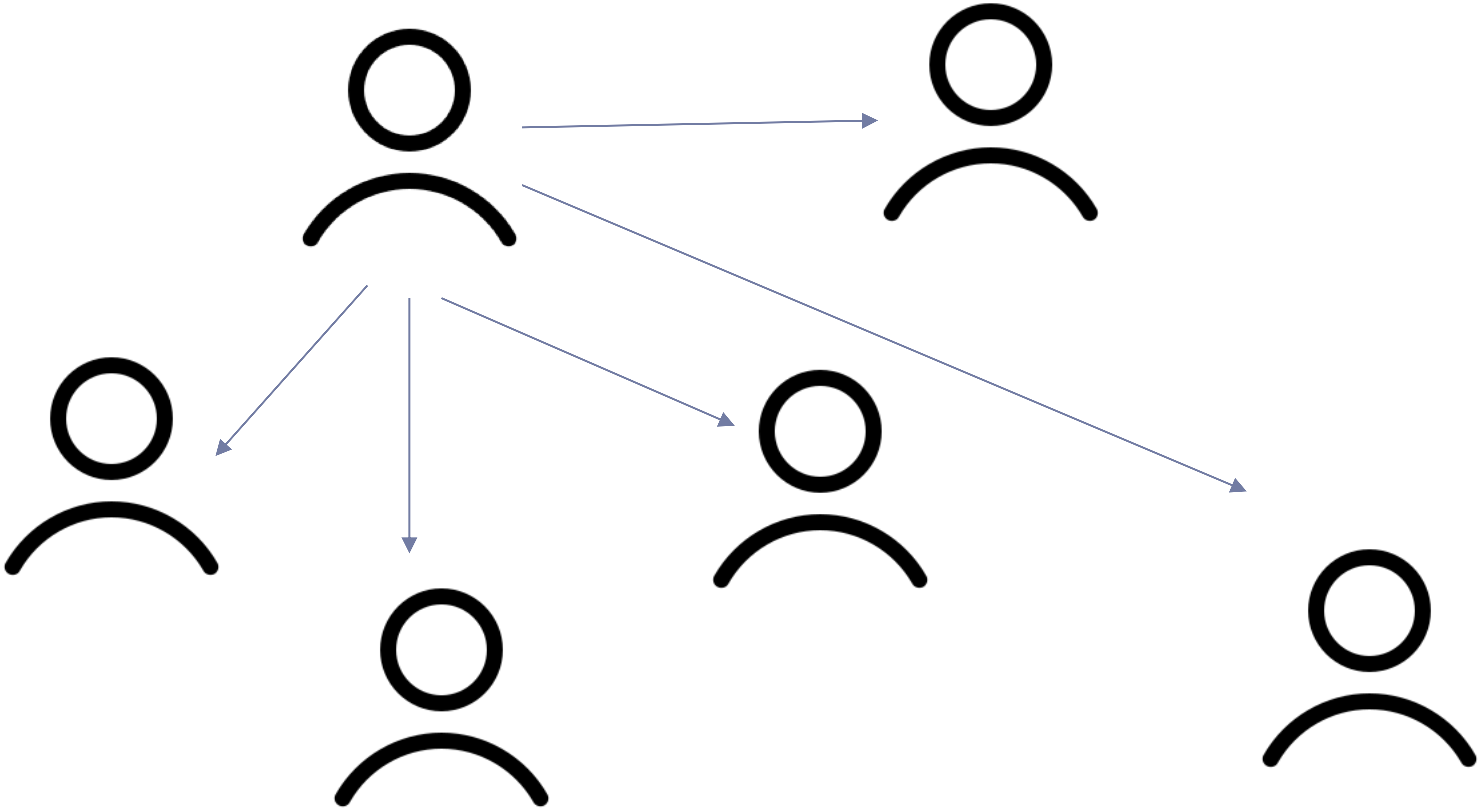
- Understanding of pitfalls in anonymizing databases
- Understanding of methods for releasing privacy-preserving statistics and their limitations
- Understanding bias in machine learning and corrective measures
- Understanding transparency (explanations) for decisions of machine learning systems

- Practical experience
  - Implement deanonymization techniques
  - Use privacy-preserving data collection/analysis tools

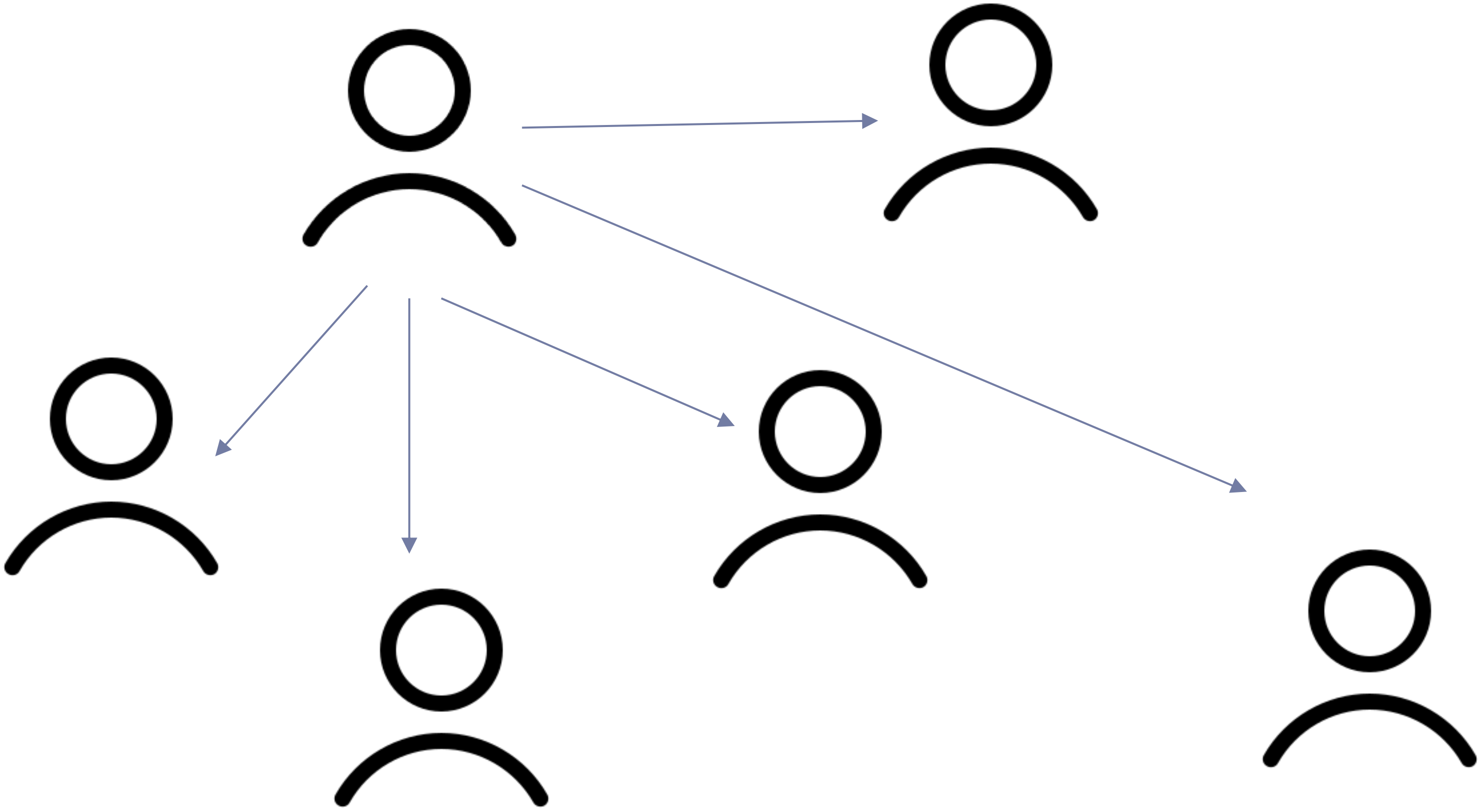
Module III: Special Topics: Cryptographic Mechanisms for Privacy Protection

Collection

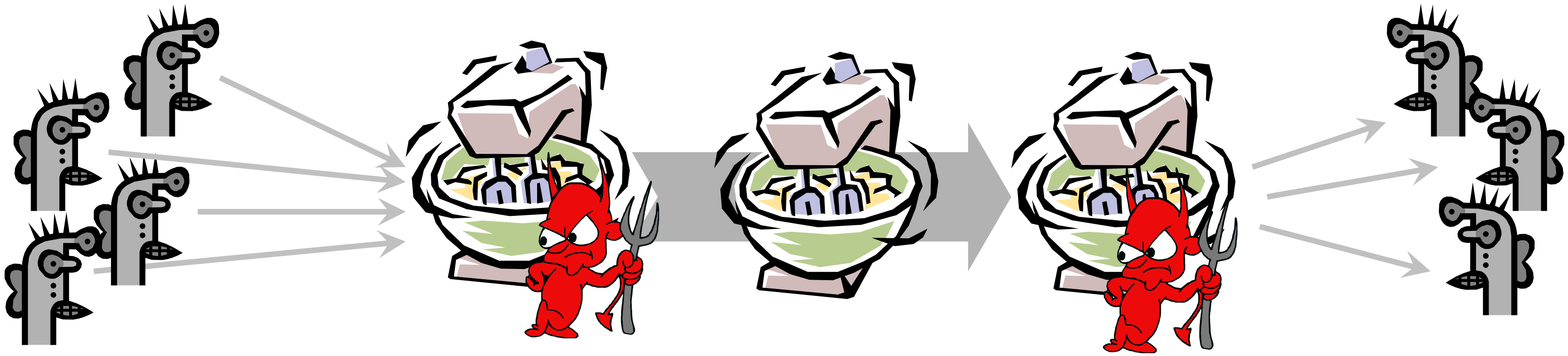
Secret-Sharing



Anonymous Communication (One-to-many)

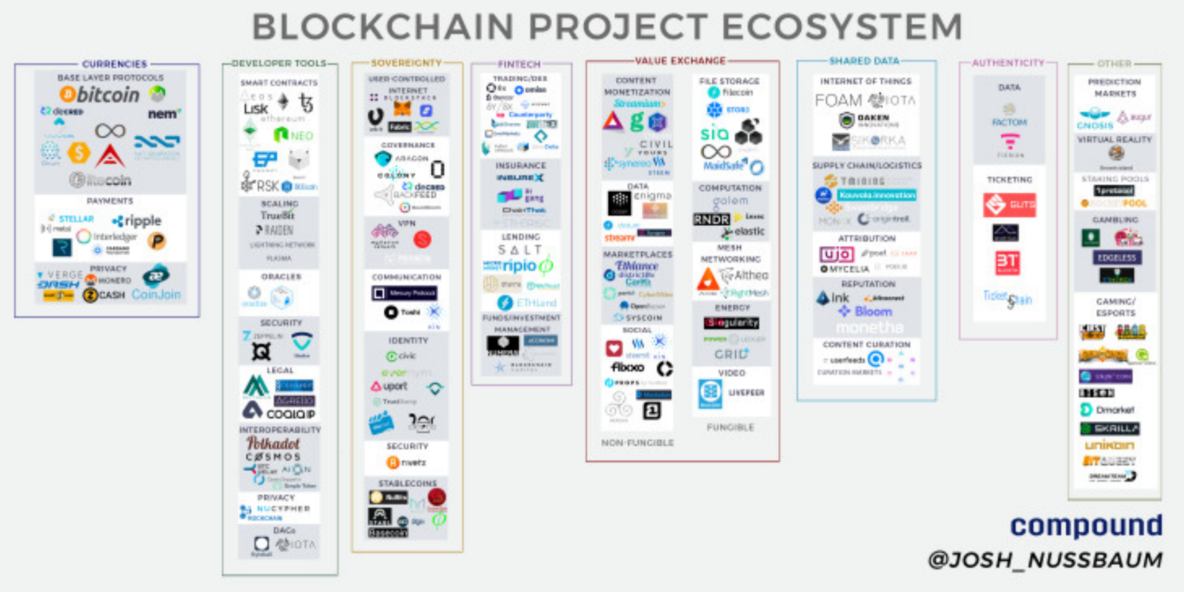


Anonymous Communication (Point-to-point)



Tor

Anonymous Cash: Blockchains



- Distributed, append-only ledgers
- Basis for cryptocurrencies and smart contracts
- Challenge: How to ensure privacy?

Module III: Learning Outcomes

- Understanding of cryptography behind
  - Anonymous communication
  - Blockchain and cryptocurrencies with privacy

- Experience using tools
  - Anonymous communication
  - Anonymous e-Cash

An Organizing Viewpoint

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

Collection

Inference

Use

Dissemination

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/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

- Critically read the entire privacy policy of a Web services company of your choice
    - Examine pairs of services owned by the same company (e.g., Facebook-Whatsapp)

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.
- Are there conflicts in policies of service pairs owned by the same company?

Thanks! Questions?