Introduction to Privacy

Asst. Prof. Sinem Sav

Some slides/ideas adapted from: Jean-Pierre Hubaux, Erman Ayday,George Danezis, Carmela Troncoso, Bart Preneel, Claudia Diaz, Seda Guerses, Bryan Ford

Goals of this and the next lecture

- Privacy? Come again?
- Some scandals (privacy as a scandal-driven topic)

- Understanding that privacy is not only an individual-oriented problem:
  - Privacy is a security property
  - Privacy is key to maintain democratic societies
  - Privacy crosses individuals' boundaries (analogy: road safety)

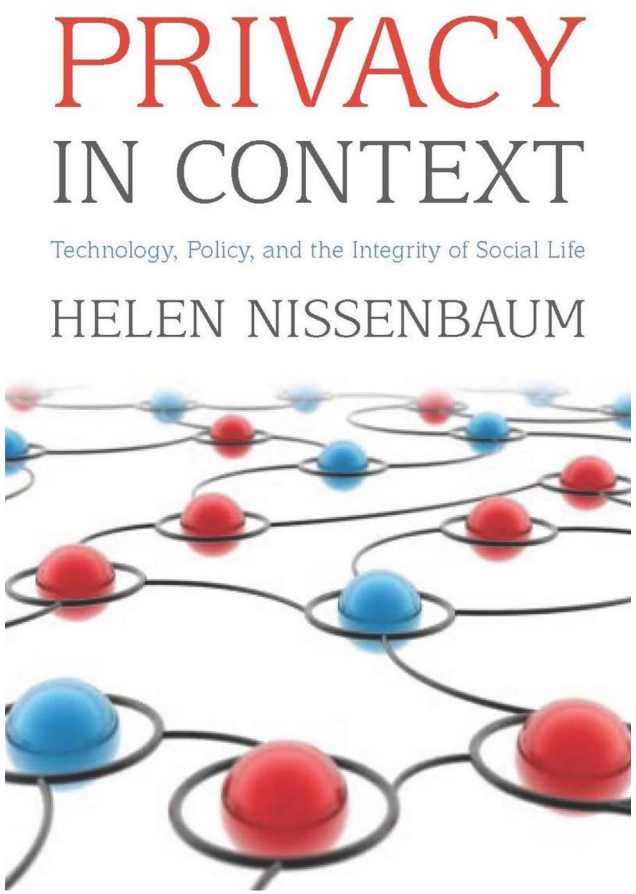
- There are different conceptions of privacy depending
  - on the privacy paradigm: what does it mean to protect privacy
  - on the adversary model: recognizing and modeling privacy adversaries

- Learning examples of Privacy Enhancing Technologies suitable for adversary models

What is privacy?

Privacy: Definition

- **Privacy control** is the ability of individuals to determine when, how, and to what extent information about themselves is revealed to others
- **Tool:** Privacy-enhancing technologies
- **Goal**: Let personal data be used only in the context it has been released



Personal Data

- Any kind of information (a single piece of information or a set of information) that can personally identify an individual
  - Name, address, national identification number, date of birth, a photograph, hospital records, etc.
- Protection is crucial
- In an interconnected electronic world, individual pieces of data can no longer be regarded in isolation
- Can also be used to put people under complete surveillance, in breach of their fundamental rights

The Value of Privacy

- Depends on the individual
- Common misconception:
  - If you aren't doing anything wrong, what do you have to hide?
- Privacy is not hiding the wrong!

If one would give me six lines written by the hand of the most honest man, I would find something in them to have him hanged.

- Cardinal Richelieu (1585-1642)

Security vs. Privacy

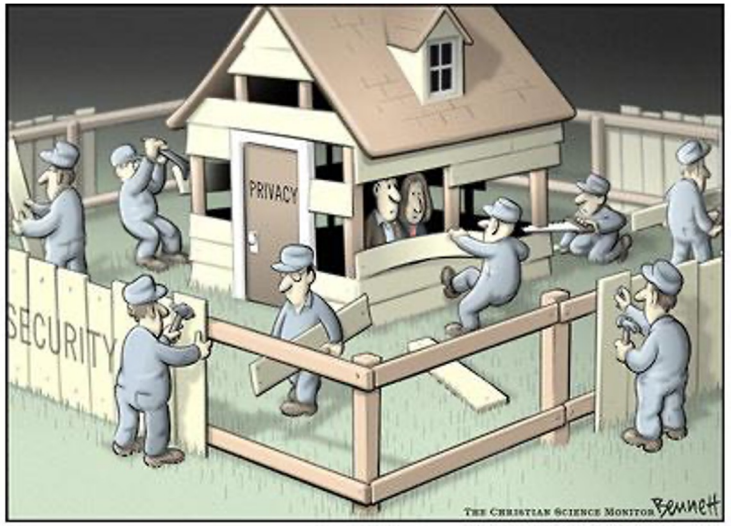
- Which is more important?
- How much privacy are you willing to give up for security?

  - How much liberty buys how much security?
  - Can we even afford privacy in this age of insecurity?

https://www.schneier.com/blog/archives/2008/01/security_vs_pri.html

Security vs. Privacy – Common Beliefs

- Increased security can be bought with reduced privacy
- Privacy and security are a zero-sum game
  - If one gains, the other loses
- Security is vital to survival
  - To the defenders of the surveillance state, security means "saving American lives"
- Privacy is unique to humans, but it's a social need
- In 2008, 51% of Americans said security is more important than privacy [1]
  - Only 29% disagreed and said privacy is more important



[1] http://www.rasmussenreports.com/content/pdf/8522

Common misconception:
we need to tradeoff security for privacy!

More surveillance -> More security -> Safer world ?

But:

Surveillance may be not **effective**: smart adversaries evade surveillance

criminals use since long Telegram, Tor, Signal, but average citizens do not!!

Surveillance tools can be **abused**: lack of transparency and safeguards

Snowden revelations: NSA spying on citizens, companies, ...

Surveillance tools can be **subverted** for crime / terrorism

Greek Vodafone scandal (2006): “someone” used the legal interception functionalities (backdoors) to monitor 106 key people: Greek PM, ministers, senior military, diplomats, journalists...

Many systems can be (stealthily) used for surveillance

Function creep: expansion of a process or system where data collected for one specific purpose is subsequently used for another unintended or unauthorized purpose.

"We will create a new system to improve X"

"We have this data, why don't we use it for Y"

A recurrent *function creep* example: identity systems

Aadhaar - India's “optional” Unique Identity identification number scheme

12-digit identity number based on their biometric information and demographic data

>1 billion people stored in the database

Goal “*promoted as providing the poor with an identity*”

It became:

mandatory for benefits system (distribution of food rations and fuel subsidies)

mandatory for buying a SIM card

mandatory for opening a bank account

pay taxes

no education without UID

Women rescued from prostitution are to put their numbers on the database to get rehabilitated!!

https://privacyinternational.org/state-privacy/1002/state-privacy-india

A recurrent function creep example: identity systems

EURODAC - fingerprint database for asylum seekers

Goal: store fingerprints from all people who cross the border into a European country without permission – asylum seekers as well as irregular migrants to help immigration and asylum authorities to better control irregular immigration to the EU, detect secondary movements (migrants moving from the country in which they first arrived to seek protection elsewhere) and facilitate their readmission and return to their countries of origin.

It became:

database for police and public prosecutors, such as Europol.

More data: in addition to fingerprints, the facial images and alphanumerical data (name, ID or passport number) of asylum seekers and irregular migrants will also be stored.

Asylum: deal to update EU fingerprinting database

Press Releases LIBE 19-06-2018 - 18:38

- EURODAC to include more data on asylum seekers and irregular migrants
- Safety of refugee children to be improved
- Europol access to EURODAC made more efficient

https://www.euractiv.com/section/justice-home-affairs/news/eurodac-fingerprint-database-under-fire-by-human-rights-activists/
http://www.europarl.europa.eu/news/en/press-room/20180618IPR06025/asylum-deal-to-update-eu-fingerprinting-database

Surveillance and Technology

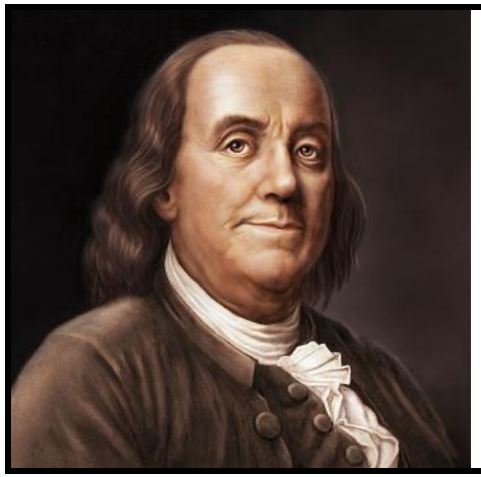
- As computer memory becomes cheaper, more and more of electronic footprints are being saved
  - 100 megabytes: to record everything the fastest typist input to his computer in a year
  - 4 to 8 gigabytes: to record everything the average user does on the Internet in a year
  - 5 gigabytes: to save the yearly phone calls of a typical person who uses 500 cell phone minutes a month
  - 200 gigabytes: to constantly record all audio of an individual per year
  - 700 gigabytes: to constantly record all video of an individual per year (life recorder)
- As processing becomes cheaper, more and more of it is being cross-indexed and correlated

NSA's data storage facility in Utah

13

Security vs. Privacy – Bottom line

- Security and privacy are not opposite ends of a seesaw
  - No need to accept less of one to get more of the other
- There is no security without privacy
- Liberty requires both security and privacy

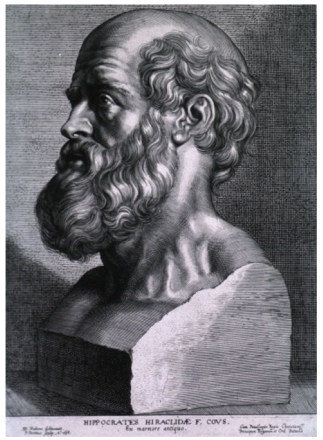


Those who would give up essential liberty to purchase a little temporary safety, deserve neither liberty nor safety

- Benjamin Franklin

History and introduction:
Privacy as a scandal-driven topic

24 centuries ago



Hippocrates
Ca. 460 to ca. 370 B.C.

Hippocratic oath

"…

All that may come to my knowledge in the exercise of my profession or in daily commerce with men, which ought not to be spread abroad, I will keep secret and will never reveal.

…"

Many centuries later…

“The Right to Privacy”

Warren and Brandeis

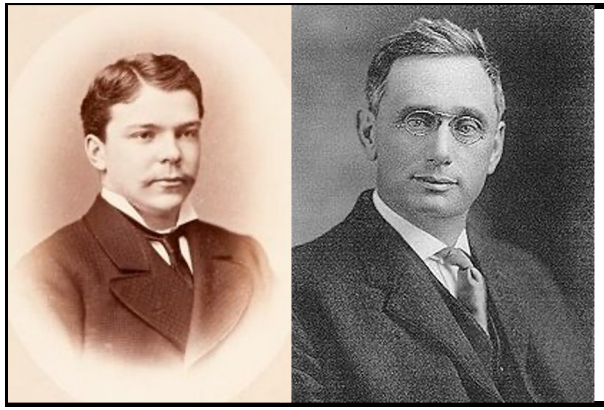
Harvard Law Review

Vol. IV Dec. 15, 1890 No. 5

First explicit declaration of a US right to privacy

Major concern: photography without consent

Privacy Concept



“The Right to Privacy” (1890)

- Laid the foundation for a concept of privacy known as control over information about oneself

William Prosser (School of Law): different interests in privacy (1960):

- Intrusion upon a person's seclusion or solitude, or into his private affairs
- Public disclosure of embarrassing private facts about an individual
- Publicity placing one in a false light in the public eye
- Appropriation of one's likeness for the advantage of another

Views on the Meaning and Value of Privacy

Alan Westin described privacy as

- *the ability to determine for ourselves when, how, and to what extent information about us is communicated to others* (1967)

William Parent defined privacy as the condition of not having undocumented **personal information** known or possessed by others (1983)

- Personal information is documented, on Parent's view, only when it belongs to the public record, that is, in newspapers, court records, or other public documents

Research milestones:

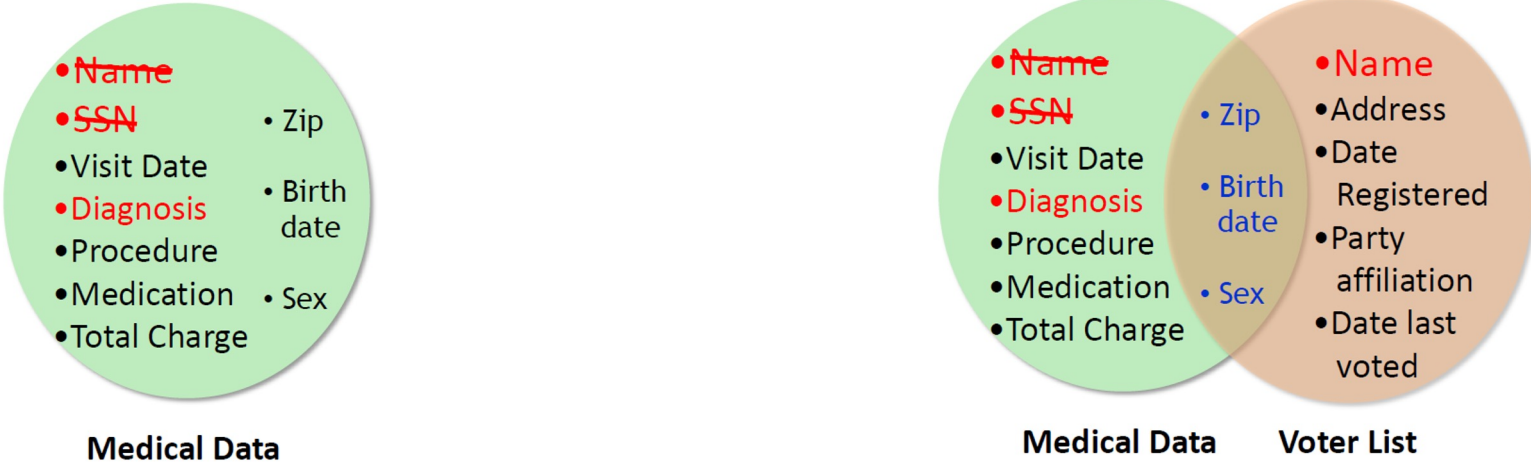
- 2002 - Massachusetts' medical data
- 2006 - AOL
- 2008 – Netflix

Major scandals:

- 2013 – Snowden revelations
- 2016 – Cambridge Analytica

2002 - The Massachusetts Governor Privacy Breach [1]

- Release of anonymized data is socially beneficial
  - E.g., the usage of hospital records greatly helps medical research
- Governor of MA was uniquely identified using ZipCode, Birth Date, and Gender
  - Additional result: 87% of US population can be uniquely identified using ZipCode, Birth Date, and Gender
- Name linked to Diagnosis



SSN: Social Security Number

[1] K-Anonymity: a model for protecting privacy - L. Sweeney, 2002

Figure: Ashwin Machanavajjhala

21

2006 - AOL Search Logs

- 20 million Web search queries collected by AOL ”anonymously” made public
- UserIDs were replaced by random numbers
- Examples (all provided by AOL)

  (AOL stands for America Online, a major US ISP)

865712345	Uefa cup
865712345	Uefa champions league
865712345	Champions league final
865712345	Champions league final 2007
236712909	exchangeability
236712909	Proof of de Finetti’s theorem
112765410	Zombie games
112765410	Warcraft
112765410	Beatles anthology
865712345	Grammy 2008 nominees

This user was not de-anonymized, fortunately for him…

A Face Is Exposed for AOL Searcher No. 4417749 [1]

- Searches by No. 4417749:
  - landscapers in Lilburn, GA
  - people with the last name Arnold
  - dog that urinates on everything
- Data trail led to Thelma Arnold, a 62-year-old widow who lives in Lilburn, GA
- AOL removed the search data from its site over the weekend and apologized for its release

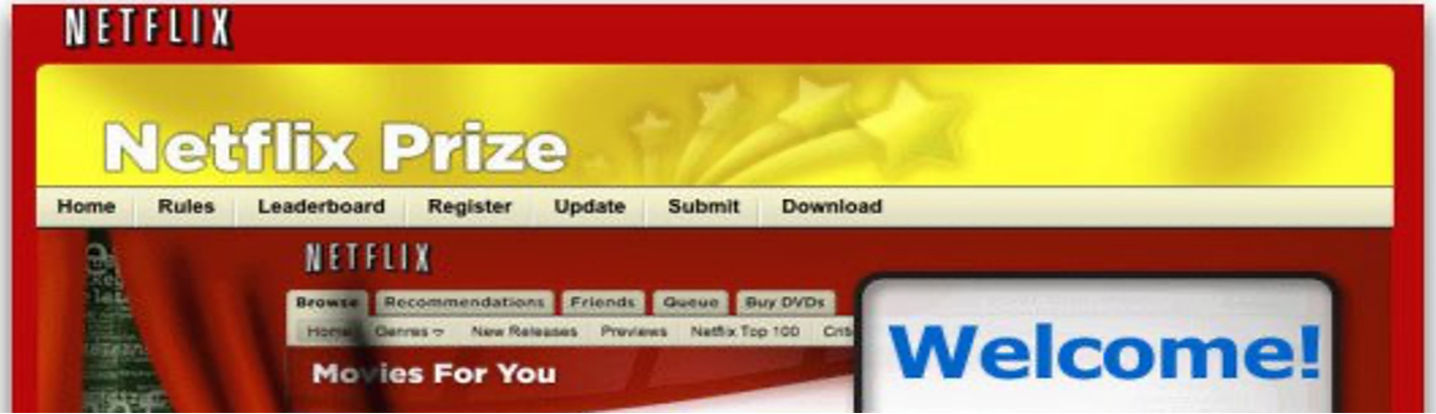
[1] The New York Times - August 9, 2006

23

2008 - Netflix Case

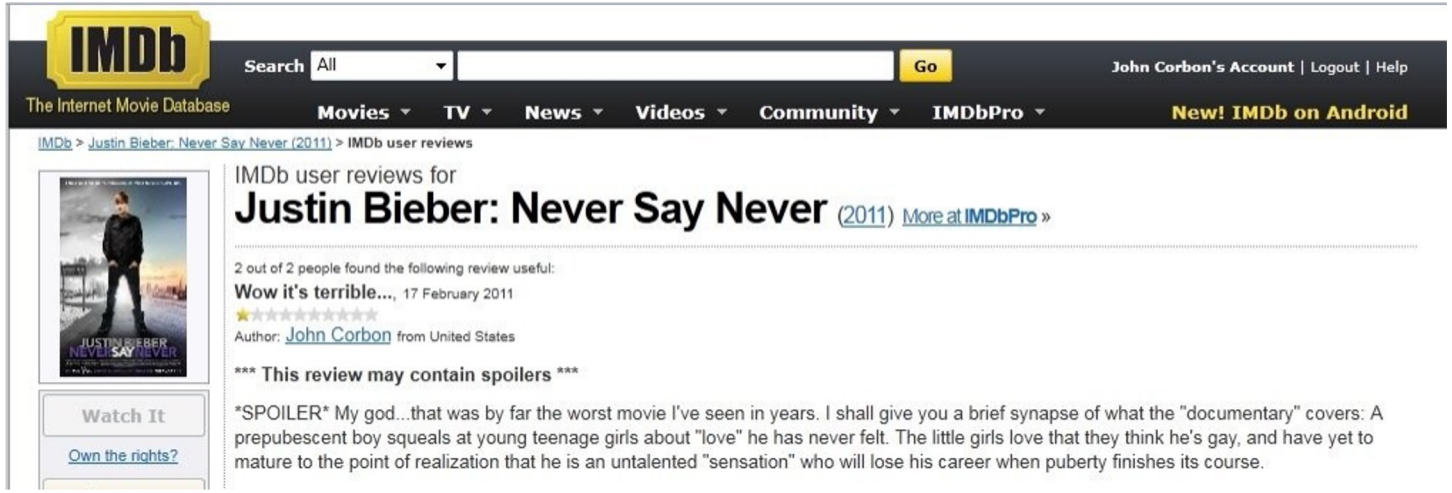
- Netflix movie recommendations

- Netflix publicly provided a training data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies



Researchers Reverse Netflix Anonymization [1]

- Researchers from UT Austin identified two people out of all “anonymized” users whose movie ratings were released by online rental company Netflix by combining this data set with another one: Internet Movie Database (IMDb)
- The collection of movie ratings - combined with a public database of ratings - was enough to identify the people



- Researchers found that one of the users had strong opinions about some liberal and gay-themed films

[1] Robust De-anonymization of Large Sparse Datasets - Narayanan and Shmatikov, 2008

25

2016: Cambridge Analytica

Context:

- 2013: CA obtained 50M records from Facebook
  - "survey" app that leaked friends' information as well as from the user answering the survey
- CA created a system that can target voters based on psychological profile
- Was used to target US voters during the 2014-2016 elections (notably US presidential elections and Brexit vote)

But how bad can targeted advertising be?

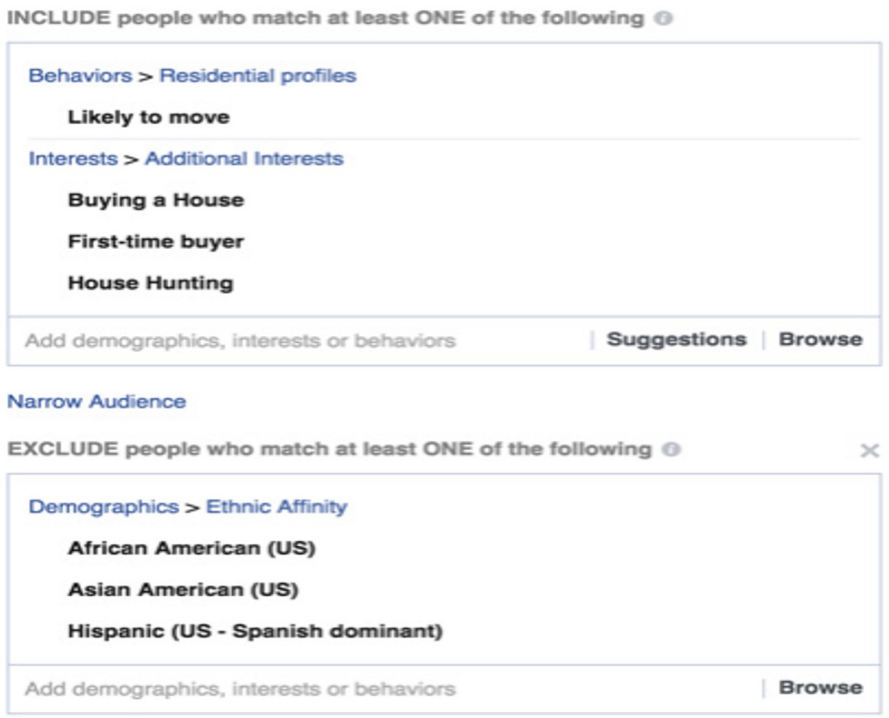
Attribute-based targeting

Facebook's users have *attributes*

- computed by Facebook
- based on Likes, 3rd-party browsing (tracking via "Like" button), etc
- bought from "Partner" companies (Data brokers)

>1200 well-defined attributes

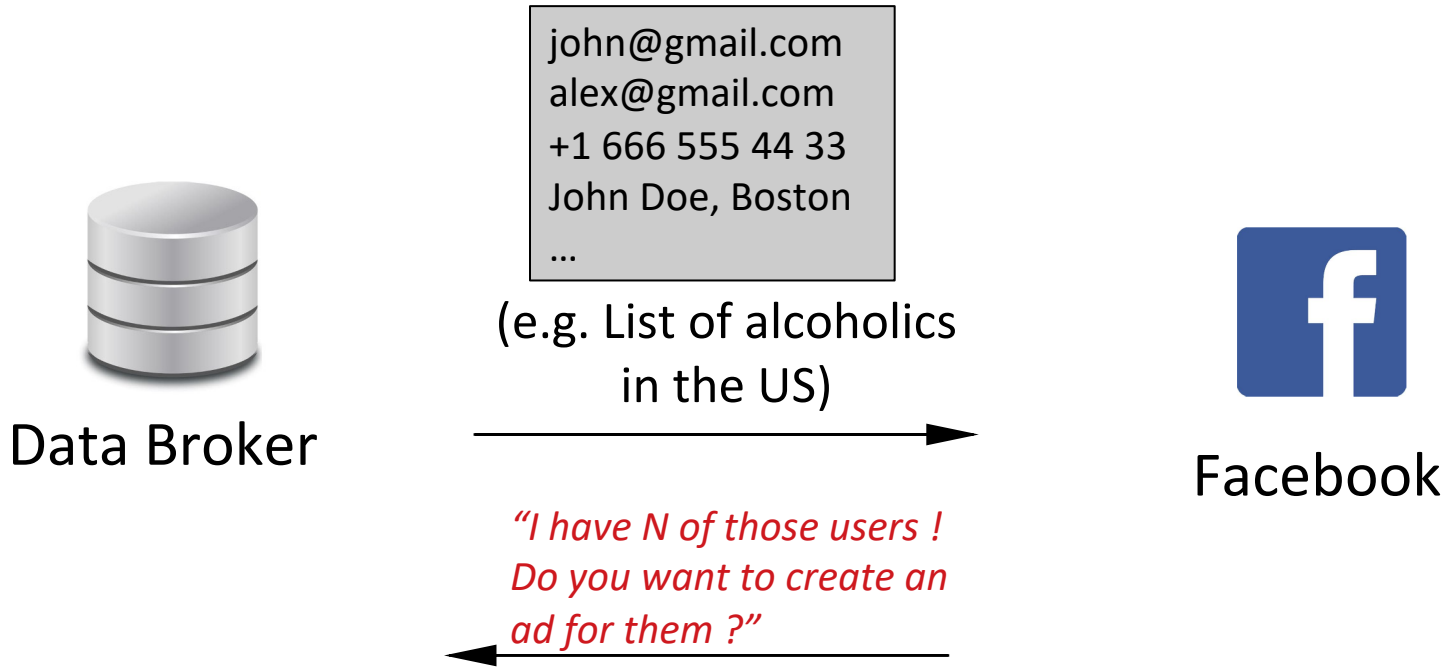
>250k "loosely defined" attributes (from text-processing)



- Advertiser selects attributes, Facebook shows ads to relevant users
- Facebook doesn't give the identity of the matching users to the advertiser

PII-based targeting

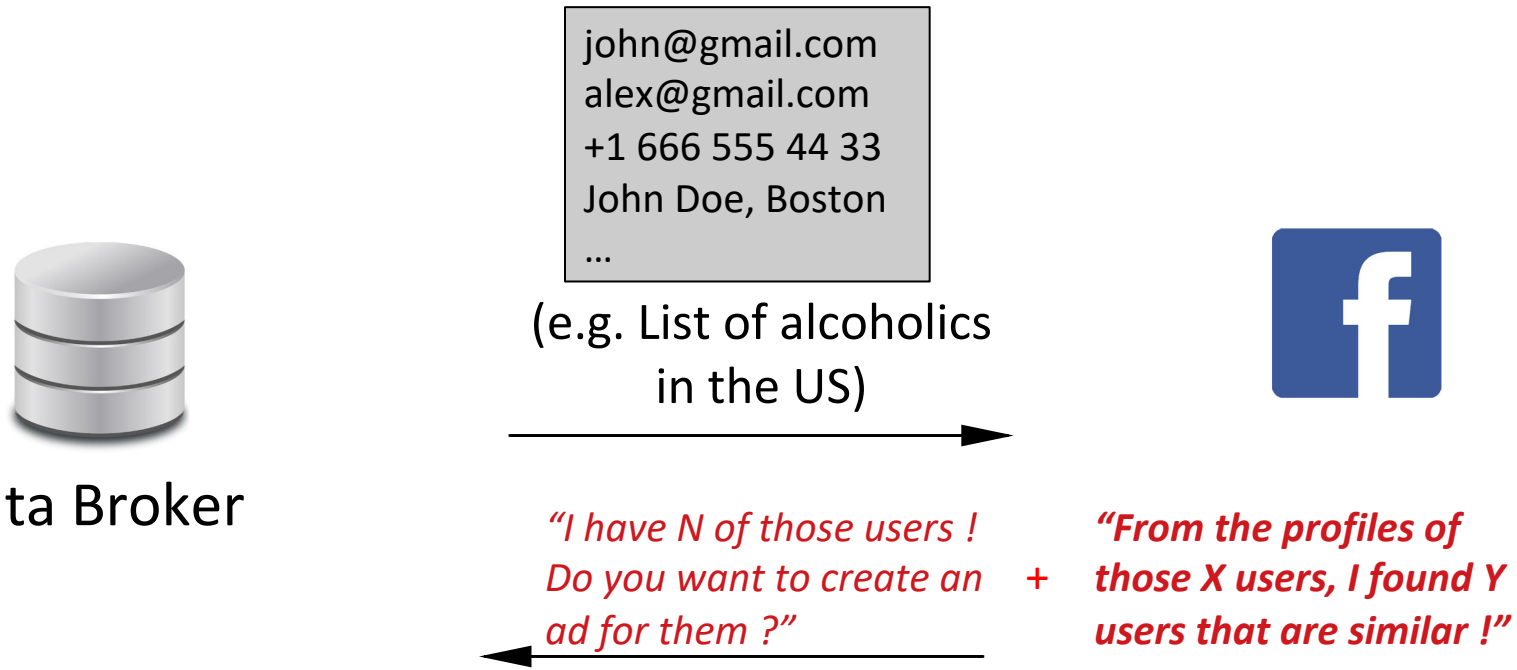
- Advertiser can bulk-upload a database (bought from data brokers, etc.) to Facebook, who tells how many users are present on the system, and allows to target them



PII: personal identifying information

Lookalike-audience

- Facebook knows how to find “similar” people (that are not listed by Data Brokers) based on interests, browsing patterns, etc

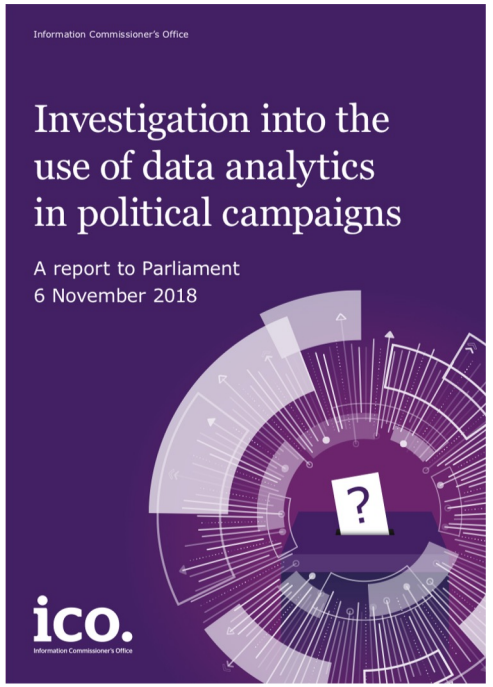


- Can be tailored per region, sex, marital status & other attributes

See e.g., Athanasios Andreou et al., Measuring the Facebook Advertising Ecosystem, NDSS 2019

Conclusion on Facebook and Cambridge Analytica

“Brexit vote” and US presidential elections
-Two leading democracies, find themselves
internally polarized, victims of home-made digital tools



Information Commissioner’s Office
(UK's independent body set up to uphold information rights)
- Fine of 500,000 Pounds for Facebook (this was before GDPR)

https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf

Will Democracy Survive Big Data Breaches?

Cambridge Analytica had around 5000 data points on each targeted voter,
provided by Facebook.

What if it had access to more?

“There is always going to be a Cambridge Analytica”

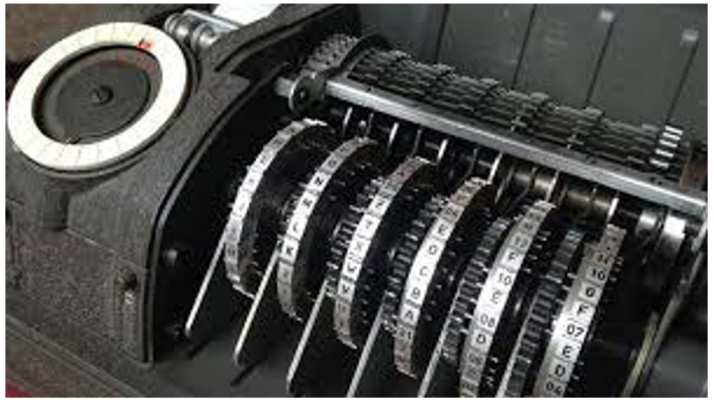
The Chinese Social Credit System

Critics argue that the extensive surveillance, data collection, and monitoring involved in the system can infringe upon individuals' privacy rights.

https://www.wired.co.uk/article/china-social-credit-system-explained (2019)
“The Chinese Social Credit System is a government-led initiative in the People's Republic of China aimed at assessing and scoring the behavior of individuals and organizations based on various factors”

Why is China building this?
It's all about building trust, says the Chinese government. The 2014 document describing the government’s plans note that as "trust-keeping is insufficiently rewarded, the costs of breaking trust tend to be low."

Crypto AG Scandal



- From 1952 to 2018, Crypto AG sold encryption devices that contained a trapdoor to a large number of countries
- For several decades, the (covert) owners of the company were the CIA and its German counterpart
- In spite of complaints from former employees, enquiries by the Swiss authorities (notably in the early 1990's) went nowhere
- By compromising the security of the encryption equipment, the CIA and BND were able to monitor communications of the countries using Crypto AG's products.
- This is considered the largest known spying operation since World War II

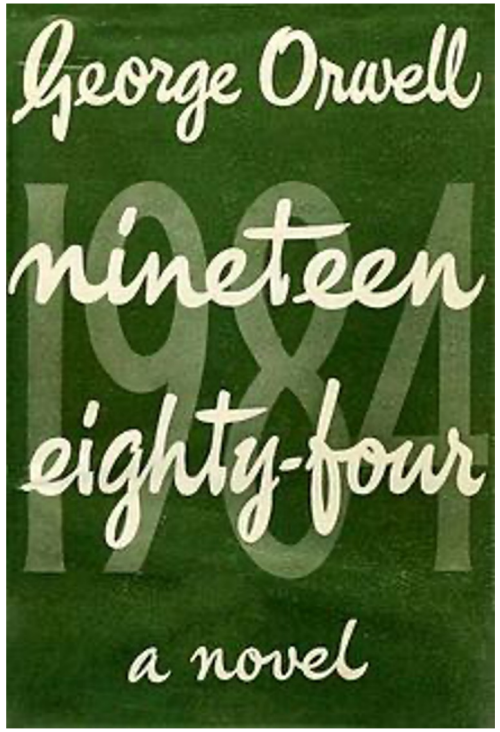
https://www.washingtonpost.com/graphics/2020/world/national-security/cia-crypto-encryption-machines-espionage/

Data Breaches

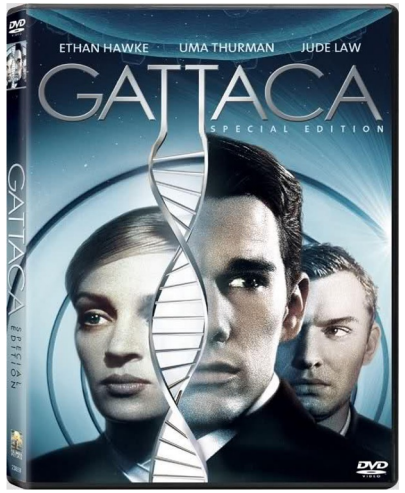
https://en.Wikipedia.org/wiki/List_of_data_breaches

34

Fiction related to privacy

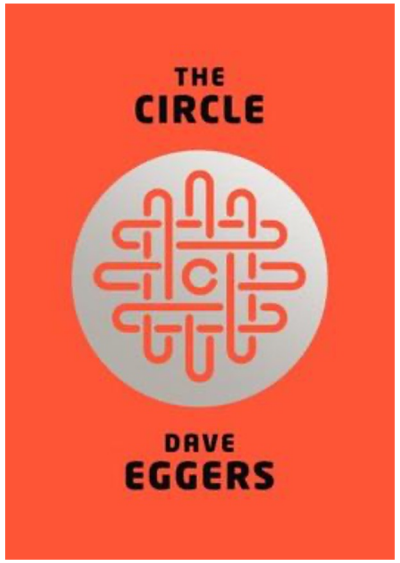


1949



1997

The Lives of Others, 2006



2013

2017

Privacy Loss as an *Overwhelming Side-Effect* of IT

- «Global warming» of Information Technology:

  The energy-based economic development has negative side-effects (pollution, depletion of fossil fuel and other resources, severe accidents - oil spills, nuclear disasters/proliferation - climate change), that are **not properly included in the benefit/cost analysis**

- Likewise, IT development paves the way to an **unprecedented assault** on privacy

Why is Privacy Tricky to Understand?

- Abstract concept
- Diverging opinions
- **Multi-disciplinary:** computer science, law, ethics, economics, sociology, politics,…
  In computer science alone: applied crypto, applied statistics, HCI, inference techniques, machine learning, databases, signal processing, networking, operating systems,…
- There is **no textbook** on the technical aspects (and the field is evolving very fast, making the writing of such a book highly problematic)

Players

- Government
- Companies
  - Advertisers
- Organizations we work for
- Colleagues
- Family and friends
- Strangers

Why is privacy difficult to promote?

Privacy Protection is often
Perceived as being at odds with:

Security (e.g., homeland security)

Usability

Business (e.g., targeted advertisement)

System performance

Medical progress

Benefits of system usage are immediate, drawbacks (in terms of privacy) usually are uncertain and come later

If You are still unconvinced about the relevance of privacy, consider this:

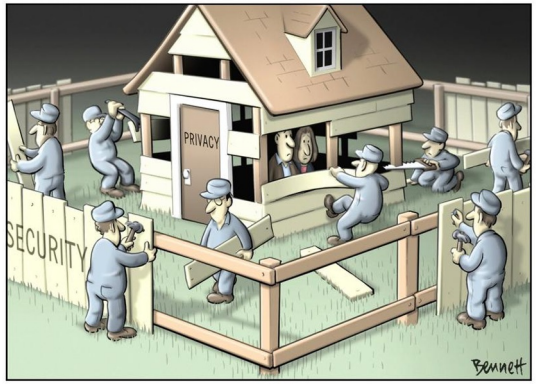
- **All** dictatorships intrude into the privacy of their citizens
- **All** self-respecting democracies have detailed regulations to protect the privacy of their citizens

The context: Availability of data

Intelligent data-based applications

- Recommendation systems
  - Movies (Netflix)
  - Products (Amazon)
  - Friends (Social networks)
  - Music (Spotify, iTunes)
- Location-based services
  - Friend finders
  - Maps
  - Points of interest
- Health monitoring
- Children/Elderly trackers
- Smart metering
- Intelligent buildings

We need privacy!



But what about security!!?!?!?!

Why Privacy is important for society

Daniel Solove,
Prof. of Law

"Part of what makes a society a good place in which to live is the extent to which it allows people freedom from the intrusiveness of others. **A society without privacy protection would be suffocation.**"

Not so much Orwell's "Big Brother" as Kafka's "The Trial":

"...a bureaucracy with inscrutable purposes that uses people's information to make important decisions about them, yet denies the people the ability to participate in how their information is used"

"The problems captured by the Kafka metaphor are of a different sort than the problems caused by surveillance. They often do not result in inhibition or chilling. Instead, they are problems of information processing—the storage, use, or analysis of data—rather than information collection."

"...not only frustrate the individual by creating a sense of helplessness and powerlessness, but they also affect social structure by altering the kind of relationships people have with the institutions that make important decisions about their lives."

The business of data exploitation

Why Do Some Companies Want Our Data?

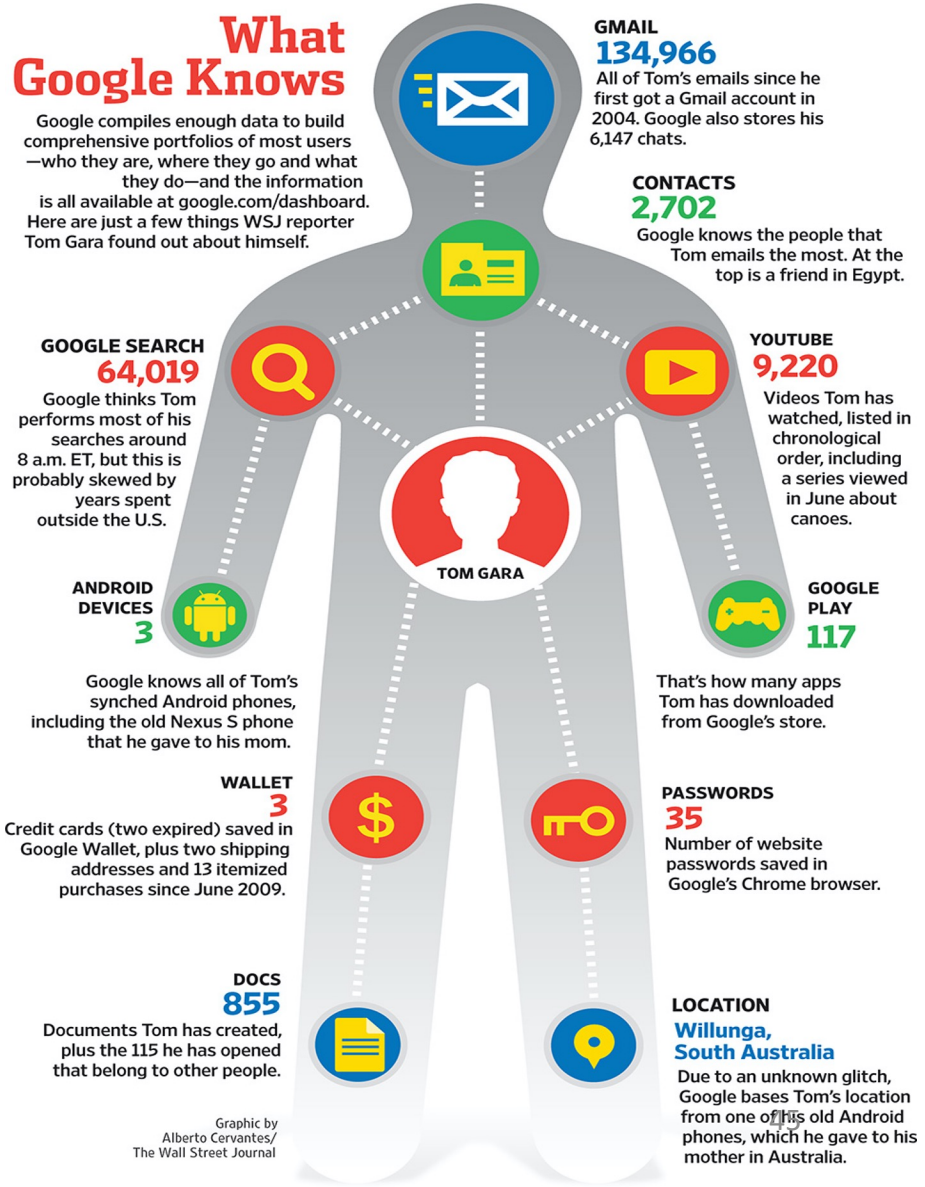
- Information and communication technologies make life better
- For applications to improve their performance and offer us better services, they require information about us
- Social networking sites and most online services are usually free to join
  - Service providers collect our data and use it for “targeted advertising”

If you're not paying for it, you're part of the product

44

What Do Companies Know About Us?

- Some companies **have very detailed profiles** of who we are
  - Can pose a problem if data is used to discriminate on the basis of assumed health status, age, gender, sexual orientation, religion, etc.
  - Even more problematic if governments seek access to and use these data
- Most of us are happy to give out personal information in exchange for specific "free" services
- What we tend to object to is the improper collection of personal information, and the secondary use of information once it's collected
  - The buying and selling of our information behind our back



45

Ways Companies Collect Our Data

- Companies collect information about us:
  - Companies such as Google, Facebook, Amazon,…, collect personal information, data about the services you use and how you use them, server logs, location information, etc.
- We voluntarily share our information
  - Facebook's 2 billion+ active users share roughly 10 billion items a day, not counting the data Facebook collects about them

46

Cloud Computing

- Companies usually store/process data on the cloud
- Worries about control of the data and their geographical location
  - Who has access to it?
  - How can it be used?
  - How easy is it to move the data from one cloud service to another?
  - How secure are they?
  - Who is responsible if the data are lost or misused?

EU-US: invalidation of the Data Protection Shield, July 2020 (“Schrems II”)

https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf

47

2013: PRISM

- Since 9/11 terrorist attack (2001), the US government has dramatically increased the ability of its intelligence agencies to **collect and investigate information** on **both foreign subjects and US citizens**

- Some of these surveillance programs capture the private data of citizens who are not suspected of any connection to terrorism or any wrongdoing

- In June 2013, a private contractor working for Booz Allen Hamilton, Edward Snowden, leaked classified presentation slides that detailed the existence of PRISM

48

PRISM – Who was Involved?

- PRISM: tool used by the NSA to collect private electronic data belonging to users of major internet services
  - Latest evolution of the US government's post-9/11 electronic surveillance efforts, which began under President Bush with the ***Patriot Act***, and expanded to include the ***Foreign Intelligence Surveillance Act (FISA)*** enacted in 2006 and 2007

PRISM - Scope

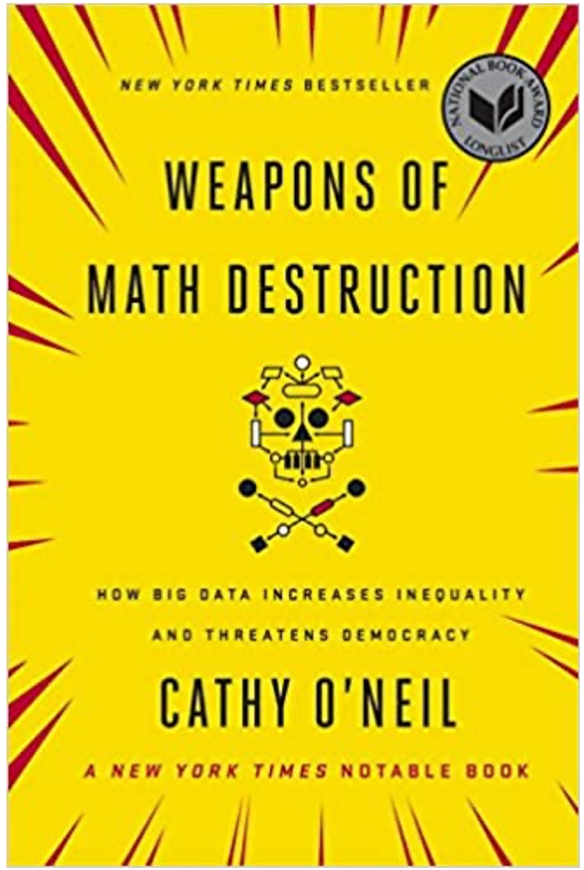
- In theory, NSA analysts are not allowed to specifically target someone "reasonably believed" to be a US person communicating on US soil
  - According to The Washington Post, an analyst must have at least "51 percent" certainty their target is foreign

- But NSA's "contact chaining" practices can easily cause innocent parties to be caught up in the process
  - An analyst collects records on a target's contacts, and their contacts' contacts

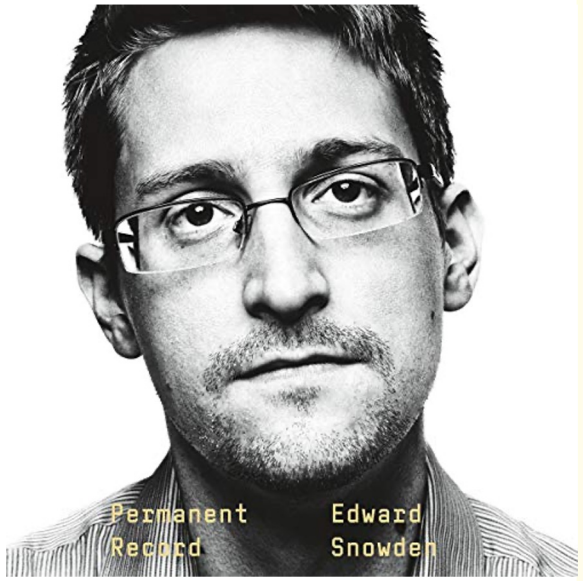
50


// A FILM BY ACADEMY AWARD® WINNER:
OLIVER STONE
THE ONLY SAFE PLACE IS ON THE RUN
SNOWDEN
JOSEPH GORDON-LEVITT // SHAILENE WOODLEY
// IN THEATERS SEPTEMBER 16
// DIRECTED BY:
OLIVER STONE
// SCREENPLAY BY:
KIERAN FITZGERALD & OLIVER STONE

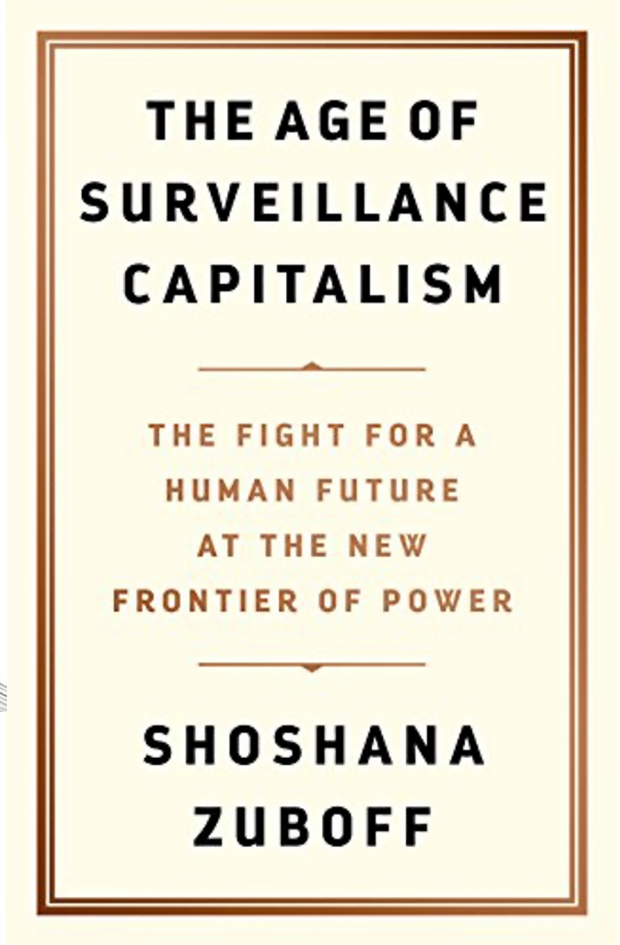
More on the topic of data exploitation / people manipulation



2016



2019



2019

2020

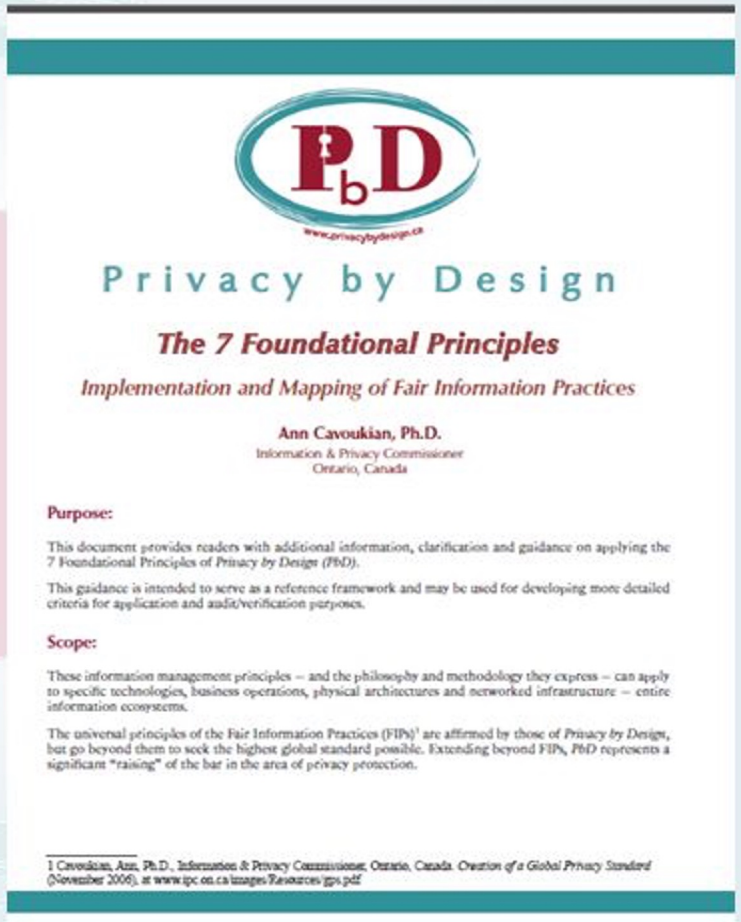
Privacy by Design

Privacy by Design (PbD) - Motivation

- To facilitate a flourishing Internet economy, consumers need to be able to trust the services they use online
- They should not need to worry about their giving companies more data than necessary for the service being used
- It is crucial to ensure that privacy protections are built into the design and implementation of the products and services

The Seven Foundational Principles of PbD

1. ***Proactive*** not Reactive: Preventative, not Remedial
2. Privacy as the ***Default***
3. Privacy ***Embedded*** into Design
4. *Full* Functionality: ***Positive-Sum***, not Zero-Sum
5. End-to-End ***Security***: Full Lifecycle Protection
6. Visibility and Transparency: ***Keep it Open***
7. Respect for User Privacy: Keep it ***User-Centric***



PbD

www.privacybydesign.ca

Privacy by Design

The 7 Foundational Principles

Implementation and Mapping of Fair Information Practices

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Purpose:

This document provides readers with additional information, clarification and guidance on applying the 7 Foundational Principles of *Privacy by Design (PbD)*.

This guidance is intended to serve as a reference framework and may be used for developing more detailed criteria for application and audit/verification purposes.

Scope:

These information management principles – and the philosophy and methodology they express – can apply to specific technologies, business operations, physical architectures and networked infrastructure – entire information ecosystems.

The universal principles of the Fair Information Practices (FIPs)[1] are affirmed by those of *Privacy by Design*, but go beyond them to seek the highest global standard possible. Extending beyond FIPs, *PbD* represents a significant "raising" of the bar in the area of privacy protection.

1 Cavoukian, Ann, Ph.D., Information & Privacy Commissioner, Ontario, Canada. *Creation of a Global Privacy Standard* (November 2006), at www.ipc.on.ca/images/Resources/gps.pdf

The Seven Foundational Principles of PbD

- **Proactive** not Reactive; **Preventative** not Remedial
  - PbD anticipates and prevents privacy invasive events before they happen
- Privacy as the **Default Setting**
  - Privacy is built into the system, by default
- Privacy **Embedded** into Design
  - Privacy is integral to the system, without diminishing functionality
- Full Functionality — **Positive-Sum**, not Zero-Sum
  - PbD seeks to accommodate all legitimate interests and objectives in a positive-sum "win-win" manner
  - Avoids misconceptions such as privacy vs. security
- End-to-End Security — **Full Lifecycle Protection**
  - All data are securely retained, and then securely destroyed at the end of the process, in a timely fashion
- **Visibility** and **Transparency** — Keep it **Open**
  - Its component parts and operations remain visible and transparent, to users and providers
- **Respect** for User Privacy — Keep it **User-Centric**

56

Privacy by Design and Default

- PbD means controllers of data take a positive approach to protecting privacy, by embedding it into both technology and into their organizational policies
- When such protections are built-in from the beginning, they can help to prevent invasions of privacy rights
- *Privacy by default*: when a user receives a product or service, privacy settings should be as strict as possible, without the user having to change them
- Sharing does not inherently mean an end to privacy
  - With effective privacy by design and by default, you can have both

Technical approaches to privacy: privacy-enhancing technologies (PETS)

a. Definition and classification
- Based on privacy paradigm
- Based on adversarial model

PETs Classification through paradigms

What is privacy: privacy paradigms

Privacy as
CONFIDENTIALITY

Privacy as
CONTROL

Privacy as
PRACTICE

Privacy as

CONFIDENTIALITY

"The right to be let alone"
Warren & Brandeis (1890)

"*the individual shall have full protection in person and in property.*"

PETs in this paradigm

1) Minimize data disclosure: every bit counts

2) Distribute trust: avoid single points of failure

3) Rely/require open source: million eyes help security

In math we believe
strong proofs of security

Privacy as

CONFIDENTIALITY

Example PETs in this paradigm

1) Minimize data disclosure:

Encrypt data in transit and/or storage

Compute while encrypted (*Homomorphic encryption*)

Obfuscate (*generalize, perturb, …*)

2) Distribute trust:

Split data and store in several places (*Secret sharing*)

Require several people to compute / decrypt (*Multiparty computation*)

Privacy as

CONTROL

"The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"
Westin (1970)

PETs in this paradigm

1) User participation: let the user decide how data will be shared

2) Transparency and Accountability: let the user know how data is used, and if against his will, point to who is responsible

3) Organizational compliance:

EU: General Data Protection Regulation (GDPR)

US: Fair Information Practice Principles (FIPPs)

Privacy as

CONTROL

"the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"
Westin (1970)

PETs in this paradigm

1) User participation:

Privacy settings

Privacy policy languages

2) Transparency and Accountability:

Secure logging

Privacy as

PRACTICE

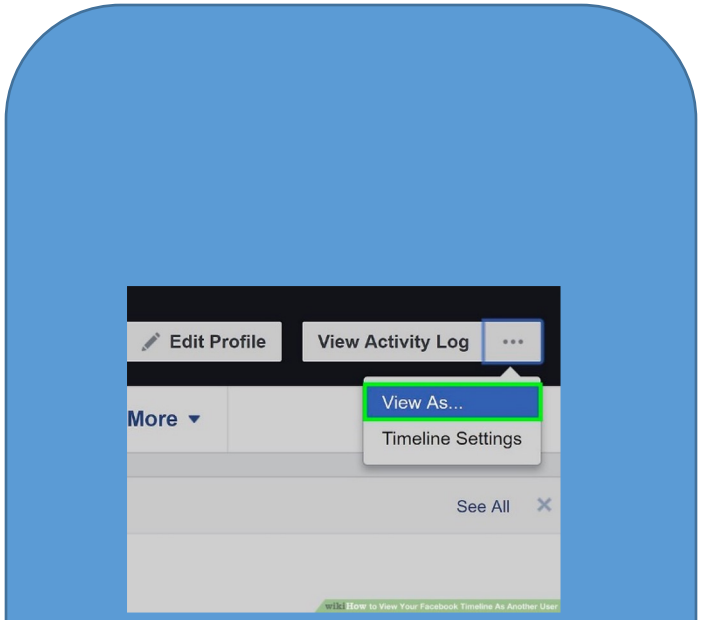
"The freedom from unreasonable constraints on the construction of one's own identity"
Agre (1999)

PETs in this paradigm

1) Improve user agency: help them negotiate privacy

2) Aid decision making and transparency of social impact: help users understand the consequences of their actions

3) Privacy as a collective practice: help identify best practices for collectives



Contextual feedback

Privacy nudges

> **"the freedom from unreasonable constraints on the construction of one's own identity"**
> Agre (1999)

PETs in this paradigm

Improve user agency; Aid decision making and transparency of social impact; Privacy as a collective practice

Contextual feedback (aka, privacy mirrors)
Recommenders for configuration
Privacy nudges

A Different PETs Classification

A different PETS classification

This classification is not better or worse than the one based on paradigms, **just different**

Paradigms are great help to understand different conceptions of privacy
- they are somehow hard to connect to privacy in real scenarios
- do not make adversarial / threat model explicit

We will now see another classification according to:
- who defines the problem (**thus defines who is the adversary**)
- what are the privacy goals (**what should be protected and what protecting means**)

They allow to see PETs limitations and the challenges they pose

Gürses, Seda, and Claudia Diaz. "Two tales of privacy in online social networks." IEEE Security & Privacy 11.3 (2013): 29-37.
Diaz, Claudia, and Seda Gürses. "Understanding the landscape of privacy technologies." Information Security Summit (2012): 58-63.
Danezis, George, and Seda Gürses. "A critical review of 10 years of privacy technology." Surveillance cultures: a global surveillance society (2010): 1-16.

Threat model, threat, harms, and vulnerabilities

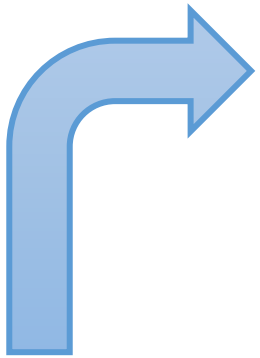
Threat model

The adversary's capabilities.
Examples:

The adversary can observe my connection

The adversary can corrupt my machine

The adversary controls a Facebook employee



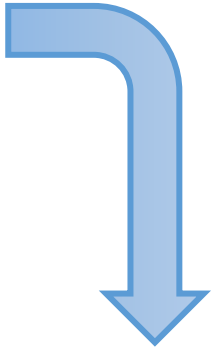
Vulnerability

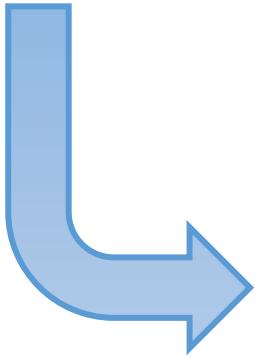
Specific weakness that could be exploited by adversaries with interest in a lot of different assets

Examples:

I am revealing personal information on online social networks

My Facebook settings are very permissive





Threat

Who might attack which assets, using what resources, with what goal, how, and with what probability.

Examples:

A consultancy company wants to profile me from my web searches to offer influential ads

A burglar wants to learn when I am traveling from my interactions in social networks

Harm

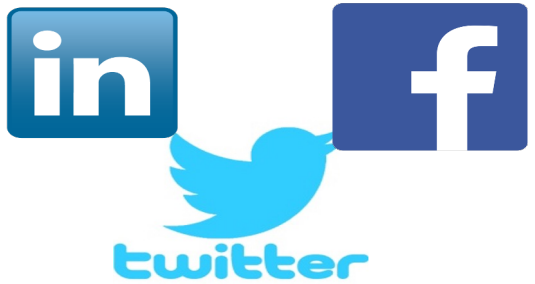
The bad thing that happens when the threat materializes

Examples:

I receive targeted advertising that aims at influencing my vote

A burglar enters in my place while I am abroad

1 – PETs for “social” Privacy



CONCERNS - The privacy problem is defined by **Users**

Technology brings problems

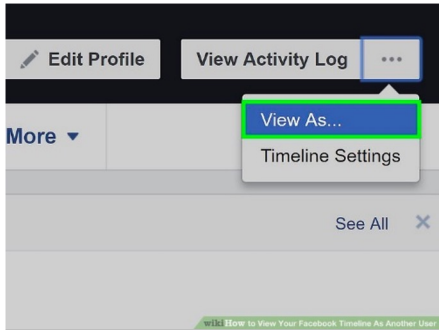
- “My parents discovered I'm gay”
- “My boss knows I am looking for another job”
- “My friends saw my embarrassing photos”

GOALS - Do not surprise the user

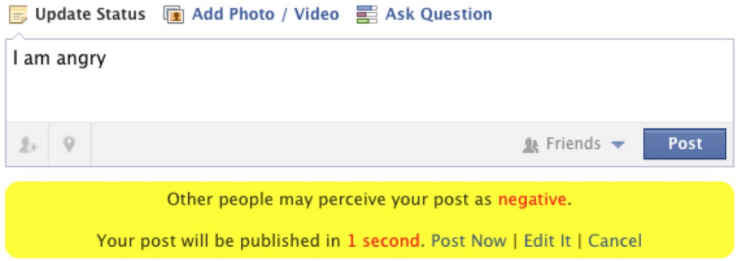
Two main approaches

- Support decision making
- Help identifying actions impact

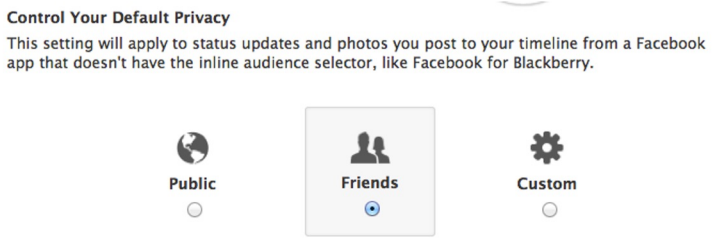
Common Industry approach: Make users comfortable



Contextual feedback



Privacy nudges



Easy defaults

1 – PETs for “social” Privacy

LIMITATIONS

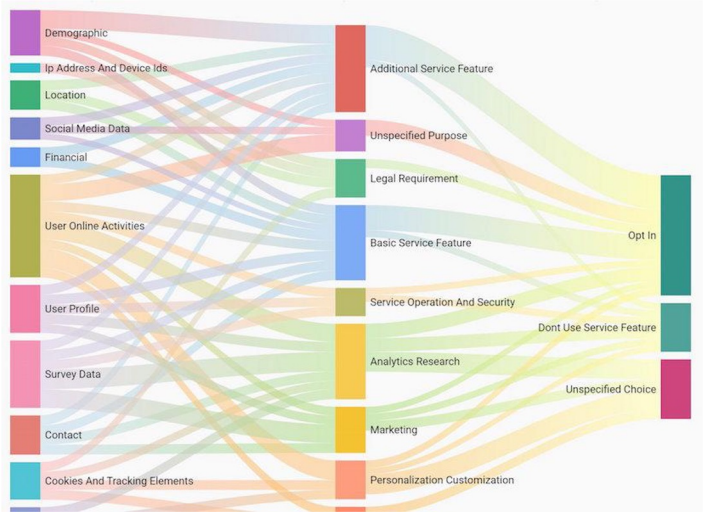
Only protects from other users: **trusted service provider**!
Limited by user’s capability to understand policies

Privacy Policy After GDPR

COMPANY	WORD COUNT	READING TIME*	READING LEVEL
Google	4036	19:11	14
Facebook	4233	20:41	13
Reddit	3414	16:39	12
Amazon	3837	18:24	13
Wikipedia	5617	27:06	14
Yahoo	2225	11:12	13
Twitter	4880	22:25	13
eBay	5666	27:32	20
Instagram	4221	20:38	13
Netflix	3417	16:39	16

* Minutes:Seconds

https://www.varonis.com/blog/gdpr-privacy-policy/



https://pribot.org/polisis

1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!
Limited by user’s capability to understand policies

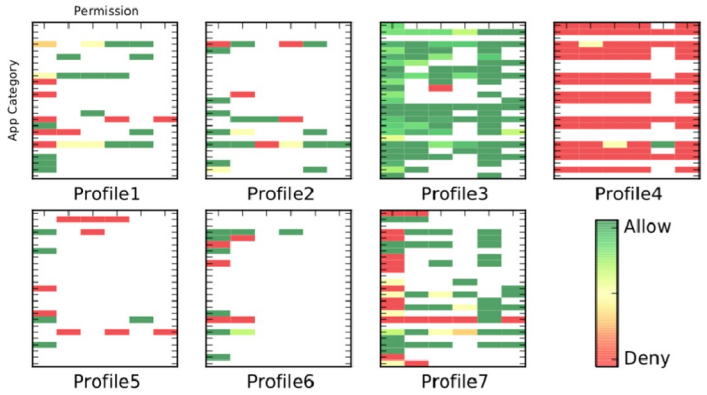


Figure 2: Privacy profiles learned from collected app privacy settings. Profile 1 is more protective on Location and Productivity apps than other profiles. Profile 2 denies phone call log permission more. Profile 3 is generally permissive. Profile 4 denies most permission requests. Profile 5 generally denies contacts, message, phone call log and calendar access, with only location and camera allowed for some apps. Profile 6 denies location and contact access of Social apps and Finance apps. Profile 7 is stricter regarding Social apps and location access in general.

Automated configuration

A. **Is good for any user**
B. **Only works for average users**
C. **Only works for outliers**
D. **Has problems for everyone**

Liu et al., Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permission, USENIX, 2016.

1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!
Limited by user’s capability to understand policies
Based on user expectations – What if the expectations are null?

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

Personal data: any information that relates to an identified or identifiable living individual.

≠

Personal Identifiable Information (PII)
NIST Special Publication 800-122

any information about an individual maintained by an agency, including
(1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and
(2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

(1) is directly sensitive data
(2) combines data from same service or from different services

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent

purpose limitation

data minimization

subject access rights

Preserving the security of data

Auditability and accountability

valid, freely given, specific, informed and active consent

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent

purpose limitation

data minimization

subject access rights

Preserving the security of data

Auditability and accountability

Data can only be used for the purpose it was collected

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent

purpose limitation

data minimization

subject access rights

Preserving the security of data

Auditability and accountability

One should only collect the data necessary for the purpose of the service (**proportionality**)

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent

purpose limitation

data minimization

subject access rights

Preserving the security of data

Auditability and accountability

One should be able to know what information is stored/processed and how. Also right to modification, deletion, etc.

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent
purpose limitation
data minimization
subject access rights

Preserving the security of data

Auditability and accountability

The controller must provide means to prove that information is treated as promised, and appoint responsibility otherwise.

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by **Legislation**

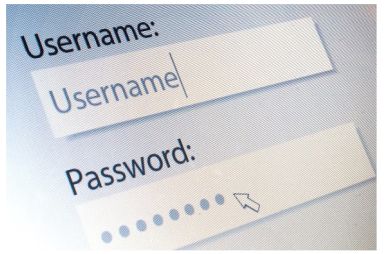
Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

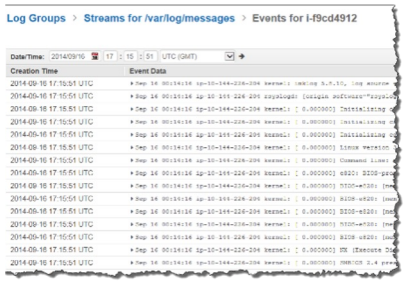
GOALS – Compliance with data protection principles

informed consent
purpose limitation
data minimization
subject access rights

Preserving the security of data
Auditability and accountability



Access control



Logging

Anonymization????

Policy{Entity (#business.name): walmart.com, . . .,
S_1 {Purpose: (current, contact [opt-in]),
Recipient: (ours),
Retention: (indefinitely),
Data: (#user.login, #user.home-info)}
S_2 {Purpose: (current, develop [opt-in], contact [opt-in]),
Recipient: (ours),
Retention: (stated-purpose),
Data: (#user.name, #user.login, #user.home-info)}}

Automated Policy Negotiation

Personal data

any information that relates to an identified or identifiable living individual.

Wouldn't it be nice if... you could take a dataset full of personal data, and transform it into one with no personal data – while keeping all the value of the data? **THIS IS HARD!!!!**

2 – PETs for “institutional” Privacy

LIMITATIONS

- Assumes:
  - collection and processing by organizations is necessary
  - organizations are (semi)-trusted and honest
    - Reliance on punishment
    - No mandated technique for the protection of the data
- Focuses on limiting misuse, **not** collection
  - Easy to circumvent minimization to collect in bulk
  - Auditing may require more data!
  - The danger of *informed consent*: if compliant is ok!

Widespread IT industry approach:
Make users comfortable
+ Legal compliance

3 – PETs for “anti-surveillance” Privacy

CONCERN: HOW TO EVADE / FOOL A GLOBAL ADVERSARY?

APPROACH: Minimize

Minimize the need to trust others

Minimize the amount of revealed information

End-to-End encryption: Signal, PGP, OTR (Of the record messaging)

Anonymous comms: Tor, mixnets

Obfuscation:
- dummy actions
- hiding
- generalization
- differential privacy

Advanced crypto:
- Private information retrieval
- Anonymous authentication
- Multiparty computation
- Blind signatures
- Cryptographic commitments

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

Privacy-preserving designs are narrow – difficult to create “general-purpose privacy”

- Difficult to evolve
- Also difficult to combine

Usability problems both for developers and users

- how do I program this?
- performance hit
- unintuitive technologies

Lack of incentives:

- Industry: loses the data!
- Governments: national security, fraud detection, …

Privacy Properties

Privacy properties

Confidentiality (in transit/storage, during processing)

- game-oriented definitions
- cryptographic definitions

Pseudonymity

Anonymity

Unlinkability

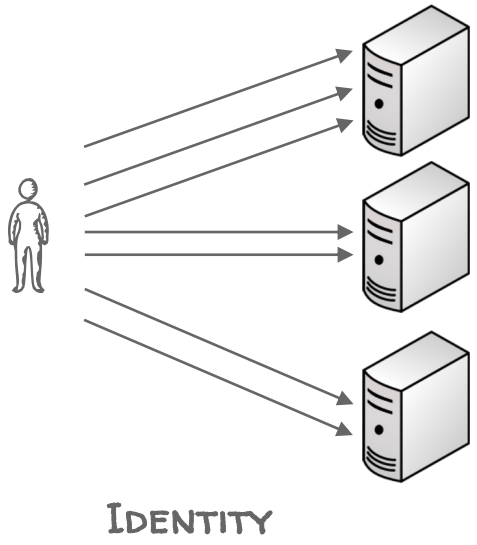
Unobservability

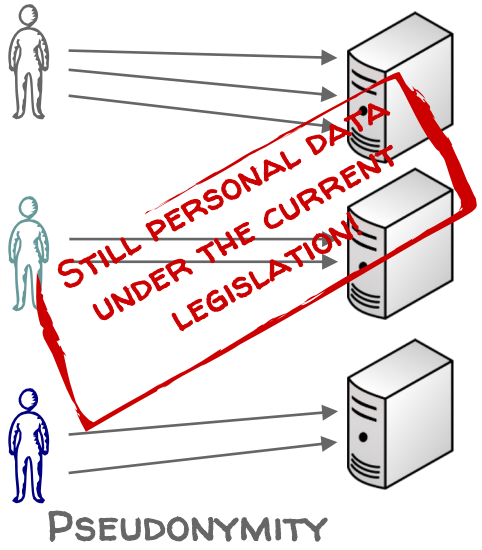
Plausible deniability

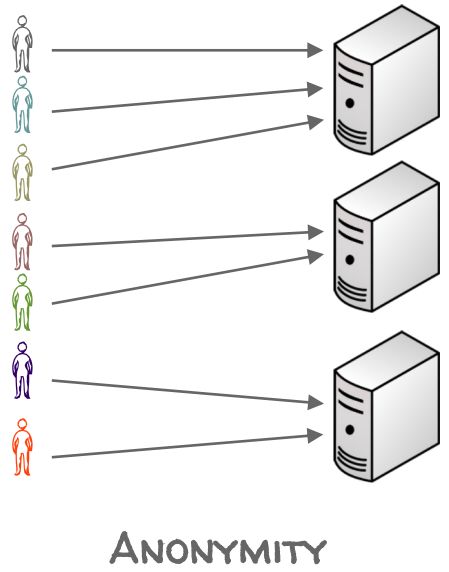
Privacy properties: **Pseudonymity**

-Pfitzmann-Hansen: "the use of pseudonyms as IDs [...] A digital pseudonym is a bit string which is unique as ID and which can be used to authenticate the holder"

-ISO 15408: "a user may use a resource or service without disclosing its identity, but can still be accountable for that use."







Pfitzmann, Andreas and Hansen, Marit. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

Examples of pseudonymity technologies

- Use of persistent (random) identifiers
- Use of hashed identifiers
- Different email addresses for the same user
- Nicknames

These techniques are very limited from a privacy perspective, more in the next lecture

- Pseudo-identifiers, Quasi-identifiers

De-identification

- Removing or obscuring information from (electronic) traces that would allow direct identification of a person

Age	Gender	Zipcode
34	male	81667
45	female	81675
66	male	81925
70	female	81931

Age	Gender	Zipcode
<50	*	816**
<50	*	816**
≥ 50	*	819**
≥ 50	*	819**

- Advantages:
  - Allows research that would otherwise not be possible due to privacy concerns
- Major misconception:
  - Governments, industry and researchers tend to claim that de-identification of personal data e.g. by pseudonymization is actually anonymization and that it can help society to ensure the availability of rich data resources whilst protecting individuals' privacy
- Unfortunately, this is simply not the case
  - Netflix, Massachusetts medical records

Privacy properties: **Anonymity**

-Pfitzmann-Hansen: “Anonymity is the state of being not identifiable within a set of subjects, the anonymity set [...] The anonymity set is the set of all possible subjects who might cause an action”

-ISO 29100: “a characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly”

Who is...

...the reader of a web page, the person accessing a service

...the sender of an email, the writer of a text

...the person to whom an entry in a database relates

...the person present in a physical location

DECOUPLING IDENTITY AND ACTION

Privacy properties: **Unlinkability**

-Pfitzmann-Hansen: "two or more items within a system, are no more and no less related than they are related based on the a-priori knowledge"

– **ISO 15408**: " a user may make multiple uses of resources or services without others being able to link these uses together "

Unlinkability ensures that a user may make multiple uses of resources or services without others being able to link these uses together

Two...

... anonymous letters written by the same person
... web page visits by the same user
… entries in a databases related to the same person
…subsequent usage sessions by the same user

DECOUPLING TWO ACTIONS FROM ONE USER

Privacy properties: **Unobservability**

– **Pfitzmann-Hansen**: “an items of interest being indistinguishable from any item of interest at all […] Sender unobservability then means that it is not noticeable whether any sender within the unobservability set sends.”

– **ISO15408:** “a user may use a resource or service without others, especially third parties, without being able to observe that the resource or service is being used.”

Hiding...

...whether someone is accessing a web page

...whether a message is being sent

...whether an entry in a database corresponds to a real person

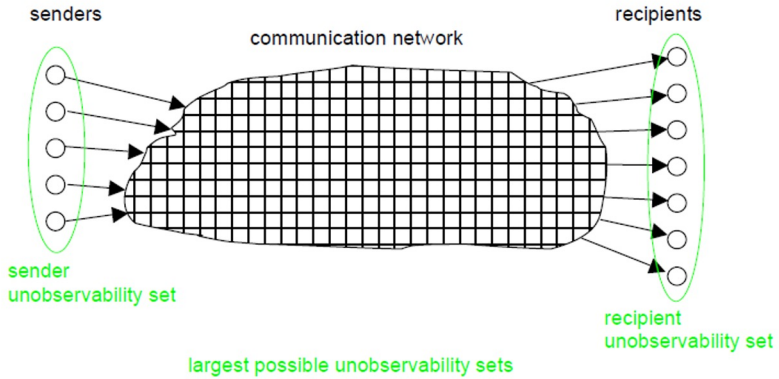
...whether someone or no one is in a given location

...

DECOUPLING OBSERVATION FROM ACTION EXISTENCE

Privacy properties: **Unobservability**

Sender unobservability means it is not noticeable whether any sender within the unobservability set sends
Recipient unobservability means it is not noticeable whether any recipient within the unobservability set receives



Hiding...

...whether someone is accessing a web page
...whether a message is being sent
...whether an entry in a database corresponds to a real person
...whether someone or no one is in a given location
...

DECOUPLING OBSERVATION FROM ACTION EXISTENCE

Unobservability vs. Anonymity

- Unobservability implies anonymity
- Anonymity does not imply unobservability
  - Anonymity only hides the identity of the sender/receiver, it does not guarantee unobservability

Privacy properties: **Plausible Deniability**

– Not possible to prove user knows, has done or has said something
  – Resistance to coercion: one can always claim ignorance
  – Resistance to profiling: one cannot filter the fake entries

Not possible to prove ...
  ... that a person has hidden information in a computer
  ... that someone has the combination of a safe
  … that a person has been in a place at a certain point in time
  … that a database record belongs to a person

DECOUPLING OBSERVATION FROM TRUE ACTION

Do not confuse privacy properties to security properties. Yet, they might serve for the same purpose

Security properties

Confidentiality (in transit/storage, during processing)	Keep information secret
Integrity	Keep information correct
Availability	Keep information available/systems running
Authenticity (aka integrity of origin)	Demonstrate authenticity of information, prevent fake information

Systematic Privacy Evaluation

Confidentiality (in transit/storage, during processing)

Cryptographic proofs

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Probabilistic expression

Anonymity - Pr[identity → action | observation]

Unlinkability - Pr[action A ↔ action B | observation]

Unobservability - Pr[real action |observation]

Plausible deniability - Pr[fake| observation]

Often times we use obfuscation as a means to achieve these properties

Obfuscation - Pr[real value| obfuscated value]

Systematic Privacy Evaluation

1) Model the privacy-preserving mechanism as a probabilistic transformation

   What is the probability that, given an input the privacy mechanism returns a given output?

2) Determine what the adversary will see

   Threat model: who is the adversary? what are her "observations"? what is her prior knowledge?

3) "Invert" the mechanism, in the way the adversary would do

   Always assume the adversary **knows** the mechanism and would try to undo its effect

4) Evaluate property after inversion

   This is the real probability the adversary can compute

5) Quantify the probability of success of the adversary

   Non trivial!

Anonymity - Pr[identity → action | observation]

Unlinkability - Pr[action A ↔ action B | observation]

Unobservability - Pr[real action |observation]

Plausible deniability - Pr[fake| observation]

Obfuscation - Pr[real value| obfuscated value]

Measuring privacy

Take a property, e.g., Anonymity - Pr[identity → action | observation]

Correctness (e.g., error) :

given the distribution, what is the probability that the adversary does not guess correctly?

Uncertainty (e.g., entropy):

given the distribution, what is the uncertainty of the adversary on the answer?

Accuracy (e.g., confidence intervals):

given the distribution, and a guess, how certain is the adversary of his inference?

Differential privacy-based:

given an observation, can we bound the information the adversary can learn?

Next lecture!

Takeaways on PETs

PETs depend on:

- The privacy paradigm: confidentiality, control, practice

- The adversary model: other users, semi-trusted service provider, everyone,…

Each type of PET offers different properties and presents different challenges

Security properties

Confidentiality (in transit/storage, during processing)	Keep information secret
Integrity	Keep information correct
Availability	Keep information available/systems running
Authenticity (aka integrity of origin)	Demonstrate authenticity of information, prevent fake information

Overall Takeaways of this lecture

- Privacy is a complex topic
- It is at the core of the relationship between citizens and authorities
- In liberal democracies, legislation mandates:
  - protection of citizens' data
  - transparency of most (but not all) government's activities
- With the advent of databases, the Web, and then ML, privacy protection has become a major concern
- PETs are the technical response to the challenge
- The topic is evolving very fast

Let’s exercise your privacy brain

You are developing a new app to rate beer bars in Lausanne. Rightfully, you decide that your target audience are **students**, your customers are the **bar owners**, and you will use a **Cloud service provider** to host the application data.

Compare the following configurations in terms of privacy from the point of view of the students. Identify possible adversaries and what can they learn.

CONFIG A: The application gathers the recommendations from the students and then: lets other students see each other recommendations, and lets the bars see the student recommendations so that they can offer discounts to students that give good ratings.

CONFIG B: The application gathers the recommendations from the students and then: lets other students and the restaurant owners see the average rating for a restaurant**.**

Let’s exercise your privacy brain

Compare the following configurations in terms of privacy from the point of view of the students. Identify possible adversaries and what can they learn.

CONFIG A: The application gathers the recommendations from the students and then: lets other students see each other recommendations and lets the bar owner see the student recommendations so that they can offer discounts to students that give good ratings.

Adversaries: other students, owners, cloud provider
What they can see: all of them can see everything – and thus can learn student behavior and preferences

CONFIG B: The application gathers the recommendations from the students and then: lets other students and the restaurant owners see the average rating for a restaurant**.**

Adversaries: other students, owners, cloud provider
What they can see: students and cloud owners cannot learn students’ behavior anymore;
The Cloud service provider still can see everything and learn student’s behavior and preferences

Let’s exercise your privacy brain

From an adversarial perspective (social, institutional, anti-surveillance), what kind of privacy technologies would you use if:

- You want to protect the students’ social network (who is friends with whom) from the students they do not know

- You do not want the bar owners to learn which bar each student has visited, only the aggregates

- You do not want the cloud provider to learn which students connect

- You want the students to be able to evaluate how much other application users know about them

For each case, what privacy paradigm have you followed?

Let’s exercise your privacy brain

- You want to protect the students' social network (who is friends with whom) from the students they do not know

PET for social privacy: access control would be enough (the question does not specify that the application is an adversary Privacy as control

- You do not want the bar owners to learn which bar each student has visited, only the aggregates

PET for social privacy: access control would be enough (the question does not specify that the application is an adversary) Privacy as control

- You do not want the cloud provider to learn which students connect

Anti-surveillance PET for social privacy: anonymous communications Privacy as confidentiality

- You want the students to be able to evaluate how much other application users know about them

PET for social privacy: privacy mirrors Privacy as practice

Let’s exercise your privacy brain

1. In these situations, what privacy properties are being achieved? (can be more than one)

A. A user accesses a web three days in a row, and the web cannot recognize it is the same user

B. A user has a program in its computer that publishes Tweets automatically. Twitter cannot prove that a given Tweet was made by the user

Let’s exercise your privacy brain

1. In these situations, what privacy properties are being achieved? (can be more than one)

A. A user accesses a Web page three times in a row, and the web page cannot recognize it is the same user

Anonymity, Unlinkability, Plausible deniabilty

B. A user has a program in its computer that publishes Tweets automatically. Twitter cannot prove that a given Tweet was made by the user

Plausible deniability

Let’s exercise your privacy brain

2. Suppose that we create our own system for providing anonymous feedback for the course. What would better hide the link between a student and his/her comment?

A. a system that always mixes the feedback from two students before presenting it to the lecturer (i.e., comment from student Bob appears sent by student Alice and vice versa),
B. a system that with 50% of the time forwards the comment of a student and 50% of the time forwards a fake comment

Let’s exercise your privacy brain

2. Suppose that we would create our own system for providing anonymous feedback for the class. What would hide better the link between a student and a comment:

A. a system that always mixes the feedback from two students before presenting it to the professor (i.e., comment from student Bob appears sent by student Alice and vice versa),
B. a system that with 50% of the time forwards the comment of a student and 50% of the time forwards a fake comment

System A provides **NO** privacy. It always swaps, thus the professor can always link the comment to the correct student.

In system B the professor cannot link the comment to the student with more than 0.5 probability. The students have plausible deniability about what they wrote.

Technical approaches to privacy: privacy-enhancing technologies (PETS)

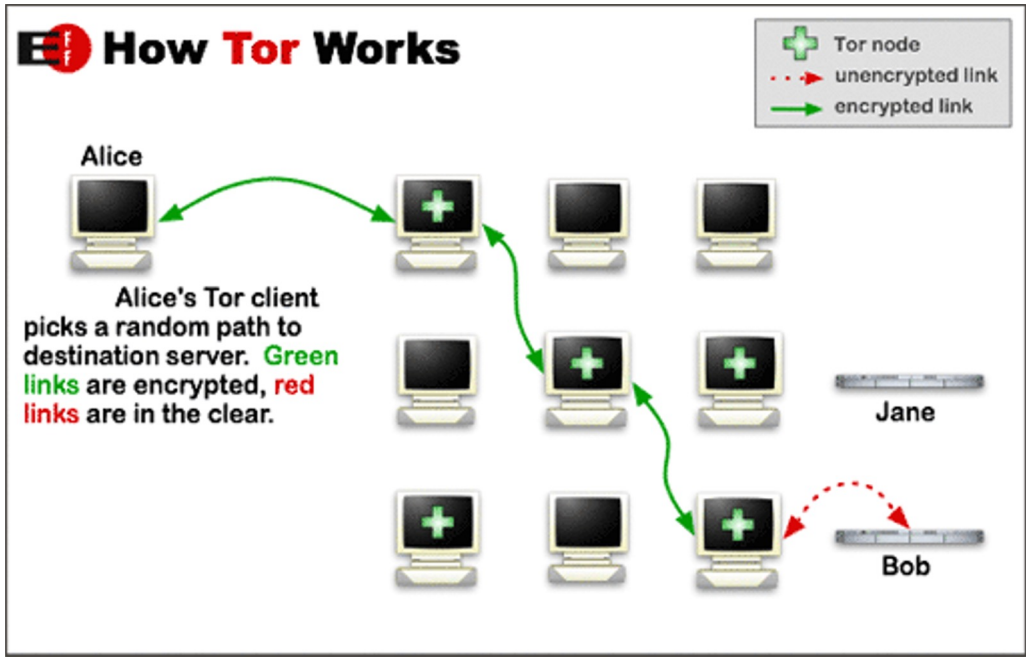
b. Crypto-based solutions

Cryptographic Mechanisms for Privacy Protection

- Anonymous communication
  - Tor
- Anonymous credentials
- Blind signatures
- Secure multiparty computation
  - Garbled circuits
- Searchable encryption
- Deterministic encryption
  - Order-preserving encryption
- Computing on encrypted data
  - Homomorphic encryption
  - Functional encryption
- Oblivious RAM
- Private information retrieval
- Zero-knowledge proofs
- Etc.

Anonymous Communication

- Anonymity of participants is usually achieved by **special routing overlay networks** that hide the physical location of each node from other participants



https://www.torproject.org/about/overview.html.en

Anonymous Credentials (1)

- Allow users to authenticate themselves in a privacy-preserving manner
- Traditional credentials:
  - Alice obtains credentials from an organization
  - At some later point, she proves to the organization (or any other party) that she has been given appropriate credentials
- Anonymous credentials:
  - Alice can do the same without revealing who she is (only that she possesses valid credentials)
  - If she uses her credentials a second time, no one will be able to tell that the two interactions involved the same user

Anonymous Credentials (2)

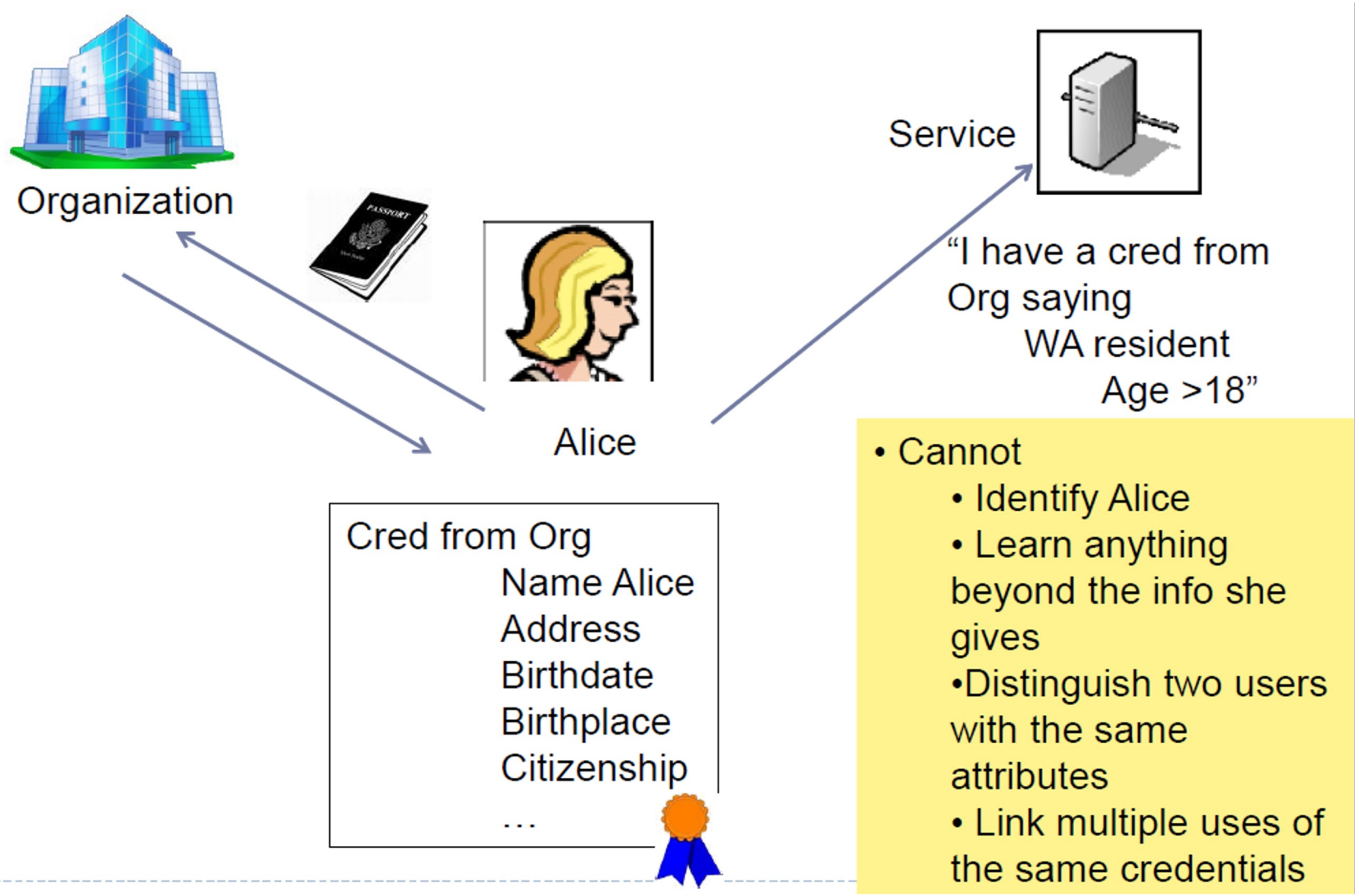


Figure: Melissa Chase – Microsoft Research

Blind Signatures

- Content of a message is blinded before it is signed
- Resulting blind signature can be publicly verified against the original (unblinded) message
- Cryptographic voting systems
  - Authority checks the credentials of the voter to ensure that he is allowed to vote, and that he is not submitting more than one vote
  - Authority does not learn the voter's selections

Secure Multiparty Computation

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

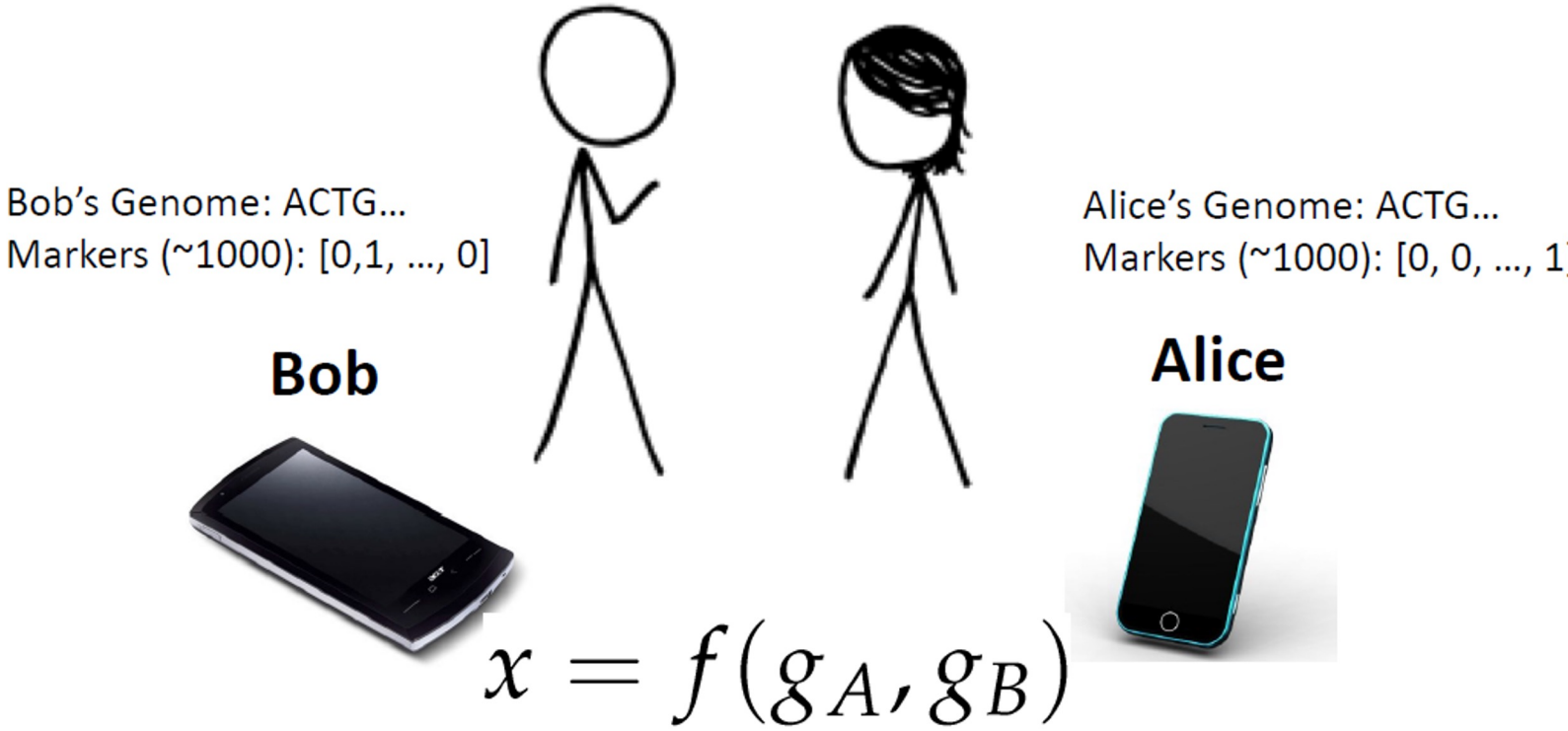
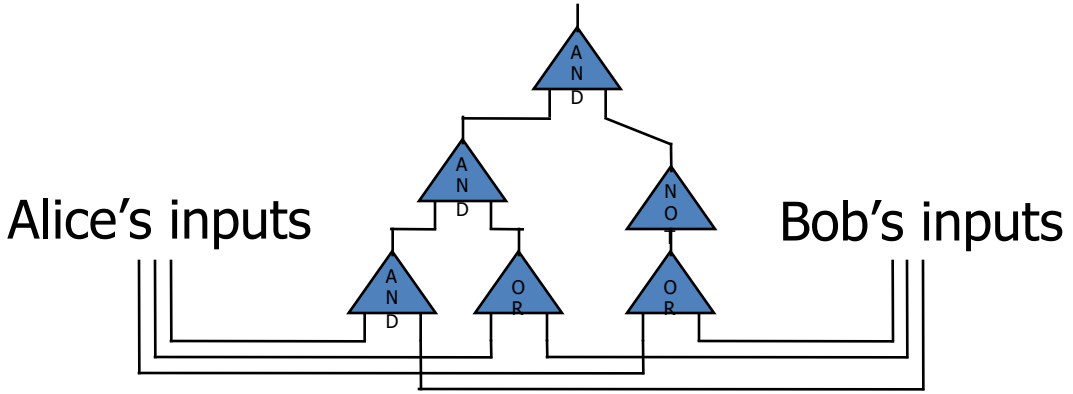


Figure: David Evans et al.

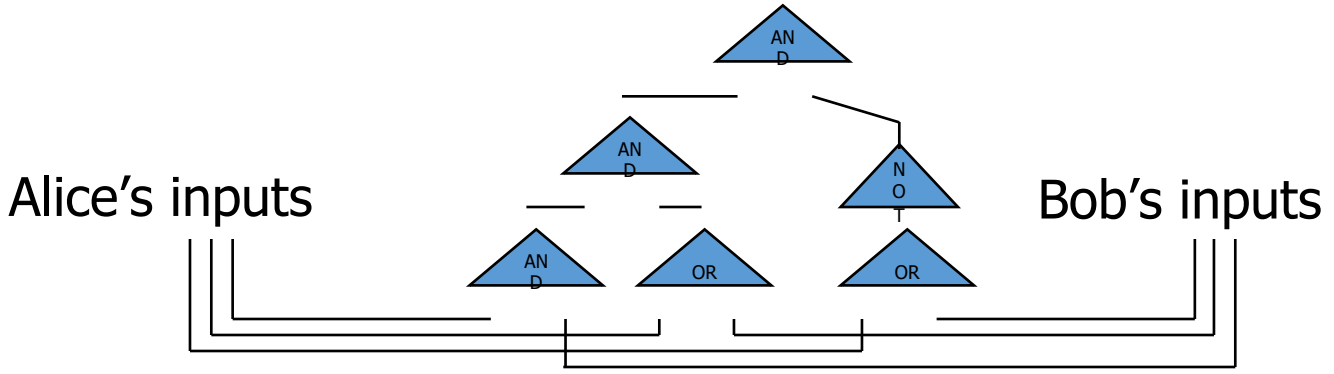
Garbled Circuits

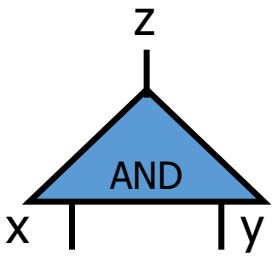
- Bob creates a garbled circuit and sends the circuit to Alice
- Alice evaluates the circuit with her inputs and returns the result to Bob
- The result of the circuit evaluation with Alice's inputs is the output of the function Alice and Bob wish to compute
- Bob does not send his inputs to Alice, instead his inputs are encoded into the garbled circuit such that Alice cannot determine what they are



Yao's Protocol

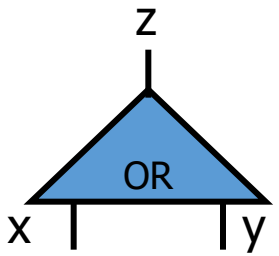
- Compute any function securely
  - … in the semi-honest model
- First, convert the function into a boolean circuit





Truth table:

x	y	z
0	0	0
0	1	0
1	0	0
1	1	1

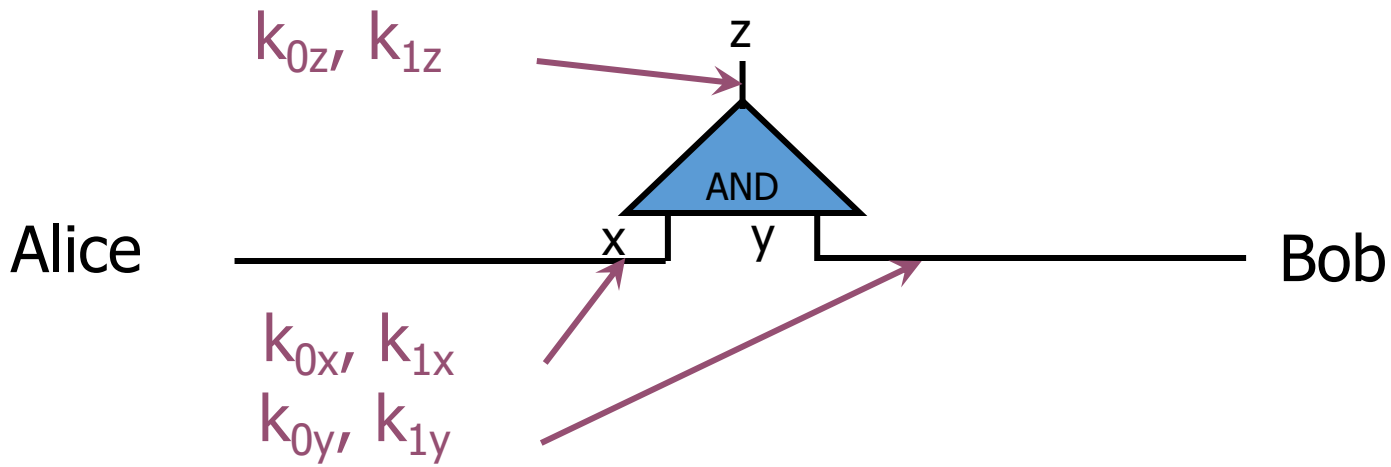


Truth table:

x	y	z
0	0	0
0	1	1
1	0	1
1	1	1

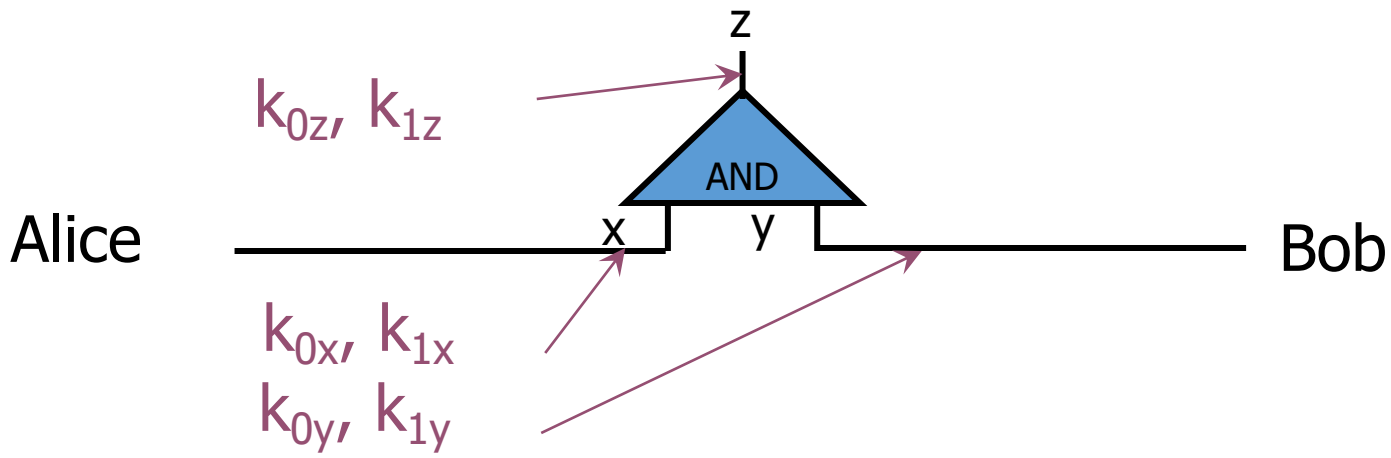
1: Pick Random Keys For Each Wire

- Next, evaluate one gate securely
  - Later, generalize to the entire circuit
- Alice picks two random keys for each wire
  - One key corresponds to “0”, the other to “1”
  - 6 keys in total for a gate with 2 input wires



2: Encrypt Truth Table

- Alice encrypts each row of the truth table by encrypting the output-wire key with the corresponding pair of input-wire keys



Original truth table:

x	y	z
0	0	0
0	1	0
1	0	0
1	1	1

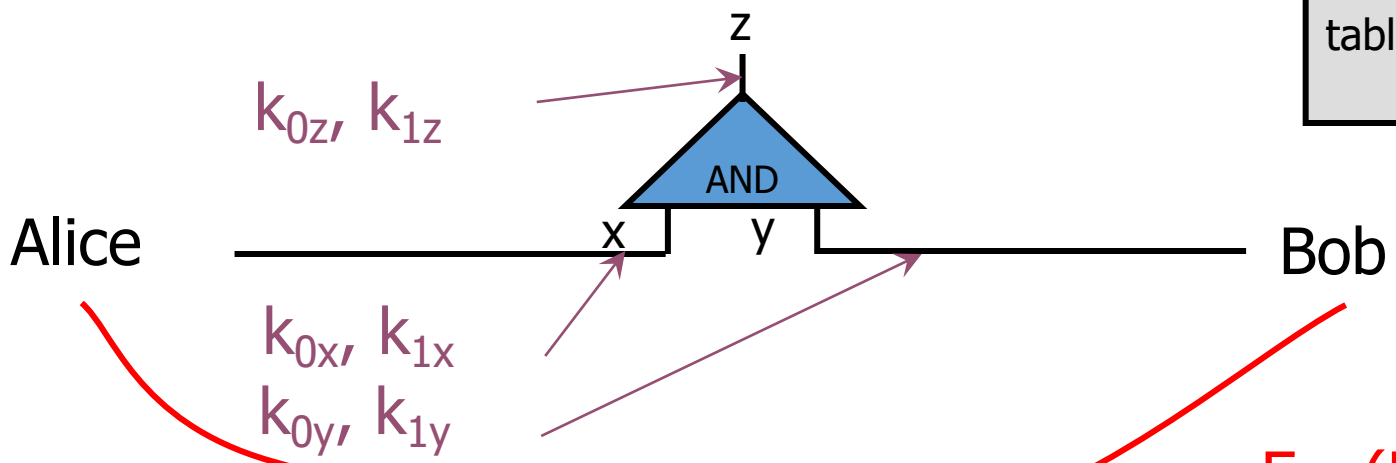
Encrypted truth table:

$E_{k0x}(E_{k0y}(k_{0z}))$
$E_{k0x}(E_{k1y}(k_{0z}))$
$E_{k1x}(E_{k0y}(k_{0z}))$
$E_{k1x}(E_{k1y}(k_{1z}))$

3: Send Garbled Truth Table

- Alice randomly permutes ("garbles") encrypted truth table and sends it to Bob

Does not know which row of garbled table corresponds to which row of original table



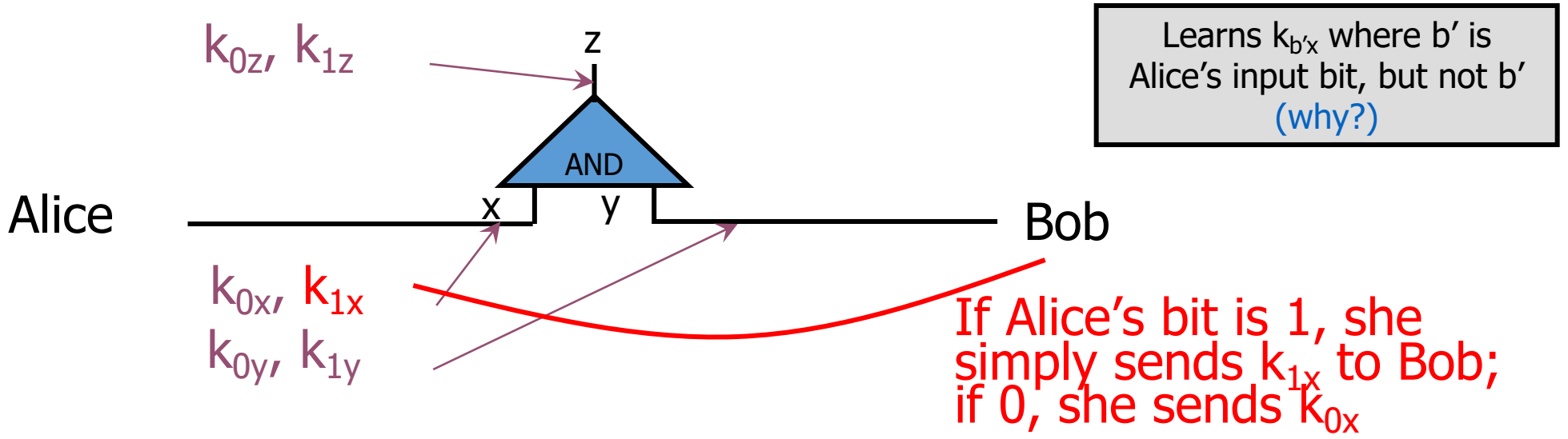
$E_{k0x}(E_{k0y}(k_{0z}))$
$E_{k0x}(E_{k1y}(k_{0z}))$
$E_{k1x}(E_{k0y}(k_{0z}))$
$E_{k1x}(E_{k1y}(k_{1z}))$

Garbled truth table:

$E_{k1x}(E_{k0y}(k_{0z}))$
$E_{k0x}(E_{k1y}(k_{0z}))$
$E_{k1x}(E_{k1y}(k_{1z}))$
$E_{k0x}(E_{k0y}(k_{0z}))$

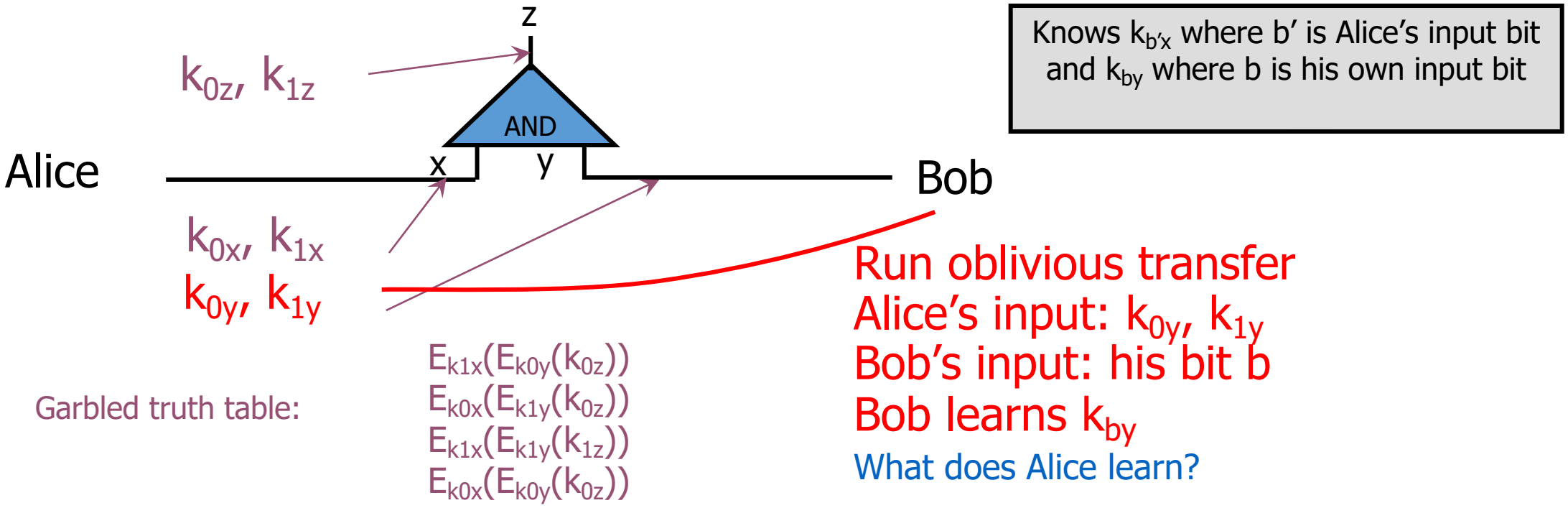
4: Send Keys For Alice's Inputs

- Alice sends the key corresponding to her input bit
  - Keys are random, so Bob does not learn what this bit is



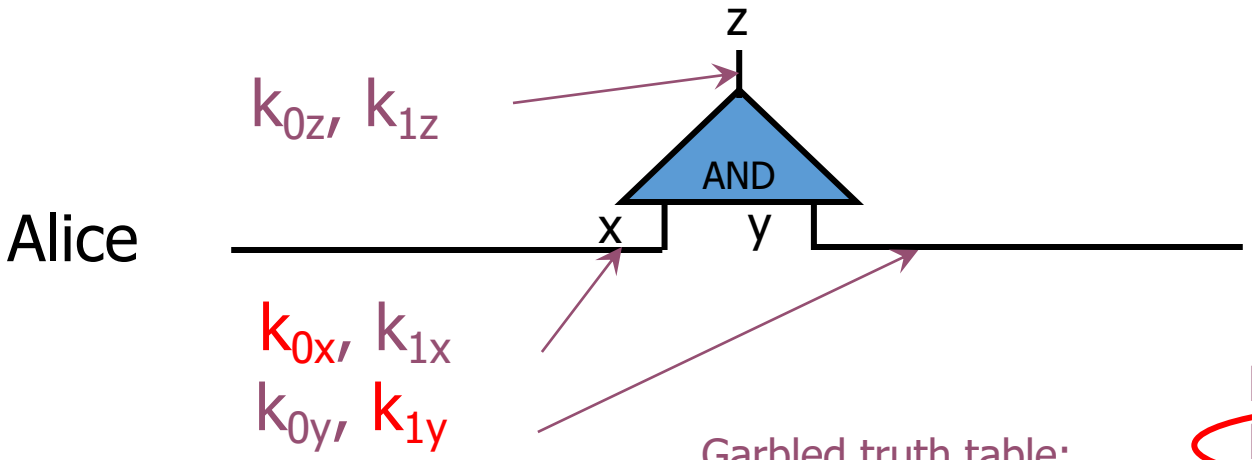
5: Use OT on Keys for Bob's Input

- Alice and Bob run oblivious transfer protocol
  - Alice's input is the two keys corresponding to Bob's wire
  - Bob's input into OT is simply his 1-bit input on that wire



6: Evaluate Garbled Gate

- Using the two keys that he learned, Bob decrypts exactly one of the output-wire keys
  - Bob does not learn if this key corresponds to 0 or 1
    - Why is this important?



Knows $k_{b'x}$ where b' is Alice's input bit and k_{by} where b is his own input bit

Garbled truth table:

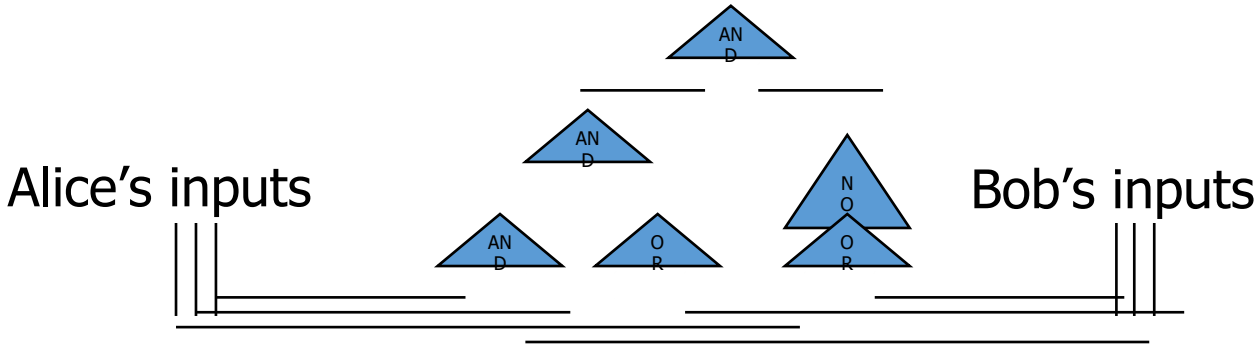
$E_{k1x}(E_{k0y}(k_{0z}))$
$E_{k0x}(E_{k1y}(k_{0z}))$
$E_{k1x}(E_{k1y}(k_{1z}))$
$E_{k0x}(E_{k0y}(k_{0z}))$

Suppose b'=0, b=1

This is the only row Bob can decrypt. He learns K_{0z}

7: Evaluate Entire Circuit

- In this way, Bob evaluates entire garbled circuit
  - For each wire in the circuit, Bob learns only one key
  - It corresponds to 0 or 1 (Bob does not know which)
    - Therefore, Bob does not learn intermediate values (why?)



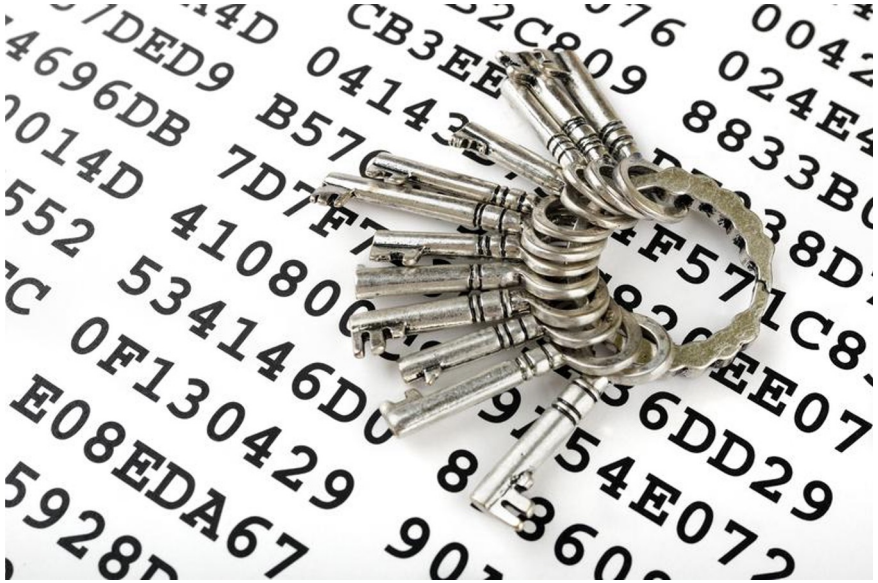
- Bob tells Alice the key for the final output wire and she tells him if it corresponds to 0 or 1
  - Bob does not tell her intermediate wire keys (why?)

Brief Discussion of Yao's Protocol

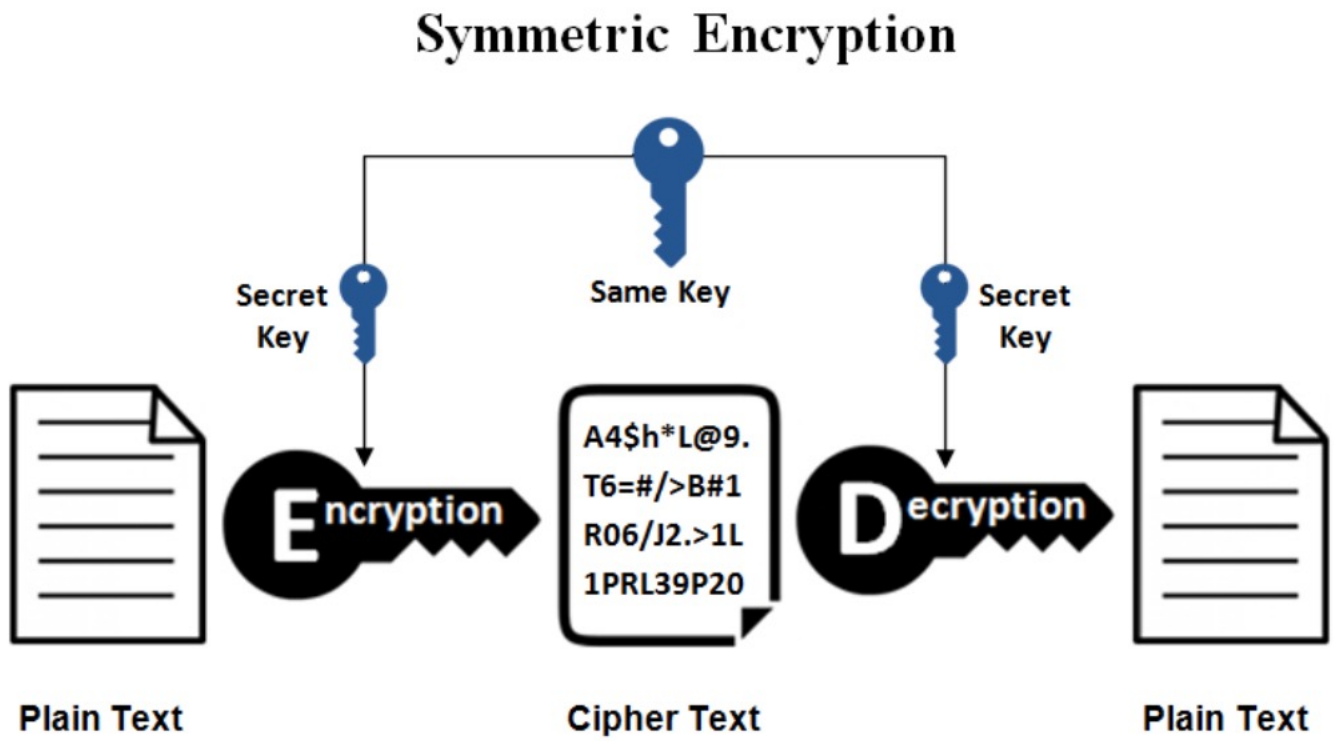
- Function must be converted into a circuit
  - For many functions, circuit will be huge
  - AES has around 30,000 gates
- If m gates in the circuit and n inputs, then need 4m encryptions and n oblivious transfers
  - Oblivious transfers for all inputs can be done in parallel
- Yao's construction gives a constant-round protocol for secure computation of any function in the semi-honest model
  - Two-round oblivious transfer protocol
  - Number of rounds does not depend on the number of inputs or the size of the circuit!

Deterministic (Symmetric) Encryption

- Always produces the same ciphertext for a given plaintext and key
  - It is efficient in searching of encrypted data
- Order-preserving encryption
  - M>N → E(M)>E(N)
  - Leaks information to an eavesdropper, who may recognize known ciphertexts
  - Does not guarantee what is known as “semantic security” -Need for **"probabilistic encryption"**

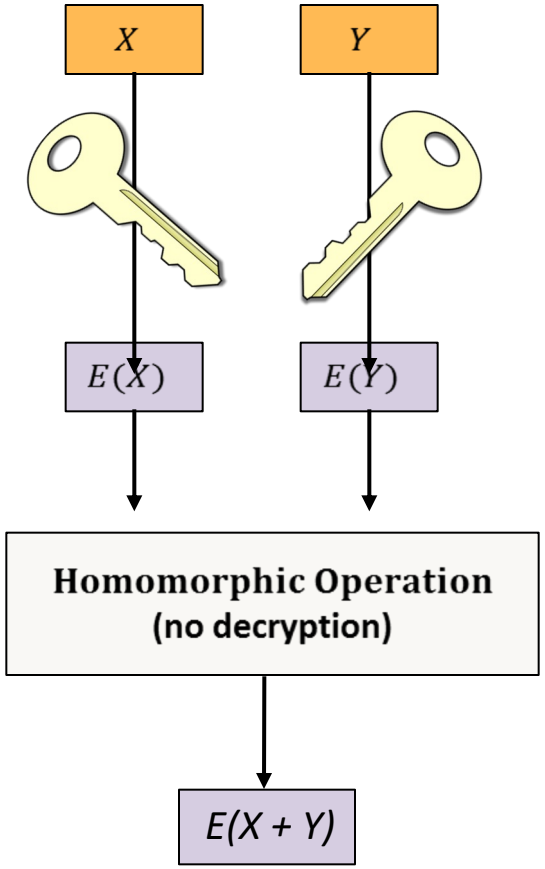


Symmetric Encryption

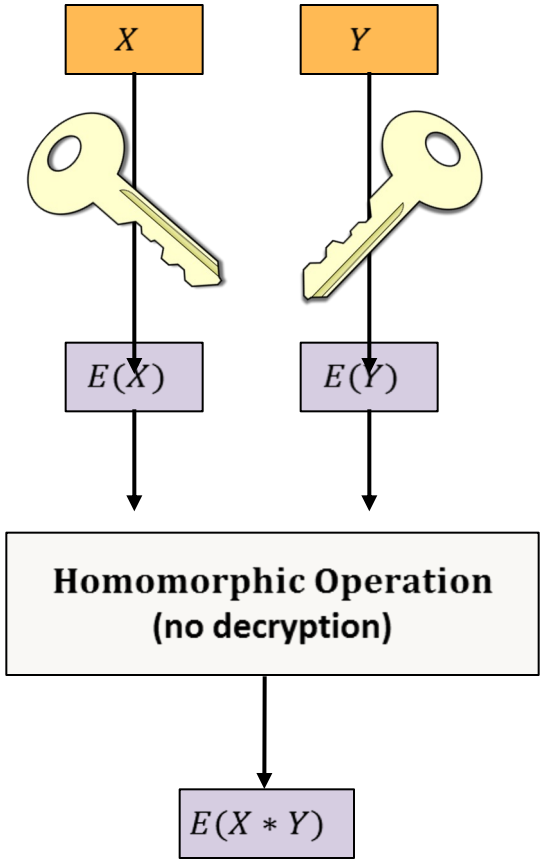


Homomorphic Encryption

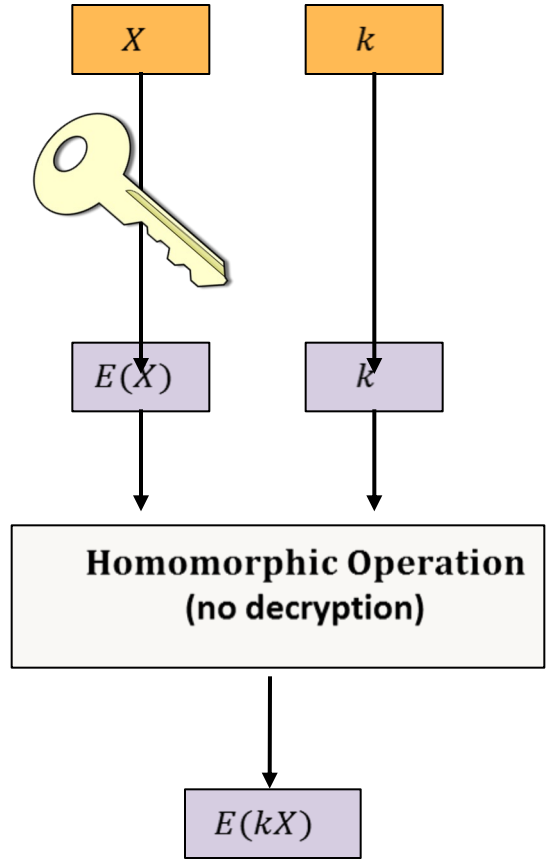
- Allows specific types of computations to be carried out on ciphertext



Benaloh, Paillier



Unpadded RSA, ElGamal



Paillier Cryptosystem

Key generation

Key generation works as follows:

1. Pick two large prime numbers p and q, randomly and independently. Confirm that $\gcd(pq, (p-1)(q-1))$ is 1. If not, start again.
2. Compute $n = pq$.
3. Define function $L(x) = \frac{x-1}{n}$.
4. Compute λ as $\operatorname{lcm}(p-1, q-1)$.
5. Pick a random integer g in the set $\mathbb{Z}^*_{n^2}$ (integers between 1 and n^2).
6. Calculate the modular multiplicative inverse $\mu = (L(g^\lambda \mod n^2))^{-1} \mod n$. If μ does not exist, start again from step 1.
7. The public key is (n, g). Use this for encryption.
8. The private key is λ. Use this for decryption.

https://blog.openmined.org/the-paillier-cryptosystem/

Paillier Cryptosystem

Encryption

Encryption can work for any m in the range $0 \leq m < n$:

1. Pick a random number r in the range $0 < r < n$.
2. Compute ciphertext $c = g^m \cdot r^n \mod n^2$.

Decryption

Decryption presupposes a ciphertext created by the above encryption process, so that c is in the range $0 < c < n^2$:

1. Compute the plaintext $m = L(c^\lambda \mod n^2) \cdot \mu \mod n$.

(Reminder: we can always recalculate μ from λ and the public key).

https://blog.openmined.org/the-paillier-cryptosystem/

Paillier Cryptosystem

Example

Here are some example values if you want to work through the algorithm:

Key generation

1. Pick $p = 13$ and $q = 17$. (They satisfy the condition.)
2. Compute $n = 221$.
3. Compute $\lambda = 48$.
4. Pick $g = 4886$.
5. Compute $\mu = 159$. (It exists.)

Encryption

1. Set $m_1 = 123$.
2. Pick $r_1 = 666$.
3. Compute $c_1 = 25889 \mod 221^2$.

Decryption

1. Compute $m_{decrypted} = 123 \mod 221$. (The same as m_1.)

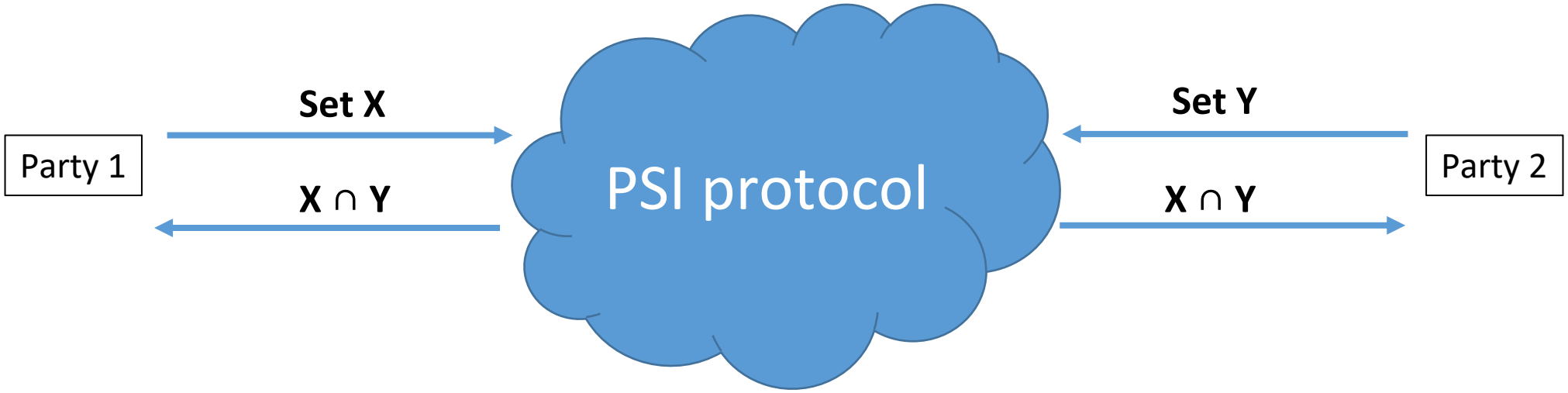
(But beware these numbers are too small to offer any real security and my random values weren't all that random.)

https://blog.openmined.org/the-paillier-cryptosystem/

Paillier Cryptosystem

Exercise!
Check homomorphic properties, e.g., addition of two encrypted numbers

Private Set Intersection (PSI)

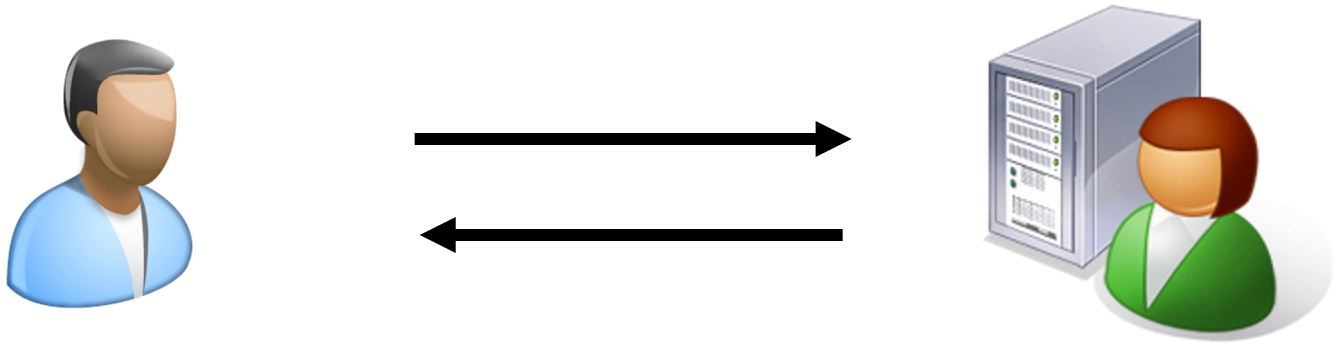


Protocol in which two parties jointly compute the intersection of their private input sets

There is no trusted third party!

Private Information Retrieval (PIR)

- Allows a user to retrieve an item from a server in possession of a database without revealing which item is retrieved



User U

wants to retrieve some data from D

Database D

shouldn't learn what U retrieved

Oblivious RAM

Same as PIR, but with R/W:

- Client outsources the storage of his data to a cloud
- Client stores only a small amount of data locally
- Client accesses (read/write) his data while hiding the identities of the items being accessed

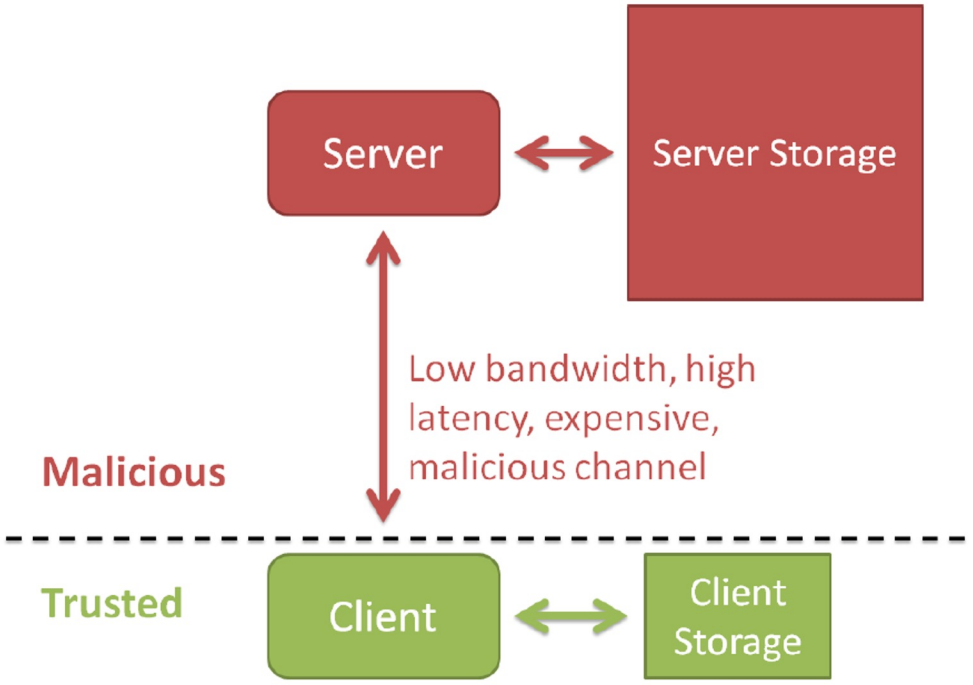


Figure: Emil Stefanov et al.

FUTURE OF PRIVACY

- Think of some modern observation tools

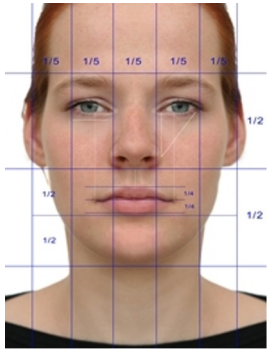
- Any issue with them? What are they?

Some Modern Observation Tools

Cellular phones

Web browsing

Online Social Networks



Face recognition

Genomic data

Challenges for Privacy

- Big databases
  - Data is never deleted, it is sometimes sold, and may be used in different contexts
  - Re-identification is becoming more and more feasible as the data is becoming detailed
- Data aggregation
  - Data is increasingly aggregated, collected and matched from multiple sources
  - Cross layer attacks
- Social networks

Privacy is rapidly becoming a collective value in that technology and market forces are making it hard for any one person to have privacy without all persons having a similar minimum level of privacy

- Priscilla Regan, 1995

- Location-based services

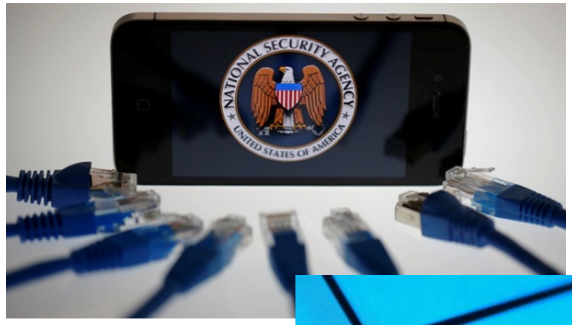
Surveillance

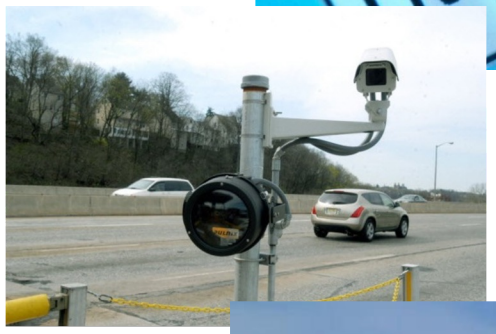
- The pervasiveness of computers has resulted in the almost constant surveillance of everyone
- Corporations and the police are both using this new trove of surveillance data
- Today's surveillance: *wholesale surveillance*
  - Previously: “follow that car”
  - Now: “follow every car”
- Tomorrow’s surveillance: even more extensive

https://www.schneier.com/blog/archives/2006/03/the_future_of_p.html

Wholesale Surveillance

- NSA can eavesdrop on most phone calls and read most e-mails
- Most of our browsing, search, and purchases are recorded
- EZ Pass can regularly record the location of our car
- Video cameras capture our images several times a day





Surveillance and Technology

- As computer memory becomes cheaper, more and more of these electronic footprints are being saved
  - 100 megabytes: to record everything the fastest typist input to his computer in a year
  - 4 to 8 gigabytes: to record everything the average user does on the Internet in a year
  - 5 gigabytes: to save the yearly phone calls of a typical person who uses 500 cell phone minutes a month
  - 200 gigabytes: to constantly record all audio of an individual per year
  - 700 gigabytes: to constantly record all video of an individual per year (life recorder)
- As processing becomes cheaper, more and more of it is being cross-indexed and correlated
- There are companies in the business of buying and reselling customer databases

Drones are Coming

- The Federal Aviation Administration (US) is relaxing its restrictions around the domestic use of “unmanned aerial systems,” leading to greater use of drones by public agencies and the private sector

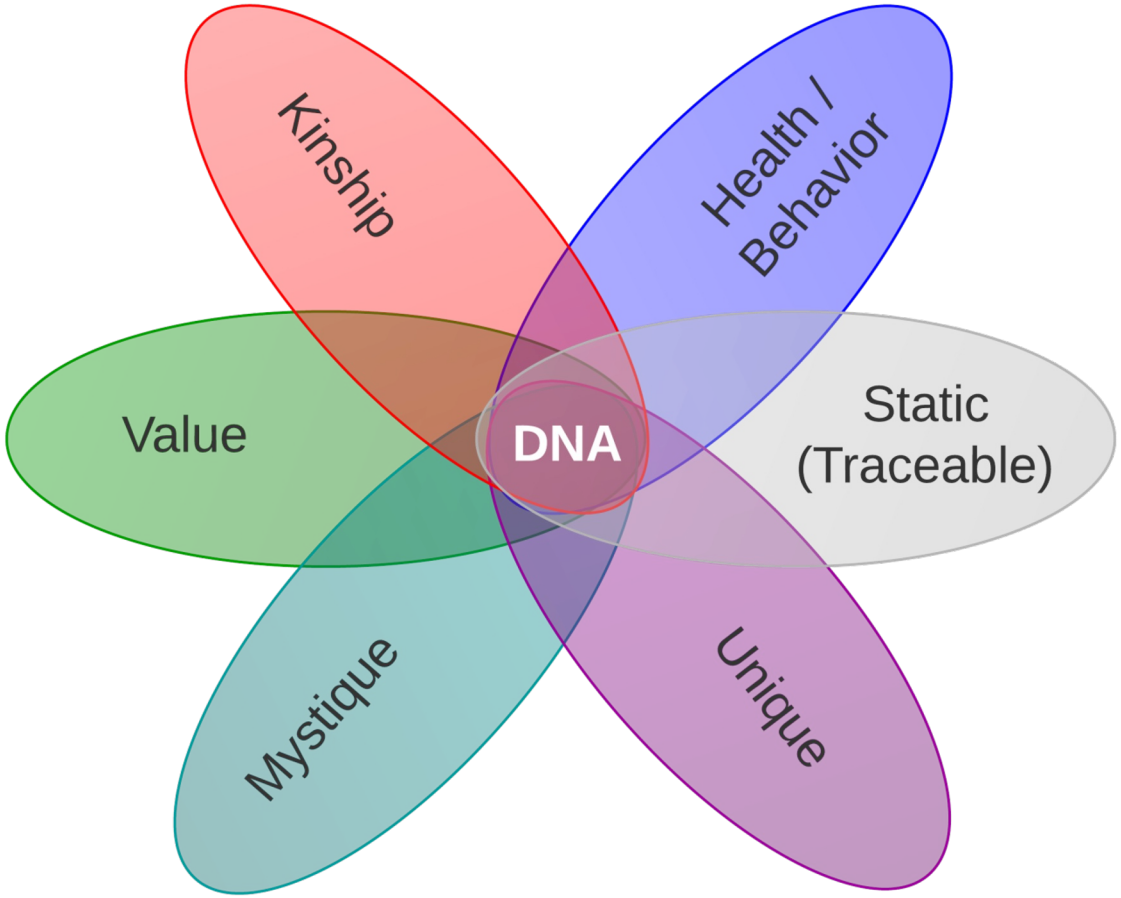
Drones are Coming

- Congress authorized the FAA to open the nation's airspace to widespread drone flights by 2015
- The FAA estimates that more than 7,000 civilian drones could be surfing the sky by 2020

New Era of Surveillance

- (Online mass surveillance: PRISM)
  - Trivial solution: do not go online?
- Supreme Court rejects GPS tracking
  - Police need a warrant before affixing a GPS device to a car and following a suspect for a prolonged period (United States v. Jones)
  - The FBI reportedly turned off thousands of GPS devices in response to the ruling
- But,
  - Drones can follow a car without the need to attach anything
  - Law enforcement can equip a drone with thermal or chemical sensors and let it loose to roam a neighborhood in search of invisible infractions such as indoor marijuana cultivation

Another New Era: Genomic Privacy

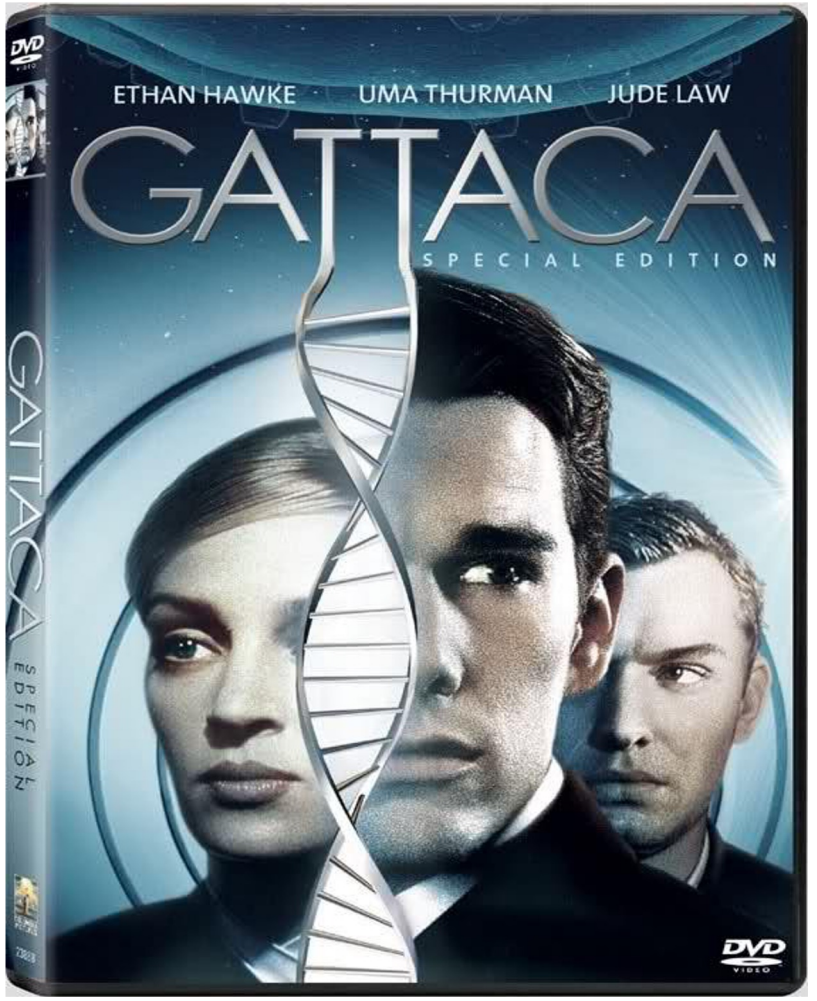


Why Protect Genomic Data?

- Genome carries information about a person's genetic condition and predispositions to specific diseases
  - Leakage of such information could cause *genetic discrimination*
  - Denial of access to health insurance, mortgage, education, and employment

Why Protect Genomic Data?

- Genome carries information about a person's genetic condition and predispositions to specific diseases
  - Leakage of such information could cause *genetic discrimination*
  - Denial of access to health insurance, mortgage, education, and employment



Why Protect Genomic Data?

- Genome carries information about a person's genetic condition and predispositions to specific diseases
  - Leakage of such information could cause *genetic discrimination*
  - Denial of access to health insurance, mortgage, education, and employment
- Anonymisation is ineffective

Identifying Personal Genomes by Surname Inference

Melissa Gymrek,[1,2,3,4] Amy L. McGuire,[5] David Golan,[6] Eran Halperin,[7,8,9] Yaniv Erlich[1]*

Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state, can be used to triangulate the identity of the target. A key feature of this technique is that it entirely relies on free, publicly accessible Internet resources. We quantitatively analyze the probability of identification for U.S. males. We further demonstrate the feasibility of this technique by tracing back with high probability the identities of multiple participants in public sequencing projects.

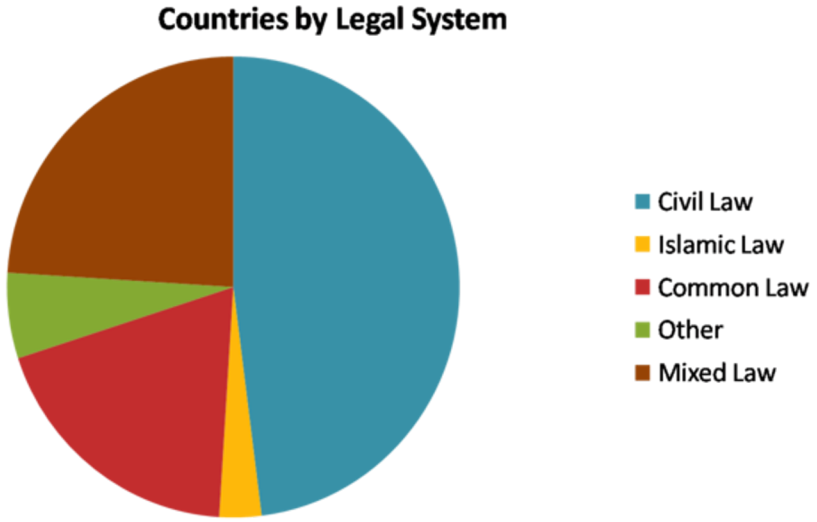
M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich. *Identifying personal genomes by surname inference*. Science: 339 (6117), Jan. 2013.

Why Protect Genomic Data?

- Genome carries information about a person's genetic condition and predispositions to specific diseases
  - Leakage of such information could cause *genetic discrimination*
  - Denial of access to health insurance, mortgage, education, and employment
- Anonymisation is ineffective
- Genome carries information about family members
  - Cross-layer attacks
    - Using privacy-sensitive information belonging to a victim retrieved from different sources

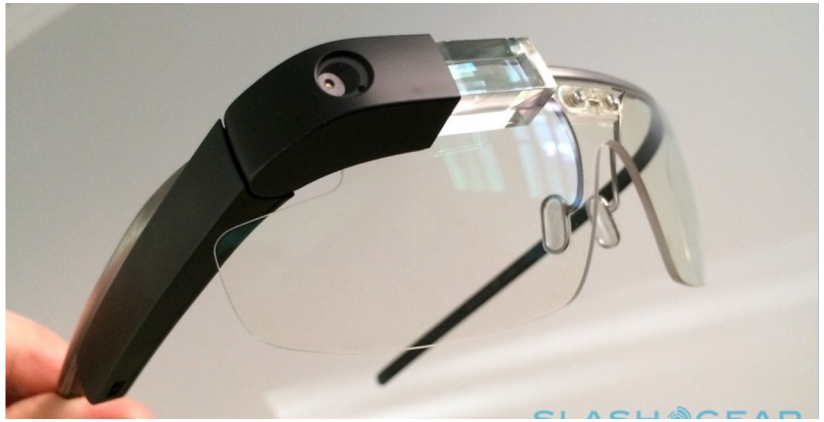
Why Protect Genomic Data?

- Genome carries information about a person's genetic condition and predispositions to specific diseases
  - Leakage of such information could cause *genetic discrimination*
  - Denial of access to health insurance, mortgage, education, and employment
- Anonymisation is ineffective
- Genome carries information about family members
  - Cross-layer attacks
    - Using privacy-sensitive information belonging to a victim retrieved from different sources
- Genomic data is non-revocable
- Law is not universal; it is hard to enforce



Other Concerns

- A recent poll, conducted by market-research firm Toluna, found 72% of Americans cited privacy concerns as the biggest reason for not wanting to wear the Glass
- Microsoft had to fully detail what the Xbox One Kinect sensor sees and sends
  - A new privacy statement page gets into what Kinect data is collected and how it will be used



http://mashable.com/2014/04/07/google-glass-privacy/

Computer Scientists Must Save the World [1]

- **Policy** is limited by that which can be expressed in words
- **IT** can provide glue technology, but heavily relies on existing technology
- Laws can change and **lawyers** often lack understanding of ways technology will continue to change
- **Computer scientists** construct tomorrow's machines, and can do so with privacy as part of their problem definition, so that new technology can be deployed and easily adopted

[1] Latanya Sweeney