ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

Information Security and Privacy (COM-402)

Introduction to Privacy

Carmela Troncoso
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

We will be using SpeakUp

- Channel 209902
- http://web.speakup.info/ng/room/5c828712a377bdbd5e9bf025

We will be using SpeakUp

- Channel 29902
- http://web.speakup.info/ng/room/5c828712a377bdbd5e9bf025

And sometimes also speaking up 😱

Warming up!

Privacy is about

A. Hiding information

B. Controlling where information goes

C. Controlling how the information is used

D. Privacy is not important, I have nothing to hide

Warming up!

Privacy in the DIGITAL world is important

A. For individuals

B. Beyond individuals (society)

C. I told you, stop asking: I have nothing to hide

D. Privacy is over (Facebook, Google, etc) why even bother…

Warming up!

Encryption is enough to solve privacy problems

A. Yes

B. No

Goals of the next three lectures

- 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

- 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 each adversary model
  - some of them with Prof. Hubaux
  - more in CS-523!

- Privacy requires protecting information beyond content: The need to protect meta-data
  - Inference attacks based on meta-data

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

Information Security and Privacy (COM-402)

Part 1: Why we need Privacy

Carmela Troncoso
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

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

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

Individual applications are legitimate

Together they become a cheap
SURVEILLANCE INFRASTRUCTURE

The context: Availability of data

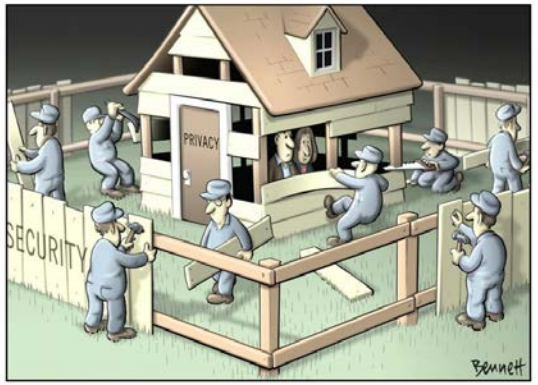
Intelligent data-based applications

- 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

Individual applications are legitimate

Together they become a cheap
SURVEILLANCE INFRASTRUCTURE

We need privacy!



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

What do you think?

Are privacy and security contradictory?

A. Yes

B. No

Common belief:
we need to tradeoff security for privacy!

"For National Security surveillance is good and privacy is bad"

(Surveillance == Security) == True ??

Common belief:
we need to tradeoff security for privacy!

"For National Security surveillance is good and privacy is bad"

(Surveillance == Security) == True ??

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

- criminals use Telegram, Threema, Signal,...
- … but we do not!!

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

- Snowden revelations: NSA spying on Americans, companies, ...
- Spanish Interior ministry spying independentist politicians

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...

And any system can become a surveillance system (aka the risk of function creep)

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

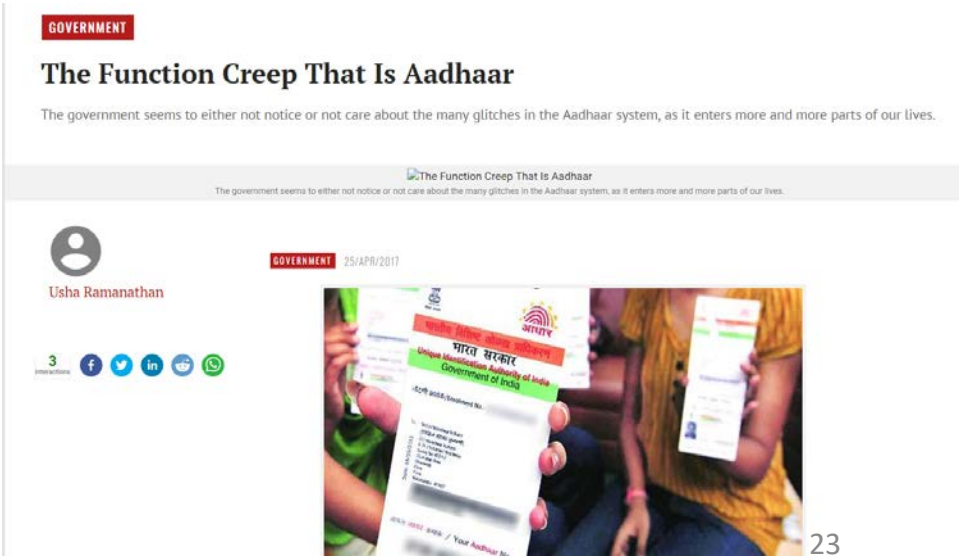
"We will create a new system to improve X"

"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*”



https://privacyinternational.org/feature/2299/initial-analysis-indian-supreme-court-decision-aadhaar
https://thewire.in/government/aadhaar-function-creep-uid

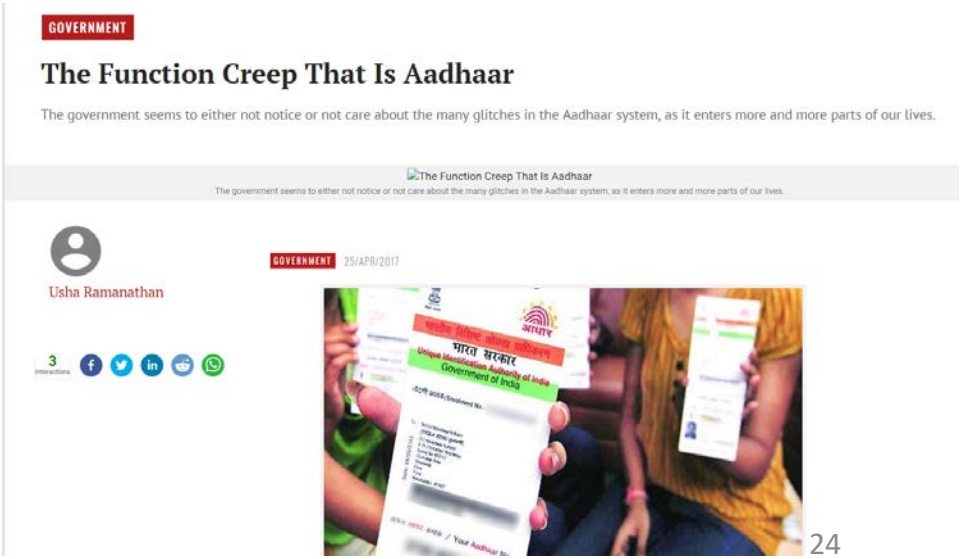
A recurrent function creep example: identity systems

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/feature/2299/initial-analysis-indian-supreme-court-decision-aadhaar
https://thewire.in/government/aadhaar-function-creep-uid

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.

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

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. Fingerprints from 6 years old.


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

Privacy **IS** a security property

For individuals

protection against crime / identity theft, control over one's information, protection against profiling and manipulation.

For companies

protection of trade secrets, business strategy, internal operations, access to patents

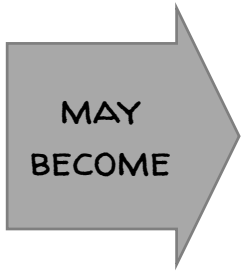
For governments / military

protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations

Privacy **IS** a security property

For individuals

protection against crime / identity theft, control over one's information, protection against profiling and manipulation.



ONE RING TO RULE THEM ALL

For companies

protection of trade secrets, business strategy, internal operations, access to patents

For governments / military

protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations

Privacy **IS** a security property

For individuals

protection against crime / identity theft, control over one's information, protection against profiling and manipulation.

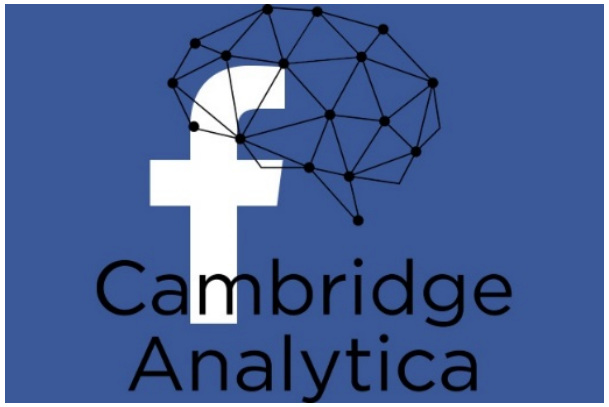
ONE RING TO RULE THEM ALL

For companies

protection of trade secrets, business strategy, internal operations, access to patents

For governments / military

protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations



100K users installed CA Facebook App

enabled **COLLECTING PERSONAL DATA** of 87+ million

public profile, page likes, birthday and current city

creation of **PROFILES** of the subjects of the data

TARGETED ADVERTISEMENTS influenced the US elections

Privacy **IS** a security property

INFRASTRUCTURE IS SHARED

Individuals, Industry, and Governments use the same applications!

Directly
(Cloud-based services, Industry 4.0, Blockchain)

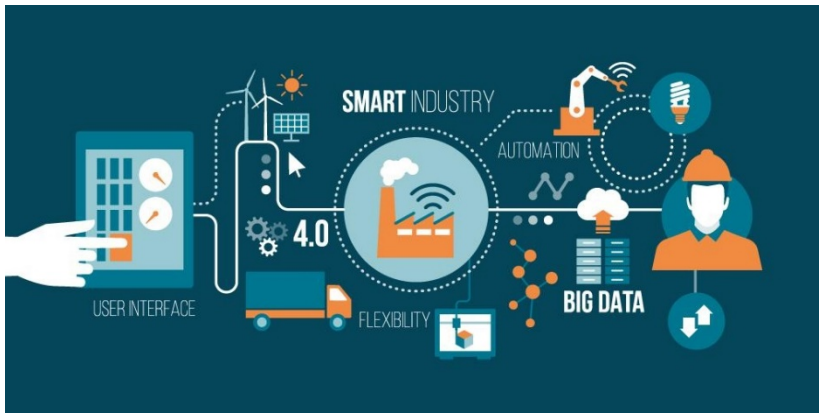
Indirectly
(employers are users)

Privacy **IS** a security property

INFRASTRUCTURE IS SHA

Individuals, Industry, and Governments use the same applications!

Denying privacy to some is denying privacy to all!!



Directly
(Cloud-based services, Industry 4.0, Blockchain)

Indirectly
(employers are users)

Privacy **IS** a security property

For individuals

protection against crime / identity theft, control over one's information, protection against profiling and manipulation.

For companies

protection of trade secrets, business strategy, internal operations, access to patents

ONE RING TO RULE THEM ALL

For governments / military

protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations

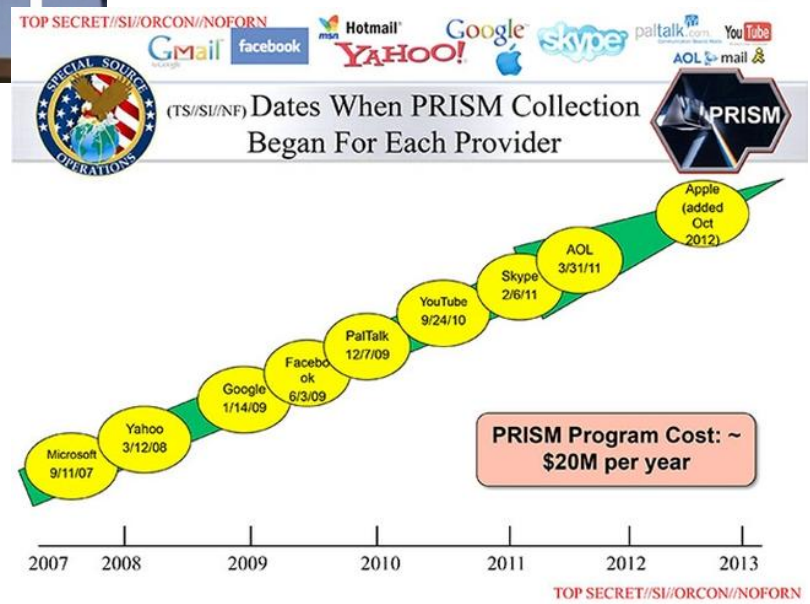
For companies

For governments

(C) Customers Can Help SID Obtain Targetable Phone Numbers

FROM: [redacted] and [redacted]
A&P Staff's Access Interface Portfolio (S203A)
Run Date: 10/27/2006

(C) From time to time, SID is offered access to the personal contact databases of US officials. Such "rolodexes" may contain contact information for foreign political or military leaders, to include direct line, fax, residence and cellular numbers.



and 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."

and 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."

MAY
BECOME

ONE RING TO RULE THEM ALL

Takeaways

Digital identities are very powerful

but enable cheap “surveillance”

Privacy is of course is about sensitive individuals’ information

but also needed for safeguard societal and democratic values

Privacy **IS** a security property

the need for a tradeoff is a fallacy

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

Information Security and Privacy (COM-402)
Part 2: Technical approaches to privacy

Carmela Troncoso
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

What is privacy in Privacy Enhancing Technologies

PETs

Not these PETs!!!

Two dimensions

1) **Privacy paradigms**: how privacy is defined

2) **Adversarial models**: who defines the problem and what are the goals

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.

What is privacy: privacy paradigms

Privacy as
CONFIDENTIALITY

Privacy as
CONTROL

Privacy as
PRACTICE

FaKebook
Social Network

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

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

What would you do?

FaKebook
Social Network

Discuss with your neighbor(s) and propose one PET

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

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

Ts 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:
General Data Protection Regulation (GDPR)
Fair Information Practice Principles (FIPPs)

Privacy as control

What would you do?

FaKebook
Social Network

Discuss with your neighbor(s) and propose one PET

> "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)

Ts 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:**
   General Data Protection Regulation (GDPR)
   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

Privacy as practice

What would you do?

FaKebook
Social Network

Discuss with your neighbor(s) and propose one PET

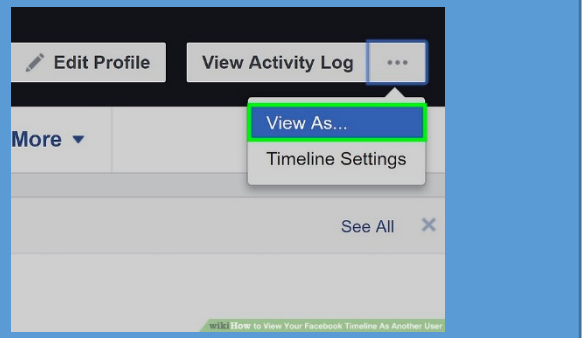
> "the freedom from unreasonable constraints on the construction of one's own identity"
> Agre (1999)

Ts 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

This classification is not better or worse that the 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 the adversary**)

what are the privacy goals (**what should be protected and how**)

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.

Refresher: Threat model, threat, harms, and vulnerabilities

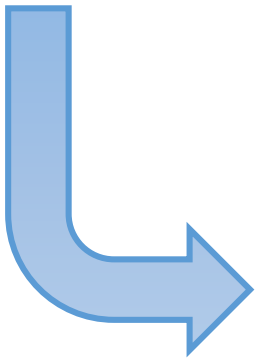
THREAT MODEL

The adversary's capabilities. Very technical term!!

The adversary can observe my connection

The adversary can corrupt my machine

The adversary controls a Facebook employee

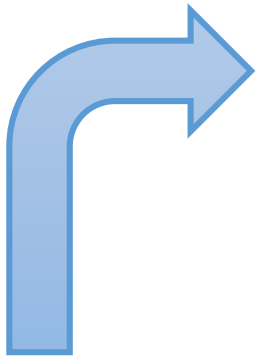


THREAT

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

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

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

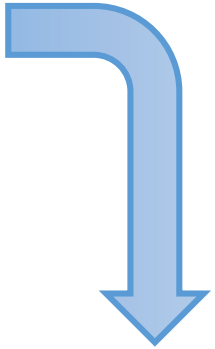


VULNERABILITY

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

My Google queries are not encrypted

My Facebook settings are very open



HARM

The bad thing that happens when the threat materializes

I receive targeted advertising that aims at influence my vote

A thief enters in my place while I am at a conference in the United States

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 other job”

“My friends saw my naked pictures”

1 – PETs for “social” Privacy

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

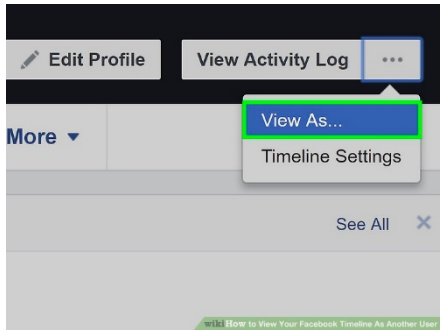
Technology brings problems

- “My parents discovered I'm gay”
- “My boss knows I am looking for other job”
- “My friends saw my naked pictures”

GOALS - Do not surprise the user

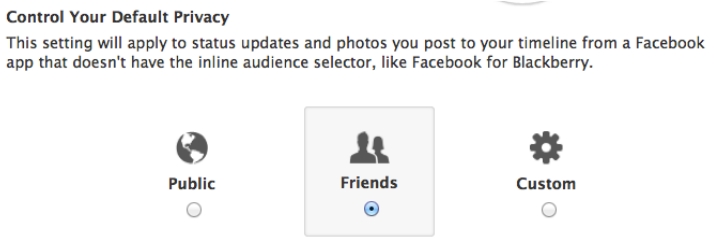
Two main approaches

- Support decision making
- Help identifying actions impact



Contextual feedback

Privacy nudges

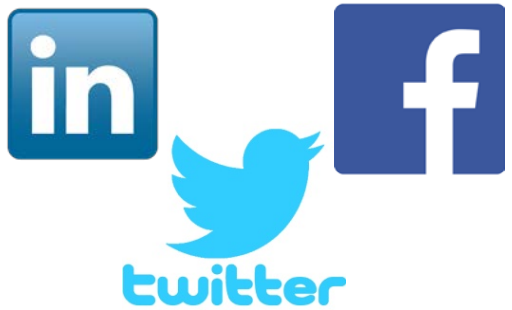


Easy defaults

1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!



Common Industry approach
Make users comfortable

1 – PETs for “social” Privacy

LIMITATIONS

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

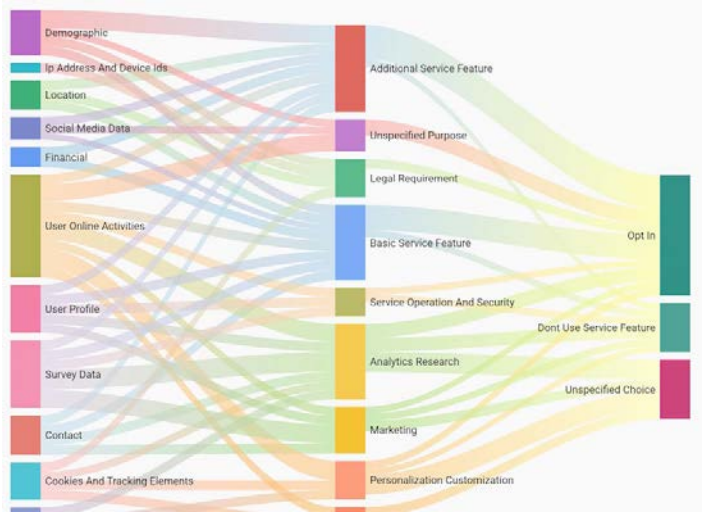
Common Industry approach
Make users comfortable

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

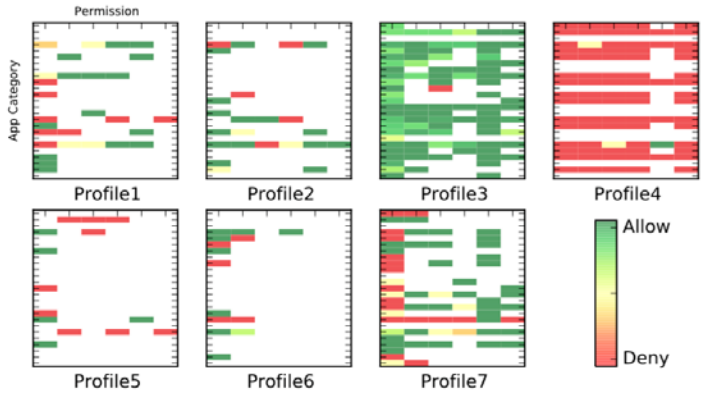


Figure 2: Privacy profiles learned from collected app privacy settings. Profile 1 is more protective on Location and Productivity apps than other profiles. Profile 2 denies phone call log permission more. Profile 3 is generally permissive. Profile 4 denies most permission requests. Profile 5 generally 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.

Common Industry approach
Make users comfortable

Automated configuration

A. **Is good for any user**
B. **Only works for average users**
C. **Only works for average apps**
D. **Only works for outliers**
E. **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?

Common Industry approach
Make users comfortable

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

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.

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

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

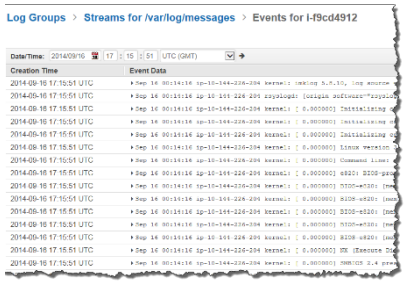
Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent
purpose limitation
data minimization
subject access rights

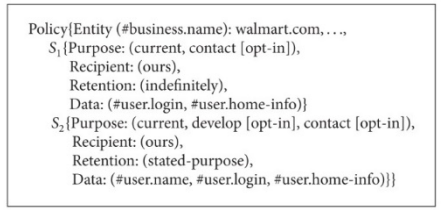
Preserving the security of data
Auditability and accountability

Access control



Logging

Anonymization????



Automated Policy Negotiation

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

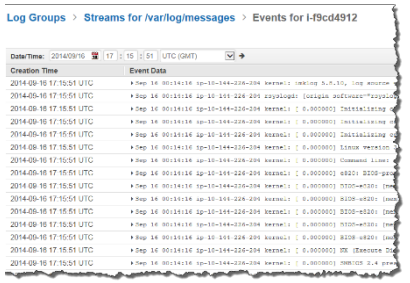
ty of data

untability

Personal data

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

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

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 technical protection of the data

2 – PETs for “institutional” Privacy

LIMITATIONS

Assumes:

- collection and processing by organizations is necessary
- organizations are (semi)-trusted and honest
  - Reliance on punishment
  - No technical 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!

2 – PETs for “institutional” Privacy

LIMITATIONS

Assumes:

- collection and processing by organizations is necessary
- organizations are (semi)-trusted and honest
  - Reliance on punishment
  - No technical 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!

Limited

- Scope (personal data != all data)
- Transparency (proprietary software and algorithms)

2 – PETs for “institutional” Privacy

LIMITATIONS

Assumes:
- collection and processing by organizations is necessary
- organizations are (semi)-trusted and honest
  - Reliance on punishment
  - No technical 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!

Limited
- Scope (personal data != all data)
- Transparency (proprietary software and algorithms)

Common Industry approach
Make users comfortable
+ Legal compliance!!

But does not prevent

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

3 – PETs for “anti-surveillance” Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

End-to-End encryption: Signal, PGP, OTR

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

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

End-to-End encryption: Signal, PGP, OTR

Anonymous comms: Tor, mixnets

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

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

COM-402 and CS-523

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

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

Difficult to evolve – do not deal well with the Agile paradigm

Also difficult to combine

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

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

- Difficult to evolve – do not deal well with the Agile paradigm
- Also difficult to combine

Usability problems both for developers and users

- how the @$%&#$Ŷ& do I program this?
- performance hit
- unintuitive technologies

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

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

Difficult to evolve – do not deal well with the Agile paradigm

Also difficult to combine

Usability problems both for developers and users

how the @$%&#$Ŷ& do I program this?

performance hit

unintuitive technologies

Lack of incentives:

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

Takeaways

One can think about privacy technologies depending on:

The privacy conception: confidentiality, control, practice

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

Each type of PETs present different challenges

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 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 bars see the student recommendations so that they can offer discounts to students that give good ratings.

Adversaries: other students, owners, cloud provider
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
They can see: students and cloud owners cannot learn students behavior anymore;
The Cloud provider still can see everything and learn student behavior and preferences

Let’s exercise your privacy brain

From the second classification, 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 what 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 what 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

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

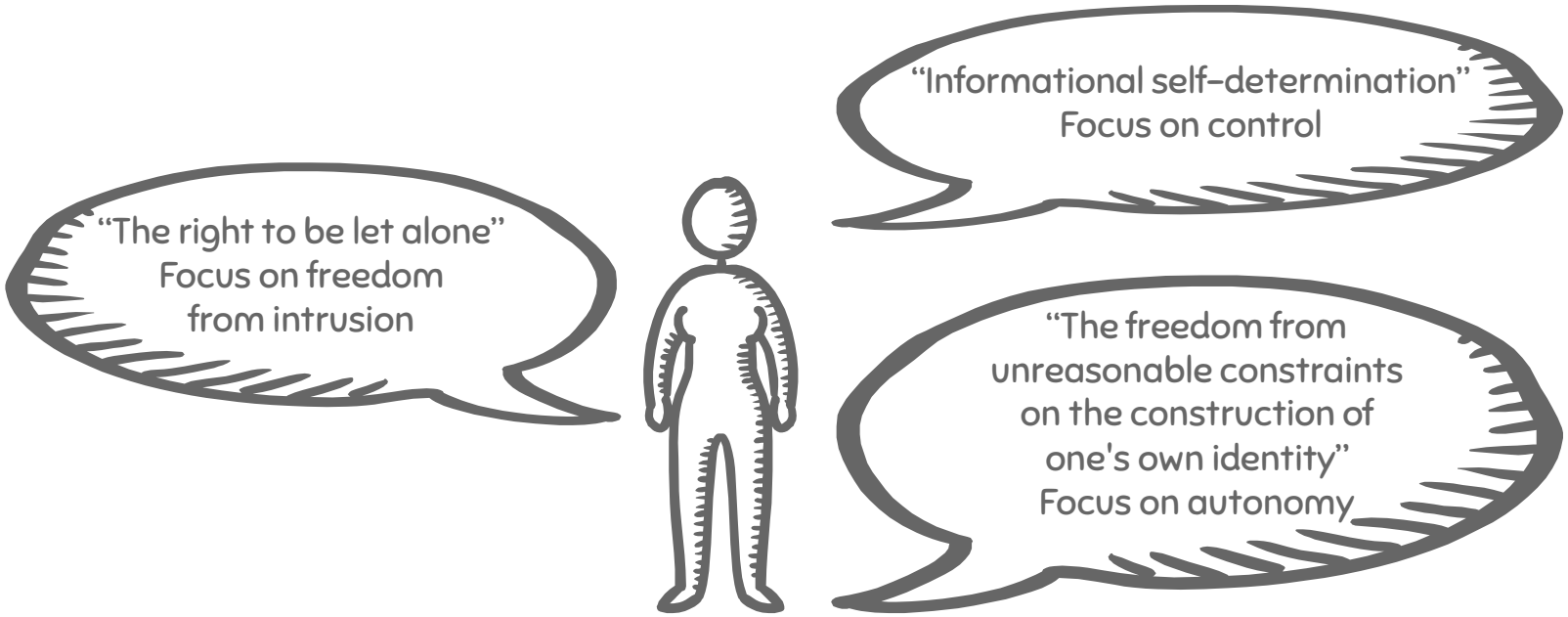
Information Security and Privacy (COM-402)

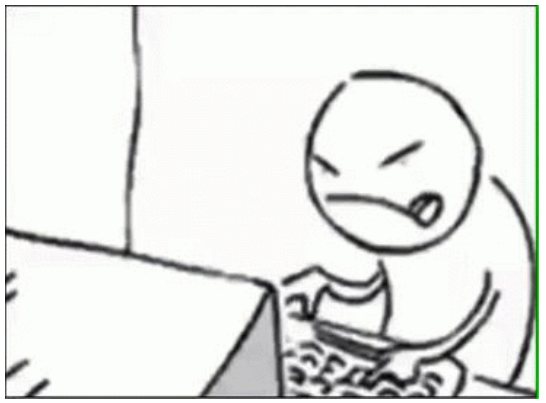
Part 3: Privacy properties and privacy evaluation

Carmela Troncoso
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

Anti-surveillance PETs – let's work on them!





We can't design / program / evaluate these!!!?!?!?!?

Anti-surveillance PETs – technical goals

Privacy properties

Confidentiality (in transit/storage, during processing)

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Anti-surveillance PETs – technical goals

Privacy properties

Confidentiality (in transit/storage, during processing)

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Game-oriented definitions
cryptographic definitions

and CS-523

and COM-501

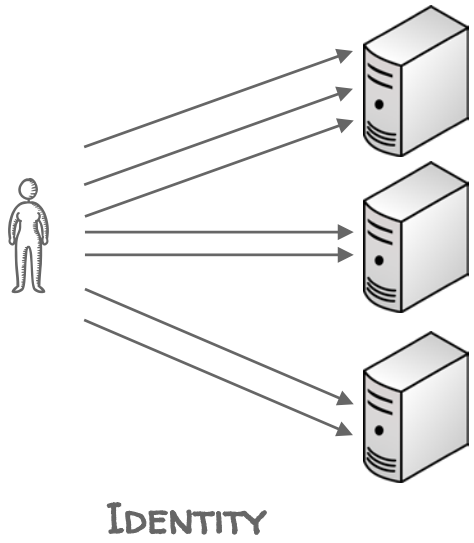
COM-402

Anti-surveillance PETs – technical goals

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.”

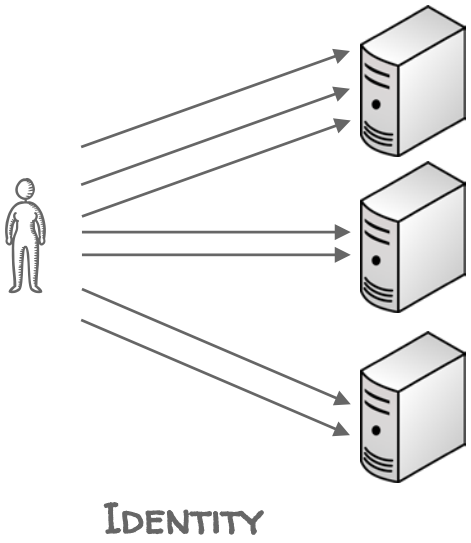


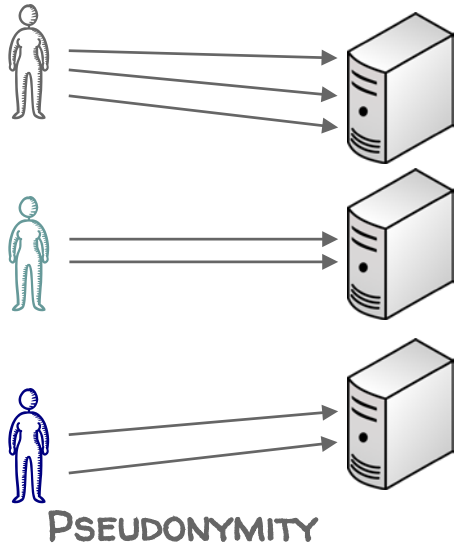
Anti-surveillance PETs – technical goals

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."





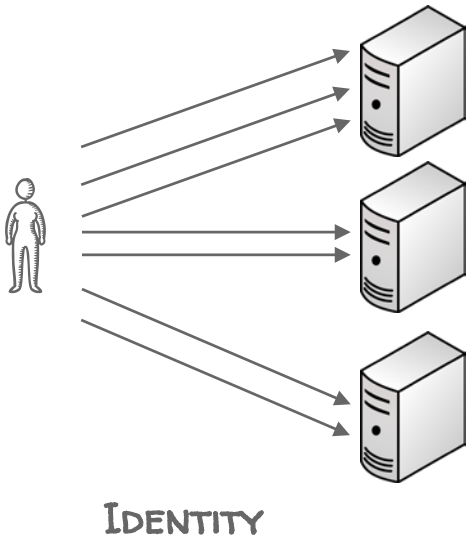
Pfitzmann, Andreas and Hansen, Marit. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

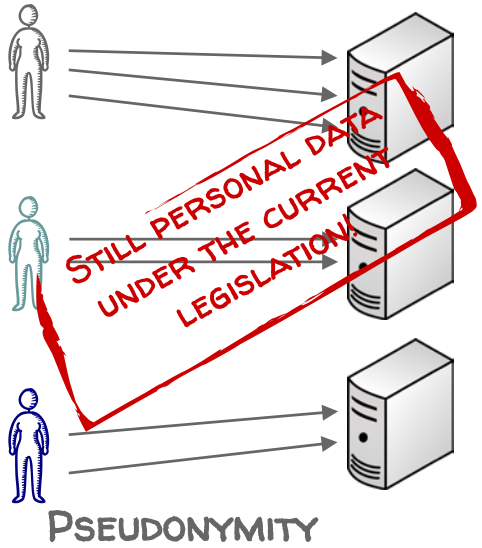
Anti-surveillance PETs – technical goals

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."





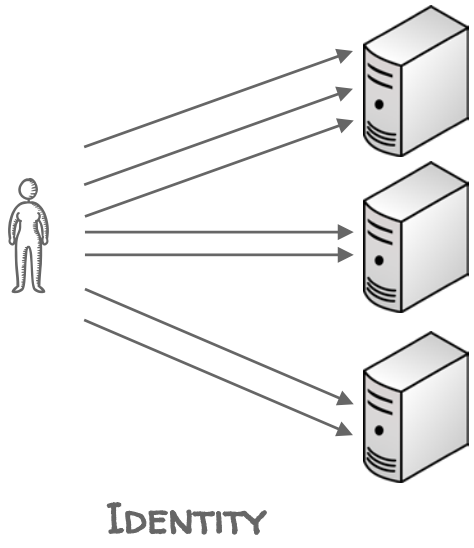
Pfitzmann, Andreas and Hansen, Marit. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

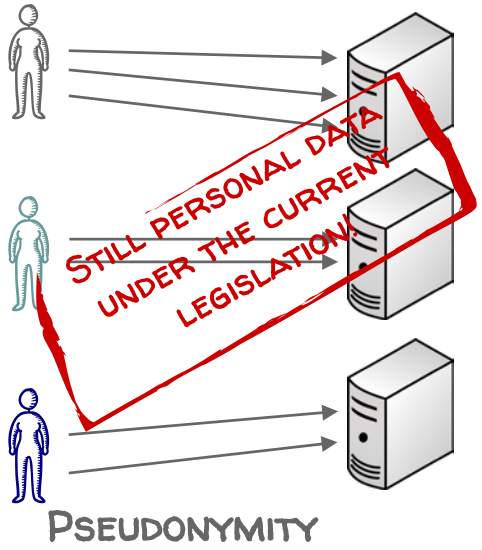
Anti-surveillance PETs – technical goals

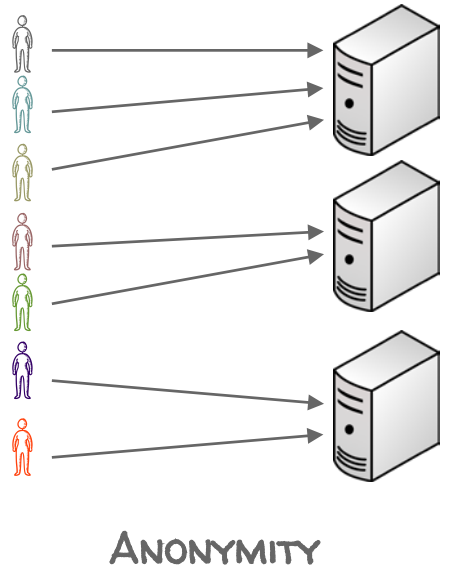
Privacy properties: **Pseudonymity**

-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.

Example of pseudonymity technologies

- Use of persistent random identifiers
- Use of hashed identifiers
- Emails
- Nicknames

These technologies are very limited from a privacy perspective, more later

- Pseudo-identifiers, Quasi-identifiers

Anti-surveillance PETs – technical goals

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!

Anti-surveillance PETs – technical goals

Privacy properties: **Unlinkability**

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

– **ISO15408**: " 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
… two people related by a friendship link
… same person spotted in two locations

DECOUPLING TWO ACTIONS FROM ONE USER!

Anti-surveillance PETs – technical goals

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!

Anti-surveillance PETs – technical goals

Privacy properties: **Plausible Deniability**

– Not possible to prove user knows, has done or has said something
  – Resistance to coercion: one can always claim one does not know
  – 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!

Systematic Privacy Evaluation

Confidentiality (in transit/storage, during processing)

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Systematic Privacy Evaluation

Confidentiality (in transit/storage, during processing) — **Cryptographic proofs**

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Systematic Privacy Evaluation

Confidentiality (in transit/storage, during processing)

Cryptographic proofs

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Probabilistic analysis

Anonymity - Pr[identity → action | observation]

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

Unobservability - Pr[real action |observation]

Plausible deniability - Pr[fake| observation]

Systematic Privacy Evaluation

Confidentiality (in transit/storage, during processing)

Cryptographic proofs

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

Probabilistic analysis

Anonymity - Pr[identity → action | observation]

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

Unobservability - Pr[real action |observation]

Plausible deniability - Pr[fake| observation]

Many 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

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 is the “observation”? what is her prior knowledge?

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 is the "observation"? what is her prior knowledge?

3) "Invert" the mechanism as the adversary would do

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

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 is the "observation"? what is her prior knowledge?

3) "Invert" the mechanism as 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

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 is the "observation"? what is her prior knowledge?

3) "Invert" the mechanism as 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

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]

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 is the "observation"? what is her prior knowledge?

3) "Invert" the mechanism as 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) Measure

   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 the observation what is the bound on the information the adversary can gain

Next block!

Takeaways

Privacy is an abstract concept, need technical definitions to build technologies

We formalize privacy as privacy properties

- You may need more than one property to capture your privacy objectives!

Privacy evaluation

- the adversary knows
- the adversary makes probabilistic inferences
- metrics consider probabilities as seen from the adversary

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
C. A user uses her university account for accessing university services and her Gmail account to access Netflix.
D. A user accesses a newspaper website using a proxy, the newspaper can only see the proxy IP address.

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
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
Unobservability, Plausible deniability

C. A user uses her university account for accessing university services and her Gmail account to access Netflix.
Pseudonimity, Unlinkability (across pseudonyms, not across all actions)

D. A user accesses a newspaper website using a proxy, the newspaper can only see the proxy IP address.
If she is the only user behind this proxy: Pseudonymity
If there are several users: Anonymity, Unlinkability, Plausible deniability

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

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 study with more than 0.5 probability. The students have plausible deniability about what they wrote.

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

Information Security and Privacy (COM-402)

Part 4: Privacy enhancing technologies

Carmela Troncoso
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

PETs for Data anonymization

Scenario:

You have a set of data that contains personal data and you would like to anonymize it to:

- not be subject to data protection while processing
- make it public for profit
- make it public for researchers

Goal:

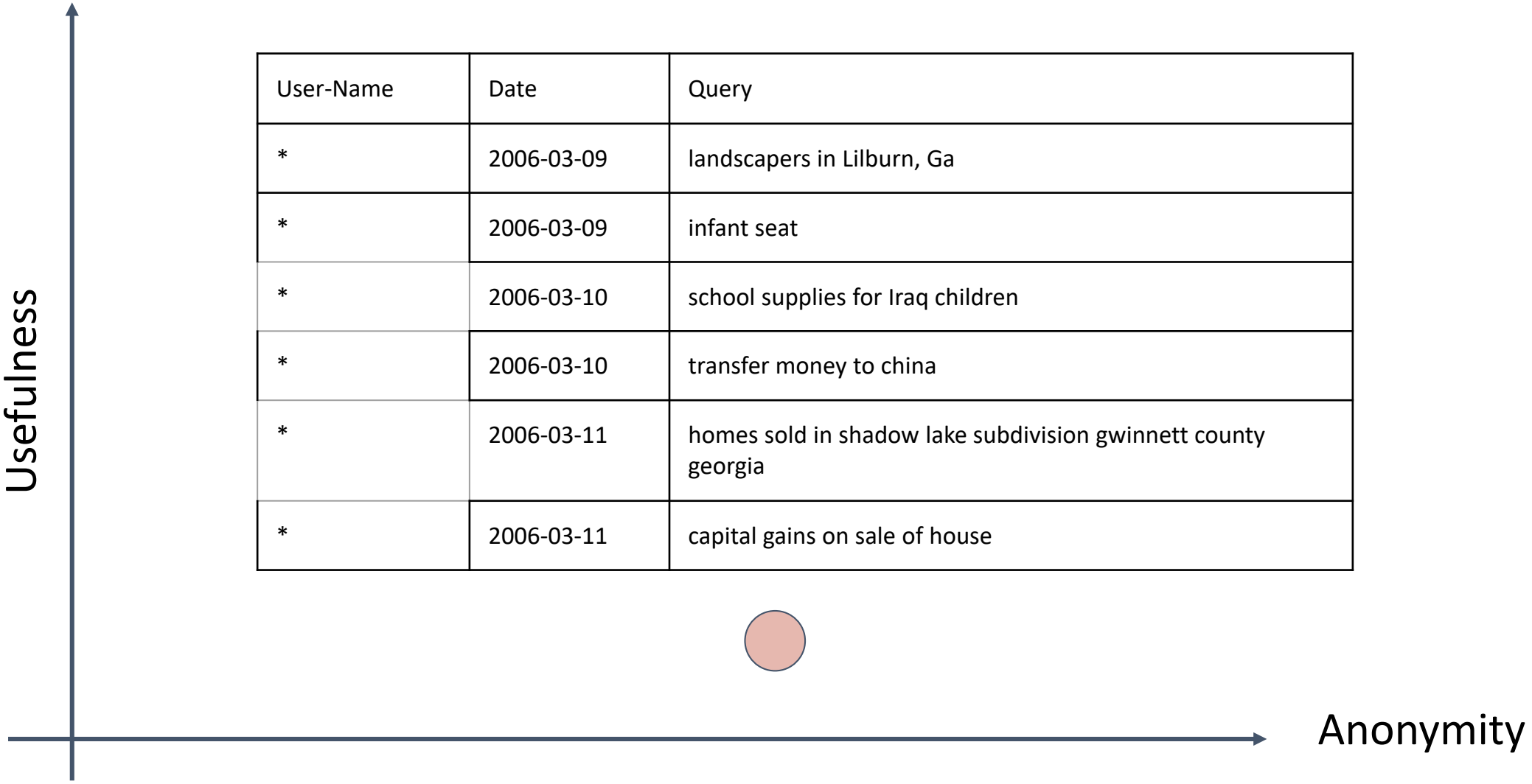
Produce a dataset that **preserves the utility** of the original dataset **without leaking information** about individuals

Part of Original AOL Dataset

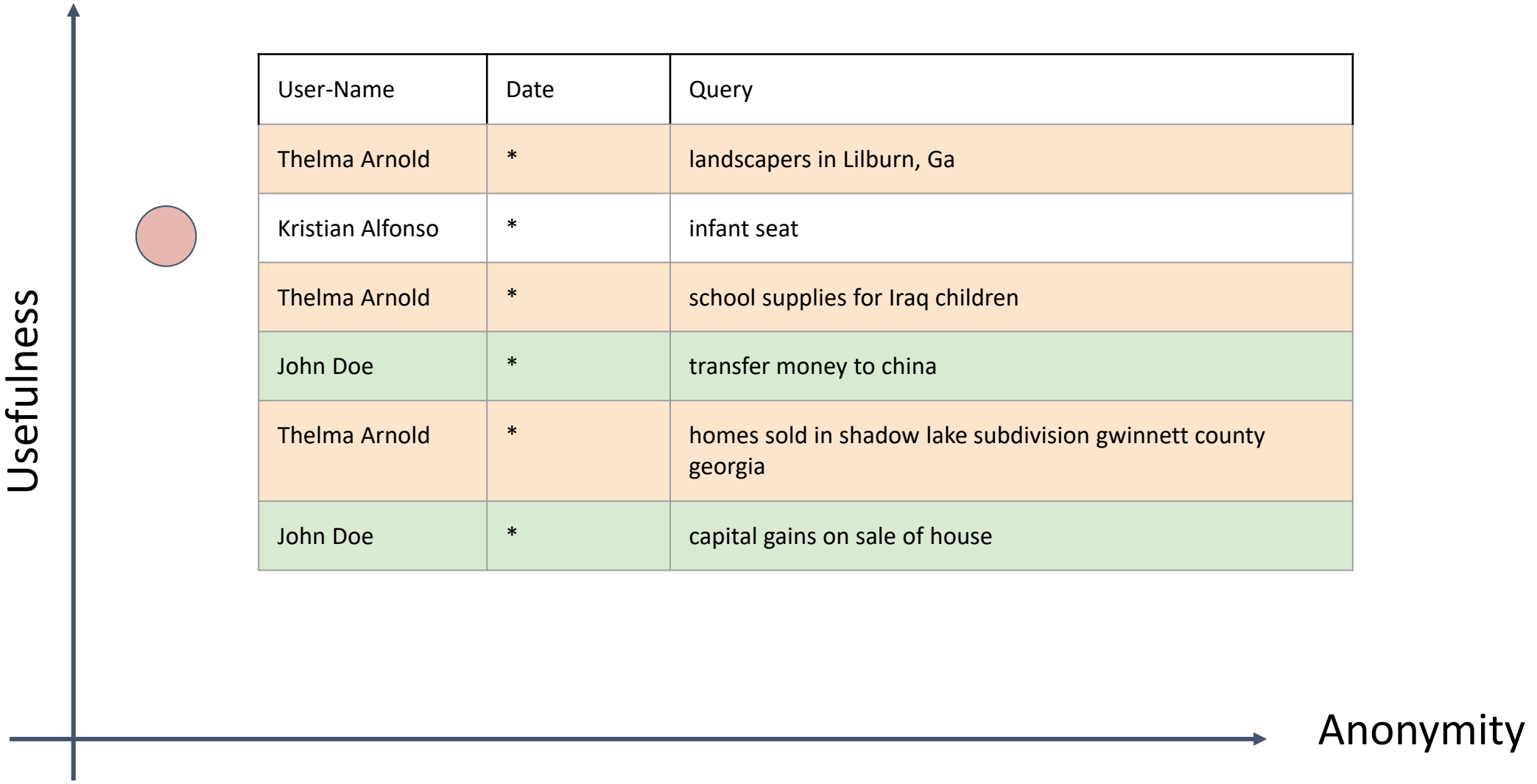
User-ID	Date	Query
Thelma Arnold	2006-03-09	landscapers in Lilburn, Ga
Kristian Alfonso	2006-03-09	infant seat
Thelma Arnold	2006-03-10	school supplies for Iraq children
John Doe	2006-03-10	transfer money to china
Thelma Arnold	2006-03-11	homes sold in shadow lake subdivision gwinnett county georgia
John Doe	2006-03-11	capital gains on sale of house

Remove User-names

User-Name	Date	Query
*	2006-03-09	landscapers in Lilburn, Ga
*	2006-03-09	infant seat
*	2006-03-10	school supplies for Iraq children
*	2006-03-10	transfer money to china
*	2006-03-11	homes sold in shadow lake subdivision gwinnett county georgia
*	2006-03-11	capital gains on sale of house



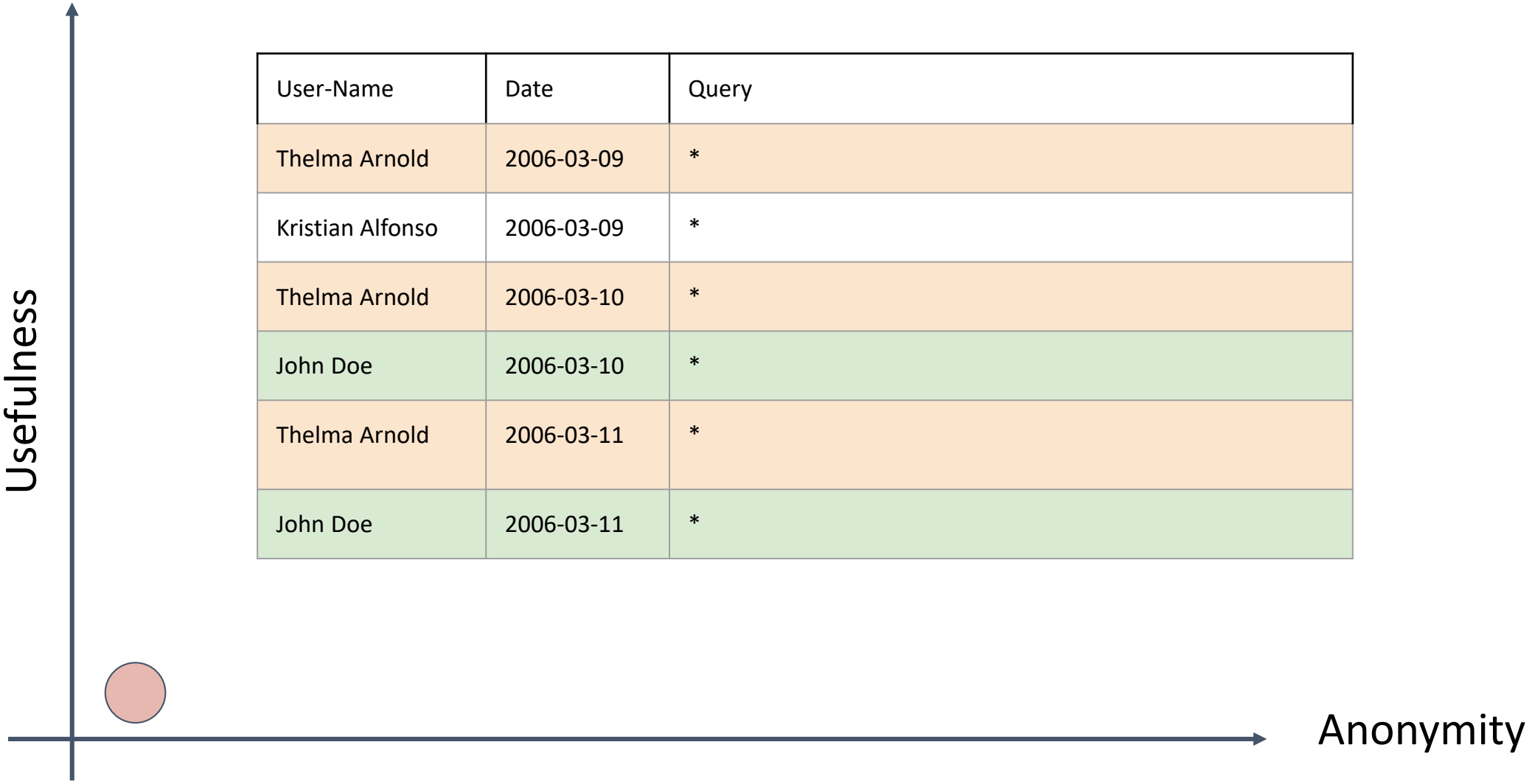
Remove Date



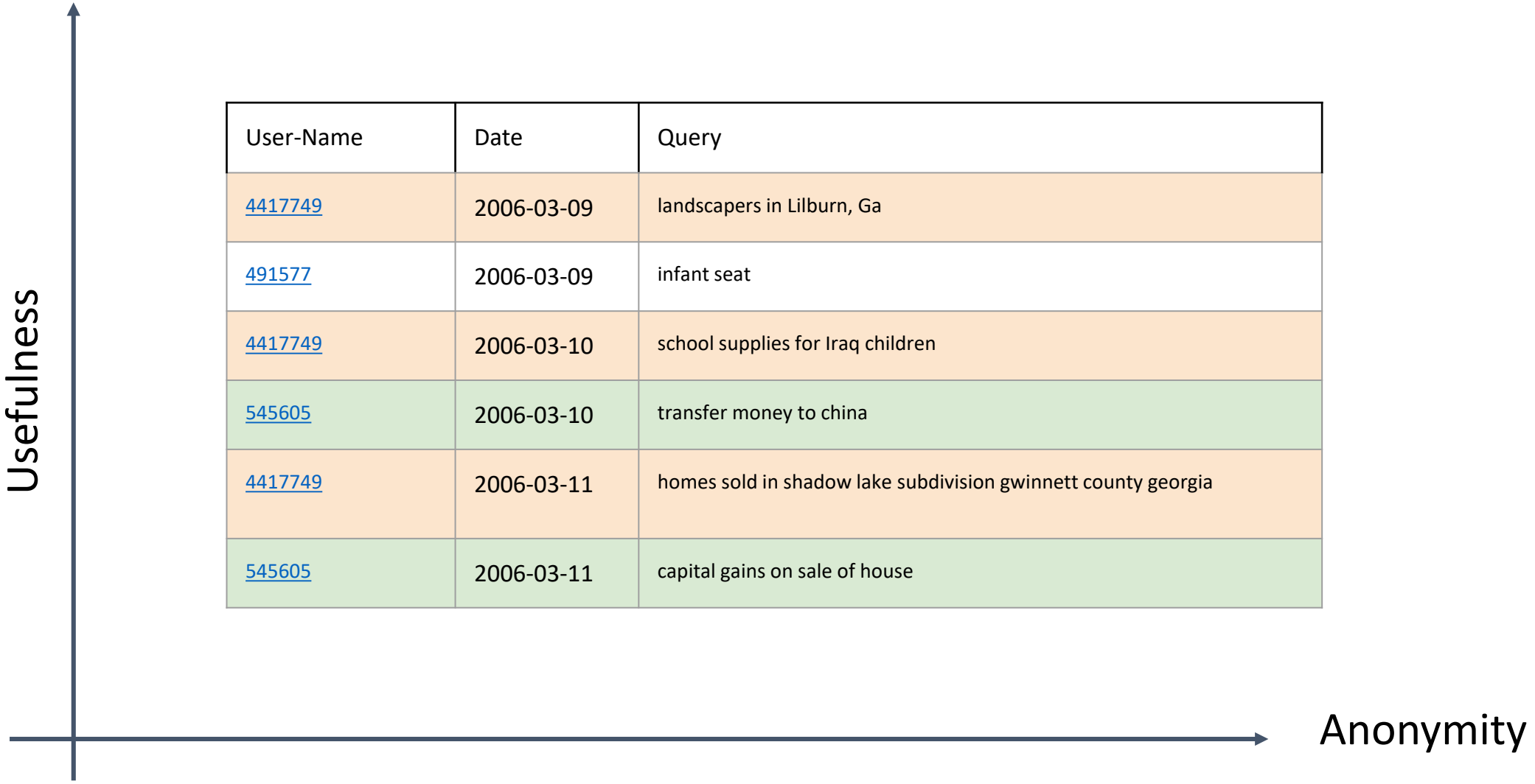
User-Name	Date	Query
Thelma Arnold	*	landscapers in Lilburn, Ga
Kristian Alfonso	*	infant seat
Thelma Arnold	*	school supplies for Iraq children
John Doe	*	transfer money to china
Thelma Arnold	*	homes sold in shadow lake subdivision gwinnett county georgia
John Doe	*	capital gains on sale of house

Remove Query

User-Name	Date	Query
Thelma Arnold	2006-03-09	*
Kristian Alfonso	2006-03-09	*
Thelma Arnold	2006-03-10	*
John Doe	2006-03-10	*
Thelma Arnold	2006-03-11	*
John Doe	2006-03-11	*



Replace User-Names with Random ID

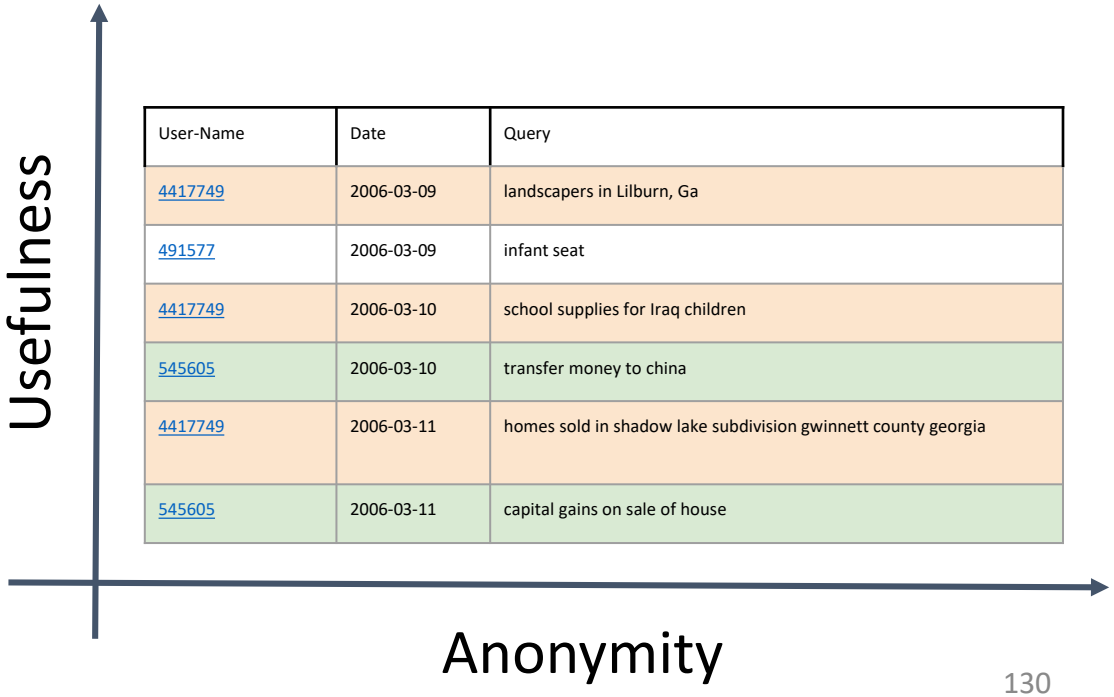


User-Name	Date	Query
4417749	2006-03-09	landscapers in Lilburn, Ga
491577	2006-03-09	infant seat
4417749	2006-03-10	school supplies for Iraq children
545605	2006-03-10	transfer money to china
4417749	2006-03-11	homes sold in shadow lake subdivision gwinnett county georgia
545605	2006-03-11	capital gains on sale of house

Is this enough to say the dataset is anonymized?

A. Yes

B. No



User-Name	Date	Query
4417749	2006-03-09	landscapers in Lilburn, Ga
491577	2006-03-09	infant seat
4417749	2006-03-10	school supplies for Iraq children
545605	2006-03-10	transfer money to china
4417749	2006-03-11	homes sold in shadow lake subdivision gwinnett county georgia
545605	2006-03-11	capital gains on sale of house

AOL Proudly Releases Massive Amounts of Private Data

*Posted Aug 6, 2006 by **Michael Arrington** (@arrington)*

Yet Another Update: AOL: "This was a screw up"

Further Update: Sometime after 7 pm the download link went down as well, but there is at least one mirror site. AOL is in damage control mode – the fact that they took the data down shows that someone there had the sense to realize how destructive this was, but it is also an admission of wrongdoing of sorts. Either way, the data is now out there for anyone that wants to use (or abuse) it.

Update: Sometime around 7 pm PST on Sunday, the AOL site referred to below was taken down. The direct link to the data is still live. A cached copy of the page is here.

AOL must have missed the uproar over the DOJ's demand for "anonymized" search data last year that caused all sorts of pain for Microsoft and Google. That's the only way to explain their release of data that includes 20 million web queries from 650,000 AOL users.

The data includes all searches from those users for a three month period this year, as well as whether they clicked on a result, what that result was and where it appeared on the result page. It's a 439 MB compressed download, expanded to just over 2 gigs. The data is available here (this link is directly to the file) and the output is in ten text files, tab delineated.

The utter stupidity of this is staggering. AOL has released very private data about its users without their permission. While the AOL username has been changed to a random ID number, the abilitiy to analyze all searches by a single user will often lead people to easily determine who the user is, and what they are up to. The data includes personal names, addresses, social security numbers and everything else someone might type into a search box.

Crunchbase

AOL

FOUNDED
1985

OVERVIEW
AOL Lifestream is a web-based application that enables users to keep track of all their comments on social networking sites. Integrated with AIM Express, AIM 7, and AIM for Mac, users can publish their statuses, reply to comments on networking sites from their Lifestream tab, and more. AOL Lifestream is a product of [AOL] (https://www.crunchbase.com/organization/aol#/entity)

LOCATION
New York, NY

CATEGORIES
Digital Media, Advertising Platforms, Content Creators, News

WEBSITE
http://www.aol.com

Full profile for AOL

NEWSLETTER SUBSCRIPTIONS

☐ **The Daily Crunch**
Get the top tech stories of the day delivered

Naive anonymization can easily be broken

In 2006, AOL released search queries of 500,000 pseudonymous users

- Days later, New York Times reveals the identity of user 4417749
- CTO and researchers responsible of sharing data fired

The same year, Netflix revealed over 100 million movie ratings made by 500,000 users after removing personal details

- Researchers de-anonymize dataset by comparing it with publicly available ratings on Internet Movie Database (IMDB)

[Credit: New York Times]

https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/

Re-identification by Linking

Medical Data

ID	QID			SA
Name	Zipcode	Age	Sex	Disease
Alice	47677	29	F	Ovarian Cancer
Betty	47602	22	F	Ovarian Cancer
Charles	47678	27	M	Prostate Cancer
David	47905	43	M	Flu
Emily	47909	52	F	Heart Disease
Fred	47906	47	M	Heart Disease

Voter registration data

Name	Zipcode	Age	Sex
Alice	47677	29	F
Bob	47983	65	M
Carol	47677	22	F
Dan	47532	23	M
Ellen	46789	43	F

Re-identification by Linking

Medical Data

QID			SA
Zipcode	Age	Sex	Disease
47677	29	F	Ovarian Cancer
47602	22	F	Ovarian Cancer
47678	27	M	Prostate Cancer
47905	43	M	Flu
47909	52	F	Heart Disease
47906	47	M	Heart Disease

Voter registration data

Name	Zipcode	Age	Sex
Alice	47677	29	F
Bob	47983	65	M
Carol	47677	22	F
Dan	47532	23	M
Ellen	46789	43	F

Takeaways

Naïve anonymization does NOT work

Attributes themselves are a quasi-identifier

Quasi-identifiers allow to link databases

Linking allows to break anonymity