EPFL

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

# CS-523 Advanced topics on Privacy Enhancing Technologies

# Introduction

**Carmela Troncoso**
SPRING Lab
carmela.troncoso@epfl.ch

Some slides/ideas adapted from: Jean-Pierre Hubaux

# Encryption is enough to solve privacy problems

A. Yes

B. No

# What is Privacy

# Privacy definitions and vocabulary

# What is privacy

**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)

# What is privacy

**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)

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

**Transparency and Accountability:** let the user know how data is used, appoint to responsible entities in case of misbehaviour

**As promoted by** General Data Protection Regulation (GDPR) / Swiss Federal Act on Data Protection (Sept 2023)

# What is privacy

**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)

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

**Transparency and Accountability:** let the user know how data is used, appoint to responsible entities in case of misbehaviour

**As promoted by** General Data Protection Regulation (GDPR) / Swiss Federal Act on Data Protection (Sept 2023)

**In practice?**
Privacy settings
Privacy policy languages
Logging

# What is privacy

**Privacy as PRACTICE**

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

# What is privacy

**Privacy as PRACTICE**

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

**Improve user agency:** help them negotiate privacy

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

**Collective wisdom**: help identify best practices for collectives

# What is privacy

**Privacy as PRACTICE**

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

**Improve user agency:** help them negotiate privacy

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

**Collective wisdom**: help identify best practices for collectives

**In practice?**
Privacy mirrors ("View as")
Recommenders for configuration
Privacy nudges

# What is privacy

**Privacy as CONFIDENTIALITY**

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

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

# What is privacy

**Privacy as CONFIDENTIALITY**

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

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

**Minimize data disclosure:** every bit counts

**Distribute trust:** avoid single points of failure

# What is privacy

**Privacy as CONFIDENTIALITY**

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

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

**Minimize data disclosure:** every bit counts

**Distribute trust:** avoid single points of failure

**In practice?**
Advanced cryptography
Data (principled) perturbation
Decentralization

# What is privacy

**Privacy as CONFIDENTIALITY**

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

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

**Minimize data disclosure:** every bit counts

**Distribute trust:** avoid single points of failure

**In practice?**
Advanced cryptography
Data (principled) perturbation
Decentralization

**We will mostly concentrate on these PETs in this course**

# **Important**: Privacy is not The Goal

## Article 12

No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

https://www.un.org/en/about-us/universal-declaration-of-human-rights

# Topics we will cover

**Attribute-based credentials**

**Anonymization**

**Differential privacy**

**Location privacy**

**Anonymous communications**

**Privacy engineering**

**Privacy-preserving cryptography**

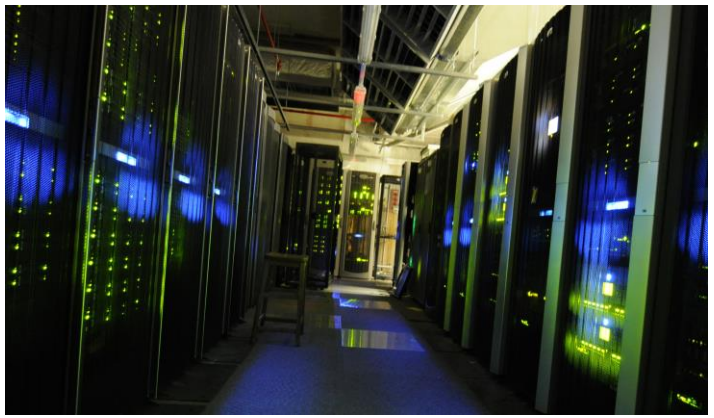
**Tracking**

**Censorship resistance**

**Machine learning & privacy**

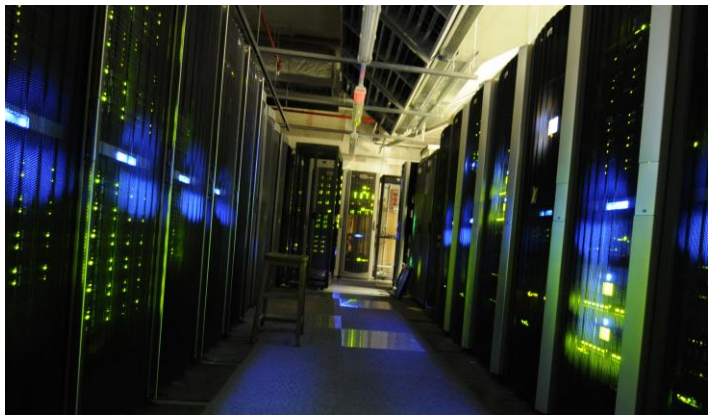
**Legal aspects**

# What is privacy: privacy adversaries



# What is privacy: privacy adversaries
# The adversary are “others”





# How Companies Learn Your Secrets

- Target knows
  - What you buy, when you buy it, how often, …

- (In the US) Target can buy data about:
  - **Online**: what webs you visit, how long, in which order, what kinds of topics you talk about online, what you like, what you share,…
  - **Offline**: Your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought your house, where you went to college reading habits, the number of cars you own, etc….

- Why? Marketing Analytics:
  - Find the customers who have children and send them catalogs that feature toys before Christmas
  - Look for shoppers who habitually purchase swimsuits in April and send them coupons for sunscreen in July and diet books in December


http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0

# When targeted marketing goes wrong

An angry man went into a Target outside of Minneapolis, demanding to talk to a manager

"My daughter got this in the mail!"

"She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

"We deeply apologize"

# When targeted marketing goes wrong

An angry man went into a Target outside of Minneapolis, demanding to talk to a manager

"My daughter got this in the mail!"

"She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

"We deeply apologize"

Manager called a few days later to apologize again

"We apologize again"

"I had a talk with my daughter"

"It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology"

https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

# What is privacy: privacy adversaries
# The provider is semi-trusted





**Data Protection Regulations**

# EU: General Data Protection Regulations (GDPR)

Aims to ensure the fair, lawful and transparent processing of personal data, by organizations operating in the EU or processing personal data of people residing in the EU, and subject to independent regulatory oversight

**Personal data**: any information relating to an identified/identifiable natural person

Assumes a **Data Controller** (the entity that determines the means and purposes of the processing) who can subcontract **Data Processors** (entities processing under the authority of the controller)

**Lawfulness**: legal ground for processing (e.g. consent, contract, law, balancing)

**Purpose limitation**: processing 'only' for limited, specific, explicit purposes

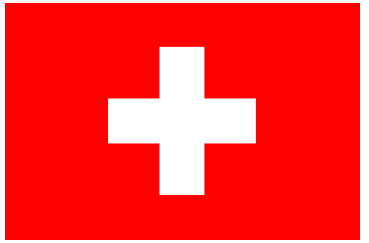
**Data minimization**: processing only data that is needed for the purpose

**Transparency**: about processing and purposes towards data subject

**Accountability**: ability to demonstrate compliance, risk management and privacy-by-design

**Data subject rights**: access, correction, object, erasure, portability, profiling

# Switzerland



- New Swiss Data Protection Act – Entered in force September 2023

- Main changes with respect to previous regulation

  **Only data of natural persons** (as opposed to legal persons)

  **Genetic and biometric data made explicitly sensitive**

  **Introduces "Privacy by Design" and "Privacy by Default"**

  **Mandatory register of processing activities** (except SMEs with limited harm risk).

  **Prompt notification** to (FDPIC) when there is a breach

  **Automated processing of personal data** part of the law!

- **Very close** to GDPR (minimize number of changes for compliance)

https://www.kmu.admin.ch/kmu/en/home/facts-and-trends/digitization/data-protection/new-federal-act-on-data-protection-nfadp.html

# Trust the provider?

Laws are **hard to scope** (… and enforce)

**GDPR**

- Purpose limitation: if the purpose is broad...
- Personal data: if data is not personal…
- Consent: if consent is given… (and there are other basis beyond consent)

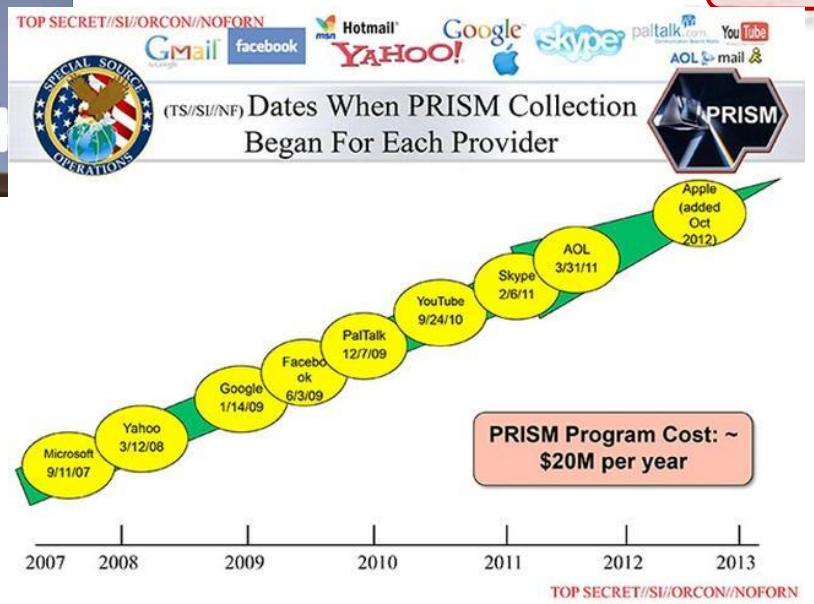
# Trust the provider?
# Snowden revelations



(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.



# Trust the provider?

Laws are **hard to scope** (… and enforce)

## GDPR

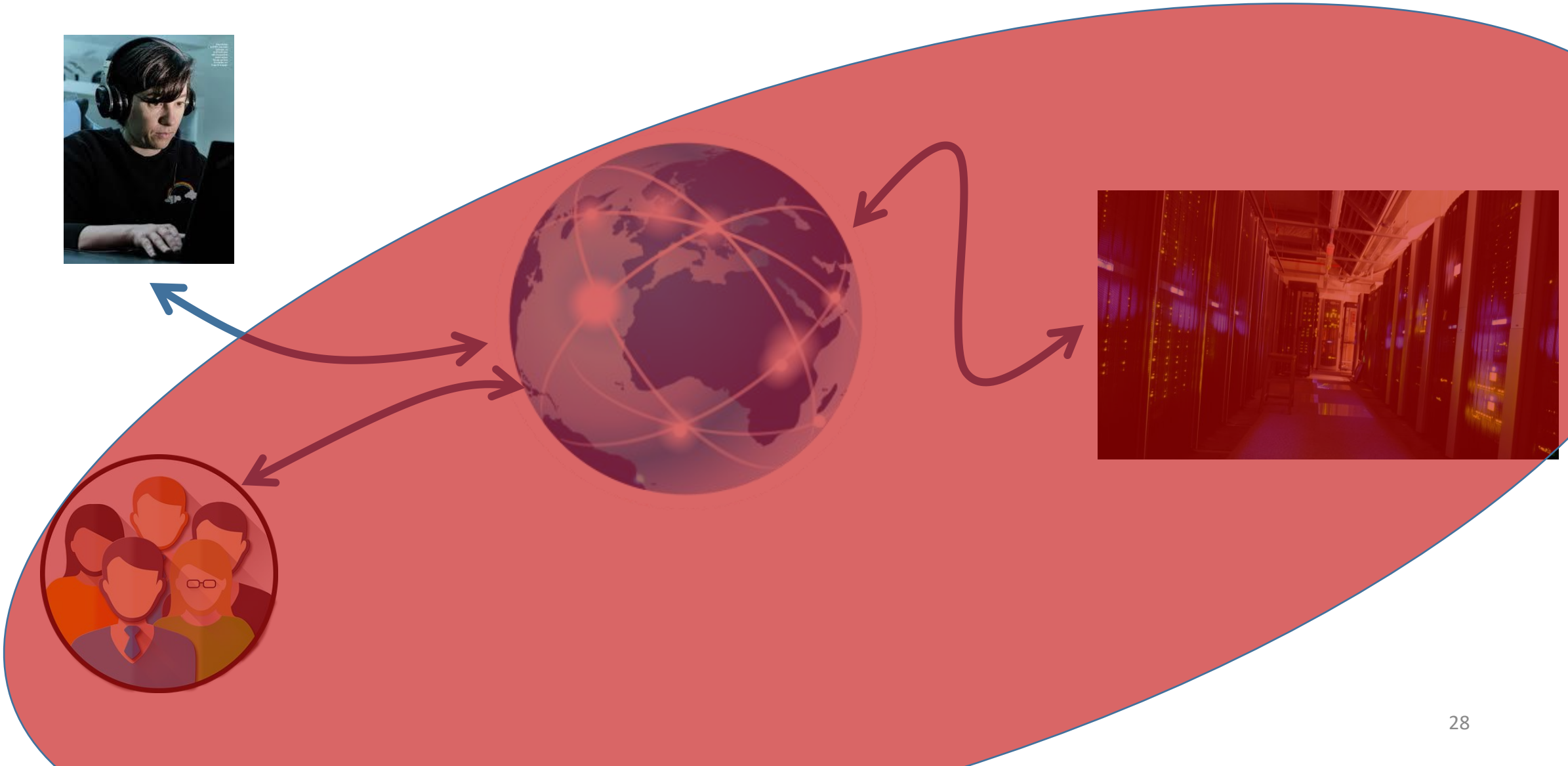
- Purpose limitation: if the purpose is broad...
- Personal data: if data is not personal…
- Consent: if consent is given… (and there are other basis beyond consent)

## United States

In theory, NSA analysts are not allowed to specifically target someone “reasonably believed” to be a US person communicating on US soil but in Snowden’s revelations it becomes clear that NSA’s “contact chaining” practices (an analyst collects records on a target’s contacts, and their contacts’ contacts) can easily cause US citizens to be caught up in the process

# What is privacy: privacy adversaries

## No-one is trusted



# Don’t trust the provider… or anyone

Don’t trust the infrastructure…

Minimize trust on any entity

Includes implicit (meta)data leaks

# Don’t trust the provider… or anyone

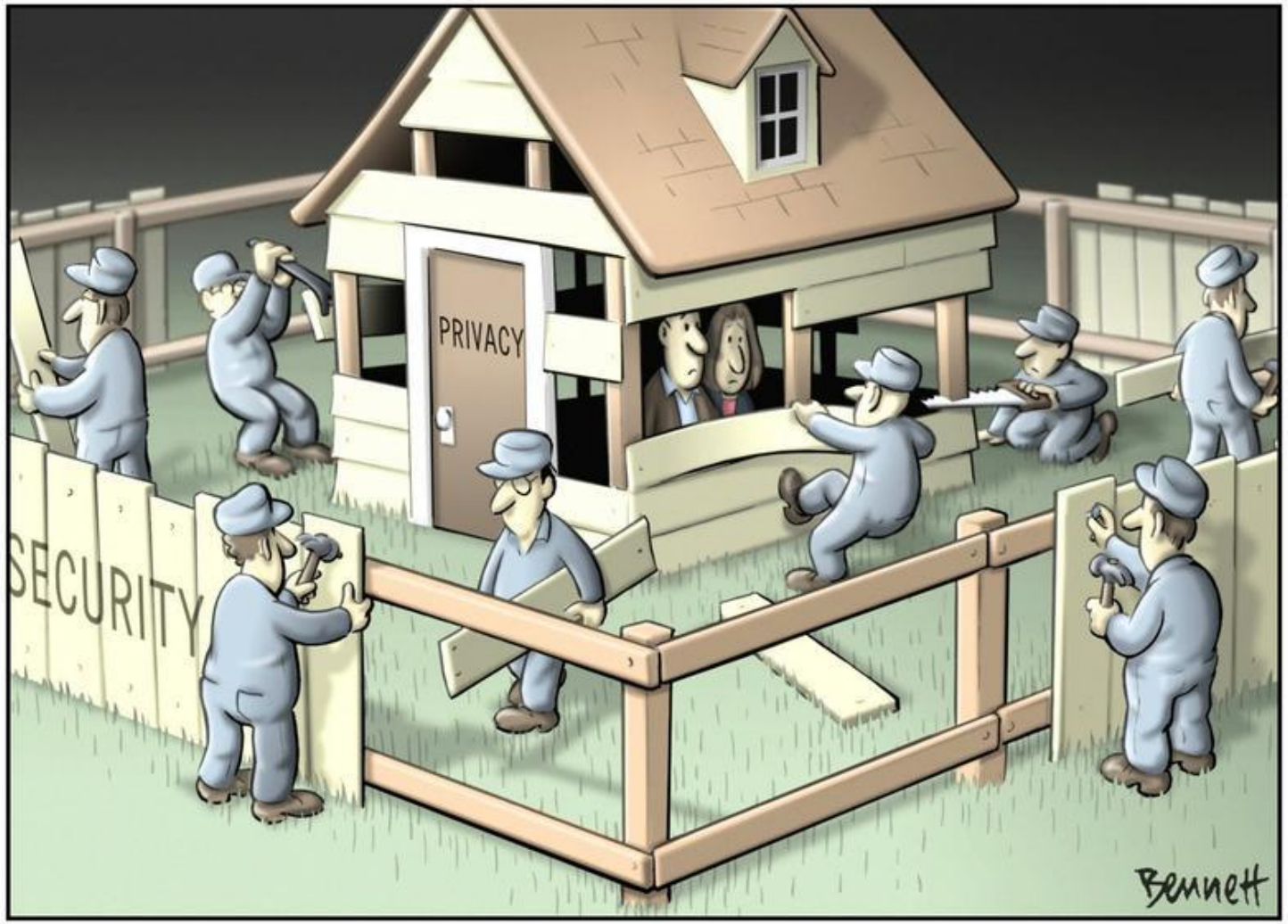
Don’t trust the infrastructure…

Minimize trust on any entity

Includes implicit (meta)data leaks

**We will mostly concentrate on this adversarial model in this course**

# Security and/or privacy?



# Are privacy and security contradictory? (more security implies less privacy and vice versa)

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** ??

*Surveillance may be not **effective***: smart adversaries evade surveillance
criminals use Telegram, Threema, Signal,... but average users do not!!

*Surveillance tools can be **abused***: lack of transparency and safeguards
NSA spying on Americans, Spanish ministry spying independentist politicians, 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

# Privacy is either for all or none

**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

**For individuals**

protection against profiling and manipulation
protection against crime / identity theft

Thus, privacy is a Security property -- *there is no security without privacy*

**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

# Corollary 1: In privacy there is a **strategic** adversary

**COM-301**

## The Resourced Strategic Adversary?

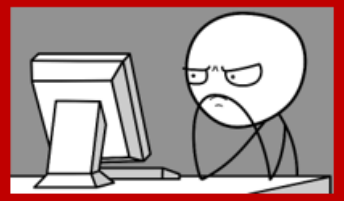
**COMPUTER SECURITY**
Properties of a computer system must hold in presence of a **resourced strategic adversary**

**THREAT MODEL**: describes the **resources** available to the **adversary** and their capabilities (observe, influence, corrupt,…)

The **adversary** is a malicious entity aiming at breaching the security policy

The **adversary** is **strategic**: the adversary will choose the **optimal** way to use her resources to mount an attack that violates the security properties

**THREAT MODELLING IS A VERY HARD TASK!!**



9

# Corollary 2:
# Security design principles apply for privacy

1. Economy of mechanism
2. Fail-safe defaults
3. Complete mediation
4. Open Design
5. Separation of Privilege
6. Least Privilege
7. Least Common Mechanism
8. Psychological Acceptability

9. Work Factor
10. Compromise recording

J. Saltzer and M. Schroeder. *The Protection of Information in Computer Systems*. Fourth ACM Symposium on Operating Systems Principles (October 1973) (Intro & Section 1) https://www.acsac.org/secshelf/papers/protection_information.pdf

# But there is more to privacy

**Cryptography → Confidentiality!**

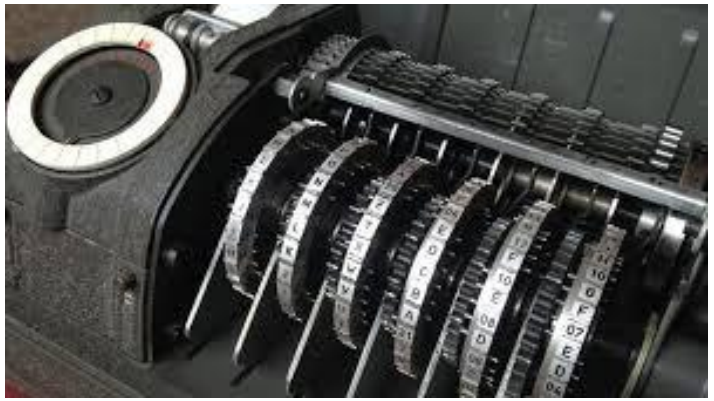
**Traditional:** computer security context

**Privacy is different than traditional confidentiality.**

**What makes Privacy Enhancing Technologies (PETs) different:**

- Threat model: **weak** actors, **powerful** adversaries.
- Susceptibility to **compulsion**.
- Cannot assume the existence of **Trusted Third Parties (TTP)**
- Also worry about **Cost, Collusion, Corruption, Carelessness**.

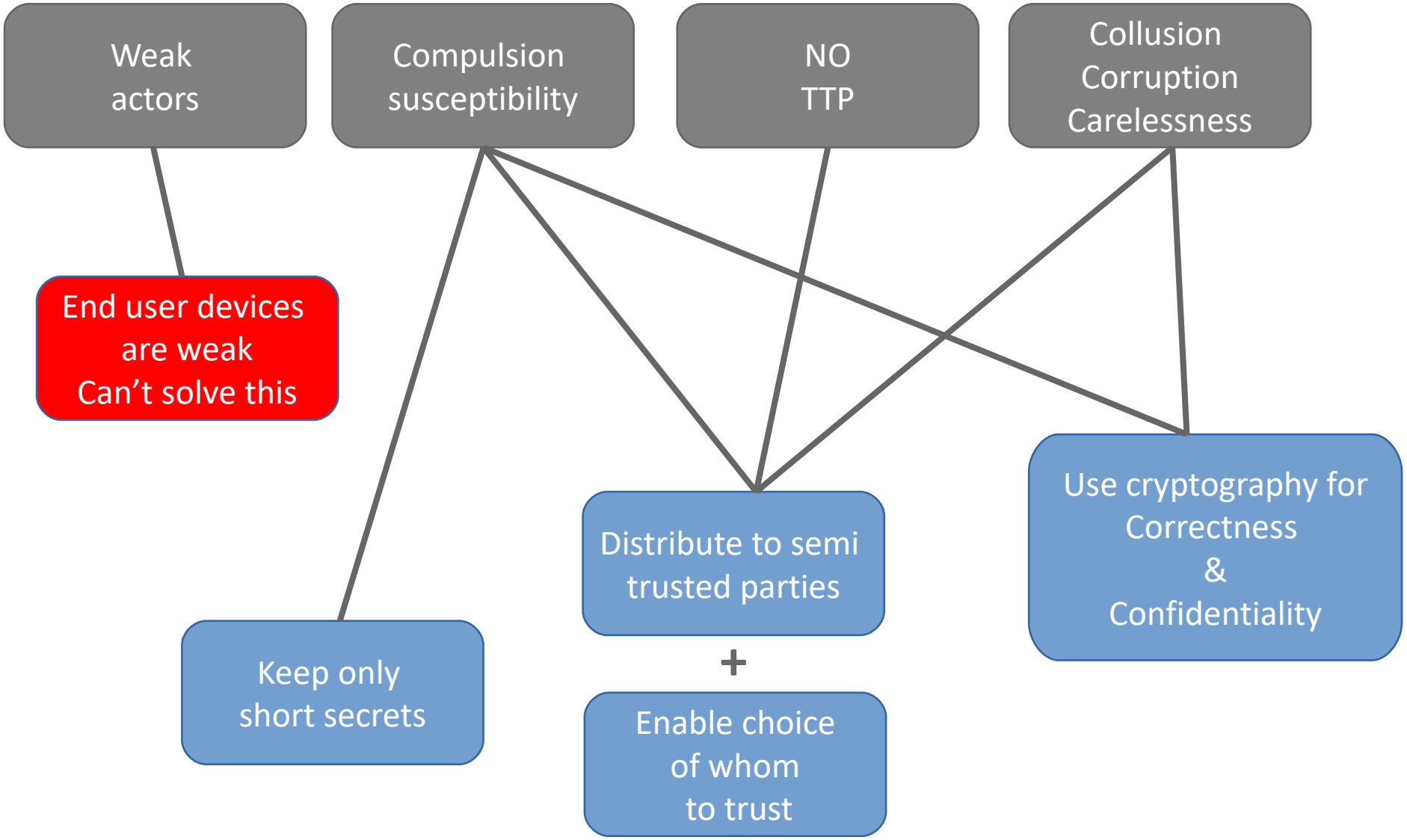
# Crypto AG scandal

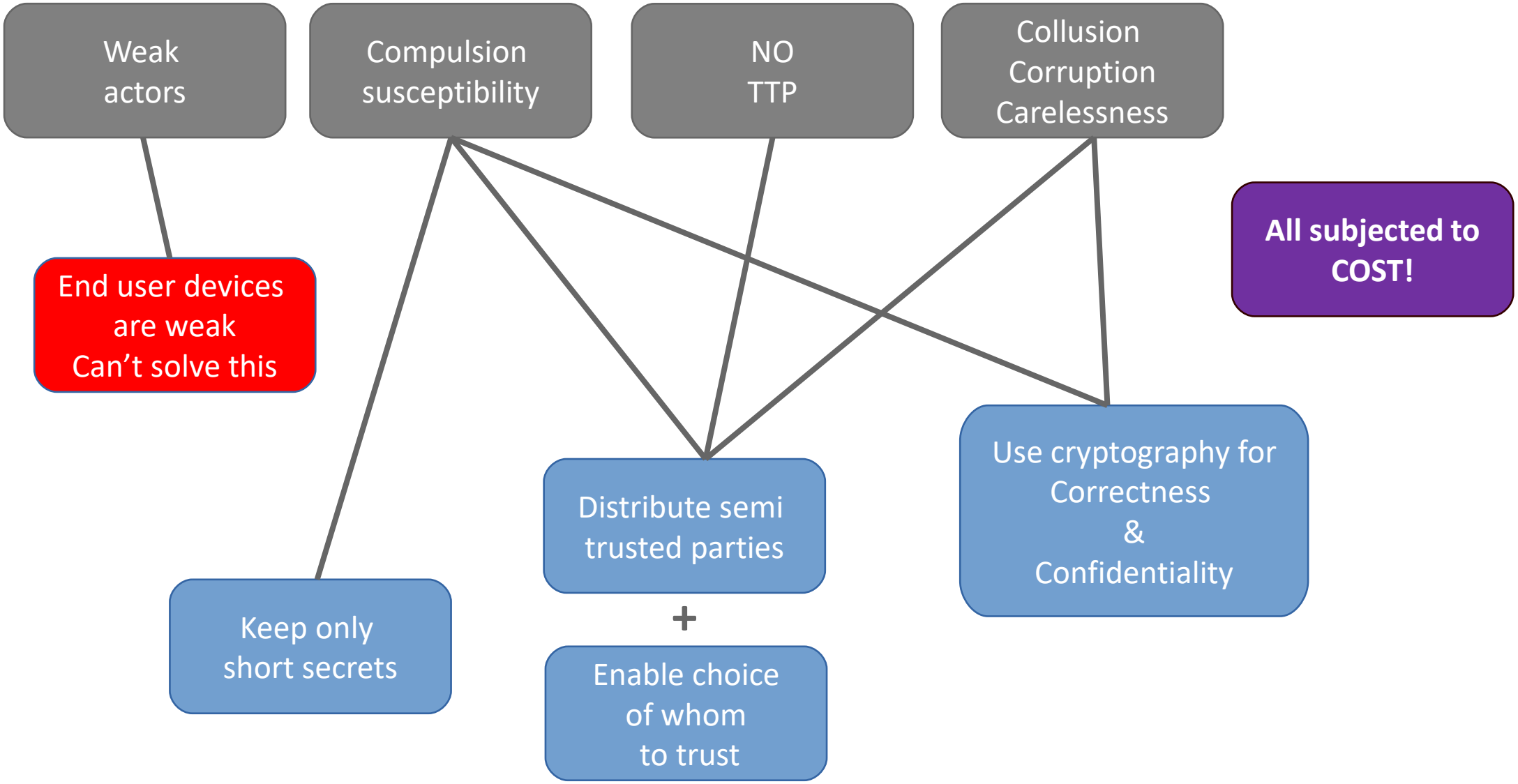
- From 1952 to 2018, Crypto AG sold encryption devices that contained a trapdoor to a large number of countries
- For several decades, the (covert) owners of the company were the CIA and its German counterpart
- In spite of complaints from former employees, enquiries by the Swiss authorities (notably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

https://pages.rts.ch/emissions/temps-present/10977257-operation-rubicon-un-reseau-d-espionnage-au-coeur-de-la-suisse.html#11012902

# PETS design principles (in this course)



# PETS design principles (in this course)



# Privacy Quantification

**No Free Lunch** Theorem [1]:

*For every algorithm that outputs a D with even a sliver of utility, there is some adversary with a prior such that privacy is not guaranteed*




[1] “No Free Lunch in Data Privacy”, Kifer and Machanavajjhala, 2011

# Privacy properties

Confidentiality (in transit/storage, during processing)

Pseudonymity

Anonymity

Unlinkability

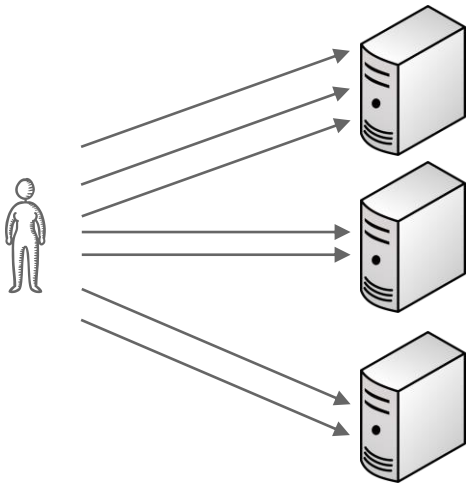
Unobservability

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.”

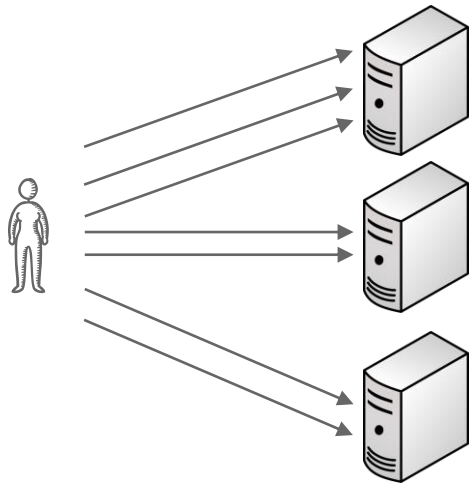


**Identity**

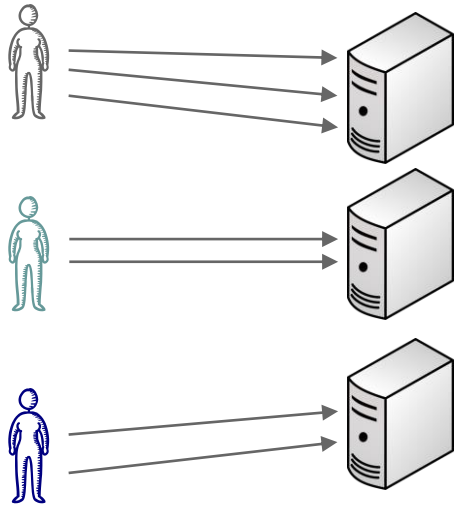
# Privacy properties: **Pseudonymity**

**-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.”



**Identity**

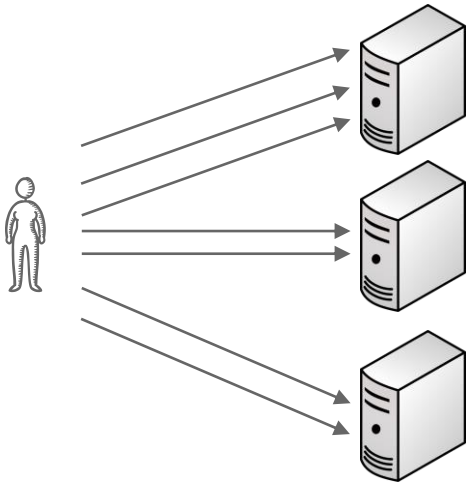


**Pseudonymity**

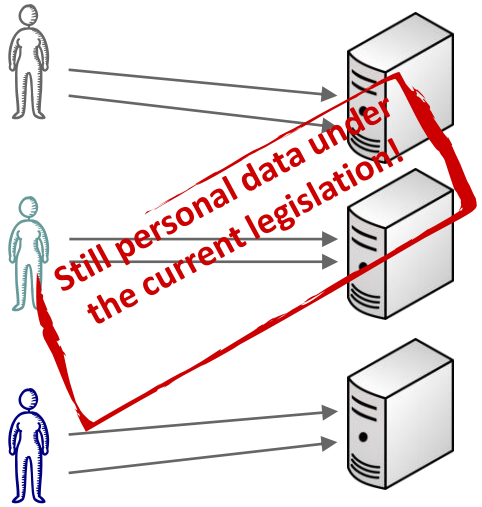
# Privacy properties: **Pseudonymity**

**-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.”



Identity




Pseudonymity

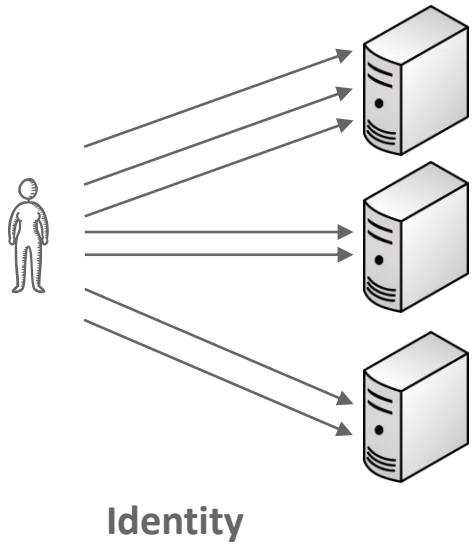


Pfitzmann, Andreas and Hansen, Marit. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

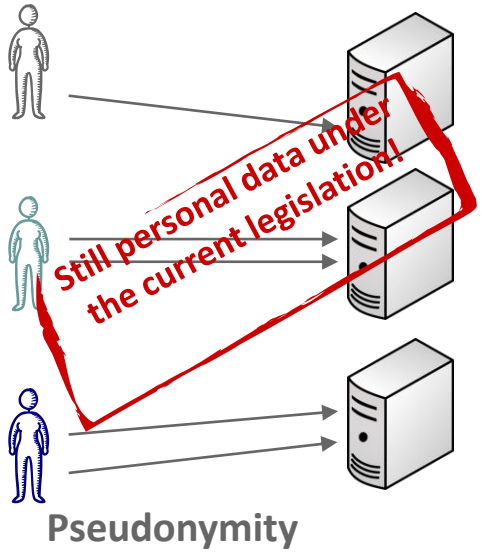
# Privacy properties: **Pseudonymity**

**-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”

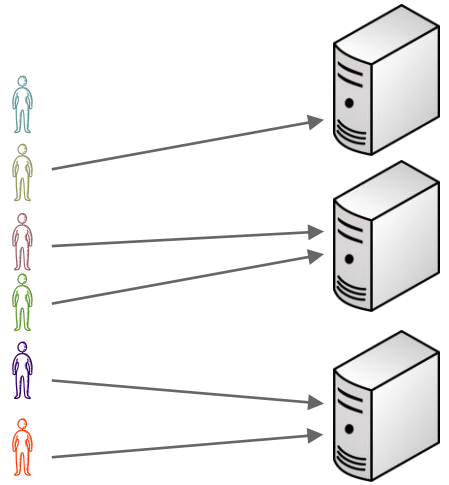
**-ISO 15408**: “a user may use a resource or service without disclosing its identity, but can still be accountable for that use.”



Identity




Pseudonymity



Anonymity

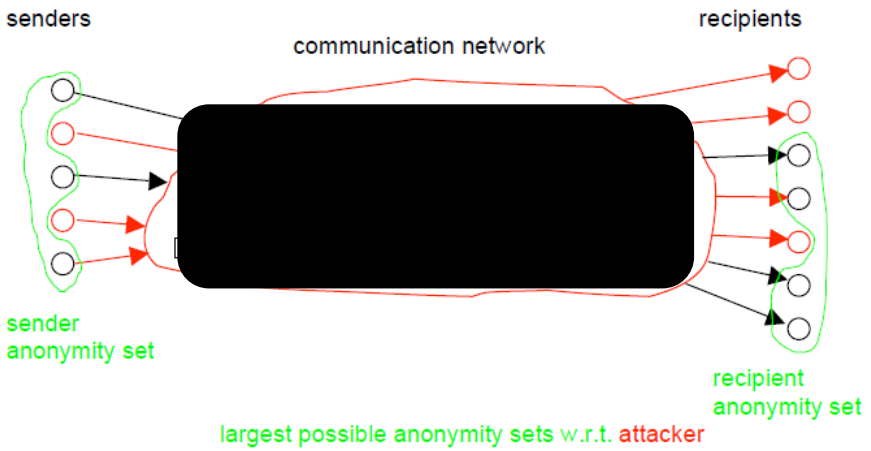


Pfitzmann, Andreas and Hansen, Marit. A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

# 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”

# 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!**

# Anonymity vs. Privacy

- Anonymity does not always imply privacy
- Example: Bob's record is indistinguishable from records of other patients that live in ZIP starting by 130 and are in their 30s
  - Yet, we can infer Bob has Cancer

| Bob | |
|---|---|
| **Zipcode** | **Age** |
| 13039 | 33 |

| | Non-Sensitive | | | Sensitive |
|---|---|---|---|---|
| | Zip code | Age | Nationality | Condition |
| 1 | 130** | < 30 | * | AIDS |
| 2 | 130** | < 30 | * | Heart Disease |
| 3 | 130** | < 30 | * | Viral Infection |
| 4 | 130** | < 30 | * | Viral Infection |
| 5 | 1485* | ≥ 40 | * | Cancer |
| 6 | 1485* | ≥ 40 | * | Heart Disease |
| 7 | 1485* | ≥ 40 | * | Viral Infection |
| 8 | 1485* | ≥ 40 | * | Viral Infection |
| 9 | 130** | 3* | * | Cancer |
| 10 | 130** | 3* | * | Cancer |
| 11 | 130** | 3* | * | Cancer |
| 12 | 130** | 3* | * | Cancer |

# 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!**

# 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!**

# Unobservability vs. Anonymity

Unobservability implies anonymity

Anonymity ***does not*** imply unobservability

- Anonymity only hides the identity of the sender/receiver, it does not guarantee the action is invisible

# 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) — **Cryptographic proofs**

Pseudonymity

Anonymity

Unlinkability

Unobservability

Plausible deniability

# Systematic Privacy Evaluation

1) Model the privacy-preserving mechanism

   typically: 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 strategic 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!!

# Why is privacy important?

# **Reminder**: Privacy is not a goal

## Article 12

No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

https://www.un.org/en/about-us/universal-declaration-of-human-rights

# 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":

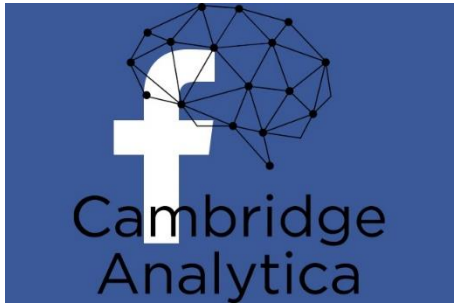
"...a bureaucracy with inscrutable purposes that uses people's information to make important decisions about them, yet denies the people the ability to participate in how their information is used"

"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 Chinese Social Credit System


https://www.wired.co.uk/article/china-social-credit-system-explained




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

# Take aways

Many flavors of privacy: in CS-523 we study the strongest protection

Privacy is a security property

- There is an adversary and they matter
- Security design principles apply, but there is more to it
  - Privacy adversaries are very strong: ensuring privacy is hard

When building Privacy Technologies we need properties to formalize the design objective

**Privacy is not the goal**, it is a key ingredient to maintain our societal and democratic values