

# Class Outline

<https://team.inria.fr/privatics/team-members/mes-cours/>

- **C1- Data Privacy**

- Introduction to surveillance and privacy
- Data Privacy Impact Assessment - DPIA (lab)

- **C2- IA & Human Decision Making**

- How IA can help humans to make decisions
  - ADS (algorithm decision making systems)
- How IA can be used to manipulate human decisions
  - Data manipulation
- IA and Policy

- **C3- Introduction to Internet and Data security (if time)**



# An Introduction to Privacy, Dataveillance, Data Manipulation

**Claude Castelluccia**

**[Claude.castelluccia@inria.fr](mailto:Claude.castelluccia@inria.fr)**

**2021**

**<https://team.inria.fr/privatics/team-members/mes-cours/>**

Janv.2019

# What is Privacy?

# What is Privacy?

- A couple of popular definitions:
  - “The right to be let alone”
    - Focus on freedom from intrusion
  - “Informational self-determination”
    - Focus on control
- Privacy is a fundamental right!
  - Universal Declaration of Human Rights of UN (article 12), 1948
    - “NO ONE SHALL BE SUBJECTED TO ARBITRARY INTERFERENCE WITH HIS PRIVACY, FAMILY, HOME OR CORRESPONDENCE”.
  - GDPR (since may 2018)



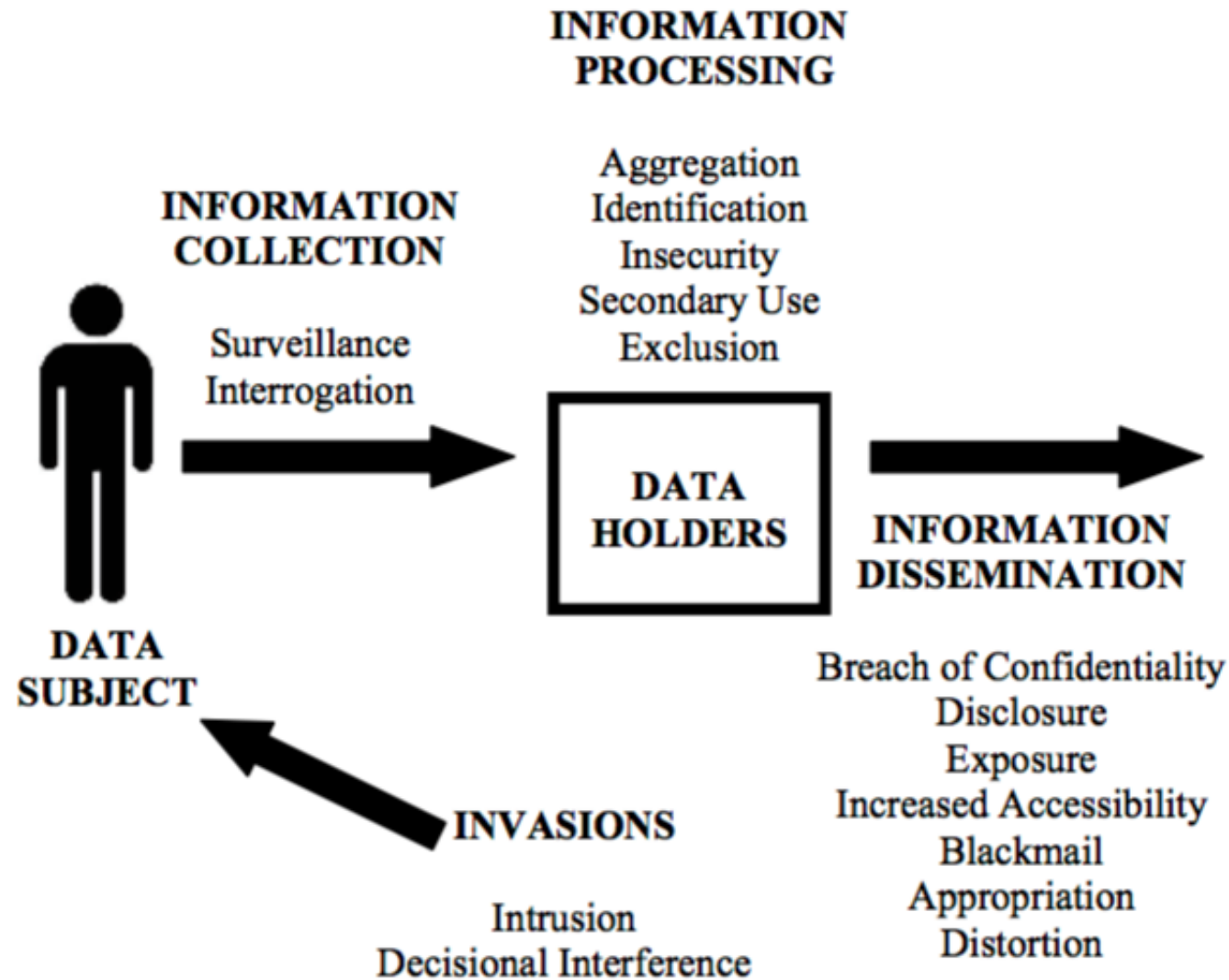
# Privacy: An Evolving Concept...



Roman citizens engaged in conversation in a public restroom

# Privacy is more than Personal Data Protection

- Privacy is often considered as similar to Personal Data Protection
- But it is also
  - control
  - Transparency
  - Lack of Interference (censorship,...)
  - Autonomy
  - ...



*A Taxonomy of Privacy by Daniel Solove*

# Popular Argument against Privacy

- The “nothing-to-hide” argument
  - “If you care so much about your privacy it’s because you have something to hide”

# Popular Arguments against Privacy

- The “nothing-to-hide” argument
  - “If you care so much about your privacy it’s because you have something to hide”
  - Really???

# Popular Arguments against Privacy

- The “nothing-to-hide” argument
  - “If you care so much about your privacy it’s because you have something to hide”
  - Really???
  - Then give me your passwords!!

**PRIVACY != SECRECY**

# This information is not necessarily secret, but do you want to disclose it?

- **Identity attributes** : Name, age, gender, race, IQ, marital status, place of birth, address, phone number, ID number...
- **Location**: Where you are at a certain point in time, movement patterns
- **Interests / preferences** : Books you read, music you listen, films you like, sports you practice
- **Political affiliation, religious beliefs, sexual orientation**
- **Behavior**: Personality type, what you eat, what you shop, how you behave and interact with others
- **Health data**: Medical issues, treatments you follow, DNA, health risk factors
- **Social network**: Who your friends are, who you meet when, your different social circles
- **Financial data**: How much you earn, how you spend your money, credit card number,...



# *An Introduction to Surveillance Studies*

- Not a new research area, but mostly studied by Sociologists [DL]!
  - Work from Bentham (*Panopticon*), David Lyon, M.Foucault,...
  - We live in a monitored world...
  - Workers are watched by their bosses, students by their professors, children by their parents, CCTV cameras, ...
- Why are we monitored?
  - Security, Business, Sorting (i.e. insurance),...
  - But it is mostly about **POWER**
    - Asymmetry of information gives a lot of power to the watcher
    - Important, for example, in negotiations (economic intelligence )
- Consumer (by Industry) vs GOV surveillance

*[DL] David Lyon, Surveillance Studies, An Overview.*

# From Surveillance to Dataveillance

- More and more data generated and collected (with its benefits)...
- Today the smallest details of our daily lives are tracked and traced more closely than ever before (**liquid** surveillance or **Dataveillance**)!
- We leak data, leave traces when we browse the web or use our phones...
  - On the **visible** web
  - On the **invisible** web
- Things got worse since 9/11, recent attacks!
  - See New French Law on SIGINT (Signal Intelligence) [P15]
- With IoT, it will even get worse [R15]

[P15] J. Parra-Arnau, C. Castelluccia, "Dataveillance and the false-positive paradox," in *Proc. 1st International Workshop on Privacy and Inference, (PrInf 2015)*

[R15] Ron Deibert, *The Geopolitics of the Cyberspace After Snowden*

# Dataveillance on the « Visible » Web

- Foursquare knows where you are
- Flickr knows what you see
- Facebook knows what you do
- Linkedin knows what you've done
- Twitter knows what you say
- Amazon knows what you buy
- Google knows what you think



# Dataveillance on the “Invisible” Web

- “**Meta-data**” (as opposed to data/content)
- Tags, Web bugs, pixels and beacons that appear on Websites to track and profile users
- Allows trackers to **build profiles of users** (mostly for advertisers!).



# Dataveillance on the “Invisible” Web



# Meta-Data = Surveillance Data

- What are the metadata?
  - Signaling messages,...
  - Caller and receiver, current location
  - Length of call, Visited sites...
- Surveillance  $\neq$  Eavesdropping
- Metadata is data of surveillance by nature [Schneier]!
  - An private detective that put a subject under surveillance will collect where he went, what he did, who he spoke to, for how long, what he read....
  - This is meta-data!
- Metadata can not be encrypted! Furthermore, traffic analysis is still possible on encrypted data!



# Consumer Surveillance: Why Are We Profiled?

- For profit
  - Data is believed to be the new oil!
- For targeting users, as in **targeted** Advertising!
  - Targeted Ads are believed to increase Click-to-Rate (i.e. Ad Network revenues)
    - AdNetworks only get paid when a user clicks!
    - Increased bidding prices (see RTB)



# Targeted Ads

## Targeted Ads

- Ads are more and more targeted to your interests

## Example:

- you want to buy a shirt and visit [www.laredoute.fr](http://www.laredoute.fr)





# Targeted Ads

The screenshot shows a web browser window displaying the Washington Post website. The address bar shows the URL: [http://www.washingtonpost.com/wp-dyn/content/article/2011/02/08/AR2011020805595.html?tid=nn\\_twitter](http://www.washingtonpost.com/wp-dyn/content/article/2011/02/08/AR2011020805595.html?tid=nn_twitter). The page features the Washington Post masthead and a navigation bar with links like 'Hot Topics', 'Verizon iPhone', 'Jules Verne', 'J. Paul Getty III dies', 'CPAC 2011', 'Spider-Man' play', 'Keith Olbermann', and 'Carolyn Hax'.

Below the navigation bar, there is a section for 'Esprit' clothing items, including an 'Esprit blouse' for € 29,95, an 'Esprit pullover' for € 49,95, and an 'Esprit T-shirt' for € 25,95. A red arrow points from this advertisement section towards the main article.

The main article is titled 'Twitter data privacy in dispute in WikiLeaks case' by MATT APUZZO, published by The Associated Press on Tuesday, February 8, 2011, at 8:35 PM. The article text reads: 'WASHINGTON -- Three people associated with the website WikiLeaks are asking a federal judge not to force the social networking site Twitter to turn over data about whom they communicate with online. In court documents unsealed Tuesday, the three challenged a court order forcing Twitter to tell the government the names of those they talk to privately and who follow their posts. Attorneys argued that violated their freedom of speech. The documents capture the heart of the WikiLeaks debate because the U.S. is'.

On the right side of the page, there is a 'Network News' section with a 'Recommend' button and a 'Tweet' button. Below this is a 'TOOLBOX' section with links for 'Resize', 'Print', 'E-mail', and 'Reprints'. Further down, there is an 'Advertisement' section titled 'NEW BY HowLIFEWORKS' featuring three articles: 'The 3 Things Your Auto Insurance Company Doesn't Want You to Know', 'How to Make Your PC as fast as the Day You Bought It', and 'How to Regain Your Memory Sharpness After Age 40'. At the bottom right, there is a 'FEATURED ADVERTISER LINKS' section with links for 'Gas Drilling Lawsuit: PA, Mesothelioma, Ya...', 'Landfill odor, Crohn's, DePuy hip surgery, /', 'Lung Cancer, Hydroxycut, Zimmer Flex Fail', and 'Femur Fracture, Paxil birth defect'.

You are one step away from helping Aissata  
and other local health workers help more kids.

Find out all the ways you can help >> See Where the Good Goes™



Save the Children.

AN ENCYCLOPÆDIA  
BRITANNICA COMPANY



m-w.com

Word Games

Word of the Day

New Words & Slang

Video



Dictionary

Thesaurus

Spanish-English

Medical



TOP 10 LISTS >



## The Language of Love (And Related Emotions)

Top 10 Words for Valentine's Day



TREND WATCH >



### "Inclement"

As a major snow storm approached the  
midwest and northeast U.S. ... [more >](#)



### "Mercurial"

After Keith Olbermann suddenly  
announced ... [more >](#)



### "Tawdry"

When ... the word to describe  
his passionate ...



Word of the Day

FEBRUARY 09, 2011

**vicissitude**

An unexpected change or fluctuation

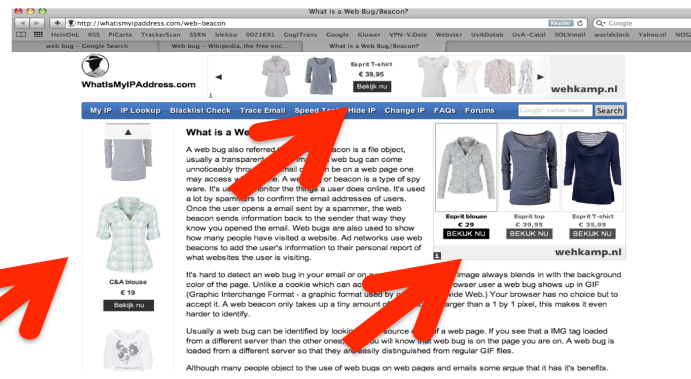
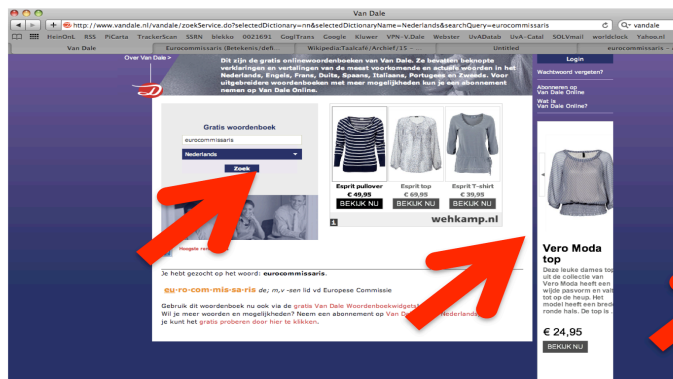
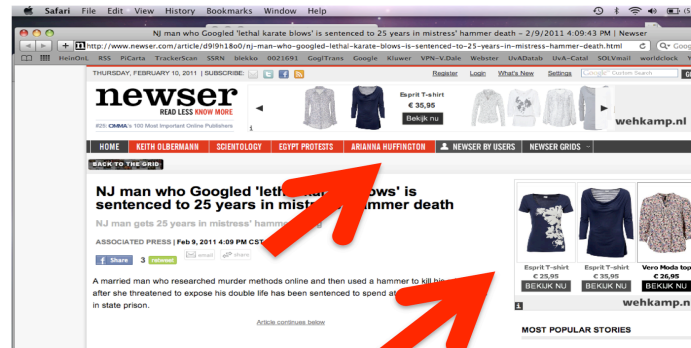


Our Free  
iPhone App  
with  
Voice Search  
**get it now!**

[Learn More  
About Our  
Privacy Policy](#)



# Sticky Ads ;)

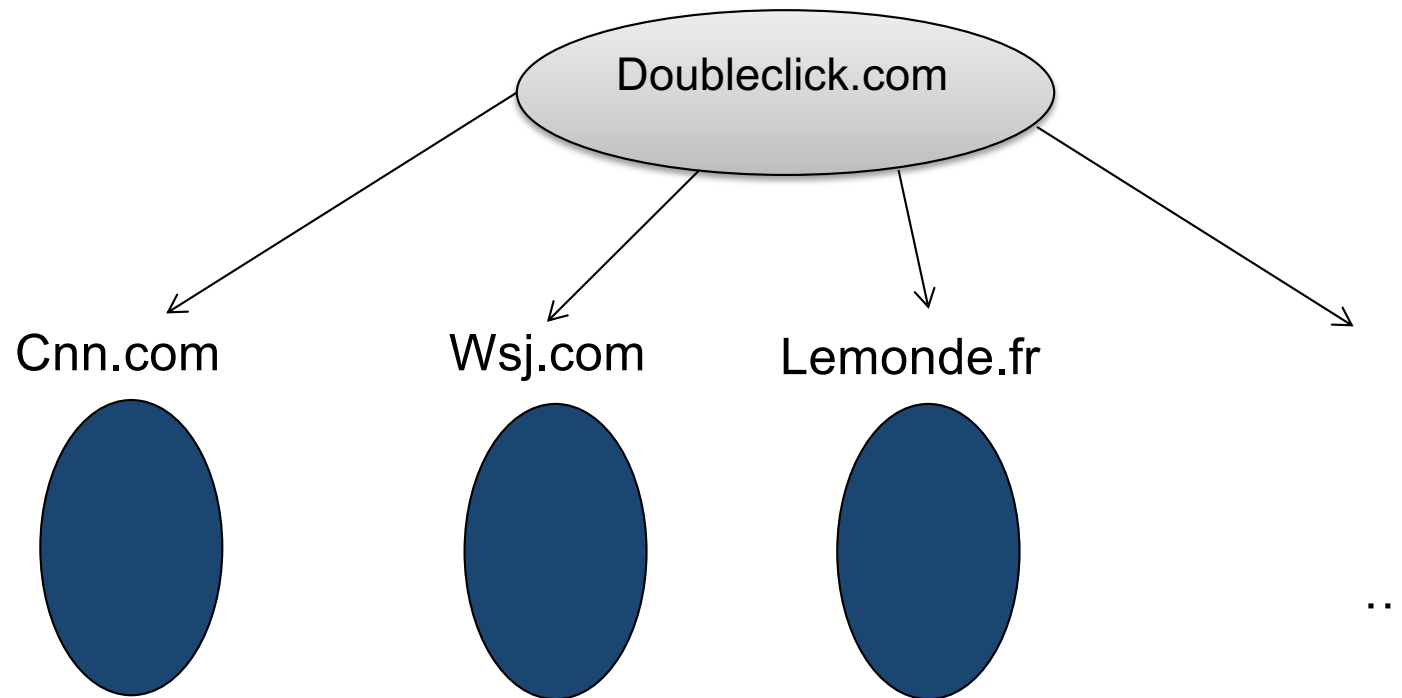


# ***How Are we Being Tracked?***

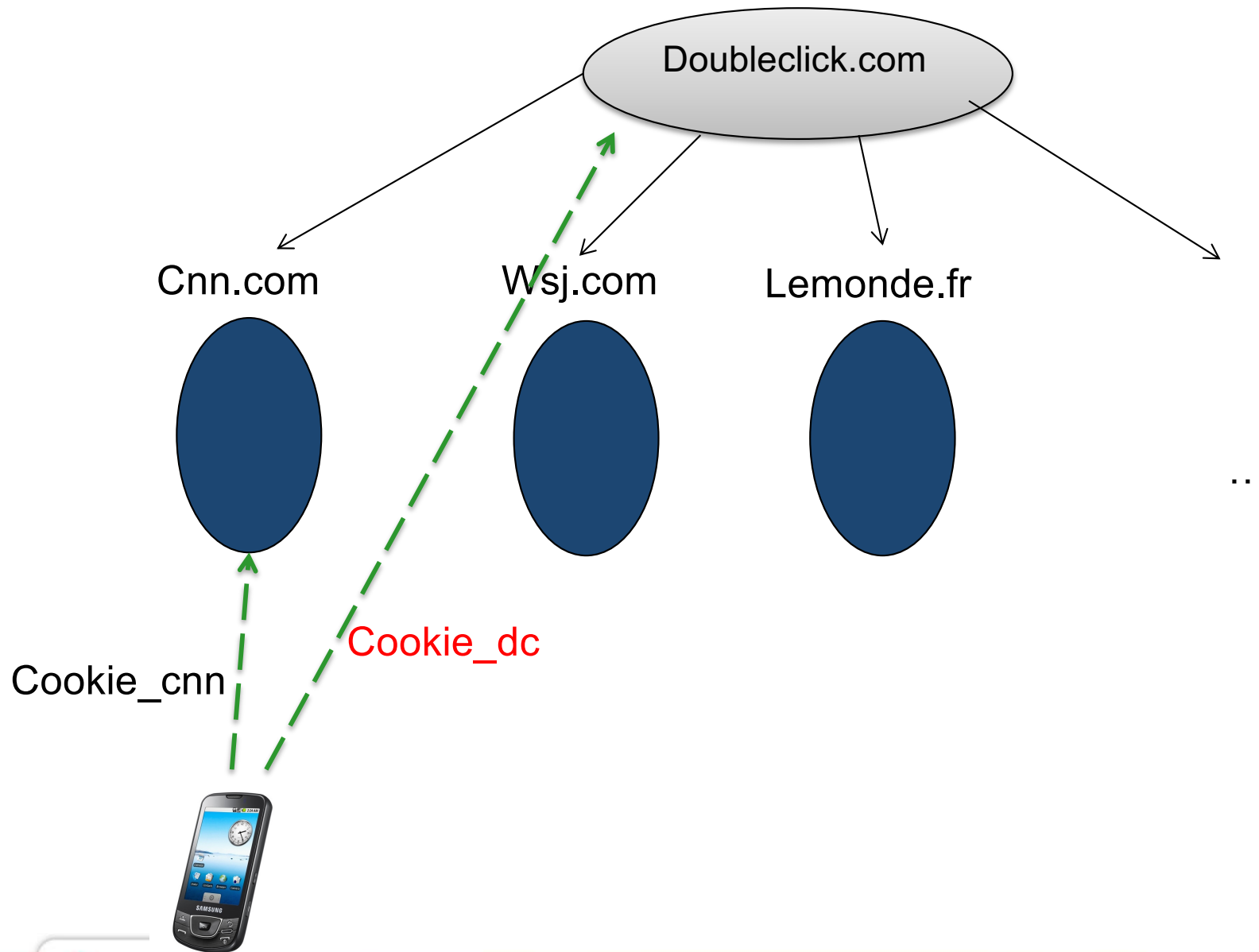
- HTTP cookies
- Local Shared Objects (Flash Cookies)
- Invisible pixels
- Scripts
- Fingerprinting (passive)
- Real Time Bidding (RTB)
- Phones
- Many different ways (see [MISC15])

*[MISC15] Souriez, vous etes tracés, <http://www.miscmag.com>*

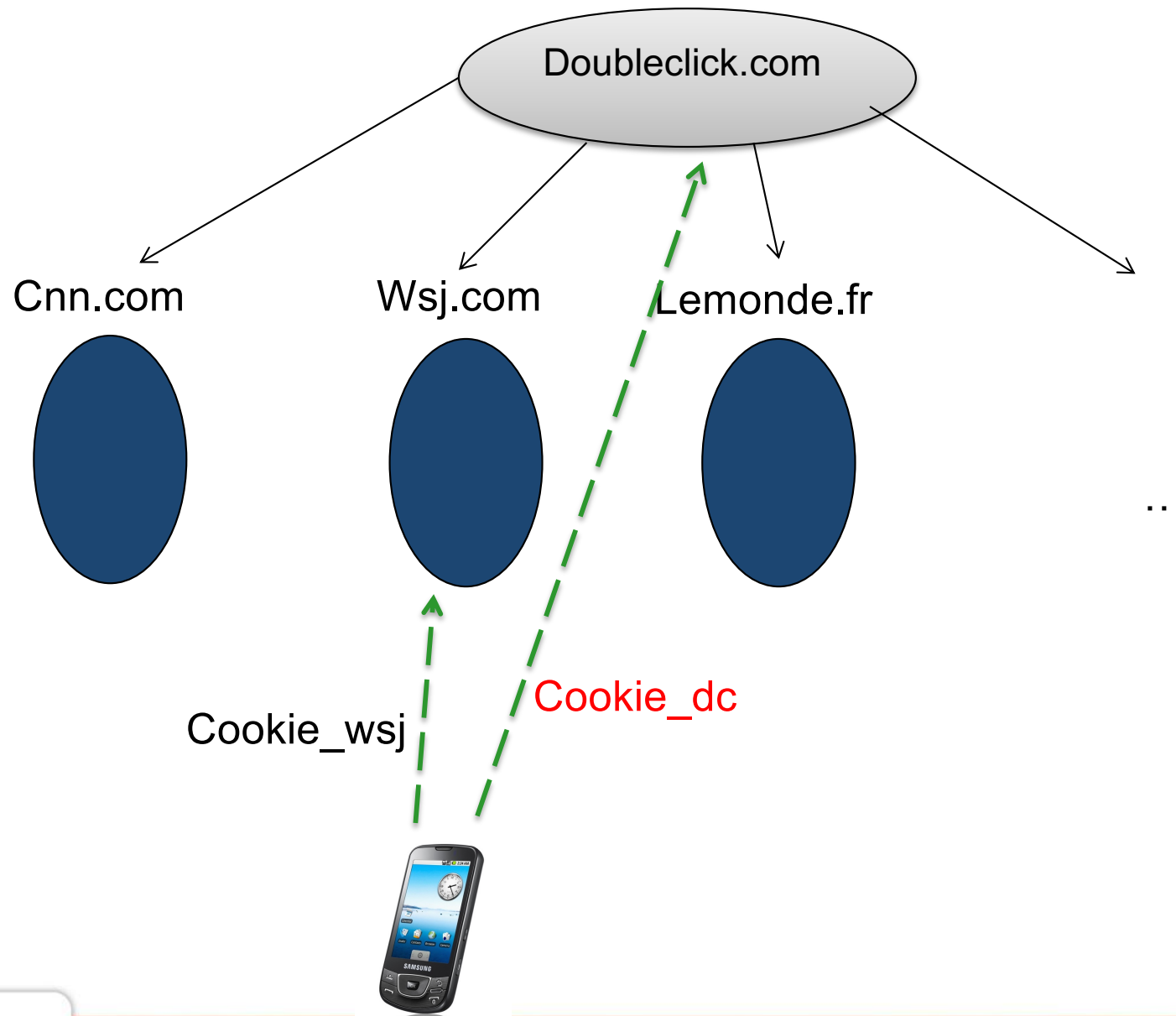
# Web Browsing Profiling: How?



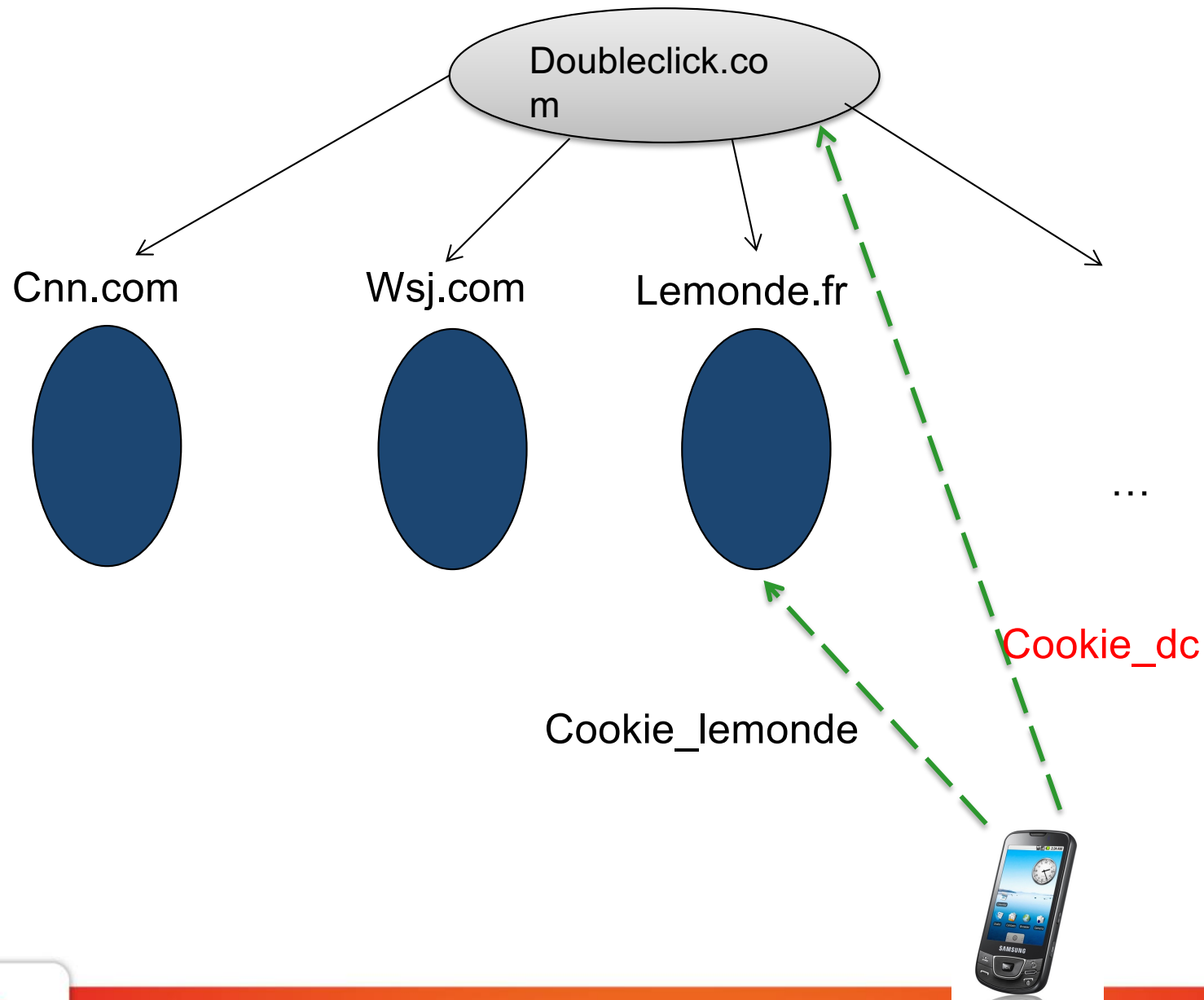
# Browsing Profiling: How?



# Browsing Profiling: How?

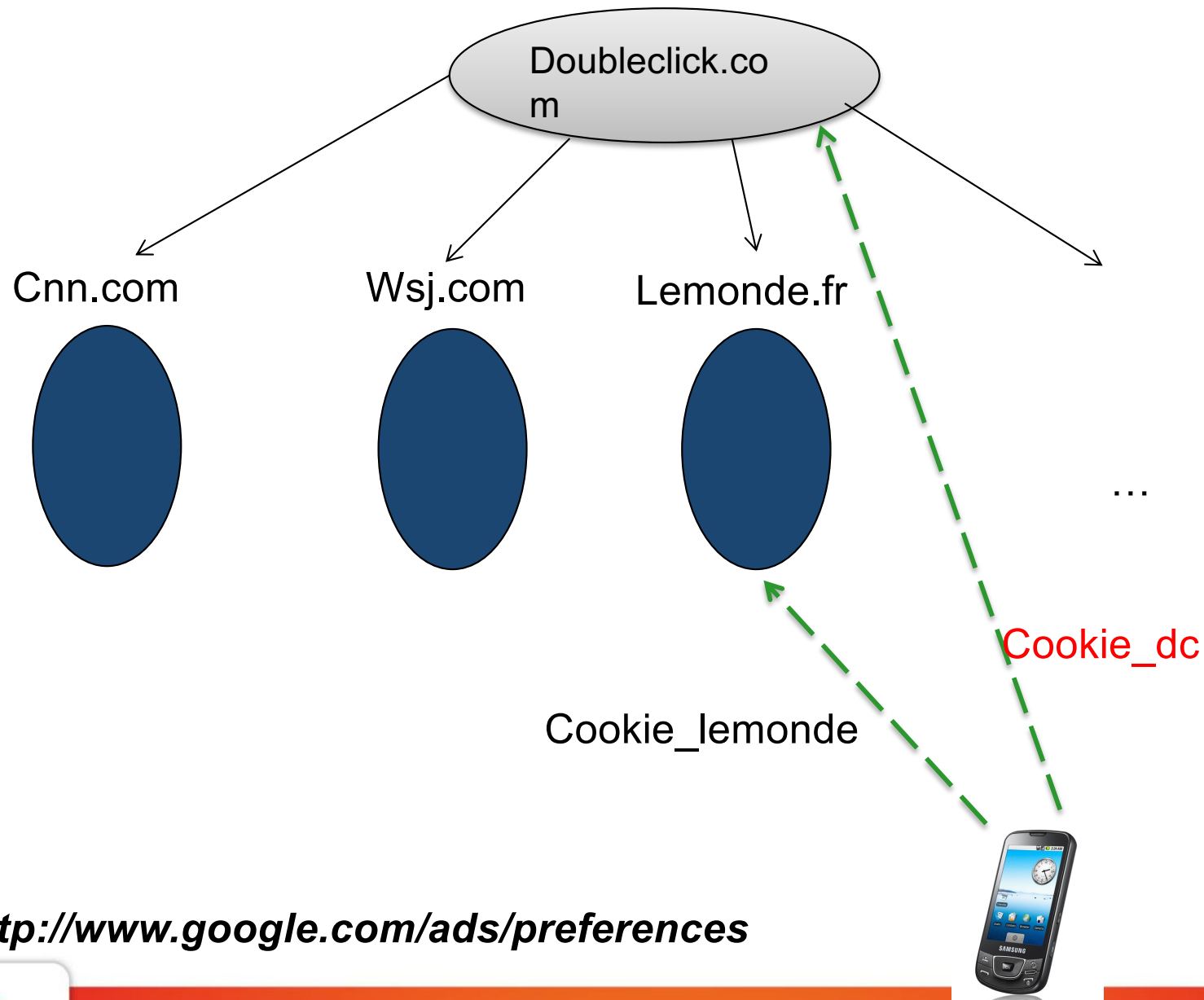


# Browsing Profiling: How?





# Browsing Profiling: How?



<http://www.google.com/ads/preferences>

# Google Profile

## Your interests

- |                                                              |                                                             |                                                        |
|--------------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------|
| <input checked="" type="checkbox"/> Action & Adventure Films | <input checked="" type="checkbox"/> Adventure Games         | <input checked="" type="checkbox"/> Air Travel         |
| <input checked="" type="checkbox"/> Arts & Entertainment     | <input checked="" type="checkbox"/> Bicycles & Accessories  | <input checked="" type="checkbox"/> Books & Literature |
| <input checked="" type="checkbox"/> Classic Rock & Oldies    | <input checked="" type="checkbox"/> Computers & Electronics | <input checked="" type="checkbox"/> East Asian Music   |
| <input checked="" type="checkbox"/> Hair Care                | <input checked="" type="checkbox"/> Hygiene & Toiletries    | <input checked="" type="checkbox"/> Internet & Telecom |
| <input checked="" type="checkbox"/> Live Comedy              | <input checked="" type="checkbox"/> Music Videos            | <input checked="" type="checkbox"/> News               |
| <input checked="" type="checkbox"/> Online Video             | <input checked="" type="checkbox"/> Politics                | <input checked="" type="checkbox"/> Rap & Hip-Hop      |
| <input checked="" type="checkbox"/> Reggaeton                | <input checked="" type="checkbox"/> Song Lyrics & Tabs      | <input checked="" type="checkbox"/> Soul & R&B         |
| <input checked="" type="checkbox"/> Sports                   | <input checked="" type="checkbox"/> TV Networks & Stations  | <input checked="" type="checkbox"/> TV Reality Shows   |
| <input checked="" type="checkbox"/> TV Talk Shows            | <input checked="" type="checkbox"/> Theme Parks             |                                                        |

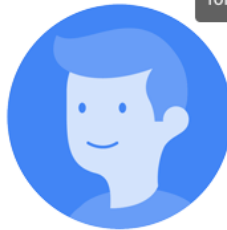
+ ADD NEW INTEREST

VIEW FEWER

WHERE DID THESE COME FROM?

These interests are derived from your activity on Google sites, such as the videos you've watched on YouTube. This does not include Gmail interests, which are used only for ads within Gmail. [Learn more](#)

## Your Google profile



Gender



Age

# Facebook Profile

## DEMOGRAPHICS

### LOCATION

Enter one or more countries, counties/regions, cities, ZIP/postal codes, addresses or designated market areas to show or exclude your ad to people in those locations. Location targeting is not available in all countries.

Everyone in this location | People recently in this location | People traveling in this location

**Note:** You can even drop a pin in a location anywhere on the map provided.

### LANGUAGES

Leave this blank unless the audience you are targeting uses a language that is not common to the location you have chosen.

Type in any language to get started

### EDUCATION

Education Level

- In high school
- High school grad
- In college
- Some college
- College grad
- Associate degree
- Professional degree
- In grad school
- Master's degree
- Doctorate degree
- Unspecified

Field of Study

Type in a field to get started

Schools

Type in a school to get started

Undergrad years

Type in a date range to get started

### FINANCIAL

### AGE

Select the minimum and maximum age of the people who will find your ad relevant.

13 | 65+

### GENDER

All | Men | Women

**Note:** Choose "All" unless you only want your ads to be shown to either men or women.

### RELATIONSHIP

Interested in:

- Men
- Women
- Men & Women
- Unspecified

Relationship Status:

- Single
- In a Relationship
- Married
- Engaged
- Civil Union
- Complicated
- Open Relationship
- Domestic Partnership
- Separated
- Divorced
- Widowed
- Unspecified

### WORK

Employers:

Type in an employer to get started

Job Title:

Type in a job title to get started

## Facebook Profile

### FINANCIAL

Income: \$30K | \$500K+

Net Worth: Liquid Assets | Total Value

### HOME

Home Type:

- Apartment
- Condo
- Multi-Family Home
- Single
- Square Footage
- Year Home Built
- Home Value
- Property Size

Home Ownership:

- First-Time Home Buyer
- Homeowners
- Renters

Household Composition:

- Family-based Households
- Grandparents
- Housemate-based Households
- New Parents
- New Teen Drivers
- Veterans in Home
- Working Women
- Young & Hip
- Young Adults in Home

### ETHNIC AFFINITY

African-American | Asian-American | Hispanic

### PARENTS

All Parents:

- New Parents
- Parents with Toddlers
- Parents with Preschoolers
- Parents with Early School-Age Children
- Parents with Preteens
- Parents with Teenagers
- Parents with Adult Children

Moms:

- Big-City Moms
- Corporate Moms
- Fit Moms
- Green Moms
- Moms of Grade School Kids
- Moms of High School Kids
- Moms of Preschool Kids
- New Moms
- Soccer Moms
- Stay-at-Home Moms
- Trendy Moms

### POLITICS [U.S.]

Liberal | Moderate | Conservative

Very Liberal | Self Reported | Very Conservative

Likely to engage in politics: conservative, liberal

### GENERATION

Baby Boomers | Generation X | Millennials

### LIFE EVENTS

Anniversary within 30 Days

Away from Family

Away from Hometown

Long-Distance Relationship

New Job

New Relationship

Newly Engaged: 3mo, 6mo, 1yr

Newlywed: 3mo, 6mo, 1yr

Recently Moved

Upcoming Birthday

Friends of [1 of the above]

### INDUSTRIES

Administrative

- Architecture & Engineering
- Arts, Entertainment, Sports & Media
- Business & Financial Operations
- Cleaning & Maintenance
- Community & Social Services
- Computer & Mathematics
- Construction & Extraction
- Education & Library
- Farming, Fishing, & Forestry
- Food Preparation & Services
- Government Employees
- Healthcare & Medical
- IT & Technical
- Installation & Repair
- Legal
- Life, Physical, & Social Science
- Management
- Military
- Nurses
- Personal Care
- Production
- Protective Service
- Retail
- Sales
- Temporary & Seasonal
- Transportation & Moving
- Veterans

### OFFICE TYPE

Home Office | Small Business | Small Office

# Facebook Profile

## INTERESTS

Reach specific audiences by looking at their interests, activities, the pages they liked and closely related topics. Combine interests to expand your ad's reach.

### BUSINESS & INDUSTRY



### ENTERTAINMENT



### FAMILY & RELATIONSHIPS



### FITNESS & WELLNESS



### FOOD & DRINK



## BEHAVIORS

Reach people based on purchase behaviors or intent, device usage and more. Some behavior data is available for U.S. audiences only.

### AUTOMOTIVE

- Motorcycle Owners
- New Vehicle Buyers [near market]
- New Vehicle Shoppers [in market, max in market]
- Purchase Type
- Used Vehicle Buyers [in market]

### CHARITABLE DONATIONS

- All Charitable Donations
- Animal Welfare
- Arts & Culture
- Cancer Causes
- Children's Causes
- Environmental & Wildlife
- Health
- Political
- Religious
- Veterans
- World Relief

### EXPATS

- Multiple Countries

### JOB ROLE

- Corporate Executives
- Financial Professionals
- Farmers

### MOBILE DEVICE USER

- All Mobile Devices by Brand
- All Mobile Devices by OS
- All Mobile Devices
- Feature Phones
- Network Connection
- New Smartphone & Tablet Owners
- Smartphone & Tablet Owners
- Smartphone Owners
- Smartphones & Tablets
- Tablet Owners

### TRAVEL

- All Frequent Travelers
- Business Travelers
- Casino Vacations
- Commuters
- Cruises
- Currently Traveling
- Family Vacations
- Frequent International Travelers
- Frequent Flyers
- Leisure Travelers
- Personal Travelers
- Returned from Trip [1 week, 2 weeks ago]
- Timeshares
- Used Travel App [2 weeks, 1 month]

### BUSINESS-TO-BUSINESS

- Seniority
- Company Size
- Industry

### DIGITAL ACTIVITIES

- Operation System Used
- Canvas Gaming
- Console Gamers
- Event Creators
- Facebook Payments
- Facebook Page Admins
- Internet Browsers Used
- Operating System Used
- Photo Uploaders
- Primary Email Domain
- Small Business Owners
- Technology Adopters [early, late]
- Unity Plugin

### FINANCIAL

- Banking
- Investments
- Spending Methods [line of credit]

### MEDIA

- Radio
- Television

### PURCHASE BEHAVIOR

- Business Purchases
- Buyer Profiles
- Clothing
- Food & Drink
- Health & Beauty
- Home & Garden
- Household Products
- Kids' Products
- Pet Products
- Purchase Habits
- Purchase Types
- Sports & Outdoors
- Stare Types
- Subscription Services
- Technology

### RESIDENTIAL PROFILES

- Length of Residence
- Likely to Move
- New Mover
- Recent Home Buyer
- Recent Mortgage Borrower

### SEASONAL & EVENTS

- Baseball
- College Football
- Cricket
- Fall Football
- Professional Football
- Rugby

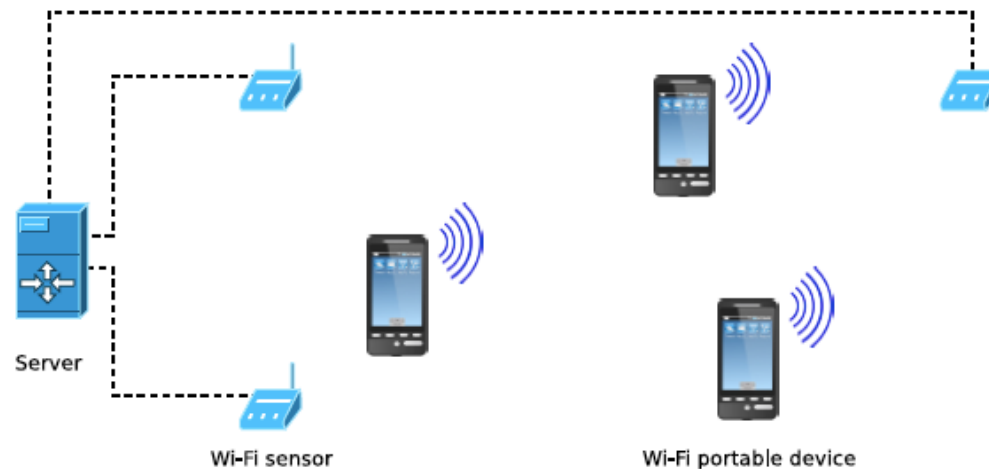


## *Facebook Profile*



# From Online to Physical Tracking

- Wi-Fi enabled smartphone: portable personal beacon
  - Broadcast a unique ID (Mac addr.) periodically to discover AP
  - Sometimes broadcast messages contained list of visited SSID
    - Leak personal information
  - Range: several 10s meters
- Perfect Monitoring tool
- Of course, GPS can also be used as well
- Things will get worse with IoT, Sensors, connected watches,...



# Physical Tracking/ Applications

- Physical analytics: Frequency and length of visit, number of visitor,..
- Profiling & Targeted advertisement
- Count (and track) people during demonstrations

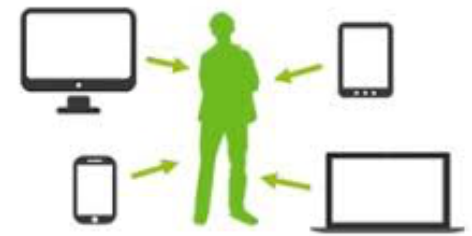


Shopping Center Monitoring

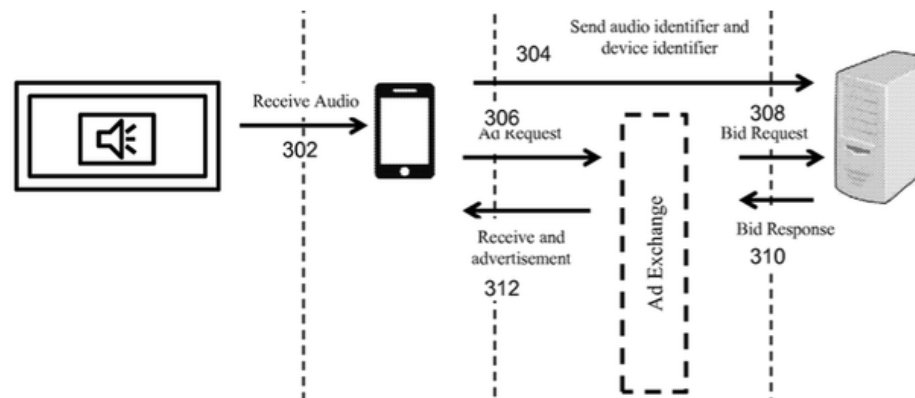


London Wifi Bins: target ads based on user profile

# Cross Device/Apps Tracking



- Industry is highly interested in tracking users across devices or across applications (to improve profiles)
- By using unique identifiers, such as MAC addresses, behavioral data and so on...
- Or using inaudible sound ☹ (Silverpush Tech.)....
  - Library that can be used by apps, web sites to pick up near-ultrasonic sound embedded in another devices (TV, web browsers,...) ... to link devices.



How it works ... the transfer of sound-encoded information from a TV to a phone to a backend server



# ***Don't Worry! Our Tracking is Anonymous!***

- Advertisers claim that the profiles they build are « **anonymous** », because they don't collect name, email
  - Btw they actually mean « pseudo-anonymous »

# What is Data Anonymization?

**Sanitization**: *process which increases the uncertainty in the data in order to preserve privacy..*

⇒ Inherent trade-off between the desired level of privacy and the utility of the sanitized data.

**Typical example**: public release of data.



Examples drawn from the “sanitization” entry on Wikipedia

# What is Data Anonymization for Computer Scientists?

*Data are anonymised if all identifying elements (all quasi-identifiers) have been eliminated from a set of personal data. No element may be left in the information which could, by exercising **reasonable** effort, serve to re-identify the person(s) concerned.*

- ☐ *Where data have been successfully anonymised, they are no longer personal data.*

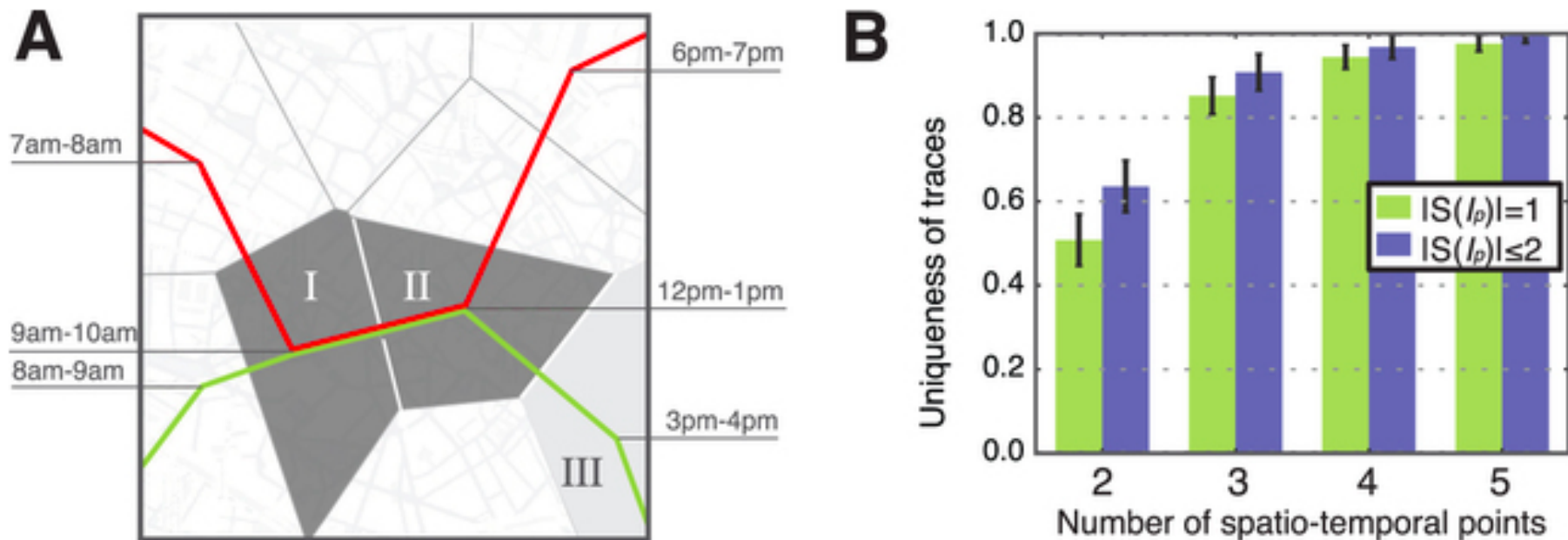
# Why is Data Anonymization Difficult?

- ❑ Quasi-identifiers are difficult to identify exhaustively
- ❑ Many combination of attributes can be used to « re-identify» a user
- ❑ We are all unique by different ways, we are full of Q.I.
  - ❑ See « Unicity me! \*»
  - ❑ Mobility pattern, webhistory, .
  - ❑ Data (content) and meta-data
    - ❑ i.e. timing can betray you!
    - ❑ Google search timing pattern can tell when you were away!

\*Unicity Me! American Scientific,

<http://www.americanscientist.org/libraries/documents/20142614253010209-2014->

# Unique in the Crowd [Nature2013]

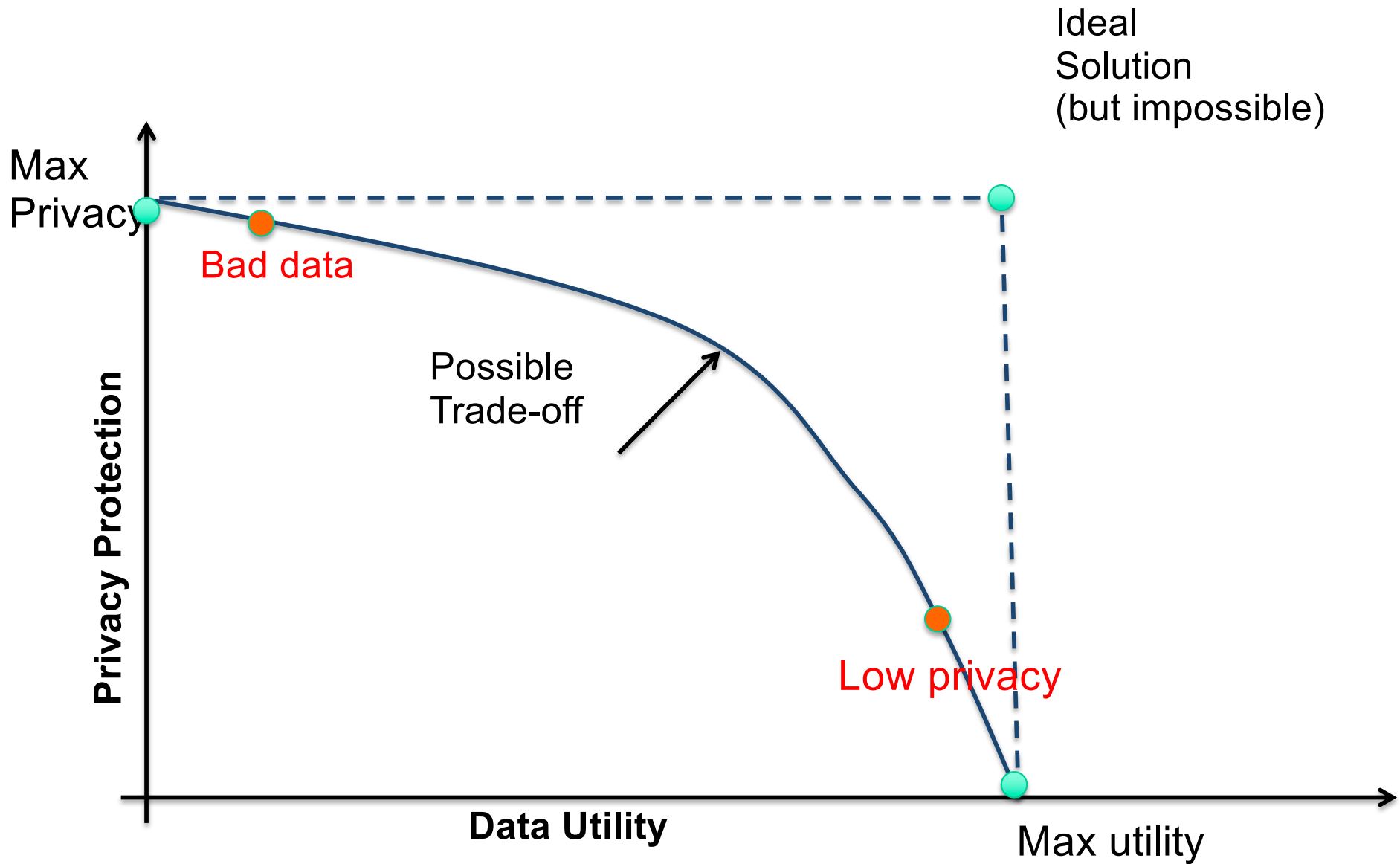


- Only 4 spatio-temporal points are necessary to uniquely identify a user with a probability  $> 95\%$  !

# Why is Data Anonymization Difficult?

- Anonymisation is a utility/privacy optimization
  - No generic solution that optimizes utility and privacy!
- Anonymisation should be performed case by case.. According to:
  - Type of data
  - Sensitivity of data
  - Type of release
  - Adversary models
  - ....
- Risk-based approach....

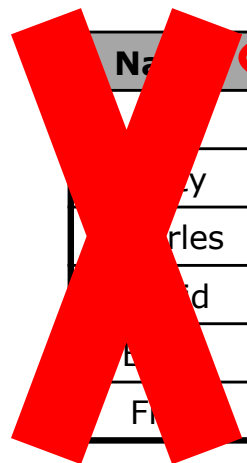
# Privacy vs Utility Tradeoff



# Pseudo-Anonymization

- ❑ What is Pseudo-Anonymization?
- ❑ *Personal information contains identifiers, such as a name, date of birth, sex and address. When personal information is pseudonymised, the identifiers are replaced by one pseudonym. Pseudonymisation is achieved, for instance, by encryption or by hashing of the identifiers in personal data.*

Microdata



Name	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



# Pseudo-Anonymization

- ❑ Why is Pseudo-Anonymization not good Enough?
  - ❑ It does not compose, i.e. several Pseudo-Anonymized data can be combined to de-anonymize...
  - ❑ External Information can also be exploited.
- ❑ Very weak protection...
  - ❑ Could be used as a security measure
  - ❑ But pseudo-anonymized data are still personal data!
    - ❑ See GDPR article
- ❑ We need schemes that also alter the quasi-identifiers (not only the identifiers)
  - ❑ K-anonymity
  - ❑ Differential Privacy
  - ❑ ...

# De-Identification: K-anonymity

- **Privacy guarantee**: in each group of the sanitized dataset, each individual will be identical to a least  $k - 1$  others.
- Reached by a combination of generalization and suppression.
- **Example of use**: sanitization of medical data.

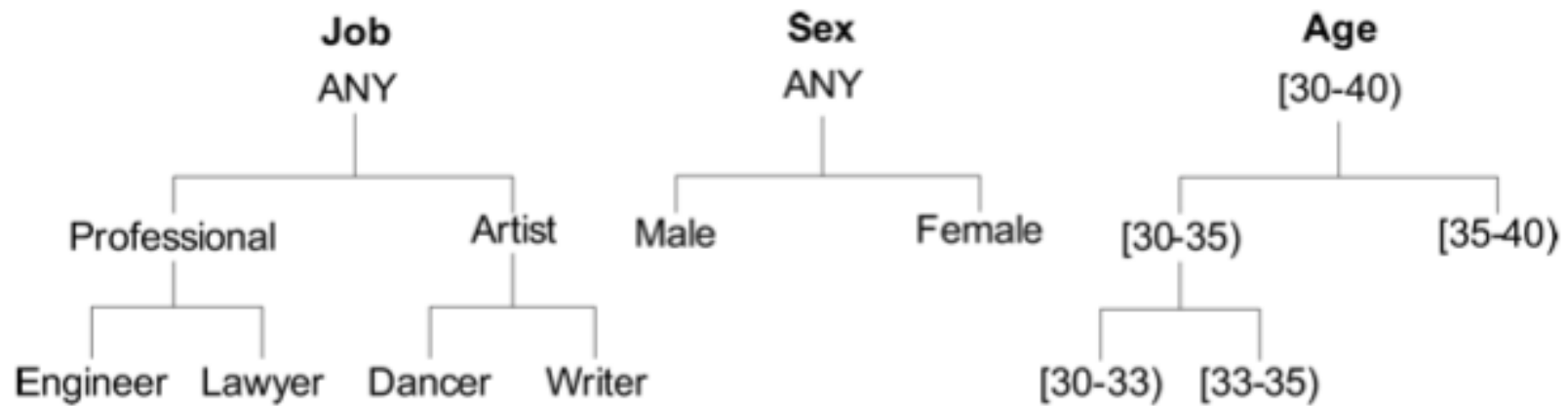
	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

Figure 1. Inpatient Microdata

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	$\geq 40$	*	Cancer
6	1485*	$\geq 40$	*	Heart Disease
7	1485*	$\geq 40$	*	Viral Infection
8	1485*	$\geq 40$	*	Viral Infection
9	130**	3+	*	Cancer
10	130**	3+	*	Cancer
11	130**	3+	*	Cancer
12	130**	3+	*	Cancer

Figure 2. 4-anonymous Inpatient Microdata

# Generalization Methods



## But K-Ano. does not compose ☹!

- **Question**: suppose that Alice's employer knows that she is 28 years old, she lives in ZIP code 13012 and she visits both hospitals. What does he learn?

	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	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer


(a)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

(b)


## But K-ANO does not compose ☹!

- **Question**: suppose that Alice's employer knows that she is 28 years old, she lives in ZIP code 13012 and she visits both hospitals. What does he learn?



	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	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)



	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

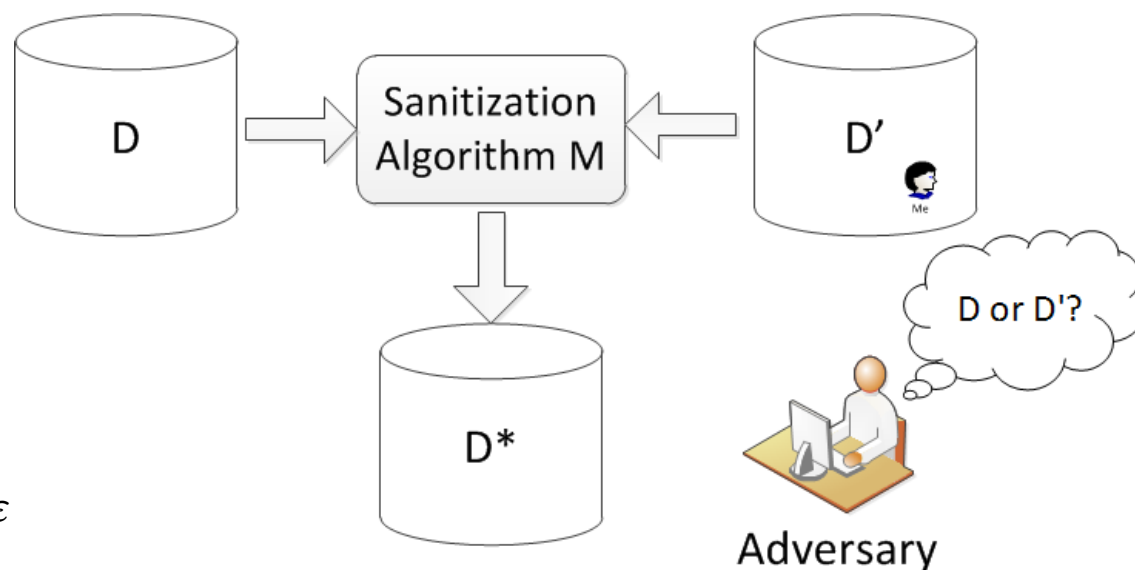
(b)

# Toward « Provable » Anonymization

- ❑ Stronger schemes are necessary
- ❑ Differential Privacy (DP)
  - ❑ Provides some strong and measurable guarantees
  - ❑ Secures even with external sources of data
  - ❑ Composes
- ❑ Intuition of DP:
  - ❑ Changes to my data not noticeable
  - ❑ Output is “independent” of my data

# Privacy Model

- Differential privacy

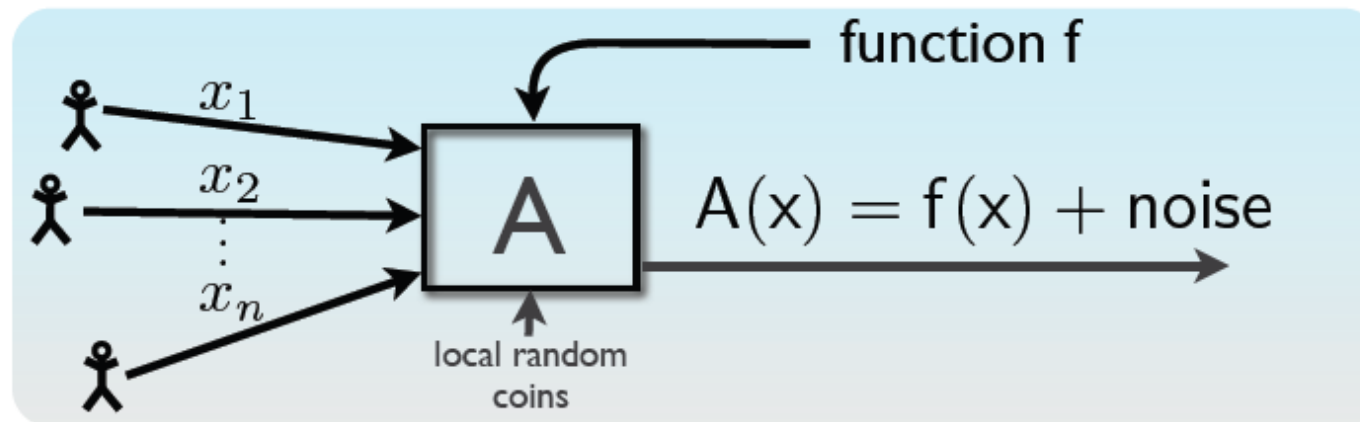


$$e^{-\epsilon} \leq \frac{\Pr(M(D) = D^*)}{\Pr(M(D') = D^*)} \leq e^{\epsilon}$$

- composes securely**: retain privacy guarantees in the presence of independent releases<sup>[1]</sup>
- Secure even with arbitrary external knowledge!
- [1] S.R. Ganta, S. Kasiviswanathan, A. Smith. *Composition Attacks and Auxiliary Information in Data Privacy*. KDD'08



# Differential Privacy



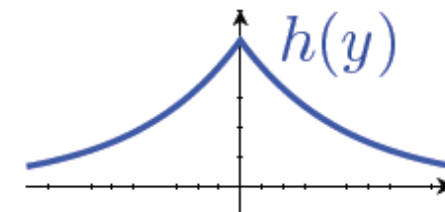
- **Global Sensitivity:**  $GS_f = \max_{\text{neighbors } x, x'} \|f(x) - f(x')\|_1$

➤ Example:  $GS_{\text{proportion}} = \frac{1}{n}$

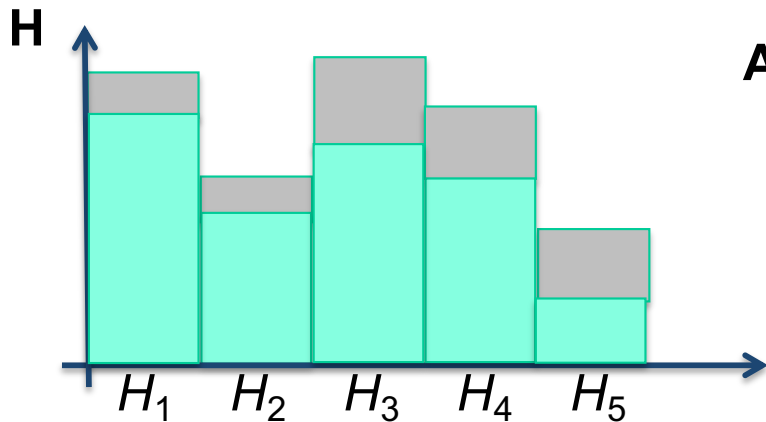
**Theorem:** If  $A(x) = f(x) + \text{Lap}\left(\frac{GS_f}{\epsilon}\right)$ , then  $A$  is  $\epsilon$ -differentially private.

➤ Laplace distribution  $\text{Lap}(\lambda)$  has density

$$h(y) \propto e^{-|y|/\lambda}$$



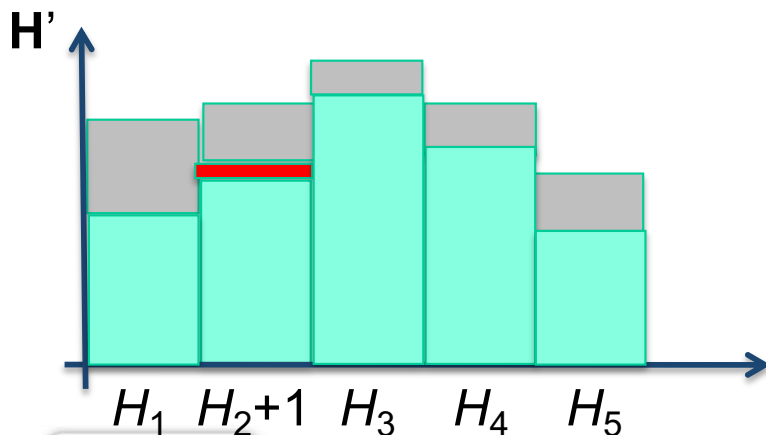
# Histogram Release with Laplace Mechanism



Add random Laplace noise to each bin before publishing!



$$\frac{\prod_i \Pr(H_i + \text{Laplace}(\lambda) = H_i^*)}{\prod_i \Pr(H'_i + \text{Laplace}(\lambda) = H_i^*)} \leq \exp\left(\frac{\sum_i |H_i - H'_i|}{\lambda}\right) = e^{\frac{1}{\lambda}}$$

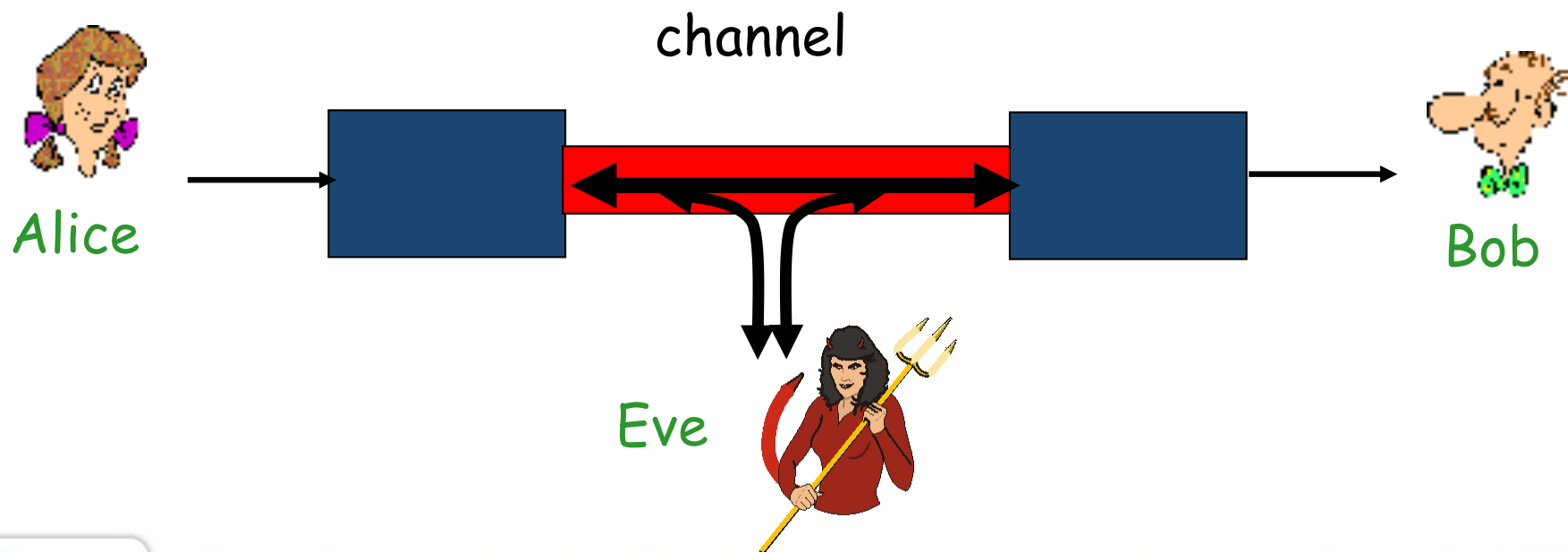


- **Global sensitivity:**  
 $\Delta H = \sum |H_i - H'_i|$
- For histograms:  $\Delta H = 1$
- If  $\lambda = \Delta H / \epsilon$ , we have  $\epsilon$ -differential privacy

# Why not using Cryptography?

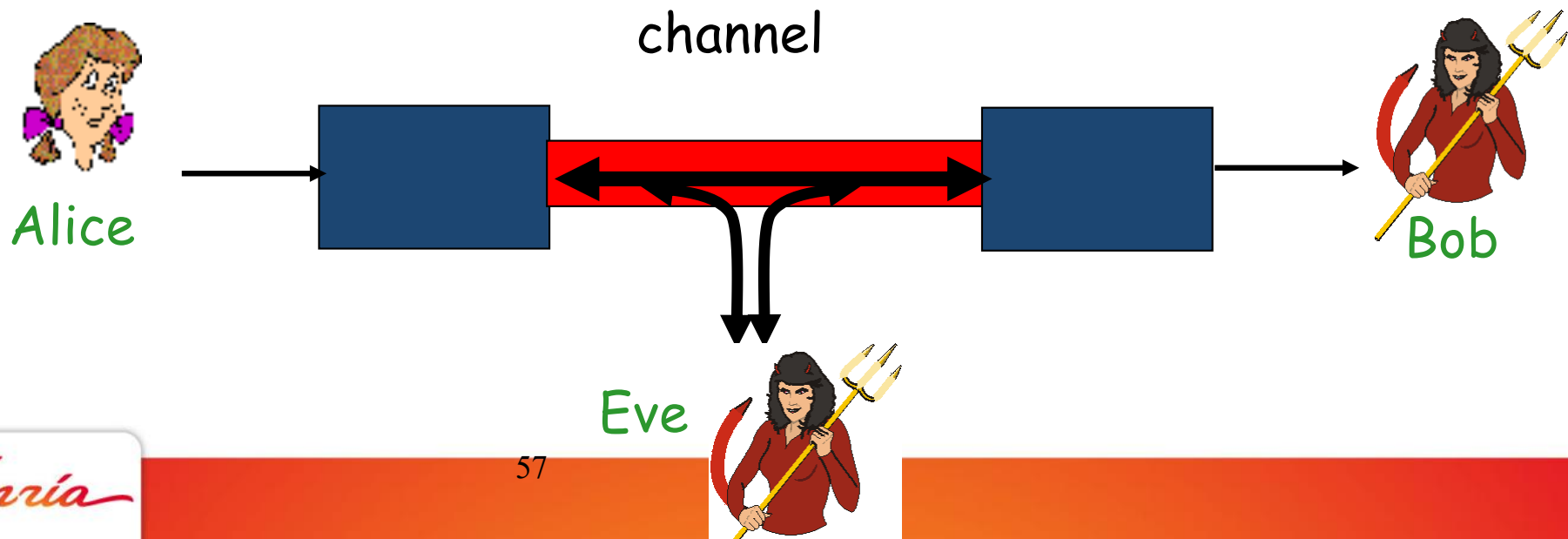
# Why not using Cryptography?

- ❑ The Trust models are different!
- ❑ In cryptography, sender and receiver trust each others:
  - ❑ Alice sends a dataset to Bob
  - ❑ Alice encrypts to protect from Eve, the eavesdropper
  - ❑ But Bob can decrypt and recover the original dataset!
  - ❑ The **adversary is Eve!**



# Cryptography and Anonymization

- With Data anonymization, the sender does not trust the receiver
  - ❑ Alice anonymized a dataset to “hide” some (usually personal) information and sends it to Bob (possibly after encryption).
  - ❑ Bob recovers the anonymized dataset. It can process it to compute some statistics/inferences...but can't recover the hidden information (identity or attribute).
  - ❑ Bob is also the adversary!



# Data Anonymisation Exercise

( <https://team.inria.fr/privatics/claude-castelluccia/skema-data-anonymization-exercices/> )

- A subset of the UCI Adult dataset was anonymised with k-anonymity using the [ARX anonymisation tool](#) and different k values (k=5, 10, 20, 50, 100, 500). All the files (README, original and anonymized datasets are available here).
- The goal of these exercises is to manipulate anonymized datasets, understand some of their limitations and practice Python coding.

# Data Anonymisation Exercice

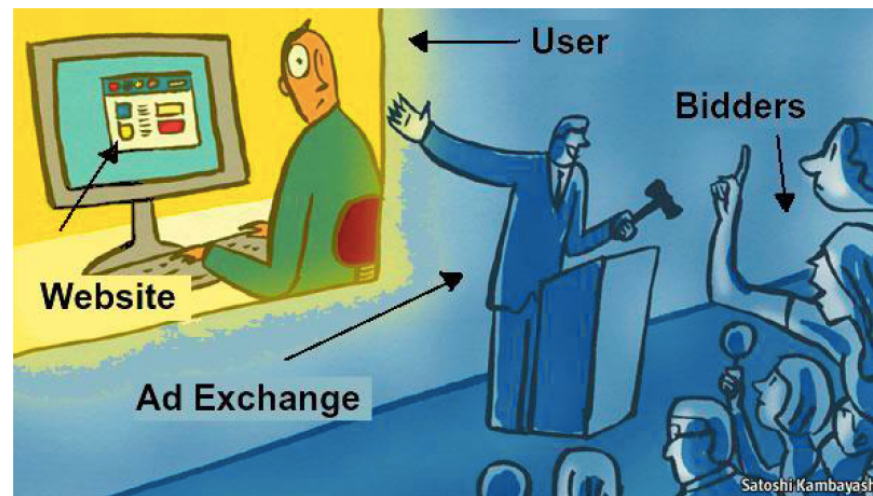
( <https://team.inria.fr/privatics/claude-castelluccia/skema-data-anonymization-exercices/> )

- Compute the unicity level of each record (i.e. how many records are unique, how many of them appear 2 times, ..., how many of them appear k times) in the original dataset. Display results as an histograms (one per anonymized dataset). What do you conclude?
- Compute the unicity level of each record in the anonymised datasets. Display the results as histograms (one per anonymized dataset). What do you conclude about the quality of the anonymization process?
- Predict the salary\_class of these people using the different anonymized datasets? The prediction is performed by computing that number of records that correspond to the salary\_class classes  $\leq 50K$  and  $< 50k$ . Compute the probability for each of these classes for the following queries.



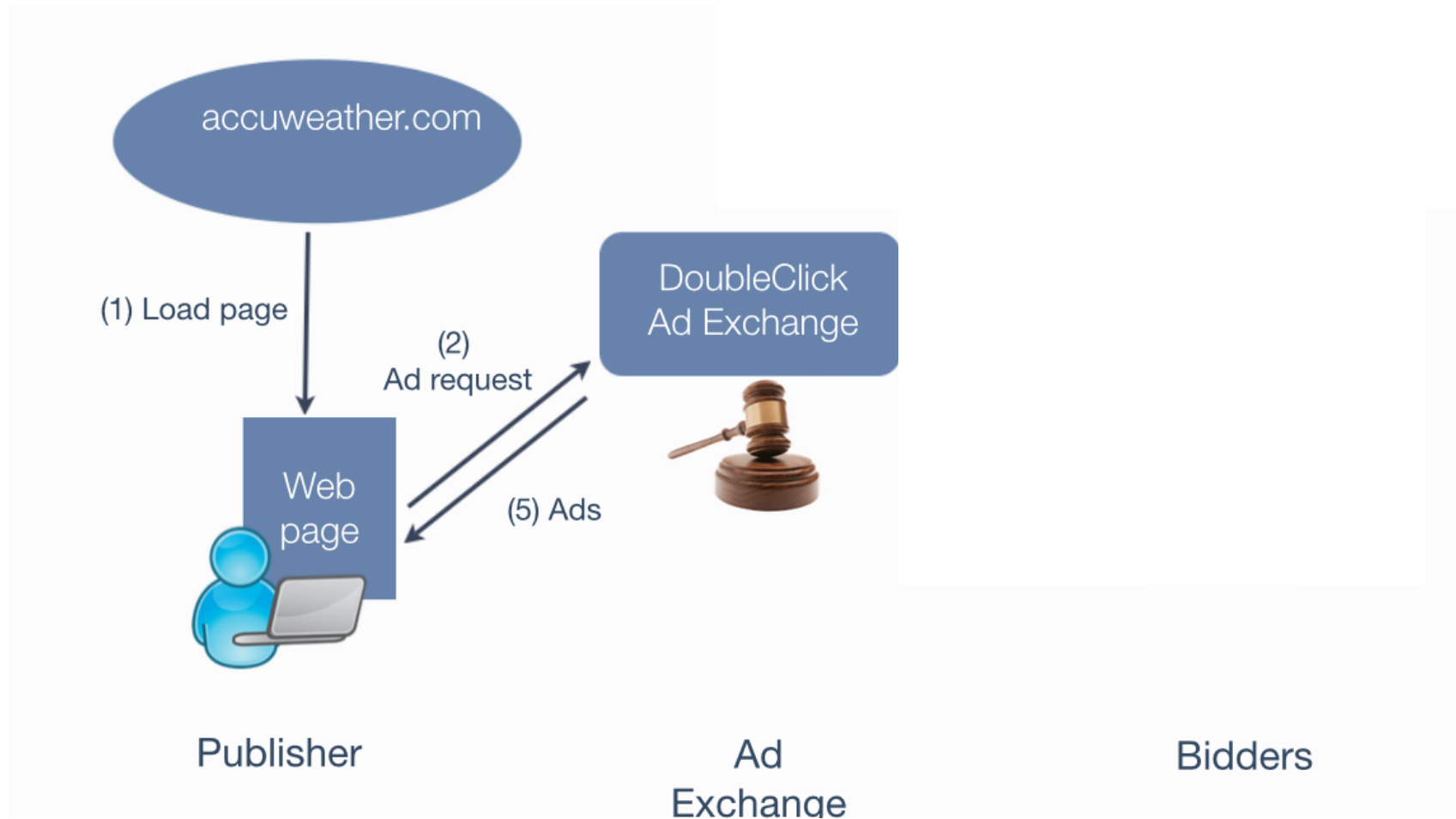
## ***Dataveillance 2.0***

- Private user data:
  - resource just as crude oil or iron that fuel digital economies.
  - It is now sold (off) in via auctions as any other products!...
- ***Real Time Bidding***: a means by which ads (and **our profile**) is bought and sold via programmatic instantaneous auction, similar to **financial markets**
- RTB is rapidly growing, expectedly accounting for:
  - 27% of total display advertising sales in the US, 25% in EU

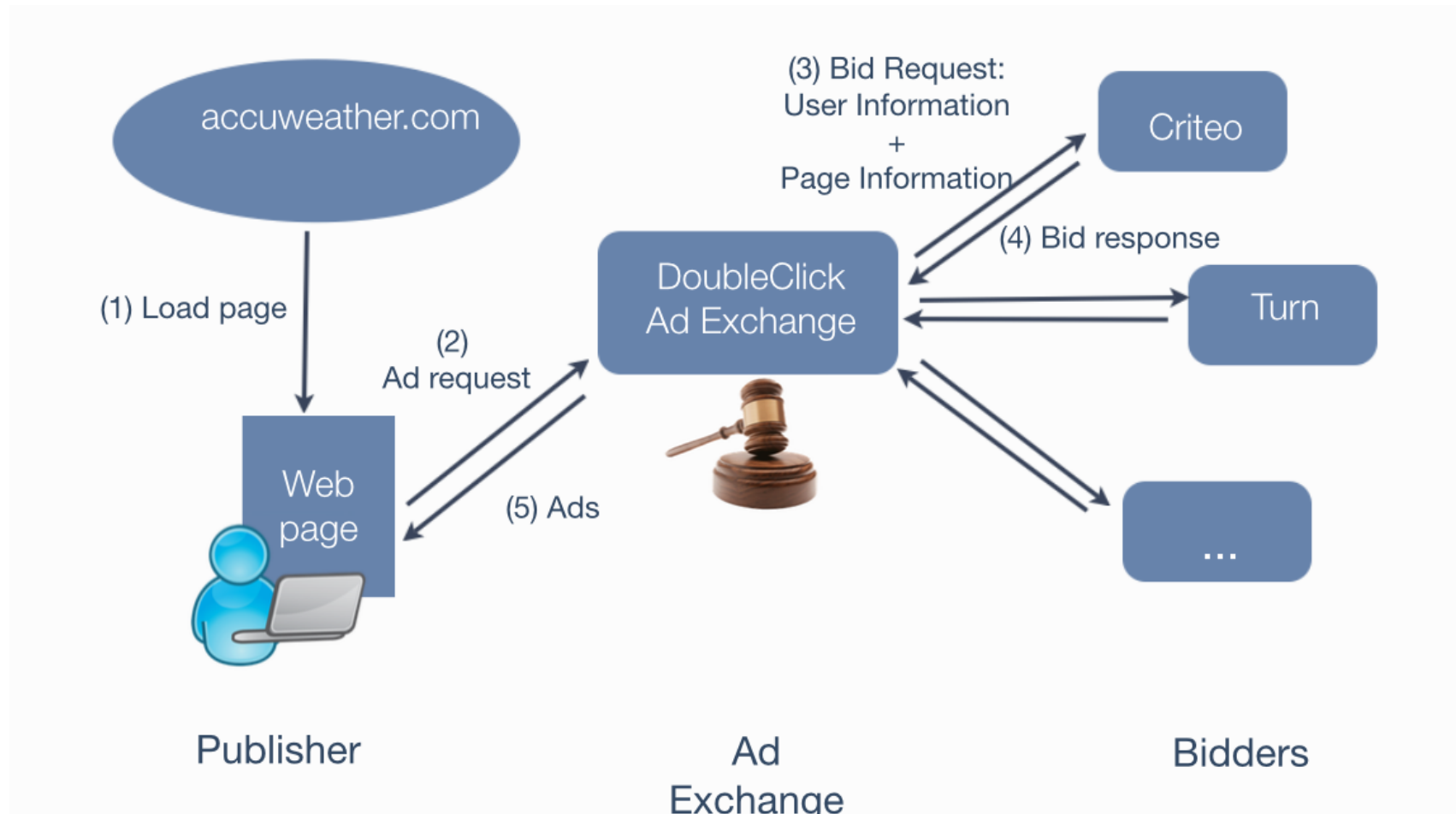


Satoshi Kambayashi

# ***RTB (Real Time Bidding): How does it work?***



# ***RTB (Real Time Bidding): How does it work?***





## Personal data in bid requests



- What you are reading, or watching, or listening to.
- Categories of the content.
- Unique pseudonymous ID.
- Unique ID matched to ad buyer's existing profile of you.
- Your location (can be your exact latitude and longitude).
- Granular description of your device.
- Unique tracking IDs / cookie match.
- Your IP address.\*
- Data broker segment ID\* when available.

\*Depending on the version of “real time bidding” system

**Source:** Johnny Ryan – BRAVE – Feb. 2019

# HUNDREDS OF BILLIONS OF RTB BID REQUESTS, EVERY DAY.

## Leading RTB exchanges, daily bid request estimates

Index Exchange	50 billion. <sup>ii</sup>
OpenX	60 billion+. <sup>i</sup>
Rubicon Project	Unknown. Claims to reach 1 billion people's devices. <sup>iii</sup>
Oath/AOL	90 billion. <sup>iv</sup>
AppNexus	131 billion. <sup>vi</sup>
Smaato	214 billion. <sup>v</sup>
Google DoubleClick	Unknown. DoubleClick is the dominant exchange.

i. "OpenX Ad Exchange", OpenX (URL: [https://www.openx.com/uk\\_en/products/ad-exchange/](https://www.openx.com/uk_en/products/ad-exchange/)).

ii. "Tour IX's Amsterdam and Frankfurt Data Centers", Index Exchange, 2 July 2018 (URL: <https://www.indexexchange.com/tour-ix-amsterdam-frankfurt-data-centers/>).

iii. "Buyers", Rubicon Project, (URL: <https://rubiconproject.com/buyers/>).

iv. "Maximize yield with Oath's publisher offerings", Oath, 3 April 2018 (URL: <https://www.oath.com/insights/maximize-yield-with-oath-s-publisher-offerings/>).

v. 500 Billion / 29.6 = 18.6 billion impressions per day. Using AppNexus 1:11.5 ratio, this is 214 auctions per day. 500+ impressions figure cited in "Optimize your mobile strategy", Smaato, (URL: <https://www.smaato.com/>).

vi. "Transacting at a peak of 11.4 billion daily impressions, our marketplace handles more traffic each day than Visa, Nasdaq, and the NYSE combined" at <https://www.appnexus.com/sell>. Note that in 2017, AppNexus said in "AppNexus Scales with DriveScale", 2017, (URL: [http://go.drivescale.com/rs/451-ESR-800/images/DRV\\_Case\\_Study\\_AppNexus-final.v1.pdf](http://go.drivescale.com/rs/451-ESR-800/images/DRV_Case_Study_AppNexus-final.v1.pdf)) that 10.7 billion "impressions transacted" came as a result of running 123 billion auctions. The impressions transacted to auctions ratio appears to be roughly 1:11.5. Therefore, the 11.4 daily impressions reported in 2018 equates to 131 billion auctions per day.

**Source:** Johnny Ryan – BRAVE – Feb. 2019

# The Dangers

- Profiles reveal user's interests (religion, health,...) and is therefore very sensitive!
- **Surveillance :**
  - We move into a surveillance society companies/gov. gather a huge amount of information about users with all associated risks.
- **Discrimination/Sorting:**
  - Profiling may reveal that a user is suffering from a certain disease.
  - Insurance might then deny insurance
- **Personalization:**
  - Filter bubble
  - Manipulation**



# Danger #1: Gouvernemental Surveillance

- July 2013: Snowden reported that:
  - NSA had collected phone records from over 120 million Verizon subscribers.
  - That Xkeyscore allows analysts to search with no prior authorization through vast databases containing emails, online chats and the browsing histories of millions of individuals.
  - And then Prism, Upstream and many more!

*[http://en.wikipedia.org/wiki/Global\\_surveillance\\_disclosure](http://en.wikipedia.org/wiki/Global_surveillance_disclosure)*





facebook



Hotmail®

YAHOO!



(TS//SI//NF) PRISM Collection Details



### Current Providers

- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple



### What Will You Receive in Collection (Surveillance and Stored Comms)?

It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:  
Go PRISMFAA



facebook



Hotmail®

YAHOO!



paltalk.com

YouTube

AOL mail



(TS//SI//NF) **FAA702 Operations**  
*Two Types of Collection*



## Upstream

- Collection of communications on fiber cables and infrastructure as data flows past.  
(FAIRVIEW, STORMBREW, BLARNEY, OAKSTAR)

**You  
Should  
Use Both**

## PRISM

- Collection directly from the servers of these U.S. Service Providers: Microsoft, Yahoo, Google, Facebook, PalTalk, AOL, Skype, YouTube, Apple.

## The Danger#2: Sorting



- The Chinese government is launching a Social Credit System.
- The main idea is to assign a ***citizen score*** to each Chinese, from their daily activities, what they buy at the shops and online, where they are at any given time, who their friends are and how you interact with them, how many hours they spend watching content or playing video games, and what bills and taxes they pay (or not)...
- This score would tell everyone whether or not you were **trustworthy**.
  - used to determine your eligibility for a mortgage or a job, where your children can go to school - or even just your chances of getting a date.
  - will influence a person's rental applications, their ability to get insurance, a loan, a job and even social-security benefit

<http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>



# CHINA'S SOCIAL CREDIT SYSTEM

It's been dubbed the most ambitious experiment in digital social control ever undertaken. The Chinese government plans to launch its Social Credit System nationally by 2020.

## WHAT'S THE AIM?

The system intends to monitor, rate and regulate the financial, social, moral and, possibly, political behavior of China's citizens - and also the country's companies - via a system of punishments and rewards. The stated aim is to "provide the trustworthy with benefits and discipline the untrustworthy."

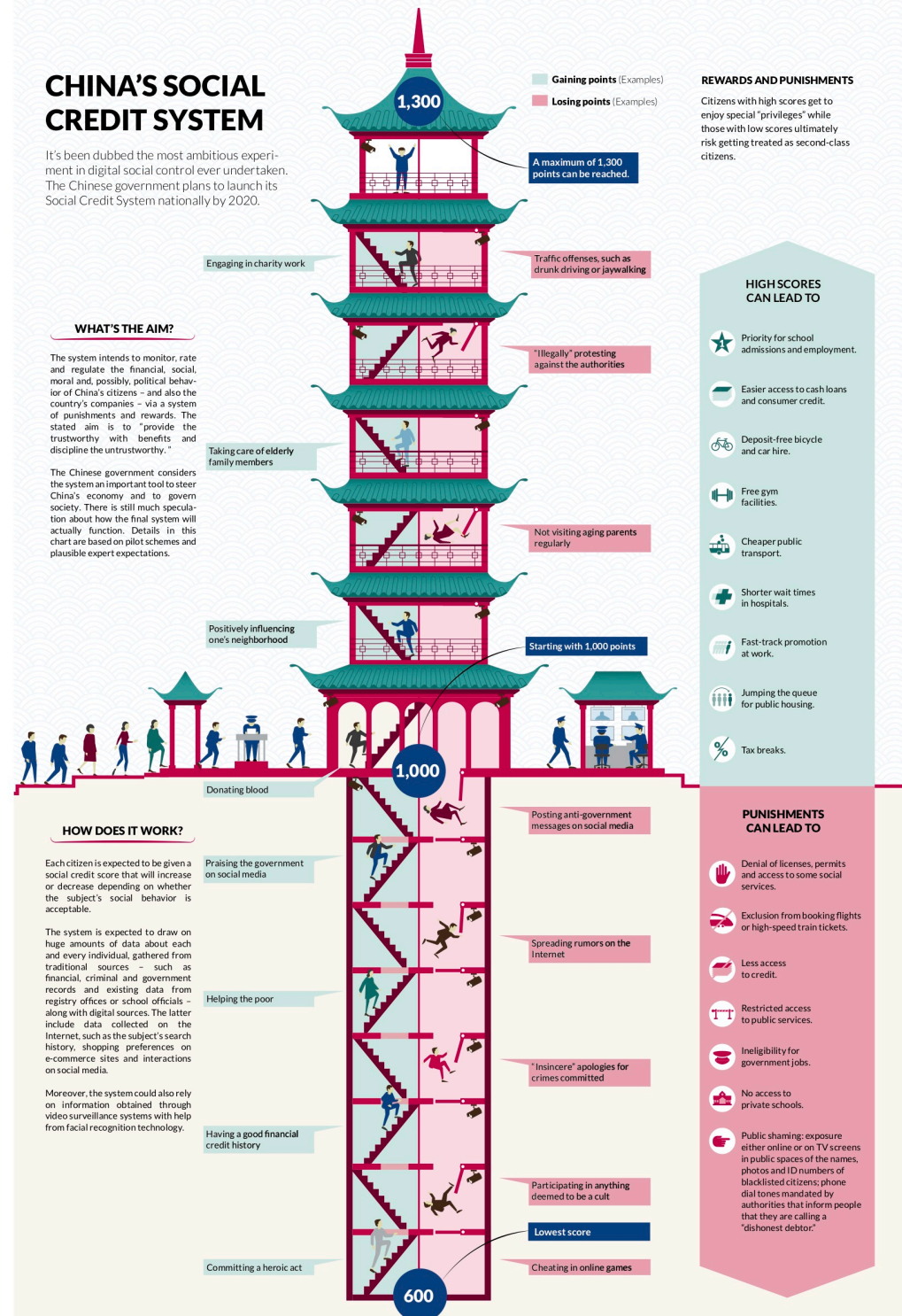
The Chinese government considers the system an important tool to steer China's economy and to govern society. There is still much speculation about how the final system will actually function. Details in this chart are based on pilot schemes and plausible expert expectations.

## HOW DOES IT WORK?

Each citizen is expected to be given a social credit score that will increase or decrease depending on whether the subject's social behavior is acceptable.

The system is expected to draw on huge amounts of data about each and every individual, gathered from traditional sources - such as financial, criminal and government records and existing data from registry offices or school officials - along with digital sources. The latter include data collected on the Internet, such as the subject's search history, shopping preferences on e-commerce sites and interactions on social media.

Moreover, the system could also rely on information obtained through video surveillance systems with help from facial recognition technology.



## The Danger#2: Sorting... some remarks

- Is the situation much better in Europe/US?
- We are constantly scored by various services
  - data brokers such as Experian trace the timely manner in which we pay our debts, giving us a score that's used by lenders and mortgage providers.
  - eBay has a rating on shipping times and communication
  - Uber drivers and passengers both rate each other; if your score falls too far, you're out of luck.
  - And more: AirBnb, BlablaCar...
- It is all about “**building trust**” which is useful for business
  - But what are the dangers?
- China's social credit system expands that idea to all aspects of life, judging citizens' behaviour and trustworthiness

## The Danger #3: (Online) Manipulation by Data

- Data are more and more used to manipulate people!
  - Fake news...
- Data can be used to manipulate:
  - Emotion
  - Memory
  - Attention
  - Perception
  - ...
- The final goal is often to manipulate Decisions (for example during an election)

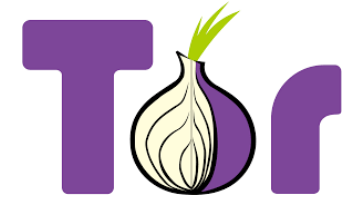
# Datapulation. What is New?

- **Scale:** Large number of users can be manipulated.
- **Personalized:** Can be highly targeted/personal.
  - 10,000 different ads to different audiences in the months leading up to Trump election
- **Adaptive:** Can be adaptive. Possibility of testing and improving.
- **Efficient and automated:** algorithm-based.
- **Hidden:** Can be Surreptitiously.
- **Affordable:** Can be performed remotely and in a distributed manner. It is affordable, not reserved to governments (low budget required).

**So can we protect ourselves?**



# What Can be Done Against Data Surveillance (short term)?



- Use **Self-Defense Surveillance** tools
- A (growing) list from EFF (Electronic Frontier Foundation)
  - Encrypt your emails with **PGP**
  - User Tracker blocker (**Privacy Badgers**)
  - Use **TOR** (an anonymisation network) when necessary to protect your metadata
  - Use OTR/**Signal/Ricochet** for secure messaging (end-to-end encryption)
  - Generate strong password and use a password manager (KeePassX)...



# ***What Can be Done Against Data Surveillance (longer term)?***

- **Education, information**
- **More transparency**
  - Of collected Data
  - Of targeting and Decision systems
  - (Political) Ads has to be public - <https://whotargets.me/en/>
- **Better Legal Protections** (see GDPR and ePrivacy regulations)
- ***Privacy-preserving*** systems that improve:
  - **Transparency + Accountability**
  - **User Control** (*MyTrackerChoices, YourRealOnlineChoice*)
  - **Privacy by Design**

## TERRITORIAL SCOPE

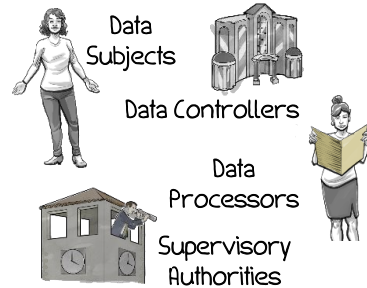


EU Establishments

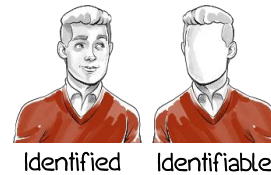
Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

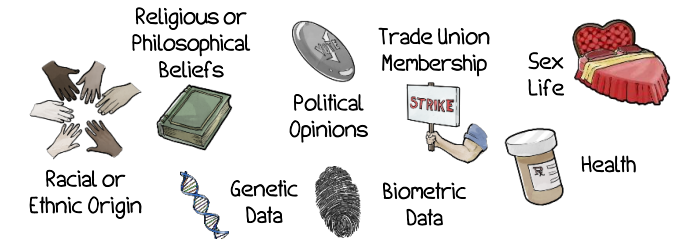
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.

Data Protection by Design

built in starting at the beginning of the design process

Data Impact Assessment

For high risk situations

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests



## CONSENT



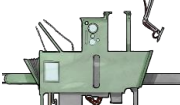
Consent must be freely given, specific, informed, and unambiguous.

## RIGHTS OF DATA SUBJECTS



Transparency

Automated Decision Making



"Right not to be subject to a decision based solely on automated processing, including profiling."

Access and Rectification



Right to Erasure



Purpose Specification and Minimization



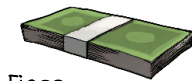
Right to Data Portability



## ENFORCEMENT

Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies:

compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



Privacy Shield



Model Contractual Clauses

## INTERNATIONAL DATA TRANSFER



Adequate Level of Data Protection

If likely to result in a high privacy risk → notify data subjects

Notify supervisory authorities no later than 72 hours after discovery.

TEACHPRIVACY

[www.teachprivacy.com](http://www.teachprivacy.com)

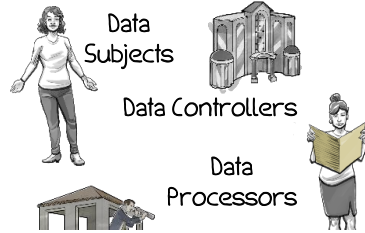
Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute

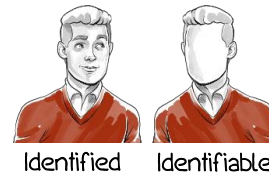
## TERRITORIAL SCOPE



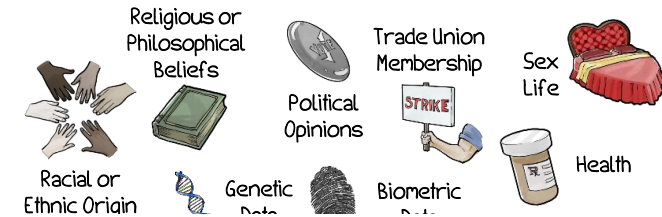
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



# General Data Protection Regulation

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for:

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests



## CONSENT



Consent must be freely given, specific, informed, and unambiguous.



Security

## Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

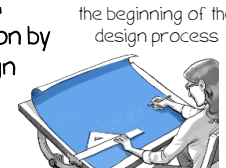
## Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.



## Protection by Design

Data Impact Assessment  
For high risk situations



## DATA BREACH NOTIFICATION



A personal data breach is "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed."

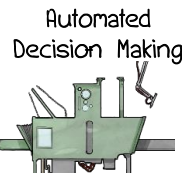
If likely to result in a high privacy risk → notify data subjects

Notify supervisory authorities no later than 72 hours after discovery.

## RIGHTS OF DATA SUBJECTS



Transparency



Automated Decision Making

"Right not to be subject to a decision based solely on automated processing, including profiling."



Access and Rectification

Right to Erasure

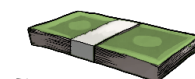


Purpose Specification and Minimization

Right to Data Portability



## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies: compensation for material and non-material harm.



## INTERNATIONAL DATA TRANSFER



Adequate Level of Data Protection

Binding Corporate Rules (BCRs)



Privacy Shield



Model Contractual Clauses

TEACHPRIVACY

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute



## TERRITORIAL SCOPE



EU Establishments

Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

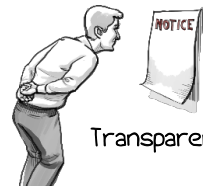
## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

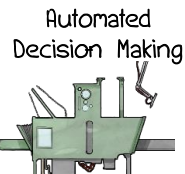
- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests



## RIGHTS OF DATA SUBJECTS



Transparency



Automated Decision Making



Access and Rectification



Purpose Specification and Minimization

Right to Erasure



"Right not to be subject to a decision based solely on automated processing, including profiling."

Right to Data Portability



## CONSENT



Consent must be freely given, specific, informed, and unambiguous.

## THE PLAYERS



Data Subjects

Data Controllers



Data Processors



Supervisory Authorities



## PERSONAL DATA



Identified

Identifiable



Religious or Philosophical Beliefs

Racial or Ethnic Origin



Genetic Data



Political Opinions

Trade Union Membership



Sex Life



Health

Biometric Data



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.



Data Impact Assessment

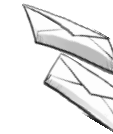
For high risk situations

Data Protection by Design

built in starting at the beginning of the design process



## DATA BREACH NOTIFICATION

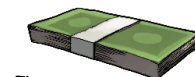


A personal data breach is "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed."

If likely to result in a high privacy risk → notify data subjects

Notify supervisory authorities no later than 72 hours after discovery.

## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies:

compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



Adequate Level of Data Protection



Privacy Shield



Model Contractual Clauses

TEACHPRIVACY

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute

## TERRITORIAL SCOPE

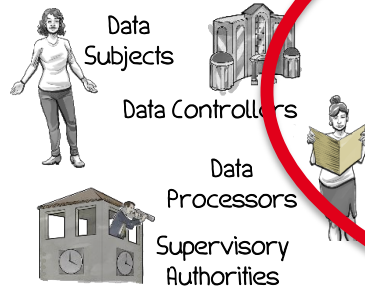


EU Establishments

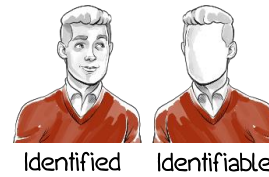
Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

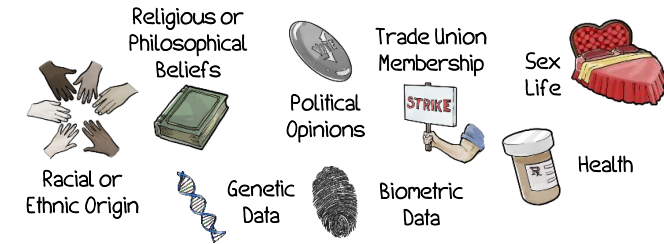
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.

Data Protection by Design

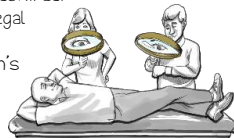
built in starting at the beginning of the design process

Data Impact Assessment  
For high risk situations

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests

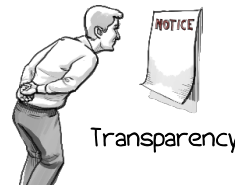


## CONSENT



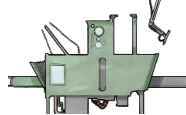
Consent must be freely given, specific, informed, and unambiguous.

## RIGHTS OF DATA SUBJECTS



Transparency

Automated Decision Making



"Right not to be subject to a decision based solely on automated processing, including profiling."



Access and Rectification

Right to Erasure



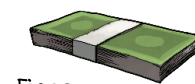
Purpose Specification and Minimization



Right to Data Portability



## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies: compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



Adequate Level of Data Protection



Privacy Shield



Model Contractual Clauses

**TEACHPRIVACY**

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute

## TERRITORIAL SCOPE



EU Establishments

Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests



## RIGHTS OF DATA SUBJECTS



Transparency



Access and Rectification



Purpose Specification and Minimization

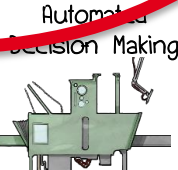
Right to Erasure



Right to Data Portability



"Right not to be subject to a decision based solely on automated processing, including profiling."



Automated Decision Making

**TEACHPRIVACY**

[www.teachprivacy.com](http://www.teachprivacy.com)

## THE PLAYERS



Data Subjects



Data Controllers



Data Processors



Supervisory Authorities

## PERSONAL DATA



Identified



Identifiable

## SENSITIVE DATA



Religious or Philosophical Beliefs



Trade Union Membership



Political Opinions



Sex Life



Genetic Data



Biometric Data



Health

## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.



Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.



Data Impact Assessment

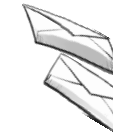
For high risk situations

Data Protection by Design

built in starting at the beginning of the design process



## DATA BREACH NOTIFICATION

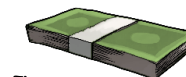


A personal data breach is "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed."

If likely to result in a high privacy risk → notify data subjects

Notify supervisory authorities no later than 72 hours after discovery.

## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies: compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



Adequate Level of Data Protection



Privacy Shield



Model Contractual Clauses

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute



## TERRITORIAL SCOPE

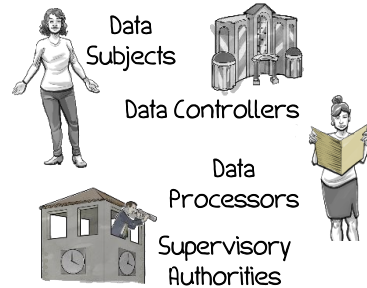


EU Establishments

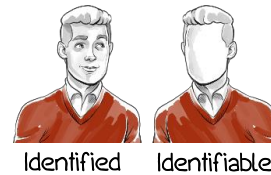
Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

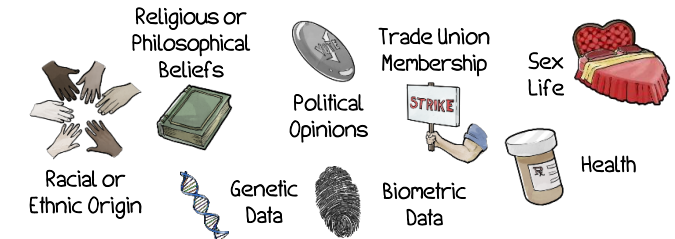
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.

Data Protection by Design

built in starting at the beginning of the design process

Data Impact Assessment

For high risk situations

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

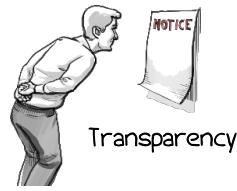
- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests

## CONSENT

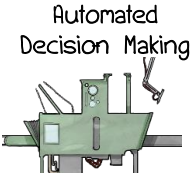


Consent must be freely given, specific, informed, and unambiguous.

## RIGHTS OF DATA SUBJECTS



Transparency



Automated Decision Making

"Right not to be subject to a decision based solely on automated processing, including profiling."



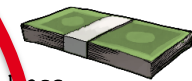
Access and Rectification

Right to Erasure

Purpose Specification and Minimization

Right to Data Portability

## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.

Effective Judicial Remedies:

compensation for material and non-material harm.

## INTERNATIONAL DATA TRANSFER



Adequate Level of Data Protection

Binding Corporate Rules (BCRs)



Privacy Shield



Model Contractual Clauses

TEACHPRIVACY

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute



## TERRITORIAL SCOPE

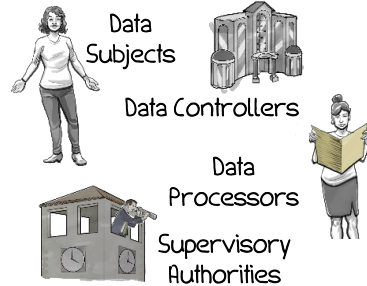


EU Establishments

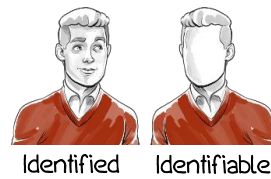
Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

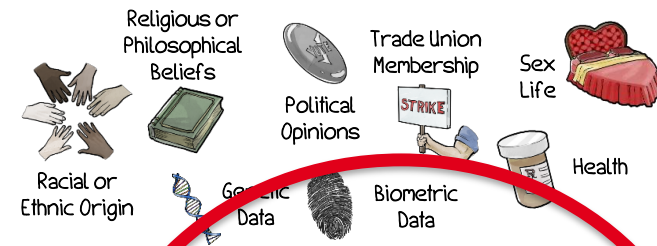
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.

Data Protection by Design

built in starting at the beginning of the design process

Data Impact Assessment

For high risk situations

## DATA BREACH NOTIFICATION



A personal data breach is "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed."

If likely to result in a high privacy risk → notify data subjects

Notify supervisory authorities no later than 72 hours after discovery.

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests



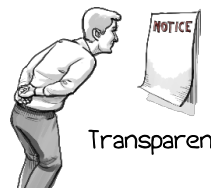
## CONSENT



Consent must be freely given, specific, informed, and unambiguous.

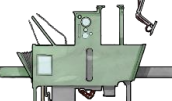
# GDPR

## RIGHTS OF DATA SUBJECTS



Transparency

Automated Decision Making



"Right not to be subject to a decision based solely on automated processing, including profiling."



Access and Rectification

Right to Erasure



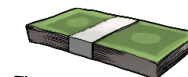
Purpose Specification and Minimization



Right to Data Portability



## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.



Effective Judicial Remedies: compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



Privacy Shield

Adequate Level of Data Protection



Model Contractual Clauses

**TEACHPRIVACY**

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute

## TERRITORIAL SCOPE

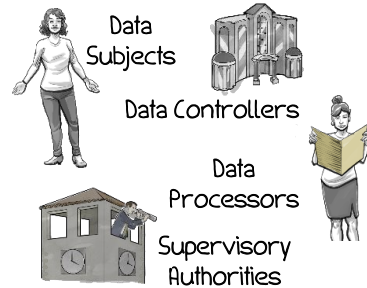


EU Establishments

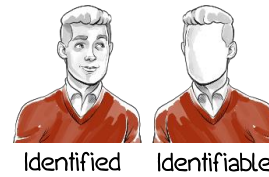
Non-EU Established Organizations

Offer goods or services or engaging in monitoring within the EU.

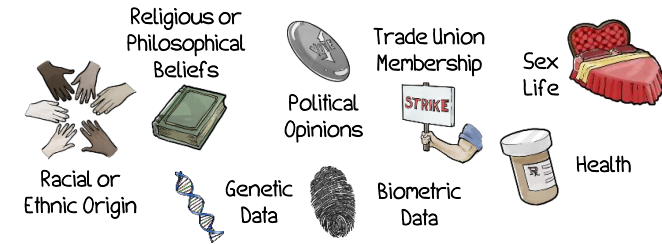
## THE PLAYERS



## PERSONAL DATA



## SENSITIVE DATA



## RESPONSIBILITIES OF DATA CONTROLLERS AND PROCESSORS

Security



Data Protection Officer (DPO)

Designate DPO if core activity involves regular monitoring or processing large quantities of personal data.

Record of Data Processing Activities

Maintain a documented register of all activities involving processing of EU personal data.

Data Protection by Design

built in starting at the beginning of the design process

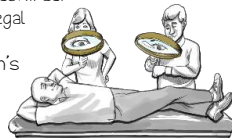
Data Impact Assessment

For high risk situations

## LAWFUL PROCESSING

Collection and processing of personal data must be for "specified, explicit and legitimate purposes" – with consent of data subject or necessary for

- performance of a contract
- compliance with a legal obligation
- to protect a person's vital interests
- task in the public interest
- legitimate interests

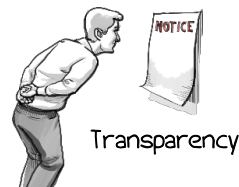


## CONSENT



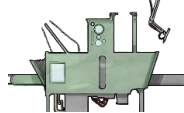
Consent must be freely given, specific, informed, and unambiguous.

## RIGHTS OF DATA SUBJECTS



Transparency

Automated Decision Making



"Right not to be subject to a decision based solely on automated processing, including profiling."



Access and Rectification

Right to Erasure



Purpose Specification and Minimization

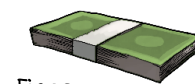


Right to Data Portability



# GDPR

## ENFORCEMENT



Fines

Up to 20 million euros or 4% of total annual worldwide turnover. Less serious violations: Up to 10 million euros or 2% of total annual worldwide turnover.

Effective Judicial Remedies:

compensation for material and non-material harm.



Binding Corporate Rules (BCRs)



## INTERNATIONAL DATA TRANSFER



Adequate Level of Data Protection



Privacy Shield



Model Contractual Clauses

TEACHPRIVACY

[www.teachprivacy.com](http://www.teachprivacy.com)

Workforce awareness training by Prof. Daniel J. Solove

Please ask permission to reuse or distribute

# TOOLS

- Legal Basis:
  - ICO tool: <https://ico.org.uk/for-organisations/gdpr-resources/lawful-basis-interactive-guidance-tool>
- PIA tool:
  - CNIL tool: <https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment>

# The CNIL First GDPR Fine!

CNIL.

## **The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC**

*21 January 2019*

---

*On 21 January 2019, the CNIL's restricted committee imposed a financial penalty of 50 Million euros against the company GOOGLE LLC, in accordance with the General Data Protection Regulation (GDPR), for lack of transparency, inadequate information and lack of valid consent regarding the ads personalization.*

# The CNIL First GDPR Fine!

## **A violation of the obligations of transparency and information**

« *Essential information, such as the data processing purposes, the data storage periods or the categories of personal data used for the ads personalization, are excessively disseminated across several documents, with buttons and links on which it is required to click to access complementary information. The relevant information is accessible after several steps only, implying sometimes up to 5 or 6 actions. »*

## **A violation of the obligation to have a legal basis for ads personalization processing**

- The collected consent is neither “*specific*” nor “*unambiguous*”

**More infos:** <https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc>



# Privacy by Design

***Privacy by Design*** is an approach to system engineering which takes privacy into account throughout the whole engineering process.

*Some (basic) principles:*

1. Proactive not reactive; build in privacy up front
2. Privacy as the default setting
3. End-to-end security
4. Visibility and transparency
5. User-centric/ user control

# Not a Really New Concept...

## *Privacy by Design in Ancient Cities (400 BC)*

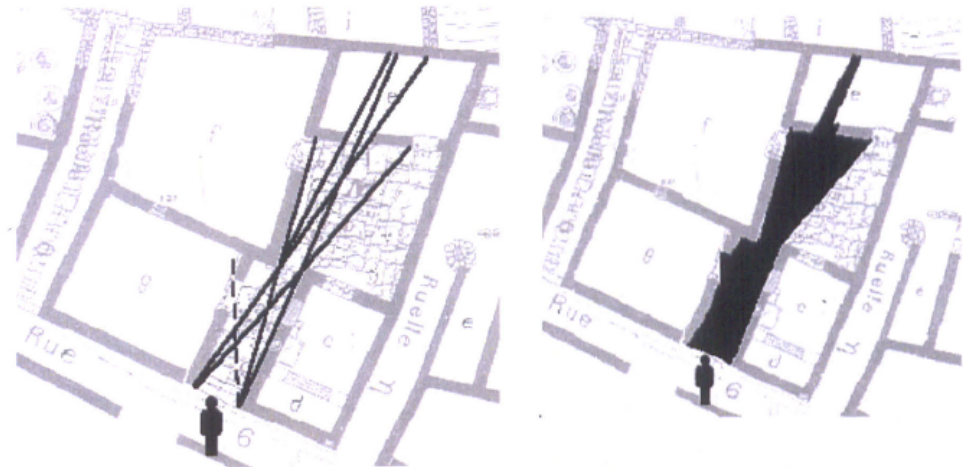
Greeks used geometry to create housing that:

- Minimize exposure to public view while
- Maximizing available light.

This is a good example of Privacy by Design!



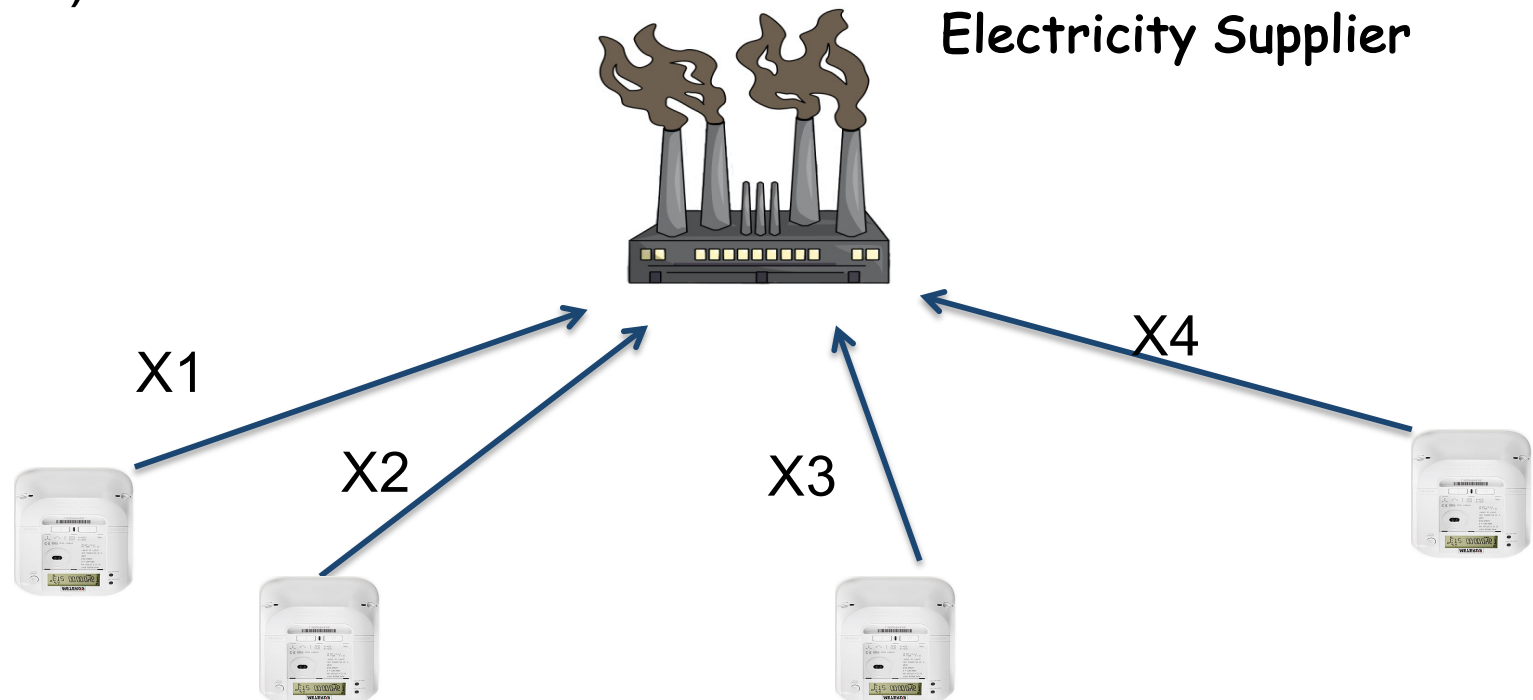
Figure 92: VI Ni window  
(Scale bar measures 0.50m, total length - left)



# Privacy by Design: *The Smart Metering Case*

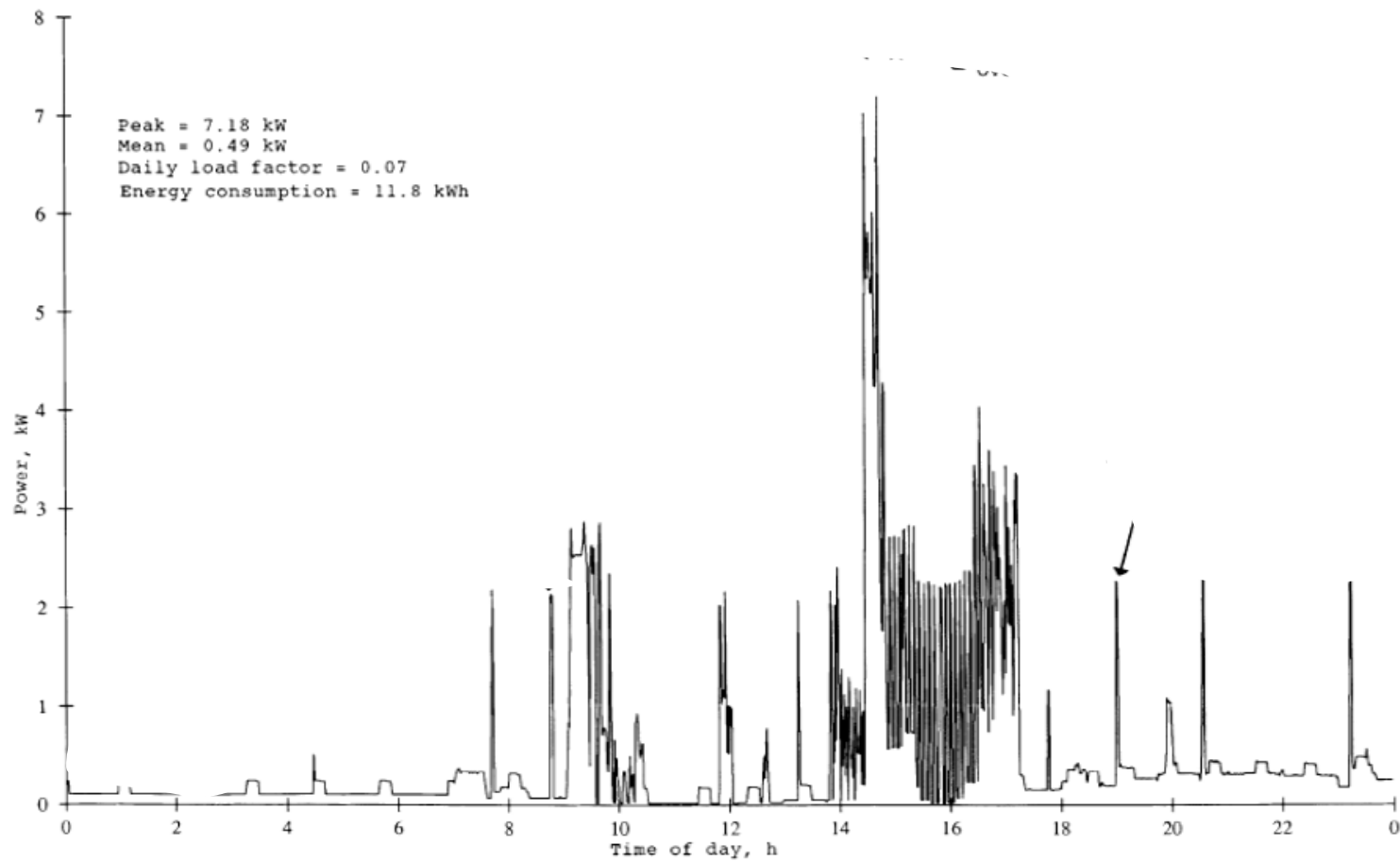
Electricity suppliers are deploying smart meters

- Devices@home that report energy consumption periodically (every 5-15 minutes).
- To improve energy management (for suppliers and customers) ...

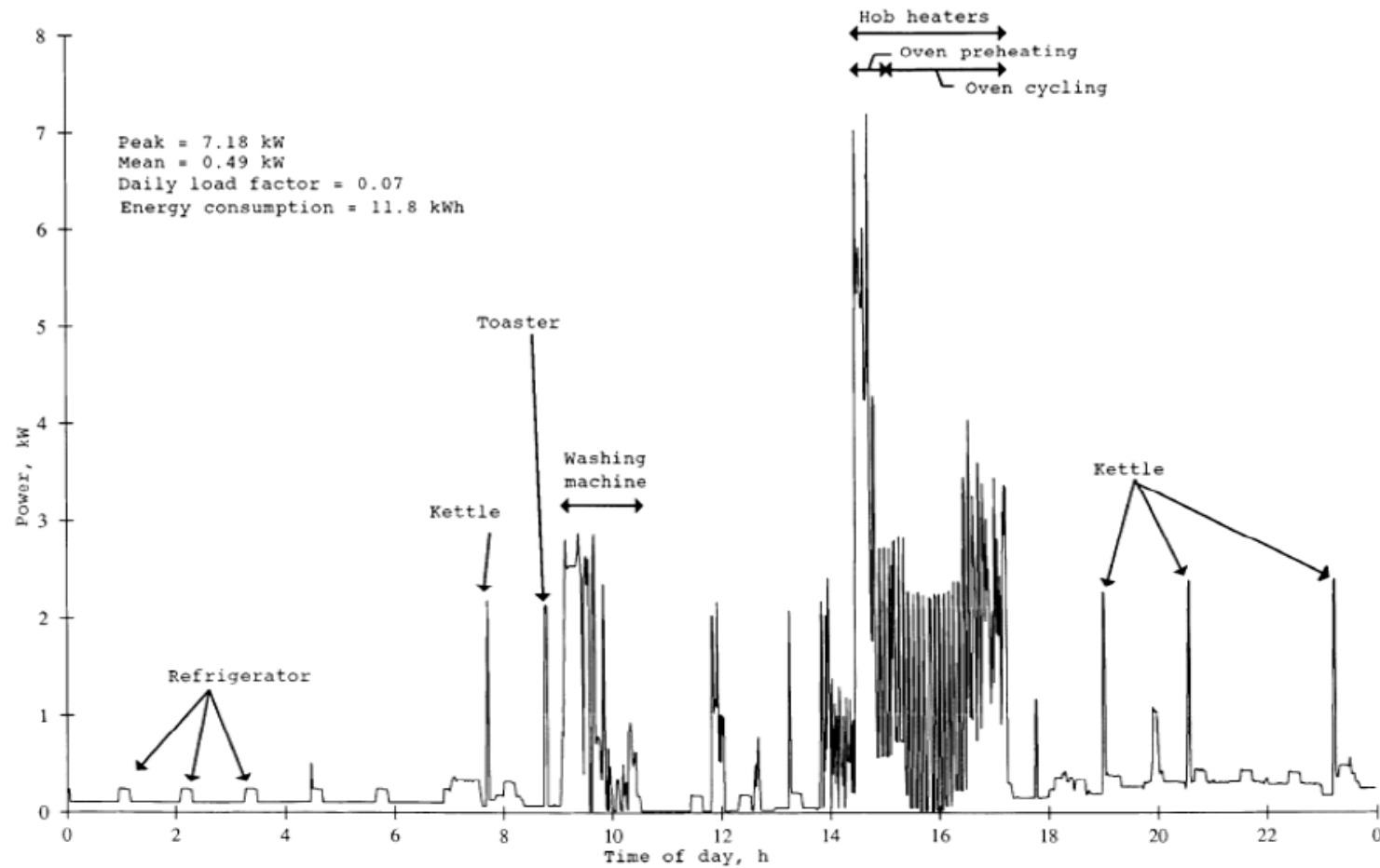




# Privacy by Design: *Privacy?*



# Privacy by Design: *Privacy?*



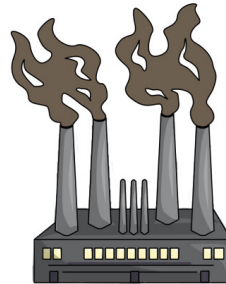
# Privacy by Design & Smart Metering

Can we design an architecture that:

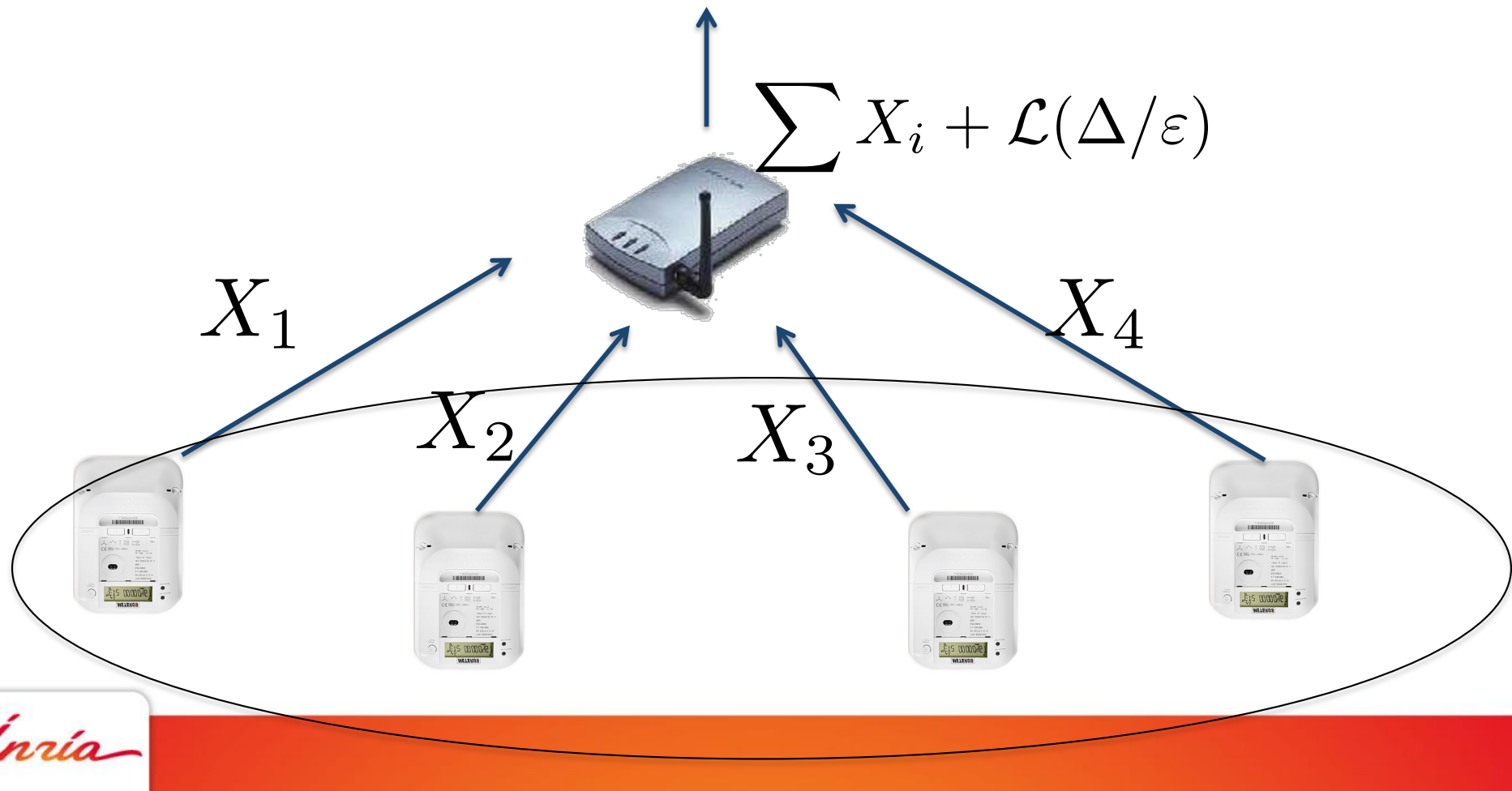
- Preserves user privacy (i.e. prevents profiling)
- While still preserving the benefits of smart metering i.e.
  - Electricity providers get enough statistical data to manage electricity efficiently
  - User gets fine-grain (possibly instantaneous) consumption information.

# A PbD Proposal: *Differentially-Private Aggregation*

Electricity Supplier

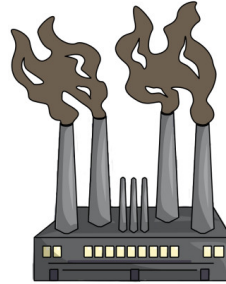


Supplier gets Differentially Private aggregated value but can't recover individual sample!



# A PbD Proposal: *Differentially-Private Aggregation*

Electricity Supplier



Supplier gets Differentially Private aggregated value but can't recover individual sample!

*Aggregator must  
Be trusted ☹*

$$\sum X_i + \mathcal{L}(\Delta/\varepsilon)$$



$X_1$



$X_2$



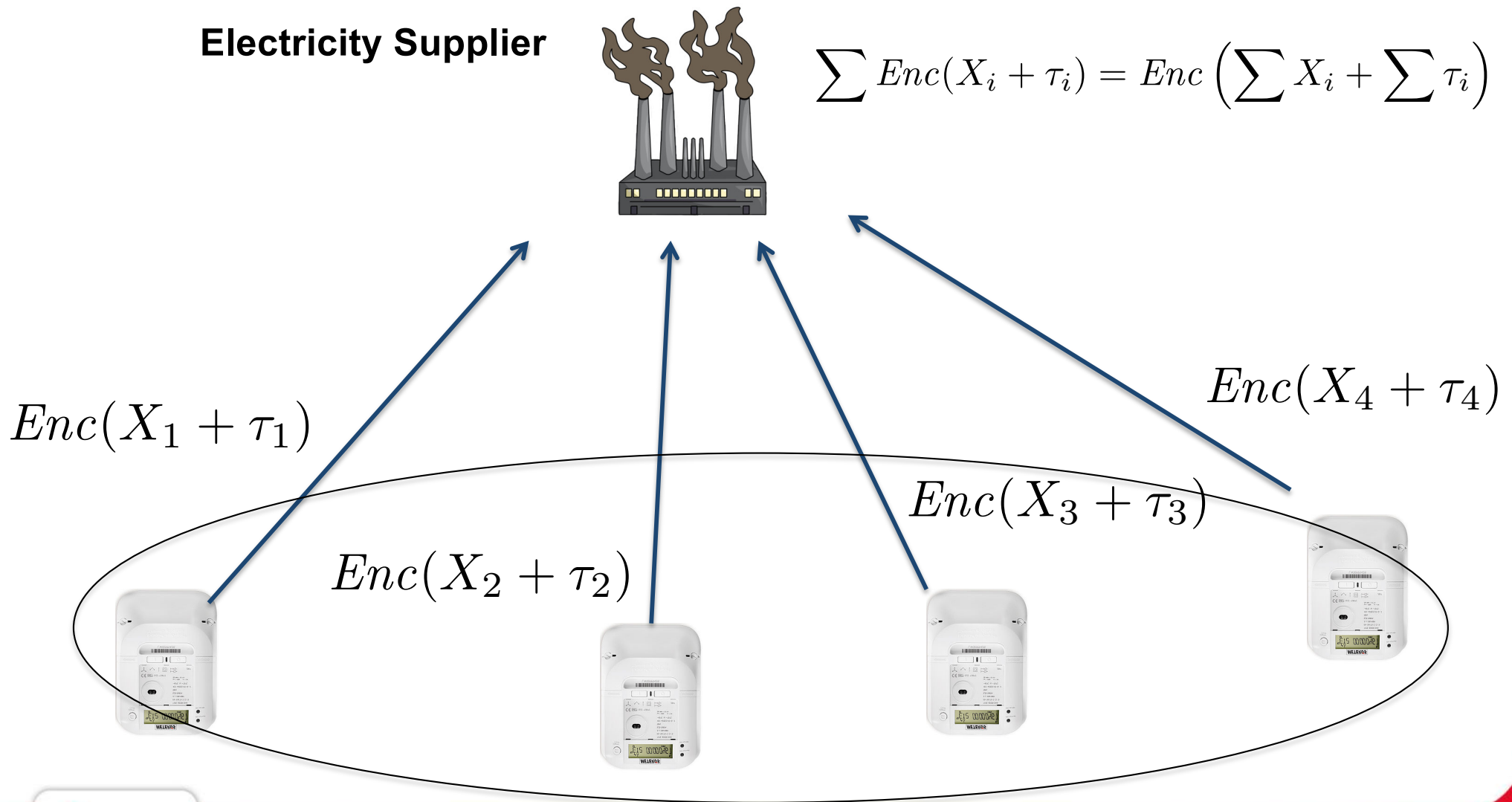
$X_3$



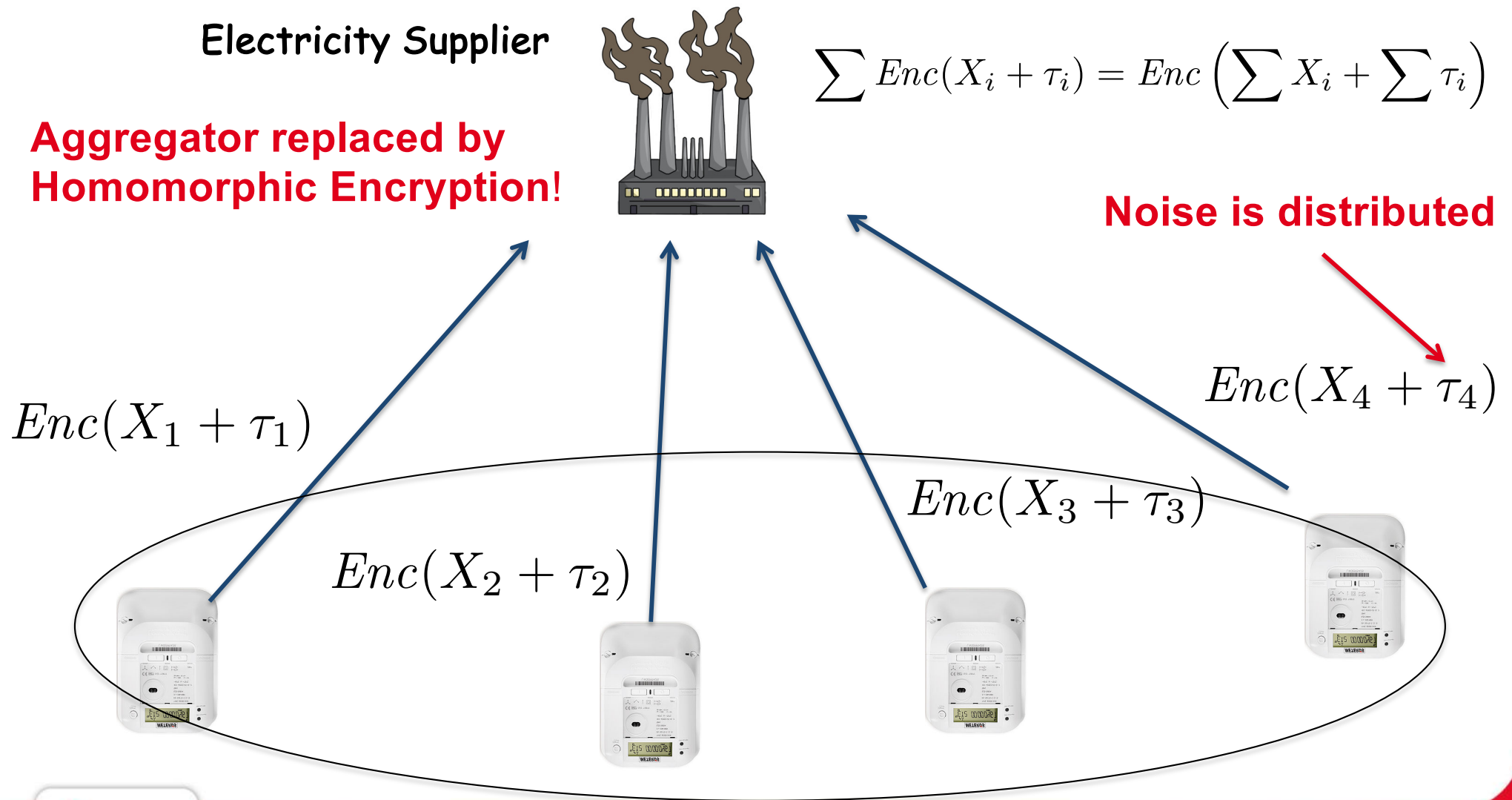
$X_4$



# Another PbD Solution: *Distributed Aggregation with Homomorphic Encryption*



# Another PbD Solution: Distributed Aggregation with Encryption



# Conclusion



- **Many Research Challenges** in the coming years...
  - We are just at the beginning of dataveillance and data manipulation...
  - The emergence of AI, Deep Fakes, IoT, smart devices will make things worse!
  - It is not only about Privacy, but also **Ethics!**
- **Responsible and Ethical AI/ Algorithm-based Decision Systems**
  - Transparency
  - Explainability
  - Biases detection/removal
  - Accountability





# MERCI!

[claudio.castelluccia@inria.fr](mailto:claudio.castelluccia@inria.fr)

The Joy of Tech™



© 2013 Geek Culture

by Nitrozac & Snaggy



joyoftech.com