

Privacy-Enhancing Technologies

Module 1: General Background

Thorsten Strufe

2.05.2022 – hybrid, KIT and TUD

Disclaimer: This lecture was prepared in cooperation with

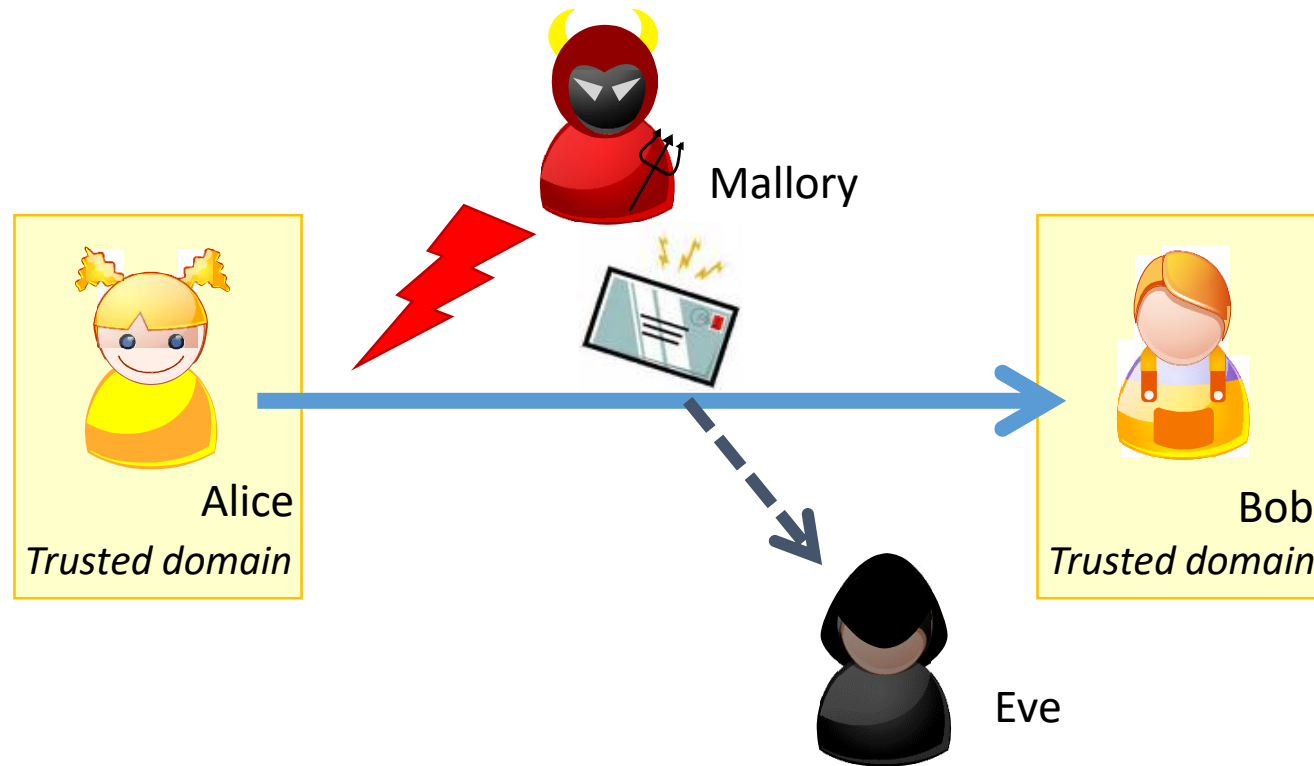
Patricia Arias-Cabarcos, Javier Parra-Arnau, and input from the people at the chair



KASTEL Security Research Labs

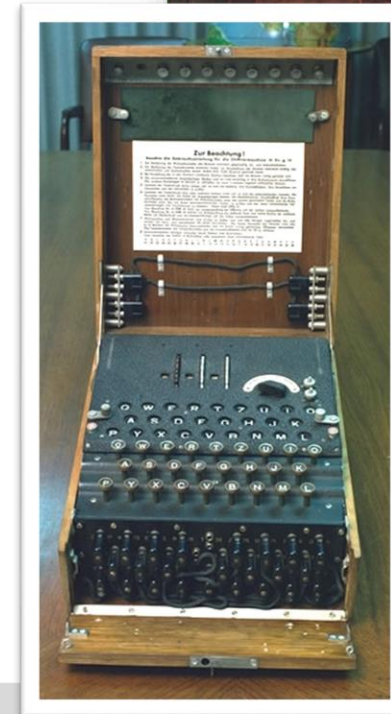


The Classical Security View



Threats!!

- Data loss
 - Data accessible to unintended parties
- Manipulation and forgery
 - Tampered, spoofed data



Classical *Security Goals* and *Adversaries*

- **Confidentiality**

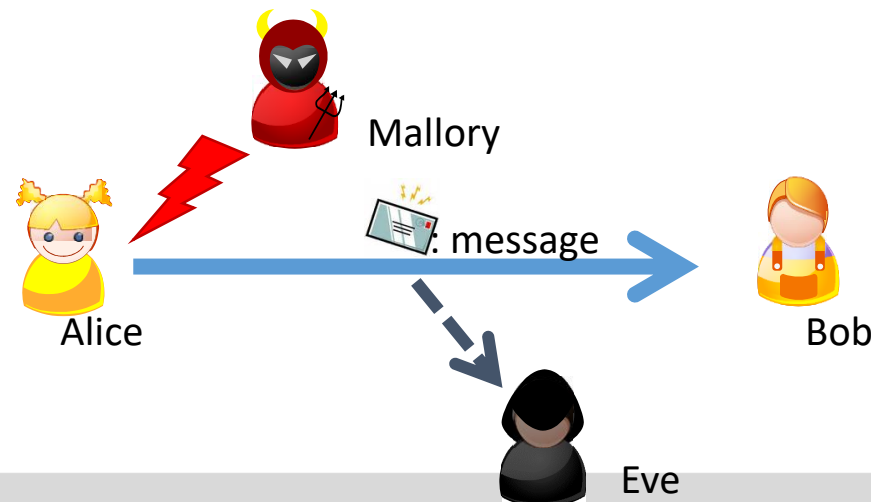
- Data transmitted or stored should only be revealed to the intended audience

- **Integrity**

- Modification of data is detected (identify source, first!)

- **Availability**

- Services should function correctly upon request





00	54	00	30	20	00
25	00	30	00	00	32
20	00	25	36	00	30
32	00	64	2E	00	20
00	00	00	73	00	60
00	6E	00	2E	00	65
00	78	00	52	74	00
00	78	00	79	65	00

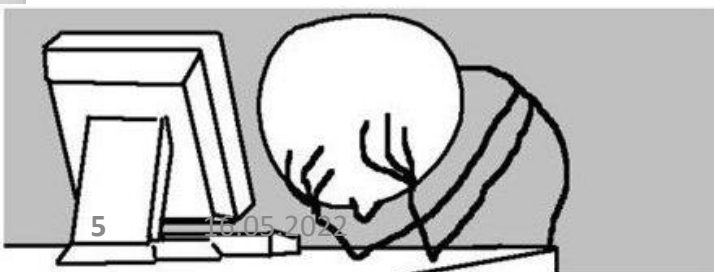
MIKE MCQUADE

ANDY GREENBERG

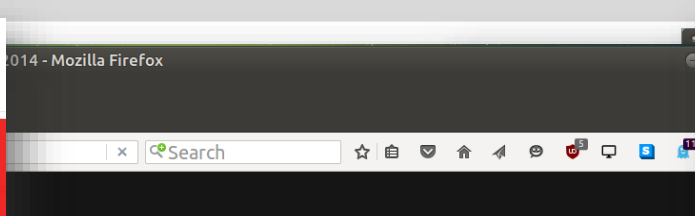
SECURITY 00.22.2018 05:00 AM

The Untold Story of NotPetya, the Most Devastating Cyberattack in History

Crippled ports. Paralyzed corporations. Frozen government agencies. How a single piece of code crashed the world.



enhagen
egan to lose



Confirmation Account

Thursday that information was stolen from... believed was a state-sponsored... the biggest cyber... stolen may have included... phone numbers, dates of... words but that unprotected... data, and bank account... have been compromised.

gest data breach ever," source Schneier.

too early to say what impact... Yahoo and its users became... in, including the identities... ers behind it.

elligence officials, who... me, said they believed t

researchers found a vulnera... attacker to unlock doors by... the mobile app.

ickstarter

ocks are sold as devices... convenient, but security re... t easy for hackers and th

2014 - Mozilla Firefox

Search

YAHOO

Discover Thomson Reuters

REUTERS

Yahoo says all three billion accounts hacked in 2013 data theft

#BUSINESS NEWS OCTOBER 3, 2017 / 10:57 PM / UPDATED 14 HOURS AGO

Jonathan Stempel, Jim Finkle

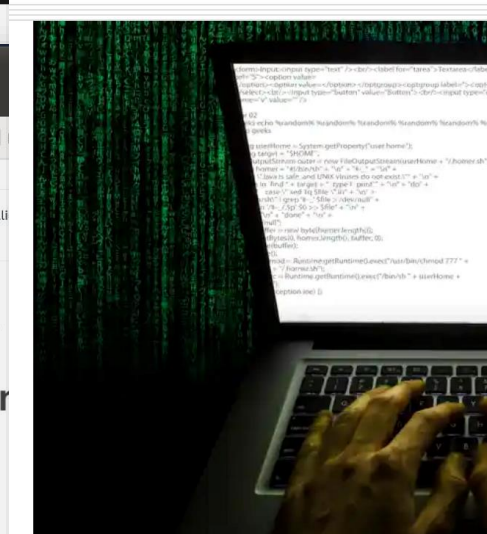
A photo illustration shows a Yahoo logo on a smartphone in front of a display of code and keyboard on December 15, 2016. REUTERS/Dado Ruvic/Illustration

Sign in

Support us →

News Opinion Sport Culture Lifestyle

Coronavirus World UK Environment Science



Hacking

DDoS attack that disrupted internet was largest of its kind in history, experts say

Dyn, the victim of last week's denial of service attack, said it was orchestrated using a weapon called the Mirai botnet as the 'primary source of malicious attack'

- Major cyber attack disrupts internet service across Europe and US

Nicky Woolf in San Fran

@nickywoolf
Wed 26 Oct 2016 21:42 BST

The cyber-attack that b... caused by a new weapo... in history, experts said... The victim was the ser



So... Privacy?

Digital Dystopias

“With the development of television, and the technical advance which made it possible to receive and transmit simultaneously on the same instrument, private life came to an end.”

George Orwell, “1984”, 1948

Privacy – Dictionary Definition

pri·va·cy | \ 'prī-və-sē , especially British 'pri- \ plural privacies

Definition of privacy

1

- a) a : the quality or state of being apart from company or observation : seclusion
- b) freedom from unauthorized intrusion <one's right to privacy>

2

- archaic : a place of seclusion

3

- a) Secrecy
- b) a private matter : secret

- **Social and legal aspect**
 - *a hard to define social concept:*
 - *social scientists, philosophers and lawyers*
- **Privacy is somewhat subjective**
 - *Understanding is a cultural construct*
 - *Changes between different societies*
 - *no precise and universal definition*

From Merriam Webster Online Dictionary

Notions of Privacy: Right to be let alone

- Samuel Warren, Louis Brandeis: **“The Right to Privacy”**, Harvard Law Review, Vol. IV, No. 5, 15th December **1890**
- **Reason:** “snapshot photography” (recent innovation at that time)
 - allowed newspapers to publish photographs of individuals without obtaining their consent.
 - private individuals were being continually injured
 - this practice weakened the “moral standards of society as a whole”
- **Consideration:**
 - basic principle of common law: individual shall have full protection in person and in property
 - “it has been found necessary from time to time to define anew the exact nature and extent of such protection”
 - “Political, social, and economic changes entail the recognition of new rights”
- **Conclusion:**
 - “right to be let alone”

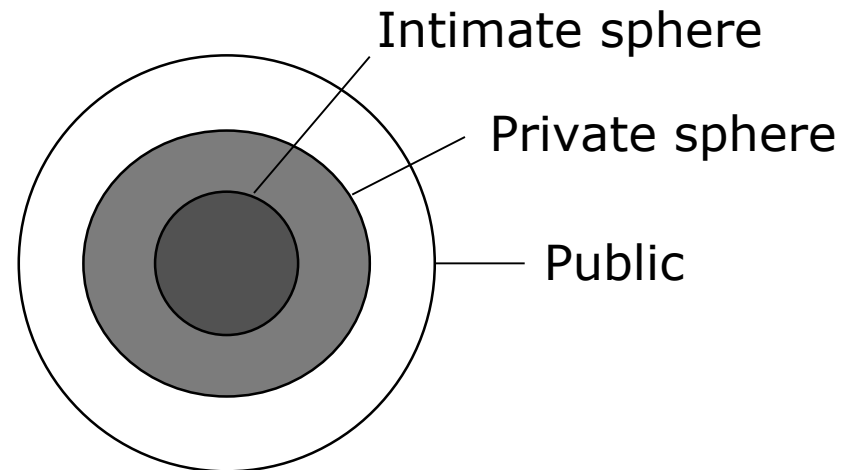
Privacy in CS and Engineering?

„the claim of individuals ... to determine for themselves when, how, and what extent of information about them is communicated to others.“

-- Alan Westin (1967)

Modelling Privacy: Spheres

- Modelling protection requirements (expectations) of classes of information as concentric circles of decreasing need for protection

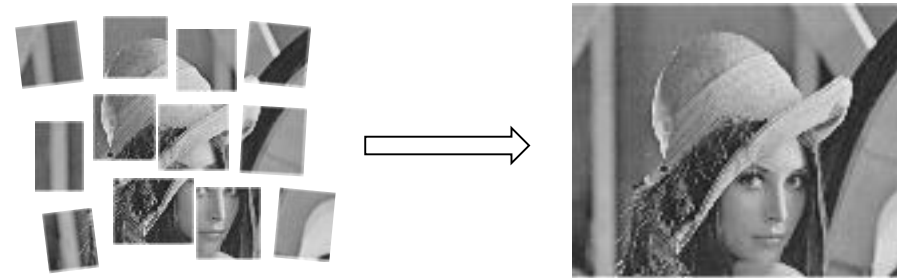


Information of an individual

- Assign data to corresponding spheres
- Assignment may depend on context and situation...

Modelling Privacy: The Human Mosaic

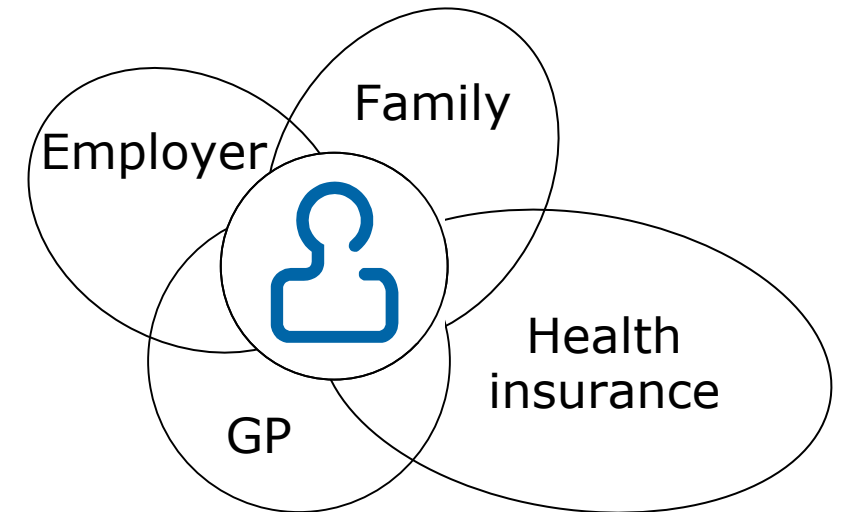
- Small snippets of information (probably) don't expose a human
- Loss (and aggregation) of several snippets lead to a mosaic of the individual
- Increasing aggregation of puzzle pieces increases detail of knowledge on the individual



- Management of pieces that initially are not considered „intimate“ possible
- Independence of the way data is lost (or: collected)
- Does not simplify determining criticality of pieces
- Considers data capture/collection, but also further data processing

Modelling Privacy: The Human in Roles

- Humans act in roles depending on their situation
- Usually specific information required to achieve certain task
- Group shared information according to context
- Personas, various levels of sensitivity
- Individual images to be restrained to context
- Transfer through 3rd parties may cause unknown leaks



Notions of Privacy: Contextual Integrity

- Helen Nissenbaum: *Privacy as Contextual Integrity*, 2004
- Idea that data is shared with specific mind set in specific context

- Two types of violation of expected privacy:
 - violation of **Appropriateness of Revelation**
 - the context „defines“ if revealing a given information is appropriate
 - **violation:** information disclosed in one context (even “public”) may not be appropriate in another (Asking a person participating in a gay pride vs. the same participating in a governmental press conference)

 - violation of **Distribution**
 - the context „defines“ which information flows are appropriate
 - **violation:** inappropriate information flows between spheres, or contexts; information disclosed in one context used in another (telling, even if first context was “public”)

Collateral Development: Ethics in Research

- Ahnenerbe Society, Head of Institute for Military Scientific Research (*Institut für Wehrwissenschaftliche Zweckforschung*), lead by Wolfram Sievers organizes human experiments
- Nuremberg Doctor's Trials
 - 20 Doctors, Sievers and 2 other Nazi officials
 - Led to 7 executions, 7 defendants acquitted
- Nuremberg Code (1947)
- Declaration of Helsinki (1964)
- Belmont Report (1978)
- Excerpt from **Nuremberg code**:
 1. Required is the **voluntary, well-informed, understanding consent** of the human subject in a full legal capacity.
 2. The experiment should **aim at positive results for society** that cannot be procured in some other way.
 3. It should be **based on previous knowledge** (like, an expectation derived from animal experiments) that justifies the experiment.
 4. The experiment should be set up in a way that **avoids unnecessary physical and mental suffering and injuries**.
 5. It **should not be conducted** when there is any reason to believe that it implies a risk of death or disabling injury.
 6. The **risks of the experiment** should be in proportion to (that is, not exceed) the expected **humanitarian benefits**.

Legal Foundations

(IANAL)



- UDHR Art. 1: *„All human beings are born free and equal in dignity and rights. [...]”*
(Also: ECHR Art. 1, Art. 1 Grundgesetz)
- UDHR Art. 12: *„No one shall be subjected to arbitrary interference with his privacy, ...”*
(Also: ECHR Art. 8)
- Charter of Fundamental Rights (CFR) Art. 8:
 - *„1. Everyone has the right to the protection of personal data concerning him or her.”*
 - *„2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law.”*
- Important: ***Prohibition with reservation of authorization***
- Introduces concept of control through responsible institutions (data protection officer)
- Fundamental Law, historically protection from state, protection of minorities
- Regulation in Europe: General Data Protection Regulation (GDPR), US: per market, 4th Amdt.

Notions: Informational Self-Determination

- European Law based on ideas from the age of enlightenment:
 - freedom of choice and freedom to reinvent oneself
 - any citizen should be their own sovereign
- Service must not coerce disclosure
- Publishing/sharing free choice of citizen
- *Important underlying idea: the sovereign (self-determined) citizen **controls collection, use, and can effectively retract even previously openly published data, upon change of mind***
- Based on principles of processing:
 - collect and process personal data **fairly and lawfully**
 - **purpose binding**
 - keep it only for one or more specified, explicit and lawful purposes
 - use and disclose it only in ways compatible with these purposes
 - **data minimization**
 - adequate, relevant and not excessive wrt. the purpose
 - retained no longer than necessary
 - **transparency**
 - inform who collects which data for which purposes
 - inform how the data is processed, stored, forwarded etc.
 - **user rights**
 - access to the data, correction, deletion
 - **keep the data safe and secure**

Informational self-determination in brief

“The claim of individuals, groups and institutions to
determine themselves,
when,
how and
to what extent
information about them
is communicated to others”
(GDPR: is processed)

EU Data Protection Directive (95/46/EC):

- “personal data” shall mean any information relating to an **identified or identifiable** natural person (‘Data Subject’);

Legalese: Personally Identifiable Information („PII“)

- **US:** Name, address (Phone, Email), national identifiers (tax, passports), IP address, driving (vehicle registration, drivers licence), biometrics (face, fingerprints), credit card numbers, date/place of birth (age, login name(s), gender, "race", grades, salary, criminal records)
- **EU:** 'personal data' means any **information relating** to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, **directly or indirectly**, in particular **by reference to an identifier such as** a name, an identification number, location data, an online identifier **or to one or more factors** specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person; [Art. 4, GDPR]



Which processing? „Anonymized“ (Pseudonyms)?

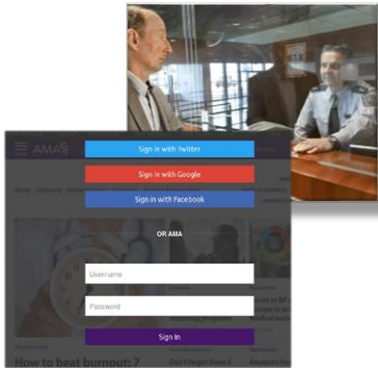
- *'processing' means **any operation** or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as **collection, recording**, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, **disclosure by transmission**, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction; [ebd]*
- GDPR's take on **pseudonymisation**:
- *'pseudonymisation' means the processing of personal data in such a manner that the personal data **can no longer be attributed to a specific data subject** without the use of **additional information**, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person; [ebd]*
- ***Pseudonymous data can be linked back to individual, and it hence is considered PII!***

Why is this suddenly so relevant?

Humanity and Cultural Practices



#TactileInternet
ceti.one

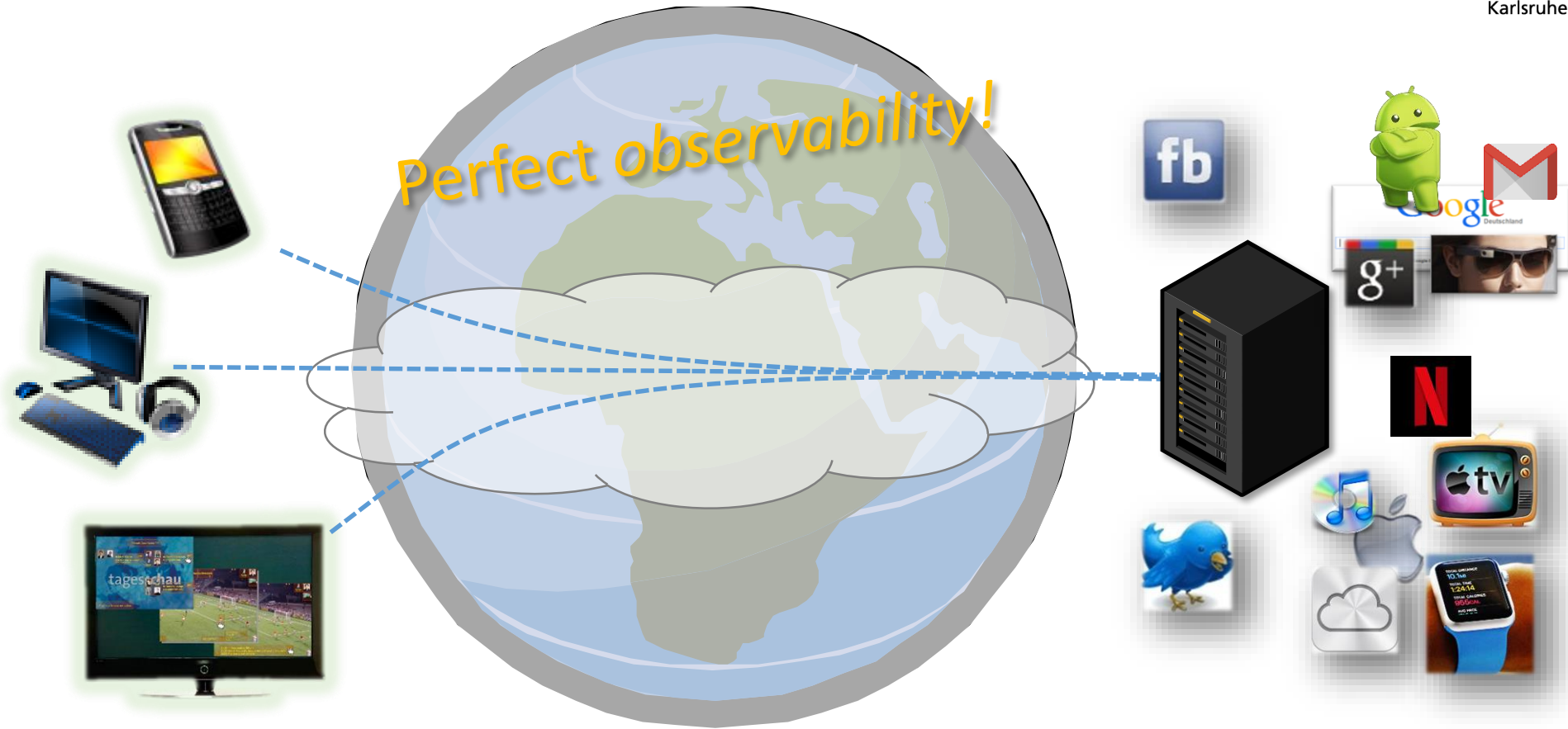


Access: Type, Scope, and Trust



- 1: Personal, unidentified
- 2: Local, decentralized
- 3: Trust in direct peer (village)

Access: Type, Scope, and Trust Today



- 1: Central, unique global login services
- 2: Global access over Internet
- 3: Trust in ... (I)SP?

Recent Anecdotes from the Trenches...

Case Study: Corona Warn-Apps



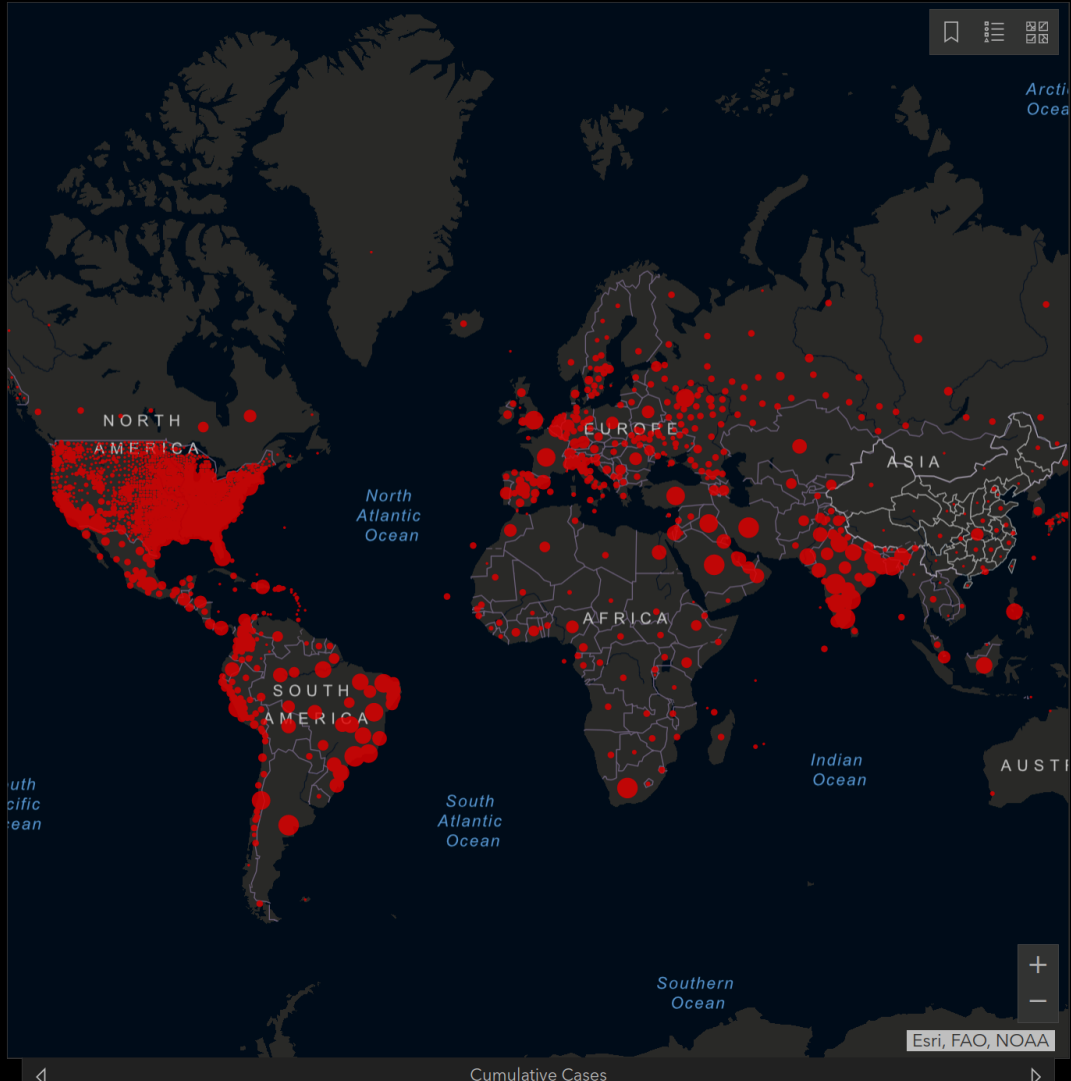
Global Cases
23,670,007

Cases by Country/Region
/Sovereignty

- 5,740,909 US
- 3,622,861 Brazil
- 3,167,323 India
- 963,655 Russia
- 611,450 South Africa
- 600,438 Peru
- 563,705 Mexico
- 551,688 Colombia
- 405,436 Spain
- 399,568 Chile
- 361,150 Iran
- 350,867 Argentina
- 328,620 United Kingdom
- 308,654 Saudi Arabia
- 297,083 Bangladesh
- 293,711 Pakistan
- 282,414 France
- 260,298 Italy
- 259,692 Turkey

Admin0

Last Updated at (M/D/YYYY)
8/25/2020, 11:28:02 AM
16.05.2022



188
countries/regions

Lancet Inf Dis Article: [Here](#). Mobile Version: [Here](#). Data sources: [Full list](#). Downloadable database: [GitHub](#), [Feature Layer](#).
Lead by JHU CSSE. Technical Support: [Esri Living Atlas team](#) and [JHU APL](#). Financial Support: [JHU](#), [NSF](#), [Bloomberg Philanthropies](#) and [Stavros Niarchos Foundation](#). Resource support: [Slack](#), [Github](#) and [AWS](#). Click [here](#) to [donate](#) to the CSSE dashboard team, and other JHU COVID-19 Research Efforts. [FAQ](#). Read more in this [Strategic Privacy-Enhancing Technologies - Intro](#)



Opinion

Lockdown Is a Blunt Tool. We Have a Sharper One.

Contact tracing helps people to protect themselves and their families.

May 5, 2020

- 177,2 US
- 115,3 Brazil
- 60,80 Mexico
- 58,39 India
- 41,51 United Kingdom
- 35,44 Italy
- 30,53 France
- 28,87 Spain
- 27,81 Turkey

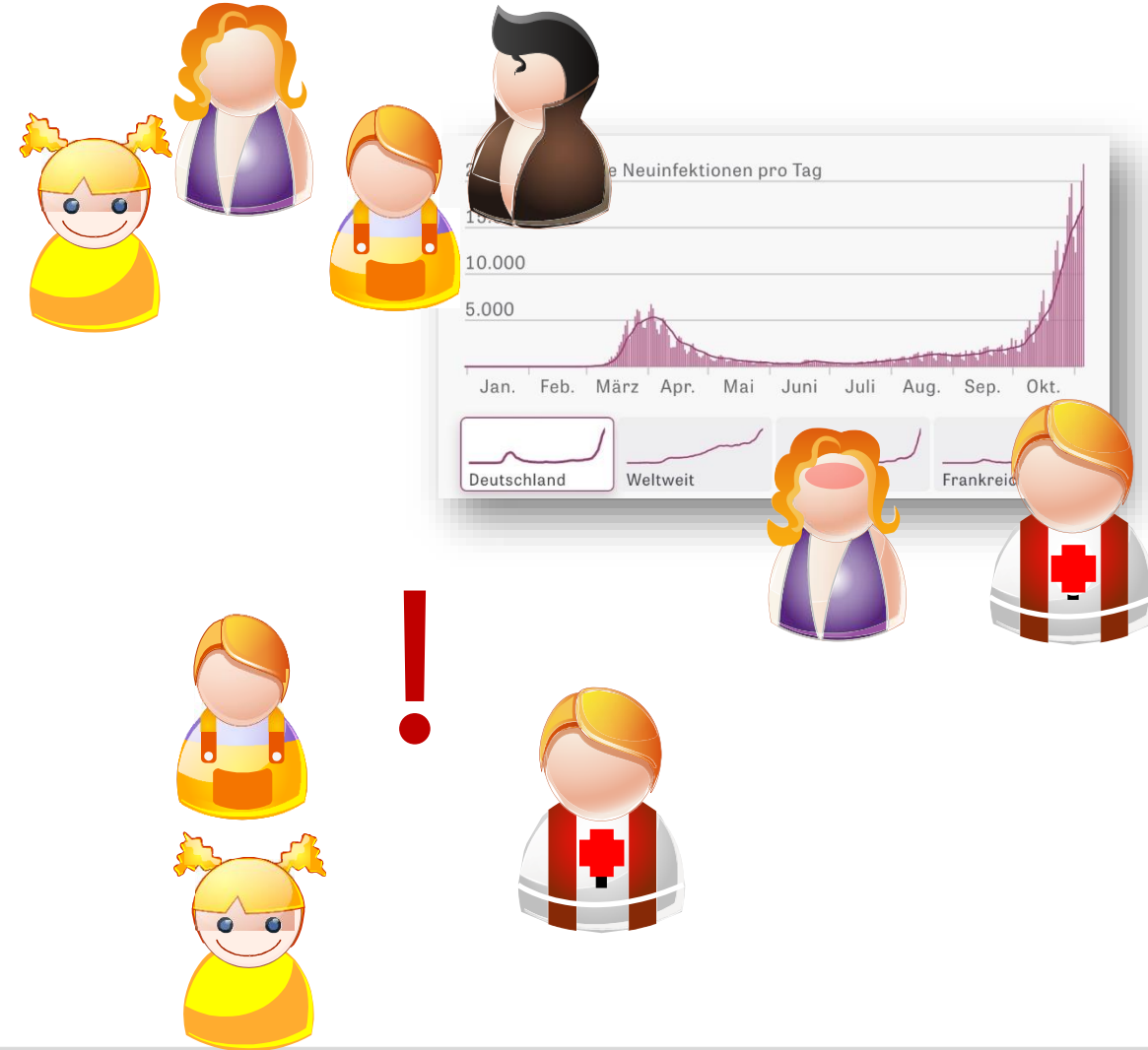


People waiting in line at a newly opened Covid-19 community testing location in New York City on Monday. Justin Lane/EPA, via Shutterstock

We've been dealt a bad hand with the coronavirus pandemic. Until we have a vaccine or effective treatment, we have limited tools to fight it. Closing large segments of our society and having people shelter at home is a blunt tool that works, but it inflicts severe hardship on individuals and the economy.

Surprisingly Difficult – Corona Warn Apps

- Encounters of people at $<2\text{m}$ distance leads to exposure risk
- Inform about risk in case of positive tests, to break infection chain
- Past: Ask who you met and call
- Now: Cell phones track encounters!



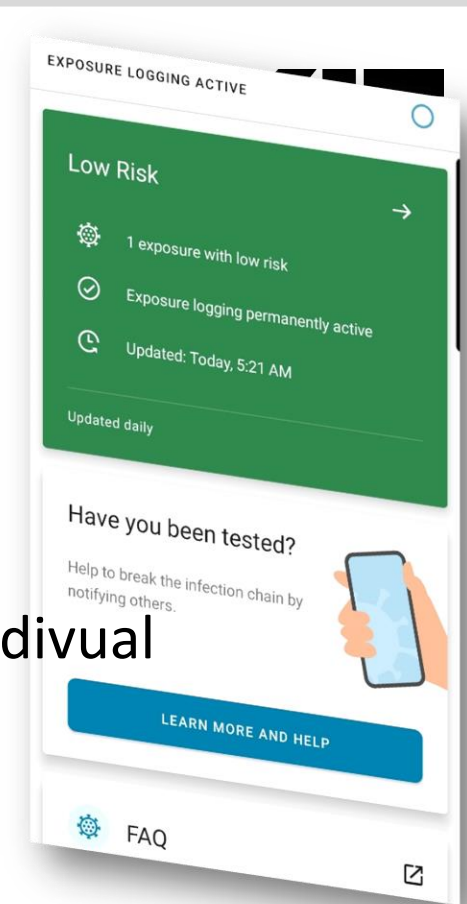
Corona Warn Apps – Overview

- **Functionality:**

- Encounter logging
- Infection reporting 1: Test positive report (verified)
- *(implicit) match encounters with non-infected individuals*
- Infection reporting 2 (risk notification): Inform potentially infected individual

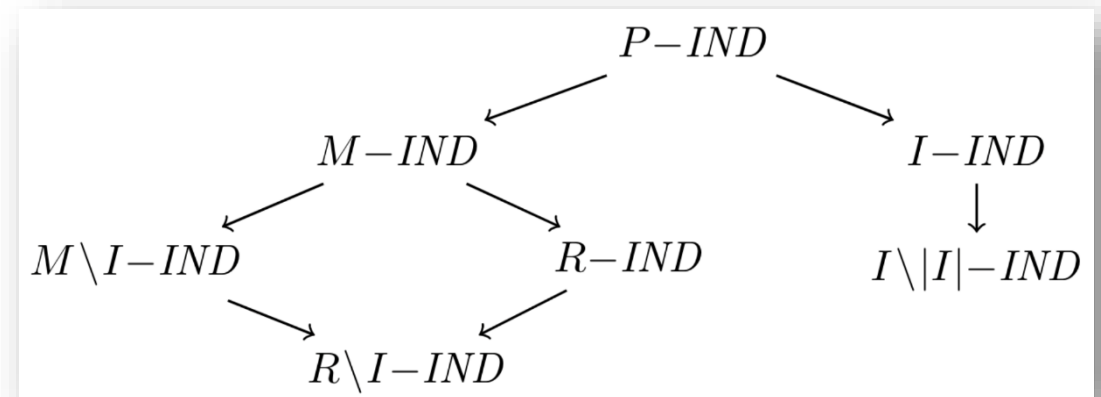
- **Necessary data:**

- (Co-)locations
- (infection risk factor)
- Infection (verified)



Corona Warn Apps – Privacy Risks/Covid Notions

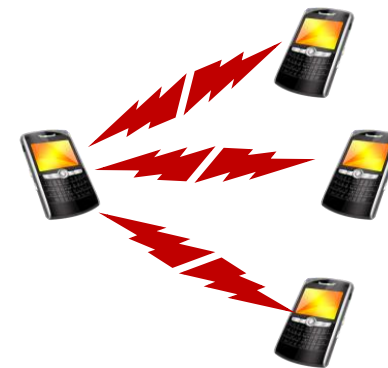
- **Location** (@time) of individuals (Beach, not the lecture)
 - Potentially linkable -> tracking
- **Social network** of individuals (Who meets and mingles)
- **Stigmatization**
 - Infection of an individual
 - Frequencies/fractions of infections in populations
- Architectural Roles
 - Users (phones)
 - Servers (yes, there are always servers)
- Conclusions about
 - Colocated others/3rd parties?
 - Infected/non-infected individuals?
- *Model as Indistinguishability game*



So where again is the Difference?

- Alternative narratives:
 - Trusted authorities, untrusted others
 - Trusted users, untrusted central party

 - Recall functions:
 1. Encounter logging
 2. Infection reporting 1 (positive test)
 3. Contact matching
 4. Infection reporting 2 (risk notification)
- In what follows: The European („Privacy Preserving“) perspective
 1. BLE broadcast shortlived pseudonym



Infection Reporting („Centralized“/“Decentralized“)

„Centralized“

■ Infection reporting 1:



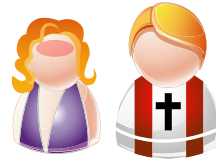
„Pseudonyms that have been colocated with an infected individual are: $RPI_a, RPI_b, RPI_c, \dots$ “

■ Infection reporting 2:



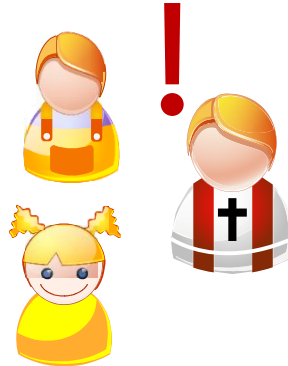
„Has someone reported colocation with RPI, \dots ?“

„yes/no“



„Decentralized“

■ Infection reporting 1:



„I am infected, my pseudonyms are: $RPI_a, RPI_b, RPI_c, \dots$ “

■ Matching and infection reporting 2:



„Tell me who's infected“

„ $RPI_a, RPI_b, RPI_c, \dots$ “

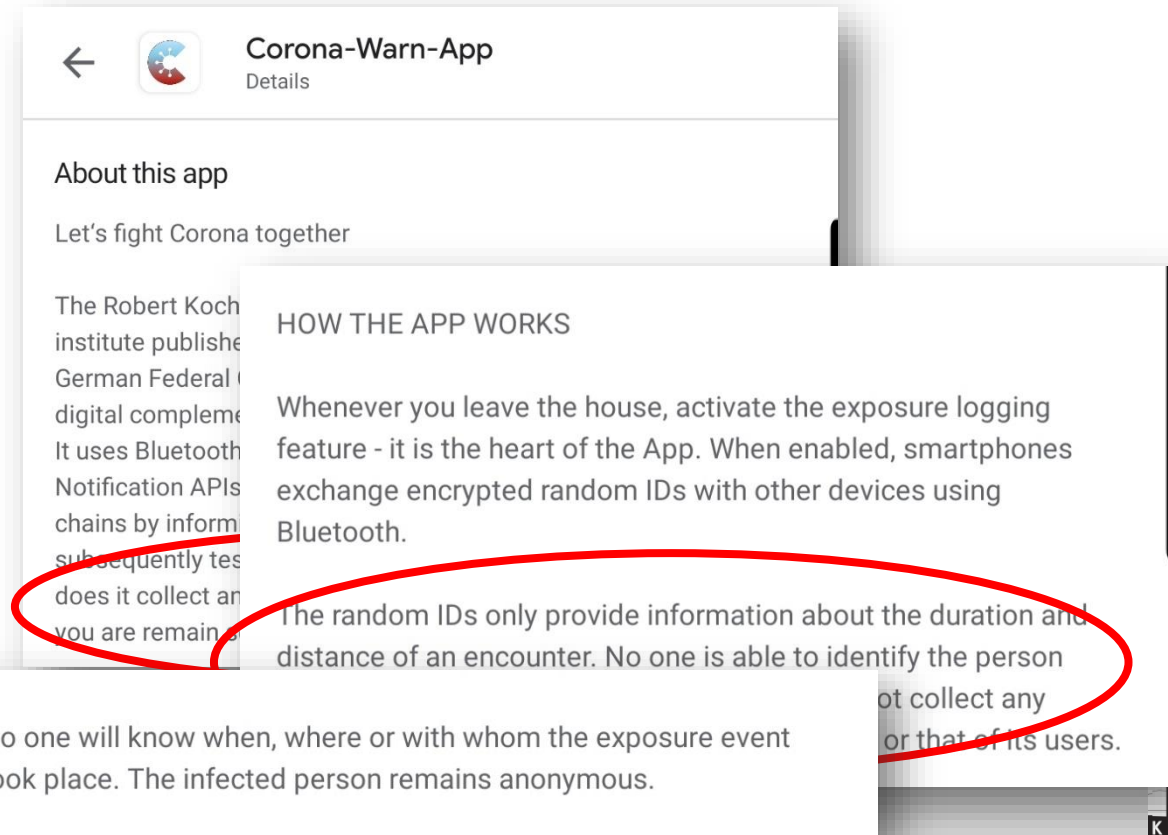
Secure hybrid approaches report colocated RPIs (1) and distribute the service provider


GAEN („decentralized“)

- *Exposure Notification API* in iOS/Android
 - Based on [CTA20]/DP3T-“simple“
 - **Encounter Logging:**
 - Phone creates daily pseudonym: tek_i
 - Derives „key“: $RPIK_i = H(tek_i, \langle \text{stuff} \rangle)$
 - Derives/publishes linkable „transaction pseudonyms“ (with rolling MAC):
 $RPI_{i,j} = AES(RPIK_i, \langle \text{stuff} \rangle | \text{time slot})$
- **Infection Reporting (1):**
 - Rename „tek“ to „diagnosis keys“ and upload to servers (with times)

■ Infection Reporting (2):

- Download all „diagnosis keys“
- Derive all $RPI_{i,j}$ and check for local coincidence, raise alarm



←  Corona-Warn-App
Details

About this app

Let's fight Corona together

The Robert Koch institute publishes the German Federal digital completion. It uses Bluetooth Notification APIs chains by inform subsequently test does it collect an you are remaining

HOW THE APP WORKS

Whenever you leave the house, activate the exposure logging feature - it is the heart of the App. When enabled, smartphones exchange encrypted random IDs with other devices using Bluetooth.

The random IDs only provide information about the duration and distance of an encounter. No one is able to identify the person or that of its users.

No one will know when, where or with whom the exposure event took place. The infected person remains anonymous.

GAEN: Critical Assessment

- What data is collected and processed?
 - Personally identifiable, linkable, health-related information
- What is disclosed to whom?
 - Pseudonyms of infected individuals, linkable throughout 24h, to the public
- How difficult is re-identification?
 - Trivial, Apps, Web pages have existed before Telekom/SAP got contract:

https://github.com/oseiskar/corona-sniffer

README.md

BLE contact tracing sniffer PoC

How "anonymous" is the semi-decentralized BLE contact tracing proposed by [Apple & Google](#) or [DP-3T](#)?

It is as anonymous as the simulated picture below - and this data is technically accessible to **any 3rd party who can install a large fleet of BLE-sniffing devices**. This is because all beacons signals broadcast by infected individuals are published to essentially all users of the system when an individual voluntarily uploads their positive infection status.

This repository contains a Proof-of-Concept implementation of a BLE-sniffing system that could uncover this data.

Simulation

The image shows the results of a simulation where 400 BLE-sniffing devices would have been deployed in a 20x20 grid over an area of 1500x1500 m². The movement of 300 people around the area have been (crudely) simulated as random walks.

COVID-LENS

List Exposure Notification Se

Click this button to see a list of device contact tracing app install

List Nearby Device

3 Get notified when they get covid

CORONA DETECTIVE

Add people Share Settings About

COVID19 amongst us

You



-SAP commented on Jun 15

If you please read
Mit freundliche
SW
Corona Warn

or if something is missing according to GDPR, and Freedom of Information.

Member ...

Mandatory DPIA (slightly paraphrased):
"If anybody were to deviate from the protocol that would be a crime. Can't do anything against crime, might as well ignore that problem."

GAEN: Critical Assessment

- What data is collected and processed?
 - Personally identifiable, linkable, health-related information
- What is disclosed to whom?
 - Pseudonyms of infected individuals, linkable throughout 24h, to the public
- How difficult is re-identification?
 - Trivial, Apps, Web pages have existed before Telekom/SAP got contract:
- Are these disclosures unavoidable?
 - „Any proximity tracing system that notifies users that they are at risk enables a motivated attacker to identify the infected people“ [1]
 - Not quite, and by far not that easily...
- For further vivid examples of corporate/academic campaigning, visit:
https://en.wikipedia.org/wiki/Exposure_Notification

Case Study: Ovulation Tracker App

Once upon a time last month...

- Case study: Ovulation tracker

- Necessary functionality:

- Record some regular observations on user's body
 - Extract some comp. simple cyclic regularities



May be...

Take photos of some supported products and check info online...

Uhm...

Allow for discussions on a forum

Upload data to the cloud (???)

- Clearly

- personally identifiable data
 - **sensitive** even sexual preferences and health related data..





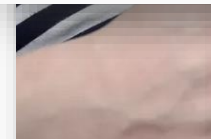
AppsFlyer, the global attribution leader, empowers marketers to grow their analytics solutions. Built around privacy by design, AppsFlyer takes a customer partners make better business decisions every day.

2. END USER DATA RECEIVED AND PROCESSED BY APPSFLYER

When a Customer uses the Services, the following End User information may be received and processed by AppsFlyer (collectively, “**End User Data**”).

- i. “**Technical Information**”: this refers to technical information related to an End User’s mobile device or computer, such as: browser type, device type and model, CPU, system language, memory, OS version, Wi-Fi status, time stamp and zone, device motion parameters and carrier.
- ii. “**Technical Identifiers**”: this refers to various unique identifiers that generally only identify a computer, device, browser or Application. For example, IP address (which may also provide general location information), User agent, IDFA (identifier for advertisers), Android ID (in Android devices); Google Advertiser ID, Customer issued user ID and other similar unique identifiers.
- iii. “**Engagement Information**”: this refers to information relating to the Customer’s AppsFlyer use and analytics. For more information, see our big data analytics [AppsFlyer, 04](#); [VentureBeat, 01](#).

- Ok, ok, those won’t be bad!?
 - Attribution, retargeting, immutable Ids?



Google CrashLytics >
Google Firebase Analytics >
Google Tag Manager >

BUT WE'RE NOT SENDING ANYTHING TO CHINA, ANYMORE!

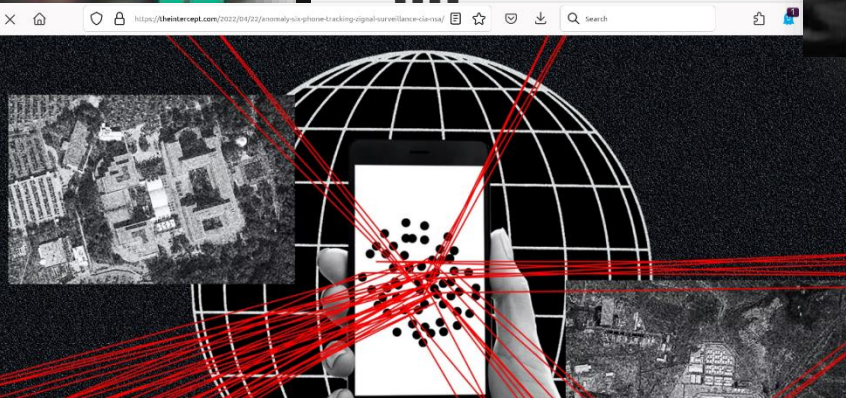
Certainly just bad examples!



Technology
Is your pregnancy app sharing your data with your boss?



PRIVACY
Brokers' sales of



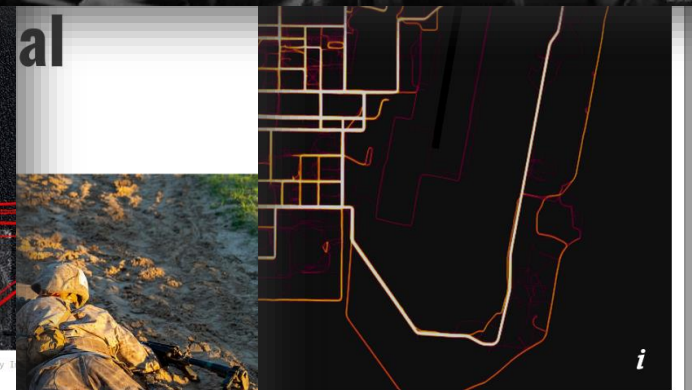
privacy a
Court ruled 3-0 ag
By Colin Lecher | @colinlecher

AMERICAN PHONE-TRACKING FIRM DEMO'D SURVEILLANCE POWERS BY SPYING ON CIA AND NSA

Anomaly Six, a secretive government contractor, claims to monitor the movements of billions of phones around the world and unmask spies with the press of a button.

Sam Biddle, Jack Poulson
April 22 2022, 1:00 p.m.

In partnership with
Tech Inquiry



...ves away location



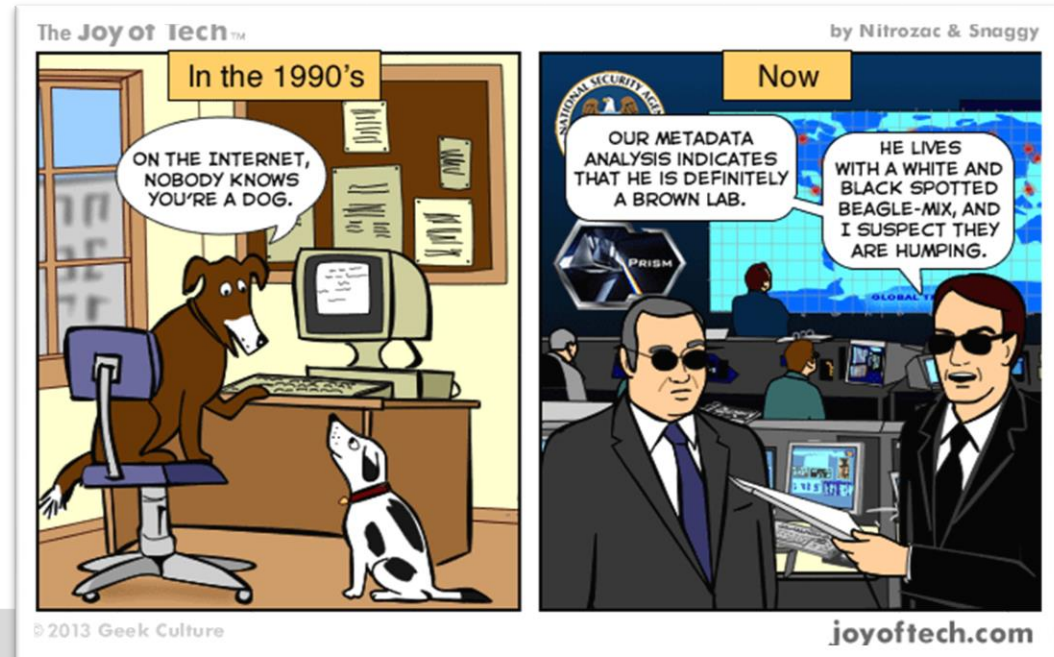
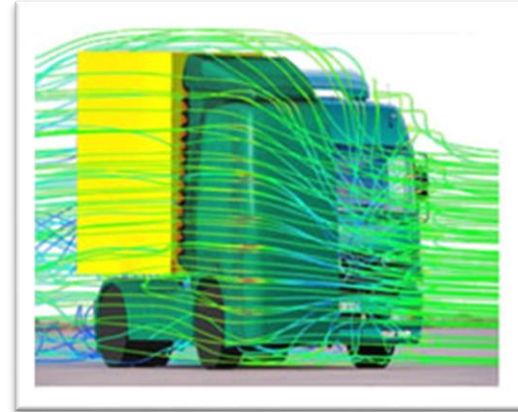
Data Processing and Threats

What's required to process PII?

- GDPR underlines
 - Legally: obtain *individual informed consent* or **anonymize** data
 - Honestly: Individual *informed* consent infeasible
- What does all that mean?
 - anonymos – without calling the name/unnamend
 - pseudonymos – with (any) pretense name (identifiable)
- Anonymity according to the GDPR 101:
 - Information is **anonymous** if it is **not pseudonymous**.
 - A **pseudonym** is any unique piece of **information** *corresponding to an identity (quasi id)*
- Process data in EU and/or of EU citizens: remove anything that makes data linkable to an (even seemingly unknown) individual

Types of Data

- Data without any *relation* to *individuals*
 - Simulation data
 - Measurements from experiments
- Data *with relation to individuals*
 - Types
 - Content
 - Meta data
 - Revelation
 - Consciously
 - Unconsciously



Case Study: Social Media

- Explicit

- Created content
- Comments
- Structural interaction (contacts, likes)



- Inferred

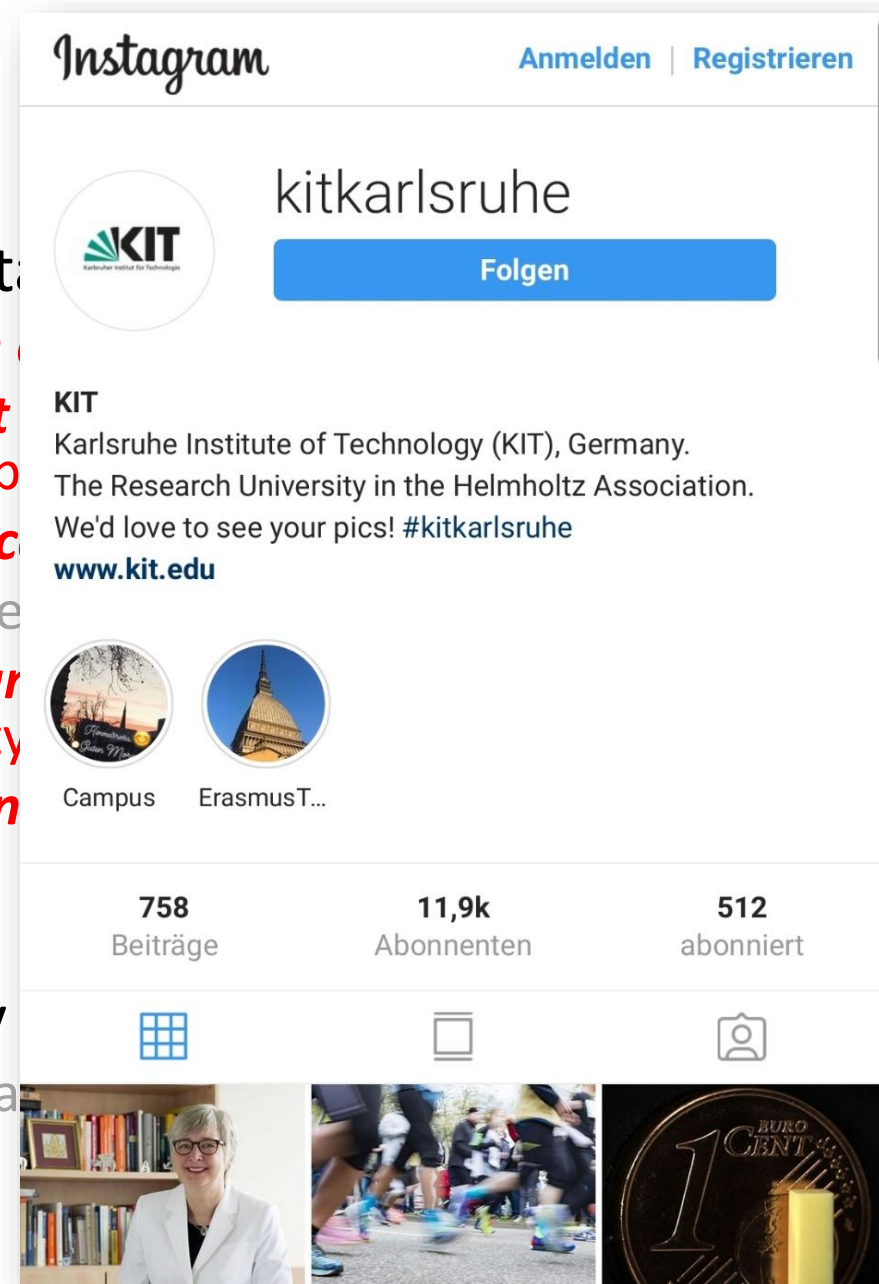
- Preference- and
- Image recognition models
- Personal details

- „Meta data“

- **Session**
- **interest**
- **in group**
- **influenc**
- Clickstre
- **commur**
- **intensity**
- **location**

- Externally

- Observa



Case Study: Social Media

■ Explicit

- Created content
- Comments
- Structural interaction (contacts, likes)



■ Inferred

- Preference– and
- **Image recognition models**
- **Personal details**

■ „Meta data“

- **Session artifacts** (time of actions)
- **interest** (retrieved profiles; membership in groups/participation in discussions)
- **influence**
- Clickstreams, ad preferences
- **communication** (end points, type, intensity, frequency, extent)
- **location** (IP; **shared**; gps coordinates)

■ Externally correlated

- Observation in ad networks

Perceived Adversaries

- *If you asked somebody on the road, they're worried about...*

