Digital contact/presence tracing in FR/Europe: lessons learned after two years

Vincent Roca

Digital contact/presence tracing in FR/Europe: lessons learned after two years

V. Roca for the PRIVATICS team

PRIVASKI seminar, March 8th, 2022

Some background

France: TAC

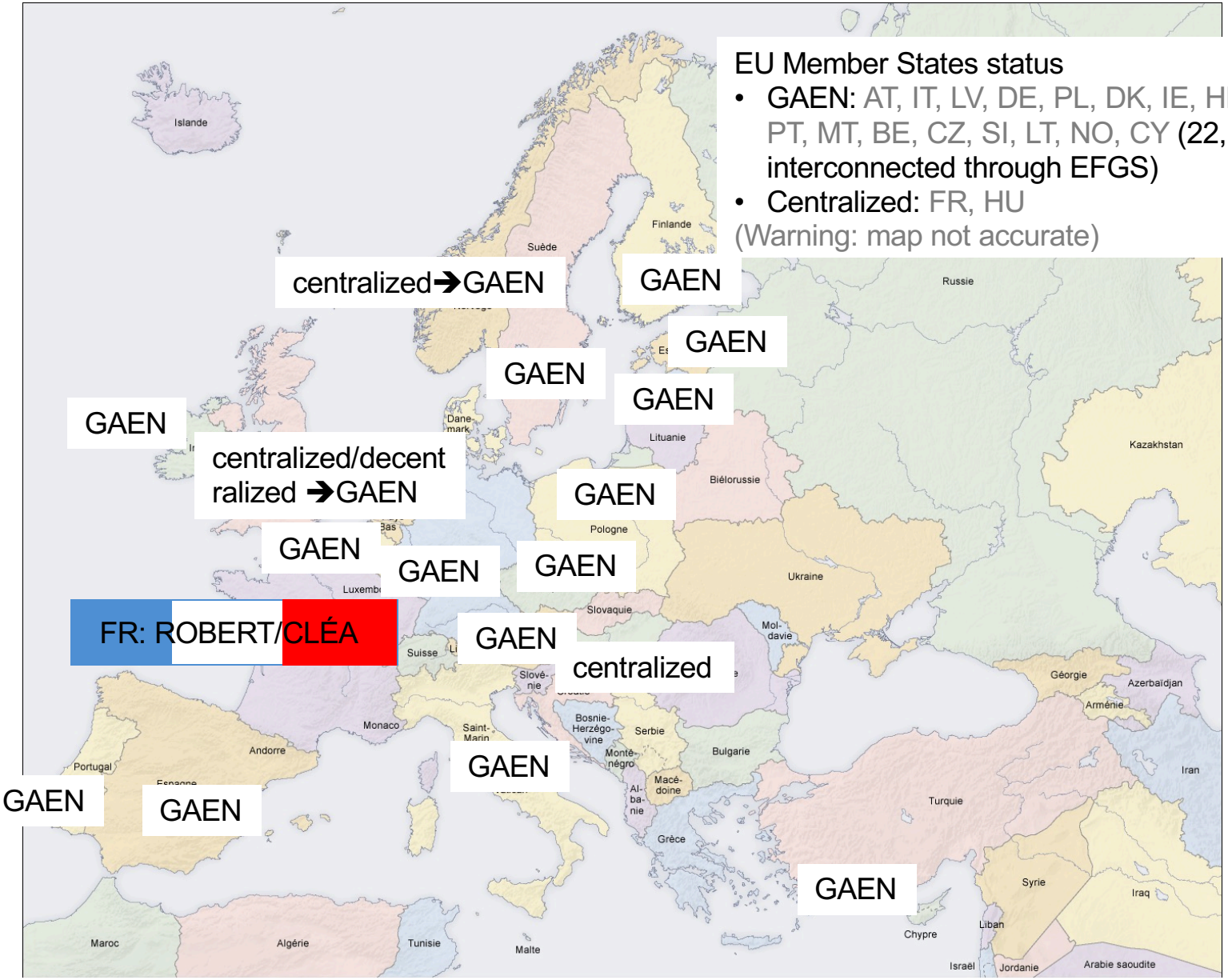
Consortium led by Inria
ROBERT & CLÉA protocols designed by **PRIVATICS**

Rest of the world (almost): GAEN

Google

Exposure Notifications: Help slow the spread of COVID-19, with one step on your phone

The easy, on-the-shelf solution, supposedly at no risk...


EU Member States status
• GAEN: AT, IT, LV, DE, PL, DK, IE, HR, ES, NL, FI, EE, PT, MT, BE, CZ, SI, LT, NO, CY (22, of which 19 are interconnected through EFGS)
• Centralized: FR, HU
(Warning: map not accurate)
centralized➔GAEN
GAEN
GAEN
GAEN
GAEN
GAEN
centralized/decentralized ➔GAEN
GAEN
GAEN
GAEN
GAEN
FR: ROBERT/CLÉA
GAEN
centralized
GAEN
GAEN
GAEN
GAEN
Islande
Suède
Finlande
Russie
Danemark
Lituanie
Biélorussie
Kazakhstan
Pologne
Ukraine
Slovaquie
Moldavie
Suisse
Slovénie
Géorgie
Azerbaïdjan
Arménie
Monaco
Saint-Marin
Bosnie-Herzégovine
Serbie
Monténégro
Bulgarie
Andorre
Portugal
Espagne
Albanie
Macédoine
Iran
Turquie
Grèce
Syrie
Iraq
Liban
Chypre
Israël
Jordanie
Arabie saoudite
Maroc
Algérie
Tunisie
Malte

Contact… or presence tracing?

Contact Tracing: did you **met** one or more COVID+ persons sufficiently close, sufficiently long?

➔ **ROBERT** (FR), **GAEN**

Presence Tracing: have you been in a **location** where one or more co-located people were later tested COVID+?

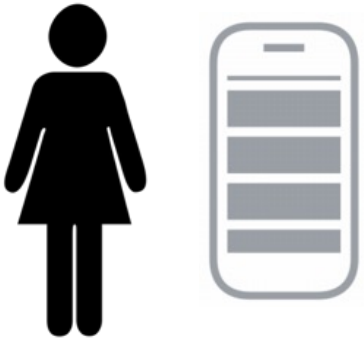
➔ **CLÉA** (FR), LUCA and CWA Event Registration (DE), CrowdNotifier (Swiss)…

(in future?)

Contact… or presence tracing? (2)

Step 1: remember…

Alice



CT: keep a 14-days history of **contacts**
{contact pseudo, time, distance}

PT: keep a 14-days history of **locations**
{QRcode (includes location pseudo), time}

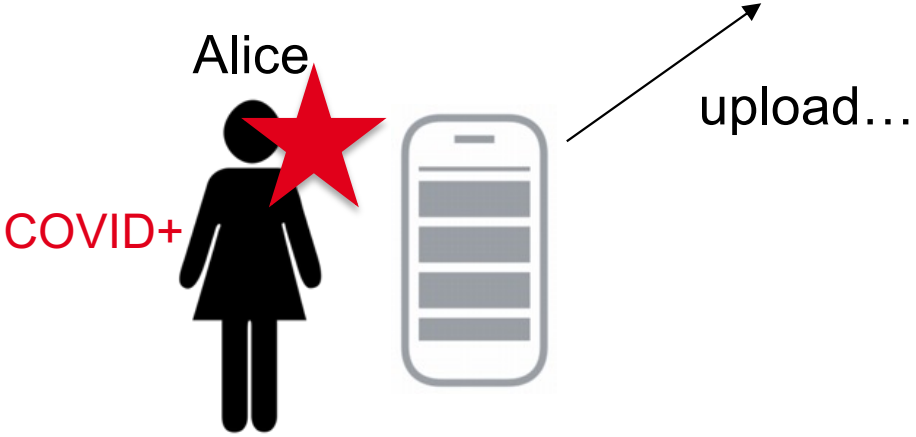
Step 2: check risk…

decentralized: **locally** (**CLÉA** for PT and **GAEN** for CT)

centralized: by polling a **server** that computes risks (**ROBERT** for CT)

Contact… or presence tracing? (3)

Step 3: share if COVID+…



if Alice agreed, upload to server :

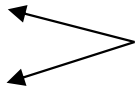
GAEN for CT: Alice pseudos (only)

CLÉA for PT: scanned QR codes (only)

ROBERT for CT: contact list (only)

Contact… or presence tracing? (4)

- different **epidemiologic** assumptions
  - direct virus transmission (close to a COVID+ person), or
  - virus remains in a location in the air for some time…



Q: where's the main risk?

- different **technological** requirements
  - QRcode scanning is less demanding than BLE

Forward tracing or backward tracing?

Assume Alice is tested COVID+…

Contact Tracing →

- **Forward:** who might have been contaminated by Alice when contagious?
  - ➔ inform her contacts

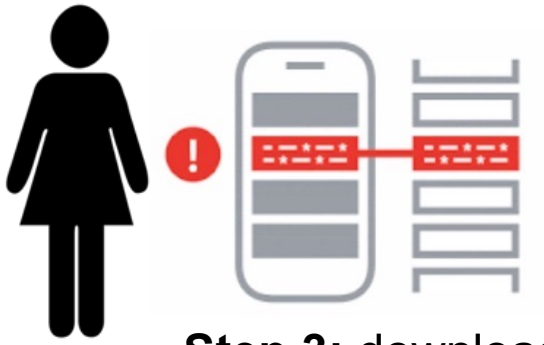
Presence Tracing (does both ☺)

- **Backward:** on what occasion/where could Alice be contaminated?
  - ➔ inform people who were there
  - ➔ (potentially) identify who is key index, inform him (was he asymptomatic?) and his contacts

Let’s talk about threats and privacy

Decentralized risk analysis requires sharing data ☹

Step 1: server collects (GAEN) or computes (CLÉA) a black list

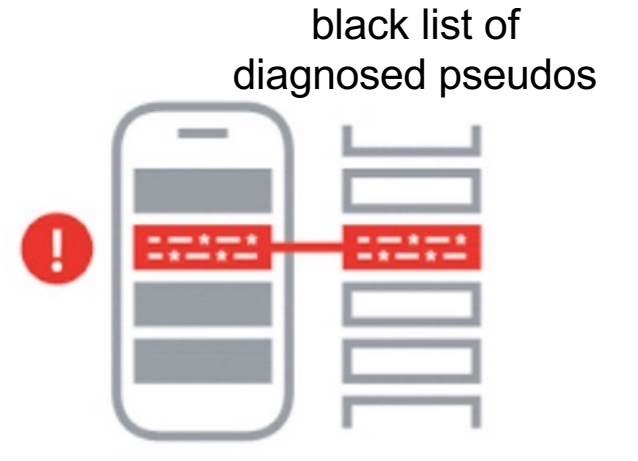


Step 2: distribute publicly the black list through a CDN (w/o access control)

Step 3: download the black list and compare with local history

Key question: nature of information in the black list?

Decentralized CT threat and privacy



Decentralized risk analysis (e.g., **GAEN**) requires a smartphone to know pseudonyms of all COVID+ users

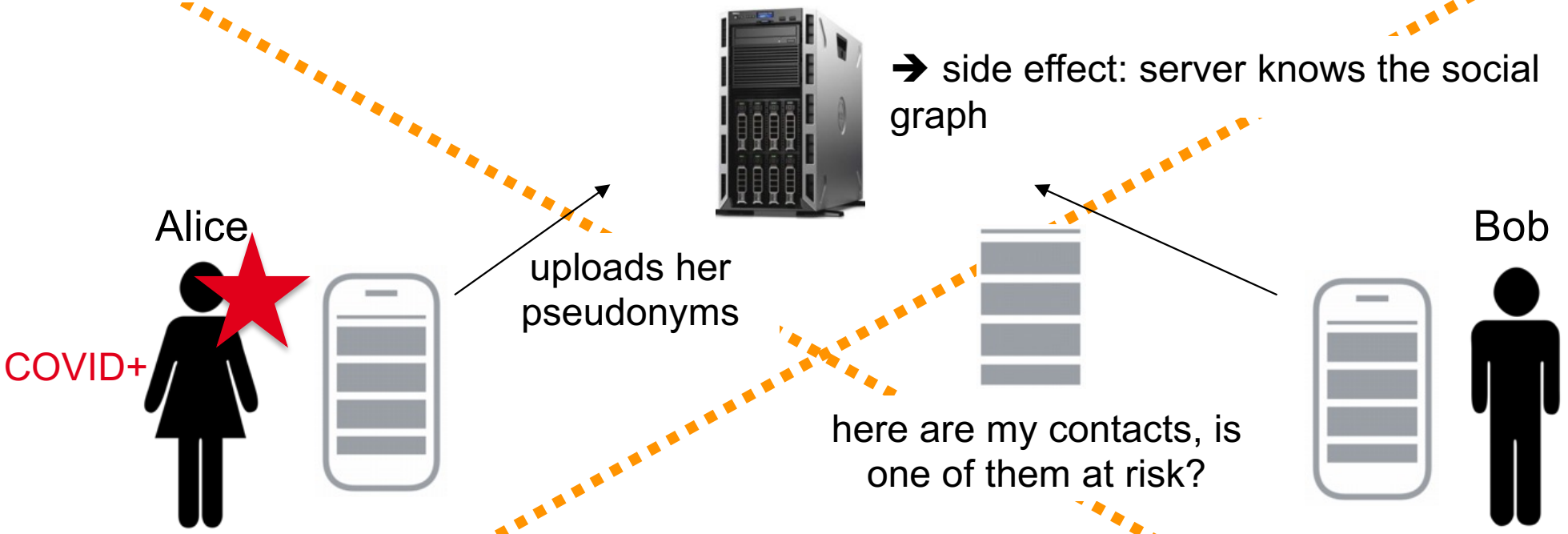
↓

The GAEN black list of “diagnosed pseudos” is **public**
(see https://hal.inria.fr/hal-02899412).
No effective and scalable way to add access control

↓

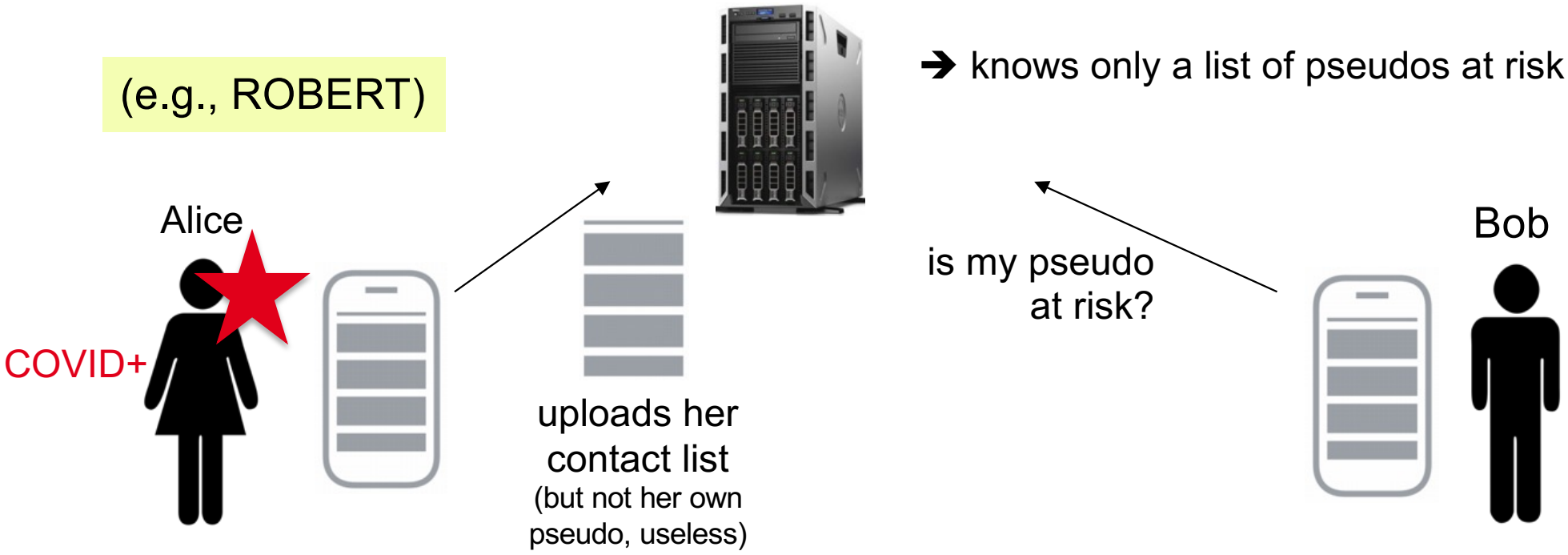
Not GDPR friendly… and creates discrimination risks
(see https://coronadetective.eu)

There are bad centralized CT designs ☹



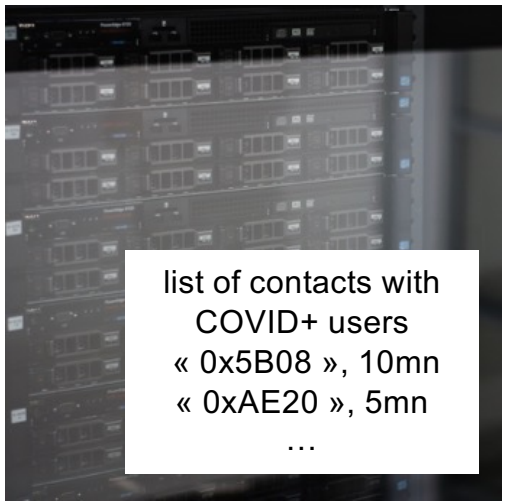
- Problem: Bob exposes his full contact list to the server ☹

But there are good centralized CT designs ☺



- Server only knows an unordered list of pseudonyms at risk, nothing else
- Bob does not reveal anything to the server (idem GAEN ☺)

Centralized CT threat and privacy



Centralized risk analysis (e.g., **ROBERT**) requires the server to know the contacts of COVID+ users

↓

No public black list (neither “diagnosed pseudos” nor “at risk” pseudos)

↓

GDPR compliant ☺

↓

Yet there are **major** assumptions regarding the threat model!

About CT threat and privacy modesl

threat model

```
if ((you can trust the institutions of your country)
 && (the use of the app is optional)
 && (a trusted 3rd party audits the whole system)) {
    /* health authority has more control, no sensitive
     * data shared publicly, GDPR compliant, safe */
    use a centralized CT;
} else {
    /* risk of abuse cannot be mitigated,
     * GDPR non-compliancy is perhaps secondary */
    use a decentralized CT;
}
```

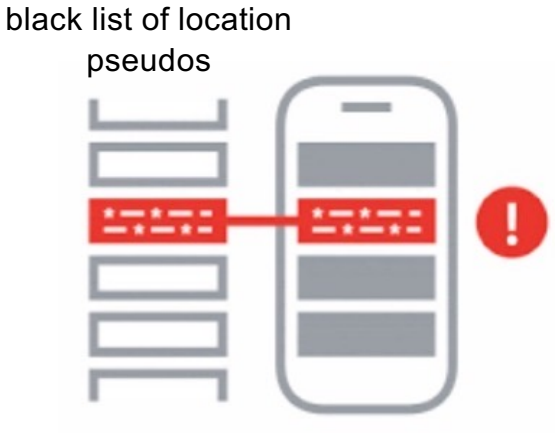
privacy model

Decentralized Presence Tracing threat and privacy

Decentralized risk analysis (e.g., **CLÉA**) requires a smartphone to know locations at risk (pseudo)

The CLÉA black list of “location pseudos” is **public**

It’s not sensitive health data ☺ and all locations Bob visited remain local (unless COVID+)



Centralized PT threat and privacy models (2)

Centralized risk analysis **LUCA** collects name/phone number of participants to an event for the HA to contact them by phone if needed

↓

Potentially helpful from an **epidemiologic** viewpoint…

↓

…but risky from a **privacy** viewpoint.

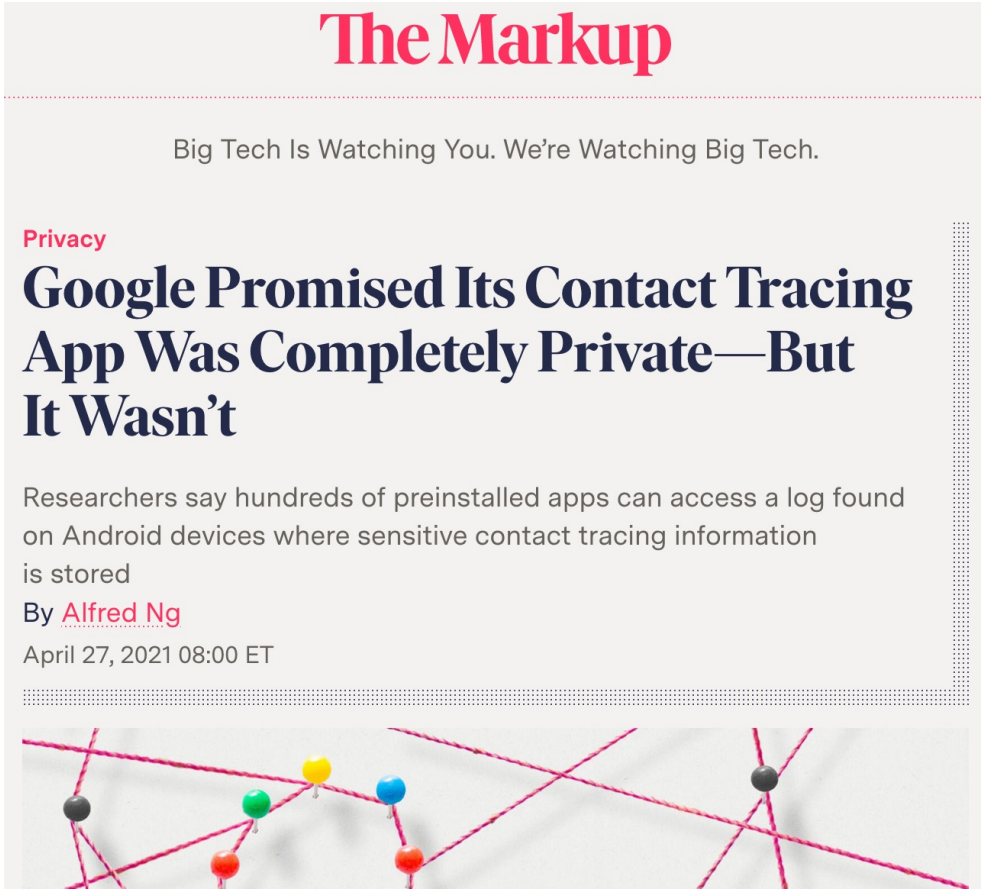
(see: abuse of LUCA by police to find witnesses 2022-01)

On-the-shelf Google's GAEN must be audited too!

- **None** of the EU Members States checked GAEN security sufficiently
  - **data breach** during > 1 year
    - 400 pre-installed apps had access to everything
  - **no fix** during 2 months after AppCensus responsible discl. whereas the fix is trivial

https://themarkup.org/privacy/2021/04/27/google-promised-its-contact-tracing-app-was-completely-private-but-it-wasnt

https://blog.appcensus.io/2021/04/27/why-google-should-stop-logging-contact-tracing-data/



The Markup

Big Tech Is Watching You. We're Watching Big Tech.

Privacy

Google Promised Its Contact Tracing App Was Completely Private—But It Wasn't

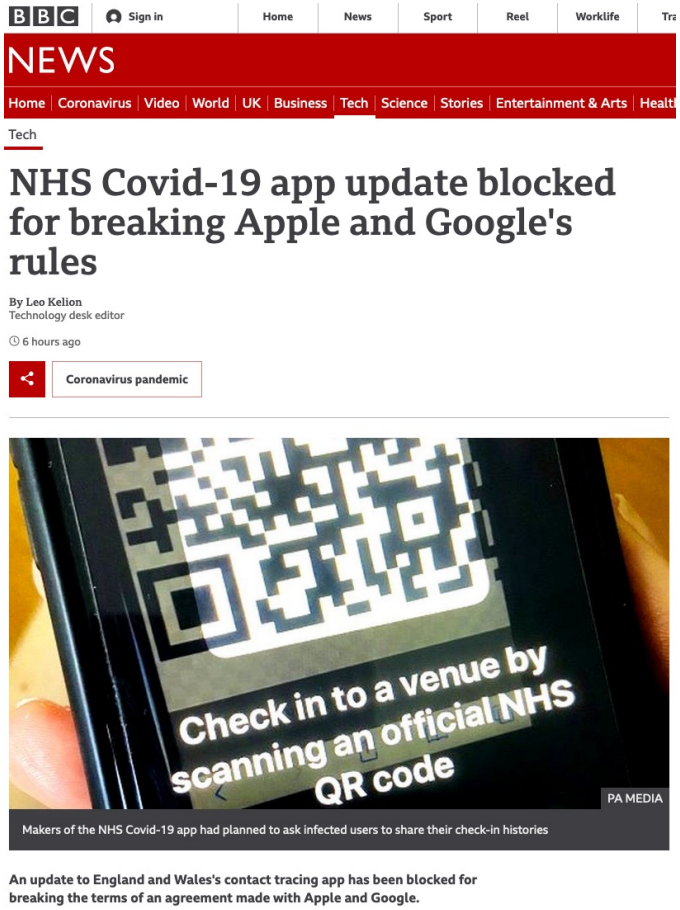
Researchers say hundreds of preinstalled apps can access a log found on Android devices where sensitive contact tracing information is stored

By Alfred Ng

April 27, 2021 08:00 ET

Sovereignty

- What place for the national Health Authority?
  - at the center vs. periphery
- Who should decide what features are authorized?
  - should G&A decide in place of democratic gov?
- Don’t you think the OS be neutral?
  - iOS restrictions on use of BLE in background make CT very hard. Apple always refused to white list the FR app. It had epidemiologic consequences. Normal?



https://www.bbc.com/news/technology-56713017

« if we're faced with any kind of trade-off one of the guiding insights that we've used through this process is that **social graph is more sensitive and more privacy risky than infection status data**.»

G. Hogben (Dir. privacy Android) Oct. 2020: https://www.youtube.com/watch?v=0gqZJXOO9Ko [offset 3:35]

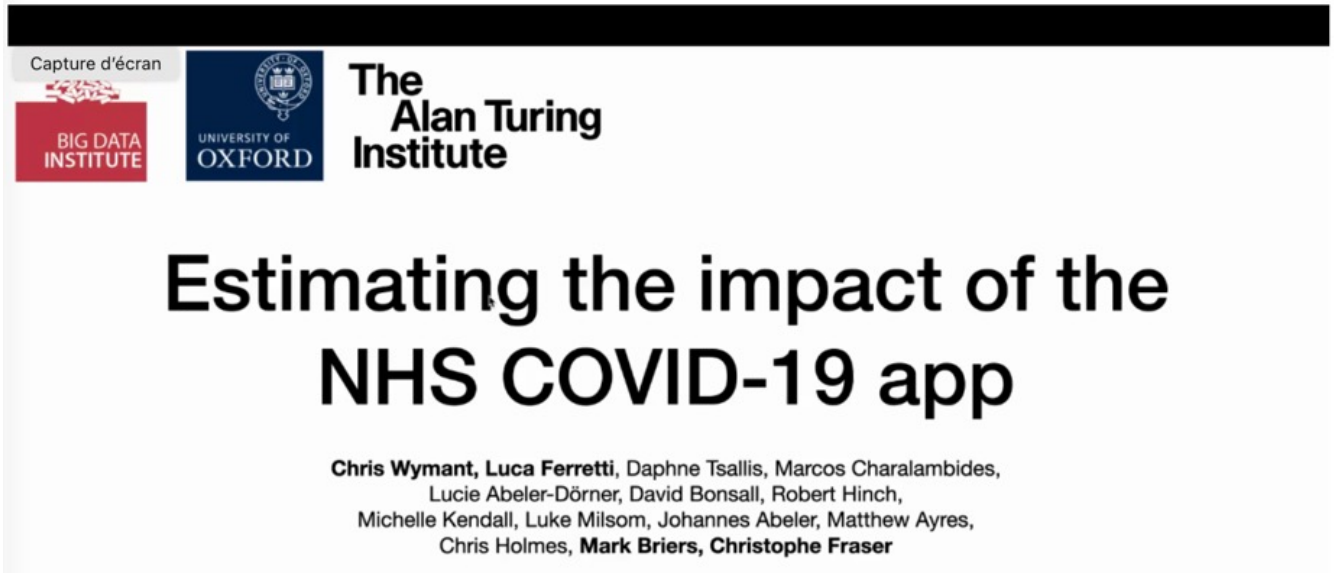
- Who should decide that GDPR compliance is of lesser importance? G&A?
  - CNIL preferred ROBERT to GAEN
  - what would have happened if this key question had been discussed in April 2020?
  - why does the "joint statement on CT" (April 19th, 2020) totally omit this trade-off?
- No longer valid in case of higher death rates
  - GAEN prohibited

Did it work?

Epidemiologic and technical efficiency

Epidemiologic: Yes, CT worked in UK

- In March 2020 there were suspicions it could work (simulations)
- One year later [1] proved CT **saved 4200 - 8700 lives** in UK in Sept.-Dec. 2020
  - to compare to the 32 500 deaths during period
  - but **~28%** of regular users within population



[1] "The epidemiological impact of the NHS COVID-19 app", Nature, Vol 594, 17 June 2021. https://www.nature.com/articles/s41586-021-03606-z

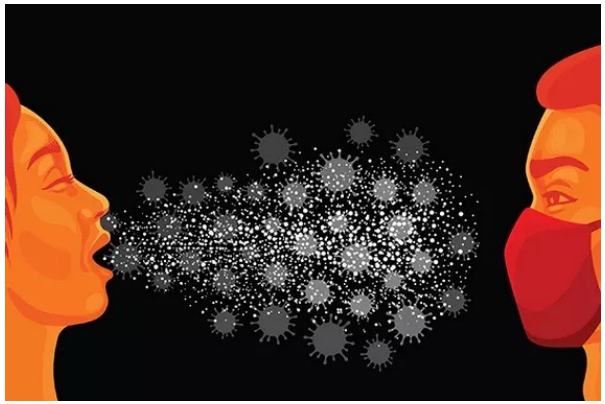
Epidemiologic: In FR it’s hard to conclude

- Missing a sound epidemiologic study!
  - Usage analytics since June 2021 only (required a decree, published Feb. 2021)
  - The existing efficiency evaluation document is limited in its methodology ☹
  - Raises the question of evaluation-by-design…

Epidemiologic: no clue for Presence Tracing

- No evaluation as far as I know
- In FR, after one month, use of PT compromised by generalization of sanitary pass in bars/restaurants/sport centers

- Important to know since CT and PT rely on different virus spreading models… Which one is valid?

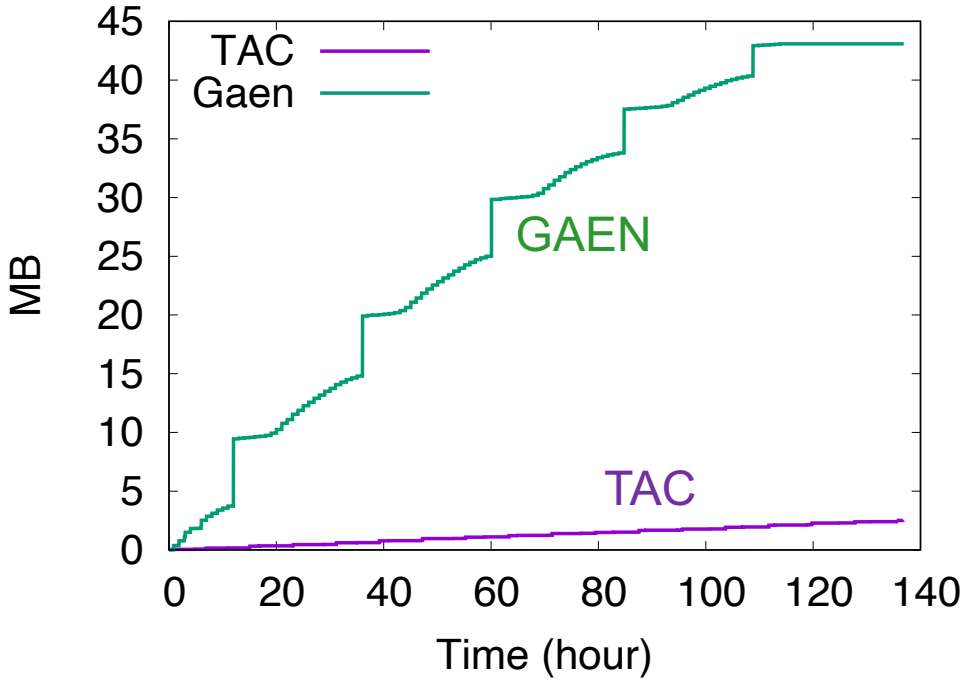


wearing masks or not?

Technical: Centralized/ROBERT more efficient

- **Scalability** in peak contamination phases
  - one of the 3 goals of Digital CT
  - ROBERT scales better (computation on server + no need to download a black list)

download traffic over days TAC vs. GAEN (CWA DE app)
NB: collected Feb. 17-22nd, 2022 during Omicron wave, keeping only traffic from TAC (resp. CWA) servers



Technical: Centralized/ROBERT more efficient (2)

- Being centralized enables **symmetric cryptography**… prohibited with GAEN
  - keep a shared key on the server per registered app
  - enables efficient use of 16 bytes of payload for BLE messages

  ➔ ROBERT BLE messages have (limited) **anti-replay** + encrypted **country** code

  ➔ ROBERT avoids the “single world” syndrome of the pan-European interconnection of GAEN apps (EFGS service): each app downloads all diagnosed keys of all EU countries

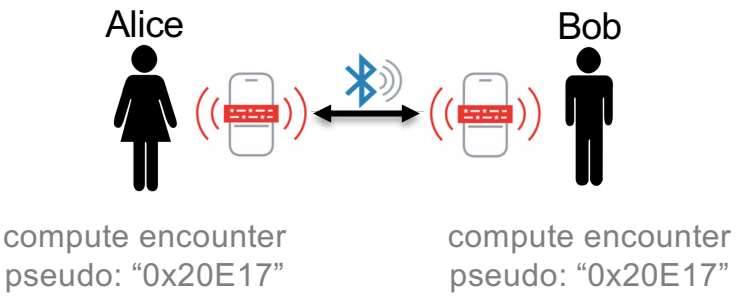
Takeout

- There's **no single universal solution**… Depends on:
  - the threat model WRT institutions versus Google/Apple
  - the threat model WRT data that decentralized systems share
- GAEN acceptable for COVID… But not with a virus with **higher death rate**
  - GAEN not being GDPR compliant is an issue… see: www.coronadetective.eu
- Result is anyway a trade-off as different stakeholders have different expectations
- Digital CT proved to be **efficient** (UK / GAEN / 28% of usage)
  - essentially a matter of user acceptance + a few tricks (easy upload, etc.)
  - however efficiency is complex question… What about **evaluation-by-design**?
- Centralized scales better and enables technical optimizations
- Acceptance raises different questions… ask **sociologists**, not only privacy experts

DÉSIRÉ 3rd way CT: “identify encounters, not devices”

A paradigm shift: from “public device pseudonyms” (GAEN, ROBERT) to “private encounter pseudonyms”:

- that **change** across time and devices
- are **private** to users who met (eavesdroppers powerless)



And major benefits:

- better **privacy** protection
- highly **flexible**: centralized or decentralized risk evaluation, at a country own discretion
- full **interoperability** across different deployments (≠ GAEN or ROBERT)

“DESIRE: Leveraging the best of centralized and decentralized contact tracing systems”, ACM DTRAP, 2021.
https://hal.inria.fr/hal-03476799/en

Thank you…

vincent.roca@inria.fr