On Bluetooth-Based Contact-Tracing Smartphone Applications: Principles and Controversies

Léo Perrin

On Bluetooth-Based Contact-Tracing Smartphone Applications

Principles and Controversies

Léo Perrin
@lpp_crypto

2nd of June, 2020
GdT Maths4covid19

In this talk

"Refusing digital contact tracing is accepting more death"

Cédric O (paraphrasing)

In this talk

"Refusing digital contact tracing is accepting more death"

Cédric O (paraphrasing)

… but then why is everyone at each other's throat?

In this talk

"Refusing digital contact tracing is accepting more death"

Cédric O (paraphrasing)

... but then why is everyone at each other's throat?

How is a bluetooth-based contact-tracing application supposed to work?

Why is this topic controversial?

In this talk

"Refusing digital contact tracing is accepting more death"

Cédric O (paraphrasing)

... but then why is everyone at each other's throat?

How is a bluetooth-based contact-tracing application supposed to work? **Theory**

Why is this topic controversial? **Practice**

First Things First

- I work at **Inria**, which is in charge of the development of `StopCovid`...
- ... but I am not involved in this project, I don't have any "insider knowledge".

First Things First

- I work at **Inria**, which is in charge of the development of `StopCovid`...
- ... but I am not involved in this project, I don't have any "insider knowledge".
- The only way to assess the security level of a protocol/algorithm is to try and identify the ways it could be abused. Such an analysis is "standard procedure" in computer security research.
- I am a co-author of `https://risques-tracage.fr`

Outline

Outline of this Section

Outline of this Section

What is Contact-Tracing?

To prevent the spread of the disease, we must isolate the sick and those they have been in contact with.

Definition (Contact)

(here) An interaction that may have caused a contamination.

It can be done “manually” (by people) or it can be done using digital tools, in particular smartphone applications.

Contact Tracing Application

The Oxford Study (principle)

A study from Oxford[1] is being quoted a lot. It studies the impact of 3 quantities on the speed of spread of the virus.

Efficiency of isolation: ($\epsilon_I \in [0,1]$) quantifies how well/how many people isolate when they develop symptoms.

Efficiency of contact tracing: ($\epsilon_T \in [0,1]$) quantifies how much of an impact contact tracing has.

Days to isolation and contact quarantine: (t) time between the appearance of the first symptoms and the implementation of counter-measures.

[1] *Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing*, Ferretti et al (2020). `https://doi.org/10.1101/2020.03.08.20032946`

The Oxford Study (results 1/2)

x-axis

The efficiency of isolation ϵ_I.

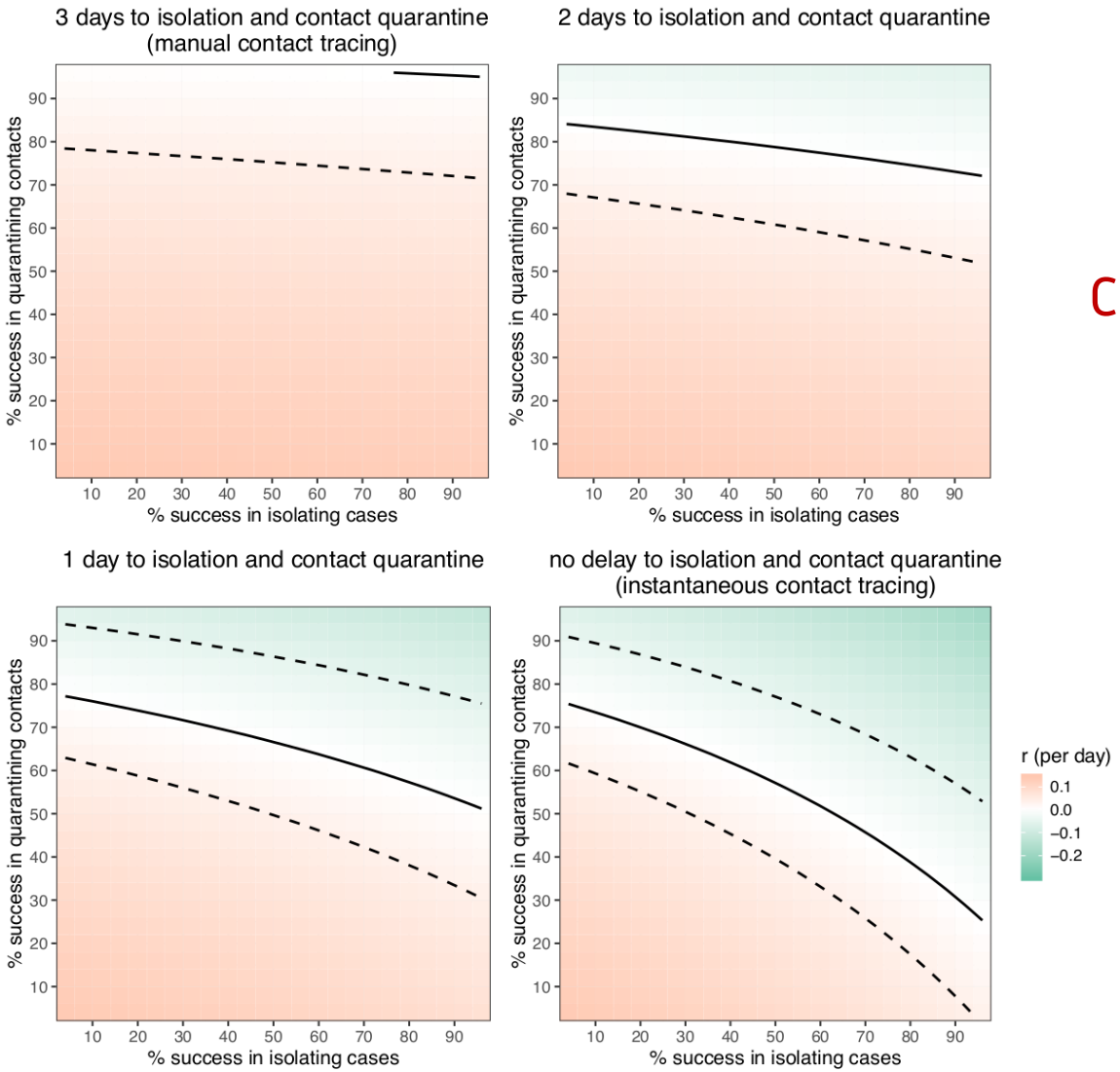
y-axis

The efficiency of contact tracing

$$\epsilon_T = U^2 \times D \times c$$

- U: proportion of app users,
- D: proportion of contacts successfully detected,
- c: fractional reduction in infectiousness resulting from being notified as a contact.

The Oxford Study (results 2/2)



Conclusions

- ϵ_T must be very high
- t must be very small
- "manual" contact tracing cannot keep up
- Some automation is necessary!

Scope Statement of an Application

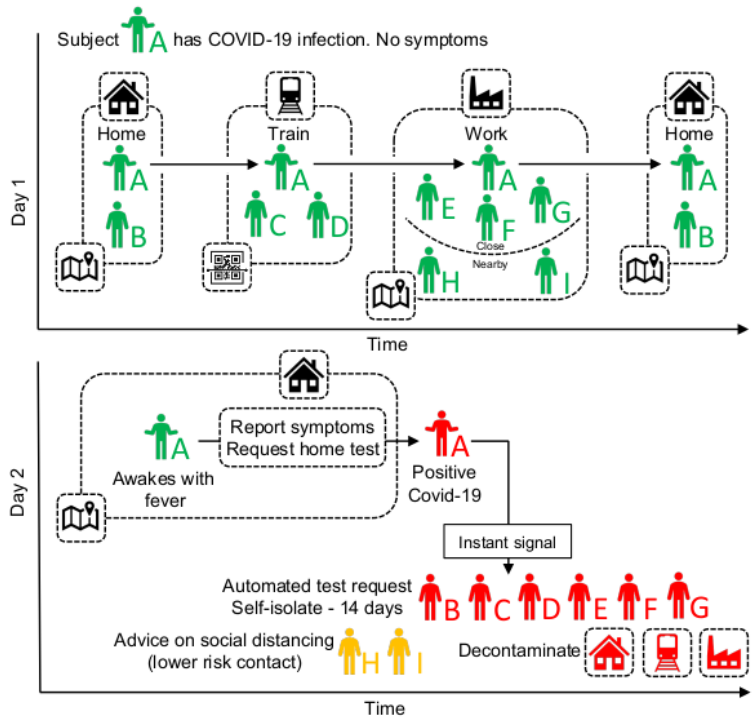


Fig. 4. A schematic of app-based COVID-19 contact tracing. Contacts of individual A (and all individuals using the app) are traced using GPS co-localisations with other App users, supplemented by scanning QR-codes displayed on high-traffic public amenities where GPS is too coarse. Individual A requests a SARS-COV-2 test (using the app) and their positive test result triggers an instant notification to individuals who have been in close contact. The App advises isolation for the case (individual A) and quarantine of their contacts.

1. GPS-based
2. With added “check-in” in “high-traffic public amenities”
3. Test is made at home, very quickly

Outline of this Section

Gone with the GPS

Basic civil liberties considerations impose modifications of this scope statement:

- users must be convinced to use the tracking device,
- GPS is usually[2] put aside: few would agree to be litterally traced by the state,
- ease of use/privacy $\Longrightarrow$ no QR codes in high traffic areas.

[2] Iceland, a democracy, decided to use it anyway.

Bluetooth?

In practice, digital CTs under consideration in Europe only track physical proximity of people (not time delayed surface contacts).

Bluetooth?

In practice, digital CTs under consideration in Europe only track physical proximity of people (not time delayed surface contacts).

Bluetooth

It is a wireless technology to exchange information over “short” distances (1 to 100m). The bluetooth standard is a Rube Goldberg machine with *many* subsections...

Bluetooth?

In practice, digital CTs under consideration in Europe only track physical proximity of people (not time delayed surface contacts).

Bluetooth

It is a wireless technology to exchange information over "short" distances (1 to 100m). The bluetooth standard is a Rube Goldberg machine with *many* subsections...

Definition (BLE)

Bluetooth Low Energy is a variant of the Bluetooth protocol aimed at minimizing energy consumption. On smartphones, it runs in the background even when bluetooth is de-activated.

Bluetooth-based Contact Tracing

1. Each device has a long term “pseudonym” in order for users to be *pseudonymous* ($\neq$ anonymous).

2. Each device has a set of short-term “crypto-identifiers” that are broadcast for 15 min.

Bluetooth-based Contact Tracing

1. Each device has a long term “pseudonym” in order for users to be *pseudonymous* ($\neq$ anonymous).
2. Each device has a set of short-term “crypto-identifiers” that are broadcast for 15 min.
3. Other devices in the vicinity receive these crypto-identifiers.

Bluetooth-based Contact Tracing

1. Each device has a long term "pseudonym" in order for users to be *pseudonymous* ($\neq$ anonymous).
2. Each device has a set of short-term "crypto-identifiers" that are broadcast for 15 min.
3. Other devices in the vicinity receive these crypto-identifiers.
4. If a user later turns out to have the COVID-19, then those who received one of his crypto-identifiers are potential contacts and must self isolate.

Bluetooth-based Contact Tracing

1. Each device has a long term "pseudonym" in order for users to be *pseudonymous* ($\neq$ anonymous).
2. Each device has a set of short-term "crypto-identifiers" that are broadcast for 15 min.
3. Other devices in the vicinity receive these crypto-identifiers.
4. If a user later turns out to have the COVID-19, then those who received one of his crypto-identifiers are potential contacts and must self isolate.

How long and how close does a contact need to be so as to be potentially dangerous?

False Positives

Bluetooth:

- can go through walls,
- does not care if a mask is worn, and
- was never intended to measure distances.[3]

A CTA is bound to have false positives.

Such false positives do not matter as far as pandemic containment is concerned... but they may damage everyday lives/the economy. They are not accounted for in the Oxford study.

[3] For distance, we also must assume that the COVID-19 models are correct.

False Negatives

Bluetooth:

- was never intended to measure distances,
- cannot figure out if a potentially contaminated surface was touched.

A CTA is bound to have false negatives.

This matters for pandemic containment: if there are too many false negatives then CTA will be useless. This effect is captured by D in the Oxford model.

Outline of this Section

No Electronic Monitoring Bracelet

At this stage, electronic bracelets are not available (also, PR...)

Instead, digital CT will rely on a very common tracking device: smartphones.

Distance measurement with Bluetooth

The further you are, the weaker the signal $\Longrightarrow$ we can use signal intensity to estimate distance...

Distance measurement with Bluetooth

The further you are, the weaker the signal $\Longrightarrow$ we can use signal intensity to estimate distance...

... in theory. In practice, chipsets/phones will produce/detect signals with varying efficiencies.

Apparently, it kinda works (though it is held together with virtual duct tape).

Implementation Issues

On smartphones, access to BLE is restricted: the operating system needs to allow it.

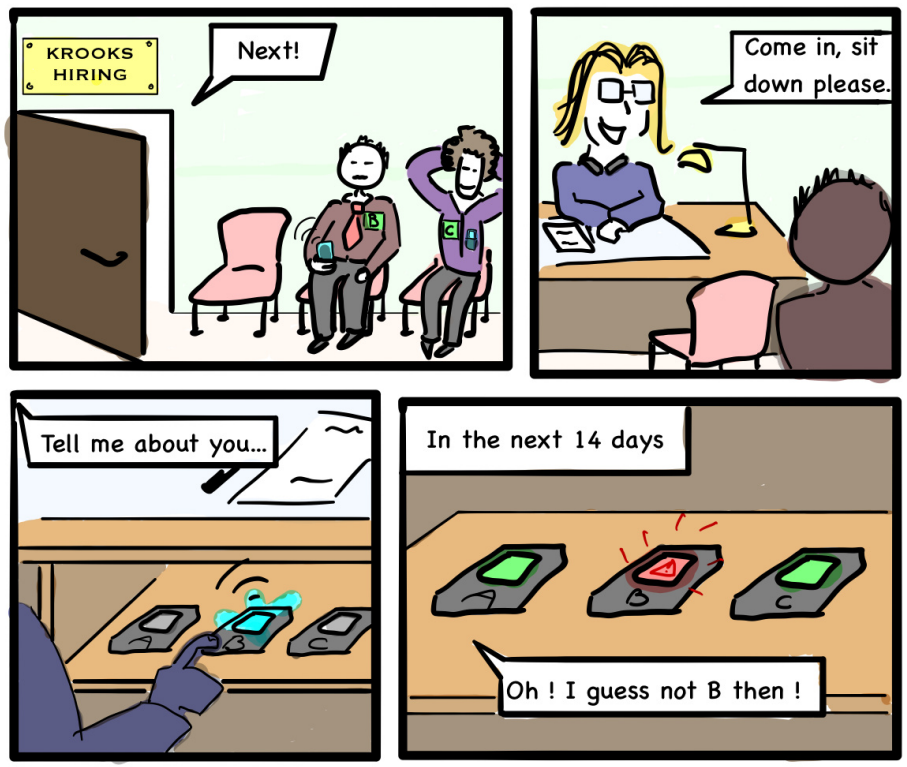
If Google or Apple wants to discourage the use of a specific CTA on their smartphone, they can (by denying it access to their application store).

If they want to ease the implementation of a specific app, they can (by providing highly optimized libraries in the operating system itself).

Outline of this Section

Outline of this Section

"Alternative" Uses of a CTA



https://www.risques-tracage.fr/

Tracking Users

Since CTAs “yell” identifiers, we can detect their users easily.

A tool already exists for `StopCovid` !

`https://github.com/rgrunbla/Stop_Covid_Detector_3000`

Single Use Instance

1. Create a new CTA instance (i.e. set it up on an otherwise unused phone)
2. Wait until you are close to your target only (e.g. alone in a room with them)
3. Once the contact was registered by the app, turn it off and do not let it be in contact with anyone else.
4. If you get a notification, you know that your target is sick.

Single Use Instance

1. Create a new CTA instance (i.e. set it up on an otherwise unused phone)
2. Wait until you are close to your target only (e.g. alone in a room with them)
3. Once the contact was registered by the app, turn it off and do not let it be in contact with anyone else.
4. If you get a notification, you know that your target is sick.

In practice, you do not need a "physical" smartphone for each target, you could emulate one with dedicated software.

Forced Isolations

1. Hide a smartphone somewhere where it will be in "contact" with the phones of your target(s),
2. obtain a COVID-19 diagnostic (bribery/extorsion/hacking...) for this phone,
3. all your targets have to self isolate!

This can be used to close schools (hiding the phone in a teachers' room), factories (by the coffee machine/locker room), to get a competitor out of the way...

Faking Diagnoses

In the Oxford study, they consider that a preliminary diagnostic is done via an in-app test—and its result is broadcast to the relevant contacts. The alert is lifted if the actual test is negative.

Exercice for the viewer

How would you abuse such a mechanism?

Faking Diagnoses

In the Oxford study, they consider that a preliminary diagnostic is done via an in-app test—and its result is broadcast to the relevant contacts. The alert is lifted if the actual test is negative.

Exercice for the viewer

How would you abuse such a mechanism?

Crucial question

More generally, how safe is the infrastructure that handles diagnoses?

Third Party Applications

Bluetooth-based CTAs do not use GPS. But.

What prevents an ill-intentioned programmer from creating another app which:

1. receives crypto-identifiers,
2. associate them to GPS data/the output of a camera, or prompts the user to give the identity of the person (if they know them),
3. stores/somehow uses the result

without anyone (other than its users) noticing it?

Third Party Applications

Bluetooth-based CTAs do not use GPS. But.

What prevents an ill-intentioned programmer from creating another app which:

1. receives crypto-identifiers,
2. associate them to GPS data/the output of a camera, or prompts the user to give the identity of the person (if they know them),
3. stores/somehow uses the result

without anyone (other than its users) noticing it?

Answer:

Third Party Applications

Bluetooth-based CTAs do not use GPS. But.

What prevents an ill-intentioned programmer from creating another app which:

1. receives crypto-identifiers,
2. associate them to GPS data/the output of a camera, or prompts the user to give the identity of the person (if they know them),
3. stores/somehow uses the result

without anyone (other than its users) noticing it?

Answer: Nothing.

Replay Attacks

How to sell positive diagnoses on the black market?

1. have the buyer send you their upcoming identifiers,
2. set up a long distance antenna with falsified distance information next to a testing center. Simulate contacts between everyone going there and your buyer's identifiers.
3. next time the buyer's app checks if he is at risk, he will be if anyone in the testing center was positive.

The specifics of the attack depend on the protocol used but the principle is the same.

How to Track a Specific Device

CTAs **can** be used to physically track someone. The Bluetooth chipset has a MAC address which changes over time. If the change of MAC is not perfectly synchronized with the change of crypto-identifier, then we can figure out that a sequence of crypto-identifiers corresponds to a unique person.

General vs. Specific Attacks

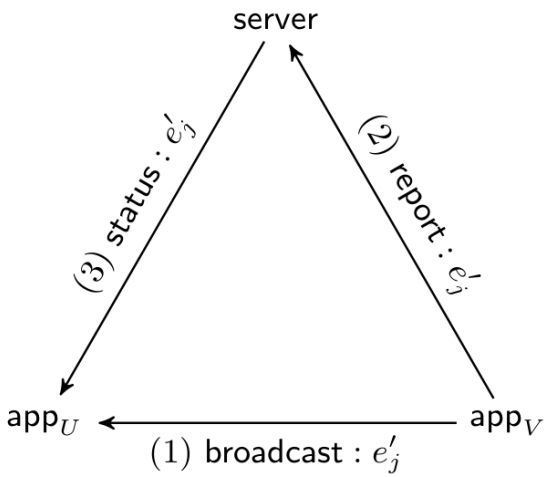
These attacks work on all Bluetooth-based CTAs!

To see other attacks, we need to look at the specifics of the different approaches considered.

“decentralized” vs. “centralized”

Outline of this Section

"Decentralized"?



source: "Centralized or Decentralized? The Contact Tracing Dilemma", Vaudenay (2020).

`https://eprint.iacr.org/2020/531`

1. Each user generates crypto-identifiers e'_i that they broadcast.
2. If a user becomes sick, they send all the crypto-identifiers they generated to a central server which adds them to their list of "sick identifiers".
3. Those who received "sick" identifiers now know they are "at risk".

Motivation

The central server knows very little.

Even if the actor running the central server is ill-intentioned, there is not much that they can do to harm/de-anonymise the users.

Topic of an international petition: "Joint Statement on Contact Tracing: Date 19th April 2020"[4]

[4] https://www.esat.kuleuven.be/cosic/sites/contact-tracing-joint-statement/

A Specific Attack (1/2)

The temporary crypto-identifiers of all infected people are public.

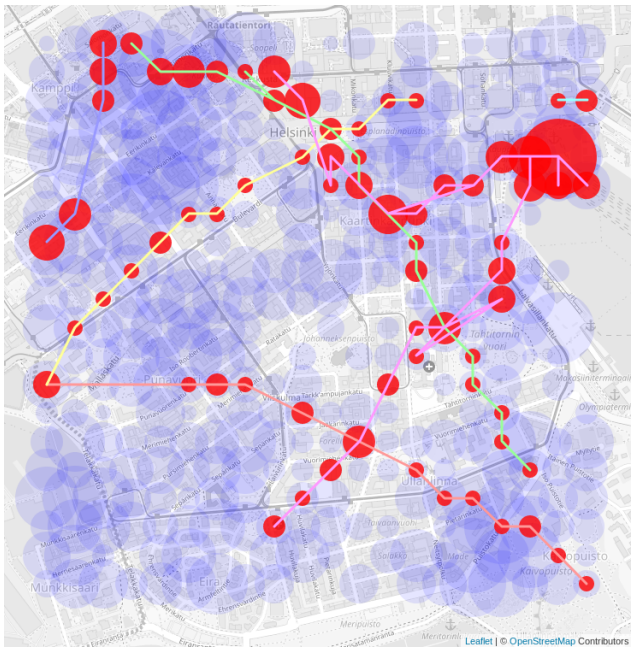
As a consequence, anyone capable of de-anonymising a person can figure out if they are sick!

1. Meet someone you can identify (you actually know them, they used a credit card in your shop...);
2. store their temporary crypto-identifier when you are close to them;
3. if said crypto-identifier shows up, you know that specific person is infected!

This could be scaled up, e.g. a supermarket could place bluetooth receivers at all the checkout desks.

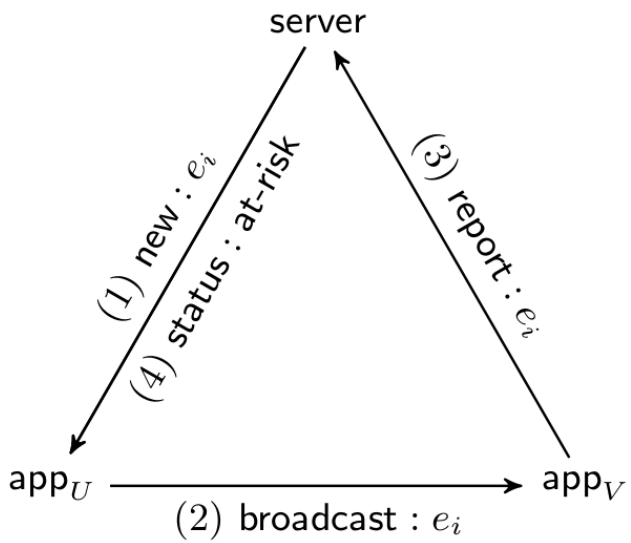
A Specific Attack (2/2)

A similar approach can be used to track the physical location of sick people.
`https://github.com/oseiskar/corona-sniffer`



Outline of this Section

"Centralized"?



1. The server generates and distributes crypto-identifers for each user,
2. Each user broadcasts the identifiers thus obtained,
3. If a user becomes sick, they send all the crypto-identifiers they received from others to the central server.
4. The server knows who these belong to and warns "at risk" users.

Motivation

The nefariousness of ill-intentionned users must be minimized.

In particular, the previous attack targetting centralized systems does not work: the attacker would only know if they have met at least one sick person, not who or how many.

A Specific Attack (1/2)

Of course, a State which is not a democratic State would have a very powerful tool for massive surveillance

[5] "Investigating Third Ways for Exposure Notifications in Europe", https://github.com/3rd-ways-for-EU-exposure-notification/resources/blob/master/A_Contribution_to_Third_Ways_in_Europe.pdf

A Specific Attack (1/2)

> *Of course, a State which is not a democratic State would have a very powerful tool for massive surveillance* *(Bruno Sportisse[5])*

The state-managed central server knows *a lot*:

- anonymity cannot really be guaranteed since we can link a permanent identifier with an IP adress,
- it knows the permanent identifiers of all those in "contact" with each sick person $\Longrightarrow$ the state will know large chunks of the social graph.

This problem is supposed to be solved using a mix-net. That would be quite a feat; it is not the case right now.

[5] "Investigating Third Ways for Exposure Notifications in Europe", https://github.com/3rd-ways-for-EU-exposure-notification/resources/blob/master/A_Contribution_to_Third_Ways_in_Europe.pdf

A Specific Attack (2/2)

More generally, the security level of a centralized system hinges on the temporary/permanent identifier correspondance.

A single cryptographic key protects the secrecy of this correspondence: if it it is recovered/leaks/is misused, then the anonymity of the whole system is compromised!

Outline of this Section

Short Time

CTAs had to be developed very quickly, mostly during the lockdown: kids running around, improvised home office, harder collaboration, (+ baseless accusations because of the centralized/decentralized war).

The code produced cannot be great (not because the developers involved are not good, because it is impossible given the circumstances).

Politics (Pressure)

Had we not worked on StopCovid, we would have been blamed for ignoring CTAs.

(Cédric O, paraphrasing)

Let's not forget what the world was like even only a month ago. In my opinion, that quote is correct.

Politics (Sovereignty)

Apple and Google want to deploy CTAs based on a "decentralized" approach.

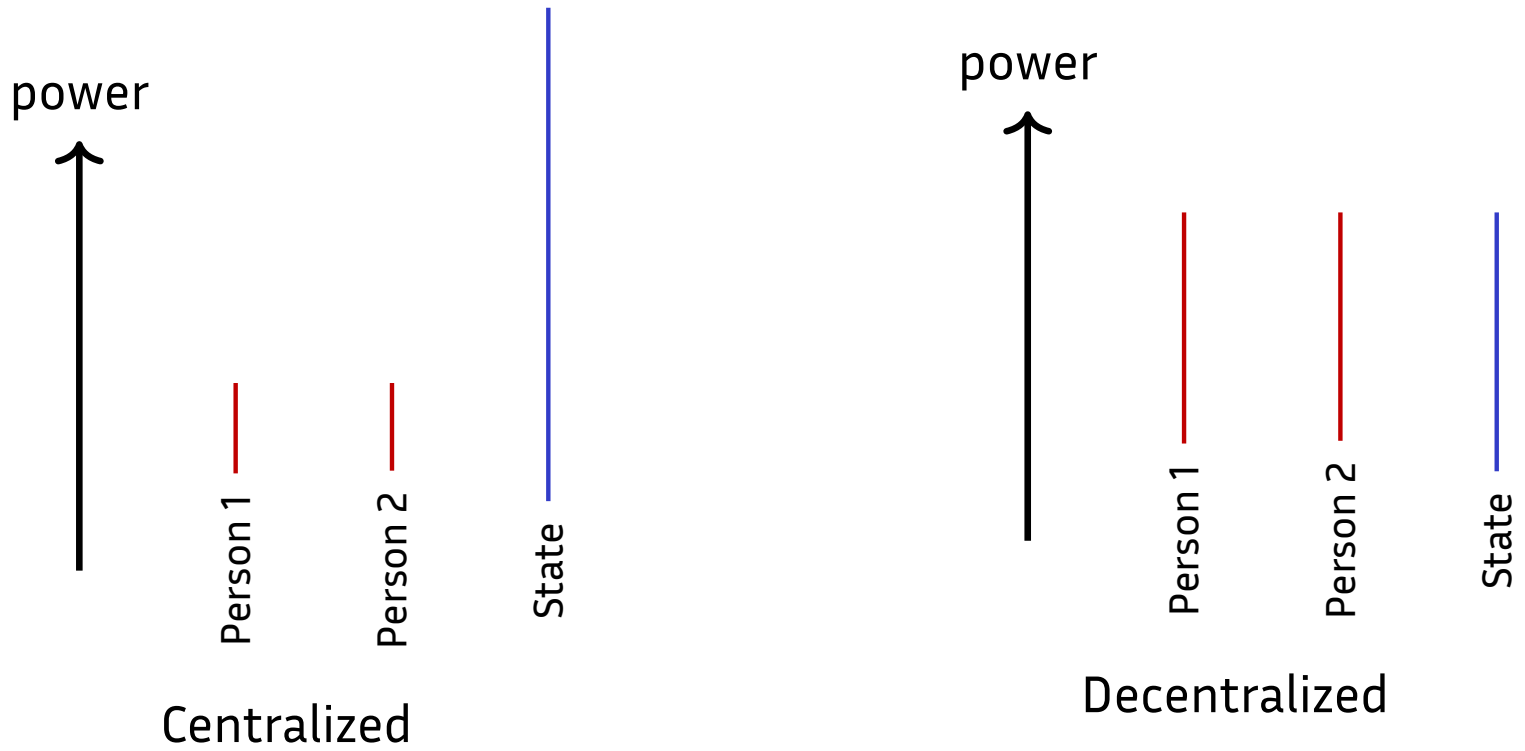
Is it their right to make this decision?

Philosophically, debatable.

Practically, yes: they chose to greatly facilitate the implementation of *decentralized* solutions. Centralized solutions are then much less efficient.

`StopCovid` cannot work properly on iOS because Apple is restricting the access to BLE.

Politics (Centralized vs. Decentralized)



Preferences between "centralized" vs. "decentralized" boil down to the core axioms of one's political view, hence religion-like passion.

Outline of this Section

Back to Tracing Efficiency

$$\epsilon_T = U^2 \times D \times c$$

In France, we have

- $U \leq 0.75$,
- $D \leq 0.80$ (according to Cédric O[6], but ignoring surface contact because it is BT-based)

[6] https://www.francetvinfo.fr/sante/maladie/coronavirus/testee-grandeur-nature-par-une-soixantaine-de-militaires-l-application-stopcovid-est-prete-et-jugee-suf 3981357.html

Back to Tracing Efficiency

$$\epsilon_T = U^2 \times D \times c$$

In France, we have

- $U \leq 0.75$,
- $D \leq 0.80$ (according to Cédric O[6], but ignoring surface contact because it is BT-based)

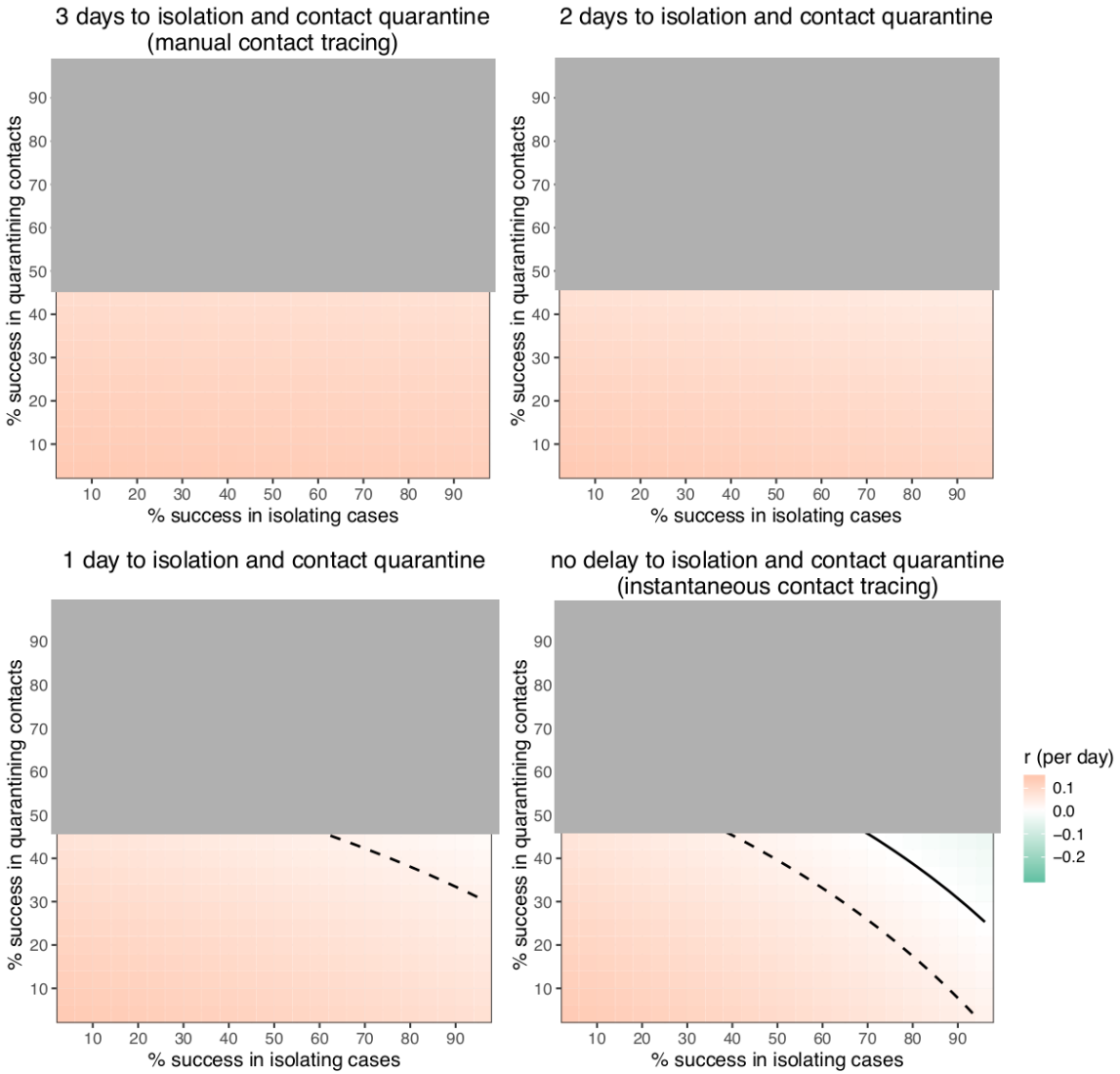
If a contact tracing app successfully detects **all that it can** (be it `StopCovid` or a DP3T-based one), then $\epsilon_T \leq 0.56$

In the case of `StopCovid`, we know D, so

$$\epsilon_T \leq 0.45$$

[6] https://www.francetvinfo.fr/sante/maladie/coronavirus/testee-grandeur-nature-par-une-soixantaine-de-militaires-l-application-stopcovid-est-prete-et-jugee-suf 3981357.html

Maximum **Possible** Efficiency (France, best case)

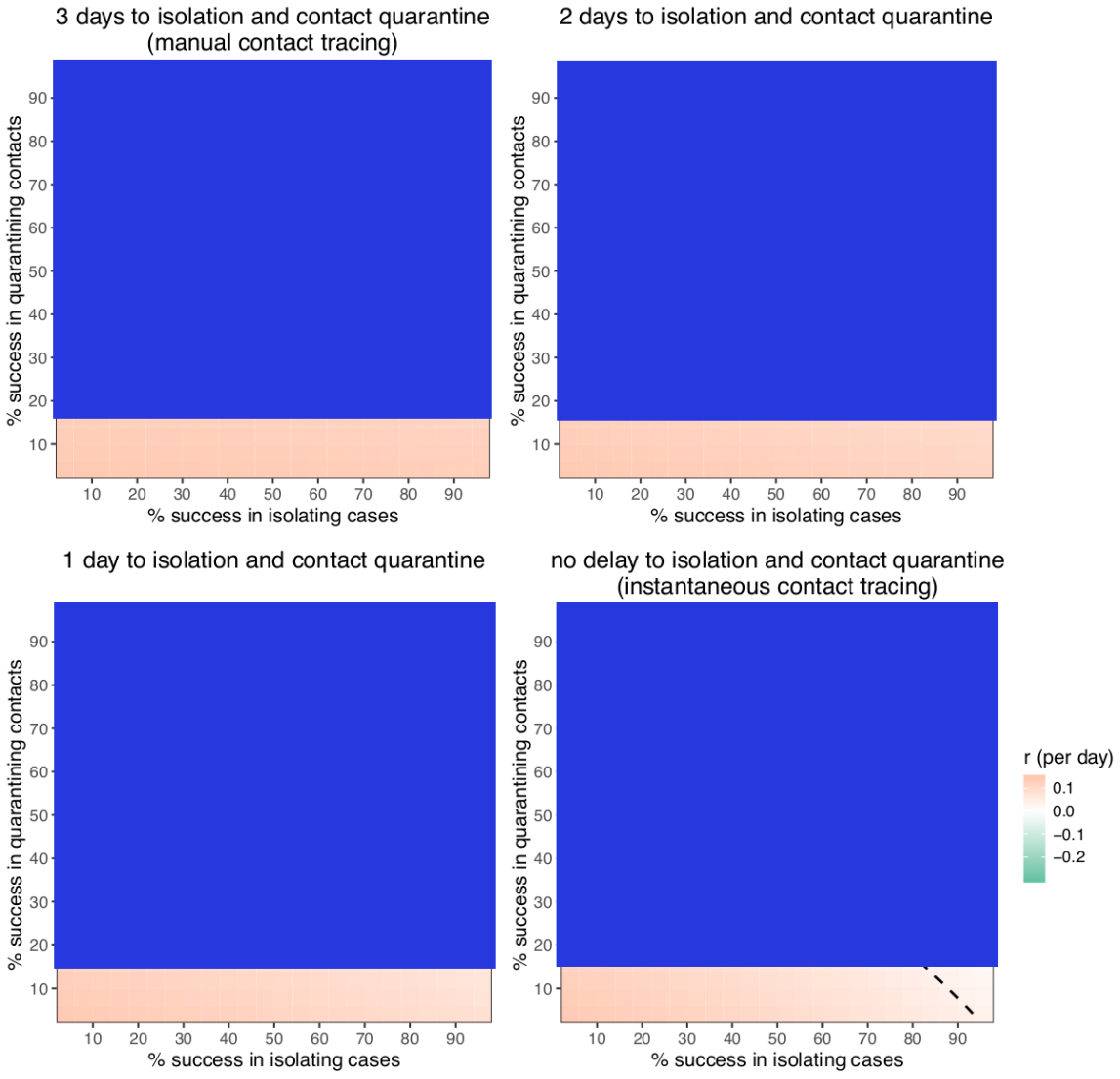


France

$U \leq 0.75$ $D \approx 0.8$

$\epsilon_T \leq 0.45$

Maximum Efficiency (Iceland)



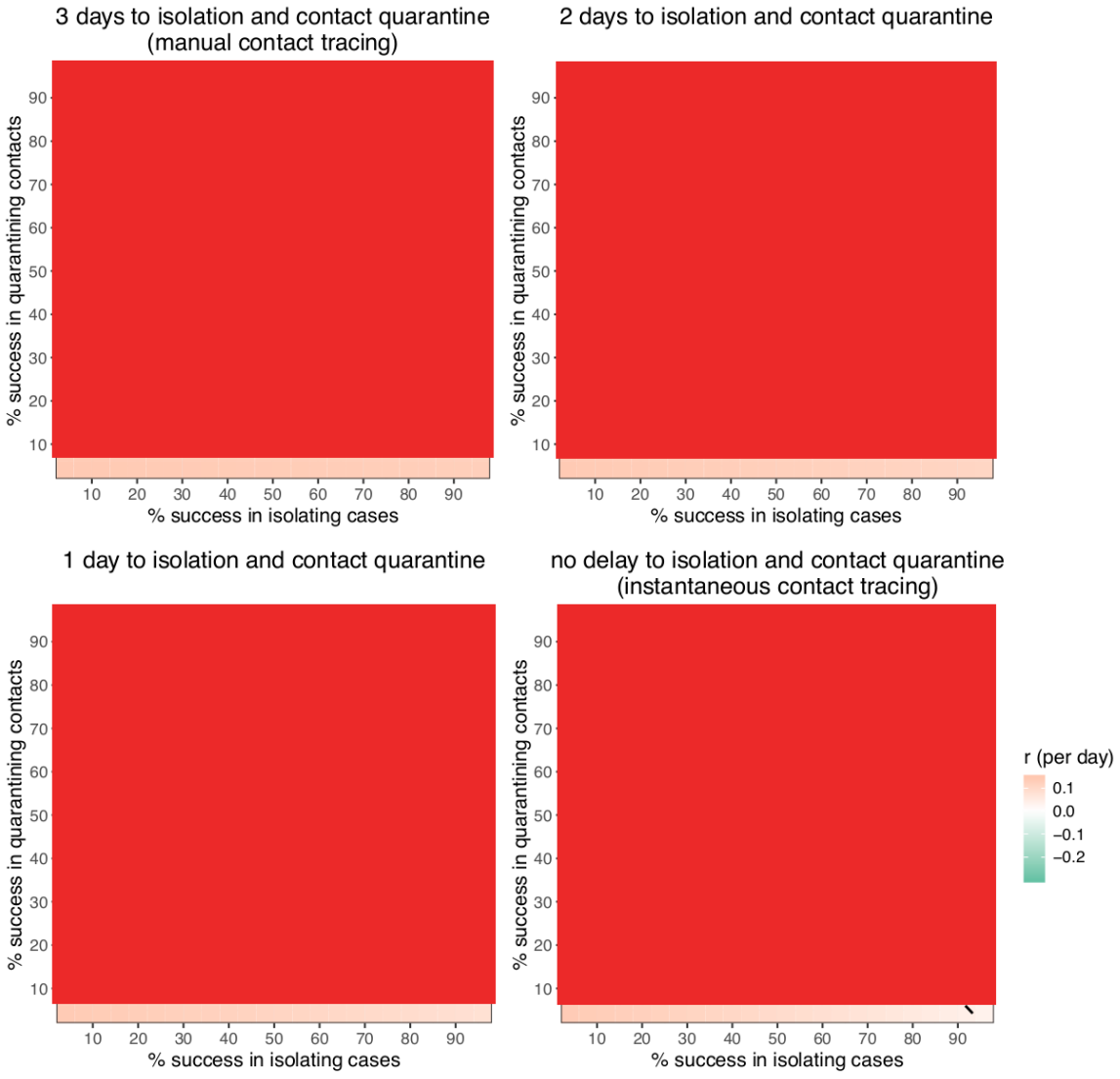
Iceland

$U \approx 0.40$

D: no idea, so ≈ 1

$\epsilon_T \leq 0.16$

Maximum Efficiency (Singapour)



Singapour

$U \approx 0.20$

D: no idea, so ≈ 1

$\epsilon_T \leq 0.04$

What Can we Hope From an Application?

Using the Oxford model:

1. the point at which a CTA is self sufficient is out of reach **in practice**,
2. the impact of a CTA is severely limited by:
   - the time taken to get a test, and
   - the adoption rate: $\epsilon_T \leq U^2$.

What Can we Hope From an Application?

Using the Oxford model:

1. the point at which a CTA is self sufficient is out of reach **in practice**,
2. the impact of a CTA is severely limited by:
   - the time taken to get a test, and
   - the adoption rate: $\epsilon_T \leq U^2$.

In order for a CTA to be able to detect half of all contacts, we need that at least $\sqrt{0.5} \approx 71\%$ of the whole population uses it (assuming the app does not miss any contact).

For `StopCovid`, we know $D \leq 0.8$, so we need $U \geq \sqrt{0.5/0.8} = 79\%$, which is more than the smartphone-equiped population!

Model vs. Reality

Even a few percent of users will already make a difference.

(Cédric O, paraphrasing)

[7] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.

`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

Model vs. Reality

Even a few percent of users will already make a difference.
(Cédric O, paraphrasing)

This quote most likely comes from a misreading of a (admittedly *very* misleading) follow-up of the same study[7] which compares CTAs with the absence of tracing...

[7] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.
`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

Model vs. Reality

Even a few percent of users will already make a difference.
(Cédric O, paraphrasing)

This quote most likely comes from a misreading of a (admittedly *very* misleading) follow-up of the same study[7] which compares CTAs with the absence of tracing...

CTAs were considered $\approx$ useless in Iceland ($U = 40\%$) and Singapour ($U = 20\%$) where they were used in conjunction with manual tracing..

If reality does not match a model, then we might be cautious about implementing policies based on this model...

[7] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.
`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

Outline of this Section

Outline of this Section

Replacing all Counter-Measures, Including Tests

Claim

With a high enough adoption rate, a CTA can replace all counter-measures (masks, social distancing) **including tests**!

(Cédric O, paraphrasing)

[8] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.

`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

Replacing all Counter-Measures, Including Tests

Claim

With a high enough adoption rate, a CTA can replace all counter-measures (masks, social distancing) **including tests**!

(Cédric O, paraphrasing)

Given that the whole point of (digital) contact tracing is to propagate the information from a test... No.

I suspect that the follow-up[8] struck again: in this report, it is assumed that users are allowed to self diagnose. There are tests, just not proper ones.

[8] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.
`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

Replacing all Counter-Measures, Including Tests

Claim

With a high enough adoption rate, a CTA can replace all counter-measures (masks, social distancing) **including tests**!

(Cédric O, paraphrasing)

Given that the whole point of (digital) contact tracing is to propagate the information from a test... No.

I suspect that the follow-up[8] struck again: in this report, it is assumed that users are allowed to self diagnose. There are tests, just not proper ones.

Highly misleading.

[8] *Effective Configurations of a Digital Contact Tracing App: A report to NHSX*, Hinch et al.
`https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf`

APIs

Claim

DP3T-based apps will give medical information to Google/Apple.

Using an API provided by Google/Apple/Inria does not imply that we are giving information to these entities.

On the other hand, if we assume that Google/Apple will purposefully try and get this data using dishonest approaches, then they can do it with StopCovid as well since it runs on Android/iOS!

True for all or none of the CTAs

It is too late

Claim

CTAs arrive too late, they are pointless now.

Who knows what the future holds?

Outline of this Section

Phone Contact $\neq$ Contact Case

Claim

StopCovid will know and interact with contacts stored in your phone.

(Jean-Luc Mélenchon (among others), paraphrasing)

Phone Contact $\neq$ Contact Case

Claim

StopCovid will know and interact with contacts stored in your phone.

(Jean-Luc Mélenchon (among others), paraphrasing)

- Contact tracing apps (try and) use bluetooth to identify physical proximity.
- In fact, accessing the contacts on a phone requires an explicit authorization from the user.

Phone Contact $\neq$ Contact Case

Claim

StopCovid will know and interact with contacts stored in your phone.

(Jean-Luc Mélenchon (among others), paraphrasing)

- Contact tracing apps (try and) use bluetooth to identify physical proximity.
- In fact, accessing the contacts on a phone requires an explicit authorization from the user.

No. (or you would notice)

The GAFAMs and StopCovid (1/2)

Claim

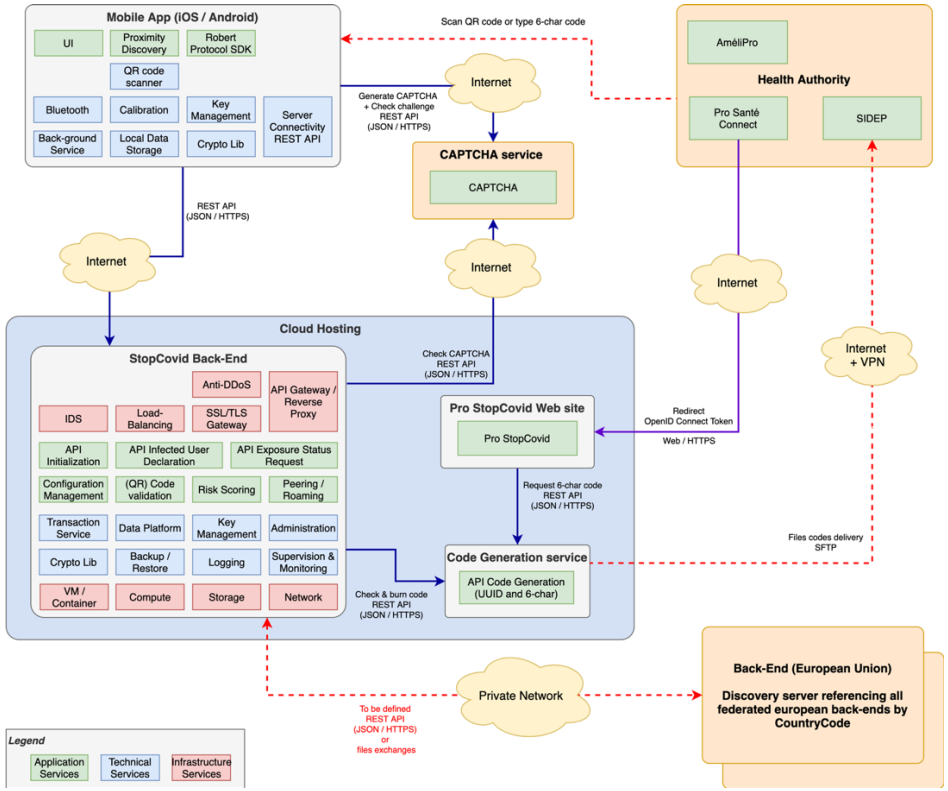
The GAFAMs are not absent from StopCovid.

The application runs on the smartphones of Apple and Google, but (as we saw when discussing API), it is not that big a deal. Unless they decide to kick StopCovid from their app stores (unlikely).

StopCovid uses a CAPTCHA from Google, which will leak some info to them. It is supposed to be temporary.

For Microsoft, it is more complicated.

The GAFAMs and StopCovid (2/2)



source: StopCovid documentation on gitlab

StopCovid seems to be interfaced with the SIDEP database which is on the Health DataHub, thus hosted by Microsoft (on their servers).

Claim

StopCovid opponents need to pick a side, they can't claim that the app is both inefficient and dangerous, it is one or the other.

(Philippe Latombe, among others)

It is possible to have all of those at once:

- the app turns out not to be of any practical help to the tracers,
- malevolent users exploit it to trigger false quarantines,
- the library developed for StopCovid (and/or the one developed by the DP3T) that uses BLE to estimate distances is used by a dictatorship to monitor its citizens even more closely.

Claim

StopCovid opponents need to pick a side, they can't claim that the app is both inefficient and dangerous, it is one or the other.

(Philippe Latombe, among others)

It is possible to have all of those at once:

- the app turns out not to be of any practical help to the tracers,
- malevolent users exploit it to trigger false quarantines,
- the library developed for StopCovid (and/or the one developed by the DP3T) that uses BLE to estimate distances is used by a dictatorship to monitor its citizens even more closely.

False dichotomy.

Outline of this Section

Conclusion

"Centralized" vs. "decentralized" is only a small part of the problem.

We have provided a risk assessment. An informed decision can only be based on an assessment of the benefits of the application considered, **one that takes into account the context in which it is deployed**.

`http://risques-tracage.fr/`

`http://attention-stopcovid.fr/`

`https://www.acm.org/binaries/content/assets/public-policy/europe-tpc-contact-tracing-statement.pdf`

Conclusion

"Centralized" vs. "decentralized" is only a small part of the problem.

We have provided a risk assessment. An informed decision can only be based on an assessment of the benefits of the application considered, **one that takes into account the context in which it is deployed**.

`http://risques-tracage.fr/`

`http://attention-stopcovid.fr/`

`https://www.acm.org/binaries/content/assets/public-policy/europe-tpc-contact-tracing-statement.pdf`

Thank you!