1 Legend

Dear Florian,
here's how to read this.

I use the following color code to indicate different levels of incompleteness or uncertainty (usually with regard to the language, not the facts):

- Not sure
  Doesn't sound weird exactly but some doubt lingers and I'm simply not sure (punctuation, grammar, wording ...).
- Pretty sure it's wrong or could be better, at least
  Might be poor style or I don't know which word to pick or something else that feels not quite right/appropriate.
- *Comments*
  Notes, remarks and such like
- TODO
  Something to check or fix. Mostly bibtex references (see footnotes starting with "**BIB-REF**:").

If there's anything to say beyond the color, I put it in a footnote. These footnotes are colored the same way the corresponding phrase is and bracketed to set them apart from ordinary footnotes – there's indeed at least one of those hidden somewhere ;D...

Section headings try to summarize what a passage is (mainly) concerned with – think of them as kind of working titles. The sections are in no particular order.

Hope you can get anything out of it – in parts it's rather notes than ready-to-use texts (and less than I expected/hoped for) :/

Use it, call me or send me an email in case you have any questions, send me a picture of scribbles you put on a print out and I'll send you (hopefully) ready-to-use revisions of stuff you want to use, whatever suits you ... just let me know :)

LG

PS: I honestly don't think the development of the German corona app took remarkably long – I seem to remember reading something like that in your introduction. Judging from my daily (professional) experience I'm sort of amazed they got it done so *fast*, i.e. almost in time ;) – sadly enough, I really mean that... I don't doubt it was costly, though.

2 Related works: DP-3T and PEPP-PT

This paper is mainly concerned with the protocol architecture proposed by the *Decentralized Privacy-Preserving Proximity Tracing* project (DP-3T). The main reason to focus on this particular family of protocols is the *Exposure Notification Framework*, jointly published by Apple and Google, following/that follows a number of core principles of the DP-3T proposal. This API is not only utilized in/by the German *Corona-Warn-App* (CWA) but has the potential of being widely adopted in future app developments (we might see/that might emerge), due to the reach of players like Apple and Google.

There are, however, competing architectures noteworthy, namely protocols developed under the roof of the *Pan-European Privacy-Preserving Proximity Tracing* project (PEPP-PT), e.g. PEPP-PT-ROBERT[1] and PEPP-PT-NTK[2].

Neither DP-3T nor PEPP-PT is synonym for just a single protocol. Each project endorses different protocols with unique properties in terms of privacy and data protection.

Yet, on a more abstract level/a higher level of abstraction, it seems feasible to distinguish two basic architectures: Protocols as endorsed by PEPP-PT might be characterized as centralized architectures whereas DP-3T-inspired protocols follow a (more) decentralized approach.[3]

"DP-3T is a free-standing effort, originally started at EPFL and ETHZ." Some of its members participated in the PEPP-PT project but have since resigned from the initiative.[4]

3 Basic DP-3T protocol: How does it work? How does it relate to ENF/CWA? (broad strokes)

Upon installation, the app generates secret daily seeds to derive so called *Ephemeral IDs* (EphIDs) from. EphIDs are generated locally with cryptographic methods and cannot be connected to one another but only reconstructed from the secret seed they were derived from/corresponding secret seed.

During normal operation each client broadcasts their EphIDs via Bluetooth whilst[5] scanning for EphIDs broadcasted by other devices in the vicinity. Collected EphIDs are stored locally along with associated metadata such as signal attenuation and date. In DP-3T the contact information gathered is never shared but only evaluated locally/on the device.

[1] **BIB-REF**: ROBERT github repository

[2] **BIB-REF**: NTK ???; mentioned in "Response to Analysis of DP-3T"

[3] As we will see, DP-3T involves a central backend server. It is decentralized with regard to collection and evaluation of contact information: In centralized architectures the server provides a risk scoring services, whereas decentralized approaches rely on local risk assessment and, thus, do not need to share contact information with the backend.

[4] github:DP-3T/documents#april-8th-2020-the-relationship-between-dp-3t-and-pepp-pt

[5] Too pompous? Is "while scanning..." better?

After patients are diagnosed (officially), they are entitled to upload specific data to a central backend server. This data is accumulated by the backend[6] and redistributed to all clients regularly to provide the means for local risk scoring/assessment, i. e. determining whether collected EphIDs match those broadcasted by now-confirmed COVID-19 patients during the last e. g. 14[7] days.

In the most simple (and insecure) protocol proposed by DP-3T [8] this basically translates into publishing the daily seeds used to derive EphIDs from. Aside from additional security measures[9] like signing content the server provides as a general rule, the protocol implemented by ENF and, hence, CWA follows this low-cost design.[10] DP-3T proposes two other, more sophisticated protocols that improve privacy and data protection properties to different degrees but are more costly/complicated[11] to set up.

I have a nice drawing to illustrate this (including some possible mitigation techniques outlined by CWA and DP-3T, respectively). I'll scan it when I'm in the office this week.

4 Formal verification using HOL: Why bother?

Sorry, loose notes is all I got on this one...

- Quote[12]: "Such a design builds on strong, mathematically provable support for privacy and data protection goals [...]"

  DP-3T on their design and how wisely chosen it is ;) We have to use this! Something along the lines of: Despite strong claims with regard to mathematical support, there is, as of yet, no formal verification (known to the authors). So we just thought, we throw our hat in the ring.

- Explore different concepts of refinement: Maybe it is possible to pin down (i.e. exemplify) different concepts of refinement (data refinement, action refinement, trace refinement (aka spec refinement?) ...) – and perhaps combinations thereof – with concrete attack scenarios. I'm sure there'd be something to learn from that (esp. with respect to IsabelleAT).
  Sorry, can't specify it any further :(

[6] Is "backend" a valid ellipsis or is it "backend server"?

[7] I can't figure out how to express that it is some fixed number of days that might be 14 but doesn't have to be (without a lot of lengthy blahblah or getting all formal using some N and "where ...").

[8] **BIB-REF**: Low-cost decentralized proximity tracing; DP-3T Whitepaper p. 14ff

[9] In short: I want to name it but don't talk about it (since we're not really concerned with this level of detail). But this sounds just sh...

[10] **BIB-REF**: CWA solution architecture (→ github:CWA); probably ENF specification

[11] "complicated" sounds lame but "costly" is, perhaps, to simple/literal

[12] **BIB-REF**: DP-3T Whitepaper p. 2

5 What about attacks? Which of them do we consider? ~~Why is that?~~

There is a variety of interesting/noteworthy privacy and security issues that justify/deserve formal consideration (and verification). The debate emerging between advocates of centralized architectures and those in favor of a decentralized approach in particular yields a lot of interesting material in terms of attack scenarios and possible mitigation strategies, e.g. [[13]].

"Analysis of DP-3T" for example addresses the following attack scenarios (categories of attacks):

- *False Alert Injection Attacks*
  - *Backend Impersonation / False Report*
  - *Replay and Relay attacks*
- *Deanonymization/Tracking[[14]] Attacks (the authors don't call it that)*
  - *Based on establishing a connection between collected EphIDs and (target) users, an attacker might:*
    - *Track user's positions by tracking their Bluetooth signal (as recognized by the EphIDs they broadcast).*
→ Location Tracking
    - *Disclose a user's (positive) test status.*
→ Test Status Disclosure
    - *Disclose private encounters.*

  We focus on what we call Deanonymization Attacks.

Although addressed in "Analysis of DP-3T", the two attacks we're concerned with, are classified by DP-3T as generic and inherent risks, respectively[[15]], ~~and, hence, only~~ ~~briefly addressed~~ ← no they're not: *PSRE actually contains a quite thorough analysis of how infected individuals might be identified. (With pretty much the same conclusion I drew.)*
I only realized yesterday, what a treasure trove PSRE really is (or would have been)... Most (if not all) of the stuff I was thinking about the last couple of weeks is noted in there... Apparently I didn't read the right document :/ ... To be honest, this is a bit discouraging...
See: Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems, p. 5f – IR 1: Identifying infected individuals

[13] **BIB-REF**: Analysis of DP-3T (→ github:PEPP-PT); Response (→ github:DP-3T); PSRE and (perhaps) Analysis of PEPP-PT by DP-3T (somewhere on github:DP-3T, I guess; links can be found in Response to Analysis of DP-3T)

[14] Not sure you could really call it "Deanonymization attack". I'm not even sure "deanonymization" is a proper word...

[15] **BIB-REF**: Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems: GR 5 and IR 1

5.1 Why choose Bluetooth Beacons as starting point for building a formal representation?

- Simple and thus easy to formalize.
- Deanynomization attacks start here.
- The Bluetooth Low Energy Beacon mechanism is employed in DP-3T and PEPP-PT protocols alike.
  Perhaps the only thing they can agree on ;)

  There is one significant difference, though:

  - DP-3T: The architecture requires that reported users can be recognized by the apps of their encounters (to some extend; e. g. mitigation by cuckoo filters).

    Quote [16]: “Based on that, the only way to mitigate this attack is to deny the same privileges as the apps to individuals.”

    (IMHO) peculiar English; even so, a direct quote
  - PEPP-PT: The backend server issues EphIDs and never shares any secret seeds, i. e. deanynomizing EphIDs arguably requires more effort (= breaking into the server). In return, more sensitive data (= contact time data) is shared with the server, to begin with.

5.2 Deanonymization Attacks

5.2.1 Different goals, different attacks, different levels of mitigation

Regularly changing EphIDs might prevent location tracking effectively, or, at least, limit it to an acceptable degree. While (possibly) having little to no effect[17] on/in scenarios where the goal is to learn if some particular user was infected.

<u>Idea:</u>
In order for location tracking to succeed, a valid/active EphID (i. e. one that is actually/currently broadcasted) is required. As a consequence, the time frame in which (additional) location information can be gathered is limited by the frequency at which EphIDs change.
A positive test result, on the other hand, can be reported a (fixed) maximum number of days after the compromised EphID was used and still yield the desired information. This maximum time frame is independent of any frequency at which EphIDs may change.

I have two (pretty high level, similar – i. e. comparable) drawings depicting Location Tracking and Test Status Disclosure, respectively.

[16] **BIB-REF**: Analysis of DP-3T, p. 10

[17] Maybe that’s too much; better (?): “significantly/much less impact/effect on”

Could help to illustrate this "strategy X works in this context pretty good, but in that one not so much" idea.

5.2.2 Tracking someone's location

If an attacker can connect EphIDs to specific users or devices, he can track their whereabouts: For as long as a deanonymized EphID is valid, i. e. is broadcasted by the respective client, scanning for this id would suffice to locate this client. The range of Bluetooth reception can be improved by sufficiently good/ high-performance antennas. While stationary set up hardware can free an attacker of the need for actual, physical presence.

5.2.3 Revealing someone's (positive) test status

Like any notification about e. g. a STD[18] might inform a spouse about the infidelity of the other, given the right circumstances, being notified about a possible COVID-19 exposure might reveal illicit information to the recipient like who it was that tested positive.
This is an inherent vulnerability of any such (notification) system, whether it is analogous or digital/digitally aided. The likelihood of said circumstances to occur, however, might increase when notifications where carried out automatically.
This is a quite interesting question to investigate; its formal treatment could benefit considerably from the flexibility IsabelleAT offers: The ability to incorporate aspects of the physical world is surely of special interest in this context[19].

DP-3T:PSRE[20] contains a detailed analysis of how infected individuals might be identified. In the end, however, this attack – being (rightfully) classified as inherent risk – is dismissed: "These weaknesses are both minor and a generic problem in all designs based on Bluetooth proximity detection."[21]

If there was a chance of quantifying quality of support (i. e. a meaningful, yet abstract concept of effectiveness of proximity tracing in terms of aiding traditional contact tracing)... The question of "cost" (e. g. in the sense of increased risk) could be interesting, indeed.
Too bad, this quantification business seems rather unlikely.

[18] Again, I don't know if you can use "e. g." like that

[19] Lame and awkward!

[20] **BIB-REF**: Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems: IR 1, p. 5f

[21] **BIB-REF**: Response to Analysis of DP-3T: I: s.5.2 Deanonymizing Known Reported User, p. 2

Quote [22]: “We emphasize that these attacks work against any contact tracing system, as they rely on the core proximity tracing functionality: notifying at-risk people. Without this notification, proximity tracing system would be useless. The risk is inherent to proximity tracing.”

Suppose, however, proximity tracing could be restricted to places of particular interest, e. g. places where “mass gatherings” are more likely to occur: public transport, supermarkets and such.
Maybe (probably?) this would be worse, privacy-wise, maybe it would be beneficial. In either way, such a system wouldn’t be useless.
How much usefulness really stands to loose (by any digital proximity tracing system) is a question that still needs answering anyway. *(As far as I know – but that’s not very far, to be honest)*

[22] **BIB-REF**: Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems: IR 1, p. 6