ConTra Corona: Contact Tracing against the Coronavirus by Bridging the Centralized–Decentralized Divide for Stronger Privacy

Wasilij Beskorovajnov[1], Felix Dörre[2], Gunnar Hartung[2], Alexander Koch[2], Jörn Müller-Quade[2], and Thorsten Strufe[2]

[1] FZI Research Center for Information Technology
lastname@fzi.de
[2] Karlsruhe Institute of Technology
firstname.lastname@kit.edu

Abstract. Contact tracing is one of the most important interventions to mitigate the spread of COVID-19/SARS-CoV-2. Smartphone-facilitated *digital contact tracing* may help to increase tracing capabilities as well as extend the coverage to those contacts one does not know in person. The emerging consensus is that a decentralized approach with local Bluetooth Low Energy (BLE) communication to detect contagion-relevant proximity, together with cryptographic protections, is necessary to guarantee the privacy of the users of such a system.
However, current decentralized protocols, including DP3T [TPH+20] and the protocol by Canetti, Trachtenberg, and Varia [CTV20], do not sufficiently protect infected users from having their status revealed to their contacts, which may raise fear of stigmatization.
By taking a dual approach, we propose a new and practical solution with stronger privacy guarantees even against active adversaries. In particular, we solve the aforementioned problem with additional pseudorandom warning identities that are associated to the broadcasted public identity, but this association is only known to a non-colluding dedicated server, which does not learn to whom the public identity belongs. Then, only these anonymous warning identities are published.
Moreover, our solution allows warned contacts to prove that they have been in contact with infected users, an important feature in times of restricted testing capacities. Among other additional security measures, we detail how the use of secret sharing can prevent the unnecessary and potentially panic-inducing warning of contacts that have only been around the infected person for a very brief time period.

Keywords: Digital Contact Tracing · Privacy · SARS-CoV-2 · COVID-19 · Active Security · Anonymity

1 Introduction

One of the most important interventions to contain the SARS-CoV-2 pandemic is – besides the reduction of face-to-face encounters in general – the consequent

isolation of infected persons, as well those who have been in close contact with them ("contacts") to break the chain of infections. However, tracing contacts manually (by interviews with infected persons) is not feasible when the number of infections is too high. Hence, more scalable and automated solutions are needed to safely relax restrictions of personal freedom imposed by a strict lockdown, without the risk of returning to a phase of exponential spread of infections. In contrast, *digital contact tracing* using off-the-shelf smartphones has been proposed as an alternative (or an additional measure) that is more scalable, does not depend on infected persons' ability to recall their location history during the days before the interview, and can even track contacts between strangers.

In many digital contact tracing protocols, e.g. [AHL18; CGH+20; RCC+20; CTV20; T; TPH+20; P20a; BRS20; CIY20; BBH+20; AG20], users' devices perform automatic proximity detection via short-distance wireless communication mechanisms, such as Bluetooth Low Energy (BLE), and jointly perform an ongoing cryptographic protocol which enables users to check whether they have been colocated with contagious users. However, naïve designs for digital contact tracing may pose a significant risk to users' privacy, as they process (and may expose) confidential information about users' location history, meeting history, and health condition.

This has sparked a considerable research effort for designing protocols for privacy-preserving contact tracing, most of which revolve around the following idea: User's devices continuously broadcast ephemeral and short-lived pseudonyms[3] and record pseudonyms broadcast by close-by users. When a user is diagnosed with COVID-19, she submits either all the pseudonyms her device used while she was contagious or all the pseudonyms her device has recorded (during the same period) to a server. Users' devices either are actively notified by the server, or they regularly query the server for pseudonyms uploaded by infected users.

Some of the most prominent designs following this idea are the centralized PEPP-PT proposals ROBERT [P20c] and NTK [P20b], as well as the more decentralized DP3T [TPH+20] approach, at which also the Apple/Google-API proposal [AG20] aims at. While the centralized approaches of PEPP-PT do not guarantee users' privacy against the central server infrastructure [D20b; D20c], the DP3T approach [TPH+20], as well as the similar protocol by Canetti, Trachtenberg, and Varia [CTV20], expose the ephemeral pseudonyms of every infected user, which enables her contacts to learn about whether she is infected. The interested reader is referred to [F20] for a fine grained comparison.

We argue that both, *protection against a centralized actor*, as well as *protection of infected users from being stigmatized for their status*, is of utmost importance for any real-world solution.[4] By specifying a protocol that achieves both of these goals

[3] In order to save energy and maximize devices' battery life, these pseudonyms should be as short as possible (e.g. 128 bits).

[4] This is especially true due to the desired inter-operability of solutions and/or a pan-European or even global adoption. Deploying strongly privacy-preserving solutions in democratic nations is particularly important, as these solutions will likely be deployed in states with fewer safeguards against mass surveillance and less democratic oversight

(under certain assumptions), and detailing the corresponding (modular) design choices, we aim to contribute to the ongoing discussion on privacy-preserving digital contact tracing.

1.1 Contribution

We propose a new and improved protocol for privacy-preserving contact tracing, which enjoys the following main properties:

- Our protocol does not allow anything non-trivial[5] to be learned on who is infected, even towards contacts that are warned by their app. This is done by splitting the broadcasted identifiers into two unlinkable pseudorandom identifiers, where one is used for broadcasting, and the other for publication by the server, in case the broadcasted identifier is uploaded by an infected individual.
  Additionally, we discuss approaches to preventing Sybil attacks (where an attacker creates multiple accounts to observe which of them is warned), based on the literature on the topic. Such attacks were deemed to be inherent by [D20a, IR 1: Identify infected individuals]. We believe, however, that Sybil attacks can be effectively mitigated by existing solutions.
- Our protocol makes use of a strict server separation concept to mitigate the threat to users' privacy posed by data collection on centralized servers. In our protocol, the medical professional uploads signed and encrypted public identifiers (without learning them) to a dedicated "matching" server, which does a lookup of the respective registered secret identity, which is an encryption of an (also pseudorandom) "warning identity". These can then be decrypted by a dedicated warning server that publishes them after de-duplication. This separation does not lead to a significant increase of computation on the side of the smartphone.
  Note that the specialized servers may be distributed amongst well-known independent institutions.[6] Thus, in order to compromise the server architecture to the full extent, multiple institutions would have to cooperate maliciously.
- Our protocol allows warned users to securely prove the fact that they have been warned, e.g., towards medical professionals that administer tests for COVID-19. Without this feature, anybody who is curious about their status

as well. In the same vein, while some are willing to give up protecting the infection status of individuals towards their contacts, a more widespread adoption will lead to adverse effects in societies where a greater stigma is attached to being infected. Finally, the voluntary adoption of solutions will crucially depend on the perceived privacy protections, and any solution needs to meet a certain threshold of users in order to provide sufficient utility.

[5] Except for leakage, such as when a warned person has only been in contact with one other person.

[6] For the case of Germany such institutions may be the Robert Koch Institute (RKI), the Federal Ministry of Health (BMG) or the Data Privacy Agencies of the federal states.

but not at risk could get tested, e.g., by showing a screenshot of a warning from someone else's smartphone – which would be unacceptable in times of restricted testing resources.
– As far as possible, our protocol prevents active attackers from injecting false positive warnings and from suppressing warnings supposed to be generated by the protocol (*false negatives*).

Moreover, we identify the timing of Bluetooth beacons as a side-channel that can be exploited to link distinct public identifiers, and propose a concrete solution for adding timing jitter that is compatible with an update of identifiers.

Joint Statement on Contact Tracing of April 19th 2020 [K+20]. The current fast-paced discussion about concepts and applications of contact tracing was recently complemented by a joint statement of many researchers from all over the world, proclaiming desired design principles for contact tracing applications. We argue that our protocol follows all of these principles. The following are short, non-exhaustive notes on these principles in our protocol.

– Any data being processed during the execution of our protocol is necessary for the contact tracing functionality or for a countermeasures against one or more threats.
– The separation of duties to multiple servers, operated by distinct entities reduces the risk of mission creep.
– Our scheme is motivated by the observation that publishing the same information as was observed locally is inevitably not fully privacy preserving.
– We agree with the joint statement and [C20] that any implementations of such protocols should be open-source and amenable to public analysis.

Finally, we encourage the reader to find flaws, alternatives or optimizations for every design choice we made.

1.2 Scope

The reaction of a user to a warning by the system is out-of-scope of this report as they are not strictly dependent on the scheme. The user might be recommended to go into self-quarantine or such a warning might be an entitlement to a medical test. However, treating a warning as an entitlement should be considered carefully, as it might create an incentive for healthy individuals to colocate with infected individuals, and might influence individuals into using a digital contact tracing application they would not want to use otherwise. (For example restricting medical tests to users of the system and not giving non-users an equal chance to be tested makes the usage of the system less voluntary.)

For simplicity, we describe our protocols using concrete numbers instead of abstract parameters. It is understood that these values represent parameters that can be chosen according to the epidemiological requirements. As examples values, we describe our protocol with 15 minutes or longer as the approximate minimum duration of encounters to be tracked, three weeks as retention period for locally

observed contacts, and one day as desired precision of a contact warning. However all these parameters can be adjusted to change the trade-off between privacy, utility and performance of the scheme and to fit the contact tracing needs of the situation the scheme should be used in.

For the system to have good accuracy the distance between participants needs to be estimated in order to track only contacts within a certain close distance. Technical challenges in estimating the distance of two devices via Bluetooth LE communication are out-of-scope for this technical report, we simply assume a reasonably accurate method of estimation is available, as such a method is needed for a large number of other proposals, too.

1.3 Related Work

First note that there are far too many approaches for contact tracing to list them here fully. In a collaboratory effort, an overview of the related work and the several offered apps has been compiled and published at [scA+20]. See also [T20] for a discussion of some recent proposals of digital contact tracing. We focus here on the cryptographic discussion that has been ongoing on preprint servers, such as arXiv and the IACR eprint archive, and the git repositories of the most prominently discussed proposals of PEPP-PT [P20a] and DP3T [TPH+20], and focus on those that are most similar to our work.

First, let us note that early proposals, such as from the just-mentioned centralized PEPP-PT [P20a] project required that "No geolocation, no personal information or other data are logged that would allow the identification of the user. This anonymous proximity history cannot be viewed by anyone, not even the user of phone A." [P20a, Sect. 2]. However, the emerging consensus is that any app for the purpose of contact tracing should be transparent and open source, which makes it possible that a curious user can run a slightly modified application without any restriction on what to log. Hence, we believe that already everything that a user can observe during the protocol should not leak personal information about other users. This will exclude all more intransparent approaches from further consideration.

Canetti, Trachtenberg, and Varia [CTV20] mention an extension of their protocol using private set intersection protocols in order to protect the health status of infected individuals. However, it is unclear how feasible such a solution is with regard to the computational load incurred on both, the smartphone and the server, cf. [P20e, P3].

Whereas [TPH+20] accepts this issue as inherent by [D20a, IR 1: Identify infected individuals] and therefore does not further address possible countermeasures. Our combined protocol tackles this problem at its root and is augmented to obtain additional security and privacy guarantees, such as preventing the contacts of an infected individual to learn this status (assuming the proposed anti-Sybil protections are effective).

Chan et al. [CGH+20, Sect. 4.1] include a short discussion of protocols that upload observed identifiers in case of infections (as in our case), and propose a certain form of rerandomization of identifiers, albeit completely at the side of the

smartphone. Hence, this approach puts a regular heavy computation cost on the user's device, and is likely not practical, as it has to query all possible identity rerandomizations from the server.

Bell et al. [BBH+20] propose two solutions for digital contact tracing, with the first of them also making use of a splitting of the role of the health care provider, and separate (non-colluding) government run servers.

There are several approaches that use public keys as identities, e.g. [CIY20]. However, the maximum message length of BLE broadcasts does not permit sending entire public keys of asymmetric encryption schemes, cf. [P20e].

Besides BLE-based approaches, there are also proposals that use GPS traces of infected individuals to discover COVID-19 hot spots as well as colocation (albeit with a lower resolution), such as [BBV+20; FMP+20]. However, there is a consensus that GPS-based approaches do not offer a a sufficient spatial resolution to estimate the distance between two participants with sufficient precision.

1.4 On the Centralized–Decentralized Debate

There is a public debate on whether to prefer "centralized" vs. "decentralized" solutions for digital contact tracing. First of all, note that the terminology of this debate is slightly misleading. Schemes utilizing private set intersection (PSI) can be realized with the help of a central server architecture. Following the arguments of this debate such schemes would be flagged as insecure, although they may provide a high level of security and privacy. (However, we agree with using a simplified terminology in a public debate, to warn against certain types of inadequate solutions.)

Our contribution to the centralized–decentralized debate is, an approach that can be described as hybrid, as there are periodic uploads from all users to a submission server, but the contact history stays local to the phone. Additionally, we introduce a server architecture which features a strict separation of different tasks. This combines the privacy advantages of both worlds. More specifically – in contrast to current centralized solutions – our approach achieves almost the same privacy guarantees of current decentralized solutions, even in the case that all servers are compromised and collude with malicious participants, cf. Section 6.1. If the servers are non-colluding, we achieve a (roughly) stronger security notion than current solutions. Due to the separation of servers, one could even call our solution more decentralized than other current decentralized solutions (that also use a central server, which receives data after an infection of a user).

Hence, the centralized–decentralized categorization does not sufficiently describe the security and privacy guarantees of a contact tracing scheme. An important aspect of this debate is, of course, the concrete security and privacy guarantees offered by the respective protocol, and (in particular) whether mass surveillance is feasible by a state-level attacker or not.

1.5 Outline

We define our security model for BLE-based contact tracing in Section 2, and specify a first basic protocol that illustrates our main idea in Section 3. For this protocol, Section 4 proposes a number of useful extensions, some of which are applied to obtain our full, combined protocol presented in Section 5. A security and privacy analysis of the full protocols follows in Section 6. We conclude in Section 7.

2 Security Model

This section presents our (informal) security model. Giving a formal definition of security for privacy-preserving contact tracing protocols as well as proofs of security is out-of-scope for this technical report.

Our main goals are *privacy* (i.e. limiting disclosure of information about participating individuals) and *security* (i.e. limiting malicious users' abilities to produce false positives and false negatives). For privacy, we consider the following types of private information:

- where users have been at which point in time (i.e. their location history),
- whom they have met (and when and where),
- whether a user has been infected with SARS-CoV-2,
- whether a user has received a warning because she was colocated with an infected user.

We refer the interested reader to [KBS20] for an extensive systematization of different privacy desiderata.

Participants of our protocol assume one or more of the following roles:

Users, who have installed the contact tracing application on their mobile phone in order to receive a warning if they have been in close proximity to one or more infected persons, and who want to warn other users in case they are infected themselves.

Medical Professionals, who administer tests for the SARS-CoV-2 virus and medical treatment as appropriate.

(Centralized) Servers, operated by health authorities or other responsible organizations.

Medical professionals may be users, too.[7] We assume all communication between the servers uses secure (i.e. confidential and authenticated) channels. We assume the attacker cannot permanently prevent communication between other parties.

We assume centralized servers and medical professionals to be trusted with respect to security (but only partially trusted regarding privacy), i.e. we assume

[7] Since their job may require them to treat infected persons, and hence stay in proximity to them, we assume they deactivate the application when wearing personal protective equipment while on duty, in order not to continuously receive warnings.

they will not engage in actions undermining the availability or correct functioning of the protocol. This entails they will not deviate from the prescribed protocol in order to cause fake warnings or suppress due ones. Furthermore, we trust medical professionals to not disclose data regarding the users who are under their care, as is their duty under standard medical confidentiality. The centralized servers are assumed not to cooperate in breaking users' privacy, cf. Section 6.1. Users not being medical professionals are not trusted to adhere to the protocol.

When considering the distance of corrupted users to another user A, we use a slightly relaxed notion of "proximity": We consider the attacker to be close to A iff she is able to communicate with A's mobile phone via the BLE protocol (potentially using dedicated equipment such as high-gain antennas). This includes situations where the attacker only communicates with A's device via relays that are in "proximity" to A, as considered by [P20d].

Given this model, we strive to attain the following important security and privacy goals. (See below for a justification of the limitations in our goals.)

1. A coalition of malicious/corrupted users must not learn private information about uncorrupted users, except for information that can be observed via other means (e.g. by one malicious user being in close proximity to the victim).
2. The same holds for medical professionals, except that medical professionals may be aware of the health conditions of users under their care.
3. Even if all of the central servers are compromised and colluding with malicious users, the privacy impact for the users must be as low as possible.
4. Users should be able to prove they have been in proximity with an infected individual and received a warning by the application.
5. A coalition of malicious users (not including medical professionals) must not be able produce a *false negative* result, i.e. they must not be able to cause a situation where an uncorrupted user who has been in colocation with an uncorrupted infected user (close enough, for long enough) does not receive a warning (or cannot prove possession of a warning), unless a corrupted user has been suppressing communication between the victim and the infected person during the colocation.
6. (Coalitions of) malicious users (not including medical professionals) must not be able to cause a warning to be delivered to an uncorrupted user, unless the victim has indeed been colocated with an infected user or the attacker has been colocated both with an infected person and with the victim (*false positive* regarding an uncorrupted user).
7. (Coalitions of) malicious users (again, not including medical professionals) must not be able to prove possession of a warning, unless one of the malicious users has in fact been in colocation with an infected user (*false positive* regarding a corrupted user).

We do not consider adaptive corruptions, i.e. users are either honest or corrupted, but this does not change during the protocol execution. Dealing with adaptive corruptions is out-of-scope for this technical report. We do not distinguish between "the attacker" and corrupted, malicious, or compromised parties.

We believe the limitations mentioned in the above goals are somewhat fundamental for an app that is based on tracking colocation:

- Regarding goals 1 and 2, observe that corrupted users can always learn information about other users by other means. For example, a corrupted medical professional who is administering tests or treatment to a potentially infected user will obviously be aware of the user's medical situation, and hence know whether the user is infected. We consider such information leakage inevitable.
- Medical professionals may always be able to cause suppression of a due warning or delivery of an undue warning by tampering with the test procedure. Again, this seems inevitable, so we only consider corrupted users not being medical professionals for goals 5 to 7.
- If an infected user is colocated with a corrupted user, the corrupted user can always simply choose to ignore a warning delivered to her (and/or uninstall the application and delete all associated data). Thus, it is unnecessary to provide guarantees of delivery of warnings to corrupted users in goal 5.
- If an infected, corrupted user wants to suppress warnings being delivered to colocated uncorrupted users, she can simply uninstall or deactivate the application. This limitation is reflected in goal 5.
- If a corrupted user is present during the meeting of two uncorrupted users, one of whom is infected, the attacker can easily prevent wireless communication between the uncorrupted users by jamming their signals. Hence, in goal 5, we only aim to provide protection from false negatives when the attacker does not prevent the communication between the uncorrupted users.
- We do not fully address the issue of replay/relay attacks as discussed by Vaudenay [V20] and [P20d]. In such an attack, a corrupted user records broadcast messages sent at one time and place and replays them at another time and/or place. This may lead contact tracing applications to register an encounter between users A, B who have not actually been in contact with one another, and hence lead to a false positive warning if A is infected. Thus, we only aim to provide security against false positives when the attacker is not close to both A and B (see goal 6).[8]
  We believe that any contact tracing protocol which is strictly based on non-interactive sending and receiving of local broadcasts without considering the actual time and location will be vulnerable to such an attack. As proposed by Vaudenay [V20], executing an interactive protocol between two users having an encounter may provide a way to achieve better security regarding these attacks. Evaluating the design space for practical interactive protocols is left for future work.
- If a user (uncorrupted or corrupted) has been colocated with an infected person, it is legitimate for the user to receive a warning, and to be able to prove "possession" of this warning. Goal 7 allows corrupted users to "exchange"/"trade" a received warning among them. Even if there were some

[8] Our protocol in Section 5 will require an attacker to stay in proximity to both A *and* B for some time as a partial mitigation of this attack.

cryptographic mechanism binding a warning to a device, corrupted users could still simply exchange their devices in order to "trade" a warning. So, again, providing cryptographic protection against such an attack would not yield improved security in a real-world deployment.

3 Basic Protocol

As a starting point for the remainder of this work we propose the following protocol. Here, H is a hash function, where $\mathsf{H}(k\|x)$ is assumed to be a pseudorandom function (PRF) with key $k \in \{0,1\}^n$ evaluated on input x.

Generation of "Random" Identities. For every time period t, the device generates a random key $k_t \leftarrow\!\!\$\, \{0,1\}^n$, and computes $\mathsf{sid}_t := \mathsf{H}(k_t\|0)$ and $\mathsf{pid}_t := \mathsf{H}(k_t\|1)$, stores them, and (anonymously) uploads k_t to the central server, who recomputes $\mathsf{sid}_t, \mathsf{pid}_t$ in the same way. Both parties store $(\mathsf{sid}_t, \mathsf{pid}_t)$.

Broadcasting. During each time period t, the device repeatedly broadcasts pid_t. When it receives a broadcast value pid' from someone else, it stores $(\mathsf{date}, \mathsf{pid}')$, where date is the current date. Every day, the device deletes all tuples $(\mathsf{date}, \mathsf{pid}')$ where date is three weeks ago or older.

Warning co-located users. When a user is positively tested for SARS-CoV-2, (or the medical personnel believes the user to be most likely infected), the medical personnel estimates the first day sdate during which the user probably was infective, taking into consideration the incubation period of SARS-CoV-2 and the time it takes for an infected person to become contagious her-/himself. Afterwards, one extracts a list of all recorded pid' from the infected user's device, where the associated date is no earlier than sdate. The user hands this list to the medical personnel, who forward the data to the health authorities, who finally upload this list to the server. (This chain of course needs to be authenticated.)

The central server infrastructure maintains a list of all pid' reported as having had contact with an infected person. Furthermore, the server has a list of $(\mathsf{sid}, \mathsf{pid})$ tuples uploaded by all users.

For each pid' reported as potentially infected, the server looks up the respective sid in his database of $(\mathsf{sid}, \mathsf{pid})$ tuples and marks the respective sid as potentially infected.

Either the server publishes a list of all potentially infected sids in regular intervals (signed by the server/the health authorities), or the server allows users to query for a given sids, answering whether the sid has been marked as potentially infected.

This protocol illustrates our idea of separating the broadcast public identities pid from the secret identities sid which are published as warnings to their owners. However, this protocol still falls short of the privacy and security goals (see Section 2) that we are trying to achieve, since a user can link the sid being published as a warning to the time and location the corresponding pid was

broadcast. Thus, the user knows when and where she met the infected person and might be able to deduce the infected user's identity.

We discuss the shortcomings and propose various extensions solving these issues in Section 4. We present our full "combined protocol", which encompasses some of these extensions, in Section 5.

4 Extensions

The simple protocol described above does not meet our requirements regarding security and privacy. This section introduces several improvements to the basic protocol.

4.1 Reusing the Secret Identifiers

In the basic protocol described above, users receiving a warning can immediately observe which of their secret identities sid was published. By correlating this information with the knowledge on when they used which public identity pid, they can learn at which time they have met an infected person, which poses a threat to the infected person's privacy. Note that the DP3T protocol [TPH+20] and the scheme by Canetti, Trachtenberg, and Varia [CTV20] succumb to analogous problems.

To mitigate this risk, we propose to associate the same secret identity sid with many public identities pid. In order to make sure corrupted users follow the protocol, we modify the process of deriving sid and pid values. Instead of choosing sid and pid as in the basic protocol, the user generates a single random key, now called *warning identifier*, once for a longer interval, e.g. one day. We propose the following concrete solution: A user generates a random warning identifier $\mathsf{wid} \leftarrow_{\$} \{0,1\}^n$ per day, and encrypts it with the server's public key pk to obtain $\mathsf{sid} := \mathsf{Enc}_{\mathsf{pk}}(\mathsf{wid})$. For each shorter time period i, the user generates a rerandomization sid'_i of sid, where the randomness is derived from a PRG,and computes $\mathsf{pid}_i := \mathsf{H}(\mathsf{sid}_i)$. The user uploads sid and the PRG seed to the server, who performs the same rerandomization, obtaining the same sid'_i and pid_i values. The user broadcasts the pid_i in random order during the day.

Note that there is a fundamental trade-off to be made here: On the one hand, users should be roughly aware of when they have been in contact with an infected person, so that they can self-quarantine for an appropriate period. Moreover, if they start to show symptoms of COVID-19 in the following days, knowing the time of infection can help medical personnel to estimate when they have been contagious more precisely, and thus enable them to give more precise warnings to other users. Additionally, switching sid values with a high frequency restricts the number of pid values that can be linked in case of a server compromise. On the other hand, if users can determine the time of contact with an infected person exactly, they may be able to deduce the identity of the infected user. In our combined protocol (see Section 5), we compromise by informing users about the *day* they have been in contact with an infected user.

4.2 Hiding Multiple Exposures

The change introduced in Section 4.1 allows to split the process of warning co-located users into three tasks for three non-colluding servers, the *submission server*, the *matching server*, and the *notification server*. This eliminates the single point of failure a single server would constitute. See Section 6.1 for a privacy analysis regarding compromised servers.

- The *submission server* collects the uploaded secret and public identifiers from different users (more precisely, it receives sid and the seed for the PRG) and then computes the $(\mathsf{sid}'_i, \mathsf{pid}_i)$ pairs using the PRG with the given seed. It rerandomizes the sid'_i values another time with fresh, non-reproducible randomness (obtaining sid''_i), and stores $(\mathsf{sid}''_i, \mathsf{pid})$ for a short period of time. When the submission server has accumulated submissions by a sufficient number of clients, it shuffles them and then sends them to the matching server.
- The *matching server* collects the $(\mathsf{sid}''_i, \mathsf{pid}_i)$ pairs and stores them. Upon receiving the pids recorded by the devices of infected users, the matching server looks up the respective sid''_is of all potentially infected users and sends them to the *notification server*.
- The *notification server* decrypts sid''_i to recover $\mathsf{wid} := \mathsf{Dec}_{\mathsf{sk}}(\mathsf{sid}''_i)$ for all potentially infected users and publishes a deduplicated list of warning identities.

4.3 Hiding Contact Information from the Medical Personnel

In the basic protocol from Section 3, the user unveils all public identities of every meeting recorded by the application to her medical personnel, who forwards it to the matching server. This puts the user's privacy at an unnecessary risk, since the medical personnel does not need learn about the user's meetings. To mitigate this issue, the application can simply encrypt the list of public identities with a public key of the matching server.[9]

4.4 Anonymous Communication Channels

When a user uploads her $(\mathsf{sid}, \mathsf{pid})$ pairs to the centralized servers, the servers can easily link these pairs with communication metadata (such as the user's IP address), which might be used to ultimately link these pairs to a specific individual. We therefore propose to use an anonymous communication channel for the submission of the $(\mathsf{sid}, \mathsf{pid})$ pairs. In practice, one might employ the TOR onion routing network [TOR] or send the submission via other publicly available proxies.

[9] This still surrenders the approximate length of the list to the medical personnel. Future work might consider further improvements in order to mitigate this remaining leakage.

4.5 Using Secret Sharing to Enforce a Lower Bound on Contact Time

The DP3T document [TPH+20] proposes splitting the broadcasted identifiers with a secret sharing scheme to ensure that malicious users cannot record identifiers that they observe for less than a specified period of time (e.g. 15 minutes). However, it does not specify how one would rotate such shared identifiers if one wishes to switch to the next public identifier. Just stopping with one set of shares and starting the next set of shares (of a different public identifier) would prevent recording of contact if the contact happens during such an identity rotation.

To solve this issue, we propose to broadcast multiple public identities in parallel with overlapping intervals. As an example we could use a 15-out-of-30 secret sharing scheme and always broadcast two identities, in such a way that the new identity starts to be broadcast when the last identity has already had 15 shares broadcast. That way every contiguous interval of 15 minutes contains enough shares of one identity to be able to reconstruct the identity.

Additionally, care has to be taken that an observer needs to know which beacons belong to the same shared identifier, in order to choose the right shares to combine.

Variant 1 to recombine shares: hardware MAC address. As the BLE-beacons are sent out with an associated Bluetooth hardware address, this address could be used to mark shares of the same public identity. For this approach to work, the application needs to be able to control the Bluetooth MAC address used for the broadcast. The application chooses a random Bluetooth hardware address for each identity to be broadcast. When multiple identities are broadcast the application switches between the hardware addresses back and forth in the time period they overlap.

Variant 2 to recombine shares: custom identifier. If the hardware address is not directly controllable by the application, a per-identity marker could be incorporated into the payload. It needs to be long enough to make local collisions unlikely. In this situation the Bluetooth hardware address should be rotated once per beacon to not provide any unnecessary linkability between multiple identities.

4.6 Side Channel Leakage by the Timing of Broadcasts

If the application sends broadcasts in strict, exact intervals, an attacker might be able to link the two public identities by observing the offset of the broadcast times to her own clock. For example, if an application sends a broadcast in exact intervals of one minute and the attacker can observe that one device is continuously broadcasting whenever the attacker's clock is 10 seconds into the current minute, the attacker may be able to link several broadcasts to the same device even if the public identities being broadcast have changed in between. This may be used to link public identities both if they are used in direct succession, and if the attacker did not observe any broadcasts for a longer period of time.

This attack applies to both cases: if the public identities are broadcast directly, and if they are broadcast in shares (as described in Section 4.5).

To mitigate this attack, we propose to add random jitter to the starting point for broadcasting identities. When applying jitter, care has to be taken to add a few more shares to each identity to still ensure that the identity can be reconstructed from any 15 minute interval. When broadcasting the identity as a single beacon the jitter adds uncertainty to the observed exposure times. In both cases there are two variants how jitter can be applied:

Jitter Variant 1: absolute. When applying jitter absolute one would start to send identity pid_i at the point in time $i \cdot \delta + \Delta_i$, where Δ_i is the jitter chosen at random (e.g. random time between -1 and 1 minute) and δ is the regular interval (e.g. 15 minutes).

Jitter Variant 2: relative. When applying relative jitter, one can think of the jitter as a random pause time between broadcasting identities. Using the notation from Variant 1, the user would start to send identity pid_i at $i \cdot \delta + \sum_{j=0}^{i} \Delta_j$. This way the jitter accumulates over time, and after a long enough period without observation the starting point for broadcasting identities will appear to be random.

As an example, consider 15-out-of-45 secret sharing, with every share being broadcast in 1-minute intervals. When a broadcast is started a random time between 15 and 30 minutes is chosen uniformly at random and after this delay the next ID-broadcast is started. Note that with this change two or three identities are being used simultaneously at every point in time. This ensures that in any 15 minute interval there is at least one public identifier broadcast completely covering the interval. Additionally this jitter accumulates very quickly to destroy the linkability of different broadcasted IDs.

4.7 Proving a Warning

In order to enable a user to prove to a third party a warning has been issued for her, her warning identity wid (or in case the extension in Section 4.1 is not used: her secret identity sid) could be chosen as the hash of a random value u. In order to (one-time)-prove that a published wid is used to warn oneself, the user can present u. This approach has the disadvantage that the receiver of this proof can now show it to someone else.

In order to reduce the trust in the verifying party, one might choose $\mathsf{wid} = g^u$ where g is the generator of a group in which the discrete logarithm problem is hard. Now the presentable proof of warning, would be a non-interactive zero-knowledge proof of knowledge (NIZK-PoK) of u. This proof should contain the current time, in order to prevent it from being reused later by the verifying party. The NIZK-PoK can be obtained quite efficiently using the Fiat-Shamir heuristic [FS86; S89]. In this case, one could include the time, with appropriate granularity,

in the computation of the hash value used as the challenge to prevent reuse at a later point in time.[10]

However the proof can still be transferred by the potentially infected user, by disclosing all secrets to the party who should receive the proof. To discourage transferring of such proofs, one could choose wid as $g^u \cdot h^{\mathsf{rid}}$, where rid is a representation of the user's real identity (e.g. her name), and h is an independent generator of the same group such that the discrete logarithm between g and h is not known. Now the proof can additionally contain the users real identity and then both, the prover and the verifier can divide wid by h^{rid} and then perform the same NIZK-PoK as before. The name cannot be changed afterwards as it is hard to find a u' for a different identity rid' such that $g^u \cdot h^{\mathsf{rid}} = g^{u'} \cdot h^{\mathsf{rid}'}$ (Binding property of Pedersen commitments).

4.8 Hiding Multiple Exposures when the Servers can be Individually Corrupted by Individual Users

In the extension in Section 4.2 the notification server can learn the number of exposures for an individual user if it is colluding with that user. In order prevent this, we introduce an additional server to the "pipeline", the *deduplication server*. The pipeline is now: submission server, matching server, deduplication server, notification server. The deduplication server and the submission server share a symmetric key. When a submission arrives, the submission server now marks all individual entries in this submission with a random identifier (the deduplication identifier) encrypted with the shared symmetric key. The deduplication server decrypts the deduplication identifier and keeps only a random sid for each deduplication identifier. Then it discards all deduplication identifiers and hands all the individual sid to the warning server.

4.9 Protecting from Encounter-wise Warning Identities and Sybil Attacks

If the extension from Section 4.2 is applied, one might additionally want to prevent users from switching their warning identity wid too quickly, because of the following attack:

Example Attack. An attacker being able to upload an unlimited number of sid values to the submission server may be able to perform an attack similar to the *Paparazzi Attack* described by [V20], as follows: After each encounter with another participant, the adversary records the time of the encounter, whom she has met, and which pids she sent during that time period. Then, the attacker switches to a pid value representing another warning identity. This way, when one

[10] Alternatively, one could choose wid to be the hash of a verification key for a signature scheme. The proof would then be a signature on the current time (with appropriate granularity) with the corresponding signing key. This approach may even offer post-quantum security if the signature scheme is post-quantum secure.

of her warning identities wid is published by the notification server, the attacker can link wid to the encounter, and thus possibly the identity of the infected person.

Rate Limiting. Preventing this type of attack requires limiting uploads of sids to one identity per app instance per day. However, an attacker might try to bypass this restriction by running a *Sybil attack*, i.e. creating multiple (seemingly) independent app instances.

A defense strategy is to force the adversary to invest additional resources for spawning Sybil instances. One possibility is to bind each app instance to a phone number during a registration process. (Note that this approach does not prevent an attacker from performing a Sybil attack on lower scale, as the attacker might own multiple phone numbers.)

Ensuring Anonymity while Binding to Phone Numbers. Binding an app to an identifiable resource (such as a valid phone number) endangers the user's anonymity, since the server might store the data linking resource to the app instance. In order to achieve rate limiting without disclosing the link between uploaded sids and an identifiable resource, we propose using the periodic n-times anonymous authentication scheme from [CHK+06] or related schemes offering this functionality. In our setting, we choose $n = 1$ and a time period of one day, i.e. the user can obtain one "e-token" per day to upload a new sid (and PRG seed) to the submission server. The token dispenser is then issued to the user during a registration process, which uses, e.g., remotely verifiable electronic ID cards or phone numbers that are verified via SMS challenges.[11]

5 Combined Protocol

We now describe the protocol that results from applying the extensions described in Sections 4.1 to 4.7 and 4.9 to our basic protocol described in Section 3. This description does not take into consideration the extension described in Section 4.8.

Let n denote the security parameter, $\mathbb{G}$ be a group of prime order q such that the decisional Diffie-Hellman problem in $\mathbb{G}$ is intractable, and let g, h be generators of $\mathbb{G}$. We assume two secure public key encryption schemes $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$: One of them having message space $\mathcal{M} = \mathbb{G}$ and rerandomizable ciphertexts, and one of them having message space $\mathcal{M} = \{0,1\}^*$. (We propose standard ElGamal and a standard hybrid encryption scheme for instantiation, respectively.) Let PRG be a secure pseudorandom generator, and H be a collision-resistant hash function.

Each device continuously and randomly partitions time into overlapping intervals. Whenever one interval begins (say, at time t_0), the application chooses a random time difference Δ (between 15 and 30 minutes, with sub-second precision) and the next interval will begin at $t_0 + \Delta$. Each interval has a duration of 45 minutes. Thus, each point in time belongs to either two or three intervals,

[11] For a low-tech version (maybe implemented in an early version of the protocol), we can also just assume a non-colluding dedicated registration server.

two successive intervals overlap by at least 15 minutes, and there are at most $24 \times \frac{60}{15} = 96$ beginnings of an interval in each day. We note that devices sample this partitioning independently of each other.

Server Setup. The matching server and the notification server each generate a key-pair for a public-key encryption scheme: The notification server for the rerandomizable scheme, the matching server for the other one. The public keys are published in a way users can retrieve them in an authenticated fashion.

App Setup. When the proximity tracing software is first installed on a user's device, the user enters her real identity rid, such as her name. (This information will only be shared with medical professionals treating her.) To avoid fears, the application should present an understandable explanation of why this is necessary (cf. Section 4.7). Additionally, for anti-Sybil measures as described in Section 4.9, the application proves possession of a phone number (e.g. via an SMS challenge) and obtains a e-token dispenser.

Creating Secret Warning Identifiers. For each day, the application generates a *warning identifier* wid as a Pedersen commitment [P91] to the user's real identity. (That is, wid is computed as $\mathsf{wid} := g^u h^{\mathsf{rid}}$, where $u \leftarrow_\$ \mathbb{Z}_q$. We assume rid is implicitly mapped to $\mathbb{Z}_q$ in a collision resistant manner.) The application stores the unveiling information u for later use, deleting it after four weeks.[12]

Deriving Public Identities. For each warning identifier wid, the application computes $\mathsf{sid} := \mathsf{Enc}(\mathsf{pk}, \mathsf{wid})$, where Enc is the encryption algorithm of a rerandomizable, IND-CPA-secure public-key encryption scheme, and pk is the notification server's public key. Additionally, the application chooses a random $\mathsf{seed} \leftarrow_\$ \{0,1\}^n$ (*rerandomization seed*) per warning identity.

The application (interactively) presents an e-token τ to the submission server via an anonymous channel (e.g. via the TOR network), and uploads $(\mathsf{sid}, \mathsf{seed})$ to the submission server via the same channel. Both the submission server and the application compute $24 \times 4 = 96$ rerandomization values $r_1, \ldots, r_{96} = \mathsf{PRG}(\mathsf{seed})$, and rerandomize sid using these values, obtaining $\mathsf{sid}'_i := \mathsf{ReRand}(\mathsf{sid}; r_i)$ for $i \in \{1, \ldots, 96\}$. The ephemeral public identities of the user are defined as $\mathsf{pid}_i := \mathsf{H}(\mathsf{sid}'_i)$ for all i.

The application saves the public identities for broadcasting during the day of validity of wid.

The submission server rerandomizes all the sid'_i values another time (using non-reproducable randomness), obtaining $\mathsf{sid}''_i := \mathsf{ReRand}(\mathsf{sid}'_i)$, and saves a list of the $(\mathsf{sid}''_i, \mathsf{pid}_i)$ tuples. When the submission server has accumulated a sufficiently large list of such tuples, originating from sufficiently many

[12] If some user A has been in contact with an infected user B during the day of validity of the respective warning identity, and even if B takes three weeks to show symptoms and have a positive test result, then A will be able to prove "ownership" of the respective warning for another week, which is sufficient time for her to get tested herself.

submissions, it shuffles the list and forwards all tuples to the matching server and clears the list afterwards.

The matching server maintains a list of all tuples it has received from the submission server, deleting each tuple after three weeks.[13]

Broadcasting Public Identities. For each time interval i, the application randomly chooses one of the public identities pid precomputed for the current day (but not used so far), computes a 15-out-of-45 secret sharing $s_1, \ldots, s_{45} = \mathsf{Share}(\mathsf{pid}_i)$, and selects a random identifier m. (m may be used as the Bluetooth MAC address if possible.)

During the respective interval, the application broadcasts the shares s_j (one at a time, with one minute between the broadcasts) together with the random identifier m.[14]

Receiving Broadcasts by other Users. The application continuously listens for broadcast messages by other users and maintains a database of these. When it receives a new broadcast (m', s'), the application checks if the database contains another broadcast value with the same random identifier m'. If it does, and the previous broadcast is less than (approximately) 60 seconds ago, the newly received message is discarded. Otherwise, the application temporarily saves the received broadcast tuple in its database. All database entries in this database are deleted after at most 45 minutes.

When the application has accumulated 15 broadcasts (m', s'_j) with the same random identifier m', it recombines the shares s'_j to recover the public identity pid' that was shared by the remote application, and records the occurrence of a meeting of its user with the user having the public identifier pid' at the current time. The information about this meeting is stored for the retention period, i.e. 21 days, and deleted afterwards.

Sending a Warning. When a user is tested positively for SARS-CoV-2 by medical personnel or the medical personnel considers an infection sufficiently likely, the medical personnel estimates the first day sdate during which the user was probably contagious. The user instructs the application to collect a list of all meetings she has had from sdate until the present, and the respective list of public identities pid'. She encrypts the list of public identities using the public key of the matching server to obtain a ciphertext c. The user sends c to the medical personnel (via an authenticated channel), who (directly or indirectly) forwards it to the matching server (again, via an authenticated channel, such that the matching server can verify c was sent from some authorized medical professional).

Centralized Processing of Warnings. When medical personnel submits a ciphertext c, the matching server decrypts the ciphertext to recover the list of public identities the application has recorded a meeting with.

[13] If some user A has been in contact with an infected user B who observes the respective pid, and even if B takes up to three weeks to show symptoms and have a positive test result, the data retention on the matching server is sufficient to deliver a warning to A.

[14] Since intervals are overlapping such that any point in time belongs to two or three intervals, the user will be sending a broadcast every 20 to 30 seconds on average.

The server looks up the corresponding secret identifiers sid in its database and sends the secret identifiers to the notification server.
The notification server decrypts the secret identifiers to recover the warning identifier wid contained in them, and regularly publishes a deduplicated list of all warning identifiers it has received during the last two weeks.

Retrieving Warnings. The application regularly fetches the list of published warning identifiers from the warning server (via an authenticated channel) and compares it with the list of warning identifiers it has used during the last 28 days itself.
If there is an overlap, it informs the user she has been in contact with an infected person on the day the warning identifier was used.

Proving Possession of a Warning. If the user reports to a hospital, asking to be tested for SARS-CoV-2, she surrenders her real identity rid and her warning identity wid to the medical personnel. Using a zero-knowledge proof (e.g., using the Fiat-Shamir heuristic), she shows her wid is a valid Pederson commitment to rid.
The medical personnel verifies the proof and verifies that wid has indeed been published by the warning server. (The medical personnel uses an authenticated channel for retrieving the list of warning identities from the notification server.)

This concludes the description of our "combined protocol".

6 Security and Privacy Analysis

We now present a preliminary, informal analysis of the combined protocol described above regarding its security and privacy. As mentioned in Section 4.9 we leave the specific choice of countermeasures against Sybil attacks for future work. We discussed naïve but efficient solutions that thwart such attacks up to a certain degree, and suggested to use the anonymous e-token system by [CHK+06] for a sophisticated but more resource-consuming solution to this problem. With these solutions an attacker can only create a limited amount of Sibyls. From a security point of view, this situation is not different from the case where the attacker is a small group of malicious users. Therefore, we exclude excessive Sybil attacks from the security analysis.

6.1 Privacy

Note that, due to the security of the secret sharing scheme, observing a public identity requires being in proximity to the user's device for approximately 15 minutes. This restrains the attacker's ability to observe public identities at specific times and places in the first place. For our privacy analysis, we assume corrupted users can link some public identities they directly observe to the real identities of the corresponding user, i.e. by accidentally meeting someone they know. This pessimistic approach yields a worst-case analysis regarding the information available to corrupted users.

A corrupted user may be able to modify the application (or use a completely different one) in order to collect additional data not typically collected. Corrupted users might also deviate from the protocol in order to obtain additional information. For our preliminary analysis, we consider the following information to be available to a corrupted user:

- her real identity as well as her own location history,
- all of her own secrets, i.e. her warning identities, the corresponding unveiling information, all encryptions of her warning identities along with the dates and times of validity, and (as stated above) her whereabouts during the time of the validity, and the rerandomization seeds used to derive the public identities,
- all of her own public identities, all secret sharings of her own identities and the corresponding random identifiers m, along with the information when and where they were used/broadcast,
- for all broadcasts she has received: the random identifier m, the corresponding share of the sender's public identity at that point in time, and when and where she has received the broadcast,
- the list of warning identities published by the warning server.

Privacy in Case of Corrupted Participants. Multiple corrupted users may work together, combining their information. Thus, we end up with an adversary in possession of a large amount of recorded broadcasts with meta data on time and location. We conservatively assume an adversary can receive BLE traffic within the entire area where a targeted person may have moved and can link some of the broadcasts to the targeted person, by e.g., meeting her in person or utilizing video surveillance.

Hiding Multiple Exposures. The frequency of exposures, i.e., warnings, may allow the warned participant to re-identify the positively tested participant, who has uploaded the received public identities pids. In Section 4.2 we describe our approach of deduplication of warning wids before publishing. The deduplication prevents a user from learning how many of her public identities have been recorded by infected users, thus hiding the frequency of exposures from the warned participant.

Location Privacy. We analyze how a coalition of malicious users might track a user's location history. For this, we assume that the attacker is in possession of one current pid of the victim individual A, e.g. because the attacker is in proximity to A, therefore also knowing her current (starting) location. In order to learn the subsequent location, the adversary is required to link two consecutive public identities. Without any side-channel information the adversary faces the problem of computing the pre-images of $\mathsf{pid}_i := \mathsf{H}(\mathsf{sid}'_i)$ and breaking the rerandomized encryption of sid. This problem is intractable if the encryption scheme is secure. Utilizing the timing of fixed-spaced intervals when broadcasts are sent as side-channel information, the adversary may be able to narrow down

the set of possible subsequent public identifiers/locations. Informally, this is currently thwarted by the application of random jitter, see Section 4.6, thus reducing the adversary's ability to track users. The adversary may however, even in presence of random jitter, discard some of public identifiers that are guaranteed not to be the subsequent public identifier.

If A started broadcasting her current pid value at time t_a (in minutes), she will start broadcasting her next pid in the time interval $T_A = [t_a + 15, t_a + 30]$. Likewise, if another user B started broadcasting his current pid at time t_b, the first share of B's next public identity will be sent in $T_B = [t_b + 15, t_b + 30]$. If A is close enough to B (so that the attacker cannot distinguish their BLE signal characteristics) for at least $15 + \varepsilon$ minutes ($\varepsilon > 0$), there is a positive probability that both A and B start broadcasting their next pids in the overlap $T_A \cap T_B$. If this happens, then the adversary has no advantage over random guessing in linking one of the pids to A (provided the above complexity assumptions hold).

We argue that our jitter extension provides enough "fuzziness" within a recording of multiple people broadcasting within the same location to hide the subsequent public identifier of the target "in the crowd". (Note that in case the target person is alone for a period of time, the adversary is trivially able to link consecutive public identities of this user.) We leave a rigorous discussion to future work, and only present an informal one:

In order to track a user, the adversary must monitor a large area for Bluetooth signals. Once the victim leaves this area for long enough, the attacker will not be able to link the location tracks of the user. If the attacker has background knowledge, the area can be chosen smaller to be targeted on a specific user. We discuss the two extremes of background knowledge available to the attacker:

- If the adversary has no background knowledge about the possible movement pattern of the target, the adversary needs to cover a radial area with receivers, growing in size with the victim's movement distance. Thus, the area to be covered with receivers grows quadratically in the victim's movement distance, and we expect tracking a user for a prolonged period of time to exceed the adversary's resources.
- If the adversary has complete background knowledge and only wishes to confirm the victim's movements, the number of receivers is linear in the victim's movement distance.

For a proper analysis, it would be advisable to choose a compromise between these two extremes, and to limit the amount of available background information the adversary may have about the movement pattern. (For example, using a priori information on typical movement patterns of people in public spaces is expected to significantly reduce the necessary number of receivers.) Furthermore, it is necessary to estimate the cost of resources, e.g., Bluetooth receivers, needed to cover the assumed movement area. The detection probability of a broad scale setup of Bluetooth receivers may be also taken into the account.

Privacy of Positively Tested Participants. To capture the privacy risk of positively tested participants, note that the only difference in their protocol behavior is that

they hand over an encrypted list of recorded pids of their contacts to the medical professional for an upload to the matching server. We assume that the hand over is done via a confidential channel and that the uploading happens without any reference to the users identity. Assuming the servers are uncorrupted, the only change in the attacker's view is the additional warning identities published on the warning server. Hence, this section only concerned about the privacy risk incurred by this extra information. However, the following discussion will argue that these warning identities are not linkable to any public identities broadcast by uncorrupted users (except under trivial conditions, e.g. that the user was in contact with only one other user during a wid's one-day lifetime), and hence do not pose an additional privacy risk. (In particular, this ensures confidentiality w.r.t. the user's infection status.)

As discussed above, we can assume the adversary is at most a small group of n colluding users (without additional Sybils). The one-day lifetime of these corrupted users' warning identities guarantees that no single user can (on their own) distinguish which of her encounters during the day caused her to receive a warning. (Note, by working together and making use of group testing mechanisms, the malicious users' might be able to single out the infected person from a larger group of 2^n users. This attack is inherent in the desired functionality.) Thus, our protocol is only susceptible to *inherent* attacks.

Privacy of Warned Participants. Our protocol naturally protects the privacy of warned participants and their social graph as the published warning identity is computationally unlinkable to any information that can be recorded locally, thus it is only of use to the warned application. Specifically, each honestly generated warning identifier wid is a Pedersen commitment to the user's real identity. Since Pedersen commitments are perfectly hiding, the attacker cannot infer the user's real identity rid from wid, and also deciding whether some identities belong to the same user, is impossible. Thus, the warning identity wid does not help the attacker in breaking the users' privacy.

Privacy in the Case of Compromised Servers. This section presents a preliminary analysis of the privacy guarantees offered by our protocol if servers are compromised.

Linking Public Identities used on the Same Day. If the submission server is compromised, the attacker will be able to link different public identities pid to the same secret sid, and hence can link the public identities the user is using on the same day. This situation only poses a privacy threat, if the attacker additionally has observed some of the targeted public identities pid, which requires colluding users. See Section 4.1 for a detailed discussion about this trade-off.

Similarly, if both the matching server and the notification server are corrupted, the attacker can decrypt the sid values stored by the matching server to recover the wid value, and hence again link public identities to the secret identities sid and the respective warning identity wid. We analyze how the additional capability

of linking public identities capability may help an attacker in breaking user's privacy.

An attacker (including users cooperating to break others' privacy) may be able to link public identities to times and places where these identities have been broadcast. An attacker additionally having compromised the servers may therefore be able to re-identify a user at different times and places during the same day. Thus, an attacker may be able to observe parts of the user's location history and track a user for up to one day.

We stress that even if all servers are compromised, an attacker will not be able to link public identities used on different days (assuming the use of anonymous communication and no other leakage).

Contact Information of Infected Users. Information about encounters between users is stored strictly on the user's devices. Only the meeting history (more precisely: the list of encountered public identities, without times and places of meetings) of infected users is transmitted to the central servers.

If the attacker has compromised the matching server *and* is able to link public identities used on the same day (as in the previous scenario), the attacker might be able to infer repeated meetings of the infected user, i.e. she can learn how many encounters with the same persons the infected user's device has registered within each day. If the attacker has additionally observed some of the warned public identities at specific times and places, the attacker will also learn where and when (approximately) the encounter took place, and hence learn parts of the location history of the infected user as well as the warned users.

Warnings Issued. If the attacker has compromised the matching server, she can immediately observe the public identities of all users who have been colocated with infected users. If the attacker can additionally link a public identity to a specific individual, the attacker can conclude this person has received a warning. (Note that a similar attack is possible in the DP3T protocol [TPH+20], but even without compromising a server.) Linking a public identity to a specific individual will require learning the public identity in the first place, which (again) requires staying in proximity to the user for approximately 15 minutes.

6.2 Security

We now analyze an attacker's ability to cause false negatives or false positives. As stated above, we consider a coalition of malicious users, who may be deviating from the prescribed protocol. However, we assume medical personnel as well as the central servers do not participate in such attacks, i.e. they follow the protocol.

Creating False Negatives. A false negative occurs when an uncorrupted user A has been in colocation with an uncorrupted infected user B and no corrupted user was present during the colocation, but either A does not receive a warning, or she cannot prove the possession of the warning to the medical personnel.

A warning is issued by the warning server publishing the respective warning identity, which is a Pedersen commitment to A's real identity.

Observe that once B's device has recorded an encounter with A's public identity pid, corrupted users can no longer interfere with the delivery of the warning: Once B is tested positively for the SARS-CoV-2 virus he sends a list containing A's public id pid to the medical personnel, who forwards it to the matching server, which again forwards the respective sid to the warning server, who will decrypt sid to recover the warning identity wid. Then wid is published for everyone to see.

Thus, if the warning identity is published by the warning server, the medical personnel can verify A's warning identifier has in fact been published there, and A will be able to give a zero-knowledge proof about knowing the unveiling information, while the unveiling information has not been deleted.

Thus, an attacker wanting to produce a false negative must prevent B's device from registering the encounter with A's public identity pid. This, however, requires the attacker being able to interfere with the local BLE communication between them, and thus to be in proximity to them while the encounter is taking place.

This shows our protocol achieves the required security guarantees regarding false negatives.

Creating False Positives Regarding Honest Users. An honest user A is subject of a false positive if she has not been colocated with an infected user, but she nonetheless receives a warning. Our security goal is to prevent false positives, unless the attacker has been in proximity to both an infected user and A.

In order to cause a warning for A, an attacker must have the warning server publish one of her warning identifiers, i.e. one of A's public identifiers pid must be uploaded to the matching server by a medical professional, and hence an infected user B must present a list including A's pid.[15]

If B is uncorrupted, the attacker must trick B's device into registering an encounter with A's pid. B's device will register the encounter when having received sufficiently many shares of A's pid. Since the application discards shares with the same random identifier m if they are sent too quickly, the attacker needs to be in proximity with B for approximately 15 minutes. (If B is corrupted, this part of the attack can be skipped.)

In any case, the attacker needs to learn one of A's public identities. As argued in Section 6.1, this requires the attacker to be close to A for approximately 15 minutes, due to the secret sharing scheme employed.

This concludes our argument that producing a false positive for an honest user requires proximity both to the honest user and to an infected user.

[15] Since the public identities stored on the matching server are created as hash values of (rerandomizations of) A's sid, collisions between differing public identities are very unlikely.

Creating False Positives Regarding Corrupted Users. We now analyze corrupted users' ability to prove possession of warnings. Since the medical personnel retrieves the list of warning identities from the notification server via an authenticated channel, the attacker can only prove possession of warnings regarding warning identities published by the notification server. Let wid be the warning identity the attacker wants to prove ownership of.

If wid was generated by an honest user, it is a Pedersen commitment to the real identity rid of the honest user created with a decommitment value u. Since honest users never share the unveiling information (not even with medical personnel), we consider it unlikely an attacker learns the value u.

Thus, proving the ownership of wid would require the attacker to present her real identity rid' and a zero-knowledge proof showing she knows a corresponding unveiling information u'. However, since the Pedersen commitment scheme is computationally binding (under the discrete logarithm assumption in $\mathbb{G}$), and if the zero-knowledge proof system is sound, the attacker will not be able to forge a proof.

If wid was generated by a corrupted user, the attacker may be able to prove ownership of the respective warning identity wid. In this case, however, the attacker will have to make sure one of the public identities pid derived from wid is reported to the matching server by medical personnel, and hence an infected user must hand out a list containing pid to the medical personnel.

If the infected user is corrupted herself, this is trivial. if the infected user is honest, causing her to output pid requires her device registering an encounter. Since the application is rate-limiting the reception of shares of public identities, this requires the attacker to stay in proximity with the infected user for approximately 15 minutes.

This concludes our discussion of the combined protocol's security properties.

7 Summary

We showed a modular approach with several alternatives and trade-offs to achieve contact tracing in a privacy preserving manner. Leakage of private information that was previously thought as inevitable has been shown to be unnecessary after all, by decoupling the identities used for warning at-risk users from the information that is broadcast locally and can be observed by other users. However, our improvements introduce new challenges (e.g. our improvements reinforce the need for protection against Sybil attacks). To address these challenges, our approach requires some additional protective measures, and we highlighted how these can be implemented using existing techniques.

Moreover, we introduce protection against a side-channel assisted linking of different broadcasts by randomizing the starting points of broadcast blocks of secret shares. In order to reduce the required trust into the central server components, we described how the server's functions may be separated by distributing core functions to different organizations, which removes the single point of failure regarding privacy of a purely centralized contact tracing application.

We argued that, even if all servers are compromised and colluding with malicious participants, our protocol still achieves almost the same privacy guarantees as previous works, such as [TPH+20]. Thus, in conclusion we argue that our protocol represents an overall improvement regarding security and privacy, while still being relatively practical.

However, many questions remain open. Finding an "optimal" trade-off between utility, privacy, robustness and performance for contact tracing applications is a delicate question which requires a careful consideration, not just by scientists, but by society as a whole.

Acknowledgements

We would like to express our gratitude to Michael Klooß for helpful comments. The authors were supported by the Competence Center for Applied Security Technology (KASTEL).

References

[AG20] Apple and Google. *Privacy-Preserving Contact Tracing*. 2020. URL: https://www.apple.com/covid19/contacttracing/ (visited on 04/17/2020).

[AHL18] T. Altuwaiyan, M. Hadian, and X. Liang. "EPIC: Efficient Privacy-Preserving Contact Tracing for Infection Detection". In: *2018 IEEE International Conference on Communications, ICC 2018*. IEEE, 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422886.

[BBH+20] J. Bell, D. Butler, C. Hicks, and J. Crowcroft. "TraceSecure: Towards Privacy Preserving Contact Tracing". In: *ArXiv e-prints* (Apr. 8, 2020). ID: 2004.04059 [cs.CR].

[BBV+20] A. Berke, M. Bakker, P. Vepakomma, R. Raskar, K. Larson, and A. '. Pentland. "Assessing Disease Exposure Risk with Location Data: A Proposal for Cryptographic Preservation of Privacy". In: *ArXiv e-prints* (Mar. 31, 2020). ID: 2003.14412 [cs.CR].

[BRS20] S. Brack, L. Reichert, and B. Scheuermann. *Decentralized Contact Tracing Using a DHT and Blind Signatures*. Apr. 8, 2020. Cryptology ePrint Archive, Report 2020/398.

[C20] Chaos Computer Club e.V. *10 requirements for the evaluation of "Contact Tracing" apps*. Apr. 6, 2020. URL: https://www.ccc.de/en/updates/2020/contact-tracing-requirements (visited on 04/06/2020).

[CGH+20] J. Chan, S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, S. Singanamalla, J. Sunshine, and S. Tessaro. "PACT: Privacy Sensitive Protocols and Mechanisms for Mobile Contact Tracing". In: *ArXiv e-prints* (Apr. 7, 2020). ID: 2004.03544 [cs.CR].

[CHK+06] J. Camenisch, S. Hohenberger, M. Kohlweiss, A. Lysyanskaya, and M. Meyerovich. "How to win the clonewars: efficient periodic n-times anonymous authentication". In: *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006*. Ed. by A. Juels, R. N. Wright, and S. D. C. di Vimercati. ACM, 2006, pp. 201–210. DOI: 10.1145/1180405.1180431.

[CIY20] H. Cho, D. Ippolito, and Y. W. Yu. "Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs". In: *ArXiv e-prints* (Mar. 25, 2020). ID: 2003.11511 [cs.CR].

[CTV20] R. Canetti, A. Trachtenberg, and M. Varia. "Anonymous Collocation Discovery: Harnessing Privacy to Tame the Coronavirus". In: *ArXiv e-prints* (Mar. 30, 2020). ID: 2003.13670 [cs.CY].

[D20a] DP-3T Project. *Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems*. Apr. 21, 2020. URL: https://github.com/DP-3T/documents/blob/master/Security%20analysis/Privacy%20and%20Security%20Attacks%20on%20Digital%20Proximity%20Tracing%20Systems.pdf (visited on 04/21/2020).

[D20b] DP-3T Project. *Security and privacy analysis of the document 'PEPP-PT: Data Protection and Information Security Architecture'*. Apr. 19, 2020. URL: https://github.com/DP-3T/documents/blob/master/Security%20analysis/PEPP-PT_%20Data%20Protection%20Architechture%20-%20Security%20and%20privacy%20analysis.pdf (visited on 04/21/2020).

[D20c] DP-3T Project. *Security and privacy analysis of the document 'ROBERT: ROBust and privacy-presERving proximity Tracing'*. Apr. 22, 2020. URL: https://github.com/DP-3T/documents/blob/master/Security%20analysis/ROBERT%20-%20Security%20and%20privacy%20analysis.pdf (visited on 04/25/2020).

[F20] Fraunhofer AISEC. *Pandemic Contact Tracing Apps: DP-3T, PEPP-PT NTK, and ROBERT from a Privacy Perspective*. Apr. 27, 2020. Cryptology ePrint Archive, Report 2020/489.

[FMP+20] J. K. Fitzsimons, A. Mantri, R. Pisarczyk, T. Rainforth, and Z. Zhao. "A note on blind contact tracing at scale with applications to the COVID-19 pandemic". In: *ArXiv e-prints* (Apr. 10, 2020). ID: 2004.05116 [cs.CR].

[FS86] A. Fiat and A. Shamir. "How to Prove Yourself: Practical Solutions to Identification and Signature Problems". In: *CRYPTO 1986, Proceedings*. Ed. by A. M. Odlyzko. Vol. 263. LNCS. Springer, 1986, pp. 186–194. DOI: 10.1007/3-540-47721-7_12.

[K+20] D. Kaafar et al. *Joint Statement on Contact Tracing*. Apr. 19, 2020. URL: https://drive.google.com/file/d/1OQg2dxPu-x-RZzETlpV3lFa259Nrpk1J/view (visited on 04/25/2020).

[KBS20] C. Kuhn, M. Beck, and T. Strufe. "Covid Notions: Towards Formal Definitions – and Documented Understanding – of Privacy Goals and Claimed Protection in Proximity-Tracing Services". In: *ArXiv e-prints* (Apr. 16, 2020). ID: 2004.07723 [cs.CR].

[P20a] PePP-PT e.V. i.Gr. *Pan-European Privacy-Preserving Proximity Tracing*. 2020. URL: https://www.pepp-pt.org/content (visited on 04/17/2020).

[P20b] PePP-PT e.V. i.Gr. *PEPP-PT NTK High-Level Overview*. 2020. URL: https://github.com/pepp-pt/pepp-pt-documentation/blob/master/PEPP-PT-high-level-overview.pdf (visited on 04/25/2020).

[P20c] PePP-PT e.V. i.Gr. *ROBust and privacy-presERving proximity Tracing protocol*. 2020. URL: https://github.com/ROBERT-proximity-tracing/documents (visited on 04/25/2020).

[P20d] K. Pietrzak. *Delayed Authentication: Replay and Relay Attacks on DP-3T*. Apr. 3, 2020. Cryptology ePrint Archive, Report 2020/418.

[P20e] D. Project. *FAQ: Decentralized Proximity Tracing*. 2020. URL: https://github.com/DP-3T/documents/blob/master/FAQ.md (visited on 04/28/2020).

[P91] T. P. Pedersen. "Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing". In: *CRYPTO 1991, Proceedings*. Ed. by J. Feigenbaum. Vol. 576. LNCS. Springer, 1991, pp. 129–140. DOI: 10.1007/3-540-46766-1_9.

[RCC+20] R. L. Rivest, J. Callas, R. Canetti, K. Esvelt, D. K. Gillmor, Y. T. Kalai, A. Lysyanskaya, A. Norige, R. Raskar, A. Shamir, E. Shen, I. Soibelman, M. Specter, V. Teague, A. Trachtenberg, M. Varia, M. Viera, D. Weitzner, J. Wilkinson, and M. Zissman. *The PACT protocol specification*. Apr. 8, 2020. URL: https://pact.mit.edu/wp-content/uploads/2020/04/The-PACT-protocol-specification-ver-0.1.pdf (visited on 04/25/2020).

[S89] C. Schnorr. "Efficient Identification and Signatures for Smart Cards". In: *CRYPTO 1989, Proceedings*. Ed. by G. Brassard. Vol. 435. LNCS. Springer, 1989, pp. 239–252. DOI: 10.1007/0-387-34805-0_22.

[scA+20] stop-covid.tech, covid-watch.org, M. Ardron, et al. *Unified research on privacy-preserving contact tracing and exposure notification for COVID-19*. 2020. URL: https://bit.ly/virustrackertracker.

[T] TCN Coalition. *A Global Coalition for Privacy-First Digital Contact Tracing Protocols to Fight COVID-19*. URL: https://tcn-coalition.org/.

[T20] Q. Tang. *Privacy-Preserving Contact Tracing: current solutions and open questions*. Apr. 14, 2020. Cryptology ePrint Archive, Report 2020/426.

[TOR] The Tor Project, Inc. *TOR Project*. URL: https://www.torproject.org/ (visited on 04/22/2020).

[TPH+20] C. Troncoso, M. Payer, J.-P. Hubaux, M. Salathé, J. Larus, E. Bugnion, W. Lueks, T. Stadler, A. Pyrgelis, D. Antonioli, L. Barman, S. Chatel, K. Paterson, S. Capkun, D. Basin, J. Beutel, D. Jackson, B. Preneel, N. Smart, D. Singelee, A. Abidin, S. Gürses, M. Veale, C. Cremers, R. Binns, and C. Cattuto. *Decentralized Privacy-Preserving Proximity Tracing*. Apr. 12, 2020. URL: https://github.com/DP-3T/documents/raw/master/DP3T%20White%20Paper.pdf.

[V20] S. Vaudenay. *Analysis of DP3T*. Apr. 8, 2020. Cryptology ePrint Archive, Report 2020/399.