Contact tracing solutions for COVID-19: applications, data privacy and security

Gustavo Betarte, Juan Diego Campo, Andrea Delgado, Laura González, Álvaro Martín, Rodrigo Martínez
Instituto de Computación, Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay, 11300
{*gustun, jdcampo, adelgado, lauragon, almartin, rodmart*}*@fing.edu.uy*

and

Bárbara Muracciole
Instituto de Derecho Informático, Facultad de Derecho, Universidad de la República,
Montevideo, Uruguay, 11200
barbara.muracciole@fder.edu.uy

Abstract

Since the beginning of 2020, COVID-19 has had a strong impact on the health of the world population. The mostly used approach to stop the epidemic is the application of controls of a classic epidemic such as case isolation, contact monitoring, and quarantine, as well as physical distancing and hygienic measures. Tracing the contacts of infected people is one of the main strategies for controlling the pandemic. Manual contact tracing is a slow, error-prone (by omission or forgotten) process, and vulnerable in terms of security and privacy. Furthermore, it needs to be carried out by specially trained personnel and it is not effective in identifying contacts with strangers (for example in public transport, supermarkets, etc). Given the high rates of contagion, which makes difficult an effective manual contact tracing, multiple initiatives arose for developing digital proximity tracing technologies. In this paper, we discuss in depth the security and personal data protection requirements that these technologies must satisfy, and we present an exhaustive and detailed list of the various applications that have been deployed globally, as well as the underlying infrastructure models and technologies they used. In particular, we identify potential threats that could undermine the satisfaction of the analyzed requirements, violating hegemonic personal data protection regulations.

Keywords: COVID-19, proximity tracing, protocols, applications, security, privacy.

1 Introduction

COVID-19 is having a strong impact on the health of the world's population. At the end of May 2021, more than 165 million people have been infected, more than 3.5 million have died from this disease and more than 1.765 million vaccines have been inoculated in recent months. As of December 2021, more than 281 million people have been infected, more than 5.4 million have died, 8.99 billion doses have been administered globally, and 57.4% of the world population has received at least one dose of a COVID-19 vaccine. It is a highly contagious virus, for this reason, quarantine and social distancing strategies have been followed throughout the world, with varied success and high social and economic costs. Although worldwide vaccination campaigns began at the end of 2020, reaching the percentages required to achieve herd immunity is a distant goal, and almost at the end of 2021, confinement measures are being strongly continued, mostly in Europe.

In Latin America, most countries find themselves in a complex situation both due to the health of their citizens and due to the impact on the economy and on people's lives. Additionally, vaccination progressed slowly in Latin America in the first semester of 2021; as of June 2021, Chile was the most advanced country

with 40.88% of its population vaccinated with two doses, followed by Uruguay with 28.29%, and several countries below 10% of its population vaccinated with two doses [1]. This is still notoriously insufficient, to the point that Uruguay was the country with the most deaths per million inhabitants in the world in the last weeks of May,[1] and several Latin American countries were in the top ten.

By December 2021, vaccination has advanced consistently, including a third dose in several countries, with only 8.3% of people in low-income countries receiving at least one dose. The percentage of people fully vaccinated against COVID-19 in Latin America is around 85% in Chile and Cuba, and 77% in Uruguay, being United Arab Emirates 91%, Portugal 89%, Malta 85%, Spain 81% and Iceland 83% over the most vaccinated countries. However, new variants of COVID-19, such as the recently discovered Omicron, but also Alfa, Beta, Gamma and Delta before, have led countries to several waves of new contagions and the application of restriction measures again, also increasing deaths mostly within non vaccinated people.

Tracing the contacts of infected people is a key part of this disease control strategy. Traditionally, it is carried out almost exclusively from interviews with infected people. Manual contact tracing is a slow, error-prone process, and weak in terms of security and privacy. Furthermore, it needs to be carried out by specially trained personnel and it is not effective in identifying contacts with strangers (for example in public transport, supermarkets, etc). For this reason, during 2020 there have been multiple initiatives for the development of technological solutions that improve the efficiency of this process.

The development of applications for smart cell phones capable of doing digital proximity tracking is the widely accepted and advanced technological solution worldwide. These applications take advantage of GPS and/or Bluetooth technologies to identify cell phones that have been within a short distance for a certain period of time. If a cell phone carrier is diagnosed with COVID-19, it will then be possible to warn those who have been close to the infected in order to consult the health services early.

The technological companies Apple and Google announced in April 2020 the joint development of a contact tracing technology, known as *(Google/Apple) Exposure Notification (GAEN)*, to help governments and health agencies reduce the spread of the virus [2]. Proximity digital tracking (PDT) provided great potential to improve contagion control, but faced privacy and personal data protection challenges. In no case, the information that these applications collect should be used for purposes other than the control of COVID-19. The best way to minimize risks is to design the technology in such a way that the information collected is the minimum necessary to fulfill its specific mission of fighting the pandemic. However, the landscape of contact tracing applications is wide and diverse, and the privacy and security risks that they may pose are numerous and not always evident, what motivates our present analysis.

In this article we present an analysis of technological solutions based on mobile devices for digital proximity tracking, in the context of the COVID-19 pandemic, with focus on data privacy and security. It is an extension of [3] where we discussed in depth the security and personal data protection requirements that these technologies must satisfy and we presented an exhaustive and detailed list of the different applications that have been deployed globally. Likewise, the potential threats that put at risk the satisfaction of the requirements analyzed are precisely identified, thus violating principles stipulated by hegemonic regulations for the protection of personal data, such as the General Data Protection Regulation (GDPR) [4].

The contributions in this extended paper are twofold: i) we have incorporated a completely new subsection in Section 3, which reviews and discusses the identified risks for digital contact tracing solutions establishing a correspondence with the different stages and specific steps defined in the scenarios 1 (decentralized) and 2 (centralized) also presented in that section and adding three tables to illustrate this correspondence; ii) we revisited the sources of the applications´ survey we had presented in the original paper, extending the survey up to December 2021. The survey now includes 25 new digital contact tracing applications, and presents an analysis on initial results from identified studies on the contribution of the adoption of this type of applications to help reducing and controlling the pandemic.

The rest of the document is structured as follows: Sections 2 and 3 present and discuss location technologies used by the applications, the main proposed design alternatives, and the risks and security and privacy challenges presented by these technologies. In Section 4 the survey and analysis of digital proximity tracking applications deployed worldwide is presented, as well as the analysis on initial results of their adoption. In Section 5 we discuss related work. Finally, in Section 6 some conclusions are drawn up.

2 Security and privacy risks of digital contact tracing solutions

All the digital contact tracing (DCT) technologies we study in this paper are based on the use of Bluetooth technology, more precisely *Bluetooth Low Energy (BLE)*, and follow a similar tracking pattern, which we describe in what follows:

[1] `https://ourworldindata.org/covid-deaths`

1. each participating mobile device constantly transmits, through its short range communication devices, a random number (ephemeral identifier) that changes every few minutes; simultaneously, each device records the identifiers received from neighboring devices,

2. as soon as the owner of a device is notified that has been diagnosed positive and was potentially contagious for a certain period of time, it is uploaded, with the owner's consent, a record containing the identifiers that his/her device transmitted (or received) during that period of time,

3. making use of this record it is possible to verify if a user was in contact with an infected person. If it is determined that there was a contact, the person is notified that he/she was exposed to the virus and the steps to follow (request a test, stay in quarantine, or whatever the health authority deems appropriate depending on the case).

Although the reviewed proposals differ in several aspects, such as the generation of the identifiers or the calculation of the exposure risk, this high level description makes it possible to identify privacy risks inherent to all solutions of this type. Numerous researchers in the domain of cryptography, computer security and computer law, among them the authors of DP-3T [5] and ROBERT [6], have published in the beginning of year 2020 reports that present security and data privacy analysis of contact tracing solutions [7, 8, 9].

As an initial result of our research we have consolidated a catalog of the most relevant threats and risks that have been identified in those works. They are described in subsections 2.1 through 2.7. The risks are summarized in the Table 1, using the categorization defined in [7]: Inherent Risks (IR) of proximity tracking systems, specific risks (SR) of any system that records identifiers received by BLE, general risks (GR) of any proximity tracking system that uses BLE, and risks originated by the use of network systems (NR). For ease of reading, the risk descriptions summarized in the table include a reference to the subsection in the text where each risk is presented in detail (except for IR2, whose description is self contained). It is noteworthy that most of the risks we shall describe are completely independent of the particular applications and in many cases do not require any particular computer skills.

2.1 Positive COVID-19 certified patients

A first consideration to make is that there must be a mechanism which certifies that a user actually suffers from the disease. Otherwise, if people have the ability to communicate to the system that they are carriers of the virus but this statement is not properly verified, we would be in the presence of a potential risk of abusive use of the system, which could generate, among other damages, false positives. This type of behavior would damage the credibility of the solution and discourage its adoption.

For this is reason person's infection must be confirmed by a test or a medical professional before the system broadcast the corresponding information. Although most of the studied solutions foresee the use of this type of mechanisms, they generally do not precisely specify the expected behaviour and the definition is left to the implementation carried out in each country.

Whatever the mechanism, in [7] it is discussed how an attacker can find ways to enter the system as infected:

- to be actually infected or subsequently become infected having carried out an attack.
- to pay a person who has symptoms and suspects to be infected to take the attacker's cell phone to the hospital in place of his/her own when going to take the test.
- unauthorized access to authority systems health or bribing an official to receive an infection certificate.

2.2 Data anonymity

To be able to warn an application's user that he/she has been in contact with an infected person, it is necessary for the system to handle some type of registry populated with user data. For example, the system could maintain a database of names and contact information of those people. This idea has been discarded from the beginning by the creators of the analyzed solutions for not satisfying basic privacy properties. Instead, the registry contains the ephemeral user identifiers (which may be transmitted by or received by infected persons, depending on the type of solution), or some key that allows them to be rebuilt.

These records are *pseudonymised*, which means that patients are not identified by name or some other unique identifier, but instead by a code or a number that is independent of the real identity. In the proposed systems, patient registries with COVID-19 are pseudonymised using cryptographic mechanisms. However, that number could be de-anonymized, combining it with other information in the database (the identifiers of the people who have been in contact), or outside the database (for example, collected with a Bluetooth

Risk	Impact	Description
IR1	Identify infected users	It is possible for an attacker to obtain the identity of infected people by combining the following information: 1) with whom he/she interacts at all times (obtained outside the application) and 2) the fact of having been in contact with an infected person at a specific time (obtained from the application). See Section 2.3.
IR2	Prevent Notifications	It is possible for an attacker to prevent (some) users from being notified that they were at risk of exposure, even though they have been exposed. For this, you simply decide not to participate in the proximity tracking, deactivate the application temporarily (or bluetooth) or do not send the proximity data registered in the application even though it has been diagnosed as positive.
IR3	Fake Exposure Risks	It is possible for an attacker to falsely enter the system as infected, causing, for example, other users to receive false alarms. See Section 2.1.
SR1	Reveal social interactions	If you have access to the number of IDs that were recorded in a certain time window by a device, it is possible to estimate the number of people you were in contact with during that time. If you have access to the IDs registered by a device, it is possible to confirm that you were in contact with a third party by knowing an ID issued by them during the contact. See Section 2.6.
SR2	Recognize risk of exposure	If locally recorded information is accessed on a device, it is possible to use it to counteract risk of exposure, potentially without consent, which can lead to discrimination against individuals. See Section 2.6.
GR1	Cause false alarms via BLE range extensions	It is possible for an attacker to connect their device to a powerful antenna and / or transmitter to increase the range of BLE, allowing distant devices to register it as nearby. It should then be reported as positive (see Section 2.1) to ensure that these interactions are marked as exposures at risk. See Section 2.4.
GR2	Cause false alarms through relay attacks	It is possible for an attacker to retransmit BLE signals from people who have a high probability of being diagnosed positive, for example from those who are in a testing center. See Section 2.4.
GR3	Identify locations with infected people	It is possible for an attacker to identify for example houses of infected people walking or circulating in a vehicle through an area of interest, to associate locations (houses) with infected people (see Section 2.3). There is also the risk of an RDP system being exploited to roughly geographically track infected users (see Section 2.5).
GR4	Interrupt contact discovery	It is possible for an attacker with a bluetooth jammer to interrupt communication between system users, preventing proximity contacts from being established. See Section 2.7.
GR5	Track a device with Bluetooth enabled	Activating Bluetooth has risks such as: 1- the device is traceable if the operating system does not implement MAC address randomization and disables advertisements; 2- poor synchronization between MAC address randomization and Bluetooth identifiers make the device traceable as long as the attacker stays in range. Point 1 is resolved in most operating systems; point 2 would be solved with the Apple/Google proposal. See Section 2.7.
GR6	Reveal use (or non-use) of the tracking application	The activation of Bluetooth and transmission of specific identifiers reveals to any observer (for example an attacker) that the tracking application is installed. This would not be particularly sensitive, being seen as a contribution to social good in many societies, but it would also reveal the fact that the application is not being used. See Section 2.7.
NR1	Identify infected users via network identifier	If the data of users with a positive diagnosis is uploaded directly from their mobile device, the identity of the user is exposed to a system administrator or central server. It can be mitigated by a proxy, eg. a hospital [7, pp 10]. See Section 2.3.
NR2	Identify infected users through network traffic analysis	If the data of users with a positive diagnosis is uploaded directly from their mobile device, it is possible to detect that data is being uploaded to the server and infer that it is someone with a positive diagnosis . It can be mitigated by using encryption and having healthy users upload empty data that is discarded by the server [7, pp 10]. See Section 2.3.

Table 1: Inherent, specific, general and network risks of DCT solutions.

antenna), or by use of IP addresses. It is not, therefore, a completely anonymized registration of personal data.

Certain legal frameworks, such as the General Data Protection Regulation of the European Union (GDPR), understands that personal data that has been subjected to pseudonymisation, which could be attributed to a natural person through the use of additional information, should be considered as information about an identifiable natural person. This is provided by the aforementioned Regulation in its Recital 26. This means that a pseudonymised database contains personal data and therefore the governing rules on the matter must be applied to it. The Regulatory and Control Unit of Personal Data from Uruguay has expressly collected the aforementioned criterion in Resolution No. 68/2017 of April 26, 2017, through which it approves the Criteria for Dissociation of Personal Data. In this sense, it has stated that the data processing that chooses to use pseudonymisation techniques is not excluded from the application of the regulations on personal data protection, as there would be processes of dissociation or anonymization of the data. Confusion between the terms pseudonymisation, dissociation and anonymization is common. However, from the legal point of view they have different meanings and consequences. Clearly identifying the technical process used by the technologies under study is decisive for knowing what the applicable legal regime will be and assess its compliance by the person responsible for processing the information. Under this hypothesis, additionally, we are dealing with health data, sensitive and especially protected by the law, whose treatment is strongly restricted, subject to prior, express and written (in some jurisdictions, such as Uruguay) informed consent, or protected by certain exceptions assessed. Accordingly, the law requires strong security measures to be in place. In this regard, can we affirm that the pseudonymisation of data constitutes a strong security measure? At least from the point of view of the aforementioned regulations, the answer is no.

2.3 Identification of infected people

Although the pseudonyms of patients do not directly reveal their identities, users can in some cases infer information about other users as soon as they find out that a person who they have been physically close to in the last two weeks has become ill. It is important to note that it is not necessary for a malicious user to have absolute certainty of the identity of an infected person to take some harmful action, such as disclosing the identity of the suspected infected person among her/his acquaintances or on social networks. Two simple examples (scenarios 2 and 3 in [9]) illustrate the limitations inherent to this type of system. In the former, a person who only leaves his/her house to go to the warehouse receives an exposure notification; the user concludes (perhaps mistakenly) that the grocer is infected. In the latter, a person receives a notification and begins inquiring with acquaintances and coworkers who could be the patient. From these inquiries concludes (perhaps erroneously) that it is a neighbor doctor and that infected the entire neighborhood. All the reviewed contact tracing systems can also reveal if a person in particular is infected. To know this information about a specific person we can use a phone in which we install the application and that we only use it when we are in contact with this person. This phone will record the contact, and if the person is diagnosed positive, our phone will receive an alert. For the purpose of illustrating this type of abuse of the system, the following example is presented in [9]. A company intends to recruit a temporary employee. They want to make sure the candidate does not get sick between the job interview and the signing of the contract. Therefore, they use a dedicated phone that is turned on only during the interview and that will receive an alert if the candidate later tests positive for the disease. Some slightly more sophisticated attacks reported consist of: locating a Bluetooth receiver together with a video camera in a crowded place, registering at the same time the BLE and the image of the person [8]; businesses that abuse of customers who install their online shopping application to cross data with contact tracing information (scenario 13 in [9]); malware that is installed on user devices (scenario 14), among others.

In addition to identifying infected users, using the same idea it is also possible to identify places inhabited or frequented by infected people [7]. To do this, an attacker travels through an area of interest, ideally during low traffic hours, registering the identifiers received by BLE and associating them with specific physical locations. When the attacker receives a contagion risk notification, he/she can associate this notification with a particular location.

Finally, there is a class of risks related to the use of communication networks to transmit sensitive information. Any contact tracing system in which infected individuals upload data to a central server from their device can potentially reveal, to both a central server administrator and an observer with the ability to monitor traffic to the server, the individual's health status.

2.4 False alarms

In all the reviewed systems it is possible to generate false alarms so that users who are not at risk receive a contact alert. One possibility, described in [7], is for a malicious user to extend the range of devices with

powerful bluetooth antennas. With this method one can spread its own identifiers in large areas (which will generate a large amount of false positives if one is infected or get the system to record it as such (see Section 2.1)). It is also possible to re-transmit the identifiers of people suspected of being positive, for example those who enter a laboratory where the test is performed. In [9] other scenarios that require less technical skill from the attacker are described. It is possible, for example, that a user who suspects being infected is bribed, forced, or simply sells the service of giving his/her cell phone to a malicious user, who then uses it to generate false positives. You can visit very crowded public places with this cell phone (generating a large number of false positives), or specific places obtaining some particular benefit (for example, generating an alert in a sports competitor, in another candidate for a position in a job interview, in a whole class so that an exam gets cancelled, etc.).

2.5 Geo-localization of infected persons

In a decentralized contact tracing scheme (such as DP-3T or GAEN), it is possible to geolocate infected users from the information exchanged using Bluetooth in the system. One way, presented in [9], is that in addition to the identifiers, users record the location where the contact was made, for example with the use of an application designed for such purposes. If many users record and share this information it is possible to observe with certain clarity the movements of the people who later become infected.

The same effect is achieved by mass placement of Bluetooth receivers in a city. This last case is studied and modeled in [10]. The result of a simulation of an attack of this style is depicted in Figure 1, where circles represent the location of users that have emitted ephemeral identifiers, estimated from the Bluetooth signal received by sensors deployed at several places by an attacker, and red circles correspond to ephemeral identifiers that are later reported as coming from infected users. This information can also help to identify infected people, deducing their identity from possible movements through the city.[2]

It is worth mentioning that in decentralized schemes carrying out this type of attack only requires a short contact time, enough to record the issuance of an identifier, but in general less than the time from which a high probability of contagion is estimated. For this type of attack to be effective in a centralized scheme it must be performed by people with access to the central server (either because they are the administrators, because they have the authority to enlist administrators, or because they manage to infiltrate the system). In this case you can know for sure which of the identifiers (and their associated locations) correspond to the same person, although the identity is unknown in principle.

2.6 Social interactions

All the contact tracing systems we have analyzed store identifiers received from other participants on the user's device, along with temporary information on when they were received. This data alone reveals potentially sensitive private information. For example, the fact that at a certain moment a large number of identifiers have been received from other devices can be used to infer that the user participated in an event with high attendance of people, for example, a demonstration, a show or a social event.

If an attacker has access to the logs of a certain device, he/she can try to infer more specific social contacts. For example, if you know some of the identifiers issued by a third-party device, you can determine if they were ever found. The attacker could also carry out a calculation of the risk of contagion from the information contained in the device, different (for example more sensitive) than that calculated by the contact tracing system, what could lead to situations of discrimination.

2.7 Risks inherent to the use of Bluetooth

The fact that the operation of a contact tracing system depends on Bluetooth technology in itself establishes risks, some of which have already been described in the previous sections. Additionally, an attacker can prevent the operation of the system in a certain area using a Bluetooth signal blocker (also known as a Bluetooth jammer). The mere fact of requiring the use of Bluetooth technology may present a security risk for users: this technology present vulnerabilities that can be exploited to carry out computer attacks on devices or violate the privacy of users (eg identifying and tracking them [11]). This is why it is not recommended to constantly keep the Bluetooth functionality of the devices turned on as a contact tracing solution would require. On the other hand, it is possible to track Bluetooth devices using their physical address, which is why most modern devices use fictitious addresses that change every few minutes. In the context of a contact tracing solution, which itself issues temporary identifiers, both the physical address and

[2]The ability to associate locations with (anonymous) users, as shown with different colored strokes in Figure 1, depends on how a decentralized contact tracing system is implemented specifically. This association is in fact simple in the implementations of DP-3T (design 1) and GAEN.

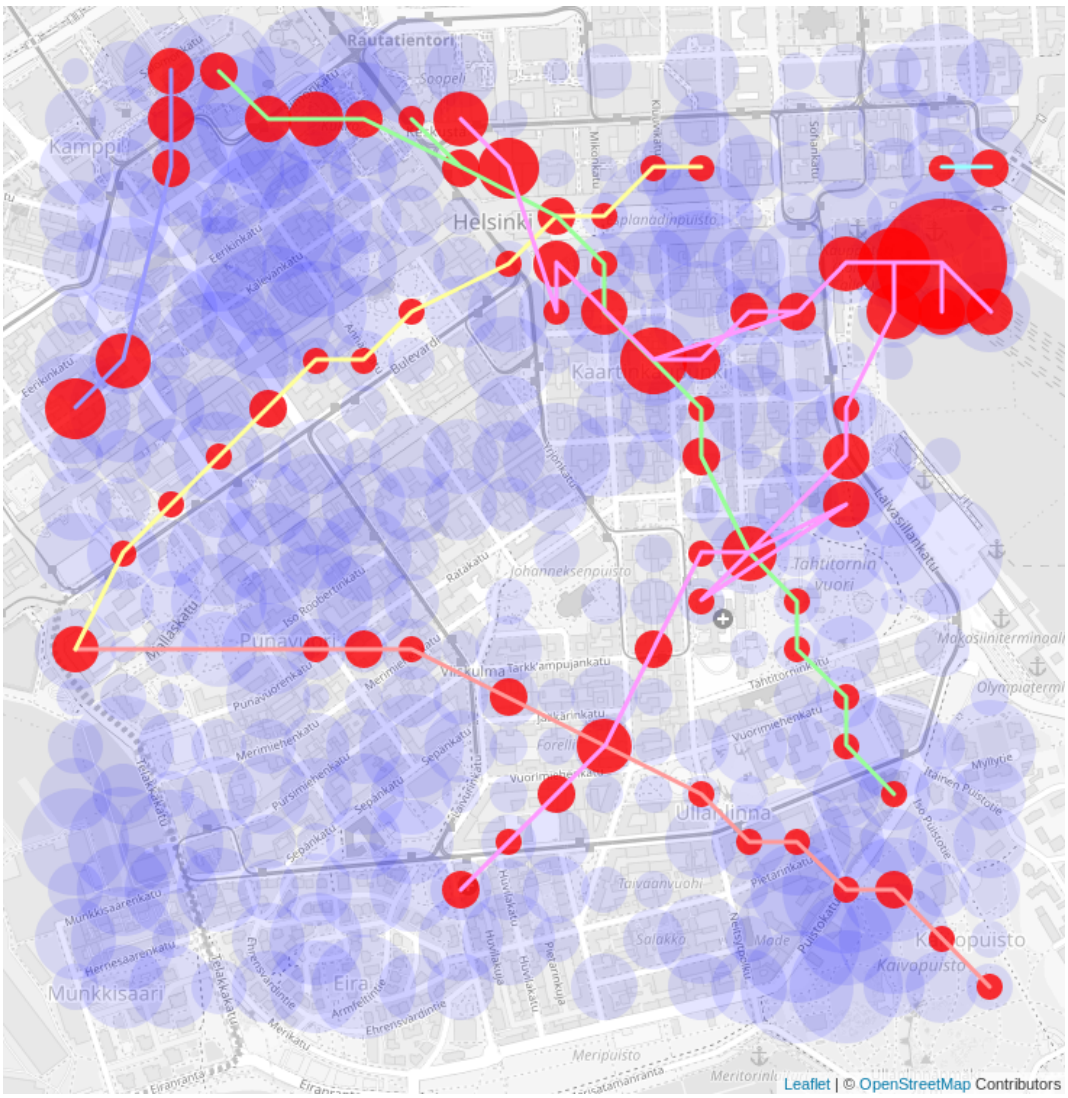

Figure 1: Simulation showing the movement of infected users in a decentralized scheme [10].

the identifier change need to be done in a synchronized manner. Otherwise, it is possible for an attacker to track the device as one of the two data changes. Since physical address management is done at the operating system level, this problem can only be properly resolved if the device's operating system supports this. For example, in mobile devices, only solutions that make use of Google and Apple technology can adequately solve this problem.

Finally, as the information broadcast via Bluetooth by contact tracing systems is visible to any observer, this reveals whether someone is using the tracking application or not. Therefore, the mere fact of putting one such system into operation puts the privacy of the individual's decision to participate or not in the system at risk.

3 Threat analysis of DCT protocols and solutions

During the first months of 2020 many different initiatives have emerged with respect to the design, implementation and deployment of DCT technologies. Some of them are being widely used by the citizens of many countries that have adopted them.

In most proposals, users broadcast their own identifiers and store identifiers received from other users through the *Bluetooth Low Energy (BLE)* interface of their mobile devices, and exchange information with a central server, following certain rules that are referred to as a *protocol*. All proposals that were implemented support IOS and Android systems. The GAEN initiative is an Application Programming Interface (API) that can be used to develop specific applications.

In general, these solutions broadly follow one of the two main design alternatives, decentralized or central-

ized, which are described in Section 3.1. In order to discuss concrete implementations of these architectures, we present an example of each one: decentralized (DP-3T) in Section 3.2 and centralized in Section 3.3.

3.1 Design alternatives

In this section we present the general characteristics of two of the approaches that have become predominant, so we can establish a reference framework for the discussion of security and privacy issues in DCT systems. These two approaches are:

1. The decentralized approach, in which the exposure verification is done in the mobile device. Examples of implementations of this design can be found in [12, 13, 5, 2].

2. The centralized approach, in which the exposure verification is done in a central server. This design is followed by [14, 6].

In Figure 2 we present a diagram that describes the essence of the decentralized approach, while in Figure 3 we present the main interactions of a centralized approach.

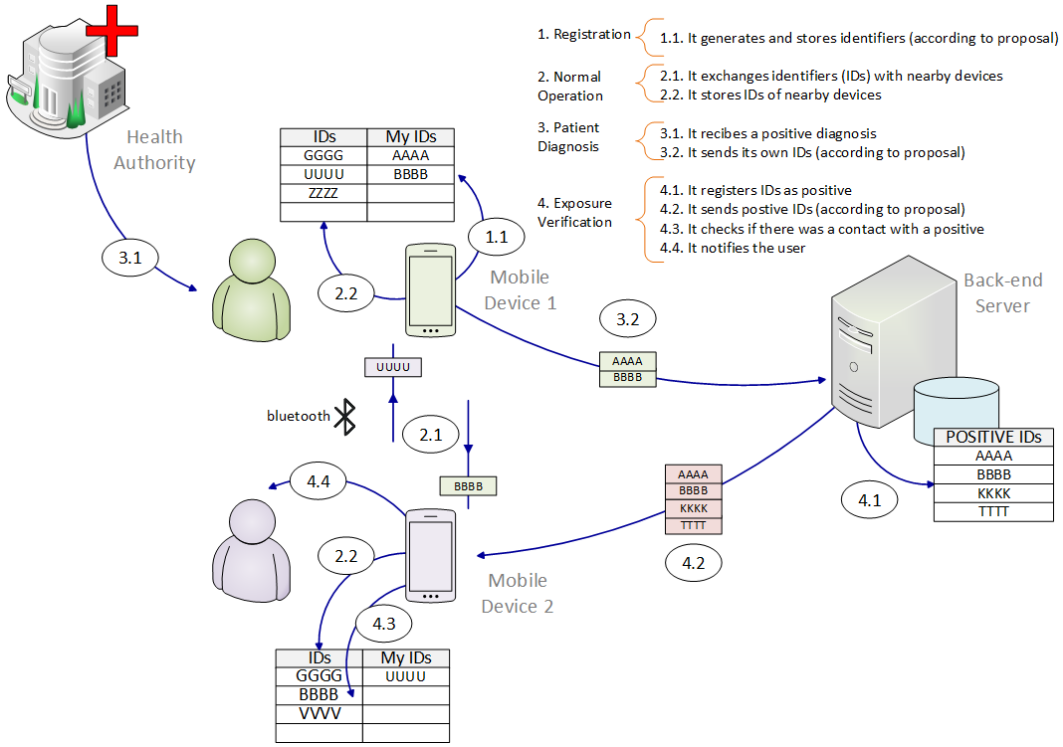

Figure 2: Decentralized DCT

In the figures we schematically show the actions that each actor performs and the interactions between them throughout the execution of each DCT system. The actors involved in both cases are: two users and their mobile devices, a central server and the health authorities. Both figures represent a scenario in which both users meet, sometime afterwards user 1 is diagnosed with COVID-19, and user 2 is notified that he/she has been exposed to the virus.

The actions and interactions are classified in four stages:

1. *Registration*: includes the steps necessary for a user to start participating in the DCT scheme.

2. *Normal operation*: refers to the operations that take place while both users are healthy.

3. *Patient diagnosis*: are the actions that are triggered once user 1 tests positive for infection.

4. *Exposure verification*: covers the steps that make user 2 aware of its exposure to the virus.

Normal operation (stage 2) is essentially the same in both schemes: user devices continually transmit and receive anonymized identifiers and locally store both of them. The ids that each device sends are generated during stage 1, and both approaches differ greatly in how they generate them. In the decentralized case,

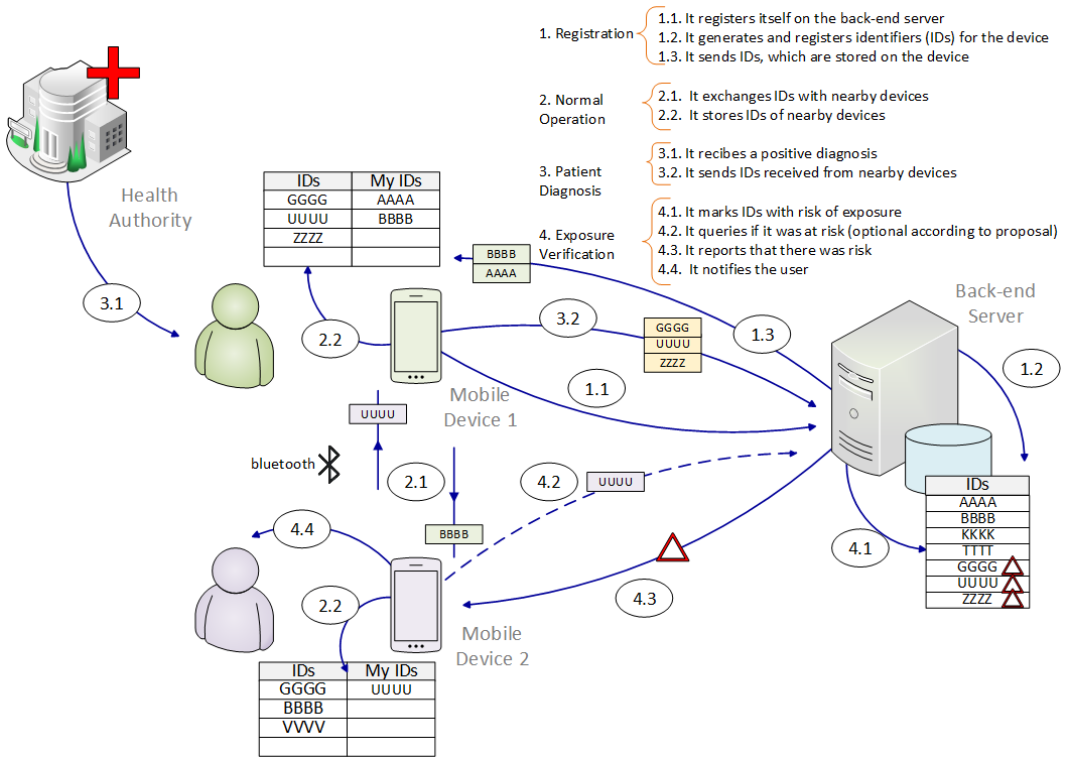

Figure 3: Centralized DCT

these ids are generated by each mobile device, while in the centralized case, it is the central server the one that generates and sends them to each device.

Important differences can also be observed in stages 3 and 4. While in the decentralized scheme user 1 sends his/her own ids to the central server when he/she is diagnosed (`AAAA` and `BBBB` in Figure 2), in the centralized scheme infected users send the ids they received from other users (`GGGG`, `UUUU`, `ZZZZ` in Figure 3). In the latter case, when the server receives id `UUUU` it detects that user 2 (anonimized) has been exposed, since it was the server that generated this id, and can notify user 2 that he/she is at risk through the mobile device. On the other hand, in a decentralized scheme the central server cannot know that user 2 was exposed given ids `AAAA` and `BBBB`, transmitted by user 1. Rather, the server broadcasts these ids to all users so that they themselves can evaluate the risk of infection. In Figure 2 the user 2 device receives id `BBBB`, detects that it has heard that id from a nearby device, and notifies the user of the situation.

the DP-3T consortium [5] (Decentralized Privacy-Preserving Proximity Tracing) has proposed a decentralized digital contact tracing solution; while INRIA and Fraunhofer AISEC developed ROBERT [6] (ROBust and privacy-presERving proximity Tracing), which is a typical example of a centralized scheme.

3.2 A decentralized DCT example: DP-3T

The DP-3T proposal includes two alternative designs, with different compromises between privacy, computing and communication requirements. We will present the first version, which is the most economic in terms of resource requirements and is very similar to GAEN.

The registration stage (stage 1) in DP-3T consists in each participating device randomly selecting a secret key, which is used to generate the ids it will send in the future. Concretely, if a certain day t a device starts participating in the scheme, it will choose a random secret key SK_t, based on which it will randomly generate a list of ids to transmit during that day, by the application of cryptographic functions to the key SK_t. This initial key will also determine the daily keys SK_i to be used from that moment on, by recursively applying the cryptographic function H:

$$SK_i = H(SK_{i-1}), \quad i > t.$$

Since each SK_i determines all ids sent by the device from day i onwards, if a user tests positive (stage 3), it is enough to publish SK_c along with the day c in which this user was first contagious in order for the rest of the users to check whether they were in contact with the infected user.

Table 2: DP-3T Analysis (decentralized scenario)

Mobile device	Back-end Server	Privacy Risks (PR)	Security Risks (SR)
1.REGISTRATION			
1.1 – Random key SK_t for the current day t is generated. The SK_t key determines a set of IDs to be issued each day from now on. Every day the device generates a random order to broadcast the IDs of that day.			
2.NORMAL OPERATION			
2.1 – Exchange of **IDs generated** with other devices in close physical proximity using BLE.		**IR1, GR3, GR5, GR6 (PR-DE1) Get locations:** Getting access to the generated IDs that are stored on a device allows the attacker to determine if the victim has been in locations he/she monitors. Partially mitigated by ID rotation and use of cryptographic tools ([7, pp 13]).	**IR2, GR1, GR2, GR4 (SR-DE1) Risk of compromised data on the device**: constant transmission and registration of IDs is required, with the system running in the background. On Apple devices this is not possible if it is not implemented using the API defined with Google; if this API is not used, workarounds, like execution with screen and device not locked and without password protection, are required. In the case of theft or request from the police, all data is exposed [15, pp 10-11]
2.2 - Storage of the IDs received from nearby devices along with the day they were received and additional information useful to assess the risk of contagion.		**SR1, SR2**	
3.PATIENT DIAGNOSIS			
3.1 - The user receives a positive diagnosis and approves sending his/her IDs.			
3.2 - Password SK_t and the first day of the contagion period are sent.		**NR1, NR2**	**IR3**
4.EXPOSURE VERIFICATION			
	4.1 – Register the keys and initial days of the contagion period voluntarily communicated by users with a positive diagnosis.		
	4.2 – Distribution of the keys and initial days of the contagion period to the rest of the users.		
4.3 – Using the keys and initial days of the contagion period of users with a positive diagnosis, it recalculates the IDs issued by these users and evaluates, comparing with locally stored information, the risk of contagion.			
4.4 – User is notified.			

During normal operation (stage 2), participating devices locally store all received ids from other nearby devices, together with the day they were received and some additional information to evaluate the risk of the exposure, such as an estimation of the distance to the other device.

In the exposure verification stage (stage 4), each user receives from the central server the keys SK_c and the first day c in which users were first contagious, published voluntarily by users that test positive. From this information, each device recomputes all transmitted ids for each day from the infected users and they

Table 3: ROBERT Analysis (centralized scenario) - Part 1

Mobile device	Back-end Server	Privacy Risks (PR)	Security Risks (SR)
1.REGISTRATION			
	1.0 – Secret key K_S, federated key K_G and country code are installed.		
1.1 – Registers with server. This requires a working demonstration (eg CAPTCHA) to avoid automated registrations or denial of service attacks. Establishes symmetric key K_A shared with the server and synchronizes times with the server. Gets IDs for the next T epochs (this step is repeated periodically).			
	1.2 - Generates unique identifier (ID_A) for the device. - Sets symmetric key K_A shared with the device. - Generates IDs that the device broadcasts during the next T epochs.	**(PR-CE1) Function creep** - making it a surveillance instrument beyond COVID-19, linking the IDs of any user, the server can: **(PR-CE1.1)** Enables tracking of individuals in time, combining with other databases to obtain the user's real identity [15, pp 2-3]. **(PR-CE1.2)** Label and classify individuals so that they can be recognized later by third parties, like the Police, without needing the server [7, pp 19 SR7].	**(SR-CE1) Key compromised** - key rotation is not specified, if keys can be obtained from the server: - secret key K_S, it is possible to know the unique identifiers, ID_A, of all users from the IDs it broadcasts. - federated key K_G, all country codes are known [15, pp 10].
	1.3 - Periodically sends IDs to registered devices to broadcast during the next T epochs.		
2.NORMAL OPERATION			
2.1 – Exchanges IDs received by BLE with other devices in close physical proximity.		**IR1, GR3, GR5, GR6**	**IR2, GR1, GR2, GR4** **(SR-CE2) Risk of compromised data on the device**: constant transmission and registration of IDs is required, with the system running in the background. In Apple this is not possible, and the new API only allows querying by IDs to verify exposure, in particular **it does not provide the functionality to send list of IDs**. Workarounds are required, like execution with screen and device not locked and without password protection. In the case of theft or request from the police, all data is exposed [15, pp 10-11].
2.2 - The IDs received from nearby devices along with the instant of time they were received are registered.		**SR1, SR2**	

are compared with the ids stored locally in the device.

3.3 A centralized DCT example: ROBERT

In ROBERT, which is a centralized DCT scheme, the central server plays a much more active role than in DP-3T. In this case, before the systems starts operation, a secret key K_S is installed in the server. This key

Table 4: ROBERT ANALYSIS (centralized scenario) - Part 2

Mobile device	Back-end Server	Privacy Risks (PR)	Security Risks (SR)
3.PATIENT DIAGNOSIS			
3.1 – Diagnosed as positive, the user approves to send his/her data.			
3.2 - Sends all the IDs received from other devices during the contagion period, along with the corresponding reception time. The registered IDs are transmitted in separate shipments (one by one), mixing the IDs communicated by other users.		**NR1, NR2** **(PR-CE2) Channel anonymization mechanism**: A mechanism must be provided to prevent the set of IDs uploaded by a user from being easily distinguishable from the IDs uploaded by another user (in [6, Section 6.1] they suggest alternatives) [15, pp 9-10].	**IR3** **(SR-CE3) False risk**: by uploading the observed IDs, the attacker could inject IDs observed by other users or devices, so that other individuals (for example, specific targets) are marked as exposed to contagion ([15] pp 10) **see also (SR-CE2)**
4.EXPOSURE VERIFICATION			
	4.1 – The IDs received by the devices of users with a positive diagnosis during their contagion periods are accessed. The unique identifier, ID_A, of the issuers of those IDs (checking the validity of each record) are obtained. Computation of the risk of exposure and mark as exposed.	**(PR-CE3) Link attacks**: allow obtaining relationships between positives that upload their records and their contacts, identify positives, co-location among non-infected. Note that in this scheme: the server knows the identifiers ID_A of the users associated with each of the IDs received from the positive ones [7, SR9 pp 21]. For this reason, this type of risk is particularly sensitive to the preservation of the pseudo-anonymity of users (**see PR-CE1.1**) **(PR-CE3.1) Binding based on timestamps**: Timestamp intersection attacks could allow estimating contacts between infected users [15, pp 4-5], [7, SR5 pp 19]. **(PR-CE3.2) Co-locations based on timestamps**: the link of infected users could also allow estimating co-locations between infected and non-infected ones, building an approximation of the social graph of contacts [15, pp 5], [7, SR6 and SR8 pp 17 and 20]. **(PR-CE3.3) Identify anonymous positives with co-locations, causality and frequency analysis**: several analyzes of the data uploaded by positives could allow to identify positives, contacts and approximate the social graph [15, pp 5- 9], [7, SR6 and SR8 pp 17 and 20].	
4.2 – Queries regularly if the user has been at risk.			
	4.3 – Reports that the user was exposed and marks the associated ID_A as notified.		
4.4 – User is notified.			

will be used to generate the ids that participating devices will transmit. For international interoperability, a federated key K_G and a country code CC_S are also configured in the server.

In stage 1, when a device enters the system, a symmetric key K_A is negotiated between the server and the device which will be used for future private communication between them. An approximate temporal synchronization is also established, so that a common temporal partition is defined between all participating devices and the server (within a small tolerance of one second at most). These intervals are called *epochs* and are enumerated from a certain initial moment in time. In addition, the server assigns a unique id to each registered device ID_A, which is not related to the real identity of the user, who should remain anonymous. Applying a cryptographic function to this id, the private keys installed in the server K_S and K_G, and the country code CC_S, the server generates and sends to the device an id to transmit in each of the next T epochs.

During normal operation (stage 2), each device transmits en each epoch the id that the server gave it for this epoch, together with a more accurate timestamp than the duration of the epoch. A cryptographic function using the key shared with the server K_A is applied to the concatenation of the id and the timestamp, and the result is included in the transmitted message. This cryptographic component is necessary to detect manipulated or replayed messages. Periodically, the device connects to the server to get new ids to transmit. Participating devices verify the temporal validity of messages they receive from others, and locally store the ids together with the time they were received.

In stage 3, when a user is tested positive, its device sends to the server the ids and reception time of all received messages during the contagious window. The sending of this information is done one by one, and mixed with the messages of other positive users, so that the server cannot know exactly which ids were received by a given user. Using its private key K_S, the server recovers the id ID_A associated to each id received by the infected user. The server then evaluates the risk of exposure of these users, and if the risk is greater than threshold, the id ID_A is marked as exposed in a local database in the server.

Exposure verification by the rest of the users (stage 4) consist in a private query (using the shared key K_A) of the device to the server. If the server detects that any ID_A assigned this device is marked as exposed, it say so to the device, which in turn will warn the user so that he/she takes appropriate measures.

3.4 Identified risks for specific solutions

In this section we review the risks summarized in Table 1 in light of each of the scenarios presented in Section 2. For this we establish a correspondence between these risks and the different stages, and specific steps, defined in scenarios 1 and 2. This mapping associates previously identified risks (see Table 1) to each specific aspect of the execution flow of each scenario, adding risks that are specific for each type of solution (decentralized, centralized). In the table we use the notation PR-DEx and SR-DEx to identify privacy and security risks, respectively, for a decentralized solution, and we use the notation PR-CEx and SR-CEx, analogously, for a centralized solution.

In Table 2 we show the specific flow of DP-3T solutions, while in the Table 3 and Table 4 the one for ROBERT. Those flows correspond to the description presented in sections 3.2 and 3.3, respectively.

In the first place, those tables show that several of the identified risks are shared by both schemes, as was already discussed in Section 2. For example, during normal operation of the protocol, both schemes are exposed to a greater or lesser extent to risks IR1, IR2 and GR1 to GR6, that is, in both it is possible to identify, track and geo-locate users, prevent notifications and cause false alarms. The specific risks associated with any system that registers identifiers received by BLE, namely SR1 and SR2, are also shared by both schemes. The same happens regarding the risks associated with NR1 and NR2 when a positive user uploads the registered identifiers to the central server.

The approaches differ, however, in the difficulty of carrying forward these attacks by different types of users. For example, for a user to be able to identify a person infected in the decentralized case it is enough to relate the identifiers received by Bluetooth with the person, for example keeping a record of who was in contact with at any given time. In the centralized case the attack requires for the user to register multiple devices in the system and, by using them, to be able to infer the moment and place of contact once he gets a notification. These differences are partly a consequence of different approaches with regard to the privacy of personal data and in the attackers' model used by each approach. In the centralized model the priority is placed on protecting the system from possible abuses by individual users, while in the decentralized model the focus is on protecting the system from possible abuse by the central authority that controls the central server or malicious actors who gain unauthorized access to it.

Additionally, there are specific risks associated with the type of scheme implemented. For example, in the case of centralized systems, where the central server plays a more important role, the permanent IDs of the users and the contacts of the positive ones are known by the server. This determines that using the recorded data a *"honest but curious server"* [15] can, for example, reconstruct the (partial) graph of social interaction around positive users and identify positive ones (PR-CE3 and sub-risks associated with SR5, SR6, SR8 and SR9 in [15]), as well as enabling tracking of locations in time and user tagging (PR-CE1

and sub-risks associated with SR7 in [15]). In this scheme there is also the risk of breaches and data leaks from the central server, which would allow an attacker to obtain the permanent IDs of the users (SR-CE1). Therefore, the security of the information located on this server becomes even more important.

4 Survey of applications

In this section we present a survey we have carried out with a specific focus on the contact tracing applications developed in the period June-October 2020, as well as the progress in existing protocols and new proposals that may have arisen. The analysis included a total of 115 applications from 65 countries on 5 continents, and 11 protocols. Although we tried to cover the widest possible spectrum, some proposals may not have been included. In this extended paper we revisited the lists of applications we used as input for the survey, extending the period to December 2021 in order to include new applications.

4.1 Survey methodology

Given the breadth of the survey to be carried out, we defined a systematic approach in order to be able to cover as many proposals as possible, based on a variety of formal and informal sources, checking the details of each proposal with specific searches.

As a first step, a search of application listings was carried out with the search strings "contact tracing apps COVID-19" and "digital contact tracing apps COVID-19" in Google, also using "applications" instead of "apps", from which the following sources were selected:

- Council of Europe [16]
- MIT Technology Review [17]
- Linux Foundation (LF) Public Health Landscape [18]
- Wikipedia COVID-19 apps [19]
- Wikipedia Exposure Notification [20]
- XDA Developers [21]

The references [16, 17, 18] were selected for their reliability, and the sources [19, 20, 21] as a possible complement and/or extension of information not provided in the previous ones. The list of applications of [16] was taken as a basis, considering only those that correspond to the "contact tracing" category since this list includes all types of applications such as information, self-diagnosis, quarantine control, among others. We copied them into our own spreadsheet with a selection of columns from that report (a), adding our own defined columns (b) that we considered relevant for the survey:

- (a) Country, application name, Origin (Government, Private, Organization, Multiactor), Protocol (name of the protocol used by the application, e.g., DP-3T, GAEN, etc.), Model (centralized, decentralized, semi -centralized, hybrid).
- (b) Continent (Africa, America as: South America, North America, Central America and the Caribbean, Asia, Europe, and Oceania, and the 7 shared countries reported as Asia-Europe), Author (government authority, organization, consortium, company, etc.), license (Apache, MIT, Mozilla, GPL, Proprietary, etc.), URL (official if possible), source code repository if released (github, gitlab or similar), source (from table 1).

In case that it was not possible to obtain the information, it was included as NA (*Not Available*). Then we extended the base list with the contact tracing applications from [17], [18], and [19] that were not included in [16] as well as information on columns not present in [16]. Then we checked with [20] and [21] that the applications based on GAEN were already identified, otherwise they were added.

When revisiting the lists of applications several were not maintained any more, such as [16] and [17].

Once the base list was generated, we carried out specific searches on each application and country to obtain detailed information on each application, from government and/or similar sources, as well as data on the release of the corresponding source code and repository, which in general were not available from the reviewed sources. The complete spreadsheet of this survey is available online.[3]

It is worth mentioning that in the column model we show the model that each application states that is followed by its implementation, with values: decentralized, centralized, semi-centralized (ej. based on a

[3] `https://doi.org/10.5281/zenodo.5806399`

centralized protocol adapted) and hybrid (based on both centralized and decentralized protocols), mostly based on the type of protocol it includes. Values for the column protocol include the two ones we described and discussed in Section 3.1: DP-3T as example of a decentralized model, and PEPP-PT as example of a centralized model, but there are also other protocols for each model, and a hybrid one. In Table 5 we present each protocol and the meaning of the acronym, when defined (links are provided in the spreadsheet).

Table 5: Protocols for each model in the survey

Model	**Protocol**
Decentralized	DP-3T - Decentralized Privacy-Preserving Proximity Tracing GAEN – Google/Apple Exposure Notification TCN – Temporary Contact Numbers Protocol Whisper Tracing Protocol PACT – Private Automated Contact Tracing OpenCovidTrace Safe2
Centralized	BlueTrace PEPP-PT - Pan-European Privacy-Preserving Proximity Tracing ROBERT - ROBust and privacy-presERving proximity Tracing ReCoVer ShareTrace
Hybrid	Herald

In addition, some applications indicate they implement a centralized or decentralized model based directly on the use of the technologies presented in Section 3 such as GPS and Bluetooth, some adding also QR codes to the application.

4.2 Results analysis

In the first survey of applications (September 22, 2020) a total of 103 applications were obtained from 65 countries on the 5 continents. The survey was updated to October 31, 2020, adding 12 applications, all in the United States. In the extension presented in this article, we revisited the lists of applications, adding 25 new applications, from 3 different continents (America, Asia and Europe), for a total of 140 applications. Table 6 shows the summary of applications by continent and country, including the new ones we added in this extension.

Table 6: Number of applications and countries with applications by continent

Continent		**Apps** IS**	**Ex**	**Total**	**Countries with Apps**	**Total countries**	**%Countries by region**	**%Countries in total**	**%Apps in total**
Africa		4	0	4	3	54	5,5	1,6	2,9
America	North America	35	12	47	3	3	22,8	4,1	37,9
	South America	5	0	5	4	13			
	Central America and the Caribbean	0	1	1	1	19			
Asia		26	9	35	28	42*	66,7	14,4	25
Europe		36	2	38	30	43*	69,8	15,5	27,1
Eurasia		6	1	7	5	7*	71,4	2,5	5
Oceania		3	0	3	3	14	21,4	1,5	2,1
Total		115	25	140	77	194	–	39,7	100

**IS=Initial survey, Ex=Extension *total of countries in Europe and Asia does not include Eurasia

The country with the higher number of applications continues to be the United States with 42 applications (December 2021) from different states and organizations including protocols, adding 13 new applications to the 29 existing applications (October 2020), followed by India with 5 applications (not adding new ones). From other countries that also have a large territory and population, such as China, Russia, Brazil, and

Mexico, information was only accessed for a few applications, including this revision, which probably are not the only existing ones.

In the first analysis, of the 115 applications, 64 follow a decentralized model (55.65%) with a variety of protocols and 29 follow a centralized model (25.22%) with a diversity of protocols as well, 1 uses a semi-centralized model based on PEPP -PT, 1 hybrid based on the Herald protocol, and 20 could not be identified (NA). In this extension, from the 140 applications, 88 use the decentralized model, increasing its adoption to 62.85%, also counting applications that used a different approach before, such as the Smittestopp ("Stop Infection") from Norway which was updated with GAEN. The centralized model is used by 31 applications (22.14%), and the rest is the same being now 19 not identified (NA). The detail of protocols for each model is presented in Table 7 for the 115 applications surveyed in October 2020, adding the 25 new ones from this extension.

Table 7: Number of applications by model and protocol used

Model	**Protocol/ Technology**	**Number** IS/Ex/Total**			**% en total IS/Ex**	
Decentralized	GAEN	39	21	62*	33,91	44,28
	DP-3T, GAEN	8	0	8	6,95	5,7
	DP-3T	2	0	2	1,74	1,43
	TCN, GAEN	1	0	1	0,87	0,71
	TCN	4	0	4	3,48	2,85
	PACT, GAEN	1	0	1	0,87	0,71
	GAEN, QR code	0	1	1*	0	0,71
	GPS, Bluetooth, QR code	7	0	7	6,1	5
	Whisper tracing	1	0	1	0,87	0,71
	Safe2	1	0	1	0,87	0,71
Total Apps Decentralized		**64**	21	88*	**55,65**	**62,86**
Centralized	GPS, Bluetooth, QR code	20	1	21	17,4	15
	Bluetrace	5	1	6	4,35	4,29
	ROBERT	1	0	1	0,87	0,71
	PEPP-PT NTK	1	0	1	0,87	0,71
	ShareTrace	1	0	1	0,87	0,71
	ReCoVer	1	0	1	0,87	0,71
Total Apps Centralized		**29**	3	31*	**25,22**	**22,14**
Semi-centralized	PEPP-PT based	**1**	0	1	**0,87**	**0,71**
Hybrid	Herald	1	0	1	**0,87**	**0,71**
Not available	NA	**20**	1	19*	**17,39**	**13,57**
Total Apps		**115**	**25**	**140**	**100**	**100**

**IS=Initial survey, Ex=Extension; *two from NA change to GAEN,one from centralized

It is worth highlighting that in the initial survey, of the 115 applications, 51 applications (44.3 %) were open source with available repositoriesmostly on github or gitlab, but in the extension only 3 new have repository and 1 released its code, totalling 55 applications (39.3 %). The licenses used are: 16 Apache 2.0, 12 MIT, 9 GNU GPL 3.0, 3 GNU AGPL, 6 Mozilla PL 2.0, 2 European UPL 1.2, 2 Mozilla PL 2.0 / European UPL 1.2, 1 Apache 2.0 / MIT, 1 Mozilla PL 2.0 / MIT. One license is proprietary (Australia), and of 2 licenses no detail was found (Brazil and ShareTrace). Of the rest with no code available in the initial survey, 64 applications, of which 60 applications have a proprietary license (52.2 %), 4 are not available NA (3.5 %) including 1 to be released (Canada without date). In the extension, the applications without available code rose to 85 of which 82 have a proprietary license (44.3 %).

Additionally, of the surveyed protocols that total 11 proposals, 7 present a decentralized model (63.6 %), 3 centralized (27.3 %) and 1 hybrid (9.1 %). 9 are open source (81.8 %) with repository, and an implementation / specification available, while two have no accessible repository. The licenses are: 3 MIT, 3 GPL 3.0, 3 Mozilla PL 2.0 and 2 Proprietary (GAEN, Safe2).

Several of the surveyed applications were questioned both by the approach and technologies used and by the data collected and their treatment, as well as the expected storage period.

In South Korea [22] the Corona 100m mobile application (not included in the lists) collected data published by the government on positive people including nationality, age, gender and locations that were visited by these people (GPS), alerting the user when approaching 100m from these areas. This application was

removed from google play. Other web applications such as CoronaMap[4] and CoronaPath published this data on a map on the web (they would not be active). In [23] privacy risks on the public data in South Korea are discussed.

Amnesty International's security lab [24] conducted an analysis of digital contact tracing applications from Europe, the Middle East and North Africa including: Algeria, Bahrain, France, Iceland, Israel, Kuwait, Lebanon, Norway, Quatar, Tunisia and the United Arab Emirates, rating them from "bad" to "dangerous" for human rights. Bahrain apps 'BeAware Bahrain', Kuwait 'Shlonik' and Norway 'Smittestopp' were singled out as highly invasive surveillance tools for people's privacy, with all three tracking people's locations in real time (or near real time), uploading GPS coordinates to a central server. Norway withdrew its first app in June [25], androlled out a completely new GAEN-based one in December 2020[5]. Important security vulnerabilities were detected in other applications.

In [26, 27] the adoption requirements for digital Contact tracing applications to be effective in contributing to the traceability of infections and the arrest of the pandemic are analyzed, which should be at least 60% of the population. As of October 2020, the highest adoption rates were around 40% in Iceland, Singapore and New Zealand, around 30% in Ireland and Finland, around 22% in Germany and the United Kingdom, over 25% in Australia, 14% in Italy, 4% in France. As of the first four months of 2021, these percentages had increased in several countries, for example, to 52% in Ireland, 33% in Germany, 28% in the United Kingdom, 15% in France, continue rising to 45% in Finland, and 38% in Norway for the new GAEN-based application deployed. In Singapore, the use of contact tracing tokens was added with a focus on older adults who do not use mobile devices. In Uruguay as of 10/26/2020 the coronavirus.uy application had been downloaded by more than 616,000 devices(around 18 %) and as of February 2021 by nearly 1,400,000 devices(around 40 %).

The European Union launched at the end of 2020 an interoperability gateway[6] so that digital proximity tracking applications can be used when traveling between European countries.

It was not possible to find official public data on the results of the use of the applications in the different countries, such as indicators of contribution to the traceability of these applications in terms of greater scope and shorter times in the contact traceability process. Although in [26] they also stated that a smaller number of users of these applications will also have a positive effect in helping avoid infections, their effectiveness is still being analyzed, with a few papers publishing reliable results on this. A systematic review [28] identified 19 studies published from January 2020 to March 2021, analyzing the effect of digital contact tracing apps on the reduction of the effective reproduction number (Reff) (i.e. "the average number of secondary cases generated by a single infection case") and on the number of infections. This review also included an analysis on the minimum percentage of people using the app (apps uptake) to control the pandemic, the relation between apps uptake level and effectiveness of digital contact tracing, and a comparison between digital and manual contact tracing, among other data.

Regarding the reduction on the number of infections, studies reported different ranks of apps uptake ranging from 50% to 80% leading to heterogeneous impact on the reduction of cases. As the studies used different metrics and models, their results are not comparable, but the review stated that "they provide evidence that a high uptake leads to the mitigation of the epidemic", since the 19 studies showed reduction of Reff and the number of infected cases. Regarding the effectiveness of digital contact tracing, it presents a quadratic dependence from the apps uptake rate, for which a high value of apps uptake is needed to control the epidemic only with digital contact tracing, being 90% the rate in most studies reporting on that. They concluded that digital contact tracing should be combined with manual contact tracing. Finally, regarding the comparison between digital contact tracing and manual contact tracing, the results are contradictory since from the eight studies reporting on that, four concluded that digital is better than manual, two only marginal gains of digital and two that manual is better than digital, for which further research is needed.

5 Related Work

The security and privacy challenges of technologies and solutions used to implement digital contact tracing have been studied and analysed in various works. In [29] the authors provide an analysis of digital contact tracing solutions in terms of methodologies/technologies and discuss open challenges. In [30] common architectures for contact tracing applications are described as well as popular applications using them. This paper also analyzes attacks that could be possibly performed in these architectures, elucidates users' concerns and discusses research directions. In addition, the paper [31] presents a review of contact tracing architectures

[4] https://coronamap.site/

[5] https://www.helsenorge.no/en/smittestopp/

[6] https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/travel-during-coronavirus-pandemic/how-tracing-and-warning-apps-can-help-during-pandemic_es

and the apps that have been deployed in Europe. That paper also describes an adversarial model in order to analyse frameworks with respect to security and data protection concerns. In [32] an anlysis is put forward of different types of security and privacy threats to which users of different digital contact tracing solutions can be exposed. In [33] the results of conducting a systematic literature review focused on this problem are presented and discussed. In particular, it is provided a review of current challenges of contact tracing apps, recommendations to address these challenges are discussed, and the paper also expore future directions as well as considerations regarding digital contact tracing technologies.

There are also other works that address security and privacy challenges of technologies and solutions to implement digital contact tracing [34, 35, 36, 37, 38, 39]. Most of these papers describe, and present in detail, the base technologies that have been used for the implementation of different contact tracing applications. Compared to our work, these papers are usually broader than ours with respect to the aspects they cover (e.g. scalability, privacy, adaptability, users' concerns). However, security and privacy issues of the centralized and decentralized schemes are not discussed in depth as we do in this paper.

In turn, as described in Section 4, various efforts have been performed in order to survey COVID-19 related applications, in particular, for digital contact tracing [16, 17, 18, 19, 20, 21]. Besides, in [30] fifteen applications/protocols are analyzed and possible attacks on each of them are highlighted. In [40] twenty applications are identified and described according to different characteristics (e.g. platform, rating on stores, country, developer, age range). In [41] forty applications are considered in order to perform an empirical assessment of contact tracing applications. In [35] twenty-eight contact tracing apps (available on Android platform) are analyzed from security and privacy perspectives (e.g. code's privileges, privacy policies). In [31] twenty-two European apps are reviewed focusing on characteristics such as type of platform, architecture, wireless technology and number of downloads.

None of those papers, however, covers as many applications as we do in this paper, namely 140. In addition, our work considers other characteristics, such as licensing and code repositories. We also provide an analysis based on these as well as other common characteristics (e.g. protocol, origin, country, architecture).

In summary, this paper presents and analyzes the two main paradigms that have been adopted by most applications deployed in different regions and countries worldwide. The security and personal data protection requirements that these technologies must satisfy are discussed in depth and the potential threats that could threaten the fulfilment of these requirements are precisely identified. Finally, we also provide a broad survey of digital contact tracing applications as well as an analysis based on the survey results.

6 Conclusion

During the year 2020, different initiatives have arisen concerning the design, implementation and deployment of contact tracing technologies for COVID-19. Most of those solutions follow one of the two predominant design paradigms: decentralized or centralized. In this article, the data flows and processes involved in the operation of these two alternatives have been precisely identified. We have also analyzed two specific implementation proposals, one for each approach: DP3T and ROBERT.

Although those solutions have been designed with focus on protecting the privacy of users, due to their operational characteristics all DCT proposals suffer from intrinsic security and privacy problems that we have put forward and discussed in this article.

It is also included in this work the result of conducting a survey of applications that were implemented and deployed in 2020 in most countries of the world as an additional element in the fight against the pandemic. We also included in this extended version of the original paper the corresponding analysis up to December 2021. On the other hand, as far as we know, the effectiveness of these applications is very uncertain, while the risks reported in the specialized literature are plausible and in many cases do not require any particular computer skills. It was not possible to find official public data on the results of the use of these applications, such as indicators of contribution to the process of contact tracing in terms of greater scope and/or shorter times.

One of the main objectives of this work is, in light of the analysis made of the privacy and security challenges posed by the DCT technologies, to motivate reflection and discussion by different actors of the society in relation to the deployment and use of this technology. The use of the data that is managed by the applications that implement digital contact tracing is a significant social issue that deserves to be presented and discussed with care and taking into account the regulations for the protection of personal data.

Acknowledgment

The authors wish to thank the comments and suggestions of the anonymous reviewers which collaborated to improve the presentation of this article. The reported work has been carried out in the context of the

PROTECT project (PRivacy Oriented TEchniques for the assessment of Contact Tracing solutions) financed by the COVID-19 fund of the Sectoral Commission for Scientific Research (CSIC), University of the Republic, Uruguay.

References

[1] MSP, "Estudio de efectividad de vacunación anti SARS-CoV-2 en Uruguay en 2021 - resultados preliminares," Ministerio de Salud Pública, Uruguay, Tech. Rep., 2021. [Online]. Available: https://www.gub.uy/ministerio-salud-publica/sites/ministerio-salud-publica/files/documentos/noticias/Informe%20efectividad%20vacunal%20al%2027_05_2021%20vf%20F.pdf

[2] Google Inc. , "Apple and Google partner on COVID-19 contact tracing technology," April 2020. [Online]. Available: https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/

[3] G. Betarte, J. D. Campo, A. Delgado, P. Ezzatti, L. González, A. Martín, R. Martínez, and B. Muracciole, "Proximity tracing applications for covid-19: data privacy and security," in *2021 XLVII Latin American Computing Conference (CLEI)*, 2021, pp. 1–10.

[4] European Parliament and of the council, "Regulation (eu) 2016/679," *Official Journal of the European Union*, 2016, Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[5] C. Troncoso *et al.*, "Decentralized Privacy-Preserving Proximity Tracing," Tech. Rep., 2020. [Online]. Available: https://github.com/DP-3T/documents

[6] PRIVATICS, AISEC, "ROBERT: ROBust and privacy-presERving proximity Tracing," PRIVATICS, Inria, France and Fraunhofer AISEC, Germany, Tech. Rep., 2020. [Online]. Available: https://github.com/ROBERT-proximity-tracing/documents/blob/master/previous_versions/ROBERT-specification-EN-v1_0.pdf

[7] The DP-3T Project, "Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems," Tech. Rep., 2020. [Online]. Available: https://github.com/DP-3T/documents/blob/master/Security%20analysis/Privacy%20and%20Security%20Attacks%20on%20Digital%20Proximity%20Tracing%20Systems.pdf

[8] Fraunhofer AISEC, "Pandemic Contact Tracing Apps: DP-3T, PEPP-PT NTK, and ROBERT from a privacy perspective," Tech. Rep., 2020. [Online]. Available: https://eprint.iacr.org/2020/489

[9] X. Bonnetain *et al.*, "Anonymous tracing, a dangerous oxymoron: A risk analysis for non-specialists," Tech. Rep., 2020. [Online]. Available: https://tracing-risks.com/

[10] Otto Seiskari, "BLE contact tracing sniffer PoC." [Online]. Available: https://github.com/oseiskar/corona-sniffer

[11] G. Celosia and M. Cunche, "Fingerprinting Bluetooth-Low-Energy Devices Based on the Generic Attribute Profile," in *IoT S&P 2019 - 2nd International ACM Workshop on Security and Privacy for the Internet-of-Things*. London, United Kingdom: ACM Press, Nov. 2019, pp. 24–31. [Online]. Available: https://hal.inria.fr/hal-02359914

[12] R. Raskar, "Private kit: Safe paths - can we slow the spread without giving up individual privacy?" March 2020. [Online]. Available: https://safepaths.mit.edu/

[13] Covid Watch, March 2020. [Online]. Available: https://covid-watch.org

[14] J. Bay *et al.*, "BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders," Government Technology Agency - Singapore, Tech. Rep., 2020. [Online]. Available: https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf

[15] The DP-3T Project, "Security and privacy analysis of the document "ROBERT: ROBust and privacy-presERving proximity Tracing," Tech. Rep., 2020. [Online]. Available: https://github.com/DP-3T/documents/blob/master/Security%20analysis/ROBERT%20-%20Security%20and%20privacy%20analysis.pdf

[16] Council of Europe, "Council of Europe - Data protection." [Online]. Available: https://www.coe.int/en/web/data-protection/contact-tracing-apps

[17] MIT Technology Review, "COVID tracing tracker." [Online]. Available: https://www.technologyreview.com/2020/12/16/1014878/covid-tracing-tracker/

[18] Linux Foundation (LF) Public Health, "Map COVID apps." [Online]. Available: https://landscape.lfph.io/card-mode?grouping=category

[19] Wikipedia, "COVID-19 apps." [Online]. Available: https://en.wikipedia.org/wiki/COVID-19_apps#List_of_countries_with_official_contact_tracing_apps

[20] ——, "Exposure Notification." [Online]. Available: https://en.wikipedia.org/wiki/Exposure_Notification#Adoption

[21] XDA Developers, "Google-Apple API list countries." [Online]. Available: https://www.xda-developers.com/google-apple-covid-19-contact-tracing-exposure-notifications-api-app-list-countries/

[22] D. Lee and J. Lee, "Testing on the move: South Korea's rapid response to the COVID-19 pandemic," *Transportation Research Interdisciplinary Perspectives*, vol. 5, no. Mayo 2020, 2020.

[23] G.Jung, H. Lee, A. Kim, and U. Lee, "Too Much Information: Assessing Privacy Risks of Contact Trace Data Disclosure on People With COVID-19 in South Korea," *Frontiers in Public Health*, vol. 8, no. Junio 2020, 2020.

[24] Amnistía Internacional, "Las aplicaciones de rastreo de contactos de Bahréin, Kuwait y Noruega, entre las más peligrosas para la privacidad." [Online]. Available: https://www.amnesty.org/es/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy/

[25] ——, "Noruega: Suspender la aplicación de rastreo de contactos para la COVID-19, un gran logro para el derecho a la privacidad." [Online]. Available: https://www.amnesty.org/es/latest/news/2020/06/norway-covid19-contact-tracing-app-privacy-win/

[26] R. H. et al, "Effective Configurations of a Digital Contact Tracing App: A report to NHSX," University of Oxford, Tech. Rep., 2020. [Online]. Available: https://045.medsci.ox.ac.uk/files/files/report-effective-app-configurations.pdf

[27] L. Ferretti *et al.*, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *American Association for the Advancement of Science*, vol. 368, no. 6491, 2020.

[28] G. Grekousis and Y. Liu, "Digital contact tracing, community uptake, and proximity awareness technology to fight covid-19: a systematic review." *Sustainable cities and society*, vol. 71, 2021.

[29] M. Shahroz *et al.*, "Covid-19 digital contact tracing applications and techniques: A review post initial deployments," *Transportation Engineering*, vol. 5, p. 100072, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666691X21000282

[30] N. Ahmed *et al.*, "A survey of covid-19 contact tracing apps," *IEEE Access*, vol. 8, pp. 134 577–134 601, 2020.

[31] T. Martin *et al.*, "Demystifying covid-19 digital contact tracing: A survey on frameworks and mobile apps," *Wireless Communications and Mobile Computing*, 2020.

[32] B. Sowmiya *et al.*, "A survey on security and privacy issues in contact tracing application of covid-19," *SN Computer Science*, vol. 2, p. 100072, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666691X21000282

[33] A. Akinbi *et al.*, "Contact tracing apps for the COVID-19 pandemic: a systematic literature review of challenges and future directions for neo-liberal societies," *Health Information Science and Systems*, vol. 9, no. 1, p. 18, 2021.

[34] A. S. Krishnan, Y. Yang, and P. Schaumont, "Risk and architecture factors in digital exposure notification," in *Embedded Computer Systems: Architectures, Modeling, and Simulation*, A. Orailoglu, M. Jung, and M. Reichenbach, Eds. Cham: Springer International Publishing, 2020, pp. 308–319.

[35] M. Hatamian, S. Wairimu, N. Momen, and L. Fritsch, "A privacy and security analysis of early-deployed covid-19 contact tracing android apps," *Empirical Software Engineering*, vol. 26, no. 3, p. 36, Mar 2021.

[36] L. Baumgärtner, A. Dmitrienko, B. Freisleben, A. Gruler, J. Höchst, J. Kühlberg, M. Mezini, R. Mitev, M. Miettinen, A. Muhamedagic, T. D. Nguyen, A. Penning, D. Pustelnik, F. Roos, A.-R. Sadeghi, M. Schwarz, and C. Uhl, "Mind the gap: Security amp; privacy risks of contact tracing apps," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 458–467.

[37] H. Wen, Q. Zhao, Z. Lin, D. Xuan, and N. Shroff, "A study of the privacy of covid-19 contact tracing apps," in *Security and Privacy in Communication Networks*, N. Park, K. Sun, S. Foresti, K. Butler, and N. Saxena, Eds. Cham: Springer International Publishing, 2020, pp. 297–317.

[38] L. Reichert, S. Brack, and B. Scheuermann, "A survey of automatic contact tracing approaches using bluetooth low energy," *ACM Trans. Comput. Healthcare*, vol. 2, no. 2, mar 2021.

[39] F. Cicala, W. Wang, T. Wang, N. Li, E. Bertino, F. Liang, and Y. Yang, "Pure: A framework for analyzing proximity-based contact tracing protocols," *ACM Comput. Surv.*, vol. 55, no. 1, nov 2021.

[40] N. Min-Allah, B. A. Alahmed, E. M. Albreek, L. S. Alghamdi, D. A. Alawad, A. S. Alharbi, N. Al-Akkas, D. Musleh, and S. Alrashed, "A survey of covid-19 contact-tracing apps," *Computers in Biology and Medicine*, vol. 137, p. 104787, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521005813

[41] R. Sun, W. Wang, M. Xue, G. Tyson, S. Camtepe, and D. C. Ranasinghe, "An empirical assessment of global covid-19 contact tracing applications," in *Proceedings of the 43rd International Conference on Software Engineering: Companion Proceedings*, ser. ICSE '21. IEEE Press, 2021, p. 173–174.