PACT: Privacy Aware Contact-Tracing

Big Data Security and Privacy in the COVID-19 Era

Mohamed TALHA
TIM Laboratory, ENSA UCA
Marrakech, Morocco
mohamed.talha@icloud.com

Jihane EL MOKHTARI
LISER Laboratory, IPI
Paris, France
el_mokhtari_jihane@yahoo.fr

Anas ABOU EL KALAM
TIM Laboratory, ENSA UCA
Marrakech, Morocco
a.abouelkalam@uca.ac.ma

***Abstract*—Tracing contacts with people identified as carrying the coronavirus is essential to control the spread of the COVID-19 pandemic. In many countries, tracing contacts is a recursive procedure based on the memory of people tested positive. In other countries, thanks to Big Data technologies, Contact-Tracing Apps have been developed. Although these applications often meet the requirements of applicable Privacy regulations, the number of people who agree to use them remains too low. This has been explained by people's lack of confidence in these apps regarding their Privacy. This article covers many aspects related to the Privacy and the techniques used to develop this type of applications. We also present a new approach to trace contacts without using personally identifiable information. The objective of this work is to regain the trust of people, especially in their Privacy, in order to be able to have a large number of users and thus contribute to effectively combating the pandemic.**

Keywords—Big Data; Data Security; Privacy; Contact-Tracing App; Coronavirus; COVID-19

I. INTRODUCTION

To contain the spread of COVID-19, numerous applications to search for the contacts of individuals have been developed in many countries. These applications, designated by Contact-Tracing Apps, are generally intended for mobile devices such as Smartphones, Smartwatches, Fitness-Trackers, etc. A Contact-Tracing App is a mobile platform that identifies people who may have been in contact with an infected person. As soon as such a contact is identified, other additional information may be collected to conduct studies and measures to limit the spread of the coronavirus.

The functioning of these applications generally depends on the technology used (Bluetooth, GPS, Wi-Fi or GSM) and the storage architecture (centralized or decentralized). Although the combination of Bluetooth and GPS improves the precision of the location of individuals, the overall quality of these applications remains insufficient. First of all, to calculate the risk of contamination, Contact-Tracing Apps are based on the distance between individuals, which should be less than two meters for example, and the duration of the contact, which should last more than fifteen minutes for example. These applications therefore do not take into consideration if the people in contact were physically separated (for example by a Plexiglas or a wall, if the people are in cars locked in a traffic jam, etc.). We must therefore take into account the contact environment to improve the estimation of the contamination risk, which is, to the best of our knowledge, not done today. Moreover, Bluetooth is used to estimate the distance between two smartphones by measuring the signal strength. Unfortunately, this technology suffers from several security vulnerabilities that allow hackers to quickly access smartphones. Keeping the Bluetooth connection open all day is not recommended for data security. Even temporarily, it has been proven that hackers can launch man-in-the-middle attacks to recover cryptographic keys securing the pairing of mobile devices. It will therefore be necessary to reinforce these applications with "correctors" in order to protect the mobiles against this type of attack. Finally, and this is what interests us for this work, the Contact-Tracing Apps launched today in many countries have not been very successful; the number of users is still too low. This is due to the lack of trust in these applications and the fear of endangering the privacy of users. Indeed, several breaches of privacy can be committed through these applications such as the collection of personal information which may lead to the identification of users, the use of the collected data in a context other than that of the health crisis or the disclosure of data to third parties for purposes different from those of combating the virus (insurance and credit organizations, market research and advertising companies, etc.).

To be able to carry out analyzes and take measures to contain the spread of the pandemic, governments must identify the contacts, their movements and even, for certain countries, the medical information of the users of the Contact-Tracing Apps. These applications then generate large volumes of personal data, sometimes sensitive, intended to be stored and exploited by Big Data technologies, which highlights the risk of violation of Privacy by such technologies.

In order to comply with applicable privacy regulations, governments promise to anonymize personal information to protect the identity of users and do not compel citizens to use Contact-Tracing Apps; only volunteers can use them. This answers the question of user consent. However, statistics show a low rate of use of these applications [1] due to the sensitivity of people's personal information. In this article, we propose a new approach which consists in reversing the control of information; it is the user who controls his confidential data. Also, no personal information will be recorded anywhere. A unique identifier, hashed and freely chosen by the user himself will suffice to operate a Contact-Tracing App.

This article is organized as follows. Section 2 is reserved for the presentation of the concept of Privacy as well as the main related laws and regulations. Section 3 briefly presents geolocation techniques including GPS, GSM, Wi-Fi and Bluetooth. We then present in Section 4 some examples of Contact-Tracing Apps in order to understand closely their operating modes and their degree of respect for Privacy. Section 5 is devoted to the presentation of our approach based on the inversion of the control of personal information. We then present in Section 6 a prototype to prove our concept. Finally, we conclude this work in section 7.

II. PRIVACY

Privacy is a complex concept and difficult to define due to the cultural subjectivity of individuals linked to this subject. It can be seen from a security aspect or confused with security. While Security deals with Confidentiality, Integrity and Availability, Privacy deals with the proper use of an individual’s information [2]. Some consider Privacy to be a derivative concept based on more basic rights such as liberty or property [3]. In many countries, there are laws governing the protection of privacy, and this since the reservation of Article 12 of the 1948 Universal Declaration of Human Rights to the right to privacy. This right is also guaranteed by Article 8 of the European Convention on Human Rights and Article 17 of the International Covenant on Civil and Political Rights.

The emergence of information technology has increased the risk of privacy breaches. In this sense, several laws and regulations have emerged to guarantee privacy and define the main areas of protection, including Regulation No. 2016/679 known as the GDPR[1] (General Data Protection Regulation) in Europe or even PIPEDA [2] (Personal Information Protection and Electronic Documents Act) in Canada. According to PIPEDA, personal information means any factual or subjective information, recorded or not, concerning an identifiable person (name, age, income, ethnic origin, political opinion, medical record, etc.). Information Commissioner's Office [3] (ICO) considers that a data is qualified as personal when it relates to a living individual and that it can be used, alone or in combination with other data, for its identification.

A set of ten principles has been defined and fully or partially taken up by several works aimed at proposing solutions for the design of privacy-friendly systems, among which we find [4], [5], [6], [7], [8] and many others. The ten privacy principles reflecting government definitions are [5]:

1. The purpose for which personal information is collected and stored must be communicated and respected,

2. The consent of the owner of the data must be explicitly given for the purposes which have been announced to him/her,

3. Limited collection is the principle which consists in collecting the minimum of personal information necessary to achieve the announced objective,

4. The limited use of the collected data in a context compatible with the purpose of the collection,

5. Limited disclosure, which consists of non-communication of collected and saved data to other parties for purposes different from those for which the donor has given his/her agreement,

6. The limited conservation of the collected data which must not go beyond the duration necessary for the achievement of the purpose at the origin of the collection,

7. The accuracy of the personal collected data must be ensured by deleting or correcting any inaccurate information,

8. The security of the collected personal information which consists in protecting it against theft or any other misappropriation,

9. The transparency which must guarantee a donor to have access to all of his/her saved data when he/she wishes,

10. Compliance with the above principles must be ensured and the owner of the data must be able to verify it.

The health crisis that the world has been going through for the past few months has prompted several countries to use the Big Data technologies to limit the spread of the COVID-19 pandemic. This involves the deployment of Contact-Tracing Apps to be installed on mobile devices (Smartphones, Smartwatches, etc.) to monitor the spread of the virus and identify possible contaminations based on frequentation between people. This type of applications can infringe on the right to protection of privacy because it is based on the collection of personal information such as the places visited, the identity of people around, phone number, state of health, medical information, etc. The excessive amount of this data implies the use of Big Data platforms to be able to store and analyze them efficiently, and their confidential nature brings us back to the question of how far Big Data techniques are capable of respecting the Privacy.

[1] https://gdpr-info.eu/
[2] https://www.priv.gc.ca/en/
[3] https://ico.org.uk/

III. GEOLOCATION TECHNOLOGIES

Geolocation is a set of techniques that allows positioning an object or an individual on a map using its geographic coordinates. These geographic positions can be transmitted in real time or saved for later use. Geolocation involves at least two entities: a positioning device and a monitoring system. The positions initially determined are transmitted to the tracking/tracing system in different ways through for example geocoder, satellite and GSM. We will present in this section the different geolocation technologies that can be used individually or in combination.

A. GPS

Thanks to signals from a network of satellites, the GPS generates data that make it possible to determine the geographic position of a terminal equipped with a dedicated chip. A position in the Global Positioning System (GPS) is represented by the latitude and longitude coordinates and its timestamp [9]. These coordinates are preprocessed by Clustering algorithms and then described in a known geometric form. The geolocation of a mobile device, for example, is done using a chip integrated into the mobile which is able to calculate the position and communicate it in real time either by SMS or via the GPRS network. This location method provides an accuracy ranging from 10 to 100 meters depending on the conditions of reception of the GPS signal [10]. Despite the fact that GPS location collection is one of the most used methods, the accuracy of geolocation can be negatively affected by the decrease or the exhaustion of the receiver power, which is the drawback of this technology. In the case of a mobile phone, the use of the GPS chip can be disabled to optimize battery consumption. Another limit of geolocation by GPS is the indoor operation (building, basement, tunnel ...) where the GPS waves do not reach the receiver, which leads to a loss of precision that can reach several tens of meters.

B. GSM

In the Global System for Mobile communications (GSM), the geographic location is determined via telephone operators' relay antennas, which are generally used to transfer data to mobile phones. Connecting a mobile device to an antenna estimates its position. This estimation can be done by multitude methods, all based on the signals sent by the mobile tag to the antennas. Among these methods, we quote:

- Angles-of-Arrival (**AoA**) measurements are based on the exploitation of the angles of signals emitted by the mobiles and intercepted by at least two surrounding antennas. The position is then estimated by the intersection of the lines passing through each reference [11]. This method particularly suffers from influences of multiple paths as well as signal dispersion phenomena.

- Time-of-Arrival (**ToA**) measurements are based on the propagation time of a signal between the instant of its emission by the mobile tag and that of its reception by the antenna. Its variant, Time-Difference-of-Arrival (**TDoA**), consists in measuring the instant of arrival of a signal at several points in space and compare the difference between the instants measured at each receiver [12]. The major drawback of time measurements is the notion of synchronization between the different nodes which must be extremely precise.

- Received Signal Strength (**RSS**) measurements are based on the power of received signals and consist in recovering spatial information based on the attenuation of the signal caused by the decrease in energy of a wave interacting with the environment during its movement from a given distance. RSS has the advantage of being simple to implement but one of the disadvantages is that the accuracy is considerably degraded by the multi-path channel [11].

- Hybrid measurements are based on the combination of two or more measurements mentioned above to guarantee better accuracy and a reduction in errors.

In summary, geolocation technology by GSM depends on cellular stations which, depending on their coverage areas, can offer data with an accuracy ranging from 300 meters in urban areas to several kilometers in rural areas [10]. The accuracy of the calculations therefore depends on the number of antennas and their absence implies the impossibility of location. Although this technology remains inexpensive in terms of consumption and quick to start, its use remains complicated given the difficulty of collecting the necessary data from operators who cannot disclose the geolocation information of their customers.

C. Wi-Fi

The Wi-Fi location is established using the identifier of the Wi-Fi terminal to which the mobile is connected. Large databases are deployed to store terminal identifiers and their positions. These Wi-Fi hotspots generally offer large ranges (up to several hundred meters depending on the model). A device detects nearby Wi-Fi connections through the built-in Wi-Fi card, which is capable of providing a measure of signal strength when receiving a packet [13]. The determination of the position of the device is therefore based on the power data of the access points. Fingerprinting is the most popular method of location due to its high accuracy compared to other methods [14]. It consists in comparing the values of the Received Signal Strength (RSS) with an RSS reference map (radiomap) which associates the values with the positions [15]. Wi-Fi geolocation is widely used today for its many advantages, including its simple implementation at low cost, the good precision it provides and its positioning capacity indoors. Connection to the Wi-Fi network is not necessary for geolocation, only network detection is sufficient. The big limitation of this technology is that it is almost unusable in rural areas.

D. Bluetooth

The birth of BLE (Bluetooth Low Energy) in 2010 has put the Bluetooth technology at the center of geolocation applications [16]. The Bluetooth allows location by relying on wireless beacons in the form of small physical boxes to be installed in signs. The range of emission of these beacons generally borders a few tens of meters and their operation does not require the Internet connection. They have the capacity to cover a field of up to 450 meters and transmit data packets, via an advertising channel, at regular intervals. These packets are retrieved by the surrounding mobile devices and each packet carries the information from the transmitting beacon making it possible to determine the distance separating it from the receiving device.

While previous versions of Bluetooth allowed a rough estimation of location using signal strength measurements, the new 5.1 version brings a multitude of advantages. In addition to reduced power consumption and improved connection speed, this version offers better advertising, thus overcoming the density problems caused by the interference of two Bluetooth devices on the same channel. It also increases accuracy by indicating the location of devices in addition to **the distance between them**. Bluetooth technology does not stop at the connection of mobile devices to the beacons; it also makes it possible to interconnect the various objects equipped with Bluetooth using radio waves on the 2.4GHz frequency band. Version 5.1 offers a greater transmission range exceeding 200 meters. Another interesting contribution of Bluetooth is that of the succession that it can take to connect a mobile device in case of absence or failure of its GPS chip to a GPS antenna. However, the weakness of this technology is the need to permanently activate the Bluetooth function on a mobile device to ensure connections, with all the security risks that this can cause.

IV. EXISTING CONTACT-TRACING APPS

Due to the ability of COVID-19 to spread quickly between people, smartphone apps have emerged as a way to facilitate the contact tracing process. How these applications work depends on the location technology used and the data storage techniques. We distinguish between two types of applications:

- **Bluetooth**: Bluetooth-based apps consist of smartphones exchanging encrypted codes via Bluetooth whenever people meet at close range and for a certain period of time. These applications, adopted in several countries, in addition to the problems related to data security due to the prolonged activation of Bluetooth, pose certain problems of protection of the privacy of individuals.

- **GPS**: this second type of application is based on GPS and location data to track the movements of users and those with whom they were in contact. These applications, although proven effective in some countries, pose many privacy problems due to the use of location data. In addition, the lack of precision in the location and the loss of the GPS signal often impose its combination with Bluetooth.

A. Bluetooth based Apps

TraceTogether, one of the first Android and iOS applications to fight against the COVID-19 pandemic, was released on March 20, 2020 by the Singaporean Ministry of Health [17]. The operating principle of this application is based on the exchange, via a Bluetooth connection, of temporary tokens between closely spaced smartphones, within a radius of two meters and for at least fifteen minutes. For each exchange, a copy of the tokens and some information about the individuals, including their identity and their phone numbers, are sent to a central server. Fig.1. illustrates how TraceTogether App works.

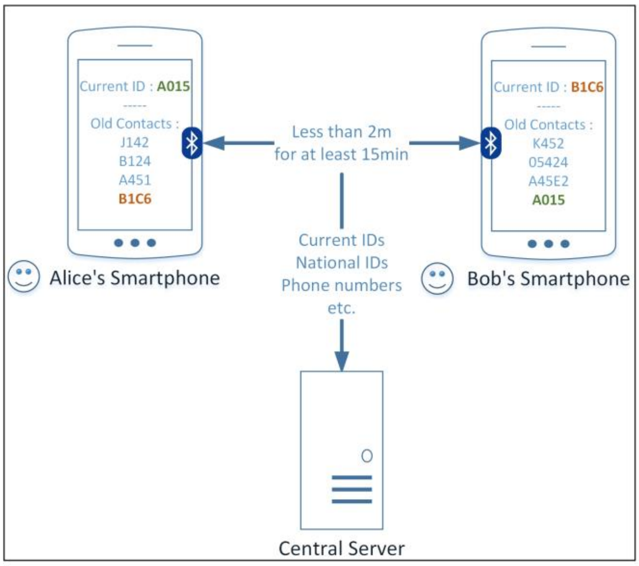


Fig.1. The COVID-19 Tracing Approach via Bluetooth

When a user contracts the virus and is confirmed to be positive, the Ministry of Health asks him to disclose his data on the application to retrieve the list of all the tokens registered in his smartphone. Then, thanks to the information of the users linked to their tokens and saved on the central server, it will be possible to find the individuals who have been in contact with the person carrying the virus. In order to protect personal data, the token is a character string generated randomly and represents a user for a short time (the time of the contact). This allows the application to keep users private from each other (unless a user links the time of receipt of a token to another user). Indeed, as a token is only valid temporarily, it is impossible to know all the tokens assigned to a person or the people assigned to the tokens received by an application. However, forcing an infected person to disclose his data, as provided by the Singaporean law [18], Privacy is no longer protected! Moreover, the collection of certain information allowing the identification of individuals worries users. This explains the low rate of use of this application in Singapore (16.7% according to [1]).

Many other applications such as **StopCovid** in France and **Wiqaytna** in Morocco were inspired by the same Singaporean model while trying to make certain modifications in order to best meet privacy protection requirements. The operating mode of these applications follows that of TraceTogether: when two smartphones remain close together for a certain time, each stores the anonymous reference of the other via a Bluetooth connection. If a user is positive for COVID-19, those who have been in contact with him will be notified by the application that they have been in contact with someone who has just tested positive without knowing who, when and where. These applications are therefore based on anonymity to preserve the identity of the users and, to comply with the applicable regulations, are intended for volunteers. In addition to the technical problems linked in particular to the lack of precision of the sensors and to the security leak due to the Bluetooth which must remain activated throughout the day, the diversion of information towards other uses could worry users and will therefore be a real challenge to succeed in convincing citizens to use these applications. Other Apps such as **NHS COVID-V19** (United Kingdom), **COVIDSafe** (Australia), and **Blue Zone** (Vietnam) can be classified in this category.

B. GPS based Apps

Healthy Together [19], launched by the State of Utah (United States), offers a variety of features to help users assess their symptoms and, with the help of local health authorities, to determine if a user should be tested. Thanks to the location data recorded by the application, if a user contracts the coronavirus, health workers can see where, when and for how long he has been in contact with other users, which could potentially identify areas spread of the virus. To have better data accuracy, this application combines both GPS location data and Bluetooth. A lot of personal information is therefore saved by this application, which risks compromising Privacy. Indeed, the fact that a user is monitored for all their trips may not be accepted by the majority of citizens. In addition, the app records patient medical data which can be particularly sensitive for many people. To reassure citizens, the government of Utah says it has made this application completely voluntary and no one is forced to share their phone number or medical information. In addition, users can delete their data at any time. Location data is also automatically deleted within 30 days and symptom data is automatically de-identified within 30 days. Despite all these measures to reassure Utahans, the application has still not been very successful (only 0.7% of Utahans use it according to [1]).

Care19 [20], launched by the State of North Dakota and also adopted by the State of South Dakota (United States), is based on the location of individuals by GPS. Each day, the application assigns a new random identifier to the user and then saves his location on his smartphone anonymously throughout the day. The special feature of this application is that only the location data of the places where a user has stayed for more than 10 minutes is recorded. Numerous testimonies from users affirm a lack of precision on the traceability of their locations [21]; the improvement of the location system is underway, in particular through the combination with the Bluetooth. In order to protect the privacy of users, these states have promised to delete all data after the pandemic, have allowed users to delete all their data whenever they want and have adopted the principles of anonymity and volunteering to use this application. Despite all these measures, the percentage of use of the application does not exceed 1.3% according to [1].

Corona Data Donation [22] is an application for Smartwatches and Fitness-Trackers launched on April 07, 2020 by the Germany's Public Health Authority. The application collects several vital signs from volunteers such as temperature, blood pressure, heart rate, sleep, etc. as well as socio-demographic data that the user can enter such as age, sex, height and weight. All these information are continuously transmitted to the RKI (Robert Koch Institute) for analysis in order to identify people who may have contracted the virus. The results are then displayed on an interactive online map allowing health authorities and the general public to assess the spread of infections. Many personal and confidential data are therefore voluntarily made available to the authorities. Given the consent of users, according to GDPR, privacy is not compromised. However, the application has still not been very successful in terms of percentage of use (less than 3% according to [1]).

V. INVERSION OF CONTROL BASED APPROACH

Table 1, adapted from [1], shows the percentage of use of various Contact-Tracing Apps.

TABLE I. PERCENTAGE OF USE OF CONTACT-TRACING APPS

Application Name	**Country**	**Users: percentage of population (%)**
TraceTogether	Singapore	16.7
COVIDSafe	Austria	10
Stopp Corona	Austria	4.5
Corona Data Donation	Germany	<3
Care19	US	1.3
Healthy Together	US	0.7
Blue Zone	Vietnam	0.1

Even for TraceTogether, which has been the most successful to date, the percentage of use of Contact-Tracing Apps remains too low. This can only be explained by the lack of confidence in these applications. Indeed, despite the promises from governments on the anonymization of personal information and the subsequent deletion of data, many factors can worry users such as:

- diversion of information to other uses,
- monitoring of movements, trips and meetings,
- sharing of medical information,
- risk of accidental access to confidential data,
- risk of malicious access to confidential data,
- etc.

This leads us to think of a model allowing the users themselves to control their information by disclosing, almost, no personal information. Our approach is illustrated in the scenario of Fig. 2.

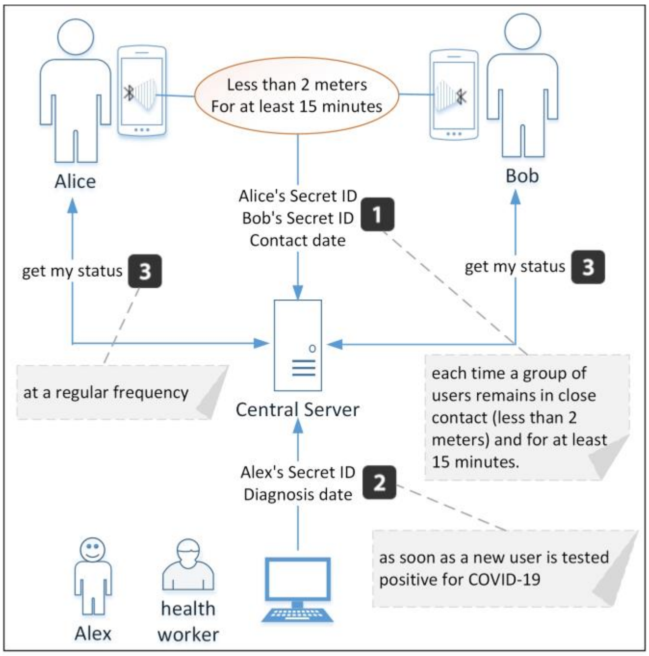


Fig.2. Privacy Aware Contact-Tracing Approach

When using the application for the first time, a user creates an account by providing only one piece of information, his identifier. It is free to choose this identifier provided that it is unique (a verification is performed on the server side). This identifier is hashed using MD5 or SHA, which is an irreversible operation, and then transmitted to the central server to create an account for the user. A user account will be linked to other information such as the number of users tested positive for the virus with whom the user was in contact. Our approach includes 3 phases:

- **Phase 1 - Contacts collection**: a contact becomes potentially contaminating when certain conditions meet, such as a close distance (for example, less than two meters), a long duration (for example, more than fifteen minutes), etc. When a contact between users becomes potentially contaminating, their identifiers and the date of contact are transmitted to the central server. In order to optimize network traffic and the volume of data transmitted to the server, the data will only be sent to the server if there is a change in the group of people in contact (no need to send the same information every fifteen minutes for example). Furthermore, at the server level, since it will receive the same information from all the users who were in contact, a single record in the database is enough to store the information by updating the column of users who have transmitted it to the server.

- **Phase 2 - Infections reporting**: when a user tests positive for the virus, with his consent, he reveals his identifier to a healthcare worker to report a new case of infection to the central server. If the user refuses to reveal his identifier, it will be possible to do otherwise. The healthcare worker requests the generation of a temporary token for the user so that he himself can report his case to the central server. Whatever the mode of declaration of the infection, this operation launches on the server a process of research/update of the accounts of the users with who the patient has been in contact during the last 21 days, the maximum time for symptoms to appear.

- **Phase 3 - Spread monitoring**: a user can consult his status at any time to get the number of direct and indirect contacts he has had with infected people. We do not consider it necessary to notify users whenever a new contact with an infected person is discovered, this will only diminish the credibility of the application. Instead, we prefer that the application requests the server, on a regular basis (for example twice a day), in order to retrieve the status of the user. In this way, control is given to the user himself: the higher the number of direct and indirect contacts with positive people, the higher the risk of contamination. The necessity and priority of screening tests will then be managed on the basis of this information.

VI. PROTOTYPE

Our approach can be implemented in different ways. We have developed a prototype, in the form of a REST API, in Java (version 1.8), Maven (version 3.6.3) and SpringBoot (version 2.3.0) on Hadoop platform (version 2.7.7) and HBase database (version 1.1.0). The project was developed on the eclipse IDE (version 4.15.0) and Ubuntu (version 18.04.2 LTS). To better understand our prototype, let's take the example of the scenario in Fig. 3. In this scenario, we consider 5 groups of contacts:

- **Group A**: a potentially contaminating contact took place on May 01, 2020 between 3 users (User-1, User-2 and User-3). 25 days later, on May 26, 2020, User-1 was identified positive for covid-19.

- **Group B**: a potentially contaminating contact took place on May 15, 2020 between 4 users (User-3, User-4, User-5 and User-6). 3 days later, on May 18, 2020, User-6 tested positive for covid-19.

- **Group C**: a potentially contaminating contact took place on May 11, 2020 between 3 users (User-6, User-7 and User-8). 7 days later, on May 18, 2020, User-6 tested positive for covid-19.

- **Group D**: a potentially contaminating contact took place on May 16, 2020 between 2 users (User-8 and User-9). For this group, no one is positive for covid-19.

- **Group E**: a potentially contaminating contact took place on May 14, 2020 between 2 users (User-9 and User-10). 12 days later, on May 26, 2020, User-10 was identified positive for covid-19.

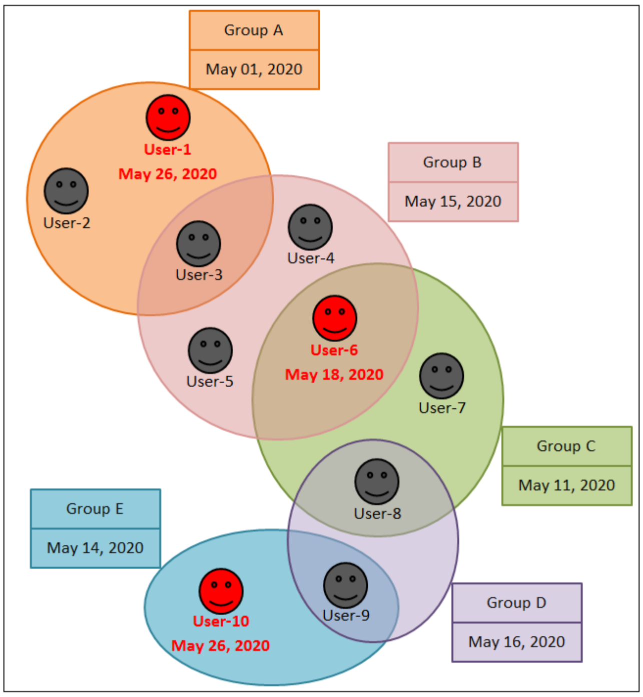


Fig.3. Example of covid-19 propagation

Note that:

- User-3 is part of two groups A and B.

- User-6 is part of two groups B and C.

- User-8 is part of two groups C and D.

- User-9 is part of two groups D and E.

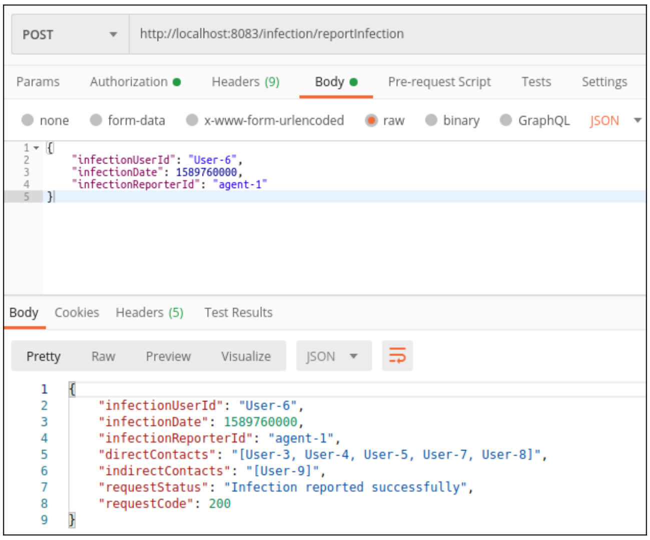
Analysis of the previous scenario leads us to the following results:

- **On May 26, 2020**, User-1 of group A is identified positive. The report of this infection must not raise any alert because the contact took place more than 21 days ago.

- **On May 18, 2020**, User-6 of groups B and C is identified positive. The report of this infection must raise two alerts: **Direct Contacts** (User-3, User-4, User-5, User-7, and User-8) and **Indirect Contacts** (User-9 through User-8 as they were in contact 2 days ago). Note that User-1 and User-2 will not be alerted as indirect contacts because they were in contact with User-3 before the latter was in contact with User-6.

- **On May 26, 2020**, User-10 of group E is identified positive. The report of this infection must raise two alerts: **Direct contacts** (User-9) and **Indirect contacts** (User-8 through User-9 as they were in contact 10 days ago).

Fig. 4. illustrates the declaration of User-6 as being positive for covid-19 on May 18, 2020 (the timestamp of May 18, 2020 at midnight is 1589760000).



http://localhost:8083/infection/reportInfection

```
{
    "infectionUserId": "User-6",
    "infectionDate": 1589760000,
    "infectionReporterId": "agent-1"
}
```

```
{
    "infectionUserId": "User-6",
    "infectionDate": 1589760000,
    "infectionReporterId": "agent-1",
    "directContacts": "[User-3, User-4, User-5, User-7, User-8]",
    "indirectContacts": "[User-9]",
    "requestStatus": "Infection reported successfully",
    "requestCode": 200
}
```

Fig.4. Web service to report user infection

Calling this web service launches the following operations:

1. Update the User-6 account to report the screening date.

2. Update the accounts of User-3, User-4, User-5, User-7 and User-8 to indicate that they were in direct contact, during the past 21 days, with a user who tested positive. For each of these users, we simply increment their number of direct contacts.

3. Update the User-9 user account to indicate that he was in indirect contact with a user who tested positive by incrementing his number of indirect contacts.

4. At a regular frequency, the instances of the application installed by all these users will retrieve and display the new updated values.

VII. Conclusion

The COVID-19 pandemic has created a health crisis around the world. Retrieving the contact history of anyone identified as a carrier of the coronavirus is essential to control its spread. This has motivated many governments to use technologies such as Big Data, Bluetooth and GPS to develop Contact-Tracing Apps. Although these applications often meet the requirements of applicable privacy regulations, the number of people who agree to use them remains too low. This was explained by the lack of confidence of users in these applications, particularly with regard to their privacy. Our proposition is based on a new approach allowing finding the history of potentially contaminating contacts without referring to any personal information. The principle of this approach consists in reversing the control of personal information; it is the users themselves who can follow, permanently and at a regular frequency, the number of direct and indirect contacts they have had with the carriers of the virus. This will then manage the necessity and priority for screening tests. However, and in order to obtain encouraging results, it is essential to deal with other problems briefly presented in this article such as the contact environment in order to better estimate the risk of contamination as well as the security risks that could be generated by the use of Bluetooth.

References

[1] J. Li and X. Guo, "COVID-19 Contact-tracing Apps: A Survey on the Global Deployment and Challenges", arXiv:2005.03599, May 2020.

[2] S. Sangeetha and G. Sudha Sadasivam, "Privacy of Big Data: A Review," A. Dehghantanha, K.-K. R. Choo (eds.), Handbook of Big Data and IoT Security, 2019, doi: 10.1007/978-3-030-10543-3_2.

[3] A. Moore, "Defining Privacy," Journal of social philosophy, Vol. 39 No. 3, 2008, 411–428.

[4] K. Barker, M. Askari, M. Banerjee, K. Ghazinour, B. Mackas, M. Majedi, S. Pun, A. Williams, "A Data Privacy Taxonomy," 26 British National Conference on Databases, BNCODth, 2009, 5588, 42-54, doi: 10.1007/978-3-642-02843-4_7.

[5] R. Agrawal, J Kiernan, R. Srikant and Y. Xu, "Hippocratic Databases," Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002.

[6] S. Vimercati, S. Foresti, G. Livraga, and P. Samarati, "Data privacy: Definitions and techniques," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2012, 20, 793-817, doi: 10.1142/S0218488512400247.

[7] J. Vincent, T. Dubin and C. Porquet, "Protection de la vie privée basée sur des ontologies dans un système Android,", APVP 2012 - Atelier Protection de la Vie Privée, 3ème édition, Jun 2012, Ile de Groix, France.

[8] A. Oglaza, "Système d’aide à la décision pour la protection des données de vie privée," Toulouse 1 Capitole University, 2014.

[9] G. Xu, S. Gao, M. Daneshmand, C. Wang and Y. Liu, "A Survey for Mobility Big Data Analytics for Geolocation Prediction," in IEEE Wireless Communications, vol. 24, no. 1, pp. 111-119, February 2017, doi: 10.1109/MWC.2016.1500131WC.

[10] M.A. Abchir, "Vers une sémantique floue : application à la géolocalisation," Intelligence artificielle, Université Paris VIII Vincennes-Saint Denis, Nov 2013.

[11] K.I. Kossonou, "Étude d’un système de localisation 3-D haute précision basé sur les techniques de transmission Ultra Large Bande à basse consommation d’énergie pour les objets mobiles communicants", Electronique, Université de Valenciennes et du Hainaut-Cambresis; Université Félix Houphouët-Boigny, Jul 2014.

[12] UIT-R, "Comparaison de la méthode de géolocalisation de signal fondée sur la différence entre les instants d'arrivée par rapport à celle fondée sur l'angle d'arrivée", Rapport UIT-R SM.2211-2, Jun 2018.

[13] D. Larrey and L. Rodier, "Géolocalisation par WiFi", inria-00112186, Rapport de recherche 2006, pp.83.

[14] S. Subedi and JY. Pyun, "Practical Fingerprinting Localization for Indoor Positioning System by Using Beacons", Journal of Sensors, 2017, 1-16, doi: 10.1155/2017/9742170.

[15] A. Pérez-Navarro, J. Torres-Sospedra, R. Montoliu, J. Conesa, R. Berkvens, G. Caso, C. Costa, N. Dorigatti, N. Hernández, S. Knauth, E.S. Lohan, J. Machaj, A. Moreira and P. Wilk, "Challenges of Fingerprinting in Indoor Positioning and Navigation", Geographical and Fingerprinting Data for Positioning and Navigation Systems, 2019, doi: 10.1016/B978-0-12-813189-3.00001-0.

[16] D. Zaim and M. Bellafkih, "Bluetooth Low Energy (BLE) based geomarketing system," 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA), Mohammedia, 2016, pp. 1-6, doi: 10.1109/SITA.2016.7772263.

[17] TraceTogether. Available online: https://www.tracetogether.gov.sg/ (accessed on May 22, 2020).

[18] Can I say no to uploading my TraceTogether data when contacted by the Ministry of Health? Available online: https://tracetogether.zendesk.com/hc/en-sg/articles/360044860414-Can-I-say-no-to-uploading-my-TraceTogether-data-when-contacted-by-the-Ministry-of-Health- (accessed on May 22, 2020).

[19] Healthy Together. Available online: https://www.healthytogether.io/ (accessed on May 22, 2020).

[20] Care19. Available online: https://www.ndresponse.gov/covid-19-resources/care19 (accessed on May 22, 2020).

[21] How accurate is the CARE19 coronavirus tracking/tracing app? Available online: https://kelo.com/news/articles/2020/apr/27/how-accurate-is-the-care19-coronavirus-trackingtracing-app/1011620/ (accessed on May 22, 2020).

[22] Germany launches smartwatch app to monitor coronavirus spread. Available online: https://fr.reuters.com/article/companyNews/idUKKBN21P1SN (accessed on May 22, 2020).