The PACT protocol specification

version 0.1 (4/8/2020)

Authors and contributors:

- Ronald L. Rivest, Massachusetts Institute of Technology <rivest@mit.edu>, Editor
- Hal Abelson, Massachusetts Institute of Technology <hal@mit.edu>
- Jon Callas, ACLU <jcallas@aclu.org>
- Ran Canetti, Boston University <canetti@bu.edu>
- Kevin Esvelt, Massachusetts Institute of Technology <esvelt@mit.edu>
- Daniel Kahn Gillmor, American Civil Liberties Union <dkg@aclu.org>
- Louise Ivers, M.D., Harvard Medical School <LIVERS@mgh.harvard.edu>
- Yael Tauman Kalai, Massachusetts Institute of Technology <tauman@mit.edu>
- Anna Lysyanskaya, Brown University <anna_lysyanskaya@brown.edu>
- Adam Norige, MIT Lincoln Laboratory <anorige@LL.mit.edu>
- Bobby Pelletier, MIT Lincoln Laboratory <bobby.pelletier@ll.mit.edu>
- Ramesh Raskar, Massachusetts Institute of Technology <raskar@mit.edu>
- Adi Shamir, Weizmann Institute <adi.shamir@weizmann.ac.il>
- Emily Shen, MIT Lincoln Laboratory <emily.shen@ll.mit.edu>
- Israel Soibelman, MIT Lincoln Laboratory <isoibelman@ll.mit.edu>
- Michael Specter, Massachusetts Institute of Technology <specter@mit.edu>
- Vanessa Teague, Thinking Cybersecurity <vanessa@thinkingcybersecurity.com>
- Ari Trachtenberg, Boston University <trachten@bu.edu>
- Mayank Varia, Boston University <varia@bu.edu>
- Marc Viera, MIT Lincoln Laboratory <mviera@ll.mit.edu>
- Daniel Weitzner, Massachusetts Institute of Technology <weitzner@mit.edu>
- John Wilkinson, MIT Lincoln Laboratory <wilkinson@ll.mit.edu>
- Marc Zissman, MIT Lincoln Laboratory <maz@ll.mit.edu>

Note on authors and contributors: Each has participated in developing this specification in their individual capacity. Affiliations are listed for identification purposes only and do not imply any kind of approval or commitment on behalf of the listed organizations.

We describe here the PACT (Private Automated Contact Tracing) protocol, a simple, decentralized approach to using smartphones for contact tracing based on Bluetooth proximity. Users of this scheme do not reveal anything about themselves, unless they volunteer to do so. In particular, users can volunteer to donate their private data to a (trusted) health authority, who can then use this data to further control the spread of the virus, but this is discretionary to the users.

This protocol has been developed by a large collaboration of researchers.

Our approach is similar in structure and many details to the following proposals:

- the Covid-Watch proposal https://www.covid-watch.org/article, and its most recent strawman protocol by the TCN coalition https://github.com/TCNCoalition/TCN
- the Canetti et al. protocol https://arxiv.org/abs/2003.13670
- the DP-3T protocol https://github.com/DP-3T/documents
- the (other) PACT protocol: https://arxiv.org/pdf/2004.03544.pdf (Note the unfortunate collision of the use of the name PACT. One could disambiguate by calling this reference the "West Coast PACT" and the current paper the "East Coast PACT". The proposals are very similar.)

We would be happy to see technical convergence on the details. Such convergence would allow for greater interoperability, resulting in less confusion about alternatives, wider adoption, and greater medical utility.

Introduction

The coronavirus pandemic motivates many (including us) to see how we can use our abilities for the good of humanity. What can we do to help shorten or diminish this crisis (and prevent future pandemics)?

One approach, followed here, is to provide automated support for tools that were once exclusively manual. In our case, we focus on the task of *contact tracing.*

The goal of contact tracing is to quickly identify individuals who may be virus carriers, before they even show symptoms, so that they may be tested, quarantined, and/or advised to monitor themselves for symptoms. By removing such individuals from the circulating population, the spread of the virus may be diminished.

The method used with traditional manual contact tracing is typically

- Begin with each individual who has tested positive for the virus (an "index case")
- Identify, through a painstaking in-person interview, as many of the contacts as possible that the index case has had since he became infectious, and

- Check with each such contactee to see if they are showing symptoms of the disease; if so, advise that they get tested immediately. If not, advise that they monitor themselves for symptoms (and typically, to self-quarantine).

The process of manual contact tracing, if applied aggressively, can slow the spread of the disease. This is true in spite of the obvious limitations of manual contact tracing: people don't remember who they have had contact with many days ago, and even if they do remember them they may not know their names or how to contact them.

An automated contact tracing system, based on smartphones, can alleviate some of these problems. Smartphones can (1) know where they are, through GPS, and (2) determine proximity to other phones, using Bluetooth. There are proposals for automated contact tracing based on both location (GPS) and proximity (Bluetooth), or both. Our PACT proposal is primarily proximity (Bluetooth) based, because GPS is generally not precise enough, and it doesn't work well indoors.

If privacy were not a concern, it would be simple to build a smartphone-based automated contact tracing system: each phone would continually send its current location to a central server, and also continually send the server the phone numbers of phones it had been close to recently (obtained via a simple Bluetooth protocol implemented for this purpose). The server can then determine, for each index case, the phone numbers of phones belonging to people the index case may have exposed; these people can then be contacted, as above for manual contact tracing.

But when one is concerned about privacy, it gets more complicated. A phone cannot continuously broadcast its phone number (or any static identifier), as then it can be identified and tracked easily. Most important, any information contained within a phone should not be disclosed without the consent of the owner.

So, the PACT protocol broadcasts constantly-changing and randomly-chosen "chirp" values, which are not useful in identifying the owner (unless the owner asserts to someone that those are indeed his own chirp values).

Furthermore, the PACT protocol satisfies the property that "*no information, aside from these chirp values, ever leaves the user's phone without his permission.*" **The privacy of the user is paramount in the PACT design.**

When a user is tested and found positive, he becomes an "index case" in the traditional terminology. In the PACT system, an index case is strongly encouraged (but not required) to make public the chirp values he has broadcast in the past three weeks, so that others may learn that they may have been exposed to the disease by being close to the index case. If someone learns that they have been so exposed, they may work with health authorities to determine the appropriate course of action.

That completes our brief overview of the PACT protocol. The following sections provide more technical detail.

High-Level Structure

Our technical approach for performing contact tracing can be broken down into three components.

Chirping layer. Each person's smartphone emits anonymous, random "chirps" on a more-or-less continuous basis. These chirps are not linkable to the device or its owner, and are changed ("rotated") frequently (on the order of minutes) so that they cannot be used to track the device. Each device keeps track of the chirps it transmits, as well as of the chirps it receives.

The chirps are emitted using the Bluetooth Low Energy (BLE) protocol, and their received signal strength is used as a proxy for the distance between the two devices. The BLE protocol is one-way: packets with a short payload are transmitted and received, but the Sender doesn't need an answer, and doesn't know if the Receiver even exists, or received the packet. Multiple Receivers may receive the same BLE packet. We model our proposal on Apple's Find My protocol (https://www.apple.com/icloud/find-my/).

The distance between the two devices is, in turn, used as a proxy for the distance between the two individuals, and it is an important factor in determining how likely one of them is to infect the other, should they happen to be infectious. Another factor in determining this likelihood is the duration of contact, which our protocol also tries to provide an estimate for (as we will see below).

Other metadata may also be helpful in determining one's risk: for example, a restaurant is a much more risky place than a bike path. Such metadata is not included in a chirp, but each device may optionally store metadata locally associated with a received chirp in its contact log.

Of course, there are many factors that smartphones cannot measure, such as the extent to which an individual is infectious, whether they are wearing a mask or other protective gear, etc. These factors come into the picture too, but not at this juncture.

Tracing layer. We need to make it possible for an individual (Alice, for concreteness) to determine her risk of contracting a disease based on whether or not, and to what extent, she has been in contact with infected individuals. This step requires that Alice interact with a database that will help Alice determine her level of exposure.

At a high level, an entry in the database corresponds to a chirp emitted by a device whose owner (let us call him Bob) has tested positive for the disease. The entry does not identify this individual or this device in any other way, and does not include any information about where the device was when this chirp was emitted.

Care must be taken in how this information gets into the database. We want to make sure that an authorized medical professional facilitates the data entry, so that indeed the chirps stored there credibly correspond to the devices of infected individuals. Moreover, even if a device is hacked, it cannot upload chirps that were sent by a different device. Finally, for the purposes of protecting an individual's privacy, we can make it possible for the device's owner to redact some of his chirps and not upload them into the database.

The chirps corresponding to Bob's device are essentially just random numbers that cannot be linked to any information identifying Bob. In fact, they cannot even be linked to each other (although for the purposes of efficiency we may group small subsets of them together, so that chirps emitted by Bob's device, say, in a one hour window of the same day are in fact linked). So in the absence of any other information (for example, if Bob did not come into contact with anyone, or to an individual who did not come into contact with Bob), the chirps do not reveal anything about Bob.

To Alice who has been in contact with Bob, these chirps do in fact communicate something: by comparing the chirps in the database to the chirps she stored in her contact log, she can see that she has in fact been in contact with an infected individual. From other metadata information, such as the chirp's signal strength and how many times Alice has heard it, as well as where she was at the time, Alice's PACT app can determine the extent of her exposure. The app can be designed to show all the metadata to Alice directly, or to simply tell Alice her exposure score and nothing else.

Interacting with medical professionals. The goal of PACT is not to subsume medical professionals, but to give them the tools that they can use in their fight with disease. It is up to medical authorities to decide what to advise individuals to do based on the information they receive via the PACT app. Many factors may come into play in determining whether or not Alice should be tested for disease -- her symptoms, her risk factors, her location -- her exposure risk score provided by PACT is just one of them. Any decisions for what Alice should do from here should be done in consultation with doctors and health authorities.

Entities

The system involves the following entities:

- *Users* of the phone app. Individuals using the application may either be *diagnosed* by a health provider as COVID-positive or *non-diagnosed*; this status can change over time.

- *Health providers* are authorized to make a positive diagnosis that an individual has COVID-19, based on a positive test or an evaluation of symptoms.
- *Testing authorities* who ensure that only information from diagnosed individuals can be uploaded.
- An *exposure database* storing information uploaded by diagnosed individuals. For efficiency reasons, the exposure database may be sharded by geographical region.

Chirping Layer

Seed generation

Each phone generates a random 256-bit seed every hour. This seed is known only to that phone. The phone stores in memory each seed along with the time it was generated, for a medically relevant period of time, after which the seed is erased. These seeds are used to generate chirps. (The “relevant period" cannot be known in advance; to build in resilience for future crises, including ones that might be adversarially designed, we might use 3 months. We’ll use 3 months in this document.)

Chirping

Every few seconds, the phone generates and broadcasts (using Bluetooth) the chirp r_t=PRF(s,t) where t is the current time (measured to one-minute precision), s is the current seed, and PRF is a pseudorandom function. The timing of broadcasts does not need to be precise or regular, but it should be fairly frequent, so that a listener can estimate the extent of the exposure. In our proposal each chirp is 28 bytes long, which is the length of an Apple *Find My* key.

We emphasize that the chirps generated by each phone are pseudorandom, and thus no one seeing the chirps of a phone can predict what any other chirps of that phone will be at a different time, and no one can tell if two chirps came from the same phone or not.

Listening

Each phone listens for such chirps, and stores any chirp that it hears for 3 months, which is long enough to combat most possible pandemics. Along with each chirp, it also stores:

- The time at which this chirp was received (measured to one minute precision); if the same chirp was received several times during the same minute it is stored only once.
- The maximum signal strength of the Bluetooth signal of the chirp it heard (we may choose to store only chirps with sufficiently strong maximum signal strength,

so that the Bluetooth contact becomes a better proxy for a possible coronavirus transmission).

- It may also store the location when the chirp was received.

We emphasize that this information is only stored locally in the receiver's phone and is never broadcast.

Logs

To summarize, we note that each phone maintains two logs:

- A *seed log*, recording all the seeds that were generated in the last 3 months, together with the time that each seed was generated (measured to one-minute precision).
- A *contact log,* recording all the chirps that were received in the last 3 months, together with the time (and possibly the location) that each chirp was received, and its signal strength.

Tracing Layer

Permission numbers

We use the permission number concept and terminology of the Covid-Watch proposal.

Each testing authority generates a list of permission numbers. The testing authority distributes a disjoint set of permission numbers to each doctor or other health provider authorized to diagnose individuals as "positive".

Each permission number is “use once”: it is used to authorize the upload of information about contact events from one diagnosed individual.

A permission number is “valid” if it is on the list of permission numbers given to health providers. Permission numbers should be random and sufficiently long so that it is infeasible for someone else to generate valid permission numbers, and it is impossible or unlikely for two permission numbers to be the same. There may be additional authorization procedures to ensure that only diagnosed individuals can upload contact information.

Uploading information about contact events

If a user is diagnosed positive, he is given an unused permission number, which authorizes him to upload his chirp logs (more precisely, the seeds that generated those chirps) to the “exposure database”. The diagnosed individual enters the permission number into the app running on his phone. The server providing the upload service checks that the permission number has not been used before and is a valid permission number (i.e., on the list of permission numbers previously given to health providers).

For each chirp he sent during the relevant period, the diagnosed individual uploads to the database (or the relevant shards of it) the tuple (s, t_1, t_2), where s is the seed used to generate the chirp, and t_1 and t_2 are the start and end times of the period (e.g., hour) during which the seed was used. (If the diagnosed individual wishes to exclude a given time period of his chirp logs from the upload, he can substitute a fresh random seed for the seed he actually used during that period.)

Detecting contact events

To learn whether she has been in contact with any diagnosed individuals, a user downloads the exposure database (or the relevant shards of it) and checks whether any chirps in her contact logs “match” any entries in the database.

Specifically, for each tuple (s, t_1, t_2) in the database, the user computes r = PRF(s, t) for each time t between t_1 and t_2, and checks whether her contact log contains r as a chirp received at approximately time t. If so, this indicates a contact with a diagnosed individual.

The number of contacts the user had with diagnosed individuals at each time, potentially combined with additional metadata in the contact logs, can be used to provide appropriate notifications and advice to the user.

The app can perform the downloads and checking of the exposure database automatically and regularly; the user need not make an active query.

Analysis

In this section, we describe some of the many design considerations that informed our design of the PACT system.

Performance

In the chirping layer, the size and rate of chirps are designed to fit within the small payload of a Bluetooth low energy (BLE) packet and for compatibility with a wide range of mobile platforms, including Android and iOS devices. PACT is purposely designed for simplicity of understanding and implementation, so that we can deploy the system on short order and add additional features later (see the Extensions section). We believe that the PACT system is compatible with the existing architecture developed by Apple for their Find My system.

The tracing layer involves a balance between two performance metrics: communication to all users in order to deliver updates to the exposure database, and local computation for each device to determine proximity with diagnosed individuals. The PACT architecture is deliberately optimized to reduce communication as much as possible while providing autonomy and privacy. Each diagnosed patient uploads only a few kilobytes of data to the exposure database, which the database service then pushes to all devices within the same geographic region.

Autonomy

PACT is designed around the principles of openness and transparency, providing users with full control over the entire lifecycle of the system. Users can freely choose whether to join the system, and a diagnosed individual controls the process of uploading their seeds to the exposure database; that is, a healthcare professional cannot perform this upload without the user's permission. Furthermore, users may temporarily "snooze" transmission of their chirps if desired, and furthermore a diagnosed individual can also decide not to upload a subset of their seeds even if they didn't use the snooze feature in real-time. This decision is privacy-preserving: nobody else, not even the user's health provider, can detect this choice. Finally, as noted by the DP-3T research team, any private contact tracing system gracefully "sunsets" once the pandemic has passed, since no more uploads will occur, while still remaining active and available to prevent future epidemics from becoming full-blown pandemics.

Privacy

We detail in this section the extent to which PACT protects the privacy of users' location and healthcare status. We stress upfront that the healthcare privacy of PACT is *not* absolute. As a simple example, if a user has only come into contact with one other person during the relevant period, then the user can learn whether the other person is diagnosed as COVID-positive based on whether she receives a notification from the PACT system. This type of privacy loss is unavoidable in any automated contact tracing system, and as a result we allow people to make autonomous decisions about the appropriate balance between social health and civil liberties.

That having been said, PACT provides the highest amount of privacy possible for an automated contact tracing system, even against a technologically sophisticated attacker with the ability to eavesdrop on all nearby BLE transmissions and to write custom software that interacts with all devices in close proximity. In particular, the chirping layer of PACT is fully private: each chirp is unconnected to the user's identity and unlinkable to any other chirp. If the exposure database is sharded by geographic region, then any user who wishes to hide her geographic region can download the entire database.

However, there is some privacy loss when a diagnosed individual participates in the tracing layer. Anybody who comes into close proximity of the diagnosed individual can recognize when the chirp they received from the diagnosed individual appears later in the exposure database. If they additionally remember the people nearby when receiving that chirp (e.g., by taking a photograph), then they may be able to re-identify the diagnosed individual. Furthermore, because each seed can be used to link all chirps transmitted within a one hour period, multiple contacts could collectively reconstruct the partial movement pattern of the diagnosed individual by combining the chirps they received. We emphasize that the diagnosed individual retains full privacy against non-contacts; in particular, the diagnosed individual's healthcare and location information cannot be recovered solely from the exposure database.

Finally, PACT only requires a small amount of trust in the exposure database; it is designed to be public since it only contains lists of randomly generated seeds that we wish to provide to all other users. We only rely on the exposure database in two ways. First, the database maintainer should not maintain any logs regarding the identities of the diagnosed users who uploaded information; alternatively, the application can perform the uploading over an anonymous network communication system like Tor. Second, the database maintainer should not work with the testing authority to link diagnosed users with the permission numbers that they submit.

Integrity

PACT protects many types of active attacks on the integrity of the proximity notifications provided by the system. Concretely, it is difficult for an attacker to fool any victim into receiving a notification when she has not come into contact with any diagnosed user, or into suppressing a notification when she has come into contact with a diagnosed user.

For protection against suppression, we rely upon the exposure database to store all uploaded seeds properly; note that this is a verifiable action, since diagnosed users and their health providers can confirm that the upload has completed successfully. As long as the database operates correctly, anyone who has come into contact with a diagnosed user will receive a chirp generated by one of the seeds posted on the database. We remark that a diagnosed individual may choose not to upload part/all of her seeds to the database; while this might be viewed as a case of suppression, we permit this activity in order to give diagnosed individuals the autonomy to make this decision.

The analysis of the converse direction is more complex. We consider below several ways in which an active attacker might attempt to mislead and scare people into incorrectly believing they were in contact with a diagnosed person.

First, suppose that the attacker is a diagnosed individual who wishes to upload seeds that do not correspond to chirps he sent. Because the attacker cannot feasibly guess the seed necessary to generate anyone else's transmitted chirps, he cannot mislead any of his non-contacts into believing they were a contact.

Second, suppose that the attacker is a non-diagnosed individual who wishes to replay chirps from other users who are either already diagnosed or may become diagnosed. Specifically, the attacker rebroadcasts at a later time any chirps that were previously transmitted by other users at an earlier time. (The attacker may have received the earlier chirps, or the attacker might recompute these chirps from the exposure database.) Here, the attacker's goal is to mislead her contacts at the later time into believing that they came into contact with the diagnosed user. Even with full knowledge of the exposure database from all diagnosed individuals, the recipients of the attacker's rebroadcast chirps will detect the attack because two chirps computed from different seeds or different times won't be equal.

Finally, a non-diagnosed individual might try to guess a valid permission number in order to upload her seeds, in order to mislead her contacts into believing they were in contact with a diagnosed individual and/or to prevent a future diagnosed individual who legitimately obtains that permission number from being able to upload his information. We protect against this by requiring that permission numbers are randomly sampled from a large enough space that it is infeasible to guess a valid permission number.

Integration with Apple *Find My*

Since it may be challenging to get a new app distributed quickly and widely enough to help reduce the spread of infection, we discuss in this subsection the prospect of making use of the Apple *Find My* functionality already deployed in many existing smartphones.

With *Find My,* each phone periodically broadcasts a 28-byte public key. This key changes randomly every 15 minutes to protect the privacy of the phone's owner. We don't need to follow the *Find My* protocol further here.

A natural extension to the PACT protocol is to use the *Find My* broadcasts as if they contained chirps, and not public keys. (More simply, we just treat the PKs as if they were chirps, and don't encrypt anything with them.) This provides a broad base of already-functioning chirps that can be used in the PACT protocol for contact tracing.

Interoperability

Some portions of this proposal need to be standardized to provide interoperability between different implementations. Such portions include:

- The format and length of chirps
- The format of seeds, as uploaded to the exposure database(s).

Extensions

The initial PACT protocol is purposely designed with simplicity in mind for ease of implementation and deployment, because we want to realize the benefits of a private contact tracing system as soon as possible. As a result, rather than being feature-full at launch, PACT is designed with extensibility in mind so that we can add medically-relevant indicators of the spread of the novel coronavirus while protecting the privacy of users.

In this section, we list some extensions of the base PACT protocol that provide improved healthcare awareness and stronger privacy guarantees.

Chirp repeaters with delay

The base PACT protocol only detects when two devices come into close contact at the same time. However, the novel coronavirus is also capable of surviving on surfaces for a moderate period of time. Here, we propose an extension to PACT that allows for contact tracing when the virus is transmitted through fomite (surface) transmissions. Our main idea is to place a device on each surface that many people commonly touch, such as the checkout counter at a store.

These devices could be smartphones that execute the PACT app as if they were an individual, thereby receiving a notification when the surface has come into proximity with a diagnosed individual. Then, the owner of the device can upload seeds for a short period of time around the contact. However, the high cost of a smartphone likely renders this plan ineffective.

Alternatively, because the surface stays at a fixed location, the device corresponding to a surface can execute a simpler protocol that doesn't require connection to the internet at all. The device can act as a repeater that simply records any chirp it hears and rebroadcasts it after a short delay (say, every minute for the next 10 minutes). Any listener should treat a chirp from these devices in the same way as it would treat any other chirp; the listener simply logs the chirp as if it came from the original user, and the rest of the PACT protocol can operate in the normal fashion. Note that our replay attack prevention technique requires moderately fresh chirps, and therefore enforces a limit to the rebroadcast delay that PACT can tolerate.

Reverse contact tracing

Sometimes, if Bob becomes unwell but doesn't know from whom he caught the infection, there's a need to search for the contacts that might have infected Bob, not only those that Bob might have infected. The method of notifying those who were in Bob's proximity at the time he might have become infected is exactly the same as the means of notifying those he might have infected, except that the time range is different.

One way to implement this: when Bob uploads his seeds, he could optionally add a flag indicating that he wants to notify possible sources of infection in the past. He could either add an explicit date range or simply post the seeds from the appropriate range. Suppose Charlie infected Bob but is asymptomatic. Then Charlie's app would detect a chirp that he has heard in the date range, which might mean he is a source of infection for Bob. The app could then display an appropriate message, such as suggesting he get a serological immunity test to see whether he has ever been covid19-positive. This approach is possible given the three month storage duration of the contact logs.

Local information hiding

When receiving a chirp, the PACT app can be augmented to store relevant metadata that might be useful to jog the user's memory and assist the health provider with determining if this was an epidemiologically significant contact. This metadata may include exact/approximate location, an RSSI measurement to indicate the distance to the diagnosed user, the phone's orientation, and so on. However, this metadata may be a valuable target for people who might try to compromise or compel viewing of the user's device.

Observe that the metadata only provides a benefit to the user if the contact is later diagnosed, and therefore uploads a seed to the exposure database. Leveraging this fact, we can prevent PACT from being co-opted for mass surveillance by encrypting metadata stored locally on the device using a secret key that is a function of the contact's seed. This involves a few simple changes to the protocols described above in order to ensure that the user holds sufficient local state to check the authenticity of her contact's uploaded seeds while not holding enough local state to decrypt the metadata herself; we omit the details here for brevity.

Conclusion

We have sketched an approach to tracing contacts using smartphones.

It provides a high degree of privacy to all parties.

The protocol requires no storage of location information on any device or at any central authority. The chirps emitted to measure contacts are kept locally on each user's phone. Nothing is stored by a central authority until an individual tests positive. Even then, the central authority knows nothing about the contacts of the infected person unless the contact voluntarily communicates with the public health authority. Even after that communication, the identity of the contact is only known to the authority or the infected person if the contact choses to disclose it.

It is technically simple.

It appears to be nearly identical to that of the Covid-Watch proposal, and close to the DP^3T proposal (which we see as a good thing).

The ability to integrate Apple *Find My* beacons, and the extensions, may be novel.

We plan to see this protocol implemented soon, by many parties, in an interoperable manner!