VTT Technical Research Centre of Finland

Rendezvous based pandemic tracing by sharing Diffie-Hellman generated common secrets

Ollikainen, Ville; Halunen, Kimmo

Published: 09/04/2020

Document Version
Publisher's final version

WHITE PAPER April 9, 2020

Rendezvous Based Pandemic Tracing

by sharing Diffie-Hellman generated common secrets

Summary

There is an urgent need to shorten the period of the current economic slowdown, caused by SARS-CoV-2 virus and COVID-19 pandemic.

In order to prevent new outbreaks, it is essential to trace potential carriers of the virus. For this purpose, mobile technologies provide the most prominent solution, since mobile devices are typically present, where viruses are transmitted in human-human encounters.

In this white paper we propose a method that keeps users persistently unidentified. Consequently, there is no kind of a registration (no mobile number or other persistent identifier), that could introduce privacy issues

Instead of having focus on people, the proposal has a focus on individual encounters, rendezvouses; hence the name 'Rendezvous Based Tracing'.

Being authority-less, this proposal will address some obvious drawbacks of distributed solutions proposed this far, providing data for spatiotemporal analytics of viral transmission, while not making any compromises on actual alerting.

However, it is essential that a single solution becomes adopted within regions where people commute or travel. The presented approach shares common elements with Identity Based methods. Therefore, the authors will continue their work by examining possibilities for integration, especially with PEPP-PT.

Authors

Ville Ollikainen, M.Sc., Senior Scientist in Applied Cryptography research team / VTT

Kimmo Halunen, Ph.D., Applied Cryptography research team leader / VTT

Foreword

There is an urgent need to shorten the period of the current economic slowdown, caused by SARS-CoV-2 virus and COVID-19 pandemic.

In order to prevent new outbreaks, it is essential to trace potential carriers of the virus. For this purpose, mobile technologies provide the most prominent solution, since mobile devices are typically present, where viruses are transmitted in human-human encounters.

In this white paper, the authors will present a method for Rendezvous Based Tracing, addressing some major drawbacks of distributed solutions proposed this far. *VTT is interested in co-operating with all partners, regardless of the chosen approach, providing its multidisciplinary expertise at the disposal of common good.*

Reference: Identity Based Solutions

Western democracies have adopted values of protecting individual privacy. Therefore, distributed approaches for virus (and people) tracing are strongly preferred over centralized methods. While there are quite a few ongoing projects globally, there are two mobile solutions that have been presented in the Finnish media above others:

A. TraceTogether[1], a Singaporean solution that reportedly saved the densely populated country from a major outbreak.

B. PEPP-PT[2], a European privacy-by-design initiative with essentially similar functionality.

These (and probably several others in the making or less publicly deployed) solutions are carefully designed, not to disclose sensitive user information without user consent. They are however based on identifiable users; therefore, we will refer to these by a general term of 'Identity Based Solutions'.

In these solutions, contacts are detected by short-range radio communication means, especially Bluetooth. Other party's identifier is recorded into user's mobile phone in an encrypted format, typically using public-key encryption. If users are diagnosed as carriers, they can upload their contact lists to a database operated by an authority (such as Ministry of Health in Singapore) capable of decrypting it. The authority will then act upon the information they get.

In order to Identity Based Solutions to work, the authority gets to know, who the individual contacts were, narrowing the amount of information that can be recorded regarding any particular rendezvous. For instance, recording location and providing it to the authorities, would disclose the location of the other party as well, since the other party's identity is known. Consequentially, the presented *Identity Based Solutions do not record location*. Role of location is of specific importance, and it is discussed later in this document.

Limitations of Identity Based Solutions

In the light of how the SARS-CoV-2 has spread in the past, Identity Based Solutions fail to address two critical phenomena:

[1] https://www.tracetogether.gov.sg Accessed in April 7, 2020
[2] https://www.pepp-pt.org Accessed in April 7, 2020

1. Role of events, based on time and location, in spreading the virus. It is notable how events, such as sports events, religious gatherings and even family celebrations have rendered into Petri dishes of viral congestion. Since Identity Based Solutions do not record location (for the privacy reasons), events can be found only indirectly after interviewing individual people and analysing the interview data separately. In other words, Identity Based Solutions provide only limited support to real-time situation awareness.

2. Nihilistic and ignorant behaviour. There is a visible minority which considers being either outside or on top of the epidemic. Despite of all alerts, this minority keeps on visiting ski resorts, going to discos and museums if possible, and meeting other people; behaving normally for the moment, not being concerned about the pandemic. Before getting sick, that is. Because the pandemic is initially not respected, we can assume them to be
    a. most effective group in transmitting the virus and
    b. least interested to download and install any APP and to be registered or recorded by any means.

Proposal: Rendezvous Based Tracing

In order to properly address the two presented concerns - in addition to uncompromised virus tracing - we propose a method that keeps users persistently unidentified. Consequently, there is no kind of a registration (no mobile number or other persistent identifier), that could introduce privacy issues.

Instead of having focus on people, the proposal has a focus on individual encounters, rendezvouses; hence the name 'Rendezvous Based Tracing'.

The proposed solution has an APP to be installed to a mobile phone. Without registration, a user can install and launch the APP privately, without awareness of nearby people or any authority. This property aims at making the APP acceptable, consequently rising APP penetration, especially among the most challenging minority groups.

The process is illustrated in Figure 1 and described step-by-step in the following text.

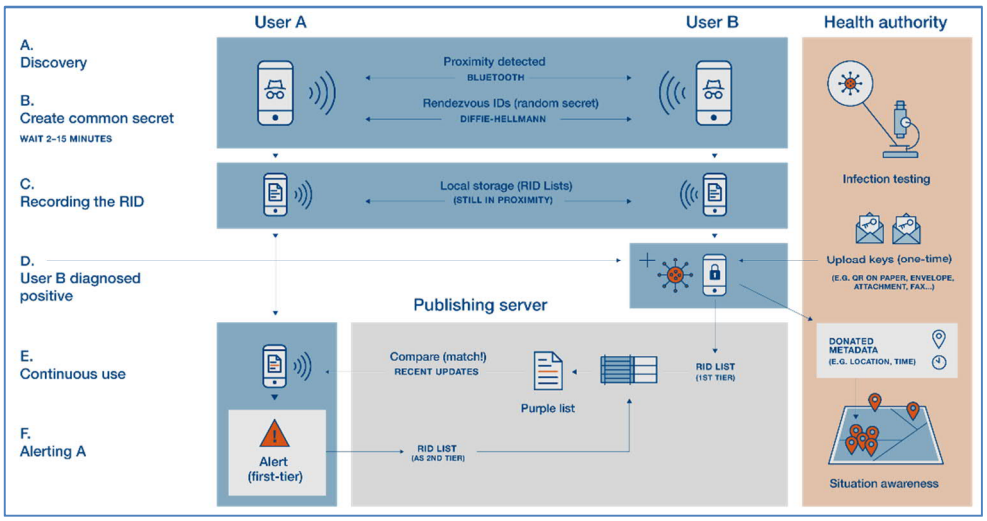

Figure 1. Process in Rendezvous Based Tracing.

A: Discovery

As in the state-of-art, Bluetooth is used to detect proximity between two users, while other data, such as estimated distance can also be extracted.

B: Creating common secret

When two users are in proximity, mobile phones calculate a common secret using the Diffie-Hellman key exchange algorithm. The result of this key exchange is hashed with a cryptographic hash function together with the exchanged Bluetooth messages:

RID = *H*(DH || exchanged Bluetooth messages),

where DH is the result of the Diffie-Hellman key exchange, RID a unique *Rendez-vous IDentifier, a common secret identifying the particular encounter*, not the parties involved. It is essential to note, that the RID is a secret that both parties share, and no one else knows.

If the two people would meet again, new random values would be in place, making the RID different for each such encounter.

Diffie-Hellman key exchange is a widely used algorithm to generate a shared secret. It has a paramount advantage, that **an eavesdropper cannot compute the outcome**. Consequentially, the eavesdropper will remain unaware what the RID was.

Because of Diffie-Hellman, parties can also negotiate and share other information, such as, where the RID will be published if either party turns out to be in a period of contagiousness at the particular time of the encounter (cf. load balancing, discussed later).

C: Recording the RID

It has been learned from previous epidemics, that longer exposure time increases likelihood of infection. Therefore, Figure 1 illustrates an *optional* delay before the RID is memorized: if the other party will not still be present after a few minutes, the RID can be discarded.

Otherwise, the RID will be recorded into a *RID list* in the mobile phone.

Obviously, RIDs don't have to be saved forever: When an incubation period is over, old RIDs can be deleted from the mobile phone. Therefore, it is also essential to store time information with each RID.

There are reasons to store time information with a certain coarse granularity, such as one day accuracy. This will be discussed later, with an illustrative example.

In addition to the RID, some metadata, such as location information or duration of the encounter, can be recorded as *metadata*. This information will be useful, if the metadata is donated for further analysis. It should be noted, that there are reasons to use asymmetric encryption for the metadata, also discussed later with the same example.

D: Positive diagnosis (of User B)

When a user is diagnosed positive, he/she (User B in Figure 1) receives anonymous a one-time disposable key, which enable

- uploading the RID list to alert others, and also
- donating possible metadata for further analysis.

It is irrelevant for the proposed method, how the upload keys are delivered, but at simplest they could be QR codes printed on paper and stored in closed envelopes in a vault at health care facilities providing test results. If the location has no upload keys at hand, keys could be delivered with secure electronic means, such as telefax if nothing else; a person skilled in the art may develop more sophisticated electronic methods.

When User B gets the key, he is allowed to upload his RID List to a Publishing Server: **only the RID List is uploaded**, no timestamps, no metadata.

Since each RID in the mobile phone list has a timestamp associated, it is advantageous to upload only the RIDs that have been generated during his congestion period.

The Publishing Server will add the uploaded RIDs to its data repository, into a *'Purple List'*. **The Purple List consists of RIDs of potentially contagious encounters**; it is publicly available, completely anonymous data.

It is advantageous to add information to each RID, whether the uploading party has actually been *diagnosed positive* ('first-tier' alerting), or has only been *in contact with an alerted suspect* yet waiting for diagnosis ('second-tier' alerting). Tier class (1 or 2) should be published with RIDs, as the only associated information.

As mentioned, each RID is a secret between the parties in an encounter: it is only the other party of an encounter, which knows the origin of the RID. Therefore, RID lists are safe to be published.

Final stages E: Continuous use and F: Alerting others

After a user has installed the APP, it begins to poll the Publishing Server, requesting updates of the Purple List.

If the same RID exists both in the Purple List and in the local memory of the phone, the APP immediately alerts the user for a possible infection.

If the alert was first-tier, i.e. originating from a positively diagnosed user, the alerted user (User A in the Figure, before any diagnosis for her) will be encouraged to disclose their RIDs for second-tier alerting. In the case of first-tier alert, the user is advantageously instructed to quickly go to nearest testing facilities, perhaps with a fast-lane ticket provided by the APP.

If the alert was second-tier, also other instructions, such as self-quarantine or social distancing could be given, depending on how much time has lapsed from local RID timestamp, when it is compared to typical incubation period.

A note on load balancing:

Obviously, if the epidemic becomes widespread, there are quite a few RIDs in the Purple List. As a simple approach it would be engineering-as-usual to create load balancing for the Publishing Server, there is also an opportunity to have a farm of publishing servers: as mentioned before, parties can also share other information after the Diffie-Hellman key exchange. Additional information after DH could be sharing, into which server in the farm they would upload their RIDs, each server having their individual Purple Lists. By doing this, the APPs may only request Purple List updates from the servers that exist in their history.

Supporting situation awareness

As mentioned above, there may be metadata recorded when creating a RID. It is fair to assume that without becoming identified, a user is likely to disclose more:

Metadata may be donated as a separate entity (i.e. without RIDs) to authorities that are maintaining situation awareness of the epidemic. The metadata enables creation of heatmaps, **visualizing spatiotemporal hotspots** for further investigation.

Some sources have mentioned that as high as 60% penetration may be needed before any app becomes effective enough for greater good. Spatiotemporal analysis will reduce the required penetration significantly, since *swift detection of emerged hotspots enable authorities to request visitor lists, ticket sales information etc. of the event that has taken place at the particular time and location*.

It should be underlined, that location is most sensitive data; there are numerous ways to combine location data with external sources to compromise the privacy of the users. **Under no circumstances will metadata be published.** As voluntarily donated, it should stay confidential.

Further notes on recording locations:

1) Google is actively tracking locations on Android phones by default. Users can switch off location tracking: If it is turned off, naturally it will not be recorded by the alerting system, either. Therefore, it is up to the user, which APPs has access to location information.
2) As long as the location belongs to a public sphere, there is only little risk of misuse. Of course, some locations may be sensitive by nature, but in that case, the user is likely to switch off location tracking completely. IF he/she is not aware of that option, it can be made visible in the APP.
3) However, if the location is in the private sphere of either party, there may be a policy not to record it at all. This would be engineering-as-usual: detecting places where people spend majority of their time may be used to disable recording location on both sides.

Further notes on recording timestamps:

1) For alerting, recording a timestamp on one-day granularity is enough, since incubation and contagious periods vary a lot. This would have a positive impact on privacy (see the next comment box).
2) For situation-awareness, on the other hand, timestamp must be accurate for precise spatiotemporal analysis. This means, that metadata timestamps should be recorded separately from RID timestamps.

It would be advantageous, it after Diffie-Hellman key exchange both parties would agree on a common timestamp and location information to be recorded as metadata. If either party denies recording, metadata would be discarded on both sides. If accepted, authorities would eventually get consistent information from both parties, making it easier to match this voluntary information.

Metadata will not be published, but it would be available for authorities creating real-time situation-awareness. If the users want to leave their contact information for authorities, that is obviously possible, but not mandatory.

Further notes on protecting metadata (with Carl and Bob example):

It is advantageous to encrypt all recorded metadata asymmetrically with a public key provided by the situation awareness authority. Why? Let's take an example:

Carl and Bob have met and they have created a RID. Suppose that the timestamp and location would be available on the mobile, and it turns out that Bob has been in contagious period. Unfortunately, Carl got infected from Bob and unintentionally passed on the disease in a family reunion, causing fatalities. If at that point the timestamp of the RID is exact, or the metadata would be available for Carl, he could reason from time/location that Bob should be held responsible for his loss. Consequences would be unpredictable.

This is why RID timestamps should have fairly coarse granularity and metadata should be appropriately encrypted in a way that only the situation awareness authority can decrypt.

Comparison to Covid Watch[3]

A list of other decentralized approached is available at gdprhub.eu website[4]. Of all the published approaches there, only *Covid Watch* by Stanford University is **essentially similar to the proposal**.

As a main difference to Covid Watch, each device in Covid Watch creates a random number on regular intervals, broadcasted and shared with others in the proximity. Both transmitted and received numbers are recorded. In contrast, the proposed method uses shared secrets created in Diffie-Hellman key exchange.

Otherwise, the alerting process is essentially the same, and benefits are very similar.

Covid Watch may scale better to densely populated areas, since individual communication between two mobiles is not necessary. However, it has two disadvantages compared to the proposed method:

1) Since the numbers are changing frequently, duration of an encounter cannot be measured. This information is valuable when it is uploaded to the Situation Awareness analytics, since we know that longer exposure increases likelihood of infection; longer durations make some hotspots even hotter than hotspots with casual passing-by.

2) Broadcasted numbers in Covid Watch are not secrets. Consequentially, they can be eavesdropped and recorded by a third party, possibly associated with data that discloses identities, such as surveillance camera footage, or credit card number at a point-of-sales. Even only from the network traffic one could use some analysis techniques to check which numbers have changed at the given refresh intervals and use this information for tracking individuals. When the server discloses numbers related to an infection, the eavesdropped data may possibly disclose the identity of the infected person. *In short, Covid Watch cannot be considered completely safe.*

However, as a whole, the authors find *Covid Watch* most interesting.

---

[3] https://www.covid-watch.org Accessed in April 9, 2020

[4] https://gdprhub.eu/index.php?title=Projects_using_personal_data_to_combat_SARS-CoV-2#Pan-European:_Decentralized_Privacy-Preserving_Proximity_Tracing_.28DP-3T.29 Accessed on April 7, 2020 at 14:00 CET

Benefits of the proposed approach

- Click-and-Use: no registration needed.
- With metadata, outburst can be swiftly located, in place and time.
    - Lower APP penetration required
- Users are in control
    - Easily explainable privacy
    - With privacy understood, users are likely to disclose more.
- The method is simple and straightforward, easy to explain and implement.

Discussion

The authors wish that this whitepaper would have a positive contribution to developing user-acceptable virus tracing. Any APP taken into use must cover as large population as possible, especially those who normally are least willing to participate.

It is fair to say that when users

- do not have to register themselves and
- have a justified feeling that no-one is able to trace them personally,

it improves acceptance of the APP.

Developers should specifically consider people with negative stance to preventing COVID-19. In the proposed approach, the selling point is obvious: as long as nothing happens, the users give out absolutely *nothing* (not even registration), only may receive *something* (even lifesaving).

It is essential that a single solution becomes adopted within regions where people commute or travel. The presented approach shares common elements with Identity Based methods. Therefore, the authors will continue their work by examining possibilities for integration, especially with PEPP-PT.

Whatever the tracing solution will be, it must always be a good servant, never a bad master.

Authors

Ville Ollikainen

Mr. Ville Ollikainen received his M.Sc. degree in Technical Physics in 1989 from Helsinki University of Technology with an academic minor and 2.5-year employment in the laboratory of Industrial Psychology, covering studies in computer assisted training and organization development. He worked previously at Technical R&D of MTV Finland. Mr. Ollikainen has been working at VTT Technical Research Centre of Finland as a senior scientist. In this role, he has since 1999 contiguously participated in projects related to new media technologies, excluding a period of entrepreneurship: he was one of the main inventors behind Envault Corporation Oy, a data security company he was establishing in 2007. He has contributed to over 30 patent applications. His current activities include privacy preserving recommendation systems, privacy-by-design targeted advertising and social networking services, most recently coordinating a European Union's Horizon 2020 funded project HELIOS (grant #825585), developing a novel peer-to-peer platform for privacy-enabled social media. Since 2019, Mr. Ollikainen has been working at VTT's cybersecurity research.

Kimmo Halunen

D. Sc. (Tech.) Kimmo Halunen, Research Team Leader at the VTT Technical Research Centre of Finland in Oulu and Adjunct Professor at the University of Oulu. He has obtained his D. Sc. (Tech.) in computer engineering on hash function security in 2012 and has over 20 publications related to security, cryptography and blockchain technology in refereed conferences and journals. In the course of his research he has contributed to several national and international projects as a researcher and as an expert of information security and cryptography in a commercialization project in authentication. He leads a project on measuring the security of cryptographic primitives in the Finnish Defence Forces research program. He has been leading cybersecurity growth initiatives at VTT from 2014 onwards. During 2014-2015 he visited the COSIC group in KU Leuven in Belgium for 9 months. He also acts as a cybersecurity advisor for Streamr.

For more information please contact

Ville Ollikainen
Tel. +358 400 841116
ville.ollikainen@vtt.fi

Kimmo Halunen
Tel. +358 40 6751836
kimmo.halunen@vtt.fi

About VTT

VTT is one of the leading research and technology organisations in Europe. Our research and innovation services give our partners, both private and public, all over the world a competitive edge. We pave the way for the future by developing new smart technologies, profitable solutions and innovation services.

We create technology for business – for the benefit of society.

VTT beyond the obvious

www.vttresearch.com