A Privacy-Preserving Protocol for Local Storage of Pseudonymous Contact Tracing Data and Asynchronous Verification Against Centralized Alerts

Maurizio Ponti

2020 | 04 | 30

Abstract. A Bluetooth Low Energy advertising protocol for contact tracing that preserves privacy through time-varying pseudonymity and by limiting contact data to be saved only locally on individual devices, and allows an asynchronous verification against periodic alerts published by government agencies.

Keywords. *Privacy-Preserving Pseudonymity, Local Storage of Contact Events, Asynchronous Verification, Centralized Alerts, Bluetooth Low Energy Advertisement Channels.*

1 Introduction

At the time of writing, Italy is nearing the day it will start loosening lockdown restrictions and it will be doing it without a clear knowledge of the actual Covid-19 contagion spread. Even worse, it might do it without a valid strategy against possible new outbreaks.

What appears to be a vital tool in the future detection and control of new cases is a collective use of a contact tracing application on mobile devices.

Steps have been made in that direction and ideas have been proposed, but they have failed to overcome privacy concerns or require time to be properly carried out due to excessive complexity.y.

In the next sections it will be presented a simple privacy-preserving protocol for contact tracing through Bluetooth Low Energy Advertisement Channels that allows users to locally store pseudonymous contact data and verify it against alerts periodically announced by public health agencies.

2 Definitions

Throughout this article and for the sake of simplicity, $\mathcal{H}$ denotes a cryptographic hash function that maps data of arbitrary size into $\mathcal{K}_1 := \{1, \ldots, 2^{128}\}$ and $\{t\}_{t\in\mathcal{K}_1}$ is a pseudorandom permutation of $\mathcal{K}_1$. Given $\mathcal{K}_2 := \{1, \ldots, 2^{256}\}$, $(pk, sk) \in (\mathcal{K}_2 \times \mathcal{K}_2)$ is a public key private key pair[1] and $n \in \mathcal{K}_2$ is a pseudorandom value.

$T0$ is a tentative title for the protocol itself[2] and a contact event describes the circumstance in which two or more individuals are at a reciprocal distance that is lower than a certain threshold defined by the competent health agency.

3 Enabling pseudonymity

Pseudonymity is the first building block of the protocol. While using Bluetooth Low Energy channels to communicate without a pairing process allows devices to avoid sharing sensitive information (e.g. GPS coordinates), pseudonymity grants an acceptable degree of opacity to protect a user identity.

When two parties, Alice and Bob, start using the $T0$ protocol to trace contacts, they respectively generate the pairs of public and private keys (pk_A, sk_A) and (pk_B, sk_B). Alice then generates a pseudorandom value $n_A \in \mathcal{K}_2$ and a pseudorandom permutation of $\mathcal{K}_1$, $\{t_A\}_{t_A\in\mathcal{K}_1}$, while Bob generates $n_B \in \mathcal{K}_2$ and $\{t_B\}_{t_B\in\mathcal{K}_1}$.

From that moment on they assume the aliases

$$\mathcal{H}\big(\mathcal{H}(pk_A + n_A) + (t_A)_j\big) \quad \text{and} \quad \mathcal{H}\big(\mathcal{H}(pk_B + n_B) + (t_B)_l\big),$$

with $j, l \in \mathcal{K}_1$ changing every 10 minutes.

Their Bluetooth enabled devices broadcast, respectively,

$$\Big(\mathcal{H}\big(\mathcal{H}(pk_A + n_A) + (t_A)_j\big), (t_A)_j\Big) \quad \text{and} \quad \Big(\mathcal{H}\big(\mathcal{H}(pk_B + n_B) + (t_B)_l\big), (t_B)_l\Big)$$

rolling to the next element of the sequences $(t_A)_j$ and $(t_B)_l$ with an interval of 10 minutes, hence changing their pseudonyms with the same frequency.

4 Tracing a contact

When Bob finds himself in the proximity of Alice, their devices listen to their respective advertisements. As the contact event occurs, Alice locally saves on her device the pair

$$\text{Alice} \leftarrow \Big(\mathcal{H}\big(\mathcal{H}(pk_B + n_B) + (t_B)_l\big), (t_B)_l\Big),$$

for a certain fixed value of $l \in \mathcal{K}_1$, according to the time Bob's advertising message has been transmitted, whereas Bob locally saves on his device

$$\text{Bob} \leftarrow \Big(\mathcal{H}\big(\mathcal{H}(pk_A + n_A) + (t_A)_j\big), (t_A)_j\Big),$$

[1] For the sake of this first draft of the protocol, only the public key will be involved.

[2] A reference to the concept of separation in topological spaces.

again for a fixed $j \in \mathcal{K}_1$, depending on Alice's setup.

No other information is stored, neither locally nor remotely, by the two parties. No pairing process is involved in the exchange.

Based on the properties held by the cryptographic hash function $\mathcal{H}$, neither Alice can determine

$$\mathcal{H}(pk_B + n_B),$$

despite knowing the right $(t_B)_l$ for the value she recorded, nor Bob can determine

$$\mathcal{H}(pk_A + n_A).$$

5 Verification against government alerts

At a later time, when Alice is tested by a public health agency and found positive, she can communicate to the health agency the value

$$\mathcal{H}(pk_A + n_A),$$

the list of pseudonyms of the contact events she stored on her device, and the (finite) sequence of $(t_A)_j$ used so far to modify her own aliases.

The last two can be used by the health agency to verify possible abuses from users claiming to have come in contact with Alice or that may fabricate fake contact events with her.

The health agency drafts on a daily basis a list of positive subjects, communicating on behalf of each of them the same key value Alice gave, i.e.

$$\mathcal{H}(pk + n) \quad \text{for certain fixed } pk, n \in \mathcal{K}_2.$$

Bob is able to download from a trusted source the updated list of positive subjects and verify on his own if he came in contact with any of them. For instance, he can now compute

$$\mathcal{H}\big(\mathcal{H}(pk_A + n_A) + t\big)$$

for any $t \in \mathcal{K}_1$ of the encounters he has stored on his device and compare the outcome with the corresponding pseudonym.

Bob finds out about his contact event with Alice and in his own health interest notifies the government agency to coordinate his next steps.

Once Alice and Bob are cured (or simply found not contagious), they can generate new random values n_A and $n_B \in \mathcal{K}_2$ to employ a different set of time-varying pseudonyms.

6 Bluetooth data transmission without a pairing process

The Bluetooth Low Energy (BLE) standard used to communicate without a pairing process limits the number of bits available for messages. Since the idea behind $T0$

is to define a protocol available to a wider range of Bluetooth enabled devices, it is advisable to make the $T0$ protocol Bluetooth 4.0 compatible. Furthermore, the sole purpose of the message sent through BLE Advertisement Channels is to be registered by listening devices and not to actually establish a connection, which allows to free up the field usually allocated for the advertisement address, reinforcing at the same time users' privacy.

Hence there are 296 bits available for $T0$ messages thorugh BLE advertising channels that can be structured as

$T0$ Flag	Pseudonym $\mathcal{H}(\dots)$	Permutation t
40 bits	128 bits	128 bits

A unique 40 bit flag needs to be agreed among the users of the $T0$ protocol, in order for devices listening for BLE advertisements to tell apart $T0$ messages from other kind of advertisements.

Each user plays both the role of the transmitter (TX) and the receiver (RX) of advertisements, alternating between the two functions TX and RX in brief and randomized time frames that allow all parties involved to properly register contact events without role collisions.

To better outline a radius under which a contact event is such, it is then possible to determine the distance between a transmitter and a receiver based on the strength of the Bluetooth signal and act accordingly.

7 Potential drawbacks and risks

A possible attack model involves a transmitter or a set of transmitters spamming an area with aliases that roll at a faster frequency to overwhelm recording resources on the receiving side, faking multiple contact events. The local nature of the application of the $T0$ protocol and the intrinsic low storage requirement (i.e. 256 bits, as of the current outline of the protocol) for saving a pseudonym $\mathcal{H}(\dots)$ and its corresponding permutation key t, already mitigate the impact of such an attack. To further reduce the risk of similar abuses it can be established a minimum broadcasting time for a $T0$ cycle of advertisements, which is also consistent with the concept of a non-negligible contact event.

At the same time, a high number of contact events could trigger an alarm on the receiving side, because it could also imply the user is in the presence of too many people and the area needs to be cleared out.

References

[1] Giorgio Corbellini, Stefan Schmid, and Stefan Mangold "Two-Way Communication Protocol Using Bluetooth Low Energy Advertisement Frames", 2015. Association for Computing Machinery. Proceedings of the 1st International Workshop on Experiences with the Design and Implementation of Smart Objects.

[2] Naresh Gupta “Inside Bluetooth Low Energy, Second Edition”, 2016. Artech House mobile communications series, Artech House.

[3] Michael Luby and Charles Rackoff “Pseudo-random Permutation Generators and Cryptographic Composition”, 1986.

[4] Philip Rogaway and Thomas Shrimpton “Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance”, 2004. Fast Software Encryption, Springer Berlin Heidelberg.

[5] Fazli Subhan, Asfandyar Khan, Sajid Saleem, Shakeel Ahmed, Muhammad Imran, Zubair Asghar, and Javed Iqbal Bangash “Experimental analysis of received signals strength in Bluetooth Low Energy (BLE) and its effect on distance and position estimation”, 2019. Transactions on emerging telecommunications technologies, Clarivate Analytics.