Considerazioni su Privacy e Security delle App di Proximity Tracing

Alessandro Armando, Università di Genova
Mario A. Bochicchio, Università del Salento
Francesco Buccafurri, Università Mediterranea di Reggio Calabria
Alberto Marchetti Spaccamela, Sapienza Università di Roma
Fabio Massacci, Università di Trento
Francesco Palmieri, Università di Salerno
Paolo Prinetto, Politecnico di Torino
Silvio Ranise, Fondazione Bruno Kessler, Trento

Considerazioni su Privacy e Security delle App di Proximity Tracing

Alessandro Armando, Università di Genova
Mario A. Bochicchio, Università del Salento
Francesco Buccafurri, Università Mediterranea di Reggio Calabria
Alberto Marchetti Spaccamela, Sapienza Università di Roma
Fabio Massacci, Università di Trento
Francesco Palmieri, Università di Salerno
Paolo Prinetto, Politecnico di Torino
Silvio Ranise, Fondazione Bruno Kessler, Trento

Sommario

Le app per il proximity tracing sono applicazioni software che utilizzano gli smartphone per facilitare il tracciamento dei contatti con persone positive al virus e possono giocare un ruolo importante nel contrasto alla sua diffusione. Non sorprende dunque che la comunità scientifica, le aziende fornitrici di tecnologia e i governi si siano attivati, ciascuno nel proprio ambito d'azione, per promuovere lo sviluppo di soluzioni che siano efficaci ma anche rispettose della privacy dei cittadini. In questo articolo illustriamo le principali soluzioni proposte e discutiamo le problematiche più importanti relative alla sicurezza e alla privacy poste dalle applicazioni di proximity tracing.

1. Introduzione

Numerosi governi – fra cui l'Italia – hanno proposto o stanno proponendo delle applicazioni per smartphone finalizzate al `tracciamento dei contatti' (*contact tracing* in Inglese). Queste applicazioni, discusse in dettaglio nel seguito, non vogliono sostituire ma aiutare il personale sanitario ad individuare i soggetti infetti e coloro che, avendo trascorso del tempo con questi ultimi, sono ritenuti a rischio di infezione.

L’obiettivo è duplice: da un lato informare le persone che sono state vicine a qualcuno che è portatore del virus; dall’altro monitorare la diffusione del virus in modo da valutare l'efficacia complessiva delle misure di confinamento. In ambedue i casi l’obiettivo finale è quello di fermare, il prima possibile, la diffusione del virus.

Le soluzioni a oggi disponibili non sono tra loro equivalenti e l’obiettivo di questo documento, pensato per i non addetti ai lavori, è di evidenziare le principali differenze e le conseguenze, anche inattese o non del tutto evidenti, derivanti dall’adozione (volontaria o obbligatoria) di una soluzione piuttosto che di un’altra.

2. Requisiti Funzionali e di Privacy

Col termine *contact tracing* si intende il processo di identificazione di persone infette o che sono venute in contatto con individui infetti. La procedura, gestita dalle autorità sanitarie, prevede l’individuazione dei soggetti infetti e dei loro contatti, la verifica dello stato di contagiosità mediante test opportuni (ad es. “tamponi”), l’isolamento (quarantena) e la somministrazione di opportuni trattamenti ai soggetti positivi al test [1]. L'obiettivo ultimo è quello di fermare la diffusione della malattia attraverso l'identificazione, l'isolamento e la cura dei soggetti infetti. Il contact tracing, da sempre affidato a operatori umani[1] opportunamente formati, è stato utilizzato con successo, in combinazione con altre strategie, per controllare ed eliminare, malattie infettive importanti, quali il vaiolo.

Nel caso del COVID-19, il processo di contact tracing prevede 5 passi: (1) un individuo, chiamato *caso indice*, viene identificato come portatore della malattia (ad esempio, il tampone risulta positivo); (2) sul caso indice vengono raccolte ulteriori informazioni quali età, sesso, residenza, domicilio, descrizione dei sintomi e loro data di inizio (se sintomatico), data in cui è stato eseguito il tampone, data in cui è stato fornito il risultato del tampone, data di isolamento, data di ospedalizzazione, data di guarigione/decesso; (3) successivamente vengono raccolti ulteriori dettagli sugli spostamenti del caso indice e sulle persone con le quali è venuto in contatto; (4) ogni contatto del caso indice viene intervistato al fine di raccogliere informazioni riguardanti la sua relazione col caso indice (ad esempio familiare, convivente, collega, amico), la finestra temporale di esposizione e il periodo di quarantena; (5) per ogni contatto del caso indice che effettua un tampone e viene trovato positivo, si riapplica la stessa procedura. Il contact tracing realizzato dagli operatori umani è strettamente integrato con l'esecuzione dei tamponi e raccoglie una grande quantità di dati personali sugli individui coinvolti, nel caso in cui questi risultino positivi al tampone.

[1] Un operatore umano, se autorizzato, può garantire l’anonimato e l’accesso alle cure anche a soggetti ad alto rischio (ad esempio immigrati clandestini o addetti alle consegne a domicilio) che, per paura di essere rintracciati o di perdere il lavoro, potrebbero voler evitare altre forme di tracciamento da parte delle autorità.

Per fronteggiare più efficacemente la pandemia di COVID-19 si è proposto di agevolare il compito degli operatori di contact tracing sfruttando la grande diffusione degli smartphone per identificare tutti i contatti con caratteristiche (di durata e distanza tra i soggetti) tali da renderli potenzialmente a rischio ai fini della trasmissione del contagio. Questa nuova funzionalità degli smartphone, detta di *proximity tracing*, consiste nel processo di identificazione di eventi di prossimità, ovvero la determinazione delle situazioni in cui due o più entità (che possono essere oggetti, veicoli o persone) si trovano a una determinata distanza tra di loro per un dato intervallo temporale. Uno dei possibili modi per implementare il proximity tracing per le persone è basato su un'applicazione per smartphone che utilizzi la tecnologia Bluetooth. In questo modo diventa infatti possibile rilevare quando due persone mantengono una distanza di circa 2 metri (o inferiore) per un periodo di tempo predeterminato (ad esempio, almeno 15 minuti). Tale evento di prossimità viene considerato potenzialmente rilevante per il contact tracing del COVID-19 nel caso in cui almeno una delle due persone venga rilevata come positiva all'infezione nel periodo successivo al verificarsi dell'evento (ad esempio, entro 2-3 settimane). L'individuo che viene a sapere di avere avuto un contatto potenzialmente rischioso ha la possibilità di tutelare sé stesso e gli altri, rivolgendosi volontariamente alle autorità sanitarie, che lo inseriranno nel processo di contact tracing. Se una percentuale significativa segue questa strada, il proximity tracing risulta particolarmente efficace in quanto gli operatori umani che implementano il contact tracing sono aiutati nel processo di identificazione dei possibili contatti di un caso indice.

Per stimolare un processo virtuoso che incentivi l'utilizzo di questo strumento, è necessario che esso offra garanzie adeguate per la tutela della privacy dei soggetti coinvolti. Nel caso dell'uso di applicazioni mobili, questo richiede, oltre all'adozione di tecniche per garantire la sicurezza e la privacy dei dati raccolti, anche opportuni meccanismi per manifestare la volontà di partecipare all'azione di proximity tracing e permettere l'identificazione degli eventi di prossimità, minimizzando i dati raccolti. Per la volontarietà, l'applicazione deve offrire la possibilità, in maniera chiara e semplice, di dare o negare il consenso al trattamento dei dati personali. Per la minimizzazione, la tecnologia più utile sembra essere quella *Bluetooth Low Energy* (BLE), che non localizza in maniera assoluta gli individui (come fa invece il GPS), ma solo la loro posizione relativa. La tecnologia BLE è stata inoltre individuata come base per la definizione di un approccio comunitario al proximity tracing finalizzato a garantire l'interoperabilità delle applicazioni mobili sviluppate dai diversi stati membri dell'Unione Europea [2]. Questo sembra essere un requisito fondamentale per permettere una rapida e sicura ripresa degli spostamenti delle persone tra gli stati membri nell'area Schengen.

L'interoperabilità pone due sfide importanti. La prima è di natura concettuale: bisogna capire come armonizzare i diversi modelli (centralizzati o decentralizzati, vedasi sez. 3.1) di gestione dei dati raccolti dalle applicazioni sviluppate dagli stati membri per rispettare i diritti degli individui, in primis quello alla privacy. La seconda sfida è di natura tecnologica e deriva dal contesto eterogeneo del mercato degli smartphone, che pone varie barriere alla possibilità di scambiare dati tra i diversi modelli e i relativi Sistemi Operativi. Per vincere questa sfida, sembra necessaria la collaborazione degli attori coinvolti senza la quale

l'efficacia delle applicazioni per il proximity tracing può risultare difficoltosa a causa, ad esempio, di incompatibilità tra diversi Sistemi Operativi; significativa, al riguardo, la soluzione recentemente sviluppata in modo congiunto da Apple e Google [3].

Occorre poi evidenziare come, nel caso dell'adozione su vasta scala delle applicazioni di proximity tracing, la quantità di dati, potenzialmente enorme, veicolata da tali soluzioni possa diventare un punto critico sia per la sicurezza sia per la privacy degli individui coinvolti: le informazioni sui contatti delle persone e sui luoghi dove queste si trovano sono infatti informazioni private ai sensi di legge. Risulta pertanto cruciale definire una strategia adeguata e in grado, da un lato, di garantire la disponibilità dei dati raccolti dalle applicazioni esclusivamente per il tempo strettamente necessario e, dall'altro, permettere un corretto ed efficace funzionamento del processo di contact tracing.

Nel seguito si analizzeranno alcuni dei problemi di sicurezza e privacy dei protocolli più diffusi per il proximity tracing basato su applicazioni mobili e tecnologia BLE; si analizzerà inoltre il potenziale impatto negativo di tali problemi sui diritti e sulle libertà fondamentali dei soggetti coinvolti.

Per quanto riguarda la privacy, un sistema di tracciamento compatibile con le norme europee in merito deve raccogliere dati che devono essere utilizzati solo per motivi di salute pubblica: ogni altro scopo privato (ad esempio marketing, guadagno commerciale) o pubblico (potenziale controllo delle persone non autorizzato dalla legge) deve essere impedito. Pertanto è necessario porre delle regole per la raccolta dei dati e dei limiti nel loro utilizzo.

L'EDPB (il board europeo dei garanti della privacy) ha sottolineato al riguardo che "il quadro giuridico sulla protezione dei dati [il regolamento GDPR] è stato progettato per essere flessibile e, in quanto tale, è in grado di ottenere sia una risposta efficace nel limitare la pandemia sia nella protezione dei diritti umani e delle libertà fondamentali" [4]. Le linee guida che qualsiasi programma proposto per tracciare la diffusione di COVID-19 deve rispettare per quanto riguarda la privacy toccano tre punti fondamentali. Innanzitutto il programma di monitoraggio deve essere implementato solo per una durata prestabilita associata alla pandemia e deve prevedere, di norma, la partecipazione volontaria; eventuali casi eccezionali in cui la partecipazione fosse obbligatoria deve essere espressamente regolata per legge. I primi due punti sono semplici da verificare, non destano ambiguità e costituiscono implicazioni di legge che non sembrano essere in discussione. Il terzo punto riguarda il tipo di informazione tracciata e il possesso delle informazioni raccolte. Per quanto riguarda il tipo di informazione tracciata, l'orientamento prevalente (che coincide con la proposta Immuni) è quello di tracciare solo i contatti di prossimità fra le persone, ma non i singoli movimenti.

3. Protocolli per il Proximity Tracing

Il digital contact tracing si fonda essenzialmente sulla possibilità di rilevare la prossimità tra gli smartphone degli utenti, allo scopo di individuare situazioni di possibile contagio dovuto a contatti diretti tra gli individui. La prossimità tra due smartphone viene rilevata attraverso l'interfaccia Bluetooth Low Energy (BLE) a bordo dei dispositivi. Per garantire la privacy degli utenti, gli smartphone si scambiano degli pseudonimi. Per ogni "contatto", ciascun dispositivo memorizza lo pseudonimo inviato dalla controparte, l'intensità del segnale (utile per stimare la distanza tra i due dispositivi) e la durata.

Come vedremo le soluzioni possono essere diverse, e vi sono potenziali minacce di sicurezza e privacy, a livello sia di protocollo sia tecnologico, che debbono essere tenute in stretta considerazione nella scelta della soluzione da adottare. È interessante quindi fornire qualche riflessione al riguardo, partendo dalla descrizione dei principali approcci esistenti.

3.1 Protocolli decentralizzati e centralizzati

I protocolli per il Proximity Tracing si dividono in protocolli *decentralizzati* e protocolli *centralizzati*.

Nei protocolli decentralizzati, tra i quali ricordiamo DP-3T (sez. 3.2) e la soluzione proposta da Apple e Google (sez. 4.1), gli pseudonimi vengono generati e memorizzati nei dispositivi senza il coinvolgimento di un server (e quindi di un'autorità) centrale. L'unica informazione che viene condivisa con un server centrale sono gli pseudonimi utilizzati dalle persone positive che volontariamente (e di concerto con l'autorità sanitaria che ne ha accertato la positività) decidono di trasmettere questa informazione. Il server rende a sua volta disponibile questa informazione alla comunità degli utenti. Il server centrale non entra dunque in possesso dell'informazione di prossimità tra i dispositivi (*grafo dei contatti*) che rimane rimane memorizzata in modo distribuito sui dispositivi degli utenti. Tali informazioni sono infatti gestite e mantenute nei singoli dispositivi conservando gli pseudonimi associati agli utenti si è verificato un evento di prossimità in un una finestra temporale assunta come significativa (attualmente 14 giorni). Il protocollo prevede che, a fronte dell'accertamento della positività di un paziente e sulla base di un meccanismo di autorizzazione gestito dalle autorità sanitarie, il paziente può volontariamente comunicare a un'entità centrale le informazioni sufficienti a generare tutti gli pseudonimi utilizzati dal paziente nell'arco della finestra temporale di contagio, di modo tale che tutti gli utenti che aderiscono al sistema possono verificare se, tra gli pseudonimi che essi hanno raccolto nell'arco della finestra temporale, siano presenti pseudonimi associati ai pazienti che hanno comunicato la positività all'infezione. Il protocollo prevede che, in base all'entità della prossimità in termini di distanza reciproca e durata, possa essere calcolato un rischio associato a ogni potenziale contatto. Pertanto l'utente che rilevi di essere stato in prossimità di un paziente positivo all'infezione, anche sulla base del rischio determinato in accordo alle precedenti

informazioni, potrà volontariamente contattare le autorità sanitarie per ricevere adeguate istruzioni sulle azioni da eseguire (ad es., indagini epidemiologiche, isolamento, tampone).

La privacy dei protocolli decentralizzati è basata sull'assunzione che dagli pseudonimi utilizzati dal dispositivo non sia possibile risalire all'identità del possessore. Qualora ciò non venisse garantito, la pubblicazione degli pseudonimi conseguente alla notifica di positività esporrebbe il paziente positivo al rischio di stigmatizzazione. Attacchi che consentono di identificare l'utente associato a un insieme di pseudonimi (i cosiddetti *attacchi di de-anonimizzazione*) sono improbabili, ma non impossibili. È infatti facile prevedere lo sviluppo di applicazioni malevole finalizzate alla raccolta degli pseudonimi dei dispositivi in prossimità e alla loro marcatura temporale. In ambienti strutturati, come ad esempio quello lavorativo, la combinazione di queste informazioni con altri dati, quali la lista dei partecipanti a incontri effettuati, può facilmente portare all'identificazione di un utente positivo. Va da sé che questa raccolta e l'uso degli pseudonimi rappresentano un reato.

Nei protocolli centralizzati gli pseudonimi vengono generati da un server centrale e assegnati ai vari dispositivi in fase di registrazione. Quando un utente scopre di essere positivo, il suo smartphone invia al server centrale gli pseudonimi dei dispositivi con cui è venuto in contatto. Quando questi dispositivi contatteranno il server, quest'ultimo comunicherà loro di essere stati in prossimità con il dispositivo di un utente positivo e il rischio associato. Il server centrale non comunica gli pseudonimi degli utenti positivi e quindi non vi è rischio di stigmatizzazione causata da un attacco di de-anonimizzazione degli pseudonimi. Per contro, il server centrale acquisisce il grafo dei contatti dei dispositivi. Per quanto questa informazione possa essere utile dal punto di vista dell'analisi epidemiologica, è evidente come questo rappresenti una minaccia alla privacy dei cittadini (ancorché portata da un'organizzazione governativa).

3.2 Il Protocollo DP-3T

DP-3T (Decentralised Privacy-Preserving Proximity Tracing) [5] è un protocollo Open Source definito da un gruppo di lavoro europeo per il contrasto della diffusione della pandemia di COVID-19 attraverso la realizzazione di sistemi e applicazioni per il proximity tracing che preservino la privacy dei cittadini.

DP-3T si basa su un modello decentralizzato e prevede due modalità, chiamate rispettivamente *Low-Cost* e *Unlinkable*.

Nella modalità *Low-Cost*, gli pseudonimi, chiamati *Ephemeral ID*, sono ottenuti attraverso l'applicazione di una funzione one-way applicata su un seed generato dal client e mantenuto segreto. Grazie alla non invertibilità della funzione, dalla conoscenza di uno o più Ephemeral ID non è possibile risalire al seed. Il seed ha una tempo di validità limitato (e.g., un giorno), trascorso il quale è possibile generare un nuovo seed attraverso l'applicazione di un hash crittografico, da cui verranno a loro volta generati i nuovi Ephemeral ID. In questo caso la

notifica dei positivi è fatta attraverso la comunicazione in broadcast del seed relativo al primo giorno della finestra temporale di contagio che il paziente che scopre la propria positività all'infezione ha volontariamente comunicato all'autorità sanitaria (a seguito di opportuna autorizzazione). In tal modo ogni client può calcolare gli Ephemeral ID di ogni giorno della finestra temporale e verificare se essi sono presenti nel proprio database. Il vantaggio dell'approccio risiede nella derivazione degli Ephemeral ID da una chiave, il che consente di risparmiare spazio in memoria del dispositivo e banda. Questo ovviamente introduce dei problemi di de-anonimizzazione (perché gli pseudonimi dei positivi sono in questo modo collegabili tra loro), che vengono però mitigati cambiando la chiave ogni giorno.

La seconda modalità, denominata *Unlinkable*, differisce dalla prima sia per il fatto che gli Ephemeral ID generati sono sempre diversi e indipendenti dai precedenti sia per il meccanismo con cui avviene la notifica dei potenziali contagi. Così come per il Low-Cost, dalla conoscenza di un Ephemeral ID non è possibile risalire al seed che lo ha generato, in quanto ogni seed genera un unico Ephemeral e i seed sono indipendenti tra loro. Il paziente positivo può quindi essere selettivo sui seed che comunica all'autorità sanitaria e, inoltre, la notifica avviene non attraverso la comunicazione dei seed, ma tramite un meccanismo che permette agli utenti di verificare se sono venuti in contatto con qualcuno degli Ephemeral ID associati a persone infette e presenti in un apposito filtro. A fronte di un maggiore volume di traffico scambiato e di maggiore spazio di memorizzazione richiesto ai client, la soluzione Unlinkable offre maggiori protezioni in termini di privacy, in quanto non viene fornito agli utenti alcun meccanismo per legare tutti gli Ephemeral ID associati alla persona infetta, come avviene, invece, nella soluzione Low-Cost.

3.3 Il Protocollo ROBERT

Il protocollo ROBERT (ROBust and privacy-presERving proximity Tracing protocol) [6], sviluppato congiuntamente da INRIA e Fraunhofer, adotta un approccio centralizzato ed è implementato nella app StopCovid [7] sviluppata dal governo francese. Anche ROBERT assume che la prossimità venga rilevata attraverso l'interfaccia Bluetooth Low Energy (BLE) a bordo degli smartphone e il calcolo del rischio è fatto in base alla potenza del segnale rilevato dall'interfaccia (come stima della distanza) e dalla permanenza del contatto.

Quando la app si registra al server centrale, quest'ultimo genera un insieme di pseudonimi per quella app e li condivide con la app stessa. Completata la registrazione, la app condividerà i propri pseudonimi (presi a caso tra quelli forniti dal server centrale) e raccoglierà quelli forniti dai dispositivi incontrati. Se un utente scopre di essere positivo potrà trasmettere al server centrale l'insieme degli pseudonimi raccolti nel periodo di possibile contagio. Il server centrale calcolerà il rischio di contagio degli utenti associati. Tale informazione sarà resa disponibile agli utenti interessati tramite le loro app.

Se, da un lato, ROBERT ha il vantaggio di non richiedere la “pubblicazione” di informazione (ancorché anonimizzata) sugli utenti positivi, dall’altro il suo svantaggio sta nel fatto che il server centrale è in grado di costruire il grafo dei contatti degli utenti positivi.

4. Sistemi per la notifica delle esposizioni

In questa sezione vengono brevemente introdotti il sistema di notifiche di esposizione proposto da Apple e Google e la app Immuni.

4.1 Il sistema notifiche di esposizione di Apple e Google

Apple e Google hanno congiuntamente definito e implementato sui rispettivi Sistemi Operativi (iOS e Android) un sistema di notifiche di esposizione basato su modello decentralizzato [3]. Tale sistema è pensato per supportare l'implementazione di app per il proximity tracing offrendo alcune garanzie di sicurezza e privacy potenzialmente interessanti. Innanzitutto sia le chiavi utilizzate per la generazione degli pseudonimi sia gli stessi pseudonimi vengono generati dal Sistema Operativo. Allo stesso modo, gli pseudonimi raccolti dallo smartphone vengono memorizzati in zone di memoria riservate al Sistema Operativo e pertanto non accessibili alle app di contact tracing. Ciò mitiga il rischio di esfiltrazione degli pseudonimi rispetto al caso in cui queste informazioni fossero gestite direttamente dalle app. Queste ultime sono infatti applicazioni complesse, pensate prevalentemente per gestire l’interazione con l’utente e non si può escludere che soffrano di vulnerabilità che consentano a un attaccante di esfiltrare i dati in esse contenuti o di inserire pseudonimi corrispondenti a pazienti infetti con i quali non siamo venuti in contatto. Nel primo caso l’attaccante potrebbe determinare se siamo stati a contatto con un paziente positivo, nel secondo potrebbe indurci a credere che lo siamo stati anche se ciò non è avvenuto, generando così un falso allarme.

4.2 Immuni

Come tutte le app di contact tracing basate sul sistema di notifiche di esposizione offerto da Apple e Google, Immuni delega la generazione e la gestione degli pseudonimi (inclusi lo scambio e la memorizzazione) ai servizi offerti dal Sistema Operativo. L’altro aspetto delicato per la privacy dell’applicazione è il caricamento delle chiavi per la generazione degli pseudonimi sul server. Immuni effettua il caricamento delle chiavi senza associare a esse informazioni che possano ricondurre all’identità dell’utente. In particolare alle chiavi viene associata la provincia di residenza e altre informazioni di natura epidemiologica e operazionale che, a una prima analisi, non sembrano essere problematiche dal punto di vista della privacy.

Apprezzabile è la disponibilità del codice sorgente sia della app sia dei servizi lato server. In linea di principio la disponibilità del codice consente di ispezionare e verificare se il servizio

implementato sia in linea con la documentazione e, in caso contrario, di identificare eventuali deviazioni o vulnerabilità. In pratica, però, come esplicitamente ammesso nella documentazione che accompagna il codice di Immuni, è molto difficile verificare che la versione della app distribuita tramite i canali ufficiali di Apple e Google corrisponda effettivamente al codice distribuito. Encomiabile è l'ammissione del problema da parte degli sviluppatori di Immuni, così come l'invito alla comunità scientifica a contribuire alla sua risoluzione contenuto nella documentazione che accompagna il codice della app.

Questo problema non è di secondaria importanza. A tal proposito è opportuno osservare che, contrariamente alla prassi, la app di Immuni distribuita su Google Play, non è firmata digitalmente dallo sviluppatore, ma da Google stessa. Sorge quindi il dubbio che neppure le autorità italiane siano in grado di verificare la corrispondenza tra il codice distribuito pubblicamente e la versione distribuita su Google Play e installata sui nostri smartphone.

5. Minacce a Security e Privacy

In questa sezione analizzeremo alcune delle principali potenziali minacce a Security e Privacy, derivanti dalle varie soluzioni adottate.

5.1. Potenziali Minacce a Integrità e Confidenzialità

Come analizzato nella sez. 3, sia i protocolli che si basano sul modello decentralizzato sia quelli che utilizzano un approccio centralizzato fondano il loro funzionamento sullo scambio reciproco di identificativi (pseudonimi) che avviene tra dispositivi che si trovano in prossimità. Vale la pena di soffermarsi con maggiore dettaglio sulle potenziali minacce che possono derivare, direttamente o indirettamente, da questo meccanismo. Esse infatti possono mettere a rischio sia la privacy degli utenti sia la corretta esecuzione dei protocolli, determinando quindi conseguenze inattese e potenzialmente lesive.
In entrambi i modelli vi è la potenziale minaccia che la raccolta degli identificativi venga effettuata da un agente avversario e malevolo (ad esempio, mediante la distribuzione, in una determinata area geografica, di un ampio numero di dispositivi preposti alla raccolta di pseudonimi prodotti dagli smartphone dei passanti) che può così memorizzarli, associandoli a informazioni di contesto (come ad esempio immagini che ritraggono le persone associate agli identificativi, il luogo, etc.). Bisogna chiedersi in quale misura tale eventualità possa comportare un leakage di privacy, visto che, in linea di principio, gli identificativi scambiati sono pseudonimi. È evidente che se la vittima dell'attacco risulta positiva all'infezione, visto che la notifica da parte del server dovrà necessariamente permettere l'identificazione del potenziale contatto a tutti gli smartphone che hanno catturato gli identificativi di tale individuo, sarà sempre possibile per l'attaccante legare le informazioni catturate all'informazione sullo stato di salute della vittima. Ma c'è un rischio ancora maggiore. In questo scenario, infatti, la vittima potrà essere anche tracciata negli spostamenti avvenuti nell'arco della finestra temporale di contagio. Se gli identificativi non sono collegabili, ciò è

possibile solo se l'attaccante collude con il server, il quale, deve necessariamente conoscere il modo per collegare gli identificativi di uno stesso soggetto.

Un'altra minaccia alla quale sono esposti entrambi i modelli (centralizzato e decentralizzato) è la possibilità, da parte del server, di collegare l'identità reale dei pazienti positivi agli identificativi (pseudonimi) usati per il proximity tracing. Dobbiamo considerare infatti che, nel caso generale, l'entità che rileva la positività di un paziente (che può essere una istituzione pubblica o privata, come per esempio un laboratorio di analisi, un ospedale, una struttura sanitaria) non abbia alcun accesso ai dati gestiti dal server sul proximity tracing. Tuttavia, non è da escludere la collusione tra le due entità. Visto che la comunicazione da parte del paziente rilevato positivo deve essere effettuata autonomamente dallo stesso paziente, proprio per evitare che si possa collegare l'identità reale con le identità fittizie, è però necessario che tale comunicazione venga autorizzata presso il server dall'entità che rileva la positività del paziente. Senza questa misura, infatti, chiunque potrebbe comunicare falsamente la positività di identità fittizie, creando allarme sociale e quarantene inappropriate. Sfortunatamente, però, l'autorizzazione di per sè potrebbe rappresentare il meccanismo attraverso il quale server e entità che rileva la positività, colludendo, possono associare l'identità reale alle identità fittizie. Il momento dell'autorizzazione alla comunicazione dei propri identificativi al server è quindi particolarmente critico e dovrebbe essere progettato con specifica attenzione nei confronti della minaccia in questione.

Il modello decentralizzato è esposto a un'altra utilizzazione anomala dei protocolli che su di esso si fondano. Le informazioni scambiate, infatti, dovrebbero avere come unico scopo la gestione dell'epidemia. Ogni altro utilizzo deve essere visto come una minaccia a sicurezza e privacy. Un possibile utilizzo anomalo potrebbe essere, per esempio, la notarizzazione preventiva, da parte di comuni utenti, di contatti avvenuti in forma pseudonima, che, una volta de-anonimizzati possono svolgere il ruolo di prova anche in procedimenti legali. In un modello centralizzato, solo il server, non quindi un utente qualsiasi, potrebbe usare tali dati in maniera non conforme alle finalità previste per il protocollo.

Vi è un'altra possibile distorsione dei protocolli basati su modello decentralizzato. Se è vero che gli pseudonimi di un individuo vengono raccolti dagli smartphone presenti in prossimità, tali pseudonimi potrebbero essere a loro volta inoltrati, anche successivamente, ad altri smartphone, impersonificando, quindi, i proprietari dei suddetti identificativi e generando, di conseguenza, falsi contatti. Similmente, ogni altro meccanismo che possa essere usato da un paziente consapevole di essere positivo per iniettare i propri identificativi fittizi nello smartphone di una vittima con cui realmente non è stato in contatto (con l'obiettivo di danneggiare la vittima inducendone l'isolamento), è una minaccia da tenere in stretta considerazione, anche tenendo conto della disponibilità di soluzioni tecnologiche che permettono di iniettare tale informazione anche a distanza.

Come considerazione finale, è opportuno osservare che il rilevamento della prossimità tra due dispositivi potrebbe non essere basato sullo scambio di informazioni che i dispositivi

effettuano perchè in prossimità, ma potrebbe invece derivare da informazioni di localizzazione. In tal caso il modello sarà verosimilmente centralizzato, considerando che centralmente è possibile derivare dalla localizzazione dei dispositivi la reciproca prossimità. Tale modello non necessariamente implica *by design* la possibilità di tracciare le posizioni degli utenti, in quanto la posizione assoluta potrebbe non essere memorizzata dal server in forma intelligibile. È tuttavia sempre importante considerare il rischio del tracciamento, che, in questo caso, può assumere le dimensioni della sorveglianza di massa.

5.2 Problemi derivanti dalla connettività Bluetooth

La connettività Bluetooth, e in particolare lo standard BLE (Bluetooth Low Energy) con cui sono equipaggiati buona parte dei moderni dispositivi, risulta essere l'unica soluzione facilmente percorribile per garantire, al contempo, un'adeguata precisione e l'anonimato nel tracciamento dei contatti. Le possibili alternative, quali la localizzazione-triangolazione via rete cellulare e l'uso della localizzazione tramite sistema GPS, oltre a problemi di imprecisione, sarebbero disastrose dal punto di vista della privacy (l'anonimato e la decentralizzazione della raccolta dati sarebbero praticamente impossibili).

Per poter partecipare al tracciamento di prossimità è dunque necessario tenere BLE sempre attivo. Anche se questo non costituisce un problema sostanziale dal punto di vista del consumo energetico (in termini, ad esempio, di durata della batteria) va considerato, dal punto di vista della sicurezza, che BLE non è un protocollo particolarmente robusto (sono state recentemente segnalate almeno una decina di vulnerabilità note di vario tipo relativamente allo stack BLE [8]). Ciò può consentire a soggetti malevoli, localizzati in prossimità di dispositivi vulnerabili, di effettuare diversi tipi di attacco, finalizzati a DoS, esecuzione di codice ostile, o addirittura privilege escalation per compromettere la disponibilità del dispositivo o infiltrare malware o esfiltrare dati, con ovvie conseguenze a livello di privacy e compromissione dell'anonimato o, peggio, di data breach.

5.3 Problemi derivanti dalla affidabilità delle misurazioni

Il meccanismo di tracciamento di prossimità si basa fortemente sulla misurazione della potenza del segnale in radiofrequenza ricevuto (RSSI) e sullo scambio di semplici messaggi periodici (beacons) che trasportano informazioni di identità (pseudonimi) attraverso opportune tecniche crittografiche. Purtroppo, quando si lavora con le onde elettromagnetiche, si opera in un ambiente estremamente "incerto" dove la potenza del segnale può essere influenzata da molti fattori, fra i quali la mobilità, l'orientamento dei dispositivi, la geometria/forma dello spazio interessato considerando, in particolare, la presenza di oggetti assorbenti (tasche, borse, corpo umano, pareti) o riflettenti (superfici metalliche, vetro), nonché disturbi di carattere ambientale (interferenza elettromagnetica). Inoltre, un ruolo non secondario è giocato da fattori dipendenti dal dispositivo utilizzato,

quali, a titolo di esempio, il tipo di antenna, i chipset utilizzati, l'eventuale incompleta aderenza agli standard.

Tutti questi elementi sono in grado di introdurre del rumore casuale nella stima delle distanze di contatto, che già di per sé sono caratterizzate da una precisione limitata rispetto al target di misura (1 metro). Gli stessi inventori del protocollo Bluetooth, Jaap Haartsen and Sven Mattisson, hanno espresso perplessità in merito all'uso dello strumento a scopo di contact tracing [9]. Va inoltre considerato che la misurazione di prossimità è cieca al contesto: è possibile trovarsi a distanza di 4 metri sottovento a una persona che tossisce nella mia direzione (ed essere ovviamente esposto al contagio) o a distanza di meno di un metro da una persona da cui sono separato da un muro/vetro quindi di fatto non esposto. Infatti, è chiaro che le onde elettromagnetiche possono passare attraverso materiali che il virus non è in grado di attraversare.

5.4 Potenziali Attacchi

Il fatto di basarsi totalmente sulla trasmissione/ricezione di onde elettromagnetiche espone il meccanismo di tracciamento a una serie di attacchi (tipicamente di "trolling"), tutti essenzialmente associati alla difficile proteggibilità fisica di risorse nello spettro radio. I più banali partono dalla potenziale attendibilità o schermatura dei segnali in trasmissione attraverso il principio della gabbia di Faraday (esistono bustine o contenitori opportunamente progettati allo scopo), che comunque hanno un senso molto limitato in un ambito di applicazione volontaristico come quello che ci coinvolge.

Analogamente è possibile utilizzare tecniche di radio jamming (il protocollo BLE utilizza la banda ISM, quindi l'hardware per realizzare attacchi di questo tipo è disponibile a buon mercato) per inibire il tracciamento in particolari aree. Viceversa, amplificando opportunamente il segnale (cosa facilmente realizzabile con hardware semplice da reperire) in modo da violare esplicitamente le specifiche di potenza massima previste dallo standard BLE, è possibile indurre la segnalazione di un contatto a rischio anche se in realtà il trasmettitore attaccante è situato a una distanza ben maggiore di quella di guardia.

Sempre in ambito Trolling/Denial of Service (DoS), il meccanismo è suscettibile sia ad attacchi di tipo "replay and relay" (messaggi di contatto associati a identificativi pseudonimizzati ricevuti da altri vengono acquisiti, memorizzati e poi re-inviati in altri contesti e ad altri dispositivi, anche molto distanti, in un secondo tempo) o di tipo "inverse sybil" (opposto di un attacco sybil, dove un dispositivo tende ad assumere l'identità di molti altri, mentre in questo caso molti dispositivi possono assumere l'identità di un dispositivo solo, magari notoriamente associato a un infetto). Questi ultimi due problemi, che evidenziano la criticità delle possibilità di intercettazione passiva e attiva in ambito radio, sarebbero risolvibili abbastanza facilmente attraverso tecniche crittografiche, ma in taluni casi a discapito dell'anonimato [10].

È anche possibile pensare a potenziali attacchi all'anonimato che possono essere facilmente alla portata sia di chi riesce a intercettare una notevole quantità di dati di contatto, sia di chi, utilizzando altri strumenti, riesce a tener traccia temporale dei propri contatti. In dettaglio, nel primo caso, partendo da una notevole quantità di dati di tracciamento completamente anonimi, senza informazioni di identità o geolocalizzazione, è possibile (rappresentando il grafo di prossimità attraverso un modello "multi-scale force-directed" [11]) ricostruire la forma dell'area da cui i dati provengono, e, a partire dalla stessa, procedere alla ri-geolocalizzazione dei punti interessati, con conseguente potenziale compromissione delle identità coinvolte. Il secondo, molto più banale, sfrutta una caratteristica attuale di alcune app, tra cui "immuni", che auspicabilmente verrà rimossa in futuro, che nel segnalare un contatto a rischio fornisce un riferimento temporale del contatto stesso [12], dando la possibilità a un utente che ha tenuto rigorosamente traccia dei propri contatti di rompere l'anonimato.

5.5 Considerazioni sulla soluzione Apple-Google

Una menzione a parte merita l'iniziativa congiunta Apple-Google su cui molte delle applicazioni (compresa Immuni) si basano. Per facilitare le operazioni di tracciamento, pur non avendo sostanzialmente modificato il protocollo BLE dal punto di vista trasmissivo, è stato introdotto un nuovo servizio di "contact detection" con relative application programming interfaces (API) integrate nel Sistema Operativo, che consentono alle app di operare sia su Android sia su iOS. Ovviamente, ciò non comporta modifiche al livello delle architetture hardware deputate alla gestione del BLE, ma sarà presente sui futuri aggiornamenti di sistema dei due brand. Per quanto sia stato abbondantemente chiarito che i dati di tracciamento non saranno resi disponibili per altri usi a livello di sistema (dove ovviamente ciò pregiudicherebbe totalmente l'anonimato) che non siano le app di tracciamento, trattandosi di Sistemi Operativi proprietari dei quali non è generalmente possibile visionare il codice, non vi è alcuna garanzia assoluta in merito. In ogni caso chiunque aggiornerà il Sistema Operativo troverà a bordo questa nuova funzionalità, anche se deciderà di non attivarla. Va capito infine, se sarà possibile tenere spenta l'Interfaccia Bluetooth senza mettere l'intero dispositivo in modalità aereo: alcune recenti scelte dei produttori sembrano poter andare in questa direzione [13].

L'adeguatezza della soluzione appena descritta presuppone un atto di fiducia nei confronti di Apple e Google. Entrambe le aziende hanno dichiarato che non utilizzeranno i dati raccolti dal servizio di notifica da loro sviluppato, ma alcune domande sono legittime. Quali garanzie abbiamo che Apple e Google onoreranno questa promessa? Ad Apple e Google converrà abusare delle nostra fiducia? Se ci limitiamo a considerare il valore economico derivante dallo sfruttamento delle informazioni di contatto, la risposta è certamente positiva. È tuttavia plausibile che il rischio (reputazionale e conseguentemente anche economico) derivante da un eventuale abuso superi di gran lunga le prospettive di guadagno portate dallo sfruttamento commerciale di questi dati.

Va infine evidenziato come l'introduzione di questo nuovo servizio di contact tracing potrà avere l'effetto di semplificare drasticamente problemi attualmente piuttosto complessi quali la co-localizzazione fine indoor, sia personale sia di gruppo, risultando, nel post-pandemia, una potenziale nuova vulnerabilità per la privacy degli utenti.

6. Considerazioni finali

Come abbiamo visto, i possibili attacchi alla privacy sono di due diversi tipi: da un lato possibili usi impropri dei dati da parte dell'autorità governativa o delle aziende private che gestiscono l'applicazione e dall'altro la difficoltà di qualsiasi tecnologia a resistere ad attacchi informatici od umani anche a basso livello tecnologico (mettere il telefono in una borsetta foderata di carta di alluminio).

Per questa ragione è necessaria la massima trasparenza delle decisioni e delle modalità operative, con cui i dati verranno processati, non solo dal governo stesso ma anche dalle società private che gestiranno tali dati. In questo senso è da apprezzare la scelta del governo Italiano di scegliere una soluzione decentralizzata, che offre maggiori garanzie rispetto alla potenziale sorveglianza invasiva, e di richiedere che tutto il codice sviluppato sia pubblico, sia per il codice che verrà installato sul telefono che, ancora più importante, quello del lato server.

Questa ultima decisione permette una verifica da parte di esperti indipendenti di quali informazioni siano scambiate durante il tracciamento e l'individuazione di possibili debolezze da parte di ricercatori e hacker etici. Molti dei possibili attacchi informatici sfruttano vulnerabilità derivanti da errori di progettazione o di programmazione non voluti. L'esame pubblico del codice e delle soluzioni proposte per la sicurezza è una forma di garanzia ulteriore e favorisce il rapporto di fiducia cittadino-governo. A lungo termine questo scrutinio, visto inizialmente dalle aziende come un `fastidio', porta a soluzioni più sicure per tutti.

La necessità di trasparenza sull'uso dei dati, da parte sia privata sia pubblica, suggerisce la necessità di un'autorità esterna che verifichi l'accuratezza e l'efficacia delle misure attuate a protezione della privacy. Le verifiche devono includere, tra l'altro, le modalità di gestione delle informazioni raccolte, l'analisi tecnica delle debolezze e delle vulnerabilità delle applicazione e la coerenza tra il software rilasciato e quello effettivamente implementato. Rimane infatti aperto il problema, già evidenziato, di garantire la continua corrispondenza tra il software reso pubblico e quello effettivamente messo in esercizio sia sui dispositivi degli cittadini che nei server delle autorità governative.

La fiducia è la condizione indispensabile per creare le condizioni per l'accettabilità sociale di qualsiasi soluzione, e quindi garantire l'efficacia di queste misure di tracciamento. Qualsiasi sistema di contact tracing risulta vulnerabile qualora le assunzioni alla base della politica

sanitaria non siano allineate con gli incentivi dei singoli e delle aziende in quanto questi ultimi cercheranno di circuire il sistema.

Uno dei rischi più maggiori è la tentazione di `*semplificarsi la vita*' da parte delle autorità sanitarie e quindi di utilizzare la tecnologia per automatizzare scelte intrinsecamente umane. L'utilizzo di meccanismi automatici di decisione ha conseguenze inattese che spesso vanificano i vantaggi portati dalla tecnologia. Il meccanismo automatico più rilevante e di cui si comincia a parlare sui mezzi d'informazione, è di utilizzare il contact tracing per mettere automaticamente in quarantena i contatti digitali di un soggetto positivo, piuttosto che informarli e farli partecipare a un processo attivo di contenimento (nel loro stesso interesse).

Illustriamo tre semplici esempi di conseguenze inattese di decisioni automatiche.

Ogni tecnologia deve fare delle scelte per necessità approssimative. Ad esempio bisogna decidere una durata prefissata della tracciatura al fine di proibire che i dati vengano conservati indefinitamente. Tale durata, oggi fissata a 14 giorni, risulta sufficiente nella maggioranza dei casi. Basta leggere i giornali per capire che alcune persone risultano contagiose ben oltre i 14 giorni. Un meccanismo automatico crea una falsa sensazione di sicurezza (non ho incontrato `nessuno' e quindi sono tranquillo) che porta immediatamente ad allentare la guardia.

Un secondo esempio è che qualsiasi tecnologia genera necessariamente dei falsi positivi, dovuti a contatti digitali che non sono necessariamente contatti fisici. Si pensi all'esempio già menzionato dell'accesso allo sportello coperto dal plexiglas (dove il virus non passa ma il bluetooth si). Laddove il numero di falsi positivi che risultassero in misure di confinamento automatico fosse troppo alto, un numero sempre più alto di utenti prenderebbe "contromisure casalinghe" per impedire che questo avvenga (la borsetta promossa a gabbia di Faraday appena discussa).

Infine un imprenditore potrebbe, ad esempio, avere interesse a chiedere a tutti i propri dipendenti di disabilitare l'app sul posto di lavoro, in quanto un lavoratore positivo determinerebbe immediatamente la chiusura dell'intera attività per quarantena automatica dei contatti digitali. Attività che invece non verrebbe chiusa se non dopo il controllo positivo del test sugli altri lavoratori che potrebbero benissimo non essere contagiati.

È quindi importante osservare che nel momento in cui il Governo volesse rendere l'installazione obbligatoria o associare il contatto digitale di un positivo a meccanismi di coercizione obbligatoria dovrebbe preliminarmente valutare se tali meccanismi non potessero trasformarsi in un costoso boomerang.

Riferimenti Bibliografici

[1] WHO - Contact tracing in the context of COVID-19 - https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19

[2] European Commission - Coronavirus: a common approach for safe and efficient mobile tracing apps across the EU - https://ec.europa.eu/digital-single-market/en/news/coronavirus-common-approach-safe-and-efficient-mobile-tracing-apps-across-eu

[3] Google - Privacy-Preserving Contact Tracing - https://www.apple.com/covid19/contacttracing/

[4] EDPB: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, Aprile 2020 - https://edpb.europa.eu/our-work-tools/our-documents/usmernenia/guidelines-042020-use-location-data-and-contact-tracing_en

[5] Decentralized Privacy-Preserving Proximity Tracing - Documents https://github.com/DP-3T/documents

[6] ROBERT -- ROBust and privacy-presERving proximity Tracing protocol https://github.com/ROBERT-proximity-tracing/

[7] R. Dillet: France releases contact-tracing app StopCovid https://techcrunch.com/2020/06/02/france-releases-contact-tracing-app-stopcovid-on-android/

[8] Singapore CERT: Multiple Vulnerabilities in Bluetooth Low Energy (BLE) Devices, 6 Mar 2020 https://www.csa.gov.sg/singcert/alerts/multiple-vulnerabilities-in-bluetooth-low-energy-devices

[9] S. Biddle: The inventors of bluetooth say there could be problems using their tech for coronavirus contact tracing, The Intercept, https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/

[10] K. Pietrzak: Replay, Relay and Inverse-Sybil Attacks on Proximity Tracing Apps, 2020 Eurocrypt

[11] M. Miculan: Shaping-anonymous-contact-tracing-data https://miculan.github.io/shaping-anonymous-contact-tracing-data

[12] A. Polimeni, M. Cazzolla: Immuni, prima analisi del codice: bene, ma attenti allo scivolone privacy,. Agenda Digitale, 25 Maggio 2020 https://www.agendadigitale.eu/sanita/immuni-prima-analisi-del-codice-bene-ma-attenti-allo-scivolone-privacy/,

[13] N. Pitzolu: iOS 11: il Control Center non spegne completamente Wi-Fi/Bluetooth. Ecco perchè!, iPhoneItalia https://www.iphoneitalia.com/648944/guida-ios-11-wi-fi-bluetooth-control-center-spegne