GARANTE
PER LA PROTEZIONE
DEI DATI PERSONALI

Valutazione d'impatto sulla protezione dei dati personali presentata dal Ministero della Salute relativa ai trattamenti effettuati nell'ambito del sistema di allerta Covid-19 denominato "Immuni" - Nota sugli aspetti tecnologici [9357972]

Valutazione d'impatto sulla protezione dei dati personali presentata dal Ministero della Salute relativa ai trattamenti effettuati nell'ambito del sistema di allerta Covid-19 denominato "Immuni"

NOTA SUGLI ASPETTI TECNOLOGICI

Roma, 3 giugno 2020

PREMESSA

La Presidenza del Consiglio, il Ministro dell'innovazione e il Ministero della salute hanno formalizzato nelle scorse settimane la scelta della piattaforma software nazionale per il contrasto all'epidemia Covid-19, individuandola nella soluzione proposta dalla società Bending Spoons S.p.A.

A seguito della scelta effettuata, il Ministero della salute, ai sensi dell'art. 36, par. 5, del Regolamento (UE) 2016/679 e dell'art. 2-quinquiesdecies del d.lgs. 30 giugno 2003, n. 196 (Codice,) ha fatto pervenire il 28 maggio 2020 la valutazione d'impatto svolta ai sensi dell'art. 35 del Regolamento (UE) 2016/679 (GDPR), per essere autorizzato ad avviare il trattamento previsto dall'art. 6 del d.l. 30 aprile 2020, n. 28.

Nel presente documento, frutto del lavoro congiunto del Dipartimento tecnologie digitali e sicurezza informatica con il Dipartimento sanità e ricerca e il Dipartimento realtà pubbliche, si presentano alcuni degli aspetti tecnologici considerati ai fini della resa del parere da parte del Garante (Provvedimento[(1)] di autorizzazione n. 95 del 1° giugno 2020), facendo cenno alle caratteristiche generali dei sistemi di tracciamento dei contatti (contact tracing ed exposure notification) proposti per il contrasto alle epidemie, e analizzando più in dettaglio gli aspetti peculiari del sistema Immuni che emergono dalla documentazione disponibile.

ASPETTI RELATIVI ALLE SCELTE ARCHITETTURALI E DI MODELLAZIONE DEI DATI

Sono stati proposti in occasione dell'emergenza epidemiologica, in Italia e all'estero, diversi sistemi di tracciamento dei contatti basati su differenti tecnologie, e talvolta su più d'una tecnologia, comprendendo approcci basati su:

- capacità di localizzazione dei dispositivi di comunicazione sulle reti cellulari da parte degli operatori di rete (capacità che è implicita nella tecnologia telefonica radiomobile) anche senza la piena consapevolezza degli utenti, potendo così fare assunzioni su eventuali situazioni di prossimità tra i terminali mobili;

capacità dei dispositivi di tipo smartphone di determinare autonomamente la propria posizione tramite l'elettronica in dotazione (ricevitori dei segnali emessi dai sistemi globali di localizzazione come GPS, GLONASS, Galileo…) e di comunicarla (con protocolli di rete Internet o con tecniche telefoniche) a un centro servizi in grado di raccogliere i dati di posizione ed elaborarli al fine di individuare eventuali situazioni di prossimità tra i dispositivi;

capacità degli smartphone di rilevare autonomamente la presenza di altri dispositivi in loro prossimità tramite la ricezione di segnali di identificazione (radio beacons) diffusi con tecnologia Bluetooth Low Energy (BLE).

Mentre in alcune aree geografiche distanti dall'Unione Europea e dai suoi standard di protezione dei dati personali sono state prese in considerazione e utilizzate tutte le diverse tecnologie citate, in ambito europeo e in gran parte del mondo la ricerca di soluzioni tecnologiche per il contrasto alla diffusione della pandemia Covid-19 (nonché di altre epidemie che potrebbero insorgere in futuro) si è concentrata sul ricorso alla sola tecnologia BLE, presente in tutti i moderni smartphone e adatta a consentire il mutuo rilevamento da parte dei dispositivi che la adottano.

Relativamente ai sistemi basati sulla sola capacità di rilevamento tramite BLE sono stati proposti differenti approcci che si possono ricondurre a due principali casistiche sulla scorta del modello dei dati assunto come base dell'elaborazione delle informazioni generate e raccolte:

Data model centralizzati
Le informazioni sugli utenti con diagnosi di positività al SARS-CoV-2 convergono su una componente di backend in cui hanno luogo anche la raccolta delle informazioni sui contatti comunque acquisiti, i confronti tra gli identificativi pseudonimizzati dei dispositivi, l'individuazione degli eventuali contatti a rischio e la segnalazione con messaggio diretto ai singoli interessati.

Data model distribuiti
Il rilevamento dei contatti avviene in modo decentralizzato, direttamente tra i dispositivi smartphone in possesso degli utenti in modalità peer-to-peer, mentre l'individuazione di quelli a potenziale rischio di contagio richiede la disponibilità di informazioni sui soggetti positivi al SARS-CoV-2 raccolte e poi distribuite da un organo centrale dell'architettura.

In astratto, nessuno dei due approcci è in assoluto preferibile all'altro dal punto di vista della protezione dei dati personali, così come da altri punti di vista, perché entrambi presentano pregi e difetti per lo più complementari, e la qualità delle soluzioni basate sui due modelli è fortemente condizionata dalle specifiche loro concrete implementazioni.

Nella comunità di data protection europea il dibattito sulle soluzioni informatiche di tracciamento non ha fatto specificamente leva su questi aspetti "architetturali", e ne sono testimonianza le posizioni assunte da alcune autorità di supervisione nazionali, dallo European Data Protection Supervisor, con due documenti sul trattamento dei dati personali relativi alla salute per scopi di ricerca(2) e sulla geolocalizzazione e altri strumenti di tracciamento, e dallo European Data Protection Board, con le Linee-guida(3) del 21 aprile 2020.

È bene in tal senso precisare come, sul piano tecnico, la distinzione tra modelli centralizzati e modelli distribuiti non sia netta, perché in tutte le proposte basate sull'uno o sull'altro data model o architettura logica in realtà è sempre presente una componente centrale, che svolge un ruolo più ridotto nei sistemi "decentralizzati" e più rilevante in quelli centralizzati.

Nell'ipotesi di approccio distribuito/decentralizzato è infatti comunque necessaria la presenza di un punto centrale che nell'architettura del sistema svolga almeno una funzione di accertamento (tramite le competenti autorità sanitarie) dello stato positivo al SARS-CoV-2 degli utenti e di raccolta di identificativi o di pseudonimi dei loro dispositivi che dovranno essere messi a disposizione di tutti gli altri utenti che partecipano al sistema.

Siamo dunque, a rigore, in presenza di due alternative tecniche: una totalmente centralizzata, in cui l'accertamento dello stato di salute e la verifica degli avvenuti contatti con soggetti positivi vengono effettuati centralmente, e una semi-decentralizzata, in cui l'accertamento dello stato di salute è effettuato centralmente ma il confronto (matching) tra gli pseudonimi dei dispositivi è effettuato localmente sugli smartphone.

L'applicazione dei principi della protezione dei dati personali nelle due alternative è differente, e non è possibile individuare una soluzione che sia in assoluto preferibile, o più conforme al dettato normativo. Vi è semmai una differente modalità di realizzazione

dei princìpi di cui tener conto in fase di valutazione, con l'effetto che, se ci si sofferma unicamente su uno di essi (a esempio, la minimizzazione dei dati), si rischia di non considerare adeguatamente il riverbero che le scelte tecnologiche proiettano sugli altri princìpi (accuratezza, disponibilità, integrità).

Questa difficoltà è ben rappresentata nel parere dello European Data Protection Board del 21 aprile 2020, che chiaramente afferma che "entrambe le opzioni sono praticabili, a condizione che siano in vigore adeguate misure di sicurezza, ed entrambe comportano una serie di vantaggi e svantaggi", aggiungendo che "la fase di progettazione delle app dovrebbe sempre prevedere un esame approfondito di entrambi gli approcci, ponderandone attentamente gli effetti in termini di protezione dei dati e privacy nonché i possibili impatti sui diritti delle persone".

Si deve inoltre dare conto che dal settore della ricerca scientifica provengono segnali più concordi in favore dell'approccio centralizzato. A esempio, alcune prestigiose istituzioni di ricerca come l'INRIA(4) in Francia e la Fraunhofer-Gesellschaft(5) in Germania hanno contestato il modello decentralizzato sostenendo che l'invio di informazioni relative a tutti gli utenti positivi costituisca un rischio in sé, in quanto tali informazioni potrebbero consentire a utenti maliziosi del sistema di re-identificarli, possibilità che deve essere evitata con ogni mezzo.

Tuttavia, le varie opinioni espresse da istituzioni e singoli studiosi non sono ancora chiaro indice di una posizione scientifica definita, stante il troppo breve periodo di concentrazione degli studi su queste tematiche, con carattere di attività di urgenza, che ha lasciato poco tempo all'applicazione piena di un metodo scientifico che notoriamente richiede, oltre alla speculazione teorica, anche tempi più lunghi per verifiche, sperimentazioni e confronto di idee.

In questo scenario, un impulso molto marcato alla realizzazione di soluzioni di contact tracing è derivato dalla definizione dell'Exposure Notification Framework da parte di Apple(6) e Google(7) (ovvero i due produttori che rappresentano oltre il 90% del mercato degli smartphone con i loro sistemi operativi iOS e Android), con la messa a disposizione di un'Application Programming Interface per l'accesso, da parte degli sviluppatori di software, a funzionalità dei rispettivi sistemi operativi specificamente concepite per favorire la programmazione di applicazioni per il contrasto alla pandemia. L'annuncio di un sistema interoperabile di comunicazione basato su tecnologia BLE ha infatti accelerato lo sviluppo di app di contact tracing disponibili alla totalità degli utenti e non a due "isole" (quella degli utenti Apple e quella degli utenti Android) sino ad allora non comunicanti in modo affidabile.

IL SISTEMA IMMUNI

Caratteristiche generali

Il Sistema di allerta Covid-19 prescelto dall'Italia e oggetto della Data Protection Impact Assessment sottoposta al Garante è denominato "Immuni" e, per le sue caratteristiche, è inquadrabile nei sistemi semi-decentralizzati di exposure notification basati sul rilevamento BLE.

Il sistema è stato progettato nel periodo antecedente alla pubblicazione del Google-Apple Exposure Notification (GAEN) Framework a supporto delle applicazioni di tracciamento, ma è stato subito dopo riconvertito all'uso delle funzionalità nel frattempo rese disponibili.

La componente periferica del sistema è rappresentata da una app che può essere installata sui moderni smartphone Apple e su quelli, di diversi produttori, funzionanti con il sistema operativo Android, tramite i rispettivi store.

La componente di backend comprende una serie di sistemi server dedicati a specifiche funzioni di raccolta dati, di diffusione di pseudonimi dei soggetti positivi, di interazione con altri sistemi informativi (come il sistema Tessera Sanitaria, cui viene fatto ricorso per l'autenticazione degli operatori sanitari), di analisi dei dati.

A corredo del funzionamento dell'infrastruttura sono posti alcuni servizi di gestione del traffico inbound che si basano sull'interposizione di una Content Delivery Network (CDN) tra la platea di utilizzatori della app e il backend.

La componente periferica rappresentata dalla app Immuni è "pronta all'uso" una volta installata e accordati i permessi necessari al suo funzionamento (essenzialmente, quello di accesso al sottosistema Bluetooth, oltre alle funzionalità di rete Internet, necessarie per l'accesso al backend anche se non continuativamente), senza richiedere la registrazione dell'utente e la creazione di account con l'immissione di dati personali, se non l'indicazione della provincia di residenza.

Dal punto di vista della logica applicativa, l'uso di Immuni è anonimo, perché il sistema è privo di anagrafica e non prevede funzionalità di comunicazione tra il backend e le app che comportino la conoscenza dell'utilizzatore: il sistema nel suo complesso è infatti privo di capacità di indirizzamento di messaggi uno-a-uno (unicast) dal centro verso i devices degli utenti, e questi ultimi possono essere soltanto destinatari di comunicazioni in broadcast (a tutti gli "ascoltatori"), peraltro realizzate con la messa a disposizione su un server dei dati scaricabili con operazioni di download dai dispositivi.

Finalità del sistema

Scopo primario del sistema, e quindi finalità del trattamento di dati personali sotteso, è la messa a disposizione agli utenti di uno strumento di allertamento (alerting) relativo all'esposizione al contagio, utilizzabile in modo partecipativo, ovvero rendendo i dispositivi mutuamente rilevabili per prossimità, e prevedendo la possibilità per un utente di qualificarsi come soggetto positivo al SARS-CoV-2 secondo una procedura volontaria assistita da un operatore sanitario competente.

In questo modo la app Immuni potrà presentare un messaggio informativo sullo smartphone degli utenti che li avvisi della rilevata vicinanza a un dispositivo segnalato, facendo assumere all'utente così avvisato la qualifica di "contatto stretto" di un soggetto positivo e informandolo sulle opportune misure di profilassi da adottare.

Inoltre, alcuni dati raccolti (analytics) tramite la componente di backend di Immuni possono essere utilizzati per finalità di sanità pubblica, ricerca scientifica o statistica e per affinare i modelli di valutazione del rischio di contagio da applicare allo stesso sistema Immuni, migliorando l'accuratezza degli algoritmi utilizzati a tal fine.

Principi di funzionamento

La app Immuni, una volta attivata sullo smartphone, consente la trasmissione in modalità broadcast di identificativi pseudonimizzati effimeri, denominati RPI (Rolling Proximity Identifiers), tramite tecnologia BLE, permettendo contestualmente di ricevere gli analoghi identificativi trasmessi dai dispositivi utilizzanti la medesima app e rilevati in prossimità tramite l'uso della stessa tecnologia.

Il sistema Immuni prevede che ciascuno smartphone trasmetta in broadcast gli RPI (potendone nel contempo ricevere da eventuali altri dispositivi in sua prossimità) e che possa comunicare al sistema (con accorgimenti specifici per tutela della riservatezza) l'informazione relativa all'eventuale di positività Covid-19 del suo utilizzatore, clinicamente rilevata in tempi successivi all'installazione e all'utilizzo della app.

Gli RPI sono ottenuti con tecniche crittografiche sfruttando la capacità di calcolo interna dello smartphone, di cui rappresentano degli pseudonimi temporanei di breve durata, ma non recano contenuto informativo immediatamente riferibile all'utente o allo stesso dispositivo che li genera, né sono riconducibili ad altri identificativi tecnici a questo associati, quali i MAC address delle interfacce di rete WiFi e Bluetooth, i codici IMEI delle interfacce GSM o i numeri telefonici.

Tali pseudonimi vengono calcolati ogni dieci minuti circa a partire da una chiave di medio periodo della lunghezza di 128 bit, denominata TEK (Temporary Exposure Key), a sua volta generata al primo uso della app e, successivamente, rigenerata con frequenza giornaliera tramite un generatore pseudocasuale crittografico.

Da ogni chiave TEK possono essere ricavati con procedimento matematico fino a 144 RPI corrispondenti (la funzione utilizzata non è invertibile, per cui non è possibile calcolare la TEK da cui un certo RPI è stato ottenuto).

L'onboarding dei soggetti positivi

In caso di sua accertata positività Covid-19, l'utente della app, su invito delle autorità sanitarie competenti, può comunicare al sistema Immuni il proprio cambiamento di stato di salute, trasmettendo un pacchetto informativo contenente gli identificativi di medio periodo TEK assunti dal proprio dispositivo negli ultimi 14 giorni, unitamente a ulteriori informazioni epidemiologiche (analytics di tipo Epidemiological Info) desunte dall'utilizzo pregresso dell'app ma non comprendenti gli RPI esterni, lato sensu riferibili ad altri soggetti, rilevati nel corso del funzionamento.

La comunicazione della variazione di stato di salute deve essere confermata da un operatore sanitario autorizzato a interagire con la componente di backend del sistema Immuni, per il tramite del Sistema Tessera Sanitaria, e in contatto con l'utente che ha

ricevuto una diagnosi di positività. Per permettere il caricamento (upload) dei dati che lo riguardano, l'utente dovrà comunicare all'operatore un codice OTP di convalida generato dalla app. Al momento del caricamento dei dati, l'operatore sanitario inserirà nel sistema Immuni anche la data di inizio dei sintomi riferita dall'utente, che verrà utilizzata per filtrare tra tutte le TEK caricate quelle relative ai giorni in cui era contagioso.

Le TEK dei soggetti che hanno scelto di contribuire al sistema con la trasmissione dei propri dati vengono raccolte nel sistema di backend e messe a disposizione delle app installate sugli altri dispositivi, che potranno scaricarli interrogando periodicamente (più volte al giorno) un apposito server che fa parte dell'infrastruttura di backend.

Comunicazione di dati ulteriori a seguito di accertata positività

Contestualmente al caricamento delle proprie TEK, l'app provvederà alla trasmissione di un pacchetto di Epidemiological Info contenente una serie di informazioni tecniche e altre riferibili all'utente accertato positivo:

1) provincia di domicilio;

2) Exposure Detection Summary, ovvero informazioni sintetiche relative a tutti gli eventuali contatti a rischio rilevati negli ultimi 14 giorni attraverso il raffronto delle TEK scaricate con gli RPI memorizzati all'interno dello smartphone, che comprende:

a) numero di contatti a rischio rilevati;

b) numero di giorni trascorsi dal più recente contatto a rischio;

c) durata aggregata dei contatti a rischio (misurata in multipli di 5 minuti fino a un massimo di 30 minuti), distinta per tre intervalli di attenuazione del segnale Bluetooth (correlato alla distanza);

d) indice di rischio più elevato tra quelli relativi ai contatti a rischio;

3) Exposure Info, ovvero un set di informazioni relative a ciascun contatto a rischio avvenuto negli ultimi 14 giorni (rilevato attraverso il raffronto delle TEK scaricate con gli RPI memorizzati all'interno del dispositivo), che comprende:

a) data del contatto;

b) durata del contatto (misurata in multipli di 5 minuti fino a un massimo di 30 minuti);

c) l'attenuazione del segnale Bluetooth, registrandone il minimo valore nell'intervallo di osservazione;

d) la somma delle durate delle esposizioni, raggruppate in tre intervalli di attenuazione;

e) il rischio di contagiosità associato alla TEK (al momento non utilizzato);

f) l'indice complessivo di rischio associato al contatto.

Tali informazioni, che non possono essere considerate quali dati anonimi potendo consentire, in certi contesti, a fronte di tassi di morbilità bassi come quelli auspicati per il prossimo futuro, la reidentificazione dei soggetti positivi in ambiti territoriali a bassa densità di popolazione, devono essere accuratamente protette nel backend del sistema per evitare ogni forma di loro riassociazione a interessati identificati o identificabili.

Il rilevamento e la notificazione dell'esposizione a rischio di contagio

Per quanto riguarda le modalità di rilevamento dell'esposizione a contagio, il sistema Immuni fa ricorso a un approccio decentralizzato, in cui la verifica a posteriori dell'avvenuta prossimità a soggetti risultati positivi è basata su una procedura di confronto (matching) svolta all'interno dello smartphone.

Il confronto è possibile perché la app memorizza in un'area crittograficamente protetta del filesystem dello smartphone tutti gli RPI

acquisiti tramite BLE nel corso del suo funzionamento (in una finestra di osservazione di 14 giorni) interagendo con gli smartphone rilevati in sua prossimità.

Gli RPI archiviati vengono automaticamente raffrontati con quelli ottenuti con procedimento matematico dalle TEK acquisite con le periodiche operazioni di download da un server del backend, ricorrendo a funzionalità di sistema messa a disposizione dal GAEN Framework e che tengono conto di parametri quali la durata del contatto e la distanza dai dispositivi rilevati (stimata mediante i valori di attenuazione del segnale Bluetooth), consentendo così di calcolare un indice complessivo di rischio (Total Risk Score).

Qualora un RPI archiviato localmente corrisponda a uno degli RPI scaricati e quest'ultimo abbia associato un Total Risk Score superiore a una soglia predefinita, l'app presenta all'utente una notifica di esposizione a rischio (Exposure Notification), realizzando così la primaria finalità del sistema, cioè quella di allertare sull'avvenuta esposizione al rischio di contagio.

Comunicazione di dati a seguito di Exposure Notification

Il sistema Immuni, nella sua componente di backend, non acquisisce conoscenza dell'identità degli utenti che sono stati esposti al rischio di contagio, ma fa comunque uso di informazioni trasmesse dai loro smartphone a seguito della notificazione (Exposure Notification Summary).

Al momento del rilevamento di una esposizione a rischio, infatti, oltre alla notificazione presentata sullo schermo dello smartphone, può essere trasmesso al backend un pacchetto informativo (analytics) denominato Operational Info with Exposure, utile all'elaborazione di statistiche ma che, una volta pervenuto nel backend, non sarà riferibile all'utente del dispositivo né al dispositivo che lo ha trasmesso.

Le Operational Info with Exposure comprendono l'indicazione della provincia di domicilio fornita dall'utente al momento della configurazione iniziale di Immuni, e poi modificabile in qualunque momento: tale informazione è raccolta al fine di permettere il monitoraggio statistico della ricezione degli alert sul territorio nazionale e di consentire l'individuazione tempestiva di eventuali nuovi focolai epidemici, potendo così interessare le autorità sanitarie locali e programmare l'allocazione di risorse (farmaci, dispositivi di protezione, personale…).

Eventuali molteplici contatti stretti rilevati nello stesso mese da uno smartphone comportano sempre la notificazione all'utente ma non producono alcuna trasmissione di dati al sistema di backend se già ne è stata effettuata una nello stesso mese.

Allo scopo di evitare effetti di single-out in presenza di un basso tasso di contagio o in popolazioni sparse e numericamente ridotte, in cui la mera trasmissione delle Operational Info with Exposure, se osservata da terzi con strumenti di analisi del traffico di rete, potrebbe essere in sé indizio del potenziale contagio di un interessato, il sistema Immuni prevede il ricorso alla trasmissione, anche in assenza di notificazione, di Operational Info without Exposure che hanno la funzione di introdurre una sorta di rumore di fondo nelle comunicazioni che non permetta di attribuire necessariamente un univoco significato ai pacchetti osservati (ancorché cifrati) in transito su una rete.

I pacchetti Without Exposure sono del tutto simili a quelli recanti l'informazione sull'avvenuta esposizione al rischio, ma recano soltanto quei contenuti informativi relativi al funzionamento della app presenti anche nei pacchetti With Exposure (stato di attivazione delle notifiche, stato di attivazione dell'interfaccia BLE).

La frequenza d'inoltro al backend degli analytics risulta così ben calibrata e idonea a mascherare efficacemente il traffico generato dai "contatti stretti" di un soggetto positivo, senza provocare un eccessivo sovraccarico del sistema.

Inoltre, è coerente con questo schema di mascheramento il previsto limite massimo di due pacchetti Operational Info al mese, per prevenire eventuali attacchi informatici da parte di soggetti malintenzionati che potrebbero sovraccaricare il sistema con dati spurii artificiosamente generati al fine di inquinamento della base dati, mantenendo comunque l'utilizzabilità statistica dei dati al fine di controllare l'adozione e l'uso della app e poterne stimare l'efficacia.

POTENZIALI CRITICITÀ

CRITICITÀ INTRINSECHE NEL MODELLO DECENTRALIZZATO

Come evidenziato da studiosi e istituti di ricerca in Europa e nel mondo[8], ai diversi pregi del modello decentralizzato di gestione dei dati si affiancano alcune debolezze intrinseche di cui occorre essere consapevoli anche al fine di adottare le opportune misure di mitigazione dei rischi relativi alla possibile reidentificazione degli utenti che abbiano mutato il proprio stato di salute, diventando positivi al Covid-19, conferendo al sistema i propri identificatori TEK (o Diagnosis Keys).

La riservatezza dei dati relativi agli utenti malati di Covid-19 è affidata in parte alla tecnologia ma, in parte prevalente, come sarà più chiaro da alcuni esempi di scenari critici, alla capacità di limitare al massimo le occasioni in cui gli pseudonimi di breve periodo RPI, diffusi in radiofrequenza con tecnologia BLE, possano essere rilevati da soggetti estranei al sistema Immuni in abbinamento temporale all'acquisizione di informazioni identificative, per poi essere, eventualmente, confrontati con gli pseudonimi ottenuti con procedimento matematico applicabile da chiunque alle Diagnosis Keys messe a disposizione in rete da Immuni così come da ogni sistema "decentralizzato" di exposure notification.

Particolare cura va riservata alla sicurezza del dispositivo smartphone, per prevenire l'azione di malware o di ogni altra app apparentemente innocua ma che potrebbe avere un comportamento malizioso (con o senza la consapevolezza dell'utente) acquisendo, per esempio, gli stessi dati utilizzati dalla app Immuni per poi trasmetterli all'esterno, alimentando così raccolte su larga scala di pseudonimi RPI che consentano la ricerca di corrispondenze utili a ricostruire, per esempio, spostamenti delle persone oppure a individuare gli pseudonimi dei soggetti positivi.

Analogamente, possono essere attivi dei dispositivi fisici o programmi software di scansione (sniffer) in grado di intercettare la trasmissione degli pseudonimi all'esterno del dispositivo, che avviene in radiofrequenza in modalità broadcast, ascoltabile da chiunque utilizzi l'idonea tecnologia.

Nel caso del ricorso a dispositivi elettronici specializzati, potrebbero anche essere superati gli ordinari limiti spaziali del rilevamento, dovuti alla scarsa potenza dell'emissione radio BLE, consentendo di captare gli identificativi RPI trasmessi anche in un'area vasta con l'utilizzazione di un solo "punto di ascolto" nascosto.

A questi attacchi di re-identificazione, basati sulla ricerca di corrispondenze tra gli identificativi pseudonimizzati impiegati nel sistema Immuni e altri identificativi raccolti con strumenti esterni al sistema, con l'ausilio di risorse tecnologiche ad hoc e sulla base di una motivazione specifica da parte del soggetto che a questo fine si adopera, si aggiungono tutti i casi di re-identificazione "inferenziale": questa prescinde dall'utilizzo di tecnologie e ciascun soggetto che riceverà un alert di esposizione a rischio, al pari di chi – al di fuori del sistema Immuni – riceva una diagnosi di positività, sarà portato a una ricostruzione anamnestica dei pregressi contatti intercorsi con altre persone a cui possibilmente attribuire l'avvenuta esposizione al rischio di contagio o addirittura lo stato di malattia, analizzando retrospettivamente i contatti più prolungati avuti nella finestra di osservazione attivata per la app (pari a 14 giorni).

Tale associazione potrà risultare in molti casi assai verosimile, anche tenuto conto della motivazione che chi riceverà un alert potrà mostrare in questo esercizio di re-identificazione, della perfetta conoscenza dei propri movimenti più recenti e del contesto in cui si è manifestato il contatto, senza che per effettuare la re-identificazione siano necessari sofisticati strumenti tecnologici. Giova dire che tali possibilità di re-identificazione "inferenziale" si applicherebbero anche a qualsiasi sistema centralizzato o distribuito, costituendo una criticità intrinseca della exposure notification.

In definitiva, qualsiasi soluzione di exposure notification è potenzialmente vulnerabile, e talora significativamente vulnerabile, ad attacchi di reidentificazione di tipo inferenziale, mentre le soluzioni di tipo decentralizzato sono ulteriormente esposte al rischio di attacchi basati sulla ricerca di corrispondenze.

Paparazzi attack

I sistemi decentralizzati, che comportano la necessità di diffondere gli pseudonimi dei soggetti positivi, espongono tali pseudonimi, in certi contesti, a un particolare tipo di attacco basato sulla ricerca di corrispondenze detto Paparazzi attack, che si può mettere in atto quando sia possibile acquisire, più o meno agevolmente, lo pseudonimo di un soggetto la cui identità sia nota o conoscibile.

Per esempio, potrebbe essere agevole captare i radio beacon RPI in prossimità del luogo di dimora o di lavoro della persona oggetto di specifica attenzione, oppure in ogni altro luogo in cui all'emissione in Bluetooth dello pseudonimo sia associabile un'identificazione personale. Situazioni del genere possono prodursi in esercizi commerciali all'atto di un pagamento con carta di credito, oppure al passaggio attraverso varchi di imbarco controllati negli aeroporti, o nei luoghi di lavoro con i sistemi di

rilevamento delle presenze o con sistemi di monitoraggio del traffico stradale accoppiati a impianti di ripresa video.

Si tratta quindi di quei contesti in cui un attaccante possa essere in grado di acquisire lo pseudonimo generato dal dispositivo target, a insaputa del suo possessore, essendo poi anche in grado di "etichettarlo" con informazioni direttamente identificative.

Scopo di questa raccolta ed "etichettatura" è la creazione di una base dati di pseudonimi che corrispondano a persone fisiche conosciute o conoscibili (al limite, una sola persona il cui stato di salute rivesta particolare interesse o abbia un valore in sé) e la ricerca in questa base dati delle eventuali corrispondenze con gli pseudonimi dei soggetti positivi resi pubblici dai sistemi "decentralizzati" di exposure notification.

Nel caso di Immuni, la diffusione delle chiavi TEK dei soggetti positivi e la pubblicazione del codice del programma come open source, nonché la pubblicità degli algoritmi crittografici adoperati nel GAEN Framework, consentono a chi abbia scaricato le TEK dei soggetti positivi, anche fuori dal sistema Immuni, di calcolare gli RPI associati a ciascuna di esse in un arco temporale di 24 ore e di utilizzare i risultati per la ricerca di corrispondenze nella base dati precedentemente formata con l'acquisizione occulta degli RPI "etichettati", pervenendo così all'individuazione dei soggetti positivi.

In proposito si osserva come l'algoritmo di generazione degli RPI sia pubblico[(9)], nonché basato su funzioni crittografiche anch'esse pubbliche, come la HKDF (definita in IETF RFC 5869), come la funzione di hash SHA-256 o come la funzione crittografica AES-128.

Il procedimento di calcolo degli RPI è descritto nelle Exposure Notification Cryptography Specification di Apple e Google come:

$$RPI_{i,j} \leftarrow AES128(RPIK_i, PaddedData_j)$$

in cui RPIK è una Rolling Proximity Identifier Key "matrice" degli RPI, la variabile j è il tempo misurato in secondi da Epoch (riferimento dei clock dei sistemi Unix storicamente fissato al 1° gennaio 1970) in cui lo RPI viene generato, PaddedData sono informazioni contenenti lo ENIntervalNumber utilizzato per sincronizzare la generazione degli RPI tra tutte le app partecipanti al sistema.

A mitigare questo specifico rischio vale la considerazione, in generale, della scarsa utilità dell'attacco e della carenza di aspetto motivazionale: resta tuttavia la possibilità (cui ci richiama la realtà dello scenario della cybersecurity e l'osservazione delle violazioni che da anni si susseguono in danno di sistemi informatici pubblici e privati anche senza una comprensibile finalità utilitaristica, essendo talvolta frutto di azioni dimostrative non necessariamente maliziose) che tali condotte vengano effettivamente messe in atto a scopi di disturbo, di contestazione, oppure per acquisire informazioni sullo stato di salute di soggetti relativamente ai quali la conoscenza di tali notizie abbia un valore in sé, per le implicazioni che ciò può comportare e per i vantaggi che può procurare.

Resta fermo che l'eventuale acquisizione occulta e l'utilizzo successivo di dati, da chiunque effettuata all'esterno del perimetro del sistema Immuni, configurerebbero una condotta illecita perseguibile legalmente.

MIGLIORIE TECNICHE SPECIFICHE DEL SISTEMA IMMUNI

Relativamente alle caratteristiche specifiche del sistema Immuni, per quanto descritto nel documento di Data Protection Impact Assessment e dai documenti tecnici a esso allegati, si rilevano alcuni aspetti tecnici meritori di chiarimento e possibilmente oggetto di interventi migliorativi dal punto di vista della protezione dei dati e della sicurezza informatica.

Conservazione dei dati di traffico IP

Non è chiaro se vengano conservati, da Sogei per conto del Ministero della salute, gli indirizzi IP dei dispositivi mobili che interagiscono con il backend nella fase di upload delle TEK e delle Epidemiological Info a seguito di accertata positività dell'utilizzatore della app.

Nel questionario di self-assessment rispetto ai requisiti stabiliti dall'EDPB, al punto SEC-5 è scritto "Non è prevista la memorizzazione degli indirizzi IP dei client da parte del server di backend centrale. L'indirizzo IP viene conservato dall'infrastruttura perimetrale ai soli fini di garantirne la sicurezza informatica", mentre nella DPIA è scritto che "L'indirizzo IP del dispositivo che invia i dati viene trasformato in un indirizzo fittizio attraverso tecniche di Network Address Translation (NAT) dall'infrastruttura di backend

all'atto dell'upload dei dati e non viene memorizzato né nel database né nei file di log. L'indirizzo IP viene esclusivamente tracciato temporaneamente sui sistemi perimetrali di accesso degli upload".

Occorre quindi verificare l'effettiva necessità, in questa forma, di tali raccolte, genericamente motivate con esigenze di sicurezza della piattaforma.
La motivazione è fuori discussione, ma i tempi di conservazione dei dati devono essere ricondotti allo strettamente necessario per il rilevamento di anomalie e di attacchi informatici e limitati al breve periodo.

Device attestation

Al fine di consentire al backend di Immuni di verificare l'autenticità dei dispositivi dai quali provengono i dati analytics di tipo Operational Info, per i soli dispositivi con sistema operativo iOS è previsto che vengano effettuate le seguenti operazioni:

- l'app richiede ad Apple ("DeviceCheck iOS API") l'attribuzione di un identificativo temporaneo del dispositivo, denominato device token, che consentirà al backend di Immuni di verificarne l'autenticità;

- successivamente, l'app genera, in modo casuale, un altro identificativo del dispositivo, denominato analytics token, salvandolo localmente e inviandolo al backend unitamente al device token attribuito da Apple;

- alla ricezione di tali dati, il backend di Immuni verifica con Apple ("DeviceCheck server API") la validità del device token relativo al dispositivo dell'utente; in tale circostanza, il backend si avvale anche di funzionalità rese disponibili da Apple (c.d. "DeviceCheck per-device bits") che consentono di tenere traccia di quei dispositivi mobili che, avendo assunto un comportamento anomalo nella generazione dell'analytics token, non sono autorizzati a inviare Operational Info;

- in caso di riscontro positivo da parte del servizio di Apple, il backend memorizza l'analytics token in un database, associandolo a un contatore di invii;

- ogni qual volta l'app deve inviare gli analytics di tipo Operational Info, assieme a tali dati viene trasmesso l'analytics token generato in precedenza;

- alla ricezione di tali analytics, il backend controlla se l'analytics token esiste, se è stato generato e se non è già stato utilizzato per effettuare due invii; solo se tutte queste condizioni sono soddisfatte, le Operational Info vengono accettate e salvate nel backend; in caso contrario, i dati vengono scartati.

L'analytics token cambia con cadenza mensile e viene inviato al backend al massimo tre volte al mese (all'atto della generazione, dell'invio delle Operational Info with Exposure e dell'invio delle Operational Info without Exposure), con lo scopo dichiarato di limitare "la capacità del server di reidentificare lo stesso dispositivo a cavallo di più chiamate al server".

Benché l'interposizione di Apple, che viene interrogata con l'invocazione delle funzionalità DeviceCheck iOS API, comporti un contatto tra il dispositivo dell'utente e il fornitore della tecnologia di base (smartphone, sistema operativo), occorre considerare come tale controllo, da una parte, miri ad assicurare una migliore accuratezza dei dati raccolti per finalità statistiche sanitarie e di miglioramento del sistema di allerta Covid-19, impedendo l'inquinamento della base dati con l'immissione, anche massiva e intenzionale, di dati spurii allo scopo di recare pregiudizio al funzionamento del sistema.

D'altra parte, il valore informativo per Apple derivante dall'utilizzo delle funzionalità di DeviceCheck iOS API è oggettivamente modesto, apparendo limitato, eventualmente, alla mera stima quantitativa di un campione delle app che risultano attive: tutto ciò, a fronte del ruolo che la società americana riveste quale fornitrice della piattaforma di distribuzione delle app e dei loro aggiornamenti, che consente la raccolta di più dettagliate informazioni relative all'intero parco applicativo installato: tutti gli utenti, come per qualsiasi altra app destinata a dispositivi Apple, per installare Immuni dovranno necessariamente interagire con lo app store della Società, attività che comporta la "registrazione" dell'utente con, in molti casi, l'inserimento dei dati relativi ad almeno un sistema di pagamento.

Interposizione di una Content Delivery Network

A scopi di gestione del traffico di rete e del carico dei sistemi di backend, è previsto il ricorso a un servizio di Content Delivery

Network (CDN) per la pubblicazione dei TEK Chunks che tutte le app dovranno frequentemente scaricare nel corso del loro funzionamento.

Da una parte, il ricorso a una CDN non appare comportare specifiche criticità in ordine alla riservatezza, perché serve per facilitare la diffusione di dati (le TEK o Diagnosis Keys dei soggetti positivi) che, pur essendo sempre dati personali ancorché pseudonimizzati, nell'architettura del sistema sono comunque precipuamente destinati a essere "accessibili da chiunque" tramite la app ma, per ciò stesso e ultra petita, anche all'esterno del sistema Immuni, tramite connessioni di rete ai server del backend di cui la CDN è mera propaggine tecnica.

Nella CDN possono però prodursi dati di traffico relativi alle connessioni con protocolli di rete, che sono anch'essi dati personali riferibili alla platea complessiva di utilizzatori attivi della app, il cui trattamento è necessario per la resa dei servizi ma la cui conservazione deve essere limitata al minimo indispensabile per permettere, oltre alla resa del servizio, una ragionevole capacità di rilevamento di anomalie e attacchi informatici all'infrastruttura.

Controllo degli accessi ai dati

I controlli previsti relativamente ai dati conservati nella componente centrale dell'infrastruttura appaiono ispirati al Provvedimento generale del Garante sugli amministratori di sistema, che era però rivolto alla generalità dei titolari di trattamento con strumenti elettronici.

Le misure di audit e di raccolta dei log andrebbero pertanto estese alle operazioni compiute sui dati da tutti gli operatori che interverranno nel loro trattamento nel backend, e non solo sugli "accessi" (eventi di autenticazione informatica), a iniziare dalla delicata fase di raccolta delle informazioni sui casi positivi.

Durata della conservazione dei dati

Mentre per i dati propri del sistema Immuni è prevista un perentorio expire time, con relativa cancellazione, allo scadere del 31 dicembre 2020, rimane meno chiaro il trattamento degli usage data relativi ai contatti con protocolli TCP/IP tra le app e i sistemi di backend, oltre alla già osservata criticità relativa alla CDN.

Si tratta comunque di aspetti di facile gestione, sui quali il Titolare del trattamento potrà agevolmente intervenire per garantire che le raccolte di dati siano strettamente finalizzate alla gestione degli incidenti di sicurezza o alla loro prevenzione, nei limiti di una breve conservazione.

Conclusioni

In definitiva, la soluzione Immuni appare un'implementazione di un sistema di exposure notification basata su un data model distribuito e su un'architettura semi-decentralizzata in cui le funzionalità centrali sono relative all'accertamento della positività Covid-19, alla diffusione di pseudonimi degli smartphone in uso agli utenti positivi, alla raccolta in forma aggregata e all'elaborazione di dati statistici per finalità di salute pubblica, di gestione delle risorse e di miglioramento dell'accuratezza degli algoritmi rilevanti ai fini dell'exposure notification.

Le criticità analizzate appaiono affrontabili con opportuni accorgimenti tecnici e organizzativi di cui sarà necessario verificare l'adozione e l'efficacia durante l'esercizio del sistema, mentre rimangono alcuni aspetti di potenziale rischio derivanti da caratteristiche intrinseche del data model distribuito e dell'architettura decentralizzata, e non da specificità del sistema tecnologico approntato (reidentificazione di soggetti positivi tramite ricerca di corrispondenze, a seguito di paparazzi attack), comuni a tutte le esperienze di contact tracing per il contrasto alla pandemia basate sul medesimo modello decentralizzato in corso di svolgimento in diversi Paesi.

Il dirigente del Dipartimento tecnologie digitali e sicurezza informatica

---

1. Garante per la protezione dei dati personali, Provvedimento di autorizzazione al trattamento dei dati personali effettuato

attraverso il Sistema di allerta Covid-19 - App Immuni: https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9356568

2. Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak
https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032020-processing-data-concerning-health-purpose_en

3. Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, Adopted on 21 April 2020: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf

4. «Contact tracing»: Bruno Sportisse, PDG d'Inria, donne quelques éléments pour mieux comprendre les enjeux, https://www.inria.fr/fr/contact-tracing-bruno-sportisse-pdg-dinria-donne-quelques-elements-pour-mieux-comprendre-les-enjeux

5. Proximity tracing in the context of corona: Fraunhofer's approach for Germany. The situation and technical implementation, https://www.fraunhofer.de/content/dam/zv/en/press-media/2020/april/fraunhofer-paper-proximity-tracing-in-the-context-of-corona-fraunhofers-approach-for-germany.pdf

6. Framework ExposureNotification – Implement a COVID-19 exposure notification system that protects user privacy, https://developer.apple.com/documentation/exposurenotification

7. Exposure Notifications: Using technology to help public health authorities fight COVID 19, https://www.google.com/covid19/exposurenotifications/

8. F. Baiardi, Le proprietà di un sistema per il tracing: centralizzato, decentralizzato, open source, https://www.riskmanagement360.it/analisti-ed-esperti/le-proprieta-di-un-sistema-per-il-tracing-centralizzato-decentralizzato-open-source/

9. Apple–Google, Exposure Notification Cryptography Specification – Preliminary — Subject to Modification and Extension, April 2020 v1.2,
https://blog.google/documents/69/Exposure_Notification_-_Cryptography_Specification_v1.2.1.pdf