UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
ED ELETTRICA E MATEMATICA APPLICATA

FACOLTÀ DI INGEGNERIA INFORMATICA

CORSO DI CYBERSECURITY

Relazione di progetto

DPPCT Expanded

Gruppo n° 37

Anno Accademico:
2019 / 2020

Montillo Andrea	0622700844
Sammarco Enrico	0622700857
Sicignano Andrea	0622700859
Verdoliva Enrico	0622700874

1. Introduzione

Questo documento propone un sistema sicuro per il tracciamento decentralizzato dei contatti tra le persone, noto anche come “Decentralized Privacy-Preserving Proximity Tracing” (DP3T).

Al fine di rallentare la diffusione del virus SARS-CoV-2, si vuole semplificare e accelerare il processo di notifica a tutti quegli individui, che sono stati a contatto con una persona risultata positiva. La progettazione del sistema mira a garantire la protezione dei dati di ogni singolo cittadino.

L'obiettivo consiste nel determinare chi è stato a stretto contatto fisico con un infetto, in modo da segnalare al cittadino a rischio che è in pericolo, senza però svelare le identità o il luogo di tale contagio.

Si assume innanzitutto che ogni cittadino abbia un proprio smartphone e che esso supporti la modalità di comunicazione Bluetooth Low Energy. Ciascun individuo possiede uno pseudonimo, che, inoltrato in broadcast via bluetooth, non rivela esplicitamente l’identità della persona.
Infatti su ogni smartphone è installata un’app, il cui compito è quello di trasmettere l’identificativo pseudo-casuale effimero del possessore del telefono e salvare anche tutti quegli identificativi effimeri degli individui osservati nelle vicinanze.

Ogni cittadino può effettuare richiesta al laboratorio di analisi per fare il tampone. Qualora l’individuo risulti positivo al virus, vengono caricati alcuni suoi dati anonimi (solo se vi è il consenso esplicito da parte del cittadino) su un server centrale, gestito dal governo. Un sistema proxy fa da tramite nella comunicazione tra l’infetto e il governo in modo da preservare ulteriormente la privacy.
Il server rivela alle app, installate sugli smartphone dei vari individui, i dati anonimi delle persone infette. In questo modo un utente capisce se si è trovato in prossimità fisica di una persona infetta e in quel caso la sua app provvederà a calcolare il rischio di un ipotetico contagio.

Inoltre, il sistema consente agli utenti di fornire volontariamente informazioni al laboratorio di epidemiologia, che si occupa di studiare e analizzare l’evoluzione e la diffusione del virus. In questa comunicazione è stata introdotta una seconda entità proxy, in modo da evitare la localizzazione tramite indirizzi IP di questi cittadini infetti.
Altro compito del laboratorio di epidemiologia è quello di stilare una classifica in modo da individuare i cittadini maggiormente a rischio per poter adottare misure di prevenzione.

Il sistema è infine in grado di gestire anche i cosiddetti “cittadini recidivi”, ovvero quegli individui che, dopo essere guariti dal virus, hanno riscontrato nuovamente i sintomi da Covid-19 e sono risultati ancora una volta positivi (si presuppone che stavolta siano stati infettati da un altro ceppo del virus).
E’ stata prevista questa funzionalità, in quanto è utile per un futuro in cui, se il virus persiste, ci sarà un accavallamento rilevante tra cittadini infetti normali e recidivi.

Di seguito riportiamo la proposta di progetto precedentemente presentata.

==========

Titolo Progetto: DPPCT Expanded

==========

Tema del progetto:
Realizzare un sistema di tracciamento dei contatti tra le persone
mediante pseudonimi su smartphone con tecnologia bluetooth low energy, con
decentralizzazione dei dati resi disponibili per studi epidemiologici.

==========

Entità coinvolte:

Cittadino non infetto: è un cittadino al quale al momento non è arrivata
comunicazione di essere infetto.

Cittadino infetto: è un cittadino al quale è già arrivata la comunicazione
di essere infetto.

Cittadino ad alto rischio: è un cittadino che, in base ai calcoli dell’applicazione, possiede un valore di rischio sopra una certa soglia.

Cittadino recidivo: è un cittadino che risulta essere nuovamente positivo.

Laboratorio analisi: è il posto in cui può recarsi un cittadino per
effettuare dei test e verificare la positività al virus.

Laboratorio di epidemiologia: è un’entità che raccoglie informazioni sulla popolazione in modo da studiare l’andamento della pandemia nel paese.

Proxy: sistema che fa da tramite nella comunicazione tra il cittadino e il governo in modo da preservare ulteriormente la privacy del cittadino

Governo: gestisce un server che permette di coordinare le informazioni collezionate dagli smartphone. Coordina anche i laboratori per assistere i pazienti nelle operazioni di comunicazione della loro positività al sistema.

==========

Obiettivi del progetto:
Permettere il calcolo di un fattore di rischio di un cittadino sulla base dei contatti con persone infette, tramite il quale l'utente potrebbe ottenere delle limitate agevolazioni.

Si vuole tutelare la privacy evitando:

- di rendere nota la corrispondenza tra cittadino o il suo indirizzo IP e i dati nel sistema a lui correlati.
- che i contatti avvenuti siano visibili a chi non ha una stretta necessità di venirne a conoscenza.

È prevista la raccolta di dati anonimi, che racchiudono informazioni sui contatti avvenuti con persone infette o recidive, al fine di effettuare studi epidemiologici riguardanti l'andamento della pandemia e la possibile mutazione del virus.

==========

Task previsti:

1) Descrizione dettagliata (ma non completamente formale) delle proprietà di completezza, sicurezza e privacy desiderate.

2) Progettazione di un protocollo per la comunicazione tra smartphone e di un un protocollo per la comunicazione tra smartphone e server attraverso il Proxy.
Progettazione di un protocollo per la comunicazione dell'avvenuta infezione evitando il problema di falsi positivi.
Progettazione del protocollo di comunicazione tra smartphone ed epidemiologi per la trasmissione dei dati utili.

3) Analisi euristica ma sufficientemente ricca di dettagli che validi la bontà della progettazione rispetto alle proprietà elencate nel punto 1).

4) Implementazione del protocollo di comunicazione tra il cittadino infetto, il laboratorio di analisi e il server del Governo mediante il Proxy.

2. Obiettivi di sistema e Requisiti funzionali

Il sistema consente agli utenti infetti di poter condividere i propri pseudonimi in maniera anonima e permette agli altri utenti del sistema di essere avvisati qualora fossero delle persone a rischio. In particolare, viene realizzato un calcolo della percentuale di rischio di ciascun cittadino, in base ai contatti che ha avuto con persone infette e alla durata del loro incontro. L'obiettivo dell'app è quindi evitare che utenti asintomatici diffondano inconsapevolmente la malattia.

Inoltre, l'applicazione offre agli utenti la possibilità di condividere volontariamente con il laboratorio di epidemiologia i propri dati anonimi relativi ai contatti avvenuti con persone infette o recidive. In questo modo gli studiosi possono analizzare la diffusione della pandemia, la possibile mutazione del virus e possono costruire un grafo di interazione tra utenti infetti e a rischio.
L'applicazione non si interessa di individuare "focolai" (luoghi specifici con alta concentrazione di infetti). Dunque evita di raccogliere dati sensibili sulla posizione, in modo da preservare la privacy.

Il sistema si propone di gestire le seguenti proprietà di confidenzialità ed integrità, che in seguito verranno analizzate come soggette ad attacchi di possibili avversari..

- Per tutelare la privacy delle persone si evita di rendere nota la corrispondenza tra il cittadino, il suo indirizzo IP e i dati nel sistema a lui correlati. Così facendo, il sistema raccoglie, archivia e utilizza dati anonimi, che non sono collegabili all'identità di una persona.

- Minimizzazione dei dati e prevenzione dell'abuso di dati: ciascuna entità autorizzata apprende solo le informazioni minime e strettamente necessarie. In questo modo si fornisce ai cittadini la sicurezza e la fiducia che i loro dati sensibili siano protetti e usati in modo accurato.
  Ad esempio, il server centrale osserva solo gli id anonimi delle persone infette; il laboratorio di epidemiologia, se l'utente accetta la condivisione dei dati, ottiene le informazioni minime relative solo ai contatti con persone infette.

- Impedisce il tracciamento degli utenti. Nessuna entità, può tracciare gli utenti. Per i cittadini positivi è previsto un nuovo identificativo, in modo tale da non avere nel sistema ID effimeri direttamente collegabili a persone infette.

Le analisi, che sono state condotte in termini di sicurezza e privacy, hanno tenuto conto di ipotetici attacchi, realizzati da diverse tipologie di avversari, ovvero:

• Utente standard, che una volta installata l'applicazione, tramite la sua interfaccia grafica, tenta di dedurre informazioni private di altri utenti.

➢ Fallisce, non deve essere possibile ricavare informazioni sugli altri utenti in base ai dati scambiati tra i dispositivi mobili e memorizzati dall'applicazione.

• Utente esperto, che è in grado di modificare il codice sorgente dell'applicazione.

➢ Un utente in grado di modificare l'applicazione potrebbe riuscire a creare false informazioni nel sistema, ma non deve riuscire a ottenere info sensibili su altri utenti né tantomeno riuscire a creare falsi positivi nel sistema.

• Intercettatore, che osservando il traffico di comunicazione sulla rete, può determinare se un utente è a rischio, infetto o recidivo.

➢ Non deve essere possibile a questa tipologia di utente capire il contenuto dei messaggi inoltrati sulla rete e dunque determinare lo stato degli altri utenti.

• Laboratorio di analisi, diagnostica la patologia ai cittadini e funge da garante alle informazioni inviate al server del governo dai cittadini.

➢ Non deve essere in grado di associare le identità dei cittadini con i loro identificativi nel sistema.

• Laboratorio di epidemiologia, analizza i dati sulle interazioni tra utenti infetti e a rischio; è principalmente interessato al grafo di primo livello.

➢ Non deve essere in grado di ottenere informazioni aggiuntive a quelle condivise dai cittadini come ad esempio l'indirizzo IP o informazioni geografiche o anagrafiche.

• Governo, può accedere a tutti i dati memorizzati sul server centrale, conosce gli identificativi effimeri delle persone infette.

➢ Non deve essere in grado di ottenere informazioni aggiuntive a quelle condivise dai cittadini come ad esempio l'indirizzo IP o informazioni geografiche o anagrafiche.

• Gestore del Proxy server, funge da tramite nelle comunicazioni sensibili tra i cittadini e il laboratorio di epidemiologia o il server del governo.

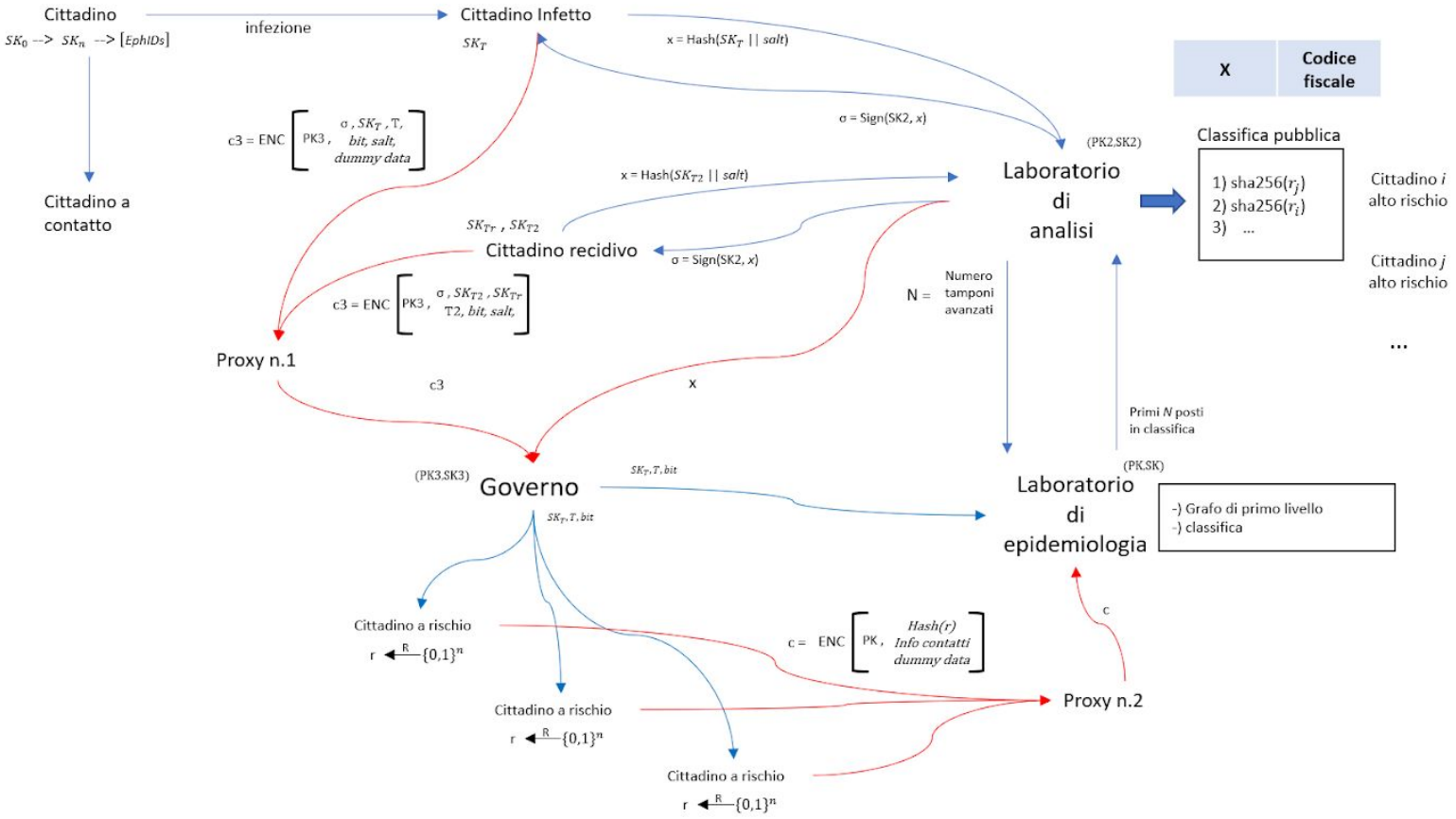
➢ Non deve essere in grado di associare un indirizzo IP alle informazioni sensibili contenute nei pacchetti, né deve essere in grado di capire lo stato di salute del cittadino con cui avviene la comunicazione.

In particolare, per queste ultime quattro entità, si assume che non collaborino tra loro scambiandosi informazioni al fine di intaccare la privacy degli utenti. Con questa assunzione è possibile mantenere la proprietà di confidenzialità precedentemente descritta.

3. Progettazione del sistema

Il sistema proposto è caratterizzato da diverse soluzioni a problemi specifici, per questo motivo risulta comodo vedere il sistema come composizione di macroblocchi ognuno dei quali descrive aspetti delle relazioni tra una o più entità citate precedentemente. Passiamo quindi all'analisi di ogni macroblocco, ai fini di leggibilità ci si riferirà ad ogni macroblocco con le relazioni tra le entità coinvolte nello stesso.

I dettagli di ogni schema di crittografia o algoritmo utilizzato verranno invece illustrati nella sezione 3.1.



Cittadino - cittadino

Per quanto riguarda la relazione tra cittadini nel sistema, ovvero la modalità attraverso la quale i cittadini calcolano e si scambiano identificativi effimeri partendo dalla propria chiave segreta del giorno, si fa esplicito riferimento al white paper del protocollo DP3T che si trova al seguente link:

https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf

Cittadino infetto (o recidivo) - Laboratorio di analisi - Governo

In questo paragrafo si analizzeranno le comunicazioni che avvengono tra il Cittadino infetto , il Laboratorio di analisi e il Governo. Le comunicazioni tra l'app del Cittadino infetto ed il server del Governo avvengono mediante un Proxy, un'entità scollegata dalle altre e considerata trusted nell'inoltro dei pacchetti, in modo che il Governo non abbia alcuna informazione riguardante la provenienza del pacchetto ricevuto (ad esempio indirizzo IP, indirizzo MAC…). Consci del fatto che non basta un singolo server Proxy per nascondere un indirizzo IP/MAC da eventuali avversari esperti nel settore, l'intento è quello di offrire un livello minimo di sicurezza e privacy in questo fronte e concentrarsi anche sugli altri aspetti del sistema.
L'applicazione cifra i dati da inviare al Proxy attraverso la chiave pubblica del Governo in modo che questo sia l'unico in grado di interpretare i messaggi che giungono dai cittadini, impedendo quindi al gestore del Proxy di ottenere queste informazioni.

Si consideri ora un attacco, basato sull'analisi del traffico, da parte di un ipotetico avversario che si pone in ascolto nelle comunicazioni in ingresso al Proxy. Poiché le uniche comunicazioni che vanno dal cittadino verso il Proxy (e dunque server del Governo) sono quelle per condividere il proprio SK_t, in quanto il cittadino è risultato essere infetto, allora l'avversario è in grado di ottenere informazioni come indirizzo IP/MAC del cittadino infetto.
La soluzione che è stata adottata per proteggere il sistema da questo attacco consiste nel far inviare all'applicazione di ogni cittadino, in istanti di tempo casuali, un pacchetto simbolico della stessa dimensione di quelli contenenti informazioni utili, contenente bit casuali e codificato con lo stesso standard di un pacchetto normale attraverso lo schema di public key encryption con la chiave pubblica del server. Il server decifrando il pacchetto ed accorgendosi che non è ben formato, lo scarta.
La medesima protezione vale anche per il Proxy posto tra il Cittadino ed il Laboratorio di Epidemiologia con la differenza che i pacchetti non sono inviati ad intervalli casuali ma regolari da parte degli utenti che aderiscono e hanno una dimensione non fissata, come vedremo successivamente.

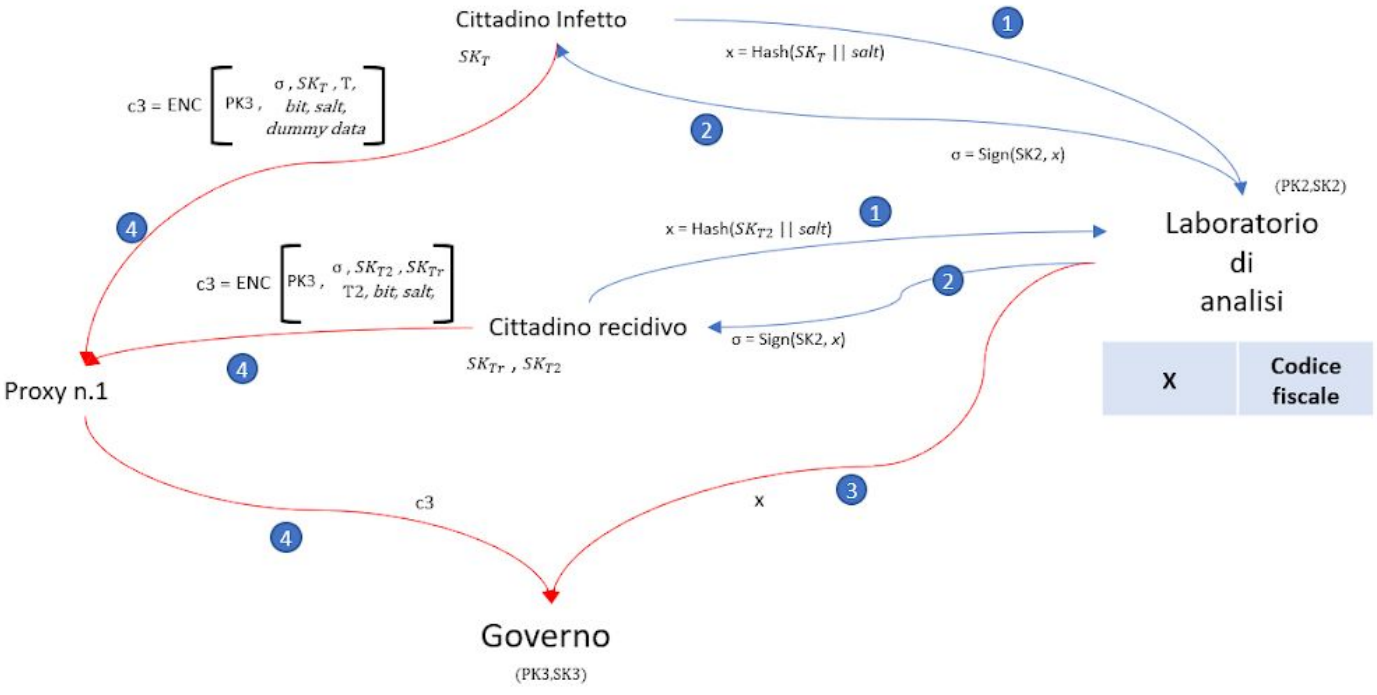
Un cittadino a cui è stato comunicato di essere positivo e che ha intenzione di mettere a disposizione i propri *EphID* alla comunità, dovrà essenzialmente avviare una comunicazione con il Laboratorio di Analisi ed una con il Proxy che indirizzerà la comunicazione verso il server del Governo. Per semplicità, nel prosieguo della discussione parleremo di comunicazione tra Governo e Cittadino infetto sottintendendo che ci sia il Proxy da tramite.

Il Cittadino che è stato sottoposto a tampone e che attende il risultato, manda al Lab. di Analisi un messaggio $x = H(SK_t \,||\, salt)$ e memorizza sul proprio dispositivo SK_t e $salt$. Il motivo dell'aggiunta del termine salt, che è una stringa casuale, è dovuto alla struttura stessa con cui si ricavano gli SK_{t+1}:

$$SK_{t+1} = H(SK_t)$$

dunque ci permette di non fornire al Laboratorio informazioni sull'SK_t ed i suoi successori. La funzione di hash utilizzata verrà illustrata nel paragrafo 3.1.

Il Laboratorio di Analisi ricevuta la x dal cittadino la conserva in un dizionario che ha la x come chiave e come valore il codice fiscale. Quando l'esito del tampone è noto, se il cittadino risulta essere positivo il Laboratorio manda la x al server del Governo e cancella la relativa riga dalla tabella, se negativo la cancella senza inviarla al Governo.
A sostegno dell'ipotesi che $salt$ debba essere una stringa random e non nota c'è la motivazione che qualora si implementasse il sistema con quest'ultima opzione, il Laboratorio di Analisi leggerebbe, a comunicazione avvenuta di positività, la SK_t del cittadino trasmessa in broadcast dal server del Governo. A questo punto, facendo la hash della SK_t con un termine concatenato noto otterrebbe la x. Avendo la SK_t e la x, potrebbe cercare nel dizionario (con chiave che ricordiamo essere proprio la x) il codice fiscale ed associarlo ad SK_t, riuscendo dunque ad associare gli EphID ad una persona fisica.



Il Laboratorio di analisi possiede una coppia *(SecretKey, PublicKey)*, grazie alla quale può calcolare la firma digitale da inviare al cittadino:

$$\sigma = Sign(Secret\ Key, x)$$

Ogni cittadino che intende inviare le informazioni al server del Governo, dovrà inviare anche σ per motivi in seguito illustrati.

Nel momento in cui il Cittadino risulta infetto, il Laboratorio invia la x relativa a quel paziente al Governo e la cancella dal dizionario. Questa operazione permette di assicurarsi che il Cittadino non cambi la SK_t da inviare al server del Governo, dal momento in cui ha mandato x al Laboratorio.
Quando il Cittadino vorrà comunicare al Governo la propria SK_t dovrà mandare i seguenti parametri: σ - SK_T - T - salt - b_r - SK_{Tr} - dummy data.

- **T** indica il numero di giorni, stimati dal Laboratorio, che intercorrono tra il primo giorno in cui il cittadino è risultato essere contagioso e la data odierna.
- **b_r** è il bit di recidività che indica se il cittadino è recidivo o meno
- ***SK_{Tr}***, questo campo non viene inviato qualora il cittadino non sia recidivo, corrisponde alla SK_t inviata al governo la prima volta che si è risultati infetti, in particolare è la SK_t del giorno precedente a quella precedentemente inviata, in questo modo il Governo potrà verificare che il cittadino è effettivamente recidivo come si vedrà successivamente.
- **dummy data**, questo campo non viene inviato qualora il cittadino sia recidivo, sono dati "spazzatura", in particolare sono 256 bit (la lunghezza di una SK_T), che servono per rendere tutti i messaggi omogenei in quanto a grandezza, in modo da rendere maggiormente difficile per un avversario cercare di interpretare la natura del possibile messaggio inviato, quindi se il cittadino è un recidivo o un infetto normale.

Il server del Governo per prima cosa decifra, attraverso la propria chiave segreta, i pacchetti inviati dal proxy in modo da ottenere le informazioni inviate dal cittadino. Ricevute tutte queste informazioni si accerta che il messaggio sia stato inviato da una persona realmente positiva in questo modo:

1. calcola $x = H(SKt \,||\, salt)$.
2. verifica la firma calcolando: $Vrfy(PKlab, x,)$
3. se il messaggio è stato firmato correttamente, si assicura che la *x* sia presente nella lista di quelle inviate dal laboratorio di analisi.

Si noti come non sia necessario per individuare cittadini che inviano informazioni al server ma che non sono realmente infetti, questi sarebbero individuati al passo 3. La sua utilità sta nel permettere al server un controllo più rapido evitando di cercare la x

nella lista durante la comunicazione con cittadini non autorizzati dal laboratorio. Ad esempio in casi di attacchi DDOS al server, che consisterebbero nell'invio di un numero molto elevato di messaggi da gestire, si ha un netto miglioramento della velocità con la quale il server scarta i messaggi in questione.

Nel caso in cui tutte le verifiche hanno dato esito positivo, il server carica le informazioni utili ed effettua l'invio broadcast, altrimenti rigetta il messaggio.

Se la richiesta è andata a buon fine l'app del cittadino calcolerà un nuovo SK_0, con cui calcolare SK_1 SK_2 … SK_t ,da utilizzare nei giorni seguenti. In questo modo il cittadino infetto evita di essere tracciato attraverso l'identificativo. Qualora risultasse essere, in futuro, non guarito o nuovamente positivo, invia le dovute informazioni ripetendo il procedimento.

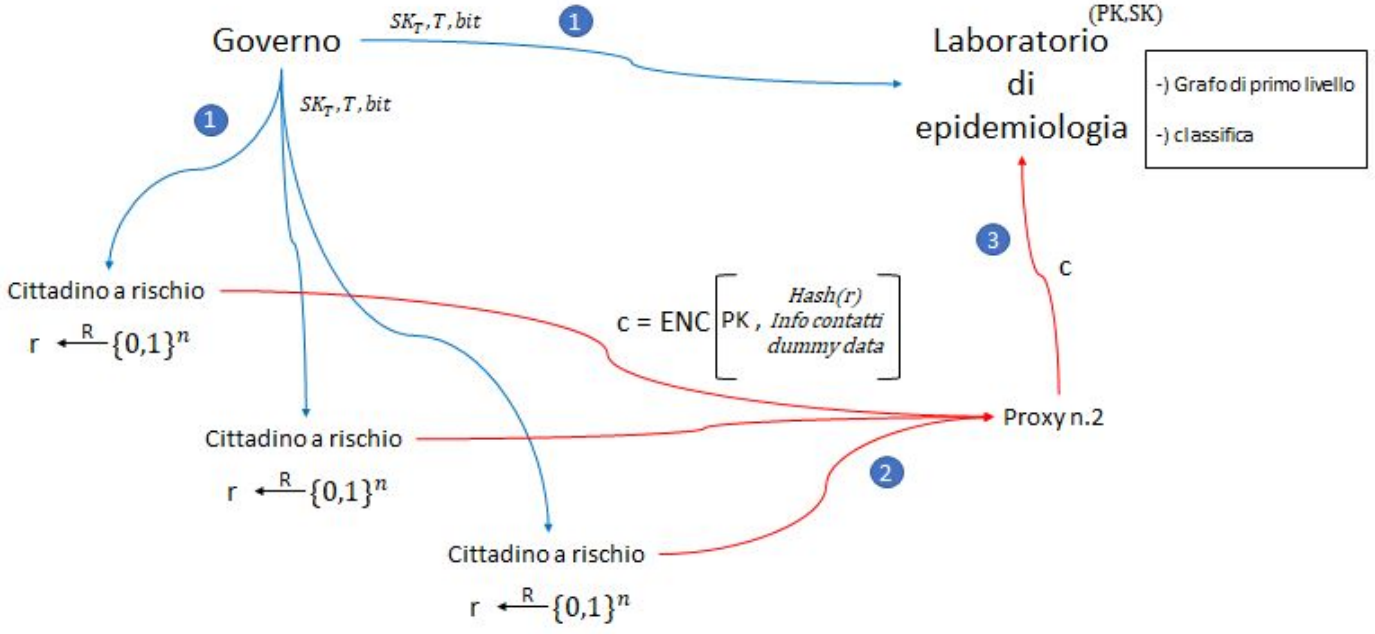
La vecchia SK_{t-1} viene salvata nell'applicazione perché risulterà utile qualora il Cittadino diventasse nuovamente positivo. Infatti, un Cittadino recidivo, si comporta come un cittadino infetto normale nella fase di comunicazione con il laboratorio e con il server del Governo (sempre attraverso il Proxy), con un'unica differenza: al posto di inviare i "dummy data" al Governo, invia l'SK_t relativa alla precedente volta che è risultato infetto, con t = T-1, ovvero invia l'SK_t precedente a quella che aveva, in passato, inviato al governo alla sua prima infezione.
Manda la *SK* relativa a T-1 e non a T perché un qualsiasi cittadino infetto potrebbe pensare di mandare come *SK* di recidività una qualsiasi di quelle che il server del Governo ha mandato in broadcast a tutti, allarmando e creando false informazioni nel sistema. Con T-1 è possibile proteggersi da questo tipo di attacco perché il server del Governo farà una hash di questa SK_{t-1} in modo da verificare che l'output è presente nella lista delle SK_t finora ricevute, e da avvisare che l'infetto è recidivo. SK_{t-1} è un'informazione che il server non possiede a priori e che possiede solamente l'utente realmente recidivo e dunque si sta effettuando un commitment.

L'individuazione di SK_t appartenenti ad utenti recidivi risulta particolarmente utile sia per un corretto calcolo del rischio da parte dei cittadini, sia per studi epidemiologici, dove può essere utile avere informazioni separate per ceppi diversi del virus.

Cittadino a rischio - Governo - laboratorio di epidemiologia

In questo blocco si analizza ciò che avviene successivamente alla pubblicazione della SK_t di un infetto nel sistema proposto. In particolare il Governo invierà a tutti i possessori dell'applicazione e al laboratorio di epidemiologia le informazioni riguardanti l'avvenuta infezione. In questo modo le applicazioni dei singoli cittadini potranno calcolare il fattore di rischio.
Il calcolo del fattore di rischio avviene in base al tempo di contatto con una persona infetta ed alla distanza stimata a cui il contatto è avvenuto. Una ulteriore distinzione può essere fatta se il cittadino infetto è recidivo, infatti potrebbe accadere che il nuovo ceppo del virus sia più o meno infettivo, dunque in broadcast il governo invia anche il *bit di recidività*. La formula che calcola la percentuale di rischio sarà governata da alcuni parametri suggeriti dal laboratorio di epidemiologia stesso. L'aggiornamento di questi parametri avviene tramite il sistema che governa le applicazioni mobili (app store dei dispositivi fisici), dunque non viene trattato nella progettazione in questione.

Di seguito viene mostrato lo schema principale di questo macroblocco.



Dallo schema si può notare come il governo invii le informazioni relative al cittadino infetto anche al laboratorio di epidemiologia, che potrà dunque calcolare un grafo dei contatti di primo livello unendo queste informazioni a quelle che gli arrivano dai cittadini.

Infatti, tutti i cittadini che accettano l'invio di dati al laboratorio, collaboreranno inviando le proprie informazioni sui contatti avuti con persone infette. In particolare, periodicamente, i cittadini invieranno la lista degli *EphID* appartenenti a persone infette e immagazzinati nella propria applicazione a seguito di uno o più contatti.
Il laboratorio dunque potrà capire quanti contatti ci sono stati tra il cittadino infetto, di cui possiede la SK_t, e altri cittadini formando dunque il grafo di primo livello.

Per proteggere la privacy è previsto l'utilizzo di un secondo Proxy, analogo a quello utilizzato nel secondo macroblocco precedentemente descritto. Infatti, conoscendo l'indirizzo IP dei cittadini che sono stati a contatto con un infetto, e conoscendo le SK_t di quest'ultimo, si potrebbe associare una SK_t ad un'area geografica, e questo non è un obiettivo del nostro sistema.
Per questo motivo, in maniera del tutto analoga all'utilizzo del proxy visto precedentemente, l'applicazione cifra i dati da inviare al proxy attraverso la chiave pubblica del laboratorio di epidemiologia cosicché questo sia l'unico in grado di conoscere le informazioni riguardanti i cittadini.
Lo schema di *public key cryptography* utilizzato sarà illustrato nel paragrafo 3.1.

Un' ulteriore accortezza di cui parlare è l'invio anche di *dummy data* da parte dei cittadini in modo da evitare analisi del traffico dati. Infatti un'analisi del genere evidenzierebbe i cittadini che sono stati maggiormente a contatto con un infetto che inviano un messaggio più lungo essendo la lista degli *EphID* da inviare più lunga. Quindi, anche se il cittadino non ha avuto alcun contatto con persone infette, invierà dummy data evitando che un avversario in ascolto sulla rete possa fare inferenza. Non essendo possibile fissare una lunghezza fissa del messaggio in questo caso, vengono inviati un numero di dummy data casuale che viene scelto al runtime dall'applicazione.

L'ultimo aspetto importante da sottolineare è la creazione di una classifica da parte del laboratorio di epidemiologia il cui utilizzo verrà esplicitato nel seguito.
Si vuole classificare ogni utente in base al proprio valore di rischio, utenti maggiormente a rischio saranno quindi nelle prime posizioni della classifica. La classifica è una classifica privata ma che, come vedremo nel prossimo paragrafo, potrebbe diventare di pubblico dominio e per questo motivo è necessario evitare di memorizzare informazioni sensibili dei cittadini. Infine, un cittadino che è presente in questa classifica deve poter dimostrare di essere effettivamente colui che si nasconde dietro quei determinati dati anonimi in una posizione.

Il sistema prevede l'utilizzo di un protocollo di *commitment*. In particolare, ogni cittadino calcola e memorizza una stringa randomica e invierà il risultato di una funzione di *hashing,* applicata alla stringa, al laboratorio di epidemiologia (sempre attraverso il Proxy). Il laboratorio potrà dunque stilare la classifica in base alla

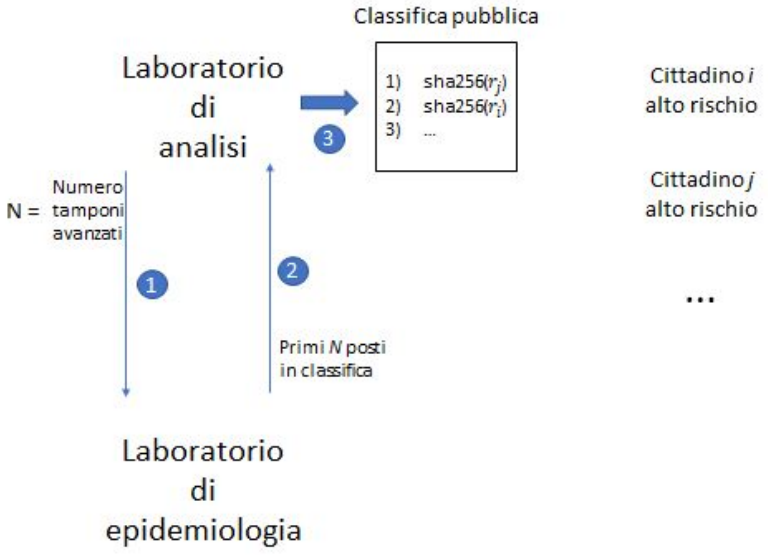
percentuale di rischio di ogni cittadino, identificato dal risultato della funzione di hash.
A questo punto il cittadino potrà dimostrare di essere colui che si trova in una determinata posizione in classifica mostrando la stringa randomica inizialmente calcolata al laboratorio di analisi che effettua l'hashing e verifica la corrispondenza con il valore in classifica.
La funzione Hash utilizzata verrà illustrata nel paragrafo 3.1.

La classifica deve inoltre essere periodicamente aggiornata e ricalcolata, difatti ogni giorno sono possibili stravolgimenti in classifica ed inoltre ogni giorno lo stesso cittadino invierà una stringa randomica diversa per evitare possibili tracciamenti non desiderati dal sistema proposto.

Laboratorio di epidemiologia - laboratorio di analisi - Cittadino

Questo macroblocco è il naturale successore del paragrafo precedente in termini temporali, infatti si analizza la comunicazione prevista tra il laboratorio di epidemiologia e quello di analisi.

Le principali funzioni esposte in questo macroblocco riguardano un futuro in cui la pandemia è in fase decrescente e quindi periodicamente risultano avanzare tamponi e altre risorse che magari andrebbero utilizzate nei confronti dei cittadini che non possiedono una grave sintomatologia ma risultano comunque essere ad alto rischio infezione.
Questi sono quei cittadini che si riconosceranno in classifica e che potranno andare al laboratorio di analisi per richiedere un tampone anche senza presentare sintomi.



In riferimento quindi allo schema mostrato in figura, notiamo come prima cosa la comunicazione del numero di tamponi avanzati, N, e la seguente comunicazione dei primi N posti in classifica, ovvero delle stringhe corrispondenti a cittadini che effettivamente possono ottenere un vantaggio da questa classifica.
Ovviamente qualora si ritenga opportuno è possibile ottenere *N* + *m* elementi della classifica e poi pubblicare solo i primi N, in modo da avere già elementi della classifica aggiuntivi qualora qualche cittadino tra i primi N non si presenti lasciando disponibile il proprio tampone per qualcun altro.

La classifica può dunque essere pubblicata dal laboratorio di analisi senza problemi non essendo presente alcuna informazione personale. A questo punto, come descritto nel paragrafo precedente, i cittadini potranno dimostrare di appartenere ad una determinata posizione in classifica in quanto gli unici in grado di calcolare, in tempo polinomiale, il valore pubblicato nella classifica, questo se si implementa una funzione di hash del tipo CRHF (Collision Resistant Hash Function) come la *sha256* presente in figura.

3. 1 Algoritmi utilizzati

Come funzione *Hash*, utilizzata in diverse parti della progettazione del sistema, è stata impiegata la terza versione di SHA, in particolare quella a 256 bit, che è l'ultima versione del noto algoritmo.

Per quanto riguarda invece lo schema di *public key encryption* è stato utilizzato uno schema di "Hybrid encryption", quindi una prima fase di cifratura asimmetrica in cui il sender invia, cifrando con la public key del receiver, la chiave privata che verrà usata nel resto della comunicazione. Infatti, essendo la cifratura a chiave simmetrica computazionalmente più veloce, risulta ottimale procedere in questo modo. In particolare, lo schema di cifratura asimmetrica utilizzato è RSA con chiave di 4096 bit, mentre quello di cifratura simmetrica è lo schema di Fernet, ovvero:

- **AES128** in **CBC mode**; ed utilizzo del padding **PKCS7**.
- **HMAC** utilizzando **SHA256**.

Infine, lo schema di firma digitale, utilizzato dal Laboratorio di analisi, consiste sempre nello schema RSA che permette, oltre che effettuare *public key encryption*, anche di ottenere *public key authentication*.

4. Analisi del sistema

In questo capitolo viene condotta un'analisi sul sistema appena realizzato, che dimostra come la progettazione in questione (trattata nel capitolo 3) riesca a conseguire gli obiettivi di privacy e sicurezza, dichiarati precedentemente nel capitolo 2.
Si analizzano prima gli aspetti legati alla tutela della privacy degli utenti e in seguito quelli inerenti alla sicurezza del sistema.

4. 1 Privacy

Analizziamo innanzitutto le entità che potrebbero sollevare preoccupazioni dovute alla diffusione di informazioni sensibili e private dei cittadini e in che modo viene risolta tale problematica. Risulta inoltre opportuno ricordare l'assunzione fatta nel secondo capitolo, secondo la quale due entità distinte non collaborino tra loro intrecciando i dati raccolti sui cittadini con lo scopo di violarne la privacy.

• Laboratorio di epidemiologia: nel grafo di primo livello sono contenute soltanto le interazioni che si sono verificate tra cittadini infetti e altri individui. Vengono rivelate unicamente al laboratorio di epidemiologia, grazie all'autorizzazione del singolo cittadino nel momento in cui installa l'applicazione sul proprio smartphone. Queste interazioni non contengono informazioni sensibili degli utenti, difatti questi inviano al laboratorio semplicemente la lista dei contatti con cittadini infetti, evitando dunque l'invio degli identificativi, le SK_t, appartenenti ai cittadini. In effetti se si associasse ad ogni cittadino un identificativo e se si inviasse l'intera lista dei contatti avuti, si potrebbe ottenere un grafo di secondo livello, maggiormente dettagliato, che però porterebbe ad una mancanza di privacy, in quanto il laboratorio sarebbe in grado di ricostruire i contatti e le interazioni che avvengono tra i cittadini non infetti.

• Cittadini infetti: nessuna entità del sistema può associare l'identità di una persona positiva al tampone con i suoi identificativi nel sistema. Infatti questi cittadini non comunicano mai la propria SK_t al laboratorio di analisi, mentre nella comunicazione con il server sfruttano il Proxy, in questo modo nessuna entità conoscerà la relazione tra dati anagrafici (ed in particolare la località geografica) e la SK_t del cittadino.
Gli altri utenti a loro volta non potranno risalire all'identità delle persone attraverso le informazioni immagazzinate a seguito di uno o più contatti avvenuti.

• Cittadini a rischio: devono essere informati del pericolo e guidati al comportamento corretto, devono inoltre essere aiutati per quanto possibile ma senza intaccarne la privacy. E' infatti lo smartphone che calcola localmente la percentuale di rischio in

base ai contatti avuti con altri utenti, e avvisa l'utente. Dopodichè se il cittadino a rischio acconsente, i dati vengono periodicamente condivisi con il laboratorio di epidemiologia, ma solo al fine del calcolo del grafo di primo livello come detto precedentemente, per questo si evita l'invio di qualsiasi informazione personale se non la lista dei contatti avuti con persone infette. Per proteggere ulteriormente la privacy, si evita un collegamento tra questa lista di contatti e l'indirizzo IP (e quindi ipoteticamente una localizzazione geografica) attraverso il Proxy numero 2. Da queste informazioni poi il laboratorio di epidemiologia stila anche una classifica per aiutare i cittadini con un valore di rischio più alto, questi infatti potranno accedere ad eventuali tamponi avanzati in modo da prevenire contagi futuri.

• Laboratorio di analisi: è in contatto con quasi tutte le altre entità del sistema. Conosce i dati anagrafici dei cittadini che risultano infetti (che hanno effettuato un tampone). Nel sistema proposto, le informazioni private dei cittadini vengono massimamente protette. Infatti, il laboratorio non è in grado di associare le informazioni anagrafiche del cittadino con i suoi identificativi nel sistema, le SK_t . Per questo motivo, durante la comunicazione con il laboratorio di analisi, il cittadino infetto cifra la propria SK_t nascondendola al laboratorio.

• Governo, gestore del server principale: conosce gli identificativi delle persone infette e manda queste informazioni a tutti gli utenti del sistema. Le SK_t di cui viene a conoscenza sono solo quelle a partire dal giorno in cui il cittadino infetto è risultato essere contagioso. Il Governo non ha né facoltà né possibilità di associare le SK_t del cittadino ad altre sue informazioni personali, si garantisce quindi il massimo livello di privacy.

4. 2 Sicurezza

In questo paragrafo si tiene conto di alcuni possibili attacchi che lo schema appena progettato potrebbe ipoteticamente subire da un avversario o da un'entità corrotta.

4.2.1 Minacce scongiurate

Partiamo elencando le possibili debolezze che risultano essere superate dal sistema ideato, alcune magari già analizzate per motivare le scelte progettuali.

• Il governo NON può apprendere alcun indirizzo IP relativo alla provenienza dei dati che gli vengono inviati dall'applicazione, in quanto vi è il Proxy numero 1, che fa da

tramite nelle comunicazioni che avvengono tra il cittadino infetto (o recidivo) e il governo.

• Il gestore del primo proxy NON può avere accesso al contenuto dei messaggi inoltrati dai cittadini (infetti o recidivi) al governo, in quanto l'applicazione cifra tali dati utilizzando la chiave pubblica del governo (schema sicuro di cifratura a chiave pubblica RSA).

• Il laboratorio di analisi NON può dedurre alcuna informazione sulla SK_t (ed i successori) del cittadino che si sottopone al tampone, infatti conosce solo il risultato della funzione hash con parametro SK_t concatenato al termine $salt$. Il $salt$, che altro non è che una stringa casuale, serve ad evitare una corrispondenza tra il risultato della funzione di hash e la SK_t nel momento in cui Il governo la rende pubblica.

• Il laboratorio di analisi NON può dedurre il collegamento tra il cittadino infetto e la sua SK_t inviata in broadcast. Infatti, un'associazione temporale tra l'esito positivo del tampone ed il caricamento delle SK_t da parte del cittadino, porta a delle conclusioni inesatte essendo il cittadino libero di caricare i dati quando lo ritiene più opportuno.

• Il cittadino infetto, dopo aver comunicato al laboratorio di analisi il messaggio x (dove $x = H(SKt \,||\, salt)$), NON può modificare la propria SK_t da inviare al governo; dovrà obbligatoriamente coincidere con quella inviata precedentemente al laboratorio di analisi.
Questo meccanismo di sicurezza viene garantito dal fatto che il laboratorio di analisi comunica al governo le x di tutti i cittadini infetti verificati.

• Un cittadino sano NON può camuffarsi e registrarsi come infetto nel server del governo, in quanto il governo effettua un controllo di verifica della firma del laboratorio, firma rilasciata solo ai cittadini infetti, e calcola la x mediante SK_t e $salt$ controllando se fa parte della lista delle x che si aspetta di ricevere.

• Un avversario in ascolto sulla comunicazione dal cittadino infetto al governo, non potrà attraverso un'analisi della grandezza dei pacchetti HTTPs scambiati, capire se il cittadino è semplicemente infetto oppure è recidivo, perché si fa utilizzo di dummy data con il fine di rendere uguali le lunghezze dei messaggi in un caso e nell'altro. In generale, un avversario di questo tipo non può nemmeno capire che il cittadino è positivo al virus (infetto o recidivo che sia), in quanto, come visto nel capitolo 3, vengono inviati anche pacchetti "spazzatura" da parte di cittadini sani.

• Il gestore del secondo proxy NON può avere accesso al contenuto dei messaggi inoltrati dai cittadini a rischio al laboratorio di epidemiologia, in quanto l'applicazione

cifra tali dati utilizzando la chiave pubblica del laboratorio di epidemiologia (schema sicuro di cifratura RSA).

• Un avversario che analizza semplicemente il traffico dati dal cittadino al proxy n.2, NON riesce ad interpretare se quell'individuo è effettivamente a rischio, in quanto i dummy data lo confondono: non fanno trapelare quali sono i messaggi contenenti una lista di contatti più lunga e quindi quali sono i cittadini a maggior rischio. In generale, non riesce a capire quali cittadini sono stati a contatto con un cittadino infetto perché ci sono cittadini che non sono stati a contatto con un infetto che inviano pacchetti "spazzatura" al laboratorio.

• Un avversario NON può fingere di essere presente nella classifica delle persone che hanno precedenza su un eventuale tampone. Infatti i dati che si nascondono in una delle posizioni della classifica appartengono a specifici cittadini che hanno effettuato un protocollo di commitment attraverso una stringa randomica calcolata appositamente. Nel momento in cui devono dimostrare di occupare una delle posizioni in classifica, devono fornire la stringa randomica al Laboratorio di Analisi che facendo la Hash dovrà poi controllare che l'output sia uguale a quello presente in classifica per fornire il tampone.

4.2.1 Possibili debolezze accettate

Passiamo adesso all'analisi degli attacchi possibili, che seppur individuati non vengono annullati, perché un'eventuale difesa porterebbe alla perdita parziale di utilità del sistema oppure perché non è interesse di questo elaborato risolverli.

• Problema dei falsi positivi: il cittadino infetto è vincolato nell'inviare al server la stessa SK_t di cui è in possesso nella fase di comunicazione con il laboratorio di analisi. Un cittadino infetto avversario può utilizzare una SK_t non di sua proprietà (ad esempio prendendo in prestito il cellulare di un altro cittadino che utilizza l'applicazione) in quella fase e caricare nel server del Governo questa SK_t facendo credere che appartenga ad una persona infetta. Ciò comporterebbe un calcolo del fattore di rischio errato nei cittadini che sono venuti a contatto con l'avversario o con la persona di cui l'avversario ha utilizzato la SK_t.

• Problema falso contatto: un avversario potrebbe munirsi di un un dispositivo che amplifica il raggio del segnale BLE emesso simulando quindi contatti tra persone a distanza non ravvicinata.

• Problema della trasmissione incompleta degli *EphID*: un utente infetto potrebbe decidere di caricare le informazioni nel server del Governo ma volendo evitare di

caricare informazioni riguardanti gli ultimi “tot” giorni. Può farlo inviando una T per cui non si arrivi a calcolare la hash della SK fino a quei giorni (esempio: un utente deve inviare informazioni degli ultimi 14 giorni ma per sue motivazioni non vuole inviare quelle relative agli ultimi 3 giorni, dunque anziché inviare la SK_t ed indicare T uguale a 14, indica T uguale ad 11). Il problema potrebbe essere risolto in diversi modi. Ad esempio:

- Fissando la T ad un numero ben definito uguale per tutti i cittadini
- Facendo inviare la T, anziché dal cittadino, dal Lab. di Analisi in associazione alla x

Tuttavia si è deciso di rendere possibile questo tipo di “attacco” perché un cittadino infetto di fronte ad un tal problema potrebbe decidere di non trasmettere affatto la SK_t a causa di un periodo limitato di tempo. Non costruendo azioni di difesa a questo “attacco”, il cittadino può mandare dati incompleti ma che risultano essere comunque utili.

• Modifica calcolo rischio: un avversario che ha accesso al codice sorgente dell’applicazione ed è in grado di modificarlo, può modificare la lista dei contatti da inviare al laboratorio di epidemiologia simulando un numero di contatti maggiore con persone infette, di cui conosce l’SK_t, e in questo modo può ottenere un posto migliore in classifica che comporterebbe con maggiore probabilità l’ottenimento di un tampone gratuito anche senza presentare sintomi. Un attacco di questo tipo, se molto diffuso, potrebbe essere identificabile confrontando le statistiche degli esiti di questi tamponi con la stima che ha portato al valore del fattore di rischio, e dunque, si potrebbe decidere di disabilitare questa funzionalità.
Questo attacco provocherebbe inoltre rumore nei dati raccolti dal laboratorio sul grafo di primo livello e sulla situazione di rischio dei cittadini.

• Proxy - Governo - Laboratori: come specificato altre volte nel corso di questo documento, un’azione combinata di 2 o più tra queste entità porterebbe ad una perdita di privacy dei cittadini.

4. 3 Conclusioni

Traendo le conclusioni rispetto le proprietà di confidenzialità ed integrità proposte nel capitolo 2, queste risultano rispettate sulla base dell’assunzione fatta che due o più entità del sistema non collaborino tra loro unendo le informazioni a loro disposizione.

5. Implementazione

La parte del progetto che si è deciso implementare è la comunicazione tra il cittadino infetto, il laboratorio di analisi ed il Governo. L'implementazione consiste nella simulazione del protocollo che regola questa parte del sistema. Il linguaggio di programmazione utilizzato è Python. Il codice, presente nella cartella "Implementazione" presenta numerosi commenti, per questo motivo in questo capitolo vengono presentati solo i concetti fondamentali.

Le entità in gioco sono:

- **Cittadino**: per prima cosa richiede un analisi al laboratorio inviando $x = H(SK_t \,||\, salt)$, quindi attende il risultato dell'analisi. Con probabilità del 50% risulta positivo e può quindi decidere se inviare o meno la propria SK_t al server, insieme alle altre informazioni descritte nel capitolo 3, cifrando attraverso la chiave pubblica del Governo.
- **Server del Governo:** server in attesa di messaggi provenienti dal cittadino, nel momento in cui ne arriva uno, effettua la decodifica e la verifica della SKt assicurandosi che sia valida la firma del laboratorio (sigma), ed inoltre che sia presente nella lista di x inviate dal laboratorio.
- **Laboratorio di analisi:** server in attesa di un cittadino, fornisce il risultato del tampone al cittadino e, qualora esso risulti positivo, invia la x al server del governo.
- **Proxy**: inoltra i messaggi HTTPS dal cittadino al governo.
- **Autorità dei certificati**: essendo parte della comunicazione basata sul protocollo HTTPS, si è implementata una entità con certificato autofirmato che fa da garante dei certificati delle altre entità in gioco.

Si noti come con questa implementazione si evitano attacchi da parte di un avversario che vuole fingersi infetto senza attendere un risultato positivo delle analisi o senza che questo risultato sia positivo. Infatti, il server scarterà tutti i dati inviati dai cittadini che non superano la fase di verifica.

Inoltre, si noti come il laboratorio faccia un "parsing dell'input" per assicurarsi che i messaggi inviati dai cittadini rispettino il format previsto, ovvero siano il risultato della funzione *sha256* e quindi siano tutti caratteri esadecimali. Lo stesso tipo di parsing è previsto dal governo quando deve ricevere le x dal laboratorio, mentre quando deve

ricevere i dati dal cittadino infetto si assicura che questi non contengano caratteri speciali non previsti.

Infine si noti come, nell'atto della comunicazione dei dati da parte del cittadino verso il server del governo, i dati vengono cifrati attraverso una chiave in maniera simmetrica con uno schema *AES* (schema di Fernet come visto nel paragrafo 3.1), a questo punto si cifra la chiave utilizzata attraverso uno schema di cifratura asimmetrica, *RSA,* utilizzando la chiave pubblica del server del governo. In questo modo evitiamo l'utilizzo eccessivo della cifratura asimmetrica che è noto essere computazionalmente più onerosa sia in termini di tempo che di risorse (è necessaria una chiave più lunga).

5. 1 Guida all'uso

Per avviare la simulazione occorre inizialmente avviare il server del governo, il proxy ed il laboratorio, presenti rispettivamente nei file *server.py*, *proxy.py*, *lab.py.*

A questo punto è possibile avviare *cittadino.py* per simulare l'arrivo di un cittadino, il quale chiederà il risultato delle analisi al laboratorio calcolando ed inviando al laboratorio la propria x.
In particolare, dopo un intervallo di tempo di 10 secondi il laboratorio mostrerà il risultato, nel caso in cui il cittadino risulti negativo, esso terminerà il programma digitando "No" ed il sistema entrerà in attesa di un nuovo cittadino.
Invece, qualora risultasse positivo potrà decidere se inviare i dati. Nel caso di una decisione negativa, nuovamente digitando "No" il programma termina e si attende l'arrivo di un nuovo cittadino. Se il cittadino risponde "Si", l'applicazione codifica i dati e li invia tramite HTTPS al proxy verificandone il certificato.

Inoltre, se il cittadino è risultato positivo, il laboratorio invia la x al server, come da protocollo, a questo punto il server potrà, dopo aver decodificato i dati, verificare la SK_t del cittadino terminando la comunicazione con il cittadino in questione.