Foreløpig rapport for kodegjennomgang av løsning for digital smittesporing av koronaviruset

09.04.2020

Bakgrunn for arbeidet

Utbruddet av COVID-19 er et alvorlig utbrudd av smittsom sykdom som kan få alvorlige helsekonsekvenser for mange mennesker. Som et tiltak for å minske smitte av dette viruset har Folkehelseinstituttet (FHI) utviklet appen Smittestopp.

Ekspertgruppen for kodegjennomgang av løsningen for digital smittesporing er satt ned av Helse- og Omsorgsdepartementet for kontroll av appen Smittestopp og bakenforliggende løsninger. Helse- og Omsorgsdepartementet (HOD) har bedt IKT Norge om å anbefale aktuelle deltakere. Ingen av deltakerne er knyttet til FHI, HOD eller Simula.

Formålet med ekspertgruppen er å ha en uavhengig part til å studere all kode i systemet og verifisere at den er forsvarlig med tanke på sikkerhet og personvern. Vi har valgt å beskrive løsningen relativt detaljert. Dette slik at det vil være mulig for brukere, tilsynsmyndigheter, folkevalgte og myndighetene å få en god forståelse av hvordan denne løsningen virker, hva slags data som samles inn, hvordan de brukes og hvordan de deles.

Vi viser til mandatet som ble publisert av regjeringen 8.4.2020.

Mandat
https://www.regjeringen.no/contentassets/82254fd2dd5f431cb98f57ac28ca1510/mandat-ekspertgruppe.pdf

Avgrensninger

Gruppen har hatt begrenset tid tilgjengelig for å evaluere løsningen. Mye av den planlagte funksjonaliteten er heller ikke på plass. I sum betyr det at vi har måttet begrense hva vi kan vurdere og uttale oss om i den foreløpige rapporten.

I den foreløpige rapporten har vi derfor fokusert på sikkerhetsmekanismer ved innhenting og lagring av data. Vi har også sett på de implementerte slettemekanismene, men kan ikke på det nåværende tidspunkt verifisere sletting tilstrekkelig i den totale løsningen.

Diskusjon rundt den valgte konkrete arkitekturen har vi ikke hatt tid til å se på i denne foreløpige rapporten, men vi vil komme tilbake til dette i den endelige rapporten. Dette inkluderer også detaljerte betraktninger rundt valg og bruk av eksisterende løsninger eller protokoller.

Vi har gjort en detaljert gjennomgang av koden som ligger bak applikasjonene på både Android og iOS. Vi har også sett på grensesnitt mellom app og backend, og på ferdig funksjonalitet i backend. Per i dag er det kun funksjonalitet for lagring av data fra app og sletting av data i skyløsningen ved bruk av sletteknappen i app. Vi har ikke hatt anledning til å teste funksjonalitet ende til ende.

Det skal også implementeres sletting av all data etter 30 dager, og sletting av data hvis det ikke kommer oppdatert informasjon fra en gitt brukers app på 7 dager (for eksempel dersom en bruker avinstallerer appen uten å slette data først), men dette er ikke på plass enda. Vi vil gjøre en evaluering av dette når denne funksjonaliteten er klar.

Vi har ikke hatt tilgang til fullstendig konfigurasjon av skyløsningen, ettersom det er flere ting som er opprettet ikke-programmatisk. Vi har heller ikke hatt tilgang til byggpipeliner eller produksjonsmiljø for å verifisere om konfidensialitet og integritet er ivaretatt med tanke på utro tjenere.

Vi har identifisert tre andre områder som vi ikke har sett på enda, men som vi kommer til å se på i de neste fasene:

- Varslings- og innsynsfunksjonalitet i webapplikasjonene til FHI (for smittesporing), med tilhørende backendkode.
- Tilgangsstyring til data og tracking av bruk av data for FHI.
- Plan for å ta ut anonymisert og aggregert data.

Dette betyr at vi ikke kan uttale oss om hvordan dataen som samles inn vil bli brukt av FHI på nåværende tidspunkt. Tilsyn og revisjon av bruk av data etter lansering er utenfor omfanget av vårt mandat. Vi vil anbefaler at det opprettes en uavhengig part som har ansvar for tilsyn med og revisjon av tilgang til dataene.

Vi er ingeniører og ikke jurister. Vi har derfor valgt å gi detaljerte beskrivelser av løsningene, som gjør det mulig for personer som har mer kompetanse på området å gjøre de juridiske betraktningene rundt personvern.

Fremgangsmåte

Ekspertgruppen ble etablert lørdag 4. april. Vi begynte å se på kildekode og rapporter søndag 5. april. Vi har tatt utgangspunkt i kildekode samt arkitekturskisser for å forstå helheten i systemet, og deretter gjort dypdykk i koden og gjennomført kodeanalyse for å identifisere mulige funn.

Vi har hatt daglige møter med Simula og FHI hvor vi har presentert våre funn. Vi har også brukt disse treffpunktene for å innhente mer data når dette har vært nødvendig. Vi kan se at Simula har gjort endringer basert på våre tilbakemeldinger.

Deler av skyløsningen er satt opp av Microsoft, og mye her er konfigurert direkte på serverne. For å få oversikt over disse delene av systemet har vi hatt møter med de som har satt opp dette hos Microsoft.

En overordnet beskrivelse av løsningen

Smittesporingsløsningen til FHI består av tre deler: Appen (Android og iOS) som brukere installerer på telefonen sin, en skyløsning som er levert av Microsoft (Azure) og webapplikasjoner som skal brukes av FHI. Appene kommuniserer kryptert til skyløsningen. I skyløsningen lagres data både i en SQL-database og i Azure Data Lake.
Tilgangsstyring, logging av tilgang, prosedyrer for sletting og aggregering av data i Azure er enda enten ikke påbegynt eller ferdig implementert. Denne funksjonaliteten er avgjørende for å kunne vurdere om personvernet er godt nok ivaretatt i løsningen.

Løsningen skal først testes ut i to kommuner. I parallell med dette vil tilgangsstyring, logging av tilgang, prosedyrer for sletting og aggregering av data i Azure bli implementert. Å teste deler funksjonaliteten selv om ikke hele systemet er klart, er vanlig.

En beskrivelse av appen Smittestopp

Kort forklart bruker appen telefonnummeret ditt til å identifisere og autentisere deg. Via GPS blir posisjonen din lagret på din enhet. Samtidig bruker appen Bluetooth til å lete etter andre enheter i nærheten som også har appen installert og estimerer distansen til disse. Dine lokasjonsdata, samt avstanden til andre telefoner i nærheten, lagres så i en sentral skyløsning. Dataene her er ikke tilgjengelige for andre enn FHI, som kan bruke informasjonen til å finne ut hvem som har vært i nærheten av en person som har blitt syk. Den lagrede dataen brukes også til forskning/analyse.

Brukeren registreres ved hjelp av tofaktorautentisering. Mer detaljert forklart skjer autentisering over HTTPS mot Microsoft Active Directory / B2C, som også binder telefonnummeret mot en intern sky-ID. Denne sky-ID-en brukes i all videre kommunikasjon med skytjenestene.

Appen sporer dine bevegelser via lokasjons-API-et på din telefon, herunder GPS og nærliggende WiFi-tilkoblingspunkt. For å oppnå høy presisjon brukes GPS aktivt og vi forventer at dette vil merkes på batterilevetiden. Dette er forsøkt å begrenses ved å skru ned presisjonen når enheten holder seg i ro. Posisjonsdata som breddegrad, lengdegrad, nøyaktighet, hastighet, høyde og nøyaktighet på høyde blir periodisk lagret i en ukryptert lokal database på telefonen.

Bluetooth-enheter i nærheten som også har installert appen annonserer i tillegg at de har installert appen (eller i tilfelle for iOS: at de er en iOS-enhet). Dette fører til en datautveksling med disse annonserende enhetene via GATT. Sky-ID-en fra de oppdagede og annonserende enhetene sendes over Bluetooth, og blir sammen med RSSI og mottakerens sendestyrke lagret ukryptert i en lokal database på telefonen. Det finnes også noe kode for å håndtere utfordringer rundt at en iOS-enhet vil lete etter andre enheter sjeldnere når den er i bakgrunnsmodus eller skjermen er slått av.

Oppsamlet posisjonsdata og Bluetooth-data blir sendt til Microsoft Azure IoT-Hub over TLS/MQTT, og blir deretter slettet lokalt. Data som er lagret lokalt, men ikke lastet opp enda, blir slettet hvis man logger ut eller blir av-autentisert.

Funn

Det er brukt vanlige patterns og arkitekturen er relativt oversiktlig. Tatt i betraktning den korte utviklingstiden er kodekvaliteten OK. Selv uten mye kommentarer er den lett å lese og relativt enkel å forstå, dog er det få automatiserte tester.

Det er heller ikke så vanskelig å forstå at man på grunn av den knappe tiden har valgt å gjøre en del konfigurasjon direkte på serverne, men over tid anbefaler vi at denne konfigurasjonen blir flyttet over i skript slik at det vil være lettere å spinne opp backenden igjen hvis det skulle skje noe. Dette øker også sporbarheten av endringer.

Appen bærer preg av at konfidensialitet er bedre ivaretatt enn integritet og tilgjengelighet. Vi har gjort funn innen skalering,robusthet, sårbarheter i tredjepartsbiblioteker, utviklingsmetodikk, bruk av protokoller, dataintegritet, datalekkasje, datavalidering, og svakheter i konfigurasjon. Det er ikke overraskende at vi har gjort funn innen så mange områder i en løsning som er laget på så kort tid.

Ingen av de tekniske funnene krever store kodeendringer eller store endringer i app / backend for å bli fikset, men vi mener at det er viktig at funnene blir rettet før appen rulles ut i hele Norge.

De dataattributtene som lagres ser ut til å være nødvendige for å oppnå formålet på den måten man har satt seg fore.

Det brukes permanente og enhets-spesifikke identifikatorer mellom enhetene. Dette vil potensielt åpne for muligheter til å utlede andres identitet eller smittestatus. Et alternativ ville være å benytte tidsbegrensede identifikator. Selv om dette kan føre til noe mer kompleks kode, er dette enkelt å implementere.

Konklusjon

I henhold til mandatet skal ekspertgruppen levere “en overordnet vurdering av om sikkerhet og personvern er forsvarlig ivaretatt”. I denne foreløpige rapporten har vi primært fokusert på sikkerhet. Personvern vil være lettere å uttale seg om når større deler av systemet er ferdig.

Appene samler inn posisjons- og Bluetooth-data som ønsket, og denne dataen sendes til og lagres i skyen med god sikkerhet. Den lagrede dataen brukes til to forskjellige ting, både smittesporing og forskning/analyse. Ved å gi brukere en mulighet til å bruke appen til kun smittesporing, vil gjerne brukere oppleve større kontroll på egen data.

Konfidensialitet er godt ivaretatt. Dataene blir samlet, sendt og lagret på en måte som gjør at risikoen for datainnbrudd er akseptabel for en app med så sensitiv data.

Vi har også sett på dataintegriteten. Dette handler om hvor lett det er å endre eller korrumpere informasjonen som lagres lokalt, og senere sendes til skyen. Vi mener at sikkerheten på dette området bør utbedres før endelig lansering av appen Smittestopp. Det er ikke grunnleggende mangler i arkitekturen, men mindre avvik som tilsammen gjør at det blir for enkelt å manipulere data.

Uten innsikt i, og mulighet til, å vurdere tilgangsstyring, logging av tilgang, prosedyrer for sletting og aggregering av data i Azure, er det vanskelig å anbefale appen for full lansering i hele Norge per dags dato. Å starte testing i to kommuner, som skissert av FHI, anser vi som et forsvarlig neste steg. Vi vil også anbefale at det kommer klart fram for brukeren at appen er i en testfase. En vanlig måte å gjøre dette på er å bruke en annen fargepalett i appen under testingen.

Vi vil anbefale at det opprettes en uavhengig part som har ansvar for å revidere (audit) all tilgang til dataene. Dette må også inkludere tilgang til kode.

Vi vil også anbefale at det etableres gode rutiner for jevnlig ytelsestesting for å sikre at applikasjonen som helhet har robust nok skalerbarhet. Vi anbefaler i tillegg at det blir utført tilstrekkelig penetrasjonstesting, før lansering.