Mikko Lindell

Tilankäyttöasteen arviointi Meshlium Scannerin ja Raspberry Pi 3 Model B:n avulla

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tietotekniikan koulutusohjelma

Insinöörityö

1.12.2017

Metropolia

Tekijä Otsikko Sivumäärä Aika	Mikko Lindell Tilankäyttöasteen arviointi Meshlium Scannerin ja Raspberry Pi 3 Model B:n avulla 27 sivua + 1 liite 1.12.2017
Tutkinto	Insinööri (AMK)
Koulutusohjelma	Tietotekniikka
Suuntautumisvaihtoehto	Tietoverkot ja tietoliikenne
Ohjaaja	Lehtori Marko Uusitalo

Insinöörityön tavoitteena oli selvittää mahdollisuus arvioida tilankäyttöastetta Raspberry Pi 3 Model B:n ja Meshlium Scannerin avulla. Työ oli tilaustyö kiinteistötekniikka-alan yritykselle.

Työssä käytettyjen mittarien oli tarkoitus pyrkiä havaitsemaan mittausalueella olevia henkilöiden käytössä olevia mobiilipäätelaitteita langattomien tekniikoiden, kuten WLAN:n ja Bluetoothin avulla. Havaittua dataa analysoimalla pyrittiin löytämään menetelmiä tunnistamaan laitteita ja sitä kautta henkilöitä, duplikaattilaitteita sekä signaalin voimakkuuteen pohjautuen laitteiden ja henkilöiden sijaintia. Analyysien lopputuloksia käytettiin puolestaan käyttöasteen arvioimiseen. Työhön kuului myös niin Raspberry Pi 3 Model B:n ja Meshlium Scannerin kuin myös WLAN- ja Bluetooth-tekniikoiden eroavaisuuksien vertailu.

Työssä keskityttiin menetelmään, jossa pyrittiin tunnistamaan yksittäiset laitteet ja minimoimaan duplikaattiesiintymät. Lisäksi tutkittiin ja pohdittiin menetelmän tarkkuutta ja tarkkuuteen vaikuttavia tekijöitä.

Vaihtoehtoisena menetelmänä pohdittiin ratkaisua, joka pohjautuu vastaanotetun signaalin voimakkuuteen. Tällöin yksittäisen laitteen tunnistus ei ole välttämätöntä, mutta mittausverkostossa on oltava huomattavasti enemmän mittareita sijainnin määrittämisen mahdollistamiseksi. Koska tässä menetelmässä laitteiden hetkelliset sijainnit ovat saatavissa, mitä tilaaja ei halunnut, ei tähän menetelmään syvennytty enempää.

Työssä tutkittu menetelmä eli henkilömäärän arviointi langattomien tekniikoiden avulla sinällään toimii, jos käytössä on vähänkin dataa henkilömäärään liittyen. Kuitenkin, suuren muuttujien määrän vuoksi, tulokset poikkeavat todellisuudesta erittäin helposti. Mitä kauempana ideaalitilanteesta ollaan, sitä enemmän tarkkuus heikentyy. Joissain tilanteissa menetelmä voi olla jopa täysin käyttökelvoton väärin arvioitujen muuttujien arvojen vuoksi. Mikäli tilankäyttöasteen arviointia yritettäisiin ilman minkäänlaista arviota todellisesta henkilömäärästä, menetelmä ei toimisi.

Avainsanat	tilankäyttöaste, Raspberry Pi 3, Meshlium Scanner, WLAN

Author Title Number of Pages Date	Mikko Lindell Estimation of space utilization using Meshlium Scanner and Raspberry Pi 3 Model B 27 pages + 1 appendix 1 December 2017
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Specialisation option	Computer Networks and Telecommunications
Instructor	Marko Uusitalo, Senior Lecturer

This bachelor's thesis was commissioned by Caverion Suomi Oy. The aim of the project was to find out whether space utilization could be estimated using Raspberry Pi 3 Model B and Meshlium Scanner.

The project focused on a method that aimed to detect single devices and to minimize duplicate devices. Additionally, the accuracy of the method and the factors affecting the accuracy were looked at. The purpose of the meters used in this project was to detect mobile devices carried by people within the scanning area. This was done by using wireless techniques such as WLAN and Bluetooth. By analyzing the detected data, the goal was to find methods to detect devices that will help to estimate the number of people and duplicate devices in the scanning area. The strength of the received signal will help to calculate the location of the devices and people. The results were then used for estimating space utilization. One part of the project was to compare the differences between Raspberry Pi 3 Model B and Meshlium Scanner as well as WLAN and Bluetooth techniques.

An alternative method was a solution that is based on the strength of the received signal instead of, for example, analyzing duplicate device entries. When this method is used, the scanner network should include a remarkably higher number of meters to make it possible to estimate the location of the devices and persons. Using this method, momentary locations are available, which was not wished for by the commissioner. Therefore, this method was not studied in more detail.

The method of using wireless techniques for estimating the number of people works as such, if any data on the number of people is available. However, due to many variables, the measurement results easily differ from real life values. The further away from an ideal measurement situation, the poorer the accuracy of the results was. In some situations, the method can even be totally useless due to wrongly set variable values. If the estimation of space utilization was carried out without any estimate of the real number of people, the method would not work.

Keywords	Space utilization, Raspberry Pi 3, Meshlium Scanner, WLAN

Sisällys

Lyhenteet

MAC-osoite	Media Access Control. Laitteen verkkokortin yksilöllinen tunniste. Koostuu kahdesta osasta: valmistaja ja verkkokortin tunniste.
Satunnaistaminen	Käyttäjän yksityisyyttä suojaava ominaisuus, jossa laitteen verkkokortin MAC-osoite muuttuu satunnaiseksi.
Duplikaattilaite	Yksi ja sama fyysinen laite, joka havaitaan usealla eri MAC-osoitteella.
Bluetooth	Langaton likiverkkotekniikka lyhyille etäisyyksille.
WLAN	Langaton lähiverkkotekniikka langattomaan tiedonsiirtoon Bluetoothia pidempiin etäisyyksiin.
Raspberry Pi 3 Model B	Raspberry Pi Foundationin kehittämä kolmannen sukupolven korttitietokone, jossa kaikki tarvittavat komponentit suorittimesta näytönohjaimeen löytyvät samalta noin luottokortin kokoiselta piirilevyltä.
Meshlium Scanner	Espanjalaisen Libeliumin moniprotokollareititin ja yhdyskäytävä, jossa on sisäänrakennettuna ominaisuudet muun muassa WLAN- ja Bluetooth-skannaukseen.
NTP	Network Time Protocol. Tietoliikenneprotokolla kellonajan synkronointiin.
RSSI	RSSI (received signal strength indicator) kuvaa vastaanotetun signaalin voimakkuutta.

1 Johdanto

Insinöörityön tavoite on selvittää, onko mahdollista arvioida tilankäyttöastetta langattomien tekniikoiden, kuten WLAN ja Bluetooth, avulla. Työ on tilaustyö Caverion Suomi Oy:lle. Yritys oli hankkinut Libeliumin Meshlium Scannerin käyttöönsä tätä tarkoitusta varten. Metropolian ohjaavan opettajan ehdotuksesta työhön otettiin mukaan myös Raspberry Pi 3 Model B -korttitietokone. Työn pääpaino on henkilömäärän arvioinnissa, jonka lisäksi osa insinöörityötä on arvioida henkilöiden etäisyys mittarista. Tavoitteena on saada käsitys henkilömäärästä alueella ja karkeahko käsitys henkilöiden etäisyydestä mittariin, ei henkilöiden sijaintia.

Päätelaitteista saatua tietoa ei järjestelmässä linkitetä muihin järjestelmän ulkopuolisiin tietoihin. Lisäksi yksilöivät tiedot, kuten päätelaitteiden MAC-osoite, pyritään joko anonymisoimaan tai salaamaan tilanteen mukaisesti. Rajauksella pyritään välttämään mahdolliset henkilötieto- tai tietosuojalain rikkomukset.

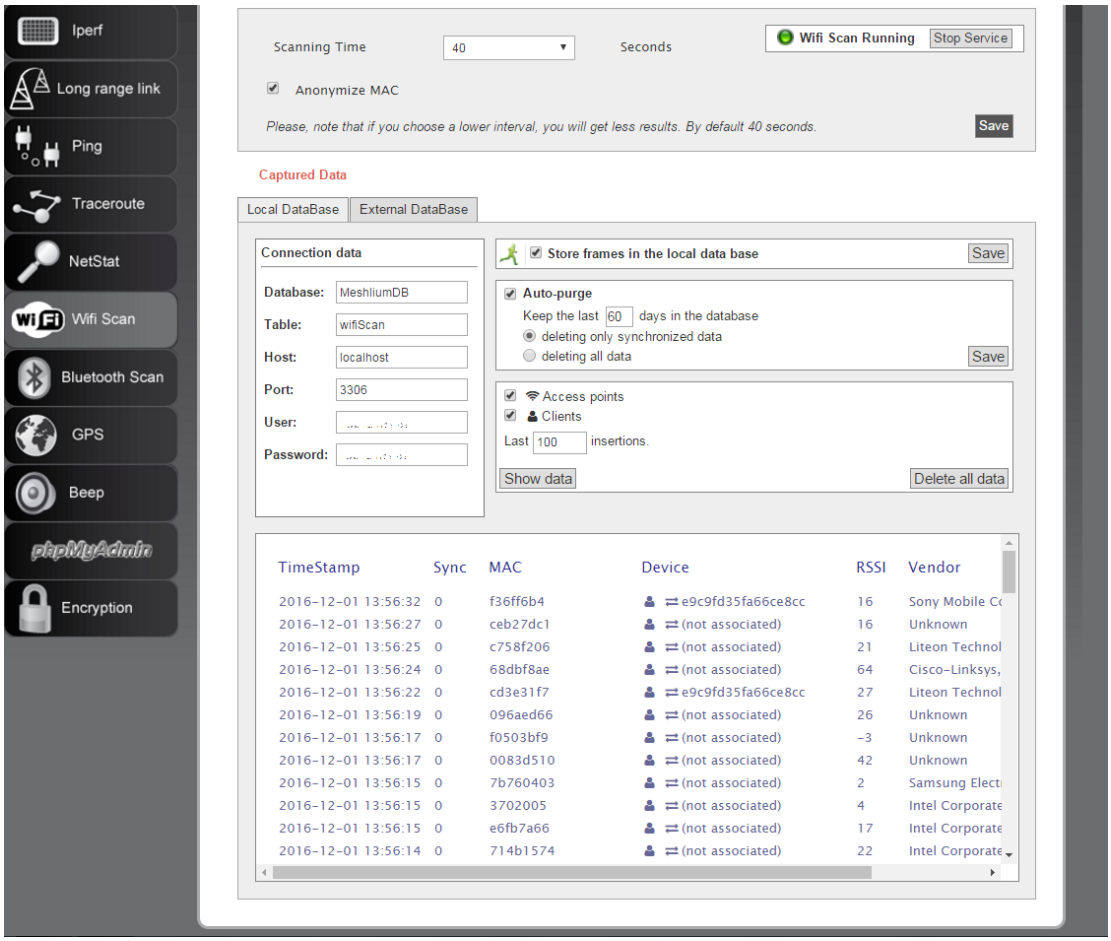
Tämä WLAN- ja Bluetooth-tekniikkoihin perustuva havainnointimenetelmä on itsessään Suomen vuonna 2017 voimassa olevan lainsäädännön mukaan täysin laillinen, mikäli data on vain mittareilta kerättyä eikä sitä linkitetä muuhun tietoon, kuten henkilötietoon, jolloin henkilötietolaki saattaisi nousta esteeksi menetelmän käyttämiselle. Lisäksi, menetelmässä ei pureta salattua liikennettä tai toimita MItM (Man In the Middle) -tyyppisesti, mikä voisi lainsäätäjän näkökulmasta olla epäilyttävää, vaan menetelmä perustuu liikenteeseen, joka on yleisesti radioteitse saatavilla.

2 Työssä käytetyt laitteet

2.1 Meshlium Scanner

Meshlium Scanner on espanjalaisen Libeliumin tuote, joka sisältää ominaisuudet sekä WLAN- että Bluetooth-tekniikoilla toimivien päätelaitteiden havaitsemiseen [1]. Valmistajan määrittelemä pääominaisuus Meshlium Scannerissa on toimia verkon yhdyskäytävänä langattomille mittareille, kuten Waspmote- ja Plug & Sense! -tuotenimien tuotteille, joita Libelium valmistaa. Meshlium Scanner tukee useita eri pilvipalveluita, joihin sen

kautta johdettua dataa voidaan ohjata. Näiden lisäksi Meshlium Scanner voi toimia langattomana tukiasemana. Kuva 1 esittää Meshlium Scannerin WLAN-skannauksen web-pohjaista käyttöliittymää.



Kuva 1. Meshlium Scannerin WLAN-skannauksen käyttöliittymä.

2.2 Rasperry Pi 3 Model B

Raspberry Pi 3 on Raspberry Pi -säätiön kehittämä tietokone. Se luokitellaan korttitietokoneeksi kokonsa vuoksi. Englanninkielisessä materiaalissa RPi:tä ja vastaavia laitteita kutsutaan nimellä ”single-board computer”, jonka vastine suomeksi on korttitietokone. Lisäksi, minitietokone-sanaan saattaa törmätä tietokoneklusteri- tai -rykelmä -yhteyksissä, kun puhutaan useammista laitteista. Version 3B koko on 85 x 56 x 17 mm. [2.]

Miksi Raspberry Pi sitten on valittu toiseksi skannerivaihtoehdoksi tässä työssä? Ensiksikin, sen sisäänrakennetut ominaisuudet kuten, WLAN ja Bluetooth, ja toiseksi edullinen hinta ja saatavuus vaikuttivat valintaan.

Valitsin Raspberry Pi:n käyttöjärjestelmäksi NexMon-nimisen järjestelmän, koska siinä on skannaukseen sopivammat ajurit verrattuna suositun Raspbian-käyttöjärjestelmän Broadcom-ajureihin. Broadcom sen sijaan on valmistanut Raspberry Pi 3 Model B:n järjestelmäpiirin, jossa on myös WLAN- ja Bluetooth-radiot. Broadcom on valitettavan tunnettu huonosta Linux-tuesta, eikä sen vakioajureissa ole esimerkiksi skannausominaisuuksia.

NexMon- kuten myös Raspbian-järjestelmät pohjautuvat Debian GNU/Linux -käyttöjärjestelmään, mistä syystä erot näiden järjestelmien välillä ovat melko pieniä. Tosin eroja voi olla, mikäli järjestelmät pohjautuvat Debianin eri versioihin.

Debian puolestaan on yksi vanhimmista vielä toiminnassa olevista Linux-järjestelmistä. Sen kehitys alkoi projektin historian [3] mukaan jo vuonna 1993.

Raspberry Pi:n moduulit

Raspberry Pi 3:lle on olemassa monia erilaisia moduuleita. Osasta niistä saattaisi olla hyötyä tämänkaltaisissa menetelmissä.

Ensimmäisenä käsittelen Pi PoE Switch HAT -nimistä Power over Ethernet -moduulia, joka mahdollistaa virran saamisen verkkokaapelin avulla. Virta siirretään 802.3af-standardin tekniikalla, joka on IEEE:n standardi virran siirtoon Ethernet-kaapelissa.
Hyötyä moduulin käytöstä syntyy, kun verkko- ja virtakaapelin sijaan tarvitaan vain verkkokaapeli Raspberrylle. Tätä kautta mittareiden asennus voi helpottua. Moduulin hyödyntäminen tosin edellyttää PoE-virtaa syöttävää verkkolaitetta, joka tukee samaa protokollaa (802.3af) kuin Pi PoE Switch HAT.

Raspberry Pi -tietokoneissa ei ole omaa kelloa, vaan kello joudutaan ohjelmallisesti simuloimaan esimerkiksi fakeClock-ohjelman avulla. Vaihtoehtoisesti voidaan käyttää RTC (real time clock) -reaaliaikamoduulia kellon ajan ylläpitämiseen. Tämä moduuli tekisi ohjelman tarpeettomaksi ja parantaisi kellonajan tarkkuutta. Vaikka käytössä olisi reaaliaikaa ylläpitävä laajennus, on RPi hyvä synkronisoida vielä NTP-palvelujen kautta. Esimerkiksi Mini RTC -moduuli on tähän tarkoitukseen sopiva. Mini RTC:n hyötynä on

myös parempi kyky säilyttää kellonaika lähellä reaaliaikaa, mikäli verkko-ongelmat rajoittavat yhteyttä NTP-palvelimiin.

Kumpikin edellä mainittu moduuli on mahdollista ottaa käyttöön samanaikaisesti, tosin yksittäisen moduulin tuomat hyödyt eivät erotu niin selvästi, mikäli kumpikin moduuli on käytössä.

2.3 Meshlium Scannerin ja Raspberry Pi 3 Model B:n erot työn kannalta

Meshlium Scannerin suurimpana puutteena on suppeampi kerätty data verrattuna RPi-hin. Esimerkiksi duplikaattilaitteiden tunnistus (sama laite, eri MAC-osoite) on erittäin vaikeaa Meshlium Scannerin tallentaman datan pohjalta, koska se ei tallenna päätelaitteen pyytämiä WLAN-verkkoja (SSID).

Meshlium Scannerin suurimpana etuna RPI3:een verrattaessa ovat valmiit ominaisuudet laitteiden havaitsemiseen. Tämä jälkeen oleellisena ominaisuutena on tieto, onko päätelaite yhdistetty johonkin langattomaan verkkoon (MAC-osoite ei muutu) vai ei (MAC-osoite voi muuttua, mikäli päätelaitteella on satunnaistaminen päällä). Etuna voi myös nähdä tuen erilaisiin pilvipalveluihin, joihin tiedon voi tallentaa ja palvelusta riippuen suorittaa tarvittaessa analytiikkaa datan pohjalta.

Raspberry Pi 3B:n suurin puute on, että kyseessä on vain korttitietokone, joka ei sisällä käyttöjärjestelmää eikä muuta ohjelmistoa. Samalla tämä RPi3:n suurin haitta on nähtävissä myös etuna, jolloin laitteiston päälle voi valita haluamansa järjestelmän sekä ohjelmistot käyttöympäristön mukaisesti.

RPi3B:llä voidaan kerätä ja tallentaa kaikki WLAN-radioliikenne, joka tapahtuu 2,4 GHz:n alueella. Tällöin mahdollisuus tunnistaa yksilöityjä laitteita on suurempi kuin Meshlium Scannerilla, vaikka päätelaite käyttäisi MAC-osoitteen satunnaistamista, ja tätä kautta on paremmat edellytykset saada tarkempaa tietoa tilan käyttöasteesta. Insinöörityössä käytetyt mittalaitteet on esitetty kuvassa 2.

Kuva 2. Vasemmalta oikealle Meshlium Scanner, Raspberry Pi 3 Model B ja tulitikkurasia mittakaavan hahmottamiseen.

3 NTP-protokolla mittareiden ajan synkronisointiin

NTP on kellonajan synkronisointiin tarkoitettu protokolla. Se on käytössä työssä niin RPi:ssä kuin myös virtuaalikoneessa, jossa testiympäristön tietokanta sijaitsee.

Mittareiden osalta NTP-protokolla on oleellinen, jotta mittareiden havaitsema data kirjoitetaan tietokantaan mahdollisimman lähellä oikeaa aikaa. Ilman NTP:tä mittarit alkavat käydä ”omaa” aikaansa, jolloin tietokannasta haettaessa dataa ei voida pitää täysin oikeana ja todellisuutta vastaavana. Sen lisäksi, että mittareilla on NTP käytössä, on oleellista, että kellonaika synkronisoidaan vähintään yhdeltä samalta palvelimelta, jonka lisäksi voi olla useita palvelimia, joista synkronisointi tehdään. NTP-protokollan tarkkuus on noin kymmenkunta millisekuntia liikenteessä, joka kulkee internetissä. Lähiverkoissa tarkkuusero on jopa alle yhden millisekuntin luokkaa oikeaan aikaan. Mikäli NTP:n tarkkuus ei riitä, tarkempaan kellonajan synkronisointiin on olemassa PTP-protokolla, jolla tarkkuus voi olla EndRunTechnologiesin julkaisun mukaan jopa 30 nanosekuntia. [4.]

Työn kannalta ─ henkilömäärän arvioinnissa ─ millisekuntien tarkkuus on täysin riittävä.

4 Työssä käytetyt langattomat tekniikat

4.1 WLAN

WLAN-verkko rakentuu yleensä tukiasemista ja päätelaitteista, mutta näin ei välttämättä aina ole; esimerkiksi Ad Hoc tai Wi-Fi Direct -tyyppiset verkot toimivat ilman perinteistä tukiasemaa [5].

Miksi tutkimukseen valittiin WLAN-tekniikka? WLAN on yksi yleisimmistä ja eniten käytetyistä langattomista tekniikoista, ja useat päätelaitteet tukevat sitä tästä syystä.

WLAN-asiakaslaitteet lähettävät luotainkehyksiä (probe frame), joita tilaan sijoitetut mittarit pyrkivät havaitsemaan. Havaitsemiseen käytetään niin Libeliumin Meshlium Scanneria sekä Raspberry Pi 3 Model B -korttitietokonetta. RPi 3:n tapauksessa pyritään havaitsemaan probe request -tyyppisiä kehyksiä, koska oleellista on päätelaitteiden eikä tukiasemien lähettämien probe response -kehysten havaitseminen. Mittarien keräämä data tallennetaan yhteiseen tietokantaan. Probe response -kehysten havaitsemisesta on hyötyä, kun havainnointia tehdään tilassa, jossa on WLAN-tukiasemaverkosto ja johon ei voi sijoittaa montaa mittaria. Probe response -kehyksistä on hyötyä muun muassa summittaisessa sijainnin määrityksessä.

4.2 Bluetooth

Bluetooth on matalan energiamäärän langaton tekniikka ─ etenkin, kun verrataan WLAN-tekniikkaan, ─ joka aikoinaan kehitettiin korvaamaan RS-232-sarjaportti-nimeä kantava tiedonsiirtoportti. Bluetooth jakautuu kahteen teknologiahaaraan, Basic Rate/Enhanced Data Rate (BR/EDR) ja Low Energy (LE). BR/EDR soveltuu ainoastaan kahdenlaitteen välisiin yhteyksiin (1:1). LE-tekniikalla voidaan luoda kahden laitteen välisiä yhteyksiä samaan tapaan kuin BR/EDR-tekniikalla, minkä lisäksi laitteesta useaan laitteeseen- tai useasta laitteesta useaan laitteeseen -tyyppiset yhteysvaihtoehdot ovat mahdollisia. [9.]

Bluetoothin osalta skannauksessa on otettava huomioon, että skannattavan laitteen Bluetoothin tila pitää olla näkyvillä (discoveryMode = on), jotta kyseinen laite voidaan standardin mukaisilla laitteilla havaita.

Tästä poiketen BlueHydra-ohjelma pystyy tunnistamaan myös niin sanotut piilotetut laitteet [6]. Nämä laitteet lähettävät dataa lähes samaan tapaan, kuin mikäli laite ei olisi piilotettuna: laitteiden lähettämä radiosignaali ei signaali- tai radioteknisesti eroa piilotetun ja piilottamattoman laitetyypin välillä.

4.3 WLAN:n ja Bluetoothin erot tilankäyttöasteen arvioinnin kannalta

WLAN-tekniikan etu on tekniikkaan kuuluvat verkon langattomat tukiasemat. Ensinnäkin, jos laite on yhdistynyt WLAN-verkkoon, sen MAC-osoite ei muutu. Toiseksi, tukiasemista on se hyöty, että oikein sijoitettuna skannausta suorittavia mittareita tarvitaan vähemmän. Bluetooth-tekniikassa ei ole tukiasemia, vaan yhteydet ovat suoraan laitteiden välisiä. Bluetooth-tekniikkaa käyttävät laitteet, joissa on satunnaistamisominaisuus, voivat satunnaistaa MAC-osoitteen, kunhan kyseinen laite ei ole Bluetoothin kautta yhteydessä muihin laitteisiin ja tekniikka on käytössä. Satunnaistamista ei tapahdu, mikäli laite on yhteydessä jonkin toisen laitteen kanssa.

Vaikka tekniikat toimivat samoilla taajuuksilla (2,4 GHz) ja niin WLAN- kuin myös BT-tekniikoiden maksimilähetysteho voi olla sama 100 mW (20 dBm), tämä koskee vain BT-tekniikan luokan 1 laitteita. Luokkien 2 ja 3 lähetystehot ovat selvästi pienempiä, jolloin niiden havainnointi on hankalampaa lyhyemmän kuuluvuusetäisyyden takia. [16.]

Lisäksi Bluetooth-skannauksessa skannaus perustuu kyselyyn eikä passiiviseen kuunteluun kuten WLAN-tekniikalla. Päätelaitteet voivat olla vastaamatta kyselyyn, jolloin havaitsemista ei tapahdu, tai päätelaitteet voivat vastata kyselyyn, jolloin havainnointi tapahtuu, mutta kaksisuuntaisen tiedonvaihdon vuoksi Bluetooth on hitaampi kuin WLAN-tekniikka. Kaksisuuntaisen skannauksen vuoksi Bluetooth-laitteiden havainnointi voi olla hankalampaa, koska päätelaitteet eivät välttämättä vastaa skannauskyselyyn.

Toisaalta havaittavan päätelaitteen sijainnin määritys voi helpottua, mikäli laite havaitaan Bluetooth-tekniikalla, jolloin etäisyyden arviointi voi olla helpompaa pienemmän kuuluvuuden vuoksi. Toisaalta, mikäli laite havaitaan WLAN-tekniikalla ja havaitusta datasta huomataan, että lähellä oleva tukiasema on keskustellut laitteen kanssa, voidaan arvioinnissa hyödyntää tietoa, että tukiasema on keskustellut päätelaitteen kanssa. Silloin

voidaan päätelaitteen etäisyyden määrittämisessä hyödyntää niin mittarin kuin tukiaseman kuuluvuusalueita. Päätelaitteen sijainti ja etäisyys tukiasemaan nähden on tukiaseman täyden kuuluvuusalueen sisällä. Kun tiedetään vastaanotetun signaalin voimakkuus, voidaan rajata säde, jolla havaittu päätelaite on, ja hyödyntää niin mittarin kuin tukiaseman leikkausta päätelaitteen etäisyyden ja sijainnin määritykseen.

Kummankin tekniikan haittapuolena tilankäyttöasteen mittaamisen kannalta on MAC-osoitteen satunnaistaminen. Bluetoothin haasteena ovat lisäksi tilanteet, joissa näkyvyys (discoveryMode) on asetettu piilotetuksi, jolloin standardin mukaiset mittarit eivät pysty havaitsemaan liikennettä. [7.]

5 Tilankäyttöasteen arviointi

Tilankäyttöasteen arviointi tehdään havaittujen laitteiden pohjalta. Havainnointi puolestaan toteutetaan RPi:n osalta skannausskriptin avulla. Meshlium Scanner toimii jo sellaisenaan, mutta jos RPi:n ja Meshlium Scannerin dataa halutaan helposti esimerkiksi vertailla, on järkevintä käyttää samaa tietokantaa, joka ei sijaitsisi kummassakaan laitteessa.

Duplikaattilaitteella tarkoitan laitetta, joka on satunnaistanut uuden MAC-osoitteen. Näitä laitteita, jotka satunnaistavat uuden MAC-osoitteen, on yhä enemmän suhteessa laitteisiin, jotka sitä eivät tee. Esimerkiksi vuosituhannen vaihteessa laitteita, jotka satunnaistivat uuden MAC-osoitteen, oli niin vähän, että ne eivät välttämättä edes olisi näkyneet arvioiduissa lopputuloksissa. Nykyään asia on hieman toisin [7], sillä esimerkiksi Apple käyttää satunnaistamismenetelmää IOS-mobiilikäyttöjärjestelmässään versiosta 8 lähtien.

IOS on Applen mobiililaitteiden, kuten iPhone ja iPad, käyttöjärjestelmä. Applen lisäksi markkinoilla on kolmannen osapuolen järjestelmiä, jotka tekevät satunnaistamisen. Lisäksi joissain laitteissa ja käyttöjärjestelmissä käyttäjän on mahdollista muuttaa manuaalisesti osoite toiseksi. Kaikki MAC-osoitteen satunnaistamisjärjestelmät tosin nollautuvat uudelleenkäynnistyksen yhteydessä, joten osoitteen satunnaistaminen pitää toteuttaa uudelleen. Tämä johtuu siitä, että piireille, joista alkuperäinen MAC-osoite luetaan, ei voi kirjoittaa, joten aina on aloitettava ns. alusta.

Mobiilijärjestelmiä käsitellään myös luvussa 6 MAC-osoitteen satunnaistaminen.

5.1 Testiympäristö

Insinöörityö aloitettiin testiympäristössä, joka koostui mittareista RPI ja Meshlium Scanner, jotka olivat kytkettynä samaan verkkoon. Mittareiden lisäksi ympäristöön kuului VirtualBox-virtualisointiympäristössä toimiva Ubuntu-järjestelmä, jossa puolestaan MariaDB-tietokanta sijaitsi. Virtuaaliympäristön ja mittareiden lisäksi käytössä oli kaksi Windows 7 -konetta, jossa toisessa kehitettiin visualisointisovellusta. Ympäristö sijaitsi Metropolia Ammattikorkeakoulussa Leppävaaran kampuksella olleessa projektitilassa. Projektin edetessä suoritettiin mittauksia useassa muussa paikassa. Muualla suoritettujen mittausten ympäristö sisälsi aina RPi:n ja joskus Meshlium Scannerin. Paikallisesta verkkoympäristöstä riippuen käytössä oli tietokantapalvelin, tai sitä ei ollut. Data analysoitiin mittauskauden jälkeen.

5.2 Testausskannaukset

Raspberry Pi 3 Model B:n testausskannauksia tehtiin niin Metropolian tiloissa kuin yksityisasunnoissa. Tulokset olivat hämmentäviä. Esimerkiksi yksityisasunnossa A skanneri tunnisti kahden vuorokauden aikana yli 30 eri MAC-osoitteella olevaa laitetta, vaikka asunnossa oli tuolloin vain neljä havaittavissa olevaa laitetta. Metropolian tiloissa mittauslukemat vaihtelivat eripituisten ja eri vuorokaudenaikaan ajoittuvien mittausjaksojen aikana 90 laitteesta lähes tuhanteen. Ongelma tuli siitä, että tarkka henkilömäärä alueella eli Metropolian Leppävaaran kampuksella olevien laitteiden määrä tai MAC-osoitteen satunnaistavien laitteiden määrä eivät olleet tiedossa.

Tämän vuoksi suoritettiin uusi testausskannaus melko eristyksissä olevalla kesäasunnolla, jossa henkilö- ja laitemäärä oli tiedossa koko skannauksen ajan. Samoin etäisyys naapureihin ja ohikulkijoihin oli sen verran pitkä, että mittari ei heidän laitteitaan pystyisi havaitsemaan tai mikäli ne havaittaisiin, ne pystyisi erottamaan vastaanotetun signaalin (RSSI) erittäin heikon voimakkuuden perusteella. Skannausaika oli kaksi tuntia. Tänä aikana mittari tunnisti uniikkeja MAC-osoitteita 13. Koska alueella oleva henkilömäärä, joka oli viisi henkilöä, tiedettiin, saadaan kalibrointiarvoksi $\frac{13}{5} = 2{,}6$. Tämä tarkoittaa, että yhdellä henkilöllä olisi keskimäärin 2,6 havaittavaa laitetta. Todellisuudessa henkilöiden

yhteenlaskettu RPi 3 -skannerilla havaittavien laitteiden määrä oli 7, ja se jakaantui henkilöstä riippuen nollan ja kahden laitteen välillä, mikä ei suoraan täsmää arvon 2,6 kanssa. Tästä voidaan todeta, että yksi tai useampi laitteista on generoinut MAC-osoitteen uudelleen yhden tai useamman kerran. Mitä lähempänä kalibrointiarvo on todellista tilannetta, sitä tarkempia tuloksia voi olettaa saatavan.

5.3 Rasperry Pi:n WLAN-laitteiden skannausskripti

Skannaus toteutettiin skriptillä, joka on esitelty liitteessä 1. Skriptin oleellisin komento on tcpdump-ohjelmaa kutsuva rivi, jossa määritellään, minkäkaltaista dataa halutaan saada tcpdump-ohjelman tuloksena, mikäli dataa on saatavilla. Liitteessä esiteltiin mm. tämä rivi: ”tcpdump -i wlan0 -G 13 -W 1 -e -tt type mgt subtype probe-req -j host -w RawData”.

Rivillä kutsutaan tcpdump-ohjelmaa [10]. Kutsun yhteydessä annetaan ohjelman lisäksi parametrejä, jotta saadaan haluttu lopputulos.

Parametreissä määritellään

- verkkoliitäntä: wlan0
- skannauksen kesto: 13 sekuntia
- lopputuloskierrosten määrä: 1
- linkkitason otsikkotiedot: -e
- aika: -tt
- haluttu data: -type
- aikaleiman tyyppi: -j
- tulosten kirjaus tiedostoon: -w

Mikäli halutaan lopputulokseen myös WLAN-tukiasemien vastaukset, on -type -parametrin arvoa muokattava esimerkiksi seuraavanlaiseksi: type mgt subtype probe-req or subtype probe-resp. Tämä voi auttaa päätelaitteen sijainnin määrityksessä.

Skripti käynnistetään automaattisesti käyttöjärjestelmän käynnistyksen yhteydessä. Automaattinen käynnistys hoidettiin lisäämällä */etc/rc.local*-tiedostoon rivit

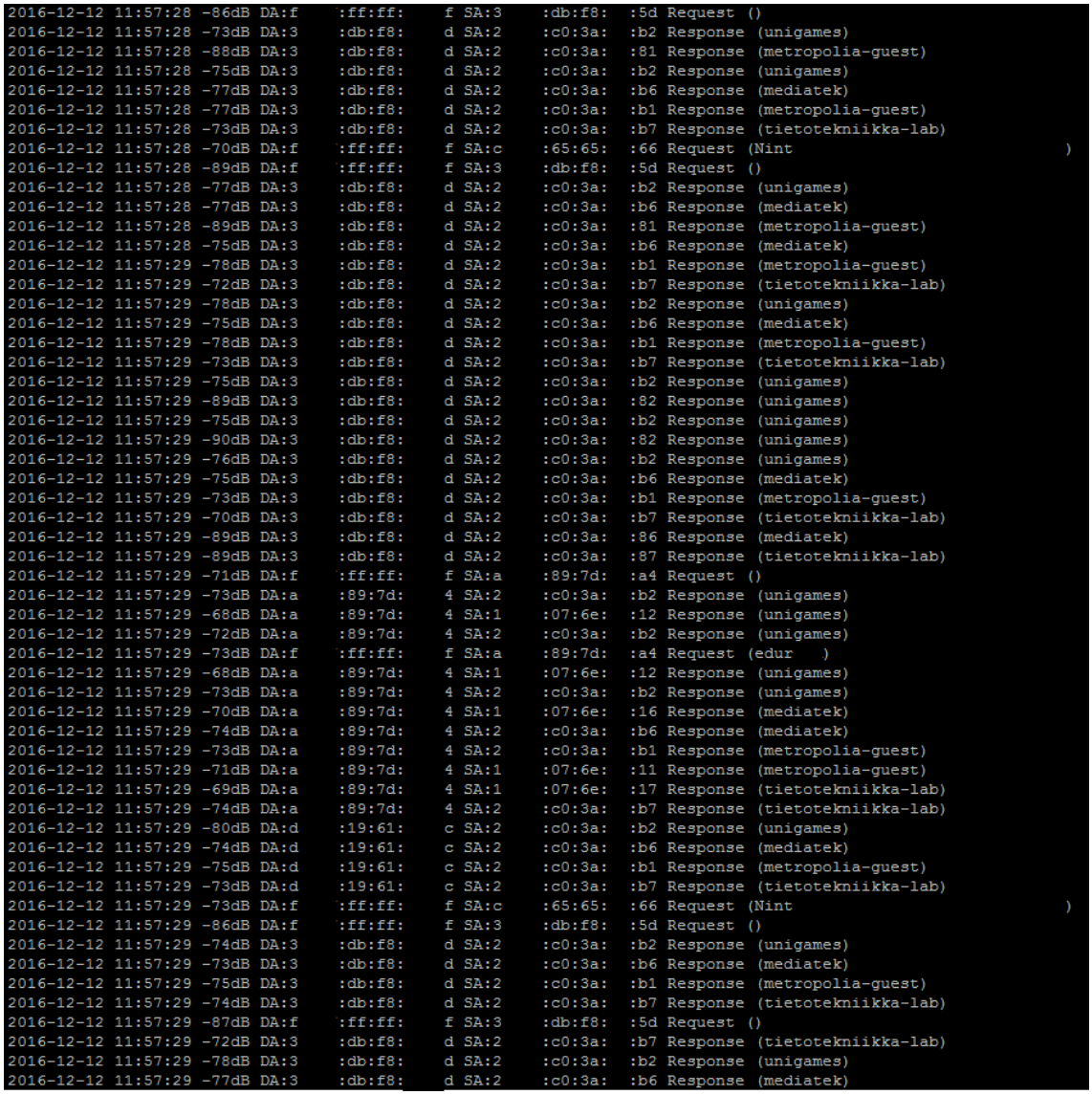
```
insmod /root/brcmfmac.ko
sleep 1
ifconfig wlan0 up
sleep 1
nice --18 sh /scan/WLANSCAN/scan.sh &
```

Ensimmäinen rivi lataa NexMon-projektin ajurit käyttöön. Toisella rivillä odotetaan yksi sekunti, jotta ajurit ovat varmasti päällä. Kolmannella rivillä kytketään WLAN päälle. Neljännellä rivillä odotetaan yksi sekunti, jotta WLAN on varmasti päällä. Viidennellä rivillä käynnistetään itse skripti.

Nice-komennon avulla voidaan asettaa ohjelman mukavuutta järjestelmän näkökulmasta, skaala on -20 – (+19). Arvolla -20 ohjelma ei ole ”mukava”, joten sille annetaan mahdollisimman paljon järjestelmäresursseja, kuten suoritusaikaa. Toisessa ääripäässä, +19, ohjelma on järjestelmän kannalta ”mukava”, joten sen suoritusajan määrä on vähäisempää kuin toisessa ääripäässä. Oletuksena ohjelmalla on prioriteetti nolla. Rivin lopussa oleva &-merkki tarkoittaa, että komento suoritetaan taustalla.

Skannausskriptillä saatuja mittaustuloksia ja arvioita tuloksista

Kuvasta 3 voidaan havaita, että vain muutaman päätelaitteen probe-request -kyselyt ja kyselyiden probe-response -vastaukset saavat aikaan muutaman sekunnin aikana melko paljon WLAN-mittarilla havaittavaa dataa.



Kuva 3. Raspberry Pi:n scannausskriptin avulla havaittua dataa.

Ensimmäisenä kuvasta löytyy päivämäärä ja kellonaika, jolloin havaitseminen mittarin näkökulmasta on tehty. Seuraavana näkyy havaitun signaalin voimakkuus desibelimuodossa. Merkeillä ”DA:” alkava kenttä, joka loppuu juuri ennen merkkejä ”SA:”, on havaitun signaalin kohde-MAC-osoite. Merkeillä ”SA:” alkava ja ennen Request- tai Response-sanoja on puolestaan lähettävän laitteen osoite. Request on päätelaitteen lähettämä viestikehys, jossa kysytään verkkoa, johon voisi yhdistää. Response puolestaan on tukiasemien vastaus päätelaitteen kyselyyn ja sisältää verkon nimen (SSID) vastauksessa. Päätelaite voi kysyä verkkoja kahdella tavalla. Kohdentamattomassa tavassa päätelaite lähettää kyselyn (probe request), jossa SSID-tieto on tyhjä (kuvassa pelkät sulkeet). Tällaisen tilanteen voi nähdä kuvan ensimmäiseltä rivitä. Toinen päätelaitteiden

kyselytapa on kohdennettu kysely, jossa päätelaite kysyy tiettyä verkkoa (SSID). Tällainenkin tilanne löytyy kuvasta, esimerkiksi riviltä 8.

Kun probe-request on kohdentamaton, kaikki tukiasemat, jotka havaitsevat kyseisen kehyksen, vastaavat siihen kertomalla itsestään; kuvassa esimerkiksi rivillä 2, 3, 4, 5 ja 6. Kun probe-request-kehykset on kohdistettu, tukiasema, joka havaitsee kehyksen, tarkistaa, ”palveleeko” se kyseistä verkkoa. Mikäli palvelee, kyseinen tukiasema vastaa. Mikäli tukiasema ei palvele kyseistä verkkoa, tukiasema ei vastaa kyselyyn. Päätelaite lähettää probe-request-kehykset aina MAC-osoitteeseen FF:FF:FF:FF:FF:FF eli broadcast-osoitteeseen, jolloin kaikki laitteet, jotka havaitsevat kehyksen, voivat sen vastaanottaa jatkoprosessointia varten.

Työn WLAN-menetelmän kannalta nämä probe-request-kehykset ovat olennaisimpia kokonaisuuden kannalta. WLAN-menetelmä on WLAN/IEEE 802.11 -standardin mukainen eikä edellytä porsaanreikien etsimistä.

6 MAC-osoitteen satunnaistaminen

Satunnaistaminen eri mobiilikäyttöjärjestelmillä

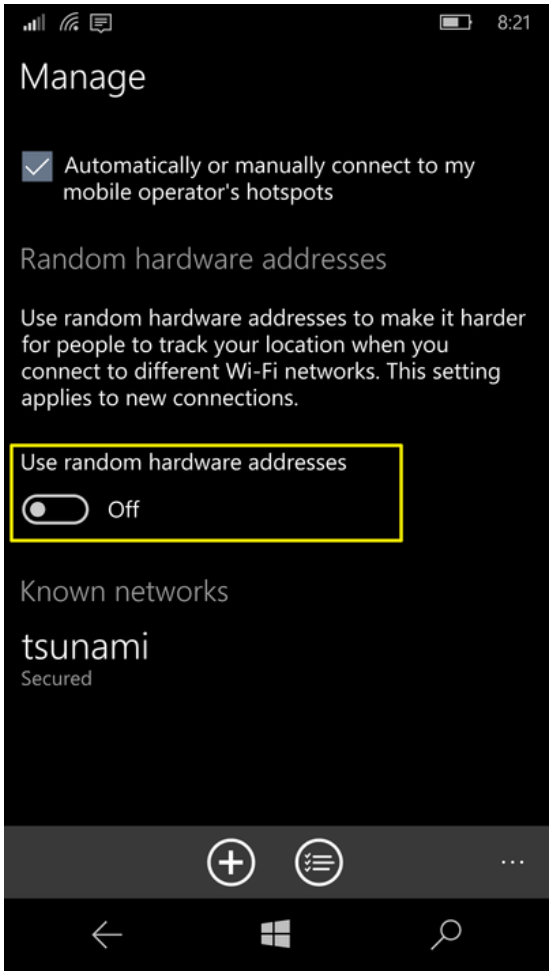
Kolmannen osapuolen sovelluksia MAC-osoitteen satunnaistamiseen on ollut markkinoilla jo jonkin aikaa, mutta vuoden 2013 aikana suureen julkisuuteen tulleiden tietovuotojen jälkeen on havaittavissa, että merkittävät kansainväliset yritykset ovat alkaneet ottaa satunnaistamisominaisuuksia käyttöön omissa järjestelmissään.

Windows 10 Mobile

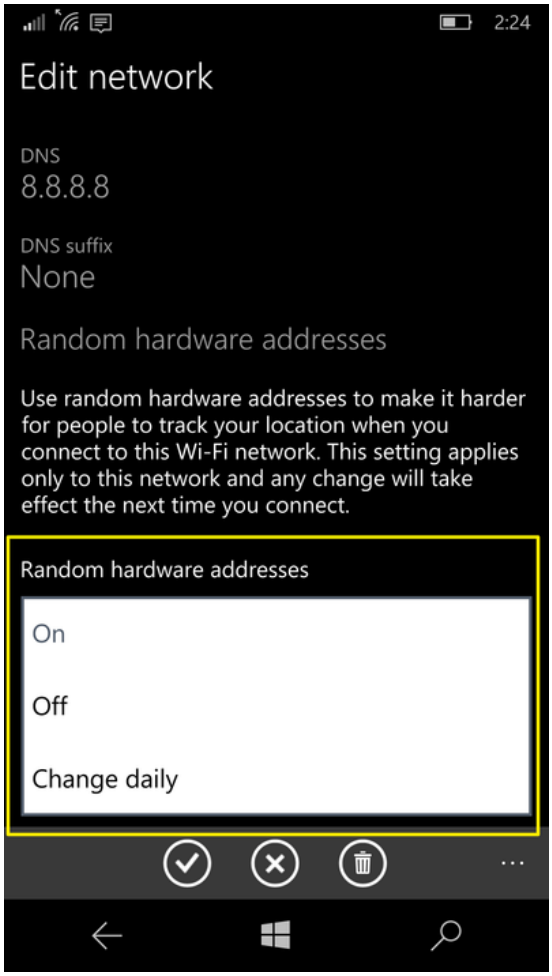
Microsoft julkaisi satunnaistamisominaisuuksilla varustetun Windows 10 Mobilen vuoden 2015 lopussa.

Windows 10 Mobile -käyttöjärjestelmässä satunnaistamisominaisuus voi olla kolmessa eri tilassa: käytössä, muuta päivittäin tai pois käytöstä [11]. Mikäli ominaisuus on käytössä, käytetään kaavaa SHA256(SSID, oikeaMACOsoite, yhteysId, salasana) MAC-osoitteen laskemiseen. Satunnaistamisominaisuuden oletustilasta ei löytynyt virallista dokumentaatiota. [12; 13.]

Tilojen muuttaminen tapahtuu kuvien 4 ja 5 mukaisesti.



Kuva 4. Satunnaistamisominaisuuden päälle/pois-kytkentä.



Kuva 5. Satunnaistamisominaisuuden tila-asetukset.

iOS

Vuoden 2014 lopussa julkaistun Applen iOS-järjestelmän versiosta 8 lähtien MAC-osoitteen satunnaistamisominaisuus on ollut olemassa.

Applen julkaisemaa virallista dokumentaatiota siitä, kuinka usein MAC-osoite satunnaistetaan, ei löytynyt. Mittausten perusteella satunnaistaminen tapahtuu satunnaistamisen ollessa mahdollista eli silloin, kun laite ei ole kytkettynä langattomaan verkkoon WLAN:n ollessa päällä. Samoin näytön pitää olla päällä, mutta ei kuitenkaan käytössä, ja sijaintipalvelujen pitää olla pois päältä. Tällöin jokaisen lähetettävän probe-kehyksen lähettävän laitteen MAC-osoite on satunnaistettu.

Virallista dokumentaatiota siitä, että MAC-osoitteen satunnaistaminen on oletusarvoisesti päällä, ei löytynyt. Syy tähän lienee, että ominaisuudelle ei ole Windows 10 Mobilen

kaltaista on/off-vipua, josta ominaisuuden tilaa voisi muuttaa, vaan kyseessä on useasta asiasta riippuvainen ominaisuus, jonka tila riippuu niin laitteen (näyttö) kuin myös asetusten tilasta (ei yhdistettynä WLAN-verkkoon eivätkä sijaintipalvelut ole käytössä).

Android

Vuoden 2015 loppupuolella julkaistusta Androidin versiosta 6 lähtien järjestelmässä on ollut satunnaistamisominaisuus sisäänrakennettuna, mutta ominaisuus on myös saatavilla versioon 5 laajennuksena (incremental patch). Satunnaistamisominaisuus on virallisen dokumentaation mukaan käytössä oletuksena. [17.] Käyttäjän mahdollisuus vaikuttaa satunnaistamisominaisuuden tilaan (päällä/pois) on pienin Android-järjestelmissä ja suurin Windows 10 Mobilessa. iOS-järjestelmien käyttäjien vaikutusmahdollisuus on tältä väliltä.

Ohjelmistotason omaisuuden lisäksi laitteiston ja laitteiston kanssa keskustelevien ajureiden pitää tukea ominaisuutta, jotta sitä voidaan käyttää. Androidin lisäksi tämä koskee myös Microsoftin ja Applen järjestelmiä. Koska Apple kehittää ohjelmistot ja suunnittelee puhelimet, ei sen laitteissa ongelmia syntyne. Microsoft puolestaan kehittää järjestelmän, mutta Windows 10 Mobile -laitepuolella ei montaa uutta laitetta ole kehitetty Microsoftin tai muidenkaan laitevalmistajien osalta: yhteensopivuusongelmia ei tästä syystä pääse syntymään. [11.]

7 Havaitun datan visualisointi

Havaitun datan visualisointi ja analytiikka toteutetaan Microsoftin ASP.NET-tekniikalla tehdyn sovelluksen avulla. Se hakee dataa tietokannasta ennalta määriteltyjen rajauksien mukaan. Sovellus voi toimia Microsoftin modernilla Windows-työasemalla, palvelinympäristössä IIS-web-palvelun avulla tai Microsoftin pilvessä (Azure) web-palveluna. Sovelluksesta lisää visualisointisovellusosiossa (s. 18).

Analytiikan toiminnan kannalta on tärkeää, että käytössä olevat mittareiden kellot on synkronoitu, jotta aikaan pohjautuvat tapahtumat ovat synkronoituja keskenään, esimerkiksi niin, että kohteessa on mittareiden lisäksi NTP-palvelin (kellonaikaa tarjoava palvelu), jotta mittareiden kellot ovat synkronoituja keskenään kyseisen kohteen alueella.

NTP-palvelulla varustettu tietokone voisi puolestaan synkronisoida kellonsa yhteisestä kellonaikapalvelusta sekä muutamasta julkisesta NTP-palvelusta.

Havaittujen laitteiden etäisyyden arviointi

Laitteen ja sitä kautta mahdollisen käyttäjän etäisyyden arviointi on toteutettavissa esimerkiksi seuraavalla tavalla: muunnetaan vastaanotetun laitteen signaalin voimakkuus metreiksi kaavan $RSSI = -(10nlog_{10}d + A)$ avulla [14]. Kaavassa n on signaalin etenemisvakio tai eksponentti, d on etäisyys lähettäjään ja A on vastaanotettu signaali 1 metrin etäisyydeltä. Vastaanotetun signaalin voimakkuus (RSSI), n ja A tiedetään, jolloin kaavaa voidaan soveltaa niin, että d saadaan selville. Kaavan antama etäisyys ei välttämättä vastaa täysin todellisuutta, muun muassa mahdollisten radiotien häiriöiden vuoksi.

Funktio toimii etäisyyden määrittämiseen, mikäli A-kenttään on syötettävissä arvo. Mitä tarkemmin A:n arvo kuvaa todellisuutta, sitä tarkemmin funktio palauttaa arvioidun etäisyyden säteen mittarista. Kaava toimii myös niin, että asetetaan eri lukuja d:n arvoksi. Mitä lähempänä kaavan palauttama RSSI on havaittua signaalin voimakkuutta, sitä lähempänä d:n arvo on oikeaa etäisyyttä.

Kaava toimii parhaiten, mikäli laitteen ja mittarin välillä on esteetön signaalin kulku ja alueella ei ole häiriölähteitä. Mikäli signaalilla ei ole esteetöntä kulkua, tarkkuus vaihtelee mm. riippuen esteen vaikutuksesta signaalin voimakkuuteen sekä mittarin ja havaittavan laitteen etäisyydestä. Yhdellä mittarilla voidaan arvioida etäisyyttä hyvin karkeasti, mutta ei suuntaa, ellei voida olla varmoja siitä, että signaalia ei voida vastaanottaa tietystä tai tietyistä suunnista.

Mikäli usea mittari havaitsee saman laitteen, suoritetaan edellä kuvatut toimenpiteet kaikille mittareille, jotka laitteen havaitsevat. Mitä useampi mittari laitteen havaitsee, sitä tarkemmaksi laitteen sijainnin määritys käy.

Ideaalitilanteessa, jossa minkäänlaisia häiriötekijöitä ei esiinny, voidaan etäisyys määrittää melko tarkasti kahdella mittarilla. Tällöin havaittu laite on kummankin mittarin arvioidun etäisyyden leikkauksessa. Mikäli havainnointi tehdään kolmella mittarilla, havaittu laite sijaitsee kaikkien kolmen mittarin arvioidun etäisyyden leikkauksessa. Mittareiden lisäys tästä eteenpäin ei lisää merkittävästi tarkkuutta etäisyyden arvioinnin osalta, mutta

kasvattaa hieman vikasietoisuutta. Lisäksi useilla mittareilla toteutettu havainnointi mahdollistaa tarkemman kolmiulotteisen etäisyyden arvioinnin, mikä voi olla tehokasta, mikäli halutaan tietää, millä vertikaalisella tasolla käyttäjä on.

Jotta tätä tietoa voidaan hyödyntää tehokkaasti, pitää tietää mittareiden etäisyydet toisistaan tai niiden sijainnit laskentaa varten.

Visualisointisovellus

Visualisointisovellus toimii käyttöliittymänä niin järjestelmän datalle kuin datan visualisoinnille. Sovellus analysoi myös havainnot.

Sovellus tarjoaa erilaisia lähinnä datan visualisointiin keskittyneitä näkymiä, joiden kautta operaattori voi katsoa esimerkiksi, mikä on tilan arvioitu käyttöaste tietyssä osoitteessa.

Duplikaattien havaitseminen rajoittuu WLAN-tekniikkaa käyttäviin päätelaitteisiin. Tunnistus on toteutettu SSID (service set identifier) -tiedon eli langattoman verkon tunnuksen havaitsemisella. Esimerkiksi, mikäli laite, jonka MAC-osoite (Media Access Control) olisi esimerkiksi 02:00:00:00:00:00, on lähettänyt probe-requesteja verkkoihin (Koti_Verkko, KaupunginJulkinenWLAN, OmaPuhelin sekä WorkNet) ja lyhyen ajan sisällä toinen laite, esimerkiksi laite MAC-osoitteella FF:FF:FF:FF:01:02, on lähettänyt probe-request-kehyksiä samoihin verkkoihin ja ensimmäistä laitetta ei havaita, voidaan olettaa, että nämä kaksi eri MAC-osoitetta ovatkin sama laite.

8 Havaitun tiedon käsittely

8.1 SQL-kyselyt

Havaittu data tallennetaan tietokantaan SQL insert -käyttökäskyn avulla. Sovellukseen data haetaan SQL-kyselyiden select-käyttökäskyn kautta. Vaihtoehtoinen tapa hakea dataa olisi käyttää SQL-näkymää datan hakemiseen. Näkymään voi suoraan yhdistää oleelliset tiedot niin tietokannan WLAN- kuin myös Bluetooth-tauluista. Näkymään lisätään lisäksi kokonaisluku kuvaamaan datan lähdettä (1 = Bluetooth, 2 = WLAN), jotta voidaan rajata kyselyt tarvittaessa lähteen mukaan. Eroavat kentät näkymässä saavat oletuksena NULL-arvon, mikäli arvoa ei ole saatavilla.

Esimerkki WLAN-tekniikalla havaittujen laitteiden hakemisesta SQL-käyttökäskykomennolla:

```
SELECT MAC, TimeStamp, UPPER(SHA2(SSID, 256)) as SSID, MeshliumID
FROM MeshliumDB.wifiScan
WHERE MAC REGEXP '^([0-9A-F]{2}:){5}[0-9A-F]{2}$'
 AND TimeStamp between ("2017-11-13 00:00:00" - INTERVAL 60 DAY) AND ("2017-
11-13 00:00:00" + INTERVAL 0 DAY)
```

Yllä on esitelty esimerkki kyselystä, jolla voidaan hakea 60 päivän ajalta kaikki WLAN-data, jossa palautettu SSID-tieto on kertaalleen anonymisoitu SHA2-tiivistefunktion kautta ja jossa MAC-osoitteen pitää olla oikeassa formaatissa. Tuotannossa anonymisoinnin voi ja joissain tilanteissa se kannattaa toteuttaa useaan kertaan, mikä hankaloittaa alkuperäisen datan tunnistamista. Anonymisointi parantaa laitteen ja sitä kautta yksittäisen henkilön anonymiteettiä erityisesti hankaloittamatta analyysiä. Useaan kertaan anonymisoidun datan ainoa varsinainen haitta on anonymisoinnin vaatima laskentatehon tarve, joka puolestaan näkyy suoraan kyselyn kestossa.

Suojattu tietokantayhteys

Mittareiden lähettämä data kulkee oletuksena suojaamattomana tietoverkossa mittarilta tietokantapalvelimelle, ellei yhteyttä ole erikseen suojattu salauksella. Salauksen voi toteuttaa useammallakin tavalla, esimerkiksi salatuilla tunneleilla tai salatuilla yhteyksillä, joista viimeistä käytettiin RPi:llä X.509-sertifikaattien muodossa.

Tarkempi käsittely ja analyysi tietoverkossa kulkevan datan salausmenetelmistä rajattiin pois työstä, sillä niillä ei ole vaikutusta tilan käyttöasteen arviointiin.

8.2 Datan kulku järjestelmässä

Tietoa voi suojata esimerkiksi salaamalla havaittu data (probe -kehykset) jo mittarissa. Tämä data lähetetään tietokantaan. Sovellus hakee dataa tietokannasta SQL-kyselyiden avulla.

Mikäli kysely palauttaa esimerkiksi vain yhden luvun, tietoa ei tarvitse uudelleen anonymisoida. Mikäli kyseessä on jotain muuta kuin vain yhden luvun sisältävää dataa, ano-

nymisoidaan data vielä kerran, esimerkiksi SHA2(str,hash_len)-funktion avulla, jolloin alkuperäisen laitteen tunnistaminen muuttuu erittäin vaikeaksi ellei jopa mahdottomaksi ilman sijaintipohjaista dataa.

8.3 Datan säilytys

Tilan käyttöasteen määrittämisen kannalta kaikkea sijaintipohjaista dataa tulisi säilyttää mahdollisimman lyhyt aika, vain ajanjakso, jolloin sijainnin määritys on oleellisinta. Tämän jälkeen data tulisi muuntaa numeeriseen tilastomuotoon, josta ei yksittäisiä laitteita tai henkilöitä voida tunnistaa. Tätä numeerista dataa voidaan säilyttää pitempiäkin aikoja ja sitä voidaan käyttää esimerkiksi tilastopohjaisten algoritmien lähteenä.

9 Menetelmän tarkkuus

Tutkitun arviointimenetelmän täsmällistä tarkkuutta on useiden muuttujien vuoksi haasteellista määrittää. Ensinnäkin mittarit havaitsevat laitteita eivätkä henkilöitä, joten 100 %:n tarkkuuteen ei päästä. Toiseksi, yhdellä henkilöllä on tuntematon määrä päätelaitteita. Lisäksi henkilön päätelaiteessa voi olla WLAN- ja Bluetooth-tekniikat kytkettynä pois päältä, jolloin laitetta ja tätä kautta henkilöä ei voida havaita.

Havaitun laitemäärän muuntaminen kävijämääräksi pitää tehdä kalibroimalla sensori kyseiseen tilaan, esimerkiksi jonkin muun mittarin, kuten kameralaskurin, avulla. Kalibroinnin tarpeen tiheys vaihtelee kohdetyyppien mukaan. Esimerkiksi toimistoympäristössä ei uudelleenkalibrointia tarvitse tehdä niin usein kuin esimerkiksi lentokentällä. Tämä johtuu siitä, että toimistotyyppisissä kohteissa väki on paikalla melko staattisesti ja samanaikaisesti päivästä toiseen. Sen sijaan lentokenttätyyppisessä dynaamisessa ympäristössä suuri osa paikallaolijoista on ”ohikulkumatkalla” ja heidän käyttäytymisensä on vaikeammin yleistettävissä. Tämän vuoksi jälkimmäisen tyyppisissä kohteissa mittauksen tarkkuus on hyvin todennäköisesti edellistä heikompi.

Esimerkiksi: kameramittarin antama mittaustulos on 1 000 henkilöä yhden kuukauden aikana. RPi-mittari samassa tilassa mittaa esimerkiksi 1 377 laitetta samassa ajassa. Tällöin saadaan kaavan $\frac{1000}{1377}$ lopputulokseksi 0,726, joka toimii mittarin kalibrointiarvona. Seuraavana kuukautena havaitaan 1 401 laitetta, jolloin kaavalla $1401 * 0{,}726 =$

1017,126, joka pyöristetään lähimpään kokonaislukuun. Järjestelmä ilmoittaa siis tunnistaneensa 1 017 henkilöä oletuksella, että duplikaattilaitteita ei esiinny. Kaava toimii mittarin koko kuuluvuus- eli skannausalueella.

Mikäli ollaan kiinnostuneita skannausalueen sisällä olevasta pienemmästä tilasta, esimerkiksi neuvotteluhuoneen henkilömäärän arvioinnista, kaava ei välttämättä toimi yhtä tehokkaasti. Tämä johtuu siitä, että kalibrointiarvo koskee koko skannausaluetta. Voidaan todeta, että järjestelmä ei ole niin tehokas kuuluvuusalueen sisällä olevan tilan henkilömäärän arvioinnissa.

Etäisyyden tarkkuus vaihtelee riippuen siitä, millä tekniikalla päätelaite on lähettänyt dataa ja kuinka moni mittari havaitsee päätelaitteen datan. Mikäli päätelaite lähettää dataa WLAN-tekniikalla ja vain yksi mittari havaitsee päätelaitteen probe request -kehyksen, voi etäisyys olla 2,4 gigahertsin taajuusalueella lähetetyllä datakehyksellä (IEEE 802.11b, IEEE 802.11g, IEEE 802.11n -tekniikat) sisätiloissa noin 46 metriä ja ulkotiloissa noin 92 metriä. Mikäli päätelaite lähettää dataa 5 GHz:n taajuudella, vähenee etäisyys, riippuen mahdollisista esteistä, noin puoleen tai jopa kolmannekseen. [15.]

Bluetoothin osalta etäisyyden tarkkuuteen eniten vaikuttaa lähetysteholuokka. Luokassa 1 (Class 1) maksimietäisyys on noin 100 metriä, toisessa luokassa (Class 2) noin 10 metriä ja kolmannessa luokassa (Class 3) etäisyys on noin 1 metriä [16]. Käytännössä hyvissä olosuhteissa havaitaan signaaleja parhaimmillaan 75 %:n etäisyydellä mainituista maksimeista. Tilanteesta riippuen havaintoarvot voivat jäädä jopa alle puoleen ideaalisesta. Mikäli usea mittari havaitsee signaalin, pienenee alue, jossa päätelaite sijaitsee, ja tämän seurauksena etäisyyden arvio paranee.

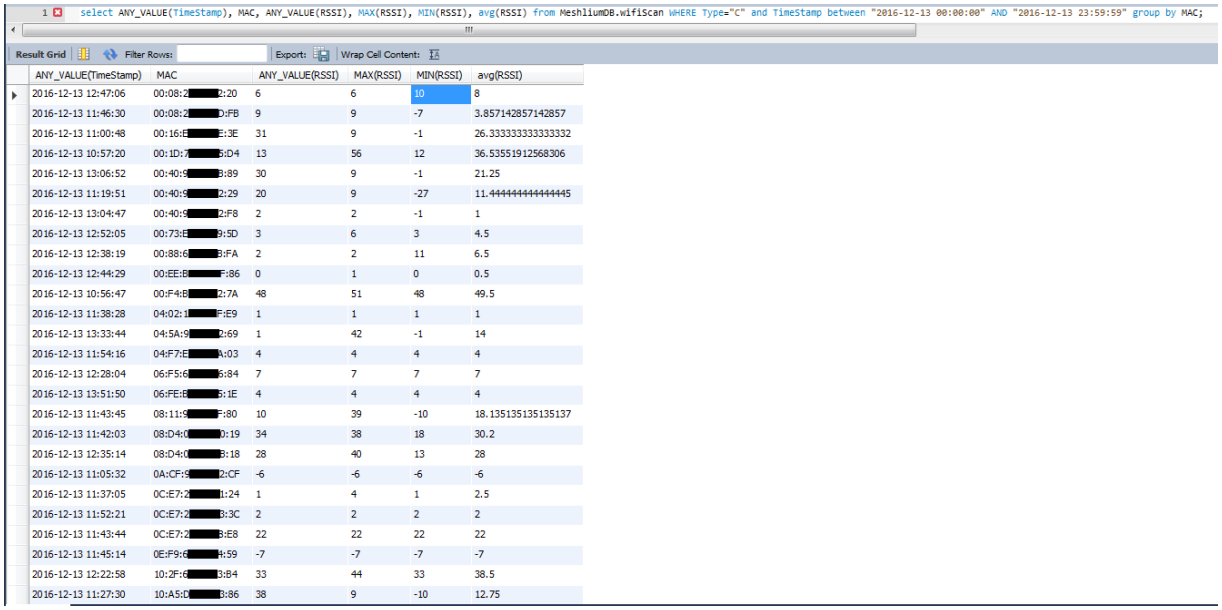
Etäisyyden arvioinnin yhteydessä on myös syytä ottaa huomioon, että vastaanotetut signaalit ei välttämättä saavu suoraan päätelaitteelta mittarille, vaan signaalit ovat voineet kimmota esteestä tilassa, esimerkiksi seinästä. Tämän seurauksena signaali heikkenee, jolloin arviointi etäisyyden suhteen vääristyy. Vääristymistä voidaan minimoida käyttämällä poikkeavuuksien havainnointiin tarkoitettuja funktioita.

Yhdessä laitteessa saattaa olla niin WLAN- kuin myös Bluetooth-tekniikat käytössä tai vain toinen tekniikoista on käytössä. Mikäli kumpikin tekniikka on käytössä ja päätelaitteen etäisyys mittaavaan laitteeseen on sellainen, että pystytään tunnistamaan niin

WLAN- kuin myös Bluetooth-tekniikoiden signaaleita, tarkkuus luetettavuusetäisyyden arvioinnissa on parempi, kuin jos havaittaisiin signaalit vain toisella tekniikoista.

Lisäksi tarkkuutta vääristävät mahdolliset saman laitteen duplikaattiesiintymät, esim. Applen modernit iOS-käyttöjärjestelmän versiolla 8 tai uudemmalla varustetut mobiililaitteet, kuten iPhone- tai iPad-brändien laitteet, eli laitteet, jotka satunnaistavat MAC-osoitteet automaattisesti. Tällä hetkellä Apple tekee näin moderneilla iOS-käyttöjärjestelmän versioilla, samoin saatavilla on kolmansien osapuolten sovelluksia MAC-osoitteen muuttamiseen. Tässä työssä keskitytään ainoastaan suurimpien brändien, Applen, Microsoftin ja Androidin, satunnaistamisominaisuuksiin.

WLAN-tekniikalla havaitut mahdolliset duplikaatit voidaan teoriassa tunnistaa hyödyntämällä päätelaitteiden SSID-pyyntöjä (request). Tarkkuutta tässäkin voidaan lisätä RSSI:n ja vastaavien arvojen perusteella. RSSI:n osalta tarkkuutta voidaan parantaa niiden laitteiden osalta, jotka ovat paikallaan, jolloin voidaan verrata aikaisemman laitteen lähettämien pakettien signaalinvoimakkuuksia (RSSI) duplikaattiehdokkaaseen. Mikäli signaalin voimakkuus on sama tai lähellä alkuperäistä, voidaan kasvattaa todennäköisyyttä, että duplikaattiehdokas on oikea duplikaatti. Mikäli signaalin voimakkuus eroaa merkittävästi alkuperäisestä, ei kuitenkaan voida suoraan poissulkea duplikaatin mahdollisuutta, mikäli kyseessä on mobiililaite. Kuva 6 esittää yhden päivän aikana havaittuja päätelaitteita ja niistä vastaanotettujen signaalien voimakkuuksien minimi-, maksimi- ja keskiarvoja.



```
1  select ANY_VALUE(TimeStamp), MAC, ANY_VALUE(RSSI), MAX(RSSI), MIN(RSSI), avg(RSSI) from MeshliumDB.wifiScan WHERE Type="C" and TimeStamp between "2016-12-13 00:00:00" AND "2016-12-13 23:59:59" group by MAC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

ANY_VALUE(TimeStamp)	MAC	ANY_VALUE(RSSI)	MAX(RSSI)	MIN(RSSI)	avg(RSSI)
2016-12-13 12:47:06	00:08:2[redacted]2:20	6	6	10	8
2016-12-13 11:46:30	00:08:2[redacted]D:FB	9	9	-7	3.857142857142857
2016-12-13 11:00:48	00:16:E[redacted]E:3E	31	9	-1	26.333333333333332
2016-12-13 10:57:20	00:1D:7[redacted]5:D4	13	56	12	36.53551912568306
2016-12-13 13:06:52	00:40:9[redacted]3:89	30	9	-1	21.25
2016-12-13 11:19:51	00:40:9[redacted]2:29	20	9	-27	11.444444444444445
2016-12-13 13:04:47	00:40:9[redacted]2:F8	2	2	-1	1
2016-12-13 12:52:05	00:73:E[redacted]9:5D	3	6	3	4.5
2016-12-13 12:38:19	00:88:6[redacted]3:FA	2	2	11	6.5
2016-12-13 12:44:29	00:EE:B[redacted]F:86	0	1	0	0.5
2016-12-13 10:56:47	00:F4:B[redacted]2:7A	48	51	48	49.5
2016-12-13 11:38:28	04:02:1[redacted]F:E9	1	1	1	1
2016-12-13 13:33:44	04:5A:9[redacted]2:69	1	42	-1	14
2016-12-13 11:54:16	04:F7:E[redacted]A:03	4	4	4	4
2016-12-13 12:28:04	06:F5:6[redacted]5:84	7	7	7	7
2016-12-13 13:51:50	06:FE:B[redacted]5:1E	4	4	4	4
2016-12-13 11:43:45	08:11:9[redacted]F:80	10	39	-10	18.135135135135137
2016-12-13 11:42:03	08:D4:0[redacted]0:19	34	38	18	30.2
2016-12-13 12:35:14	08:D4:0[redacted]3:18	28	40	13	28
2016-12-13 11:05:32	0A:CF:9[redacted]2:CF	-6	-6	-6	-6
2016-12-13 11:37:05	0C:E7:2[redacted]1:24	1	4	1	2.5
2016-12-13 11:52:21	0C:E7:2[redacted]3:3C	2	2	2	2
2016-12-13 11:43:44	0C:E7:2[redacted]3:E8	22	22	22	22
2016-12-13 11:45:14	0E:F9:6[redacted]4:59	-7	-7	-7	-7
2016-12-13 12:22:58	10:2F:6[redacted]3:B4	33	44	33	38.5
2016-12-13 11:27:30	10:A5:D[redacted]3:86	38	9	-10	12.75

Kuva 6. Havaittujen päätelaitteiden vastaanotettujen signaalien voimakkuuksia yhden päivän ajalta.

Toinen tarkkuutta parantava menetelmä on aikaleimoihin perustuva tunnistus, joka edellyttää, että mittarit käyttävät esimerkiksi NTP:tä kellonajan synkronisointiin. Menetelmä edellyttää, että niin ”alkuperäisestä” laitteesta kuin duplikaattiehdokkaasta on saatu kerättyä dataa – mitä enemmän, sitä tarkemmaksi menetelmä käy. Menetelmässä pyritään tunnistamaan kaava, millä tahdilla laite on lähettänyt dataa. Mikäli kaavat ovat samoja, voidaan pitää todennäköisenä, että laitteiden malli tai ohjelmisto tai kumpikin ovat samoja. Sen sijaan ei voida ottaa kantaa siihen, onko duplikaattiehdokas alkuperäisen vertailukohteen duplikaatti.

Poikkeamien havaitseminen

Vastaanotettujen signaalien poikkeamien havaitsemiseen työssä käytettiin seuraavaa menetelmää:

Q_1 - 1.5 * IQR > kunnollinen data < Q_3 + 1.5 * IQR,
Q_1 - 1.5 * IQR < poikkeamia datassa > Q_3 + 1.5 * IQR,
jossa Q_1 on ensimmäinen neljännes ja Q_3 on kolmas neljännes.
IQR puolestaan on neljännespisteiden väli.
Haetaan 8 viimeisintä RSSI-lukemaa tietyn MAC-osoitteen lähettämästä datasta.
Haettu data: -67, -68, -69, -59, -61, -61, -57, -59.
Muunnetaan data pienimmästä suurimpaan: -69, -68, -67, -61, -61, -59, -59, -57.

Lasketaan ensimmäisen neljänneksen arvo, tässä tapauksessa:
$(-68 + (-)67) / 2 = -67,5.$
Lasketaan kolmannen neljänneksen arvo: $(-59 + (-)59) / 2 = -59.$
Kunnollisen datan alempi alue = $Q\_1 - 1.5 => (-67,5 - 1.5 * 4) = -73,5.$
Kunnollisen datan ylempi alue = $Q\_3 + 1.5 => (-59 + 1.5 * 4) = -53.$
[8.]

Täten kunnollisen datan alue on välillä -73,5–53,5 joten alkuperäisen datan joukossa ei ole poikkeamia.

25 minuutin kuluttua tehtiin sama kysely uudelleen ja kohdennettiin se samaan laitteeseen. Tällöin data muunnettuna pienimmästä suurimpaan oli -83, -70, -68, -67, -61, -60, -59, -56.

Q_1: -69
Q_2: -59,5
$Q_1 - 1,5 * 4: -75$
$Q\_2 + 1,5 * 4: 53,5$

Poikkeamana havaitaan tällöin luku -83, joten sitä ei sijainnin määrityksessä huomioida.

Tätä menetelmää käytettiin tarkentamaan laitteen ja henkilön sijainnin määritystä. Menetelmän vaikutus sinänsä on melko pieni, mutta se joka tapauksessa myös osaltaan parantaa tilankäyttöasteen arvioinnin tarkkuutta.

Menetelmän mahdolliset erot Suomessa tai muualla maailmassa

Sillä, missä päin maailmaa menetelmää käytetään, ei ole suoraa vaikutusta, mutta erot erilaisissa seikoissa, kuten mobiililiittymien datapakettien hinnoittelussa, vaikuttavat ainakin teoreettisesti menetelmän tarkkuuteen. Suomessa mobiililiittymiä on usein myyty rajattomalla datalla, kun taas maailmalla mobiilidata on ollut liittymäkohtaisesti rajoitetumpaa. Miten tämä sitten vaikuttaa WLAN- ja Bluetooth-tekniikoilla tehtyyn havainnointiin?

Mikäli käyttäjä Suomessa pääsee Internetiin helposti suoraan mobiililaitteellaan eikä tarvitse esimerkiksi alueella olevaa WLAN:a, ei käyttäjä välttämättä pidä WLAN-radiota päällä turhaan kuluttamassa mobiililaitteensa virtaa. Käyttäjä maailmalla taas saattaa helpommin käyttää olemassa olevaa WLAN-verkkoa päästäkseen Internetiin ja pitääkseen mobiililiittymänsä datapaketin liikenteen pienempänä. [18.]

10 Johtopäätöksiä

Henkilömäärän määrittäminen WLAN- ja Bluetooth-tekniikoiden avulla Raspberry Pi:llä ja Meshlium Scannerilla voi olla toteutettavissa, mikäli mahdollisimman monet muuttujat tiedetään tai mikäli arviot muuttujien arvoista osuvat riittävän lähelle todellisuutta. Käytetään sanaa voi, sillä menetelmä, jossa pyritään havaitsemaan päätelaitteiden signaaleja eikä itse henkilöitä, on menetelmänä selvästi epätarkempi kuin esimerkiksi kameralaskuri, joka mittaa henkilöiden lukumäärää. Ideaalitilanteessa tämän menetelmän avulla saadut tulokset voivat olla hyvin lähellä oikeaa. Ideaalitilanteella tarkoitetaan tässä yhteydessä tilannetta, jossa mittausympäristönä olevasta tilasta on langaton verkko ja jossa kaikki mittarin havainnointialueella olevat päätelaitteet ovat yhdistyneet siihen. Tällöin MAC-osoitteen satunnaistamista ei esiinny. Tämän lisäksi tilan käyttäjäkunta ja sen päätelaitteiden lukumäärä on hyvin lähellä vakiota.

Epäideaalisimmassa tilanteessa yhtään signaalia tai laitetta ei havaita.

Ei-toivotussa tilanteessa alueella ei ole langatonta verkkoa, jolloin kaikki päätelaitteet, joissa on satunnaistamisominaisuus, satunnaistavat MAC-osoitteet eikä alueella ole laitteita, jotka eivät satunnaistamista osaisi. Epäideaalisimmassa tilanteessa ei ole suoritettu mittareiden kalibrointia eikä käytettävissä ole mitään tietoa henkilömäärästä.

Reaalimaailman tilanne asettuu yleensä ideaalisimman ja ei-toivotun tilanteen välimaastoon. Mittauksissa ei kohdattu epäideaalisinta tilannetta sinä aikana, jolloin tilassa oli henkilöitä. Realistisessa tilanteessa MAC-osoitteen satunnaistavilta laitteilta on erittäin vaikea välttyä, mikä ei itsessään ole ongelma. Jos duplikaattilaitteita ei pystytä tunnistamaan, tarkkuuteen vaikuttaa satunnaistamiskertojen määrä. Satunnaistamiskertojen määrään puolestaan vaikuttaa muun muassa päätelaitteen ohjelmisto.

Bluetoothin osalta havaittiin hyvin nopeasti, että tekniikka ei yksinään sovellu käyttöasteen arviointiin. Mittauksissa havaittiin Bluetooth-laitteita hyvin vähän: keskimäärin 0,4–6 % siitä, mitä WLAN-tekniikalla havaittiin. Tästä huolimatta Bluetooth-tekniikkaa ei kannata täysin unohtaa, sillä se soveltuu erityisen hyvin tilanteisiin, joissa halutaan tarkempaa tietoa etäisyydestä tai useammalla Bluetooth-mittarilla jopa sijainnin tyyppistä tietoa (mittareiden havainnointialueen leikkaus). Samahan onnistuu WLAN-tekniikalla,

mutta koska sen havainnointialue on laaja, Bluetooth-tekniikka pienempine havainnointialueineen on tehokkaampi erityisesti etäisyyksien arvioinnissa. Tosin, mikäli laitteita ei havaita, ei etäisyyden tai sijainnin määritys onnistu.

100 %:n tarkkuuteen ei tällä menetelmällä voi päästä, vaikkakin numeerisesti se saattaa olla mahdollista. Koska menetelmässä pyritään havaitsemaan henkilöiden päätelaitteiden signaaleja ja sitä kautta yksittäisiä laitteita, eivät mittarit havaitse suoraan oikeita henkilöitä, vaan henkilömäärä on puhdas arvio. Tästä syystä ei ole tämänkaltaisen menetelmän osalta oikein puhua 100 %:n tarkkuudesta.

Menetelmän tarkkuuden haaste on useiden muuttujien määrä. Menetelmän kannalta keskeisiä muuttujia ovat muun muassa

- yhdellä henkilöllä oleva laitemäärä
- laitemäärän keskiarvo henkilöä kohti
- onko havaitussa laitteessa WLAN tai Bluetooth tai molemmat päällä
- kuinka moni laitteista on havainnointialueella tai sen ulkopuolella
- lähettävätkö laitteet signaaleja ollessaan havainnointialueella.

Esimerkiksi yhden henkilön laitemäärän tietäminen ei ole kovinkaan oleellista ympäristöissä, joissa havaitaan useita satoja laitteita, varsinkin, jos laitemäärä on melko lähellä keskiarvoa, koska sen vääristymä ei välttämättä yksittäistapauksena riitä muuttamaan menetelmän arvioimaa henkilömäärää. Jos tilanne on muuten sama kuin edellä kuvattu, mutta havaittu laitemäärä on kymmeniä, vaikutus henkilömäärän arvioon on merkittävästi suurempi: tällöin menetelmän arvio henkilömäärästä todennäköisesti antaa väärän arvon helpommin.

Toisin sanoen, muuttujien vaikutukset vaihtelevat muun muassa laitemäärän mukaan. Mitä enemmän dataa, havaittuja laitteita tai henkilöitä tilassa on, sitä vähemmän poikkeukset näkyvät massasta. Tarkkuuden voi siis olettaa paranevan. Mikäli jokin arvio kuitenkin poikkeaa selvästi merkittävässä osassa dataa, tarkkuus heikkenee.

Vaihtoehtoinen RSSI-arvoon perustuva menetelmä

Tutkitulle arviointimenetelmälle vaihtoehtoinen menetelmä on signaalin voimakkuuteen perustuva menetelmä. Siinä ei tarvitse välttämättä tunnistaa duplikaattilaitteita, mutta mittareiden tarve verrattuna aiemmin esittelemääni menetelmään on moninkertainen. Mittarit pitäisi sijoittaa niin, että käytännössä lähes jatkuvasti noin kolme mittaria (jotka voivat toki vaihtua laitteen sijainnin muuttuessa) havaitsevat mahdolliset käyttäjän laitteet. Jos huomataan, että pari laitetta on havaittu samoilla mittareilla samoihin aikoihin ja että RSSI-arvojen muutokset ovat poikkeamat pois lukien samankaltaisia, voidaan olettaa, että laitteet kuuluvat samalle henkilölle. Mikäli kuitenkin huomataan, että laitteet erkanevat jossain vaiheessa ja erkanevien laitteiden RSSI-arvot eivät ole stabiileja (eli laite liikkuu), on kyseessä menetelmän näkökulmasta vähintään kaksi henkilöä, jotka sattuivat kulkemaan hetkellisesti samaa reittiä. Menetelmä ei toimi, mikäli päätelaite satunnaistaa MAC-osoitteensa jokaisen tai lähes jokaisen lähetettävän probe-kehyksen aikana.

Tässä menetelmässä seurattaisiin RSSI-muutoksia tiheämmällä mittariverkostolla, jolloin olisi teoreettiset mahdollisuudet henkilömäärän laskemiseen. Työssä ei kuitenkaan keskitytty menetelmään tämän enempää, sillä työn tavoite oli tutkia menetelmiä, joissa henkilöiden sijaintitietoja ei kulje. Tämä signaalin voimakkuuteen pohjautuva menetelmä voisi olla mielenkiintoinen jatkotutkimuksen kohde.

Lähteet

1 Meshlium technical guide. 2017. Verkkodokumentti. Libelium. <http://www.libelium.com/downloads/documentation/meshlium_technical_guide.pdf>. Päivitetty 11.7.2017. Luettu 22.11.2017.

2 Raspberry Pi 3 Model B SBC. 2017. Verkkodokumentti. RS. <http://uk.rs-online.com/web/p/processor-microcontroller-development-kits/8968660/>. Päivitetty 2017. Luettu 21.11.2017.

3 A Brief History of Debian. 2017. Verkkodokumentti. Debian. <https://www.debian.org/doc/manuals/project-history/ch-intro.en.html> Päivitetty 27.8.2017. Luettu 22.11.2017.

4 Precision Time Protocol. 2013. Verkkodokumentti. EndRunTechnologies. <http://www.endruntechnologies.com/pdf/PTP-1588.pdf> Päivitetty 12.7.2013. Luettu 22.11.2017.

5 Wi-Fi Direct. 2017. Verkkodokumentti. Wi-Fi. <https://www.wi-fi.org/discover-wi-fi/wi-fi-direct>. Päivitetty 2017. Luettu 22.11.2017.

6 Blue Hydra. 2017. Verkkodokumentti. GitHub. <https://github.com/pwnieexpress/blue_hydra>. Päivitetty 24.10.2017. Luettu 5.11.2017.

7 MAC address randomization joins Apple's heap of iOS 8 privacy improvements. 2014. Verkkodokumentti. Appleinsider. <http://appleinsider.com/articles/14/06/09/mac-address-randomization-joins-apples-heap-of-ios-8-privacy-improvements>. Päivitetty 9.6.2014. Luettu 30.8.2017.

8 Outlier Function. 2017. Verkkodokumentti. Tutorialspoint. <https://www.tutorialspoint.com/statistics/outlier_function.htm>. Päivitetty 2017. Luettu 22.11.2017.

9 what is Bluetooth? 2017. Verkkodokumentti. Bluetooth. <https://www.bluetooth.com/what-is-bluetooth-technology/how-it-works>. Päivitetty 2017. Luettu 19.11.2017.

10 Tcpdump. 2017. Verkkodokumentti. Tcpdump. <http://www.tcpdump.org/>. Päivitetty 2017. Luettu 28.6.2017.

11 Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. 2017. Verkkodokumentti. Mathy Vanhoef. <http://papers.mathyvanhoef.com/asiaccs2016.pdf>. Päivitetty 1.11.2017. Luettu 6.11.2017.

12 How to stop your Windows 10 Mobile phone's location from being tracked through Wi-Fi. 2016. Verkkodokumentti. Windows Central. <https://www.windowscentral.com/how-prevent-places-tracking-your-location-windows-10-mobile>. Päivitetty 8.1.2016. Luettu 17.11.2017.

13 Experience with MAC Address Randomization in Windows 10. 2015. Verkkodokumentti. The Internet Engineering Task Force <https://www.ietf.org/proceedings/93/slides/slides-93-intarea-5.pdf>. Päivitetty 2015. Luettu 28.6.2017.

14 Enhanced RSSI -based high accuracy real-time user location tracking system for indoor and outdoor environments. 2008. <http://s2is.org/Issues/v1/n2/papers/paper14.pdf>. Päivitetty 2008.
Luettu 28.6.2017.

15 Here's why you should use 5GHz WiFi instead of 2.4GHz. 2014. Verkkodokumentti. PocketNow. <http://pocketnow.com/2014/01/23/5ghz-wifi>.
Päivitetty 23.1.2014. Luettu 17.11.2017.

16 Bluetooth Power Classes. 2008. Verkkodokumentti. Bluetooth Insight.
<http://bluetoothinsight.blogspot.fi/2008/01/bluetooth-power-classes.html>.
Päivitetty 1.11.2008. Luettu 17.11.2017.

17 Android 6.0 Changes. 2015. Verkkodokumentti. Android.
<https://developer.android.com/about/versions/marshmallow/android-6.0-changes.html>. Päivitetty 11.4.2017. Luettu 18.11.2017.

18 Suomalainen käyttää mobiili-dataa hurjat 11 gigatavua kuussa – määrä on monin-kertainen muihin maihin verrattuna. 2017. Verkkodokumentti. HS
<https://www.hs.fi/kotimaa/art-2000005405431.html>. Päivitetty 12.10.2017.
Luettu 22.11.2017

Skannausskripti

```
while :
do
 for i in 1 6 11
 do
  iwconfig wlan0 channel $i
  tcpdump -i wlan0 -G 13 -W 1 -e -tt type mgt subtype probe-req -j host -w
RawData
  tcpdump -nle -tt -r RawData | awk '{$2=$3=$4=$5=$7=$8=$11=""; print $0}' |
cut -c 1-17,22-25,28,34-51,55-71,73- | uniq -f1 -f2 -f4 -f6 >> Data.txt
  END=$(wc -l Data.txt | awk '{print $1}')
  START='0';
  if [ $END -gt 0 ]
  then
  while [ $START -lt $END ]; do
   START=$(( $START + 1 ))
   TIME=$(cat Data.txt | head -$START | tail -1 | awk '{print $1}' | perl -
MPOSIX=strftime -e 'print strftime("%Y-%m-%d %T", localtime(<>)) . "\n"')
   MAC_SOURCE=$(cat Data.txt | head -$START | tail -1 | awk '{ print $4 }' |
awk '{ print toupper($1)}' | cut -c 1-17)
   ORGSSID=$(cat Data.txt | head -$START | tail -1 | awk '{$1=$2=$3=$4=$5="";
print $0}' | cut -c 6- | cut -c 1-34 | cut -f1 -d")" | cut -c 2- | cut -c 1-
32)
   #ORGSSID=$(cat Data.txt | head -$START | tail -1 | awk '{$1=$2=$3=$4=$5="";
print $0}' | cut -c 6-34 | cut -f1 -d")" | cut -c 2-33)
   TYPE="C"
   if [ ${#ORGSSID} -eq 0 ]
   then
    SSID=""
   else
    if [ "C" = $TYPE ]
    then
     SSID=$(echo $(cat Data.txt | head -$START | tail -1 | awk
'{$1=$2=$3=$4=$5=""; print $0}' | cut -c 6- | cut -c 1-34 | cut -f1 -d")" |
cut -c 2- | cut -c 1-32)$(echo $(cat Data.txt | head -$START | tail -1 | awk
'{$1=$2=$3=$4=$5=""; print $0}' | cut -c 6- | cut -c 1-34 | cut -f1 -d")" |
cut -c 2- | cut -c 1-32) | cksum | cut -c 1-10) -n | md5sum | awk '{print
toupper($1)}')
    else
     SSID=$(cat Data.txt | head -$START | tail -1 | awk '{$1=$2=$3=$4=$5="";
print $0}' | cut -c 6- | cut -c 1-34 | cut -f1 -d")" | cut -c 2- | cut -c 1-
32)
    fi
   fi
   unset ORGSSID=""
   AP="(not associated)"
   RSSI=$(cat Data.txt | head -$START | tail -1 | awk '{ print $2 }' | cut -c
1-3)
   VENDOR=$(less -FX /root/WLANSCAN_2017/oui20161212.txt | egrep -i $(echo -n
"$MAC_SOURCE" | cut -c 1-2,4-5,7-8) | awk '{$1=$2=$3=""; print $0}' | cut -c
4-154 | awk '{sub(/\r/,"",$NF)} {print($0)}')
   if [ -z "$VENDOR" ]; then
    VENDOR="Unknown"
   #else
   # VENDOR=$(echo -n "$VENDOR")
   fi
   echo "$TIME $MAC_SOURCE '$SSID' $RSSI $VENDOR $TYPE $AP $(hostname)"
   SRV=$(mysqladmin --host=192.168.1.2 -u TESTAAJA -pSalasana11235 --connect-
timeout=5 --ssl ping | awk '{print $3}')
   if [ "alive" = "$SRV" ]
   then
    FILE="/root/WLANSCAN_2017/toMariaDB"
```

```
    if [ -f "$FILE" ]
    then
     SIZE=$(wc -l $FILE | uniq -f9 | awk '{print $1}')
     if  [ $SIZE -gt 0 ]
     then
      LINE=0
      while [ $LINE -lt $SIZE ]; do
       LINE=$(( $LINE + 1 ))
       cat $FILE | uniq -f9 | head -$LINE | tail -1 | mysql --host=192.168.1.2
-u TESTAAJA -pSalasana11235 MeshliumDB --ssl-ca=/etc/mysql/ssl/ca-cert.pem
      done
      rm -rf /root/WLANSCAN_2017/toMariaDB
      unset SIZE=""
      unset FILE=""
     fi
    else
     echo "INSERT INTO wifiScan (TimeStamp,MAC,SSID,RSSI,Vendor,Type,AP,Mesh-
liumID) VALUES ('$TIME', '$MAC_SOURCE', '$SSID', '$RSSI', '$VENDOR', '$TYPE',
'$AP', '$(hostname)');" | mysql --host=192.168.1.2 -u TESTAAJA -pSalasana11235
MeshliumDB --ssl-ca=/etc/mysql/ssl/ca-cert.pem &
    fi
   else
    echo "INSERT INTO wifiScan (TimeStamp,MAC,SSID,RSSI,Vendor,Type,AP,Mesh-
liumID) VALUES ('$TIME', '$MAC_SOURCE', '$SSID', '$RSSI', '$VENDOR', '$TYPE',
'$AP', '$(hostname)');" >> /root/WLANSCAN_2017/toMariaDB
   fi
    unset TIME=""
    unset MAC_SOURCE=""
    unset SSID=""
    unset VENDOR=""
    unset TYPE=""
    unset AP=""
    unset RSSI=""
  done
   sleep 1
   echo ""
  fi
  rm -rf RawData
  rm -rf Data.txt
 done
done
#Skannausskriptissä käytetty oui20161212.txt tiedosto:
#Tiedoston sisältö vastaa http://standards-oui.ieee.org/oui/oui.txt -sisältöä.
```