

Vesa Koskela

WLAN FINGERPRINTING -POHJAINEN LAITEPAIKANNUS

Kandidaatintyö
Informaatioteknologian ja viestinnän tiedekunta
Tarkastaja: professori Mikko Valkama
Joulukuu 2019

TIIVISTELMÄ

Vesa Koskela: WLAN fingerprinting -pohjainen laitepaikannus
Kandidaatintyö
Tampereen yliopisto
Tieto- ja sähkötekniikka
Joulukuu 2019

WLAN fingerprinting -pohjainen laitepaikannus on WLAN-tukiasemien RSS-arvoja hyödyntämiseen perustuva paikannusmenetelmä. Menetelmä jakautuu opetus- ja paikannusvaiheeseen. Opetusvaiheessa kohdeympäristöstä kerätään sormenjälkiä. Sormenjälki on tieto, jossa fyysiseen sijaintiin liitetään kyseisessä pisteessä kuultavista WLAN-tukiasemista saatavat RSS-arvot. Kun kohdeympäristöstä on kerätty tietokantaan riittävästi tällaisia sormenjälkiä, laite voidaan algoritmin avulla paikantaa. Tämä algoritmi hyödyntää oppimisvaiheen tietokantaa sekä käyttäjän laitteen kuulemia RSS-arvoja.

Tässä työssä esitellään WLAN fingerprinting -menetelmän teoria ja siinä käytettäviä algoritmeja. Työhön kuuluu myös mittaukset, joissa WLAN fingerprinting -menetelmää testattiin käytännössä Tampereen yliopiston Tietotalossa. Tietotalossa kerättiin 292 sormenjälkeä ja paikannusta mallinnettiin MATLAB:ssa.

Avainsanat: WLAN fingerprinting, RSS, MATLAB

ABSTRACT

Vesa Koskela: WLAN fingerprinting based device positioning
Bachelor's thesis
Tampere University
Computing and Electrical Engineering
December 2019

WLAN fingerprinting based device positioning is a method for positioning a device using WLAN access points. It is possible to measure RSS from WLAN access points. WLAN fingerprinting based device positioning is using this RSS measurements and algorithms to position the device. The method consists of two different phases. The first phase is called learning phase. In this phase the fingerprints are collected. One fingerprint consists of physical location and the RSS measurements of heard WLAN access points. The second phase is called user location phase. After enough fingerprints are collected in the first phase the device is possible to locate using the database of fingerprints and algorithms.

In this thesis we explain the theory behind WLAN fingerprinting based device positioning. We also introduce some of the algorithms used in this method. Studying this method in practice was also a part of this study. 292 fingerprints were collected in Tampere University's Tietotalo. Using this database location phase was modeled in MATLAB.

Keywords: WLAN fingerprinting, RSS, MATLAB

ALKUSANAT

Kiitän kandidaatintyöni ohjaajaa ja tarkastajaa professori Mikko Valkamaa mielenkiintoisesta kandidaatintyön aiheesta sekä avusta ja neuvoista työn eri vaiheissa. Kiitän myös yliopistonlehtori Jukka Talvitietä sekä apulaisprofessori Elena-Simona Lohania avusta ja neuvoista työn tekemisessä.

Tampereella, 10.12.2019

Vesa Koskela

SISÄLLYSLUETTELO

1.	JOHDANTO	1
2.	WLAN FINGERPRINTING -MENETELMÄ	3
2.1	Oppimis- ja paikannusvaihe.....	3
2.2	Algoritmit	6
3.	MITTAUKSET	8
3.1	Oppimisvaihe.....	8
3.2	Paikannusvaihe	10
4.	VIRHEARVIO.....	16
5.	YHTEENVETO.....	20
	LÄHTEET	22

LYHENTEET JA MERKINNÄT

dBm	desibelimilliwatti
GNSS	engl. Global Navigation Satellite System
GPS	engl. Global Position System
JSON	engl. JavaScript Object Notation
KNN	engl. K-Nearest Neighbor
MAC	engl. Media Access Control
MATLAB	engl. Matrix Laboratory
NN	engl. Nearest Neighbor
RSS	engl. Received Signal Strength
WKNN	engl. Weighted K-Nearest Neighbor
WLAN	engl. Wireless Local Area Network

1. JOHDANTO

Sisätiloissa toimiviksi tarkoitetuille paikannusjärjestelmille on kasvava tarve. Paikkatiedon avulla voidaan tarjota erilaisia palveluita esimerkiksi museoissa, ostoskeskuksissa ja sairaaloissa. Paikallinen markkinointi on yksi käyttökohde. Suurissa rakennuksissa kuten ostoskeskuksissa voi olla myös tarve reittiopastukseen. [1]

Satelliittipaikannusjärjestelmät kuten GPS (Global Positioning System) toimivat hyvällä tarkkuudella ulkotiloissa. Sisätiloissa satelliittipaikannus ei kuitenkaan toimi niin hyvin signaalin ollessa heikko erilaisten fyysisten rakenteiden vuoksi. Kaupungeissa signaali voi myös heijastua rakenteista, mikä huonontaa paikannustarkkuutta. GPS voi siis olla liian epätarkka tai virheellinen joihinkin käyttötarkoituksiin sisätiloissa. Toisaalta GPS voi olla joissain käyttötarkoituksissa liian tarkka ja energiatehokkuudeltaan huono ratkaisu. Esimerkiksi paikalliseen markkinointiin GPS on turhan tarkka ja huono energiatehokkuudeltaan. On siis syntynyt tarve vaihtoehtoisille paikannusjärjestelmille, jotka toimivat hyvällä tarkkuudella sisätiloissa ja ovat energiatehokkaampia kuin GPS ja muut satelliittipaikannusjärjestelmät. [2, 3]

Uusia paikannusjärjestelmiä suunnitellessa myös kustannustehokkuus on tärkeää. Suunnittelussa pyritään ottamaan huomioon paikannustarkkuuden lisäksi aloituskustannukset ja ylläpitokustannukset. Lisäksi energiatehokkuus on tärkeää erityisesti mobiililaitteissa. Kustannuksien minimoimiseksi on pyritty hyödyntämään jo olemassa olevaa infrastruktuuria ja tekniikkaa. Yksi tällainen vaihtoehto on hyödyntää rakennuksissa jo olevia WLAN-tukiasemia (Wireless Local Area Network). Monet paikannusjärjestelmät hyödyntävät WLAN-signaalin RSS (Received Signal Strength) tietoa. RSS:n avulla voidaan selvittää, kuinka vahva WLAN-signaali on eri sijainneissa. Tätä tietoa voi hyödyntää usealla eri tavalla selvittäessä käyttäjän tai laitteen sijaintia. RSS-tieto on myös helposti saatavissa WLAN-verkosta, joten sitä on helppo hyödyntää. [2, 4]

Koneoppimisen ja syväoppimisen suosion kasvamisen myötä myös sisätilapaikannus voi olla yksi sovelluskohde, jossa voidaan hyödyntää koneoppimista. Syväoppimista onkin jo kokeiltu käyttäjän sijainnin arvioimiseen. [5]

Yksi paikannusjärjestelmä ei voi olla paras jokaisessa tilanteessa. Siksi tarvitaan useampia paikannusjärjestelmiä, joita voi käyttää erilaisissa tilanteissa. Tässä työssä keskitytään WLAN fingerprinting -pohjaiseen laitepaikannukseen. Toisessa luvussa käsitellään WLAN fingerprinting -menetelmän teoriaa ja siinä käytettäviä algoritmeja. Kolmannessa luvussa käsitellään tämän työn osana olevat mittaukset sekä niiden tulokset. Neljännessä luvussa on virhearvio. Yhteenveto on luvussa viisi.

2. WLAN FINGERPRINTING -MENETELMÄ

Tässä luvussa esitellään WLAN fingerprinting -menetelmään liittyvää teoriaa. Ensimmäisenä esitellään menetelmän käytössä tarvittavat lähtökohdat ja suureet. Tämän jälkeen esitellään menetelmän käyttämiseen liittyvät oppimis- ja paikannusvaihe. Luvussa 2.2 esitellään tarkemmin paikannusvaiheessa käytettäviä algoritmeja, joiden avulla laite lopulta paikannetaan.

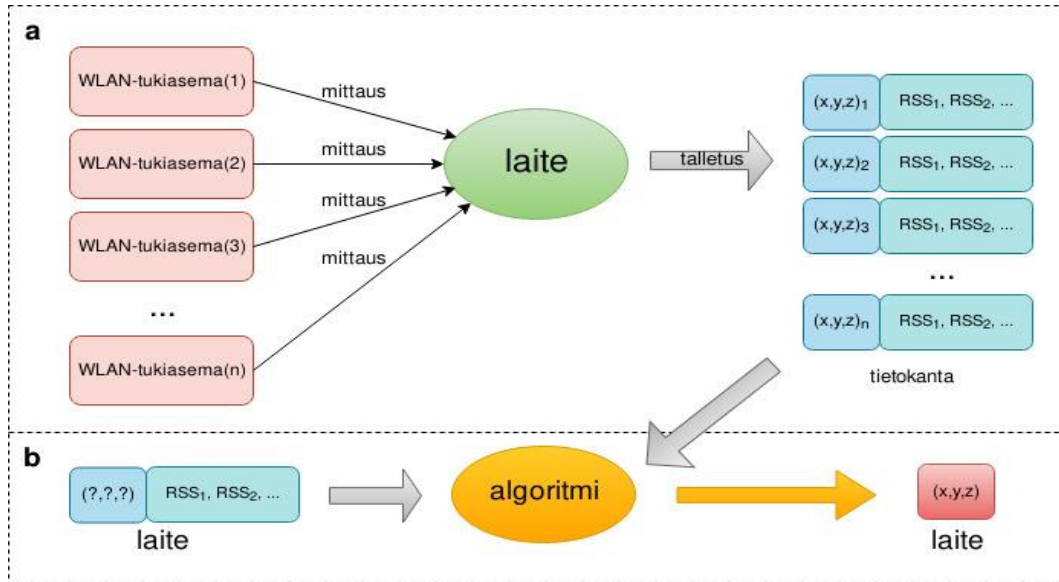
2.1 Oppimis- ja paikannusvaihe

WLAN fingerprinting -menetelmän avulla on mahdollista paikantaa käyttäjän laite käyttäen hyväksi WLAN-tukiasemia ja niistä saatavia RSS-arvoja. Menetelmässä voidaan käyttää rakennuksissa jo olevia tukiasemia, eikä menetelmän käyttäminen vaadi uusien laitteiden asentamista. Käyttäjän laitteen mittaamia arvoja vertaillaan tietokannassa oleviin sormenjälkiin (engl. fingerprint). Tavoitteena on löytää sormenjälki, jonka RSS-arvot ovat mahdollisimman lähellä käyttäjän laitteen mittaamia RSS-arvoja. Arvio käyttäjän sijainnista saadaan käyttämällä yhden tai useamman tällaisen sormenjäljen koordinaatteja. Sormenjäljellä tarkoitetaan tietoa, jossa on yhdistettynä tiedossa olevat sijainnin koordinaatit ja tässä sijainnissa mitatut RSS-arvot. Nämä sormenjäljet kerätään menetelmän oppimisvaiheessa.

RSS-arvojen käyttämisessä on monia hyviä puolia. RSS on yksi tärkeimmistä tietoliikennetekniikassa käytettävistä suureista. Se on tärkeässä osassa tietoliikennetekniikassa, ja sitä monitoroidaan jatkuvasti useissa tietoliikennejärjestelmissä. RSS-arvojen mittaukset ovat siis tärkeitä, ja on erittäin todennäköistä, että niitä tullaan käyttämään myös tulevaisuudessa. Tästäkin syystä RSS-arvojen käyttäminen on järkevää WLAN fingerprinting -menetelmässä. RSS-arvot ovat siis laajalti saatavilla ja tulevat olemaan saatavissa tulevaisuudessakin. RSS-arvot ovat myös helposti saatavilla useimmissa käyttöjärjestelmissä. [2, 6]

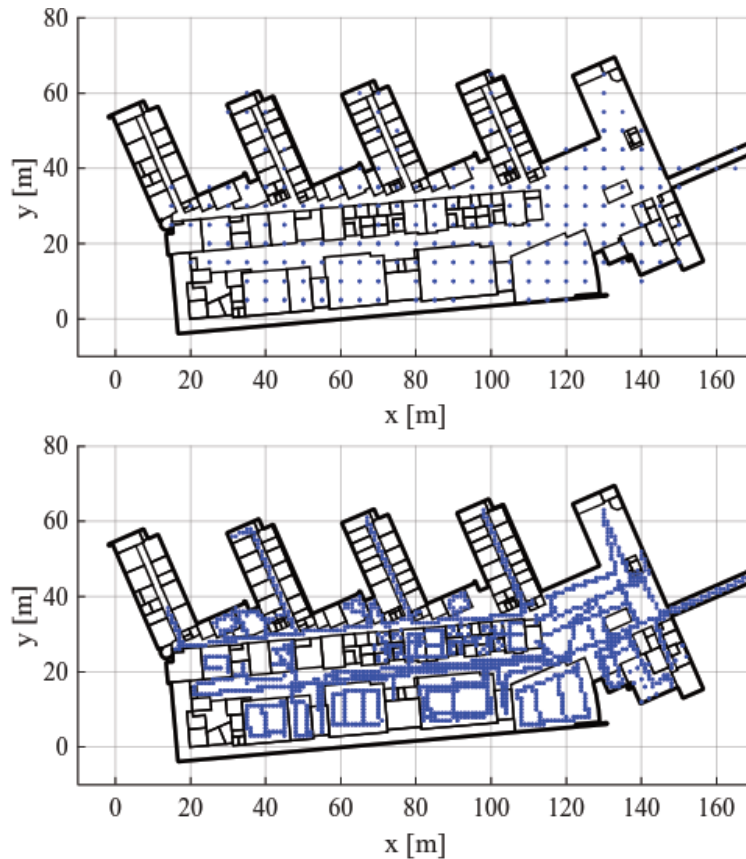
WLAN fingerprinting -menetelmä jakautuu kahteen vaiheeseen: oppimisvaiheeseen ja paikannusvaiheeseen. Oppimisvaiheessa kerätään dataa eli sormenjälkiä kohdealueesta. Paikannusvaiheessa käyttäjän laite paikannetaan käyttäen hyväksi aikaisemmin kerättyä oppimisdataa. Kuvassa 1 on kuvattu fingerprinting-menetelmän kahta vaihetta. Kuvan 1 osassa *a* on kuvattuna oppimisvaihe ja osassa *b* paikannusvaihe. Käyttäjän laitteen mittaamia RSS-arvoja vertaillaan tietokannassa

olevaan dataan ja määritetään käyttäjän sijainti. [2] Käyttäjän sijainnin määrittelemiseen voidaan käyttää useita erilaisia algoritmeja. Näitä erilaisia algoritmeja käsitellään luvussa 2.2 tarkemmin.



Kuva 1. Fingerprinting-menetelmän kaksi vaihetta [7].

Jotta käyttäjä voisi paikantaa laitteensa käyttäen fingerprinting-menetelmää, täytyy olla olemassa tietokanta, jossa on sormenjälkiä. Oppimisvaiheessa kerätään tietokantaan dataa kohdealueesta. Jokaisesta pisteestä kerätään ainakin sen koordinaatit ja pisteessä kuultavien WLAN-tukiasemien RSS-arvot. Termi tälle tiedolle on sormenjälki. Oppimisvaiheen jälkeen tietokanta koostuu sormenjäljistä, joissa on koordinaattien lisäksi RSS-arvot yhdestä tai useammasta WLAN-tukiasemasta. Yhdestä tukiasemasta voi olla yksi tai useampi RSS-arvo riippuen siitä, kuinka monta tukiasemaa kyseisessä pisteessä kuuluu. Lisäksi yhdestä tukiasemasta voidaan saada mitatuksi useampi kuin yksi RSS-arvo, jos tukiasemassa on monta antennia. Kuvassa 2 on kaksi esimerkkiä sormenjälkiruudukoista. Ensimmäisessä esimerkissä sormenjälkien välinen etäisyys on viisi metriä ja toisessa yhden metrin. Paikannusvaiheessa saatava paikannuksen tarkkuus riippuu suuresti siitä, miten oppimisvaihe on suoritettu. Jotta paikannus olisi mahdollisimman tarkka, täytyy sormenjälkiä kerätä mahdollisimman laajalti ja mahdollisimman tiheästi.



Kuva 2. Kaksi eri sormenjälkiruudukkoa Tampereen yliopiston Tietotalosta [2].

Tarvittavien sormenjälkien määrä riippuu valittavasta etäisyydestä sormenjälkien välillä. Jokaiseen sormenjälkeen liittyy koordinaatti, joka käyttäjän täytyy itse määrittää karttapohjalla. Koordinaatit jokaiselle sormenjäljelle täytyy määritellä käsin, koska sisätiloissa GNSS-pohjainen paikannus ei ole riittävän tarkka. [2, 8] Tästä syystä oppimisvaiheessa voi tulla inhimillisiä virheitä, jotka vaikuttavat paikannusvaiheessa tarkkuuteen.

Oppimisvaiheessa sekä paikannusvaiheessa mitattavat RSS-arvot ovat herkkiä ympäristön muutoksille. Arvot muuttuvat, kun ympäristössä tapahtuu muutoksia. Tällaisia muutoksia voi olla vaikkapa oven asennon muuttuminen tai huonekalujen järjestyksen vaihtuminen. Jo se, mihin suuntaan käyttäjä seisoo, vaikuttaa RSS-arvoon [9]. Muutokset fyysisessä ympäristössä vaikuttavat siis suuresti RSS-arvoihin. Oppimisdata ei siis välttämättä ole pitkään ajan tasalla. Tietokantaan kerättävän oppimisdatan kerääminen ja päivittäminen ovat siis aikaa vaativaa. Siksi yhtenä vaihtoehtona oppimisvaiheen toteuttamiseen on joukkouttaminen. Joukkouttamisessa vapaaehtoiset ihmiset keräävät sormenjäljet.

2.2 Algoritmit

Käyttäjän sijainti voidaan määrittää usealla eri algoritmilla. Kaikissa näissä on kuitenkin sama perusperiaate. Käyttäjän mittaamia RSS-arvoja vertaillaan tietokannassa olevaan oppimisdataan. Käyttäjän sijainti määritellään käyttäen yhden tai useamman parhaiten sopivan sormenjäljen koordinaatteja. WLAN fingerprinting -pohjainen paikannus onkin siis pohjimmiltaan luokitteluongelma, jossa käyttäjän mittaamat arvot yritetään luokitella oikeaan sormenjälkeen. [2]

Fingerprinting-menetelmässä käytettävät algoritmit voidaan jakaa kahteen kategoriaan, deterministisiin ja todennäköisyyteen perustuviin menetelmiin. Deterministisissä menetelmässä käyttäjän sijainnin ei ajatella olevan satunnainen, kun taas todennäköisyyteen perustuvissa menetelmissä sen ajatellaan olevan satunnainen muuttuja, jolla on todennäköisyysjakauma. Deterministisillä menetelmillä käyttäjän sijainnille saadaan arvio, kun taas todennäköisyyteen perustuvilla menetelmillä saadaan myös todennäköisyys sille, että käyttäjä on kyseisessä sijainnissa. [2, 10, 11]

Deterministisissä menetelmissä lasketaan ennalta määrätty funktio jokaisessa sormenjäljessä. Yleisesti käytetty funktio on Euklidinen etäisyys $\|.\|^2$ käyttäjän mittaamien arvojen ja sormenjälkien arvojen välillä. Tämän jälkeen valitaan sopivin koordinaatti käyttäen yhtä tai useampaa sormenjälkeä. Yleisesti käytettyjä deterministisiä menetelmiä ovat Nearest Neighbor (NN), K-Nearest Neighbor (KNN) ja Weighted K-Nearest Neighbor (WKNN). [1, 2, 7]

Determinististä menetelmistä NN on yksinkertaisin. Siinä lasketaan käyttäjän mittaamien signaalien ja tietokannassa olevien signaalien voimakkuuksien etäisyys kaikissa sormenjäljissä Euklidisista etäisyyttä $\|.\|^2$ käyttäen. Esimerkiksi kolmen kuullun WLAN-tukiaseman tapauksessa käyttäjän mittaamien arvojen [rss1, rss2, rss3] ja sormenjäljen arvojen [RSS1, RSS2, RSS3] erotus olisi

$$d = \sqrt{((rss1 - RSS1)^2 + (rss2 - RSS2)^2 + (rss3 - RSS3)^2)}$$

Sormenjälki, jolle tämä etäisyys d on lyhyin, on arvioitu käyttäjän sijainti. KNN menetelmä on muuten sama kuin NN, mutta siinä otetaan huomioon k kappaletta sormenjälkiä, joiden signaalien voimakkuuksien etäisyydet ovat pienimmät. Näistä k kappaleesta

sormenjälkien koordinaateista lasketaan keskiarvo. Saatu keskiarvo on käyttäjän sijainti. WKNN-menetelmässä näitä sormenjälkiä vielä painotetaan eri kertoimilla. Yksi tapa on painottaa käyttämällä etäisyyden laskemiseen käytetyn funktion käänteisfunktioita. [1, 2]

Yleisesti KNN:ää ja WKNN:ää käytettäessä voidaan saada tarkempia tuloksia, kuin yksinkertaisempaa NN:ää käytettäessä. Jos kuitenkin sormenjälkien välinen etäisyys on tarpeeksi pieni, NN voi suoriutua paremmin kuin monimutkaisemmat algoritmit. [2] Käyttämällä K:n arvoja 3 ja 4 saadaan parhaita tuloksia. Kaksi lähintä naapuria ei siis riitä hyvän tarkkuuden saamiseksi, koska osa hyödyllisestä informaatiosta jätetään huomioimatta. Jos taas käytetään yli neljää lähintä naapuria, tarkkuus pienenee, koska osa naapureista on liian kaukana todellisesta sijainnista. [9]

Bayesin metodi on yksi todennäköisyyteen perustuva algoritmi, jolla voidaan määrittää käyttäjän sijainti. Bayesin sääntö on

$$p(s_t|h_t) = p(h_t|s_t)p(s_t) * N$$

missä s_t on sijainti hetkellä t , h_t on havainto hetkellä t (hetkelliset RSS arvot). N on normalisointitekijä, joka varmistaa, että todennäköisyyksien summa on 1. Todennäköisyys käyttäjän olemiselle sijainnissa s havainnolla h on yhtä suuri todennäköisyys havainnon h tekemiselle sijainnissa s ja todennäköisyys olla ylipäätään sijainnissa l . Paikannusvaiheessa tämä ehdollinen todennäköisyys lasketaan kaikille sormenjäljille. Kaikista todennäköisimmän sormenjäljen sijainti on käyttäjän arvioitu sijainti.

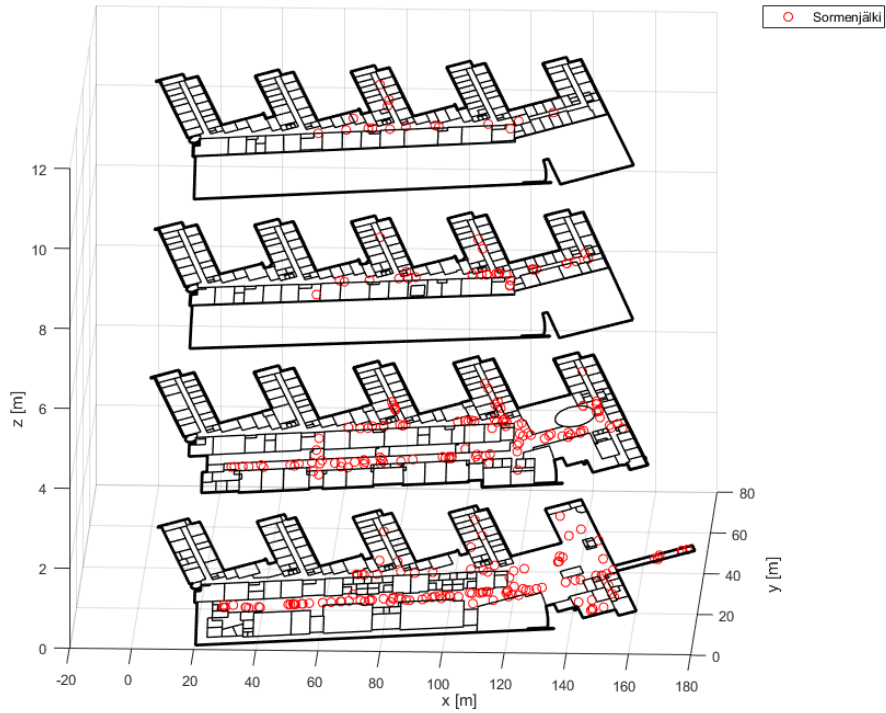
3. MITTAUKSET

Tässä työssä tehtyjen mittauksien tarkoituksena oli tutustua WLAN fingerprinting -menetelmään käytännössä. Mittauksissa kerättiin sormenjälkiä Tampereen yliopiston Tietotalosta. Mittaukset suoritettiin 2.10.–20.10.2017. Näiden sormenjälkien avulla tehtiin MATLAB:ssa erilaisia mittauksia. Oppimisvaihe, paikannusvaihe sekä virhearvio esitellään seuraavissa kappaleissa.

3.1 Oppimisvaihe

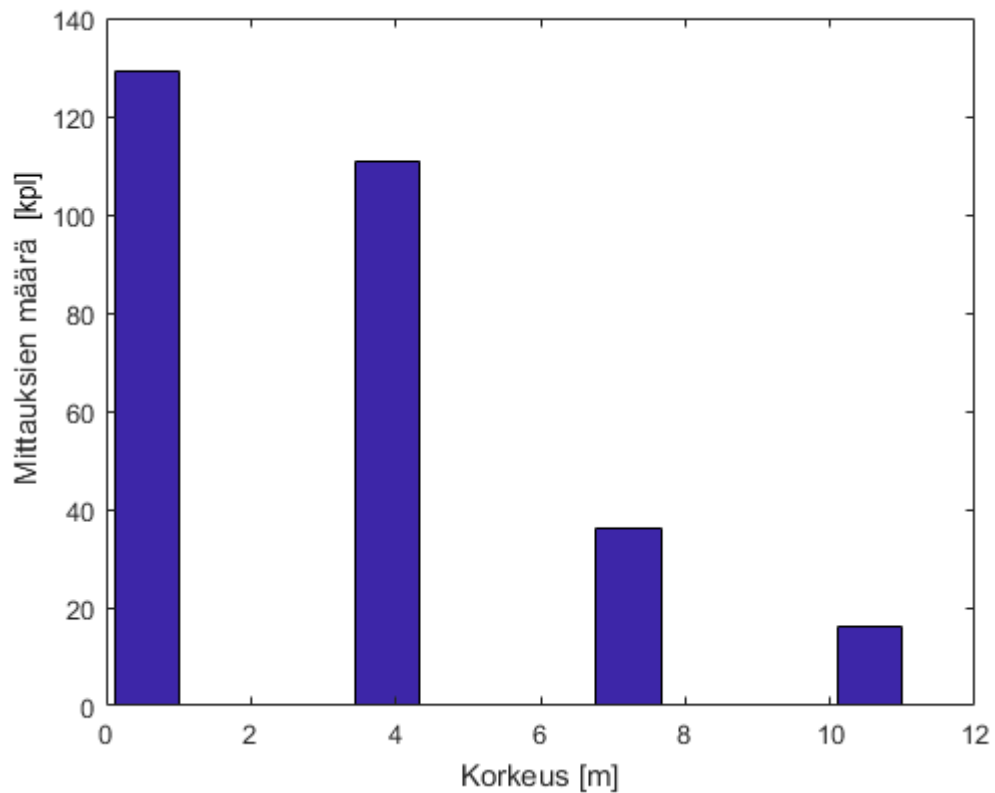
Mittauksissa käytettävä data kerättiin puhelimella käyttämällä TTY:n opiskelijoiden kehittämää sovellusta, jonka minulle toimitti Simona Lohan. Mittaukset tehtiin TTY:n Tietotalossa. Data kerättiin ympäri Tietotaloa eri kerroksissa. Painamalla nappia sovellus tallensi tiedon siinä sijainnissa kuultavista WLAN-tukiasemista ja niiden RSS-arvoista. Tämän jälkeen käyttäjä valitsi sovelluksen kartalta oikean kerroksen ja oman sijaintinsa tässä kerroksessa. Tämä käyttäjän sijainti tallennettiin yhdessä RSS-arvojen kanssa. Sormenjälkeen liitettiin vielä päivämäärä ja käyttäjän laitteen mallin nimi. Tämän jälkeen sovellus lähetti sormenjäljen palvelimelle.

Kun kaikki mittaukset oli suoritettu, sormenjäljet ladattiin palvelimelta JSON-tiedostona. Tämä JSON-tiedosto parsittiin MATLAB:ssa, jotta sitä voisi käsitellä. Palvelimelta ladatussa tiedosto sisälsi myös muiden henkilöiden tekemiä mittauksia. Jokaiseen sormenjälkeen liitetyn laitteen mallin avulla suodatettiin muut kuin tämän työn tekijän mittaamat sormenjäljet pois. Kuvassa 3 näkyvät nämä mitatut sormenjäljet. Mittapisteitä on tässä yhteensä 292 kappaletta.



Kuva 3. Mitatut sormenjäljet Tampereen yliopiston Tietotalossa.

Suurin osa mittapisteistä sijaitsevat ensimmäisessä ja toisessa kerroksessa. Kaikkiin käytäviin ei ollut pääsyä, joten näistä käytävistä ei ole mittapisteitä. Tutkimuksen kannalta olisi ollut parempi, jos mittapisteitä olisi tasaisemmin eri kerroksista sekä eri käytäviltä ja huoneista. Tällöin paikannustarkkuus olisi mahdollisesti parempi. Nyt suurin osa mittapisteistä on samasta tilasta ja useimpien eri mittapisteiden välissä ei ole seiniä tai muita esteitä. Kuvassa 4 on mittapisteiden jakauma eri kerroksittain. Jokaisen mittapisteen ja kolmen lähimmän naapurin sijainnin keskiarvon välinen etäisyys oli keskimäärin 1,9 metriä.

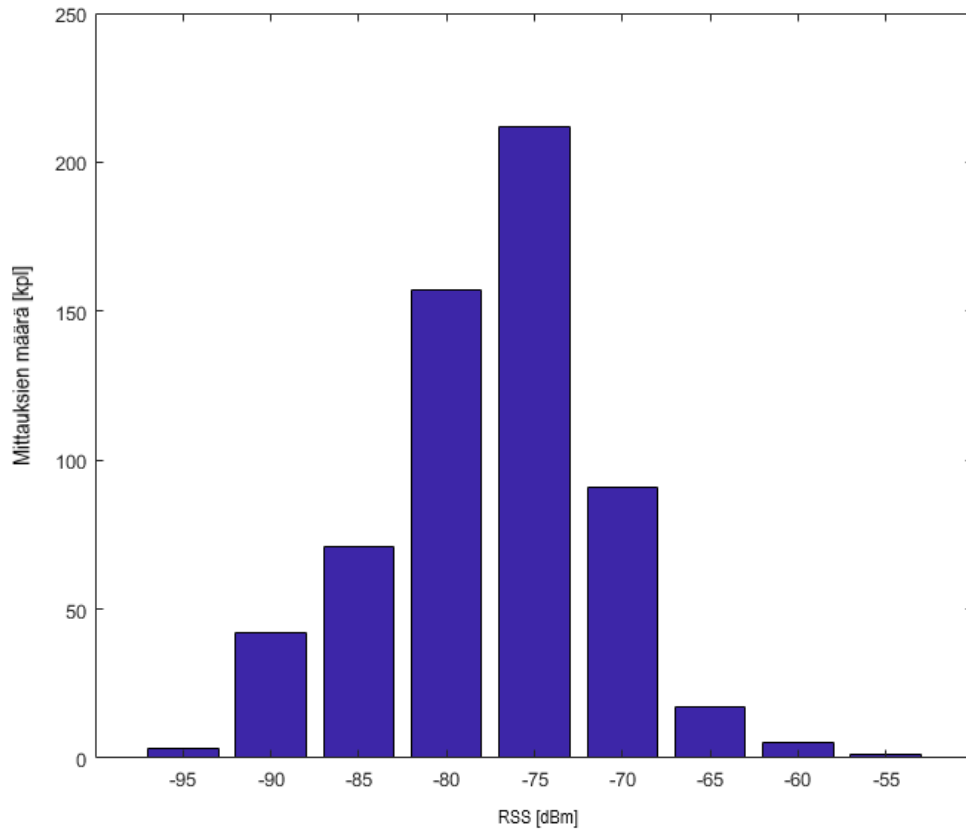


Kuva 4. Mittapisteiden jakauma kerroksittain.

3.2 Paikannusvaihe

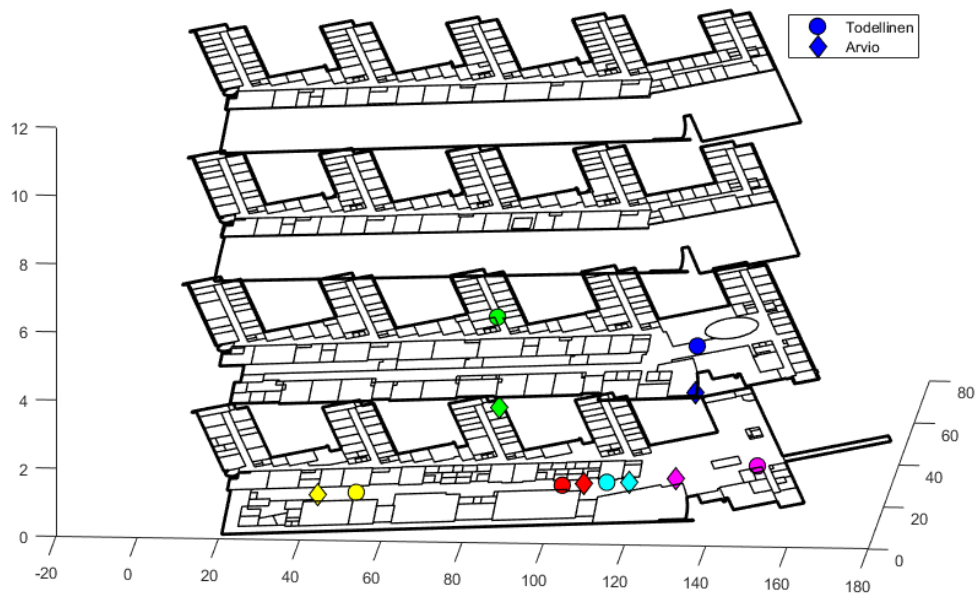
Tietokannassa oli 599 kappaletta uniikkeja MAC-osoitteita. WLAN-tukiasemia oli todellisuudessa vähemmän, koska yhdellä WLAN-tukiasemalla voi olla useampi MAC-osoite. MAC-osoitteita, jotka kuultiin vain alle kymmenessä mittapisteessä, oli 195 kappaletta. Nämä 195 MAC-osoitetta saattavat olla esimerkiksi matkapuhelimien omia WLAN-tukiasemia tai viereisten rakennusten sisällä olevia WLAN-tukiasemia.

Yhdessä sormenjäljessä kuultiin keskimäärin 71 MAC-osoitetta. Yksi MAC-osoite kuultiin keskimäärin 36:ssa sormenjäljessä. WLAN-tukiasemien keskimääräinen RSS-arvo oli -74.7 dBm. Kuvassa 5 näkyy WLAN-tukiasemista saatujen RSS-arvojen jakauma.



Kuva 5. WLAN-tukiasemista saatu keskimääräinen RSS-arvo.

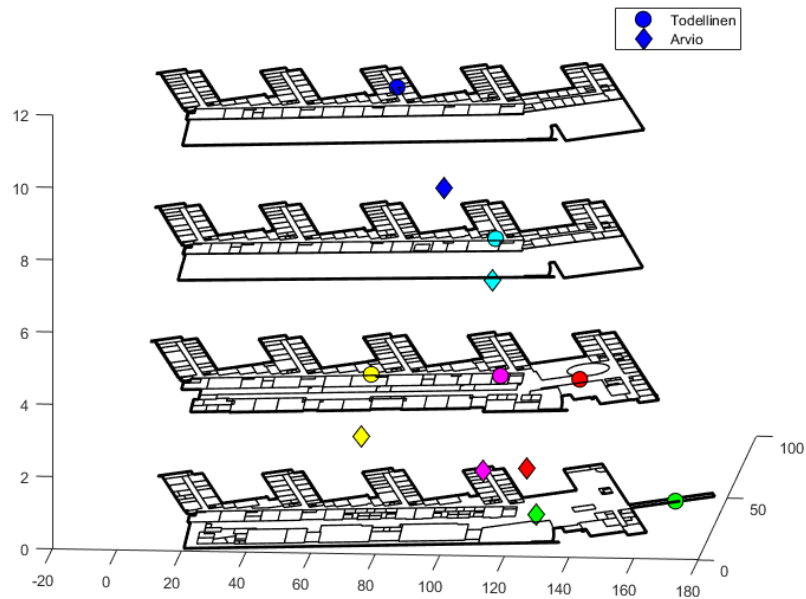
Mittauksissa käytettävät 292 sormenjälkeä jaettiin kahteen osaan satunnaisesti tietyn prosentien mukaan. Käyttämällä osaa sormenjäljistä muodostettiin oppimisdata. Käyttämällä tätä oppimisdataa, loppujen sormenjälkien avulla testattiin paikannusta ja sen tarkkuutta. Oppimisdatan määrää vaihdeltiin ja testattiin sen vaikutusta paikannustarkkuuteen. Kuvassa 6 on esimerkki paikannuksesta, kun 292 sormenjälkeä oli jaettu puoliksi oppimisdataan ja paikannusdataan.



Kuva 6. *Esimerkki paikannuksesta, kun oppimisdataa on 50%.*

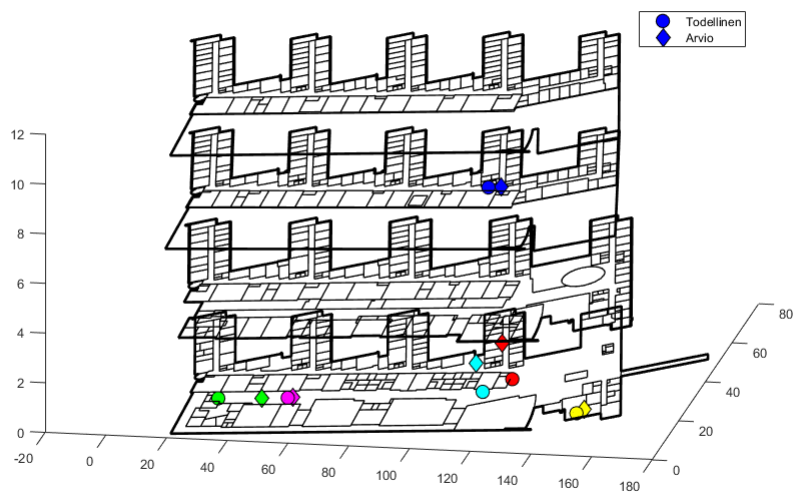
Paikannustarkkuutta mitatessa paikannusdatassa olevista sormenjäljistä tarkasteltiin vain niiden RSS-arvoja ikään kuin niiden oikeaa sijaintia ei tiedettäisi. Bayesin metodiin perustuvalla algoritmilla laskettiin todennäköisyys paikannusdatassa olevalle sormenjäljelle jokaisessa oppimisdatan sormenjäljessä. Tämän jälkeen käytettiin 3NN:n menetelmää eli näistä oppimisdatan sormenjäljistä valittiin 3 sormenjälkeä, joilla oli suurin todennäköisyys. Näiden 3 sormenjäljen koordinaateista laskettiin keskiarvo, joka oli lopullinen arvio käyttäjän sijainnista. Tämän jälkeen laskettiin erotus algoritmin laskeneen sijainnin ja todellisen sijainnin välillä.

Tällaisista paikannustarkkuuden mittauksista piirrettiin kartalle kuusi esimerkkiä, joissa näkyy algoritmin laskema arvio sekä todellinen sijainti. Kuvissa 6, 7 ja 8 on näitä esimerkkejä. Todellinen sijainti on merkitty ympyrällä ja arvio salmiakkikuviolla. Samanväriset kuviot ovat pareja keskenään. Kuvassa 7 oppimisdataa on 5 % ja kuvassa 8 oppimisdataa on 95 % kaikista sormenjäljistä.



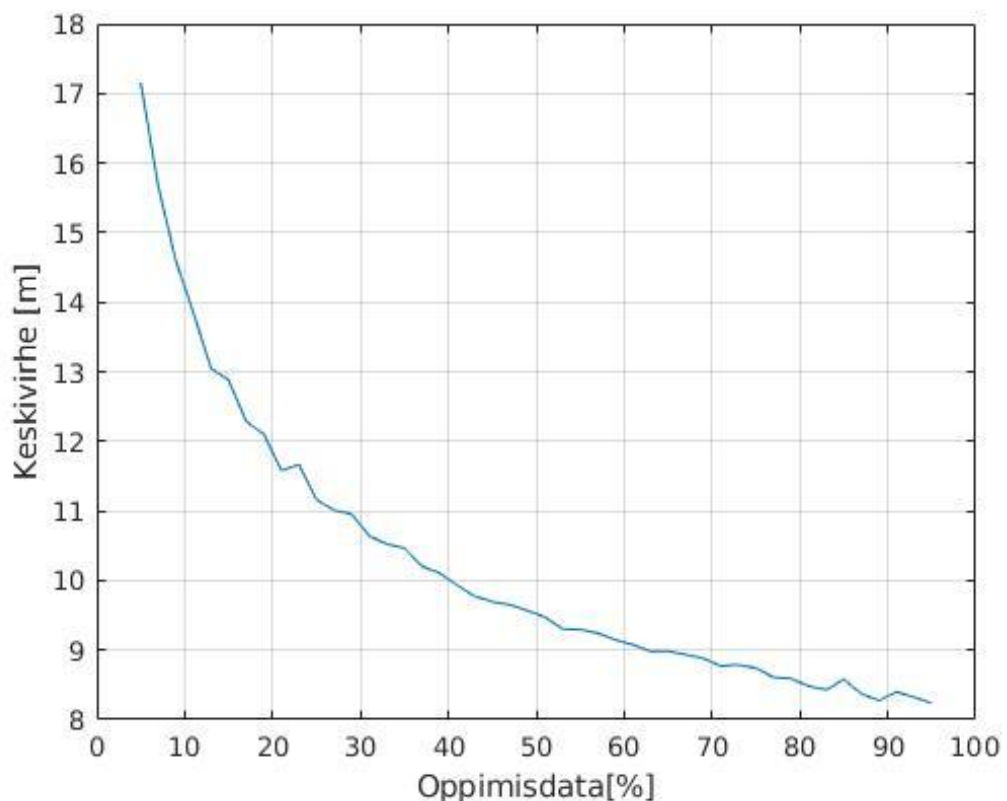
Kuva 7. Esimerkki paikannuksesta, kun oppimisdataa on 5%.

Algoritmi käyttää arviointiin 3 sormenjälkeä, joille todennäköisyys on suurin. Tästä syystä osa paikannusarviosta on kerrosten välissä. Esimerkiksi kuvassa 7 keltainen ja sininen arvio ovat kerrosten välissä. Vertailemalla kuvaa 7 ja 8 huomaa eron paikannustarkkuudessa, kun oppimisdatan määrää kasvatetaan. Paikannustarkkuus kasvaa, kun oppimisdatan määrää kasvatetaan.



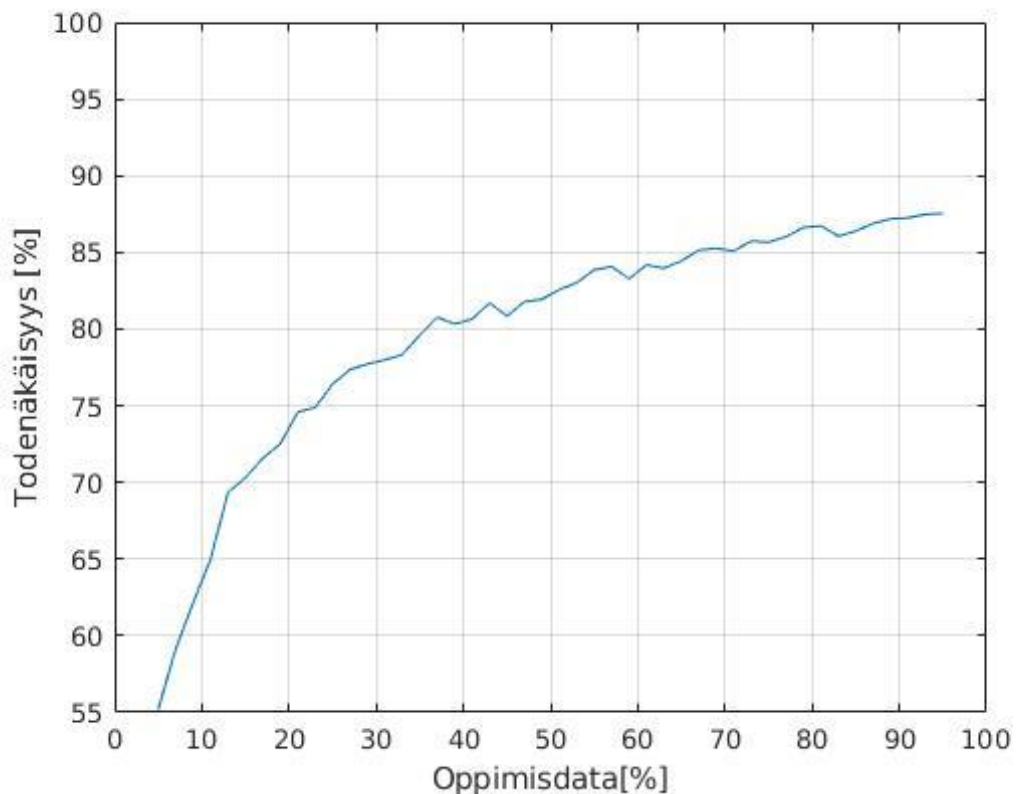
Kuva 8. Esimerkki paikannuksesta, kun oppimisdataa on 95%.

Seuraavaksi tutkittiin tarkemmin, kuinka oppimisdatan määrä vaikuttaa paikannustarkkuuteen. Tätä tutkittiin oppimisdatan prosentiosuuksilla 2 prosenttiyksiköstä 98 prosenttiyksikköön. Kun sormenjäljet oli jaettu oppimisdataan ja paikannusdataan, kaikille paikannusdatan sormenjäljille laskettiin algoritmilla sijaintiarvio. Tämän avulla laskettiin virhe sijainti-arvion ja todellisen sijainnin välillä. Sen jälkeen laskettiin keskiarvo kaikkien tapauksien virheistä. Tämä toistettiin 50 kertaa ja näistä keskiarvoista laskettiin yksi keskiarvo tälle prosenttiyksikölle. Toistoja tehtiin 50 kertaa, jotta satunnaisuutta olisi vähemmän. Tämä toistettiin kaikille prosentteille. Tulokset ovat nähtävissä kuvassa 9. Kuvasta näkee, että keskivirhe pienenee voimakkaasti, kun oppimisdatan osuus kasvaa 2 prosentista noin 20 prosenttiin. Kun oppimisdatan osuus on 2 prosenttia, keskivirhe on yli 25 metriä. Kun oppimisdatan osuus on yli 50 prosenttia, keskivirhe on alle 10 metriä. Keskivirhe ei kuitenkaan laske alle viiden metrin, vaikka oppimisdatan osuus olisi yli 90 prosenttia.



Kuva 9. Keskivirhe sijaintiarvion ja todellisen sijainnin välillä.

Tästä selviää, että Tietotalon kokoisessa rakennuksessa sormenjälkiä pitää kerätä noin 150 kappaletta, jotta keskivirhe paikannuksessa olisi alle 10 metriä. Tähän tietysti vaikuttaa myös se, kuinka tasaisesti sormenjäljet on kerätty. Tässä tutkimuksessa sormenjäljet sijaitsivat keskimäärin 2 metrin etäisyydellä toisistaan. Paikannuksessa arvioidun kerroksen oikeuden todennäköisyyttä tutkittiin samalla tavalla, kuin keskivirhettä. Tämä oikean kerroksen todennäköisyys oppimisdatan osuuteen nähden on esillä kuvassa 10.



Kuva 10. Oikean kerroksen todennäköisyys.

Keskivirhe ja oikean kerroksen todennäköisyys muuttuvat samalla tapaa oppimisdatan määrän kasvaessa. Kun oppimisdatan määrä kasvaa, keskivirhe pienenee ja oikean kerroksen todennäköisyys kasvaa. Jos oppimisdatan osuus on 2 prosenttia, algoritmi arvioi kerroksen oikeaksi vain 50 prosentin todennäköisyydellä. Vaikka oppimisdatan osuus olisi yli 90 prosenttia, oikea kerros arvioidaan kuitenkin alle 90 prosentin todennäköisyydellä.

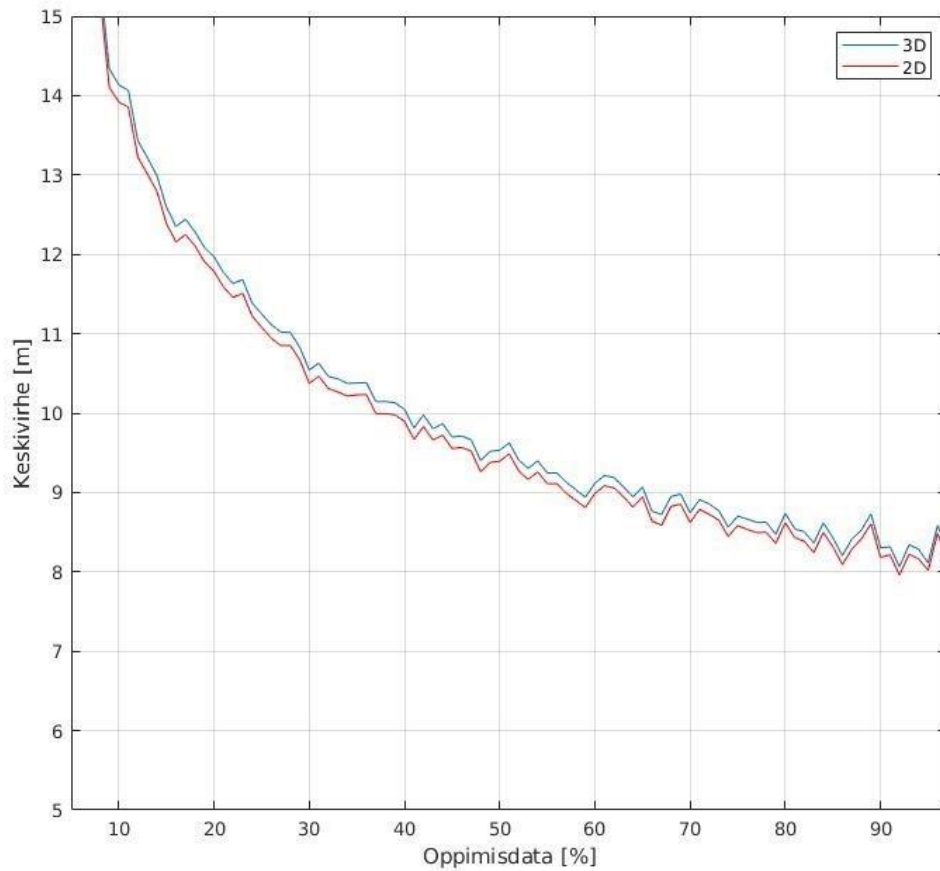
4. VIRHEARVIO

Paikannuksen tarkkuuteen vaikuttaa moni tekijä. Oppimisvaiheen toteutus vaikuttaa suuresti paikannusvaiheessa saavutettavaan tarkkuuteen. Suurin vaikuttava tekijä on paikannusvaiheessa kerättävien sormenjälkien määrä suhteutettuna kohdeympäristön pinta-alaan. Tässä työssä tehdyssä tutkimuksessa huomattiin, että alle 10 metrin paikannustarkkuuteen keskimäärin päästään, kun oppimisdataa on noin 150 sormenjälkeä. Pienempään paikannustarkkuuden virheeseen päästäisiin, jos sormenjälkiä kerättäisiin enemmän sekä tasaisemmin kuin tässä työssä tehtiin.

Myös tapa, jolla sormenjäljet kerätään, vaikuttaa paikannustarkkuuteen. Koska sormenjälkeen liitettävää todellista sijaintia ei voi määrittää satelliittipaikannuksen avulla riittävällä tarkkuudella, pitää sijainti määrittää muulla keinoin. Tässä työssä olevassa tutkimuksessa sormenjäljen todellinen sijainti määritettiin itse. Tämä tapahtui puhelimen näytöllä olevalle karttapohjalle merkitsemällä. Tämän vaiheen suorittamisessa voi tapahtua inhimillisiä virheitä, jotka vaikuttavat heti paikannustarkkuuteen. Myös ympäristön muutokset voivat vaikuttaa paikannustarkkuuteen. Jos oppimisdata on keräämisen ja paikannuksen välissä rakennuksessa on esimerkiksi purettu tai rakennettu seiniä, se vaikuttaa paikannustarkkuuteen. WLAN-tukiasemien fyysisen sijainnin muuttuminen vaikuttaa myös paikannustarkkuuteen, jos fyysinen sijainti muuttuu paljon.

Luvussa 3.2 esitellyistä tutkimustuloksista selviää, että paikannustarkkuus paranee vähemmän, mitä enemmän datasta käytetään oppimisdataan. Kun oppimisdataa käytetään 75%:n sijasta 95%, paikannustarkkuus paranee vain muutamalla metrillä. Tämä on nähtävissä kuvassa 9. Sama trendi on huomattavissa myös oikean kerroksen todennäköisyydessä kuvassa 10. Niinpä on mahdollista, että väärän kerroksen arvio huonontaa sijainnin kokonaistarkkuutta metreissä huomattavasti, koska kerrosten väli on 3.7 metriä.

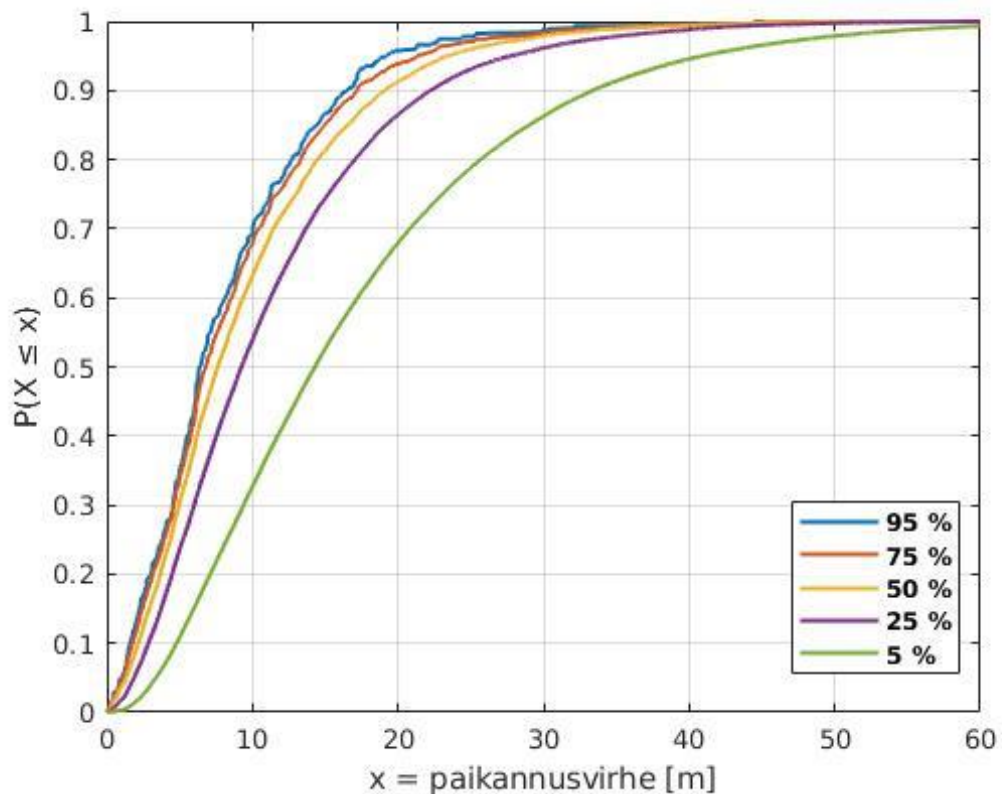
Tästä syystä tutkittiin erikseen paikannustarkkuutta 2D -tasossa eli oikeasta sijainnista sekä algoritmin antamasta arviosta sijainnille otettiin huomioon vain x ja y -koordinaatit. Tätä keskimääräistä 2D-virhettä verrattiin 3D virheeseen. Kuvassa 11 on nähtävissä tämä tulos.



Kuva 11: Keskivirhe per oppimisdataprocentti 3D- ja 2D-tasossa.

Kuvassa 11 sininen väri esittää keskivirhettä 3D-tasossa ja punainen väri 2D-tasossa. Tästä kuvasta näkee oikean kerroksen arvioinnissa tapahtuva mahdollinen virhe ei vaikuta lopulliseen virheeseen paikannuksessa juurikaan. Selvästi paikannustarkkuus paranee, jos tuloksia tarkastellaan vain kahdessa ulottuvuudessa, mutta ei merkittävästi. Ja jos käyttäjälle annettaisiin vain kaksiulotteinen tulos, tämä alle puolen metrin parannus kaksiulotteisessa tarkkuudessa, se ei olisi järkevää, koska tällöin käyttäjä ei saisi tietoa kerroksesta ollenkaan.

Paikannusvirheelle laskettiin myös kertymäfunktio käyttämällä viittä eri opetusdataprosenttia. Opetusdataprosenteiksi valittiin kertymäfunktioita varten 5, 25, 50, 75 ja 95. Näillä eri opetusdataprosenteilla 292 sormenjäljen data jaettiin tämän prosentin mukaan. Sen jälkeen lopuilla sormenjäljillä laskettiin sijaintiarvio ja virhe todellisen sijainnin ja sijaintiarvion välillä tallennettiin. Tämä toistettiin jokaiselle viidelle opetusdataprosentille 500 kertaa, jotta saatiin riittävästi dataa kertymäfunktioita varten. Näiden kertymäfunktioiden kuvaajat ovat nähtävissä kuvassa 12.



Kuva 12: Kertymäfunktio paikannusvirheelle eri opetusdataprocenteilla.

Tässä ei ollut merkityksellistä, mitkä nämä prosenttiyksiköt opetusdatalle täsmälleen ovat. Enemmän oli merkityksellistä valita tasaisesti eri arvoja väliltä 0-100. Lisäksi käyttämällä viittä eri arvoa, kuvaan saatiin enemmän informaatiota, mutta se pysyi edelleen selkeänä lukea. Tästä kuvasta on mahdollista lukea millä todennäköisyydellä paikannusvirhe X on pienempi tai yhtä suuri kuin x metriä. Esimerkiksi lukemalla kuvasta violettia viivaa, mikä kuvaa opetusprosenttia 25, on nähtävissä, että noin 55% todennäköisyydellä paikannusvirhe on alle 10 metriä. Lisäksi tästä kuvasta 12 on nähtävissä sama tieto, kuin kuvasta 9, eli paikannusvirhe pienenee, kun opetusdatan määrä kasvaa. Erona on se, että kuva 9 esittää keskivirhettä ja kuvasta 12 on mahdollista kertoa millä todennäköisyydellä virhe on alle x metriä. Kuvaan 12 käytetystä datasta poimittiin vielä erikseen 50 todennäköisyydelle raja-arvot eri opetusdataprocenteille. Sama tehtiin 68 ja 95 prosentin todennäköisyyksille. Nämä tulokset ovat nähtävissä taulukossa 1.

Taulukko 1. Kertymäfunktioista kerätyt arvot paikannusvirheelle eri oppimisdataprosenteille.

Opetusdataprosentti (%)	$P(X \leq x)=50\%$ (m)	$P(X \leq x)=68\%$ (m)	$P(X \leq x)=95\%$ (m)	keskiarvo (m)
5	14,2	20,0	42,6	17,2
25	9,2	13,2	28,8	11,3
50	7,7	11,2	23,8	9,6
75	6,9	10,1	22,0	8,7
95	6,3	9,7	19,8	8,2

Lisäksi taulukon 1 viimeisessä sarakkeessa on paikannusvirheen keskiarvo. Tästä taulukosta on nyt mahdollista nähdä esimerkiksi, että opetusdataprosentilla 25 paikannusvirhe on 95% todennäköisyydellä 28,8 metriä tai vähemmän. Ja opetusdataprosentilla 95 vastaavasti paikannusvirhe on 95% todennäköisyydellä 19,8 metriä tai vähemmän.

5. YHTEENVETO

Tässä kandidaatintyössä käsiteltiin WLAN fingerprinting -pohjaista laitepaikannusta. Tämä on yksi tapa toteuttaa laitepaikannus sisätiloissa, jolloin satelliittipaikannus ei ole sopiva vaihtoehto esimerkiksi riittämättömän tarkkuuden vuoksi. Laitepaikannusta on mahdollista toteuttaa myös toisilla tavoilla ja tekniikoilla, mutta tässä työssä käsiteltiin vain WLAN fingerprinting -pohjaista laitepaikannusta.

Toisessa kappaleessa esiteltiin teoria, johon WLAN fingerprinting -pohjainen laitepaikannus perustuu. Se jakautuu kahteen vaiheeseen, opetus- ja paikannusvaiheeseen. Opetusvaiheessa kerätään sormenjälkiä kohdeympäristöstä ja paikannusvaiheessa voidaan erilaisia algoritmeja hyödyntämällä laskea sijaintiarvio käyttäjän laitteelle. WLAN fingerprinting -pohjainen laitepaikannus hyödyntää tärkeimpänä suureenaan RSS-arvoa. RSS-arvo on suure, joka ilmaisee WLAN-tukiaseman signaalin vahvuuden käyttäjän laitteelle. Yksi sormenjälki on tieto, jossa yhteen fyysiseen sijaintiin on yhdistettynä siinä pisteessä kuultavat WLAN-tukiasemat sekä niistä saatavat RSS-arvot. Lisäksi toisessa kappaleessa esiteltiin tarkemmin paikannusvaiheessa käytettäviä algoritmeja.

Tähän kandidaatintyöhön kuului myös osuus, jossa kokeiltiin WLAN fingerprinting -menetelmää käytännössä. Tämä osuus esiteltiin kappaleessa kolme. Oppimisvaihe suoritettiin keräämällä 292 sormenjälkeä Tampereen yliopiston Tietotalon neljästä eri kerroksesta. Tämä oppimisvaiheen suorittamisen jälkeen paikannusvaihetta mallinettiin MATLAB:ssa. Tärkeimpänä tutkimuksena testattiin oppimisdatan määrän vaikutusta lopulliseen paikannustarkkuuteen. Tätä varten laskettiin keskivirhe ja oikean kerroksen todennäköisyys eri opetusdataproseenteilla. Nämä tulokset ovat nähtävissä kuvissa 9 ja 10.

Neljännessä luvussa käsiteltiin paikannuksessa tapahtuvaa virhettä tarkemmin. Luvussa pohdittiin eri syitä virheen syntymiseen. Lisäksi MATLAB:n avulla laskettiin erikseen paikannusvirhe, kun korkeuskoordinaatti jätettiin huomioimatta. Paikannustarkkuutta kolmessa ulottuvuudessa tutkittiin tarkemmin kertymäfunktioiden avulla. Kertymäfunktioista saadut tulokset kerättiin taulukkoon 1. Kun oppimisdataan käytettiin 75% kaikista 292 kerätyistä sormenjäljistä, oppimisdatana oli 219 sormenjälkeä ja paikannusvaiheessa käytettiin loppuja 73 sormenjälkeä. Tällöin paikannuksen keskivirhe oli 8,6 metriä. Ja 98% todennäköisyydellä paikannusvirhe oli 28,9 metriä tai vähemmän.

Oppimisvaihe on siis todella tärkeä, jotta WLAN fingerprinting -menetelmä toimii halutulla tavalla. Jos sormenjäljet kerätään huolellisesti ja tarkasti, helpoin tapa parantaa paikannustarkkuutta on keräämällä lisää sormenjälkiä.

WLAN fingerprinting -menetelmällä ei siis tässä työssä päästy vastaaviin tarkkuuksiin kuin esimerkiksi satelliittipaikannuksella esteettömässä ulkotilassa. Mutta WLAN fingerprinting -menetelmää voidaan hyödyntää kuitenkin tapauksissa, joissa tällainenkin tarkkuus riittää. Käyttökohteena voi olla etenkin suuret rakennukset, joissa kuuluvuus satelliittipaikannukselle on huono tai olematon. Tällöin WLAN fingerprinting -pohjaista laitepaikannusta voisi hyödyntää. Yksi tämän menetelmän etu on se, että erillisiä laitteita ei tarvita. Tällöin voidaan säästää rahassa ja energiankulutuksessa. WLAN fingerprinting -pohjainen laitepaikannus on siis toimiva paikannusmenetelmä, joka hyödyntää jo olemassa olevia tekniikoita ja laitteita.

Syväoppimisen soveltuvuus sisätilapaikannukseen yleisesti on myös kiinnostava mahdollinen kehityssuuntaus. Syväoppimista käyttämällä WLAN-tukiasemista mitattavat RSS-arvot voisivat olla yksi monista sisääntuloista neuroverkolle. Muita käytettäviä suureita sisääntulossa voisi olla esimerkiksi kiihtyvyyssanturi ja GPS. Syväoppimisen avulla voisi myös mahdollisesti paikantaa käyttäjän liikettä ja nopeutta. [5]

LÄHTEET

- [1] T. Bagosi, Z. Baruch, Indoor localization by WiFi, IEEE 7th International Conference on Intelligent Computer Communication and Processing, 2011, s.449–452.
- [2] J. Talvitie, Algorithms and Methods for Received Signal Strength Based Wireless Localization, Tampere University of Technology, 2016, s.1–69.
- [3] D. Dardari, P. Closas, P.M. Djuric, Indoor Tracking: Theory, Methods, and Technologies, IEEE Transactions on Vehicular Technology, 2015, s.1263–1278.
- [4] N. Alsindi, Z. Chaloupka, N. AlKhanbashi, J. Aweya, An Empirical Evaluation of a Probabilistic RF Signature for WLAN Location Fingerprinting, IEEE Transactions on Wireless Communications, 2014, s.3257–3268.
- [5] K.S. Kim, S. Lee, K. Huang, A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting, Big Data Analytics, 2017, s.1–17.
- [6] Google Inc WifiInfo. Saatavissa (viitattu 21.11.2019):
<https://developer.android.com/reference/android/net/wifi/WifiInfo.html>.
- [7] B. Li, J. Salter, A.G. Dempster, C. Rizos, Indoor positioning techniques based on wireless LAN, Proceedings of the 1st IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, 2006, s. 130–136.
- [8] S. Ali-Löytty, T. Perälä, V. Honkavirta, R. Piche, Fingerprint Kalman Filter in indoor positioning applications, 18th IEEE International Conference on Control Applications, 2009, s.1678–1683
- [9] P. Bahl, V.N. Padmanabhan, RADAR: An in-building RF-based user location and tracking system, Proceedings - IEEE INFOCOM, 2000, s.775–784.
- [10] K. Li, J. Bigham, L. Tokarchuk, Validation of a Probabilistic Approach to Outdoor Localization, IEEE Wireless Communications Letters, 2013, s.167–170.
- [11] P. Prasithsangaree, P. Krishnamurthy, P. Chrysanthos, On indoor position location with wireless LANs, The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002, s.72–724.