

Reaaliaikaiseen mittaukseen perustuva tieliikennemalli päästölaskennalla, loppuraportti

Ville Rantanen, Tarek Mohsen, Hanna Hagström, Johan Himberg (Reaktor)

16.12. 2020

Helsingin kaupungilta projektipäällikkönä kokeilussa toimi Juho Kostiainen (kaupunginkanslia, elinkeino-osasto)

Tiivistelmä

Kokeilussa rakensimme ajoneuvojen päästötietoja tilastoivan prototyypilaitteiston ohjelmistoinen ja testasimme sen oikeassa käyttöympäristössä. Ajoneuvojen kilven tunnistamiseen käytettiin konenäköä ja päästötiedot oli tarkoitus saada ajoneuvorekisteristä. Kehityksen aikana keräsimme paljon kokemusta siitä, mitä kaikkea tuotantolaitteistossa ja sen asettamisessa liikenteen seurantaan on otettava huomioon. Selvitimme miten ajoneuvorekisteritiedot olisi saatavissa käyttöön Traficomilla ja miten tietoja täytyy käsitellä, jotta toisaalta anonymiteetti säilyy ja toisaalta että tiedot ovat riittäviä päästötilastojen parantamiseen.

Kahden tunnin jaksolla mittaisessa kenttätestissä Helsinginkadulla (päivänvalossa) laitteisto löysi kilven ja tunnisti sen oikein 89% niistä ajoneuvoista, jotka laitteisto ylipäättään pystyi havaitsemaan (yli 90% jos jätetään ulkomaalaiset kilvet huomiotta). Löydetyn kilven oikein tunnistaminen tapahtui yli 95% tapauksista. Ajoneuvojen laskeminen osoittautui laadultaan vaihtelevammaksi. Ajoneuvoja laitteisto löysi reilun 5 tunnin mittaisella laskentajaksolla päivänvalossa noin 79% silmukan laskemista. Onnistumisprosentti vaihteli kaistan, ajankohdan valaistuksen ja liikennetilanteen mukaan.

Kokeilusta tehtiin Helsingin kaupungin tietosuojavaikutusten arviointi. Kokeilussa kertyneet henkilötiedot: kuvamateriaali ja rekisteritunnukset tuhottiin kokeilun jälkeen.

Yhteenveto kokeilun tavoitteista ja onnistumisesta

Kokeilun tavoitteet

Helsingin kaupunki arvioi liikenteen päästöjä muun muassa liikenteenlaskennalla. Ajoneuvovirran suuruus ei kuitenkaan yksinään kerro päästöjen suuruutta, koska erilaisten ajoneuvojen jakauma liikenteessä vaihtelee ajan ja paikan myötä. Ajoneuvojen tekniset ominaisuudet löytyvät Traficomin ajoneuvorekisteristä, ja konenäön avulla voitaisiin tunnistaa ajoneuvon rekisterinumero ja hakea tunnetut päästötiedot. Tiedolla voitaisiin tarkentaa pelkkään liikennemäärään perustuvaa arviota päästöistä.

Liikenteen seurannassa käytetään jo konenäköjärjestelmiä, esimerkiksi parkkihalleissa, mutta nämä järjestelmät ovat usein suljettuja tuotteita ja palveluita. Tämän hankkeen tarkoitus oli selvittää, miten edullinen ja avoimeen lähdekoodiin perustuva järjestelmä

liikenteen päästötietojen tarkempaan arviointiin voitaisiin toteuttaa turvallisesti, ja arvioida toteutettavuus ja tarkkuus todellisella aineistolla. Projektissa rakennettiin prototyypilaitteisto ja ohjelmisto rekisterikilpien tunnistamiseen ja kokeiltiin sitä liikennevirrassa Helsingin kaupungin tarjouspyyntö kuvaa järjestelmän perusratkaisun:

“ - - -Kokeilussa selvitetään, mikä olisi kustannustehokas ja riittävän toimiva laiteratkaisu ajoneuvojen rekisterikilpien tunnistamiseen ja tämän tiedon avulla päästötietojen hakuun laitteeseen ladattuun kopioon ajoneuvorekisteristä. Rekisterinumeroita ei tallenneta yksityishenkilöiden tietosuojan turvaamiseksi.

Laitteiden lisäksi kehitetään avoimen lähdekoodin ratkaisu, joka kuvaa ajoneuvot, tunnistaa rekisterinumeroita, hakee ajoneuvotiedot laitteen omasta ajoneuvorekisterin kopiosta ja lähettää anonyymit ajoneuvotiedot palvelimelle.

Palvelimella tiedot yhdistetään koko kaupungin kattavaan liikennemalliin, jossa muodostetaan ajantasaista indikaattoritietoa. Prototyyppi testataan kenttäolosuhteissa. Kokeilussa kiinnitetään erityistä huomiota henkilöiden tietosuojaan. Toteutuksen tulisi tunnistaa mahdollisimman suuri osa mittauspisteiden ohi ajavista ajoneuvoista. - - -”

Kokeilun onnistuminen

Kokeilussa rakensimme prototyypilaitteiston ohjelmistoinen ja testasimme sen oikeassa käyttöympäristössä. Kehityksen aikana keräsimme paljon kokemusta siitä, mitä kaikkea tuotantolaitteistossa on otettava huomioon. Selvitimme miten ajoneuvorekisteritiedot olisi saatavissa käyttöön Traficomilla ja miten tietoja täytyy käsitellä, jotta toisaalta anonyymiteetti säilyy ja toisaalta että tiedot ovat riittäviä päästötilastojen parantamiseen.

Konenäön käyttö rekisterikilpien tunnistamiseen

Kamera- ja laskentalaitteiston määrittely ja rakentaminen

Laitteisto (ks. edempänä) saatiin määriteltyä, ja prototyyppi rakennettua. Kokeilujen perusteella suurempia laskentatehoja tuskin tarvitaan. Kenttätestiä varten saatiin rakennettua kohtuullinen koteloitinta ja löydettiin riittävän tehokas akku noin 16 tunnin yhtämittaista mittausta varten. Kokeilussa ei rakennettu jatkuvalla virransyötöllä toimivaa laitteistoa.

Ajoneuvojen tilastointi konenäöllä

Kokeilun päätarkoitus oli tunnistaa ajoneuvoja, jotta voidaan tehdä arvioita päästöistä.

Tavoite voidaan jakaa osiin:

- ajoneuvojen rekisteritunnuksen tunnistaminen, jotta päästötiedot voidaan hakea ja
- ajoneuvojen laskenta yleensä, vaikka rekisteritunnusta ei tunnistettaisikaan

Kokeilussa ei oltu asetettu tarkkaa etukäteisvaatimusta tunnistustarkkuudelle tai sen laskentatavalle, mutta keskusteluissa mainittiin että odotukset ovat yli 90% tarkkuudelle. Kokeilu keskittyi rekisteritunnuksen tunnistamiseen (teknologiana OpenALPR), mutta myös ajoneuvotyyppien tunnistaminen ilman kilpeä on mukana (teknologiana YOLOv5). Ajoneuvoa on voitava myös seurata kuvassa, jotta sitä ei lasketa uudestaan (teknologiana SORT). Teknologista tarkemmin myöhemmin.

Yhteenveto onnistumisesta:

- Jos konenäkö löytää ajoneuvon kilven, se tunnistuu verraten tarkasti
- Ajoneuvojen laskenta on hankalampi tehtävä

Tuloksiin kokeessa vaikutti erityisesti valaistus, havaintoalueen geometria ja myös ajoneuvovirran tiheys (nopeus/ajoneuvojen peittyminen). *On myös huomattava että testi tehtiin yhtenä päivänä, yhdenlaisissa olosuhteissa.* Olosuhteet valoisaankaan aikaan tuskin olivat ideaaliset (hämärää, satoi), joten joissain olosuhteissa voidaan päästä parempaan kiini. Toisaalta olosuhteet valoisaankin aikaan voisivat toki olla huonommatkin (kirkas vastavalo, lumisade jne.)

Seuraavassa tulosten yhteenveto. Kvantitatiiviset tulokset tarkemmin liitteessä.

Rekisteritunnuksen tunnistustarkkuus konenäön avulla

Kilpien tunnistus testitilanteessa päivänvalossa (to 26.11.2020) manuaalisesti tarkastetuilta jaksoilta 10:00-10:29, 12:30-13:30 ja 14:45-15:15.

Konenäön havaitsemia oli ajoneuvoja oli yhteensä näillä ajanjaksoilla

- 1179, joista
- 1087 ajoneuvosta konenäkö löysi kilven
- 1052 tapauksessa kilpi luettiin täysin oikein

Tunnistustarkkuutta voi laskea eri tavoin:

Tunnistusprosentti "täysin oikein luetut kilvet / konenäön havaitsemat ajoneuvot on (1052/1179) **89 %**. Jos otetaan huomioon, että manuaalisessa tarkastelussa todettiin että joukossa oli mm. 43 ulkomaalaista kilpeä, joita ei pitäisikään pystyä tunnistamaan / löytämään, olisi tunnistusprosentti korkeampi.

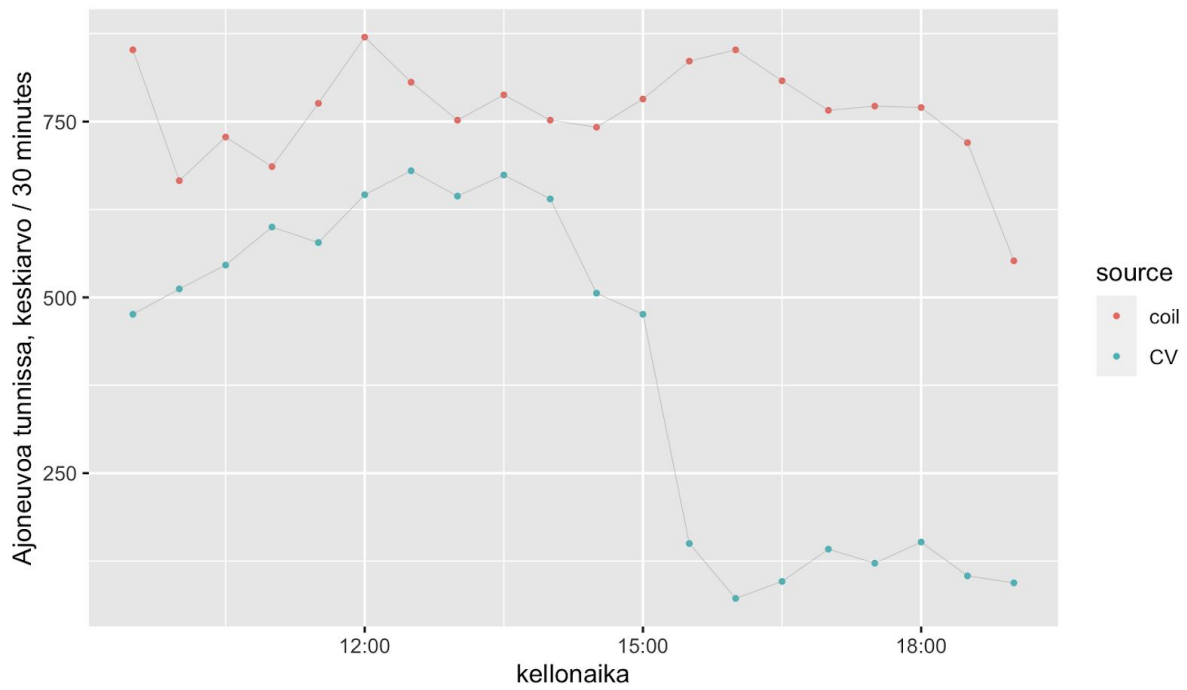
Tunnistusprosentti "täysin oikein luetut kilvet / konenäön löytämät kilvet on puolestaan (1052/1087) **97%**

Ajoneuvojen laskentatarkkuus konenäön avulla

Ajoneuvojen laskennassa verrattiin konenäön ("CV") havaitsemien ajoneuvojen määrää silmukan tuloksiin. Tarkkuus vaihtelee kellonajoittain ja kaistojen välillä. Valoisana aikana konenäkö löysi n. **79%** ajoneuvoista. (Järjestelmän ns. false positive ratea - "havaitaan ylimääräisiä autoja" - ei mitattu. Silmukan ja kameramittauksen aikaleimat eroavat toisistaan, joten saman auton identifiointi ei onnistu automaattisesti. FPR on manuaalisesti tarkastettujen jaksojen perusteella kuitenkin paljon pienempi ongelma kuin autojen havaitsematta jääminen.)

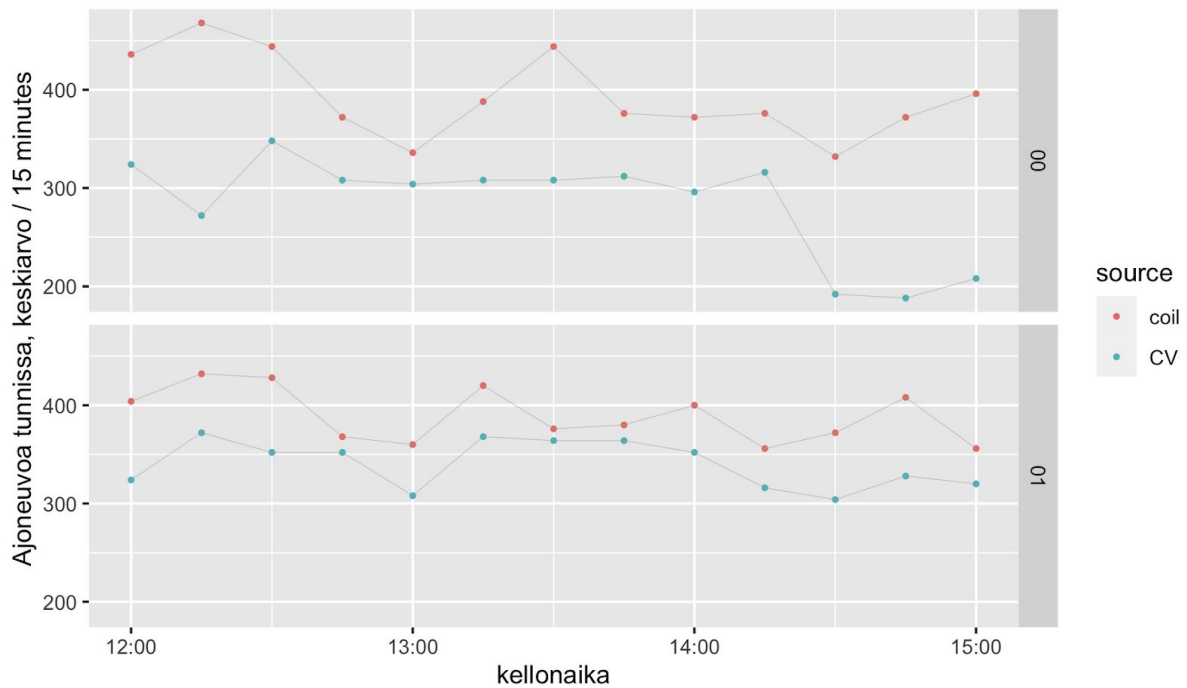
Aurinko nousi mittauspäivänä vähän ennen yhdeksää ja laski noin puoli neljältä. Valoisin ajan sää oli aamupäivällä sateinen, ja iltapäiväkin pilvinen, joten olosuhteet eivät olleet mitenkään täydelliset. Seuraavasta näkyy että laskenta ei oikein toimi hämärässä.

Mitattu keskimääräinen liikennevirta



Toisessa kuvassa näkyy että valoisana aikana, lähempänä kameraa olevalla kaistalla päästään jo lähelle silmukkamittauksen ("coil") määriä.

Mitattu keskim. liikennevirta kaistoittain



Kaista 01 on lähempänä kameraa.

Päästömallinus ajoneuvorekisterin teknisten tietojen perusteella

Kokeiluun ei saatu Traficomien tietokantaa ajoneuvojen rekisteritunnuksista ja teknisistä tiedoista. Tunnistamisen vaikutusta varsinaisen päästömallin tarkkuuteen ei siis voitu kokeilla, samoin ei voitu kokeilla validien rekisteritunnusten tietämisen vaikutusta kilpien tunnistamiseen.

Päästölaskennan arkkitehtuuri

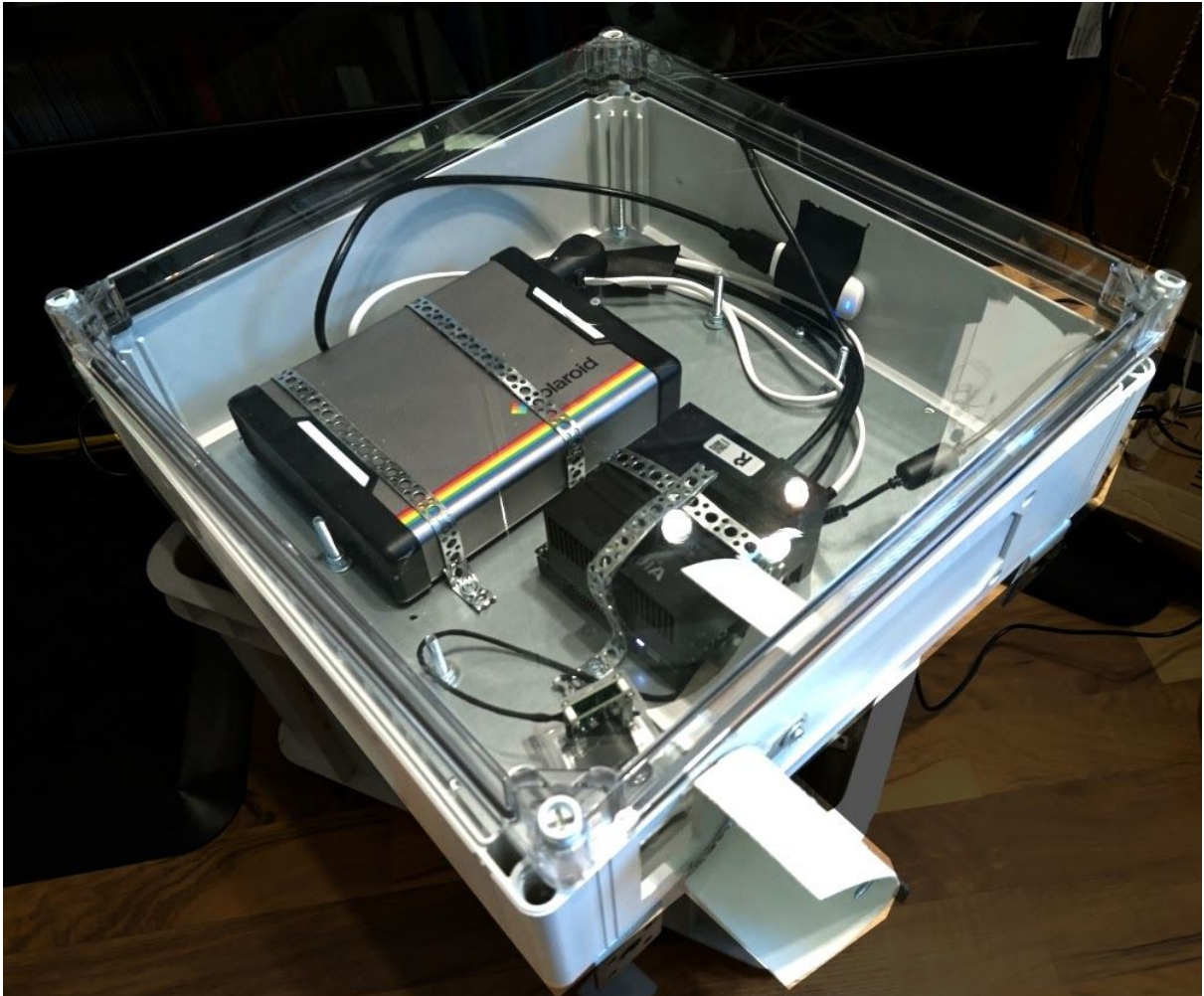
Tuotantokäyttö vaatii palvelimen, joka vastaanottaa päästötietoja ja vie tietoja päästölaskennan käyttöön. Toteuttaminen ei ollut kokeilun tavoitteille olennaista ja vaatii Helsingin kaupungin it:n suunnittelua. Sinänsä palvelu olisi suoraviivaista toteuttaa.

Kokeilun eteneminen

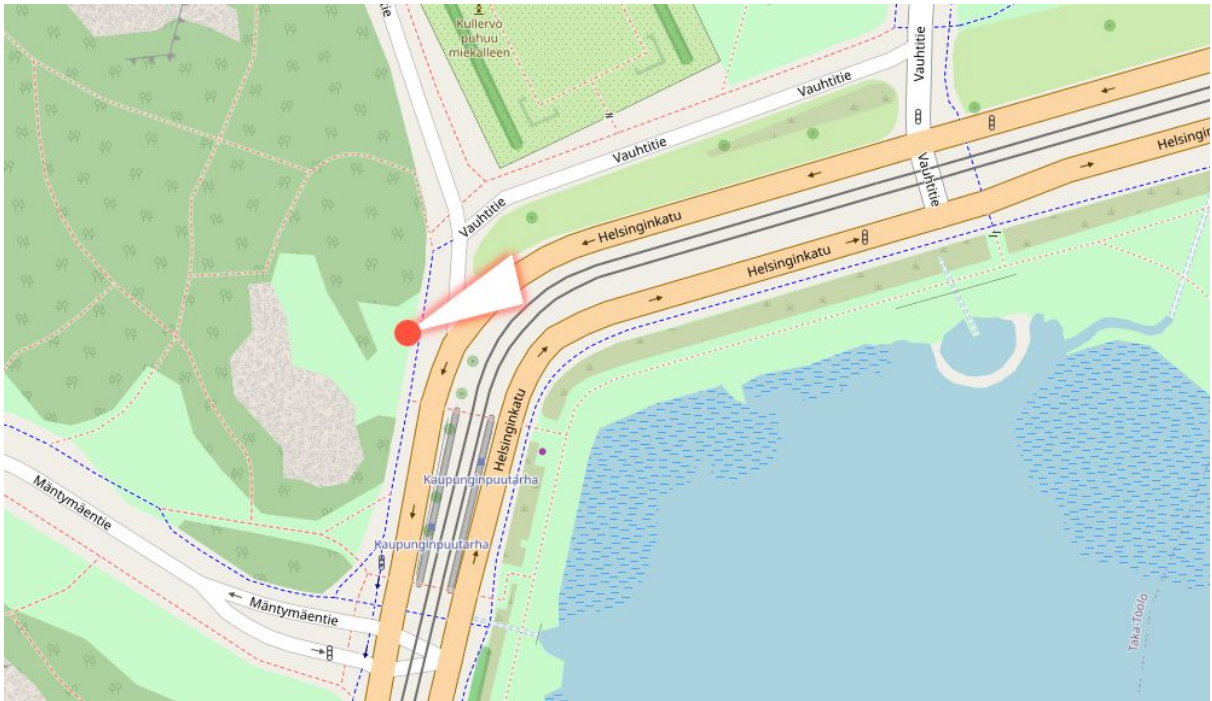
- Kick-off 18.9.2020
- Tietosuojavaikutusten arviointi 29.9. - 30.9.2020
- Määriteltiin riittävä laskentateho reaaliaikaiseen optiseen rekisterikilpien lukuun. Hankittiin ja asennettiin laitteisto, johon kuuluu laskentayksikkö, kamera ja akku.
- Keskustelu Traficomien kanssa ajoneuvorekisterin käyttömahdollisuuksista 27.10.2020
- Ohjelmistoa kehitettiin alkuun erillisellä kameralla kuvatulla videolla, ja sitten viemällä koteloitua laitetta lyhyisiin testijaksoihin kadun varteen (loka-marraskuu)
- Laitteiston kenttätesti Helsinginkadulla 26.11.2020. Paikassa saatiin ajoneuvomääristä referenssitieto kaupungin laskentapisteestä.
- Loppuraportti valmis 14.12.2020

Tiivistelmä kokeilun tuotoksista

- Rekisterikilpiä tunnistava ja ajoneuvovirtaa laskeva laitteisto
- Laitteiston open source -ohjelmisto (<https://github.com/City-of-Helsinki/remppa>)
- Tietosuojan vaikutustenarviointi
- Kenttätestauksen (26.11.) tulokset tunnistustarkkuudesta (liite)
- Loppuraportti ja esitys



Mittalaite koteloituna 400 mm x 400 mm asennuskoteloon. Ilman akkua (harmaa osa, jossa värilliset raidat) laitteen voisi laittaa huomattavasti pienempään koteloon.



Mittauspaikan sijainti ja suunta.



Kuva mittauspaikalta

Pohdintaa tekoälykokeilun näkökulmasta

Opit tekoälyn mahdollisuuksista ja kehittämisestä

Keskeinen oppi kokeilusta on nähdäksemme se että yhden tekoäly- tai konenäköongelman asemesta on ratkaistava monta, hieman eri tyyppistä ongelmaa, joista jokainen vaatii omaa paneutumistaan.

Konenäön tuloksien käyttökelpoisuuden kannalta päästömallinuksessa on ratkottavana useita kysymyksiä, esimerkiksi:

1. Halutaan tietoa ajoneuvojen *laadullisesta jakaumasta* (luokka, relevantit tekniset tiedot) muodossa, joka myös takaa anonymiteetin.
2. Halutaan tietoa ajoneuvojen *määrästä* nykyistä verkostoa laajemmin.
3. Halutaan kohtuuhintainen helposti liikuteltava ja kalibroitava laitteisto
4. Uusi päästömalli, jossa em. parantavat kaupungin liikennepäästömallinnusta
5. Edge-laitteen, ajoneuvorekisterin datan ja kaupungin järjestelmien integraation ja arkkitehtuuriin liittyvät asiat.

Projektin lähtöasetelma ja teknologia vastaa erityisesti kohtaan 1. Ajoneuvojen laadun ja päästötason kertominen ei onnistu pelkän silmukalla tehtävän määrälaskennan avulla. Tämä on teknologian tuoma lisähyöty ja rekisterikilvet tunnistuvat hyvässä valaistuksessa siten, että päästömallintamiseen saadaan uutta tietoa.

Kohta 2 - ajoneuvojen laskenta, on oma ongelmansa: miten ajoneuvot lasketaan konenäöllä, vaikka kilpi ei esim. peittymisen takia tunnistu. Kaikissa ajoneuvoissa kilpeä ei edes ole edessä. Tämä tehtävä ei ole sen helpompi kuin kilpien tekstin tunnistaminen. Ajoneuvojen määrän laskenta on kokeilussa parhaissa olosuhteissa lähellä silmukkamittausta, mutta olosuhteet vaikuttavat tähänkin voimakkaasti. Ero on muuttuva kuten kilven lukemisen onnistuminenkin. Virhe on riippuvainen valaistuksesta ja ajoneuvojen nopeudesta, kaistanvaihtoista yms. Kalibrointi mittauspaukan olosuhteisiin vaikuttaa yhtä lailla kuin 1 kohtaan.

Silmukan aikaleimat ja kameratunnistuksen aikaleimat eroavat ja siksi ajoneuvoyksilöllinen määrän vertaaminen lyhyellä aikavälillä ei ole suoraviivaista. Ajoneuvojen laskennassa ei mitattu false positive ratea (havaitaan ajoneuvoja, joita ei todellisuudessa ole). Mekanismit tällaisen syntymiseen olisi esim. saman auton rekisterikilven väärä tunnistus josta syntyy "haamuauto". Tätä oli estetty tunnistuksen yhteydessä teknisesti. Määrä lienee pieni, mutta sen mittaaminen olisi tulevaisuudessa paikallaan.

Kohta 3 on vaativa lopputulema. Laitteiston kehittämisessä testaaminen oikeassa ympäristössä oli tärkeää jo alusta lähtien. Otollisten olosuhteiden varmistaminen, laitteiston ja mittausasetelma (esim. ROI) on keskeinen osa työmäärää. Hyvässä valaistuksessa tunnistaminen onnistuu hyvin - laitteiston kohdistaminen ja tunnistusalueiden laadinta vaatii paneutumista. Valittu kilpien lukualgoritmi on ehkä alansa parhaita, mutta tulokset heittelevät silti johtuen useasta tekstin kontrastiin ja geometriaan liittyvästä seikasta. Tällaisia ovat muun muassa likainen kilpi, lumisade, vesisade, pimeys, nopea liike ja perspektiivimuunnokset kuvattaessa autoa lähes sivusta. Monissa ajoneuvoissa on myös

poikkeuksia rekisterikilven sijoittelussa ja ulkonäössä. Kalibrointiprosessin ja laitteiston jatkokehitys vaatii paljon työtä.

Riskien hallinnan näkökulmasta on hyvä muistaa, että rekisterikilpien tunnistaminen useammassa paikassa mahdollistaa ajoneuvojen melko tarkan seurannan. Siksi nykyinen toteutus ei välitä rekisteritunnuksia tai muuta identifioivaa tietoa ulos laitteelta. Ohjelmiston muuttaminen tähän tarpeeseen olisi kuitenkin helppoa.

Opit kokeilemisesta ylipäätään

- Projekti rajoitettiin onnistuneesti konenäön prototypointiin. Teknisesti valitut algoritmit ja laitteisto tuntuu alustana olevan käyttökelpoinen, ja skaalattavissa. Kenttätestaukseen saatiin rakennettua realistinen asetelma Helsinginkadulle. Testaus saatiin tehtyä paikassa, jossa on perinteiden laskentasilmuksia joten saatiin verrokkiarvo ajoneuvojen määrälle.
- Tietosuoja-asioihin paneuduttiin varhaisessa vaiheessa, ja vaikka ne ovat työläitä, ne saatiin asianmukaisesti tehtyä.
- Projektissa palaverien ulkopuolinen kommunikaatio projektin edunsaajien ja kehittäjien tiimikanavalla jäi vähäiseksi. Kokemuksemme mukaantämän tyyppisessä kokeiluissa tarvitaan muutama edunsaaja, jotka keskittyvät kokeilun tukemiseen proaktiivisesti ja iteratiivisesti valitulla ad-hoc-kanavalla. Tarkoitusta varten oli perustettu Teams-kanava, jolla kehitystiimi pyrki jakamaan asioita, mutta vaste jäi vähäiseksi.
- Iteratiivinen ja osallistava kommunikaatio korostuu, jos projekti saa päästömallinnuksen keskittävää jatkoa. Edge-arkkitehtuurin ja päästömallinnuksen suunnittelun ja jatkokehittämisen välinen problematiikka vaatii tiivistä yhteistyötä kehittäjien kanssa. Ilman kommunikointia joudutaan arvailemaan paljon ja kohdistetaan resurssit tehottomasti.

Mitä Reaktor sai projektista?

- Reaktorin konenäköasiantuntemus kasvoi. Vaikka kyse oli kokeilusta, oli asetelma kuitenkin realistinen. Helsingin kaupungin tuella tehty koejärjestely mahdollisti kokeilun todellisella, realistisella materiaalilla.
- Saatua käytännön kokemuksia edge-laskentaan liittyvästä problematiikasta voidaan hyödyntää mahdollisesti muissa projekteissa
- Viranomaisten tietokantojen lupaprosesseihin tutustuminen

Opit resursoinnista

Reaktorin resursointi

Reaktor lähti mukaan erityisesti konenäköön liittyvän edge-laskennan asiantuntijoiden kiinnostuksen ja oppimisen näkökulmasta; hankkeen rahallinen budjetti ei sinänsä mitenkään olisi voinut riittää kehitystyöhön laitteistoinen.

Projektissa oli mukana kaksi konenäköön ja kehitykseen erikoistunutta koneoppimisen asiantuntijaa ja avustamassa tietosuoja-asioiden ja liikennepäästömallintamisen osalta kolmas data scientist.

Reaktorin käyttämä työaika kokonaisuudessaan oli noin **110 htp**.

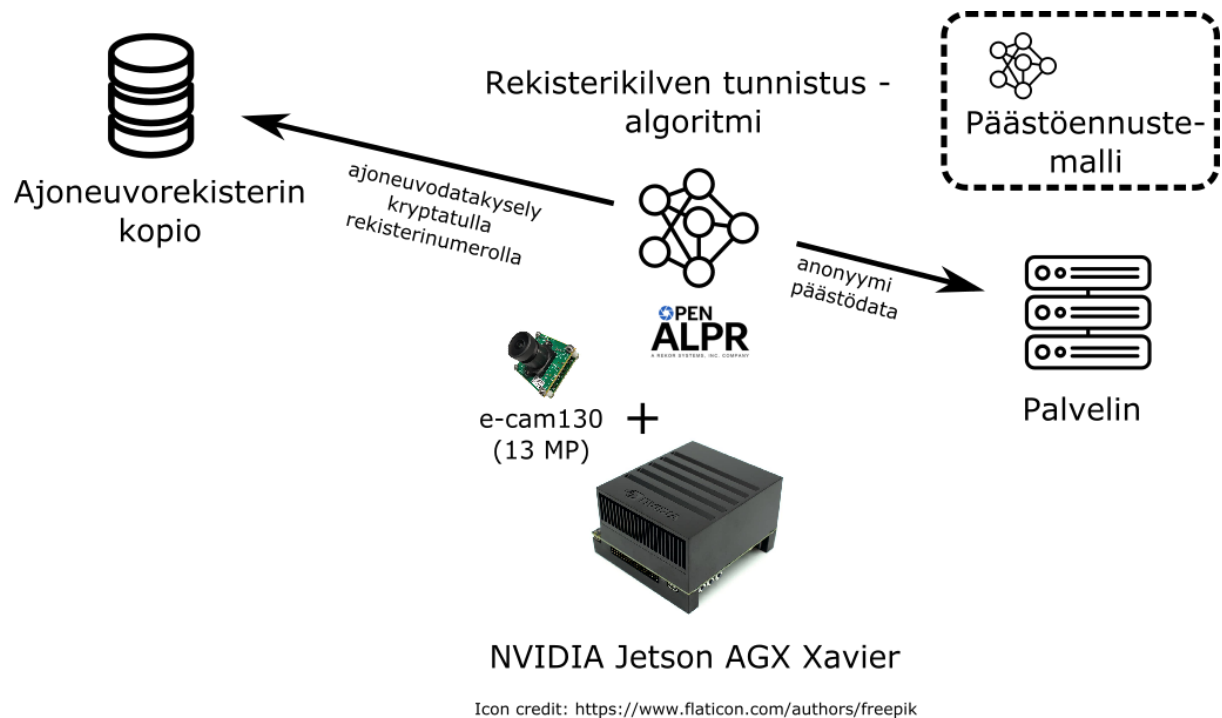
Suuri ajasta kului havainnointiohjelmiston hienosäätöön. Nyrkkisääntönä ongelmanratkaisussa 10 aikayksikköä kestävällä kehitystyöllä ratkaistaan 90% ongelmista, ja lopun 10% ratkaisuun käytetään 90 aikayksikköä. Tässä tapauksessa, omien havaintojemme mukaan, lyhyt kehitystyö ei tuonut riittävän hyvää rekisterikilpien lukualgoritmia. Projektissa kului alkuperäistä arviota enemmän aikaa myös laitteiston kalibrointiin ja sen ajureiden asentamiseen.

Helsingin kaupunki

Varsinaisen kokeiluprojektin aikana (poislukien valmistelu ja käynnistys) kaupungin puolella työpanos oli n. kaksi henkilötyöviikkoa jakautuen eri henkilöiden välille. Noin puolet työmäärästä liittyi tietosuojan vaikutustenarviointiin. Kokeilun aikana laajempi kaupungin henkilöistä koostuva seuraajaryhmä osallistui pariin yhteiseen kokeiluun suunnitelmista keskusteluun sekä tarpeiden määrittämiseen.

Kokeilun tekninen toteutus ja tulokset

Arkkitehtuuri



Laitteisto

- Kamera:
 - e-con systems 13Mpix, 350\$
 - 10° FOV linssi, 10\$
- Suoritin: NVidia AGX Xavier, 900\$
- Akku: Polaroid PS300, 358e
- Kotelointi: Muovinen laitekotelo, 100e
- Internet yhteys: USB 4G modem, 70e

Laitteiston kokonaishinta ilman hankintakuluja on noin **1250 euroa**. Kokeiluun varatussa paikassa ei ollut tarjolla sähköä, joten laitteeseen hankittiin suuri akku, jolloin kokonaishinta oli noin **1600 euroa**. Hinnat sisältävät ALV:n. Lopullisessa ratkaisussa olisi voinut käyttää myös pienempiresoluutiosta kameraa, jonka hinta noin 100\$.

Ohjelmisto

Ohjelmisto on kirjoitettu Python - ohjelmointikielellä ja siitä on käytössä versio 3.6. NVidia AGX Xavierin käyttöjärjestelmä on Ubuntu 18.04LTS, joka on asennettu JetPack 4.3 asennusmedialta.

Tietokanta: PostGreSQL

PostgreSQL on avoimena lähdekoodina jaettava tietokannan hallintajärjestelmä, joka on lisensoitu BSD-lisenssin kaltaisella lisenssillä. PostgreSQL perustuu relaatiomallille. PostgreSQL ei ole yksittäisen yrityksen tai henkilön kontrolloima, vaan perustuu kansainväliseen ohjelmoijien ja yritysten muodostaman yhteisön tekemään kehitystyöhön. PostgreSQL on tullut tunnetuksi hieman hitaampana mutta ominaisuuksiltaan parempana ja luotettavampana tietokantana kuin esimerkiksi MySQL. (Lähde: wikipedia)

Ohjelmistokirjastot

OpenALPR

Projektissa käytettiin rekisterikilpien lukemiseen OpenALPR-nimistä avoimeen lähdekoodiin perustuvaa kirjastoa. Kirjasto on kirjoitettu C++ ohjelmointikielellä ja sisäisesti se käyttää OpenCV-kirjastoa kuvankäsittelyyn liittyvien operaatioiden tekemiseen sekä Tesseract-kirjastoa varsinaiseen merkkien tunnistamiseen. Tunnistus perustuu *local binary pattern* -tekniikkaan, jolla tunnistukset pystytään tekemään nopeasti ja kohtuullisen luotettavasti erilaisissakin valaistusolosuhteissa. OpenALPR:ssä on tuki usean eri maan rekisterikilville ja se on tarvittaessa koulutettavissa tunnistamaan myös uuden tyyppisiä rekisterikilpiä. Mainittakoon, että suomalaisten rekisterikilpien osalta tunnistamista haittaa se, että esim. merkit I ja 1 muistuttavat kilpien kirjasinlajissa toisiaan. Sama tilanne on myös merkkien O ja 0 sekä S ja 5 kohdalla.

OpenALPR:n tunnistusalgoritmin parametreja voidaan muuttaa `openalpr.conf` - nimisen konfigurointitiedoston kautta. Siellä voidaan määritellä mm. seuraavanlaisia asioita:

- Maksimi kuvan koko
- Rekisterikilven maksimi tunnistuskulma
- Tunnistustarkkuuden kynnyсарvo
- Laitepohjaisen kiihdytyksen asetus
- Kirjasinlajin minimikoko

Tunnistuksen tuloksena saadaan lista eri merkkiyhdistelmistä järjestettynä niiden todennäköisyyksillä.

OpenALPR on lisensoitu GNU Affero General Public License v3.0 mukaisesti ja on siten ilmaiseksi käytettävissä avoimen lähdekoodin projekteihin, kunhan sen lähdekoodi on julkista ja käyttää samaa lisenssiä.

YOLOv5

YOLOv5 on hahmontunnistukseen käytetty nopea ja tarkka konvoluutioneuroverkko. Sitä on hyödynnetty projektissa tunnistamaan ajoneuvojen sijainnit sekä ulottuvuudet. Tässä yhteydessä YOLO:sta käytettiin sen COCO (Common Objects in Context) datalla esiopetettua versiota, joka sellaisenaan on kykenevä tunnistamaan 80 erilaista objektiota, mukaan lukien moottoripyörät, autot, bussit ja kuorma-autot.

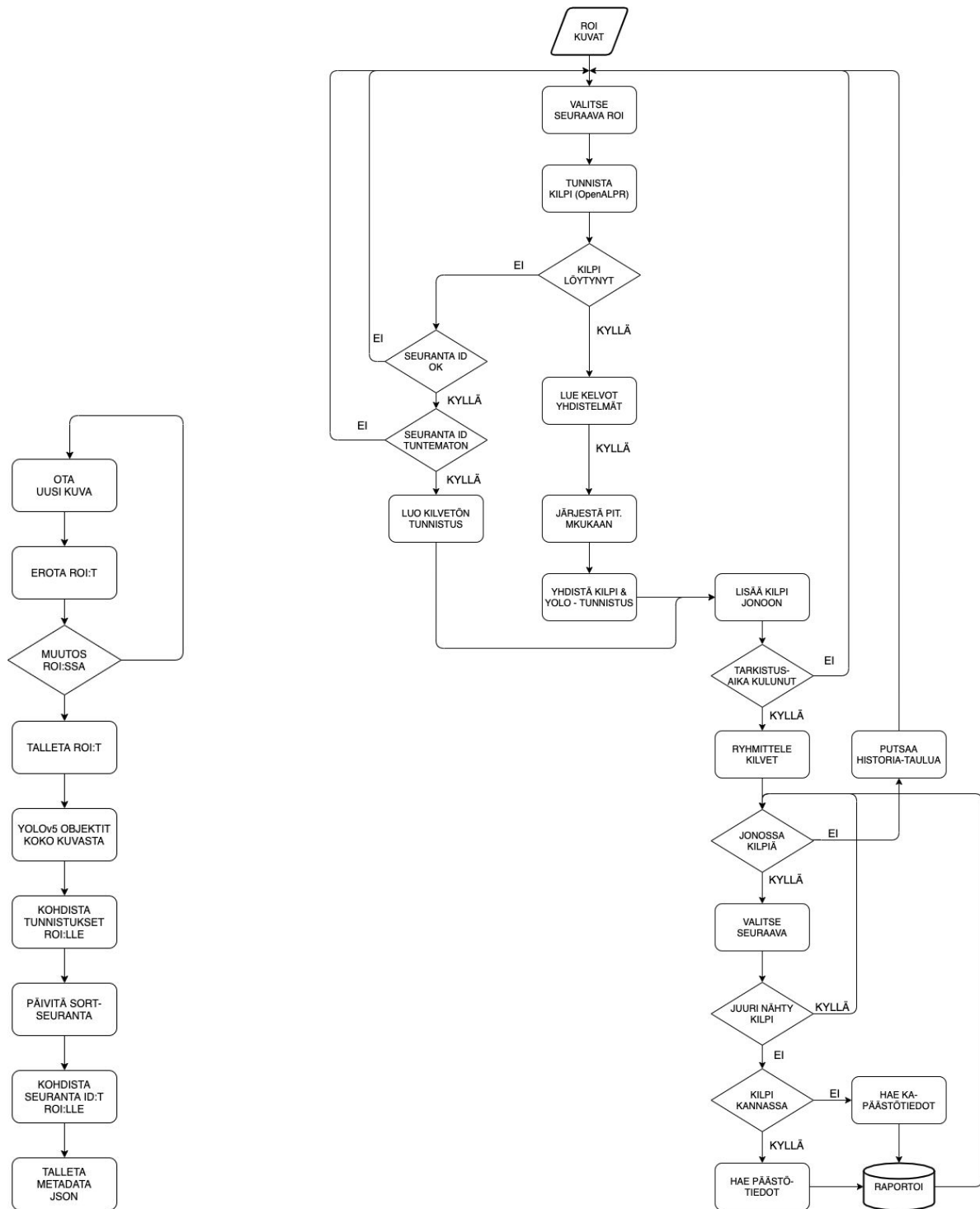
Myös YOLO-malli on tarvittaessa uudelleen koulutettavissa tunnistamaan uudenlaisia kohteita tai vahvistamaan jo tunnettujen objektien tunnistusta. Konvoluutionaalisen rakenteensa ansiosta malli rinnakkaistuu hyvin ja on ajettavissa grafiikkaprosessorilla (GPU) erittäin nopeasti.

YOLOv5 on lisensoitu GNU General Public License v3.0 mukaisesti ja on siten ilmaiseksi käytettävissä avoimen lähdekoodin projekteihin, kunhan sen lähdekoodi on julkista ja käyttää samaa lisenssiä.

SORT

SORT (Simple Online and Realtime Tracking) 2D-seuranta-algoritmia on käytetty projektissa YOLO-algoritmilta saatavien ajoneuvotunnistusten seuraamiseen, jotta kyetään identifioimaan ajoneuvoja peräkkäisten kuvien välillä. Algoritmi toimii suorakaiteiden muotoisten tunnistusten avulla ja hyödyntää Kalman-suodinta ennustamaan ajoneuvojen sijainteja peräkkäisissä kuvissa. Algoritmi valittiin sen yksinkertaisuuden ja nopeuden perusteella.

SORT on YOLOv5:n tapaan lisensoitu GNU General Public License v3.0 mukaisesti ja on siten ilmaiseksi käytettävissä avoimen lähdekoodin projekteihin, kunhan sen lähdekoodi on julkista ja käyttää samaa lisenssiä.



Processor ja Reader - moduulien vuokaaviot. Kaavioita on yksinkertaistettu ja ne eivät näytä säikeistystä.

Opit ja huomiot kokeilun laitteistoteknisestä toteutuksesta

- Sähkö tarvitaan paikan päälle, akkuteknologia tuottaa paljon epävarmuutta. Laskentayksikön sähkön saantia jouduttiin rajoittamaan sammuttamalla osa prosessoreista.
- Suuriresoluutioisen kuvan (13Mpix) käsittely on liian hidasta tälle laitteistolle. Sen sijaan, resoluutio tiputettiin 1080p tasolle, ja käytettiin tarkemmin rajaavaa optiikkaa (90 asteen avauskulmasta 10 asteen avauskulmaan)
- Kamera ei tunnista kilpiä pimeällä, koska valotusaika liian pitkä. Autoista sen sijaan saa kyllä tiedon, ovatko ne henkilöautoja, rekkoja vai busseja. On mahdollista käyttää myös valovoimaisempia kameroita, jolloin tunnistusta voidaan ainakin laajentaa hämärämpään ympäristöön.

Opit ja huomiot päästömallinnuksesta

Kokeilun varsinainen tavoite on päästöjen tarkempi arviointi - rekisterikilpien ja ajoneuvojen tunnistaminen on tarkoitettu vain teknisten tietojen hakemiseen, ja konenäkö voi tuki käyttää liikennevirran analysointiin muissakin tarkoituksissa.

Edge-laskennan, anonymiteetin ja päästömallinuksen suhde

Tekninen arkkitehtuuri ja anonymisointi vaikuttaa myös päästömallinnukseen ja sen jatkokehittämiseen. Yksinkertaisin keino varmentaa kaikki mahdollisuudet tiedon jatkokäyttöön olisi lähettää kustakin ajoneuvosta kaikki tekniset tiedot, pl. yksilöivät henkilötiedon luonteiset tiedot. Tällöin kuitenkin löytyy helposti paljon ajoneuvoja jotka ovat yksilöitävissä. Osa päästötietojen mallinnuksesta tehdään ratkaisussa edge-laitteessa ja osa myöhemmin. Edge -ratkaisun ja yksilöivien tietojen hävittämisen hyvä puoli on, että anonymiteetti on helpompi säilyttää. Ideaalitapauksessa voitaisiin lähettää pelkkiä kokonaispäästöarvoja ja tämä myös vähentää viestittävää tietomäärää. Huono puoli on, että jos laskentatapaa muutetaan, esimerkiksi kun tehdään parempi malli: aiemmin kerättyä dataa uudelleenlaskentaan ei ole. Samoin mallin virheiden analysointi jälkikäteen on hankalampaa.

Yksinkertainen keino taata anonymiteetti on tarjota riittävän karkeita tilastoja reunajakaumista, toisin sanoen erikseen vaikkapa jakauma ajoneuvojen päästöluokista riittävän pitkältä aikaväliltä, erikseen käyttövoimista, erikseen ajoneuvoluokista ja niin edelleen. Ideaalitapauksessa voitaisiin lähettää pelkkä kokonaispäästöarvo!

Huomioita

- Jo avoimesta datasta huomaa, että CO₂ -arvoja puuttuu vanhemmille henkilöautoille ja raskaille ajoneuvoille (lähes) kokonaan. Kun Traficomien koko dataa ei ole saatavilla, ei ole tiedossa paljonko
 - Jatkokehityksessä on mahdollista myös korvata puuttuva tieto arviolla kullekin ajoneuvolle (tehdä ns. imputointi)
- Liikennetilanne ja ajoneuvojen kuormaus vaikuttanee todellisiin päästöihin. Onko vaikutus huomattava?

- Liikennevirran nopeutta ja pysähdysten kestoa voi ajatella estimoitavan kuvasta, kuorman mallintaminen vaatii ulkoista tietoa. Nopeuden (ja kuorman) mallintaminen jätetään tässä nyt huomiotta ja vaatii omaa työtään

Esimerkki päästömallinnukseen tuotettavasta datasta

Keskusteluissa kävi ilmi että tässä vaiheessa päästömallinnuksessa keskeiset tiedot olisivat

- Ajoneuvoluokka (ja osin ryhmä)
- Käyttövoima
- Euro-päästöluokka (erityisesti onko Euro 6 vai vanhempi)
- Tieto onko ajoneuvo Helsingin kaupungin kriteerien mukaan vähäpäästöinen vai ei

Ajoneuvojen teknisissä tiedoissa on saatavilla myös detaljitietoja NO_x, CO_x ja hiukkaspäästöistä ja lisäksi koko teknisten tietojen kirjo, mm. ajoneuvon käyttöönottopäivä, moottorin teho, ajoneuvon massa, jne.

Kokeilussa käytettiin Traficomien avointa ajoneuvodataa kehitetyn järjestelmän tietokannan kokeilemisessa. Päästömallinnuksessa kantaa ei voinut käyttää koska siinä ei ole ajoneuvon rekisteritunnusta. Em. tiedoista löytyy ajoneuvoluokka, käyttövoima ja ajoneuvon ikä. Euro-luokkaa voi arvioida ajoneuvon iän perusteella.

Kokeilun data ja tietosuoja

Datalähteet

Videokuva liikenteestä

Laitteessa oli oma kamera, joka kuvaa liikennettä. Projektin alussa kameraa pidettiin vähän liikennöidyllä tiellä, jotta saatiin materiaalia ohjelmistokehitystä varten.

Ajoneuvotietokanta

Alkuperäinen suunnitelma oli ladata koelaitteistoon ajoneuvotietokannasta rekisteritunnus ja päästölaskentaan tarvittavat tiedot demonstraatiotarkoituksiin. Osoittautui kuitenkin, että sopivan Traficomien datatuotteen käyttöön saaminen kestäisi helposti kuusi kuukautta. Myös hinta ja hankintaan liittyvä työmäärä olisivat olleet kertaluonteiseen kokeiluun epätarkoituksenmukaisen suuria, joten keskityttiin ajoneuvotietokannan käytön teknisen ratkaisun pohdintaan yleisemmin.

Traficomilta on tuloillaan sopiva eräpäivitettävä datatuote. Tietojen välittämiseen tarvitaan myös operaattori, joka tällä hetkellä on TietoEVERY. Ajoneuvon tietojen yhdistäminen rekisteritunnukseen itse laitteessa on sinänsä suoraviivaista ja tietomäärä on varsin pieni.

Traficomien tarjoamat mahdollisuudet

- Kysely jokaista ajoneuvoa kohti (lähes) reaaliajassa.
- Datan päivitys eräajona esimerkiksi kuukausittain

Ensimmäisen vaihtoehdon kustannukset kohoavat helposti suuriksi (yksi kysely maksaa n. 7 senttiä julkiselle toimijalle) ja ajoneuvovirrat ovat mahdollisesti suorituskyvyllekin liian suuria. Jälkimmäinen vaihtoehto riittänee muutenkin hyvin käsillä olevaan tilastointitarkoitukseen. Ajoneuvokanta ei uusiudu nopeasti eivätkä tekniset tiedot ajoneuvoissa muutu niin usein että se vaikuttaisi vaativan jatkuvaa ajantasaisuutta.

Mitä Traficomien datan hankinnasta opittiin?

Traficomien ajoneuvotietokannan teknisiäkin tietoja ei rajoitetustiakaan noin vain saada tämän tyyppiseen hankkeeseen, vaikka tässä tarvittava tietomäärä sinänsä on pienehkö ja rakenne yksinkertainen. Prosessi datan saamiseksi on kuitenkin sinänsä selkeä ja toteutettavissa. Asiaan liittyy kuitenkin sen verran kustannuksia ja työtä että toteutukseen on paneuduttava erikseen. Erityisesti jos kaupunki ottaa järjestääkseen Traficomien dataa välittävän järjestelmään vaaditaan erikseen projektoitava määrä resursseja, koska se vaatii pysyvän ratkaisun sekä datan päivitykseen taustajärjestelmään, kuratointiin, transformaatioiden että välittämiseen kenttälaitteeseen ja tietosujoaan.

Päästömallin laadinta autojen teknisestä datasta on myös riippuvainen Traficomien datassa saatavilla olevista parametreista. Kun datan laatu tunnetaan paremmin, voi päästömalleista vastaavat asiantuntijat laatia vaatimukset tiedoille, joita kenttälaitteisto välittää.

Jos kokeilussamme olisi edetty Traficomien datan suhteen, realistinen malli olisi ollut se että kaupunki olisi hankkinut datan, tehnyt palvelukuvauksen ja toiminut rekisterinpitäjänä ja Reaktor toiminut käsittelijänä. Tämä ei ole ainoa vaihtoehto: mahdollisen jatkohankkeen olisi erikseen ratkaistava hoitaako ja vastaako Helsingin kaupunki vai palvelutarjoaja tietojen hankkimisesta ja operoinnista, ja yleisemmin, tuottaako kaupunki palvelun itse.

Tietosuojasta

Liikennekuva-aineisto kerättiin yleiseltä paikalta, neutraalissa tilanteessa ja vain kehitysryhmän käyttöön EU-alueella. Tässäkin tapauksessa rekisterinumerot ovat henkilötietoa, samoin videokuvasta tunnistettavissa olevat henkilöt. Kuva-aineisto oli rekisterinumeroiden tunnistamistarkkuuden selvittämiseksi käytävä osin läpi manuaalisesti. Rekisterinpitäjänä tässä hankkeessa oli Helsingin kaupunki, joten hankkeelle tehtiin Helsingin kaupungin prosessin mukainen tietosuojavaikutusten arviointi ja toimittiin kaupungin ohjeistuksen mukaan. Henkilötiedot tuhoetaan projektin päätyttyä. Selvennykseksi: tuotantokäytössä ratkaisu säilyttää kuva-aineistoa ja yksittäisiä rekisterinumeroita vain tunnistuksen tekniselle suorittamiselle välttämättömän ajan.

Projektin lopputavoitteeseen liittyviä tietosuojariskejä

Edge-laitteistoon liittyviä tietoturva- ja suojariskejä on siinä olevien henkilötietojen varastaminen, laitteiston hallintaanotto (etäkäyttö tai keino hyökätä muihin järjestelmiin) tai tietoliikenteen seuraaminen. Seuraavassa muutamia projektin alussa tunnistettuja tietosuojariskejä ja keinoja niiden pienentämiseen. Tietosuojavaikutustenarvioinnissa on käsitelty asiaa tarkemmin.

Riskit: ajoneuvorekisteri tai kuvamateriaali joutuu väärin käsiin; järjestelmää kehitetään rekisteröityjen järjestelmälliseen valvontaan

Laitteen tietoteknisen ja fyysisen suojaamisen lisäksi käytetään ja tallennetaan mahdollisimman vähän tietoa.

Kehitettävä järjestelmä ei tallenna kuvaa, eikä tunnistettuja rekisteritunnuksia, paitsi sen ajan, mikä on välttämätöntä tunnistamisen teknisen toiminnan kannalta. Paikkaan ja aikaan sidotun henkilötiedon paljastumisen riski pienenee: Mitattujen autojen rekisteritietoja tai niiden tiivisteitä ei tallenneta, joten edes laitteeseen tunkeutuva ei saa tietoa mittauspisteiden ohittaneiden henkilötietoa ainakaan pitkältä ajanjaksolta. Prototyyppi välittää ulospäin keskiarvopäästötietoa liikennevirrasta ja tilastoa ajoneuvoluokista. Ajoneuvojen tunnistetietoja ei välitetä edes pseudonymisoituina.

On kiinnitettävä huomiota myös sen estämiseen, että laite kaapattaisiin niin, että siinä olevia kameroita / tunnistettuja rekisterinumeroita seurataan etänä.

Ajoneuvorekisteristä tarvitaan vain osa: päästötietoja ja rekisterinumero yhdistämistä varten. Päästölaskentaan liittyvien välttämättömien teknisten tietojen lisäksi ei oteta mukaan muita tietoja. Erityisesti: ei rekisteritunnuksen lisäksi muita henkilötietoja tai ajoneuvon suoranaisia tunnistetietoja kuten valmistenumero.

Riski: rekisteritietokannan pseudonymisointi puretaan

Suunnitellussa päästölaskentaprototyyppissä käytettäisiin paikallista kopiota ajoneuvorekisteristä. Laitteisto sisältää kaikkien ajoneuvojen rekisterinumeron tiivisteiden ja päästötiedon. Tiivisteavaimien purkamisen uhkaa voidaan rajoittaa

- Tekemällä tiivisteavaimien purkaminen vaikeaksi: rekisteritieto tallennetaan suolattuina tiivisteinä ja pääsy avaimiin pyritään estämään teknisesti
- Tallentamalla ajoneuvon rekisterinumeron tiivisteiden lisäksi vain päästölaskentaan tarvittavat tekniset tiedot
- Laitteisto suojataan fyysisesti varkauden varalta

Tarvittaessa ajoneuvoon liittyviä tietueita voidaan käsitellä niin, etteivät ne itsessään muodosta merkittävää määrää identifioivia luonnollisia avaimia.

Riski: Tuloksen anonymisointi puretaan

Kehitysvaiheessa kiinnitetään huomiota siihen, että tuotetaan riittävän aggregoitua tilastotietoa, jotta yksittäisiä ajoneuvoja ei voisi tunnistaa, tai käytetään jopa kehittyneempiä anonymisointimenetelmiä, esim. differentiaaliyksityisyys.

Kvantitatiiviset tulokset

Tunnistustarkkuuden analysoinnista yhteenvetoa tarkemmin erillisessä liitteessä: "REMPPÄ pilottikonenäköjärjestelmän kenttätesti, tuloksia"

Jatkopäätökset ja -ideat

Jos hanketta jatketaan ajoneuvotietojen päivittäminen sekä palvelimelle että laitteeseen on ratkaistava myöhemmin. Lisäksi Traficomien teknisten tietojen kuranttiuden varmistaminen (ainakin avoimessa datassa on erilaisia puutteita ja virheitä) ja se on käsiteltävä sopivaksi päästömallintamiseen

- Puuttuvien tai selvästi virheellisten tietojen imputointi on tehtävä: säännöt / tilastolliset keinot
- Traficomien erilaisten ajoneuvoluokitus- ja päästömallinnuksen luokitus- ja luokitus, esimerkiksi ajoneuvoluokka tai käyttövoima, mäppäys. Missä määrin asioita tehdään laitteessa ja missä määrin taustajärjestelmässä?
- Laite kyselee dataa tarvittaessa kaupungin (tai palvelutarjoajan IP-yhteyden yli lähettämällä pyynnön rekisterinumeron perusteella.
 - Ajoneuvojen identiteetti paljastuu potentiaalisesti palvelimelle (tai tietoliikenneyhteyden kautta). Yksilöiden liikkeen seurannan esto täytyy tehdä palvelinpäähän.
- Jos lähdetään siitä että laitteessa on koko ajoneuvotietokanta: Kanta on määrävälein päivitettävä joka laitteeseen ja laitteen "korkkaaminen" tai fyysinen varkaus on erityisesti huomioitava riskinä.
- Ohjelmisto ei vielä tunnista tapauksia, joissa sama ajoneuvo on siirtynyt kaistalta toiselle. Lisäksi on mahdollista, että ensin havaitaan ajoneuvo ilman rekisterikilpitunnistusta ja vasta myöhemmin saadaan kilpi tunnistettua. Tällaisessa tapauksessa voi syntyä tupla-tunnistus.

Päästömallinnus on suunniteltava kun Traficomien dataa on saatu. Anonymiteetin säilyttävän edge-laskennan vuoksi mallinnusta täytyy suunnitella tiiviisti yhdessä

Laitteen määrittelyssä ei ole otettu kantaa laitteen tuotantokelpoiseen koteloointiin.

Elektroniikan kanssa toimiessa pitää ottaa huomioon toimintalämpötila-rajat, kosteuden kulkeutuminen, ilmanvaihto, linssien suojaus pölyltä, kotelon puhdistus, yms.

Toteutuksia jotka pitää vielä tehdä, jotta voitaisiin päästä tuotantotasolle:

- Pilvipalvelu useamman laitteen hallintaan ja datan vastaanottoon.
- Rekisterikilpi-data pitää pystyä toimittamaan laitteelle, jo valmiiksi "hashattyna", eli ei-selkokielisenä tekstinä
- Laitteen tietoturvan saavuttamiseksi, laitteesta tulee poistaa kaikki ulkoapäin yhteyden tarjoavat palvelut (esim SSH). Hallinta toteutetaan yllämainitun pilvipalvelun kautta siten, että laite itse hakee tiedon onko se ajossa vai ei.
- Laite voi hyvin käynnistää ohjelman käynnistyessään, mutta tätä ei toteutettu nykyisellään.
- Kalibroinnissa tarvittavat ROI-alueet pitäisi pystyä piirtämään jollain paremmalla työkalulla kuin nykyisin. Tällä hetkellä ne piirretään millä tahansa piirto-ohjelmalla, ja siirretään käsin laitteelle oikeaan hakemistoon.

Liitteet

- Tunnistustarkkuusraportti 26.11.2020 (erillinen PDF)