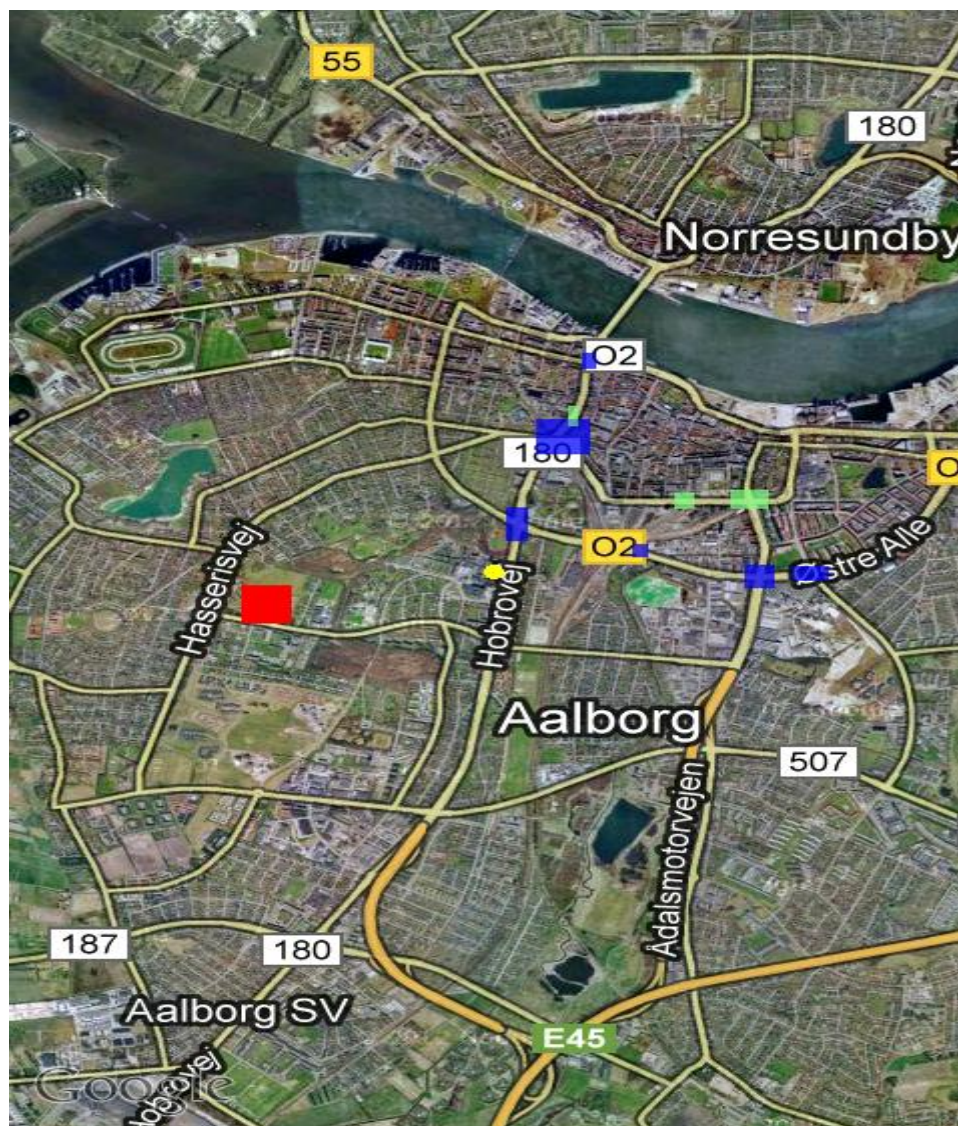


Trafik overvågning ved hjælp af Android smartphone og GPS data



Title:

Trafik overvågning ved hjælp af Android
Smartphone og GPS data

Projektperiode

1 februar – 7 juni 2012

Semester:

Master speciale

Emne:

Softwarekonstruktion speciale i database.

Projektdeltager:

Jesper Andersen

Vejleder:

Kristian Torp

Oplagstal: 4

Sidetal: 31

Typeenheder i alt: 58644

Rapporten er ikke fortrolig.

SYNOPSIS:

Igennem dette projekt har det være ønsket at undersøge om der kun udvikles et system, der kan indsamle data om hvordan trafikken og trafikbelastningen er igennem et vejkryds eller en vej strækning, ved at bruge en smartphone. I projektet er der brugt en Android smartphone.

Igennem projektet er der udviklet en algoritme der er implementeret i den klient software der er installeret på Android smartphone. Algoritmen har til formål at styre dataindsamlingen, det vil sige at den skal håndtere de regler der er opsat for hvordan data indsamlingen skal forgå. Hvor den regel der er stærkest, er reglen om at bilisten skal sikres en så høj grad af anonymitet som muligt. At denne reglen er den der vægter mest er for at sikre en bilisterne der vælger at give data til systemet skal kunne følge sig tryk ved at bruge det. Måden det gøres på er at bilisten opsætter nogle privatzoner hvor i der ikke må indsamles data. Derudover er der krav om at klient applikationen skal hvad et lavt batteri forbrug, samt at den ikke må belaste telefonnetværket mere end højst nødvendig når de indsamlede data sendes.

Til at efterprøve hvordan klienten indsamler data og om den overholder de opsatte regler, er der fortaget et feltforsøg. Hvor der er opsamlede data og bilisten kommer igennem flere privatzoner. Det er eftervist at klienten indsamler brugbare data og overholder reglen om privatzoner, samt de andre regler. De indsamlede data er b.la. i KML format så det er muligt at se den rute bilisten har kørt i Google earth.

Forord

Denne rapport og Android klient applikation er resultatet af mit master speciale. Hvor det det skulle undersøges om det var muligt at indsamle brugbare GPS data ved hjælp af en Android smartphone. Der er igennem rapporten udviklet en algoritme til at håndter dataindsamling, denne algoritme er grundbestanddelen i Android klienten.

I rapportens feltforsøg del er der brugt jeg, dette er gjort da jeg selv har gennemført feltforsøget. Det er gjort ved at jeg har haft klienten installeret og kørende på min Android smartphone og så indsamlede data når jeg er kørt til og fra mit arbejde.

Kildehenvisninger fremtræder som [kilde] og kan findes i litteraturoversigten bagerst i rapporten.

Jeg vil i forbindelse med rapporten gerne takke min vejleder Kristian Torp for hjælp og mange gode diskussioner, igennem projekt gennemførelsen

Aalborg den 7. juni 2012.

Jesper Andersen

Indholdsfortegnelse

Introduktion.....	1
Hypotese.....	1
Systemdefinition.....	2
Dataindsamlingszoner	2
Privatzoner	3
Prioritet af Dataindsamlingszonerne.....	3
Datarække	3
Systemets definitioner	3
Analyse	5
Problemområde.....	6
usecase i problemområdet.....	7
Anvendelsesområde	8
Design	9
Minimum antal rækker fra en bilist.....	9
Kan det bruges i andre byer ?.....	11
Beregning af længde for DIZ.....	12
Maksimum og minimum størrelse på en Dataindsamlingszone	13
Størrelse på en privatlivszone	13
Opsætning af Zoner	15
Dataindsamlingsalgoritmen.....	16
Generel data indsamling i en DIZ.....	16
Dataindsamling i en DIZ.....	17
Dataindsamling uden for DIZ.....	18
Flowchart Diagram	20
Pseudo kode	20
Implementering	22
Database design	23
Komponent diagram	25
Relateret arbejde.....	26
Feldtforsøg.....	26

Resultat.....	27
Videre arbejde	31
Konklusion	31
Appendiks	32
Litteraturliste	35

Introduktion

På grund af den voksende mængde af trafik på vejene, er dataanalyser af trafiksituationer blevet et interessant tema, der er igennem de sidste år lavet forsøg hvor trafikbelastningen måles på i bestemte områder[6][10]. Der er lavet forsøg med brug af satellitter til at overvåge trafikken på bestemte vej områder. Der er også foretaget forsøg hvor der er brug GPS apparater til at indsamle data om hvor en bil befinder sig [5][7][9][12]. Endvidere er der foretaget forsøg hvor der er brug en smartphone, hvor der er kørt 160 studerende rundt i San Francisco bay med 100 smartphone i 3 timer, for eksempel en Nokia N95[2][4], i stedet for specielt bygget hardware. Desuden er der foretaget forsøg med at opsætte apparater i vejsiden som detekterer om en bestemt bil kører ind i et bestemt område, dette sker ved at bilen har påmonteret en sender som kommunikerer med apparaterne i vejsiden[3].

Flere af de måder hvor der er brugt GPS enheder eller specielt bygget hardware[5][7][9][12], til at undersøge hvor stor en belastning der er af et vejnet i et bestemt område, er alle meget kostbare og tidskrævende. Udvikling og opsendelse af satellitter er kostbare og tager tid.

Igennem projektet præsenteres der en model om hvordan der kan indsamles data, hvor der gøres brug af en ganske almindelig smartphone. Samt en algoritme der håndterer dataindsamlingen samt at der tages højde for at de personer der deltager skal have en så høj grad af anonymitet som muligt, hvor dataindsamling vil forgå til en så lille omkostning for bilisten som muligt[8].

Hvor dataindsamlingen forgår ved at bilisten har en applikation installeret på brugerens mobiltelefon, som ved hjælp af satellitter får deres GPS position. Derigennem kan der opsamles data om hvor brugeren er på et givet tidspunkt, med hvilke hastighed bilisten bevæger sig med etc. På figur 1 vises det hvordan systemet virker med dataindsamlingszoner, privatzoner, bilistenes vej til og fra arbejde, samt brugernes arbejdsplads.

Igennem eksemplet sidst i rapporten, vises det hvordan der kan indsamles data ved hjælp af den udviklede mobil applikation, hvor der gøres brug af den udviklede algoritme.

Hypotese

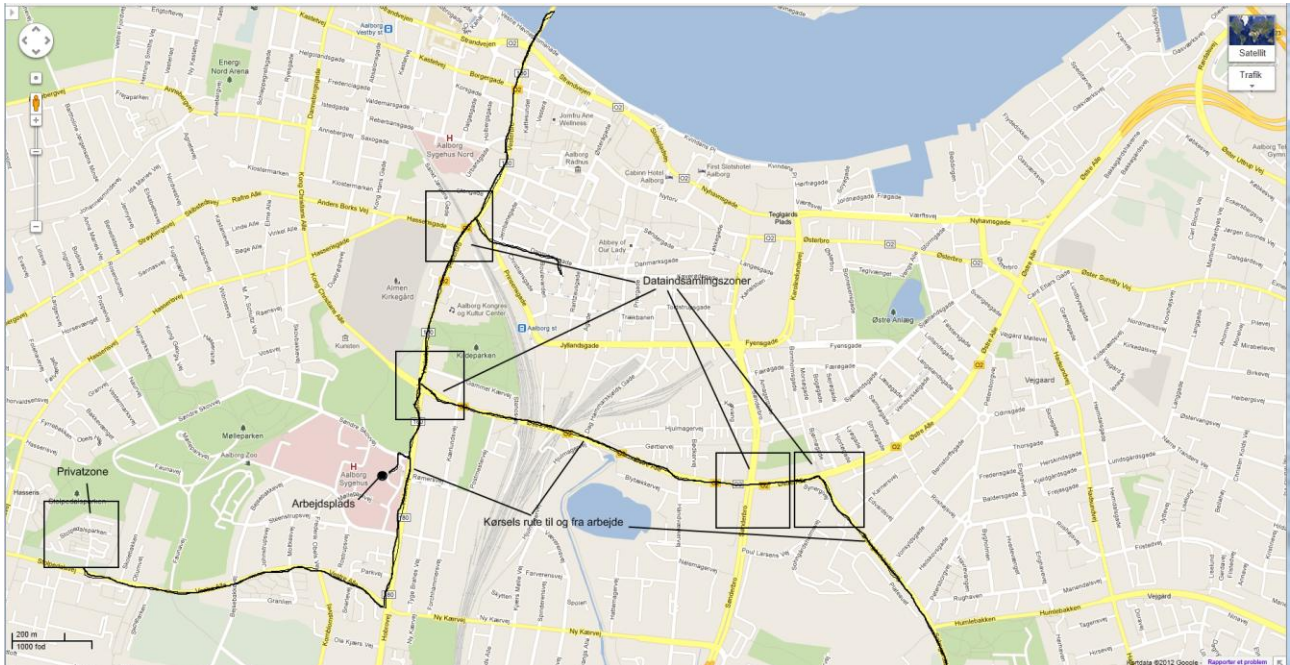
På baggrund af forsøg på at indsamle data med specielt bygget hardware, og forsøget i San Francisco bay hvor der blev brugt en Nokia N95. Har jeg en tese om at kunne indsamle data ved hjælp af en smartphone og eftervis at de indsamlede data brugbare, hvor min tese har givet følgende hypotese:

Er at kunne indsamle brugbare data, hvor disse data er indsamlet ved at bruge en almindelig smartphone, og ikke speciel udviklet hardware. Systemet skal kunne indsamle data fra prioriteret indsamlingszoner samt at det skal beskytte en bilist der er i en privatzone mod at udlevere data. For at give bilisten en så høj grad af anonymitet som muligt.

Indsamlingen skal forgå med mindst mulig gene for bilisten, det vil sige at applikationen der skal installeres på brugers smartphone, skal være en baggrunds applikation som ikke har brug for at bilisten har interaktion med applikationen når den først er startet. Desuden skal applikationen have så lille et strøm

forbrug som muligt. Indsamlingen af data skal forgå på en sådan måde, at en bilists krav for anonymitet altid skal overholdes, disse krav er ufravigelige.

For at sikre at alle systemets bilister levere indsamlede data til systemet, skal hver bilist levere en minimum mængde data om dagen, dog ikke i week-enden.



Figur1: målestoks forhold 1:10000. Figuren beskriver projektets ide hvor der skal indsamles data fra dataindsamlingszoner, samt at der ikke må indsamles data fra privatzone. På figuren er der opsat fire dataindsamlingszoner, der er også opsat en privatzone. De ruter som bilisterne køre af for at komme på arbejde er også indtegnet. Bilisternes arbejdsplads Aalborg Sygehus Syd er markeret som endepunkt for bilisterne.

Systemdefinition

I dette afsnit defineres og beskrives hvilke krav der er til systemet, såsom hvilket operativsystem den mobile applikation er udviklet til og på hvilke grundlag disse valg er truffet. Dataindsamlingszoner, Privatzone og prioritering af Dataindsamlingszonerne beskrives.

Dataindsamlingszoner

Dataindsamlingszoner (DIZ): er et område hvor trafikanalytikeren, trafikanalytikeren er en af systemets aktører, der uddybende beskrevet i anvendelsesområdet under analysen, der ønsker at der opsamles data i. En DIZ er i projektet defineret som et rektangel. Hvor et sådan rektangel er defineret ud fra breddegraderne og længdegraderne for det øverste venstre hjørne og det nederste højre hjørne. At DIZ er defineret ved en rektangel er gjort ud fra den betragtning at det er hurtigt at opsætte et rektangel ud fra to punkter. Da rektangellen er opsat ud fra breddegrader og længdegrader følger den gitteret for breddegrader og længdegrader.

Privatzoner

Privatzoner (PZ): er det område hvor bilisten har opsat regler for hvordan der ikke ønskes at sende data. En privatzone er en rektangel, der er defineret ud fra breddegraderne og længdegraderne for det øverste venstre hjørne og det nederste højre hjørne. Privatzonerne har samme problem som dataindsamlingszonerne med at de følger gitteret for breddegraderne og længdegraderne og ikke vejnette, også her vil løsningen være at definere dem som et polygon. Disse zoner kan være opsat både for at beskytte hvor man bor, men der kan også være andre steder hvor man ikke ønsker at sende data fra.

Prioritet af Dataindsamlingszonerne

Dataindsamlingszonerne er op delt i følgende tre prioriteter, et, to og tre. Hvor et er den højeste prioritet og tre den laveste. Prioritet et og to bliver bestemt af trafikanalytikeren. Når der opsamles data uden for en defineret zone, så bliver disse data opsamlet i en prioritets tre zone, det vil sige at en prioritets tre zone er en zone som ikke er opsat af trafikanalytikeren. Prioriteringen af DIZ bruges, hvis en bilist har sat en begrænsning på hvor mange rækker data der må leveres om dagen. Så hvis en bilist der har givet tilsagn om at levere en begrænset mængde data til systemet pr. døgn, først køre igennem en prioritets to zone, hvor der opsamles den mængde data som bilisten vil levere pr. døgn og derefter kommer ind i en prioritets et zone. Så skal der opsamles data i prioritets et zonen svarende til den mængde som er sat som den øver grænse for hvor meget data der må leveres pr. døgn, og der skal slettes tilsvarende mængde data fra prioritets to zonen.

Datarække

En datarække er de data der indsamles i hver gang der indsamles data, hvilket vil sige GPS data, hastighed og tidspunkt. Det vil sige at hver datarække består af følgende data:

- Breddegrad
- Længdegrad
- Hastighed
- Tidspunkt

De data der repræsenteres ved en dataindsamling bliver hentet igennem Googles egent API set.

Systemets definitioner

Til brug for dataindsamling skal der bruges en smartphone med GPS modul, i dette projekt er der valgt at bruge en Andorid baseret smartphone. Grund til dette er at Android er et åbent system, hvilket gør det nemt at afprøve kode, da det ikke skal igennem nogen kontrol før det kan ligges på smartphone. Desuden er udviklingssproget på Android Java, med Googles egne tilføjelser. Det vil sige at man skal ikke til at bruge nogle proprietær værktøjer eller et proprietær udviklingssprog.

Ved at designe de software komponenter der håndter dataindsamling samt regler for hvordan der skal indsamles data, samt verificere om en bilist er i en privatzone så generisk som muligt med et minimums brug af platforms afhængige systemkald. Vil det gøre en ønskede portering til andre software platforme forholdsvis simpel.

Den software der ligger på serveren til at fortage kommunikation mellem databasen og klienterne er skrevet i Java, sammen med at der bruges en relationel database, vil systemet kunne køre på både en Linux/Unix og en Windows platform.

At softwaren på server siden er udviklet i Java, gør at det vil være forholdsvis nemt at sætte systemet op på både en Linux/Unix eller Windows server.

Ved at have udviklet til Android platformen for klienten, gør det at klienten er rettet mod massemarkedet.

Ved at have udviklet systemet på en sådan måde at, det er nemt at installere på både en Linux/Unix eller Windows server. Med den udbredelse det forventes for Android platformen, ifølge Gartner vil der blive solgt 310088000 enheder i 2012 og 539218000 i 2013[13].

Da systemet skal bruges til at indsamle data fra bilister, hvor dataindsamling sker i forudbestemte zoner se figur 1. Så vil indsamling af data fra en bilist forgå ved at bilisten har en applikation installeret på bilistenes Android smartphone. Ved hjælp af GPS kan applikationen afgøre om bilisten er i en dataindsamlingszone, er der overlap mellem en dataindsamlingszone og privatzone, og om det begynder at sende data til systemet.

Applikation skal kunne beregne ud fra dens nuværende position hvor lang tid der er før den kommer hen til en dataindsamlingszone, det skal bruges til at beregne hvor lang tid der vil gå før bilisten er henne ved zonen. Det vil sige at den tid der går til bilisten er i en dataindsamlingszone kaldes sovetid, hvilket er den tid applikationen sover i mens bilisten kører mod en dataindsamlingszone. Når bilisten kommer ind i en dataindsamlingszone, så skal applikationen vågne op for at indsamle og sende data.

Dataindsamlingszone er defineret af en trafikanalytiker og konfigureret ved hjælp af længdegrad og breddegrad. Når en Dataindsamlingszone sættes så vil der samtidig sættes nogle regler for hvornår der skal indsamles data fra Dataindsamlingszonen.

Der skal være mulighed for at ændre / oprette nye zoner. Hvis der oprettes en ny zone så skal den kunne sendes til alle de bilister som er online. Bilisten skal have mulighed for at opsætte regler for hvor og hvornår der ønskes at sende data til systemet. Der kan være tale om at man skal være mere end en bestemt afstand fra en forud bestemt position.

Der skal kunne være mulighed for kunder at købe og gøre brug af de indsamlede data, til videre trafikanalyser.

Det vil sige det system, der ønskes, beskrives således:

Betingelser: At bilisterne finder Android applikationen nem at bruge, samt at de føler sig trykke ved at deres data ikke bliver misbrugt, og deres regler for privatzoner overholdes.

Anvendelsesområde: Igennem dette projekt fokuseres der på analyse af trafik igennem bestemte områder hvor der for disse områder er et ønske om at vide hvordan trafikken kører igennem et givet område, dette kan være et vejkrøds på et bestemt tidspunkt på dagen.

Teknologi: Applikation skal udvikles i Java på en Android smartphone med brug af Google Maps. Af database server skal der bruges en relationel database på en Windows 7 platform, på database serveren skal der være en Java server applikation der håndterer kommunikation med smartphonen. Der skal være mulighed for at opsætte trafikzoner på serveren ved hjælp af Google Maps.

Objektsystem: Indsamling af data fra en dataindsamlingszone, hvor der er mulighed for at opsætte regler for hvordan der ønskes at sende data så der sikres at de private regler overholdes.

Funktionalitet: På baggrund af krav og regler fra trafikanalytikeren skal dataindsamlingszonerne oprettes. Desuden skal der være mulighed for at bilisterne kan opsætte regler til at beskytte deres privatzoner, samt hvor de bor.

Filosofi: Systemet er et værktøj til indsamling af trafikdata i bestemte områder, hvor dette forgår automatisk.

Ovenstående punkter kan sammenfattes til følgende system definition:

På baggrund af trafikanalytikerens krav skal administratoren kunne opsætte de angive målingsområder med de regler som til disse. Der skal være mulighed for at bilisterne kan opsætte regler for hvornår og hvordan de ønsker at sende data til systemet. Opsætning af målingsområder og hvor bilisterne bor forgår ved hjælp af Google Maps. Når systemet er sat op, så vil det virke fuldautomatisk. Systemet udvikles til at bruge en relationeldatabase på Windows 7, samt en Android smartphone. Al programmer udvikling vil blive udført i Java.

Analyse

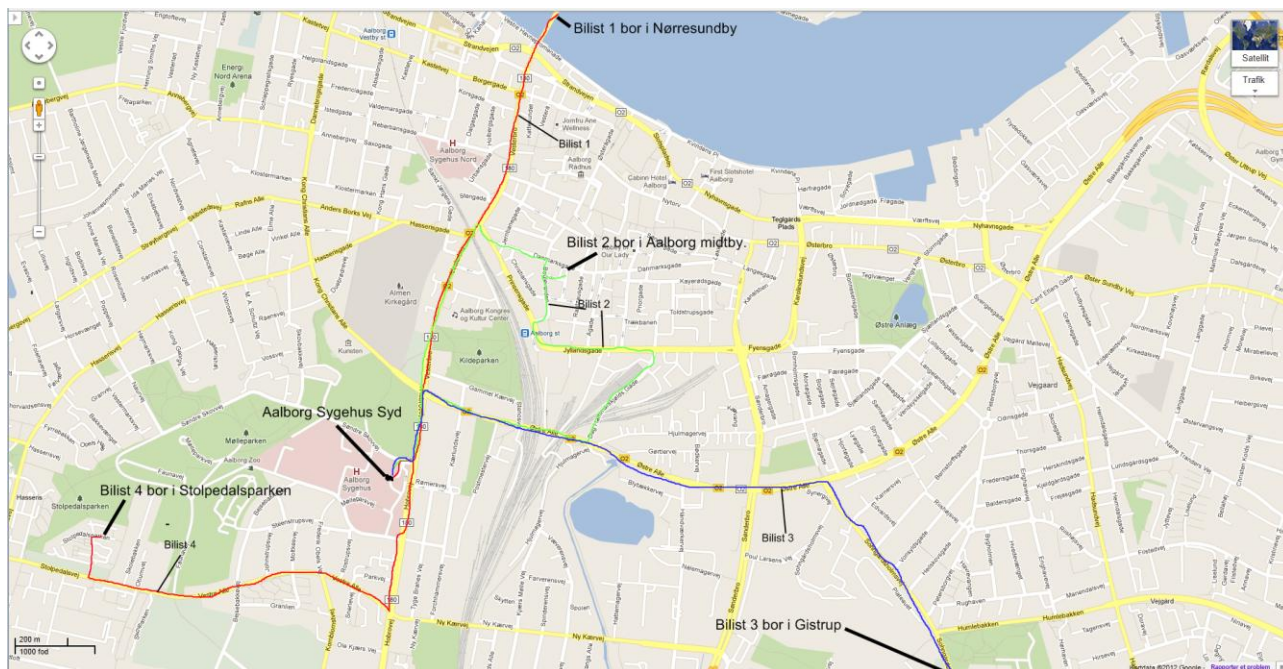
Til at belyse de ideer rapporten gør brug af, gøres der brug af et eksempel. Dette eksempel er fire bilister, som alle har Aalborg Sygehus Syd se figur 2 som deres arbejdsplads.

Aalborg Sygehus Syd er valgt som arbejdsplads da der er flere forskellige veje der leder frem til dette. En vej fra nord og en fra syd. Dette gør at der kan opsætte DIZ på kun en af de to veje. Dette giver den mulighed at der kan opsamles data fra bilister der for eksempel kommer fra nord og ikke dem der kommer syd fra på figur 2 ses de bilisters ruter til og fra arbejde.

I eksemplet er der fire bilister hvor følgende gælder.

- Bilist 1, er en person der bor i Aalborg forstaden Nørresundby og har sit arbejde på Aalborg Sygehus Syd. Denne bilist vil komme igennem flere DIZ som vist på figur 2 når der køres til og fra arbejde.
- Bilist 2, er en person der bor i Aalborg midtby og har sit arbejde på Aalborg Sygehus Syd. Denne bilist kører til og fra arbejde ad forskellige ruter hvilket gør at bilisten ikke kommer igennem de samme DIZ hver dag, dette gør at denne bilist leverer data fra forskellige DIZ som vist på figur 2.
- Bilist 3, er en person der bor i Gistrup en mindre by syd for Aalborg, og har sit arbejde på Aalborg Sygehus Syd. Denne bilist kører til og fra arbejde ad forskellige ruter hvilket gør at bilisten ikke kommer igennem de samme DIZ som vist på figur 2 hver dag, dette gør at nogle dage så leverer denne bilist data, andre dage ikke.
- Bilist 4, er en person der bor i Stolpedalsparken et parcelhus kvarter på grænsen mellem Aalborg og Aalborg bydelen Hasseris, og har sit arbejde på Aalborg sygehus syd. Denne bilist kommer ikke igennem DIZ som vist på figur 2 på vej til og fra arbejde. Bilisten har sagt ja til at lever data til systemet, så dette må foregå ved at der indsamles data når bilisten er ude af dennes private zone.

Det videre arbejde vil være opbygget omkring disse fire eksempler, algoritmen til dataindsamling vil være designet på en sådan måde at den kan håndtere de problemer der vil opstå i forbindelse med at indsamle data ud fra de ovenstående eksempler.



Figur2: målestoks forhold 1:10000, de fire bilisters ruter til og fra arbejde.

Problemområde

System skal bruges til at indsamle data fra bilister. Hvor der er opsat indsamlingszoner, disse zoner vil tit ligge i et kryds, en rundkørsel eller en motorvejs frakørsel, de kan også ligge over et stykke vej hvor der er ønske om at få et overblik af hvordan vejen er belastet på forskellige tidspunkter hen over en dag.

Indsamlingen af data forgår ved brug af en Android smartphone, som sender de indsamlede data til en database server for videre databehandling.

Der er opsat følgende krav til systemets virkemåde, endvidere er de opdelt i funktionelle- og ikke funktionelle krav:

- Bilisterne skal sikres en meget høj grad af anonymitet.
- Den mobile applikation skal indsamle data på en sådan måde at den belaster batteriet mindst muligt.
- Den mobile applikation skal være en baggrundsapplikation
- Applikationen skal have et minimums memory og CPU forbrug.
- Ved kommunikation med serveren skal applikationen belaste netværket så lidt som muligt.
- Data op loades fra klienten til serveren gang i døgnet.
- Data indsamles med en frekvens, hvor der indsamles en række i sekundet.
- En bilist skal levere minimum 20 rækker og dagen.

Bilisterne skal sikres en meget høj grad af anonymitet. Dette skal sikre at når en bilist har opsat dennes privatzoner så bliver der ikke indsamlet data fra disse zoner. Desuden gælder det at hvis en PZ har det mindste overlap med en DIZ, så vil der ikke blive indsamlet data fra denne DIZ for denne bilist.

Den mobile applikation skal indsamle data på en sådan måde at den belaster batteriet mindst muligt. Dette skal være med til at sikre at bilisten mærker mindst muligt til at applikationen er installeret, det vil sige at den skal bruge så lidt strøm som muligt.

Den mobile applikation skal være en baggrundsapplikation. Ved at gøre applikationen til en baggrundsapplikation så sikres det at bilisten ikke vil blive generet af applikationen, efter at den er blevet downloadet, installeret og konfigureret.

Applikationen skal have et minimums memory og CPU forbrug. Ved at sikre at applikationen skal have et minimums forbrug af memory og CPU, da sikres det at applikationen generer bilisten mindst muligt, samt at når der sendes til serveren da vil netværket blive belastet mindst muligt.

Ved kommunikation med serveren skal applikationen belaste netværket så lidt som muligt. Ved at sikre at netværket bliver belastet mindst muligt, sikres det at det bliver så billigt som muligt for bilisten at sende data til systemet.

Data op loades fra klienten til serveren gang i døgnet, ved at op loades data en gang i døgnet, sikres der mod datatab fra klient siden, det vil sige at hvis der sker noget på klienten så er det kun data fra indeværende døgn der vil gå tabt. Det er også med til at mindske hvor meget data der skal gemmes på smartphone. Hvilket falder godt i tråd med kravet om at applikationen skal have et minimums memory og CPU forbrug.

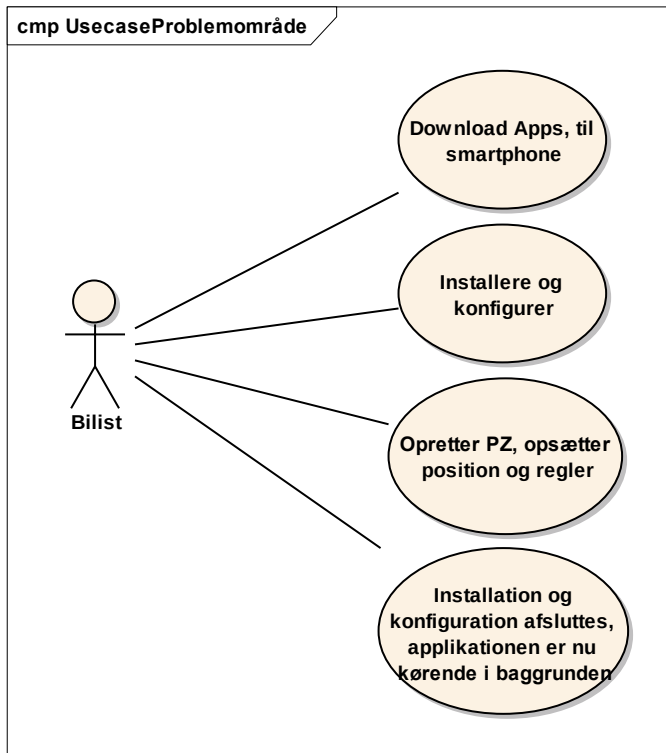
Data indsamles med en frekvens, hvor der indsamles en række i sekundet. Ved at have en fast frekvens for tit der indsamles data, så er dataindsamlingen uafhængig af bilisten hastighed.

En bilist skal levere minimum 20 rækker og dagen, hvis bilisten er kørende. Dette krav skal sikre at alle bilister som har meldt sig til system også lever data. Samt at der kan opsættes en DIZ for et kryds eller et stykke vej, som sikre at der fås tyve rækker.

Kravet om at applikationen skal tage hensyn batteriet konvergere med brug af GPS modulet, dette kan løses ved kun at hente en GPS position når der er brug for det også derefter sætte modulet i en vent position.

usecase i problemområdet

I problemområdet er der følgende usecase. Denne usecase afspejler handlingerne for at applikation downloades, installeres og konfigureret samt opsætning af PZ. På figur 3 ses de handlinger der skal til for at gennemføre applikations installations use casen.



Figur 3, use case for installation og konfiguration af klient applikationen.

1. Bilisten downloader applikationen til sin Android smartphone.
2. Bilisten installer og konfigurerer applikationen på sin Android smartphone.
3. Der oprettes PZ, hvor skal være, hvilke regler er der for deres virkemåde.
4. Skal der opsættes flere PZ, gentag fra punkt 3
5. Installation og konfiguration af applikationen afsluttes.
6. Applikationen vil nu være en baggrundsapplikation.

Anvendelsesområde

Systemet har følgende Aktører

- Administrator
- Trafikanalytiker.
- Bilist der leverer data.
- Systemejer

Administrator: opsætter og konfigurerer de forskellige DIZ der skal oprettes og udformer reglerne for disse, information om hvordan de forskellige DIZ får administratoren fra Trafikanalytikeren.

Trafikanalytiker: bestemmer hvor der skal opsættes DIZ og regler for disse. Derudover er trafikanalytikeren med til at analyser de indsamlede data.

Bilister: bilister sender data til systemet, har mulighed for at opsætter PZ, og regler for hvornår og hvordan disse PZ skal være aktive.

Systemejer: repræsenterer overvågning af bilister og deres færden. Så for at beskytte mod overvågning så skal der opsættes regler der vil give bilisten anonymitet i bestemte zoner PZ.

Design

Igennem design afsnittet vil de forskellige design kriterier blive gennemgået.

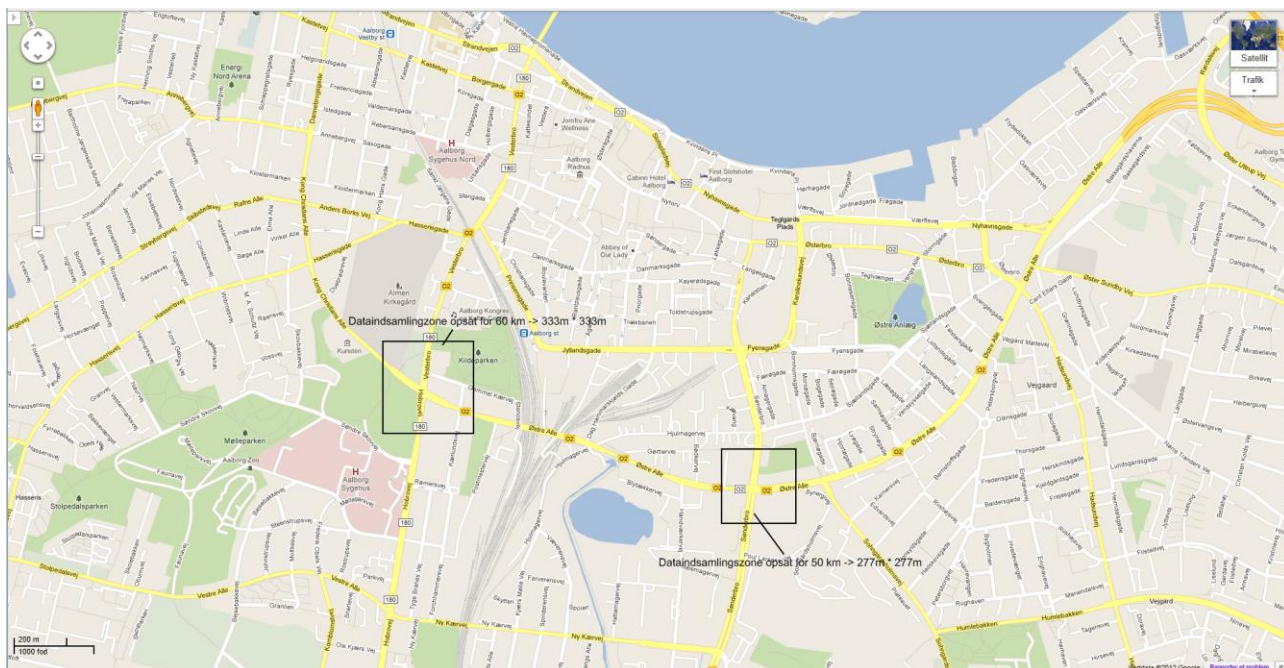
Minimum antal rækker fra en bilist

For at få et fornuftigt og brugbart datagrundlag fra blisterne. Skal hver bilist sende et minimum antal rækker dagligt. For at få et overblik over hvor mange rækker det vil være fornuftigt at sætte som det minimum antal rækker pr. bilist.

Størrelsen på en DIZ er afgørende for hvor mange rækker der minimum skal sendes, og hvor hurtigt bilisten kører. Ud fra dette vil vi se på hvor store en DIZ skal være ved en hastighed på henholdsvis 50 km og 60 km ved tyve rækker.

Ved en hastighed på 50 km vil tyve rækker kunne indsamles på en længde ad 278 meter, og ved 60 km vil tyve rækker kunne indsamles på en længde af 333 meter.

Hvis man tager midten af et vejkryds som udgangspunkt, så skal der begyndes at indsamle data i en afstand af 170 meter før og til og 170 meter efter. Se figur 4 hvor der er oprettet to DIZ, en for 50 km og en for 60 km.



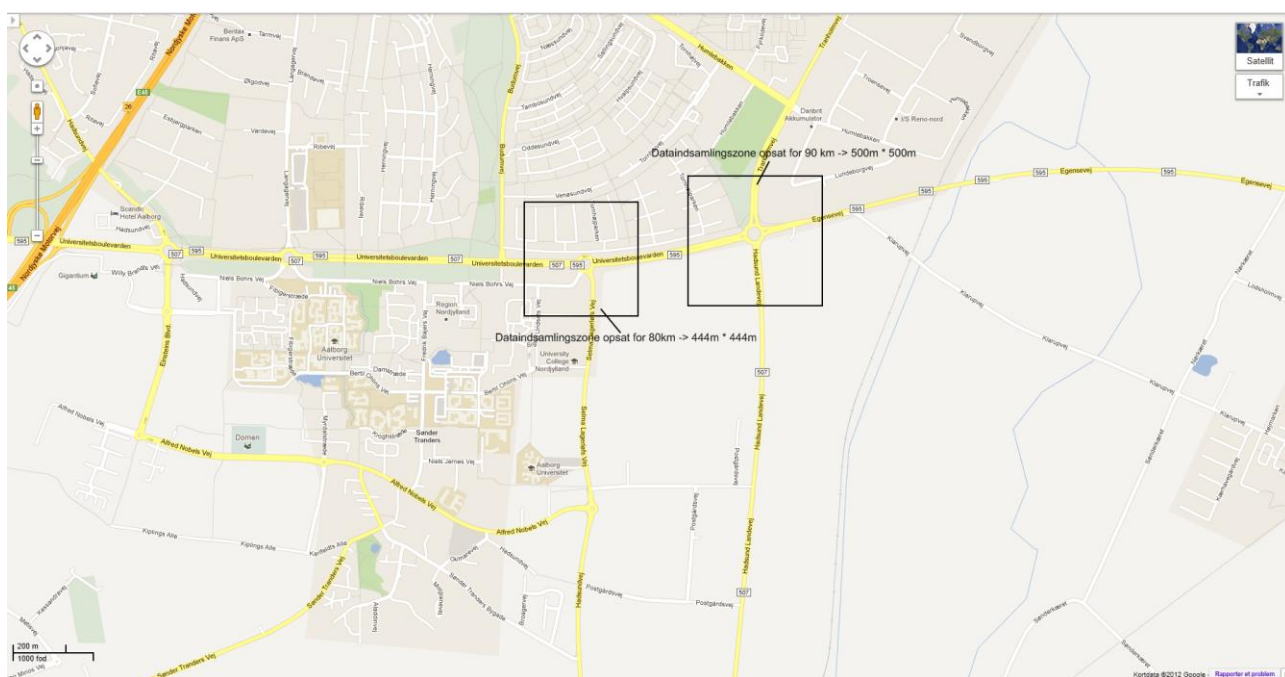
Figur4: målestoks forhold 1:10000, her ses to DIZ'er der er defineret til at kunne indsamle tyve rækker ved en hastighed på henholdsvis 50 km og 60 km.

Så i byer vil det være fornuftigt at oprette indsamlingszone som et kvadrat hvis de skal dække over kryds eller rundkørsler, med side længder på 340 meter gange 340 meter med centrum i krydses eller rundkørselens centrum.

Ved at oprette dataindsamlingszoner ud fra ovenstående størrelse, vil det ved gennemkørsel af et vejkræds med den tilladte hastighed i byer give tyve rækker.

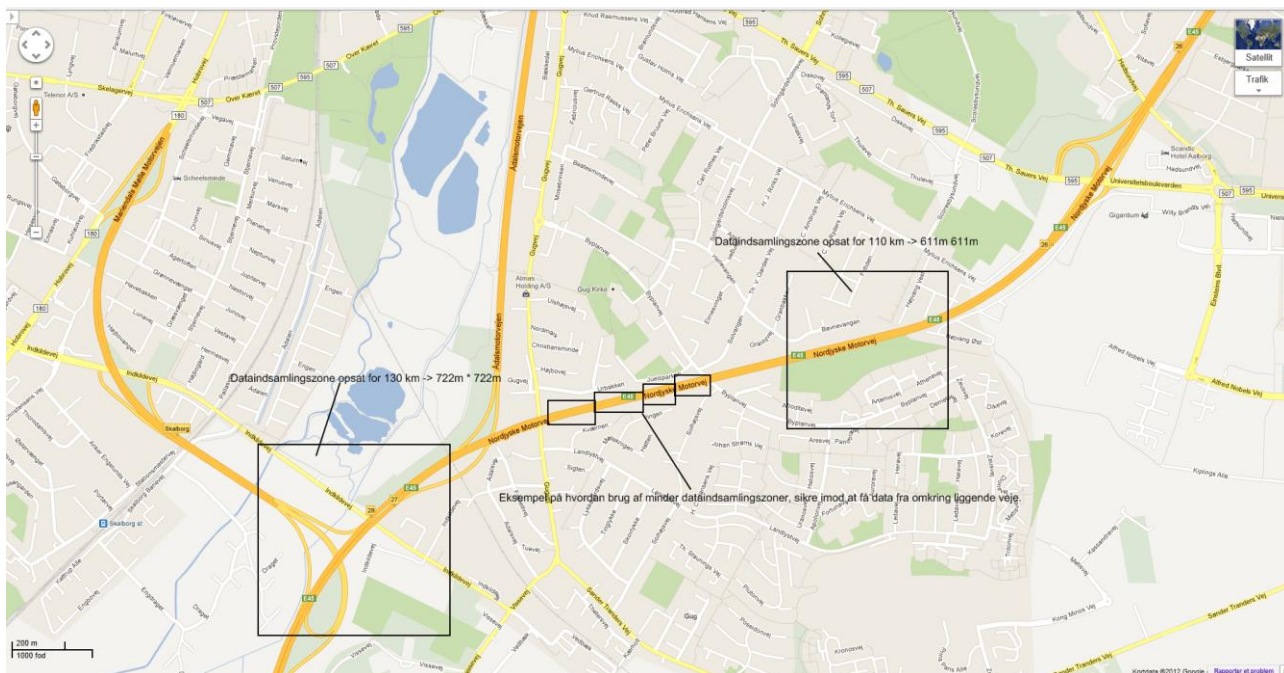
Ved oprettelse af dataindsamlingszoner i byer, skal der tages højde for afstanden mellem kryds kan være relativt kort, så for at undgå to zoner overlapper hinanden skal de være mindre eller det kan klares ved at rykke dataindsamlingszonerne i hver sin retning.

Uden for byer og på landeveje vil DIZ'erne ved 80 km og 90 km være henholdsvis 444 meter og 500 meter, se figur 5.



Figur 5: målestoks forhold 1:10000, her ses to DIZ'er der er defineret til at kunne indsamle tyve rækker ved en hastighed på henholdsvis 80 km og 90 km.

På motorveje vil DIZ'erne ved 110 km og 130 km være henholdsvis 611 meter og 722 meter, se figur 6. Når DIZ'erne når en sådan størrelse da vil de dække over flere side veje. For at undgå dette kan DIZ'erne opdeles i mindre stykker, se figur 6. Hvor DIZ'erne er opdelt i mindre stykker kan der være overlap, men det bør undgås.



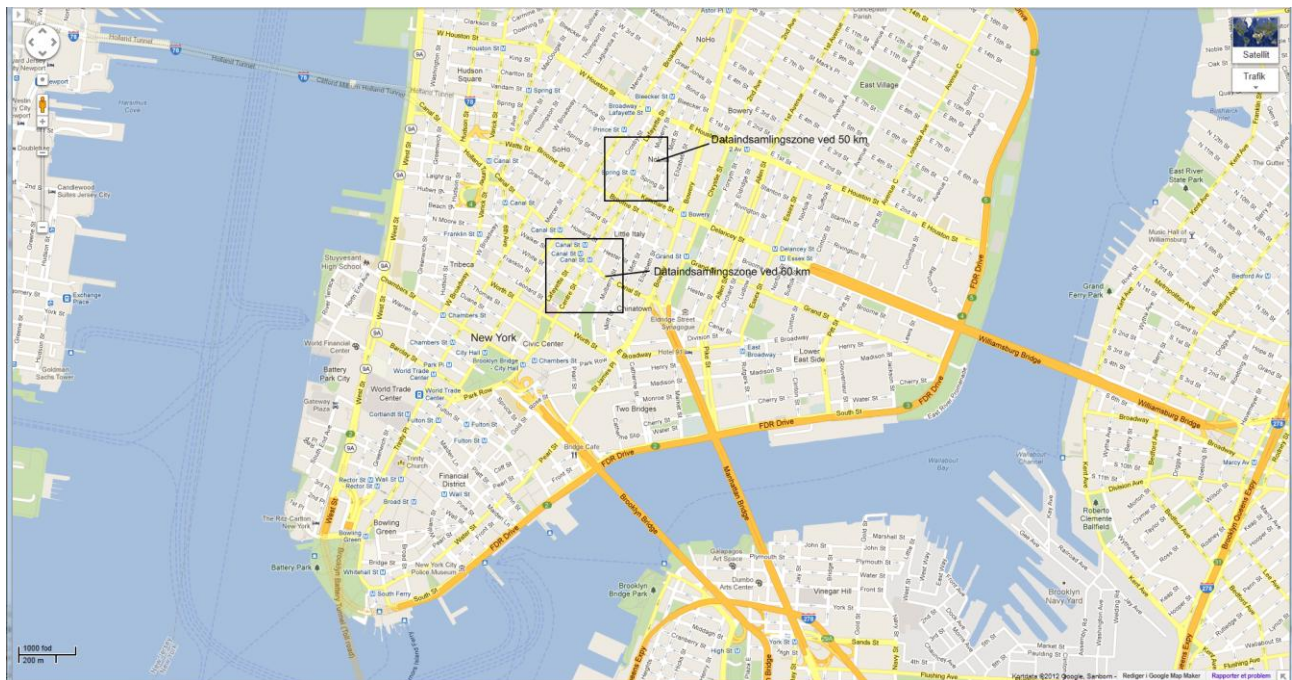
Figur 6: målestoks forhold 1:10000 her ses to DIZ'er der er defineret til at kunne indsamle to rækker ved en hastighed på henholdsvis 110 km og 130 km. Der ses også hvordan problemet med at store DIZ'er der dækker over minder veje der ikke noget at gøre med motorvejen kan løses ved at opdele DIZ'en i minder zoner.

Hvis der skal fortages sving i krydset så vil den tid en bilist er i zonen være større.

Kan det bruges i andre byer ?

Igennem dette projekt er Aalborg blevet brug som eksempel for hvordan systemet vil virke, samt for at belyse hvor store de forskellige zoner skal være. Der er også brugt en lille landsby Nøvling uden for Aalborg for at se forholdet af hvor store zonerne skal i tæt bymæssig bebyggelse og ude på landet, for at yde en så høj grad anonymitet som muligt.

For at se om systemet vil virke i en anden by er definitionen af DIZ'er der er brugt igennem projektet for 50 km og 60 km. oprettet på Manhattan, så kan de ses at der vil opstå et problem med at de vil dække over op til flere kryds se figur 7. Det vil sige at vis der ønskes at opsætte dataindsamlingszoner i andre byer rundt omkring i verden, da skal de tilpasses de forhold som de skal oprettes under, ellers vil det udemærkede kunne bruges i andre byer.



Figur 7: målestoks forhold 1:15000, Manhattan, New York, her ses det hvordan det vil se ud hvis der oprettes DIZ'er på Manhattan ud fra de forudsætninger der er brugt til at oprette DIZ'er i Aalborg. Størrelserne på disse DIZ'er er for 50 km og 60 km. På Manhattan ses det at de skal være mindre eller vil de dækker over op til flere kryds.

På figur 7 ses det at DIZ'er der oprette med den størrelse de er beregnet til at skulle have, dækker over for mange kryds. Dette vil kunne løse ved at brug polygoner i stedet for rektangler, der er tilpasse de kryds og veje hvor der ønskes at indsamle data fra.

Beregning af længde for DIZ

Der er lavet følgende beregninger ved de mest almindelige hastigheder i byer, uden for byer og motorveje, ved tyve rækker med en indsamlings frekvens på et sekund.

Kilometer i timen	Meter pr sekund	Længden i meter
30	8,33	166,6
40	11,11	222,2
50	13,89	277,8
60	16,67	333,4
70	19,44	388,8
80	22,22	444,4
90	25	500
110	30,56	611,2
130	36,11	722,2

Tabel 1: længden på en DIZ ved en given hastighed

Det vil sige ved minimum tyve rækker skal størrelserne på en DIZ være som de ses i tabel 1. Dataene er beregnet ud fra at det kan garanteres at der modtages at GPS pr. sekund. Se appendiks A1.

Maksimum og minimum størrelse på en Dataindsamlingszone

Den minimale størrelse på en dataindsamlingszone er fast sat ud fra vores krav om at vi vil have minimum 20 rækker fra hver bilist pr dag. Dette giver nogle minimums størrelser på hvor små dataindsamlingszonerne kan være, se afsnittet om beregning af længde for dataindsamlingszone, her er der beregnet hvor lang en dataindsamlingszone skal være i forhold til hvor hurtigt bilisten kører.

Der er ikke nogen øvre grænse for hvor stor en dataindsamlingszone kan være. En fornuftig maksimal størrelse vil være en zone hvor det tager 1 minut at køre igennem, det vil sige at der opsamles 60 punkter eftersom der indsamles data med en frekvens på et GPS punkt pr. sekund.

Selvom der ikke er nogen øvre grænse for hvor stor en dataindsamlingszone kan være, så vil det ikke give mening hvis de bliver så store at de dækker over flere kryds. Så hvis det ønskes at DIZ'en skal være mere end 50 % større end den størrelse DIZ'en vil have ved den tilladte hastighed, så vil det være til rådigt at at opdele DIZ'en i minder DIZ'er.

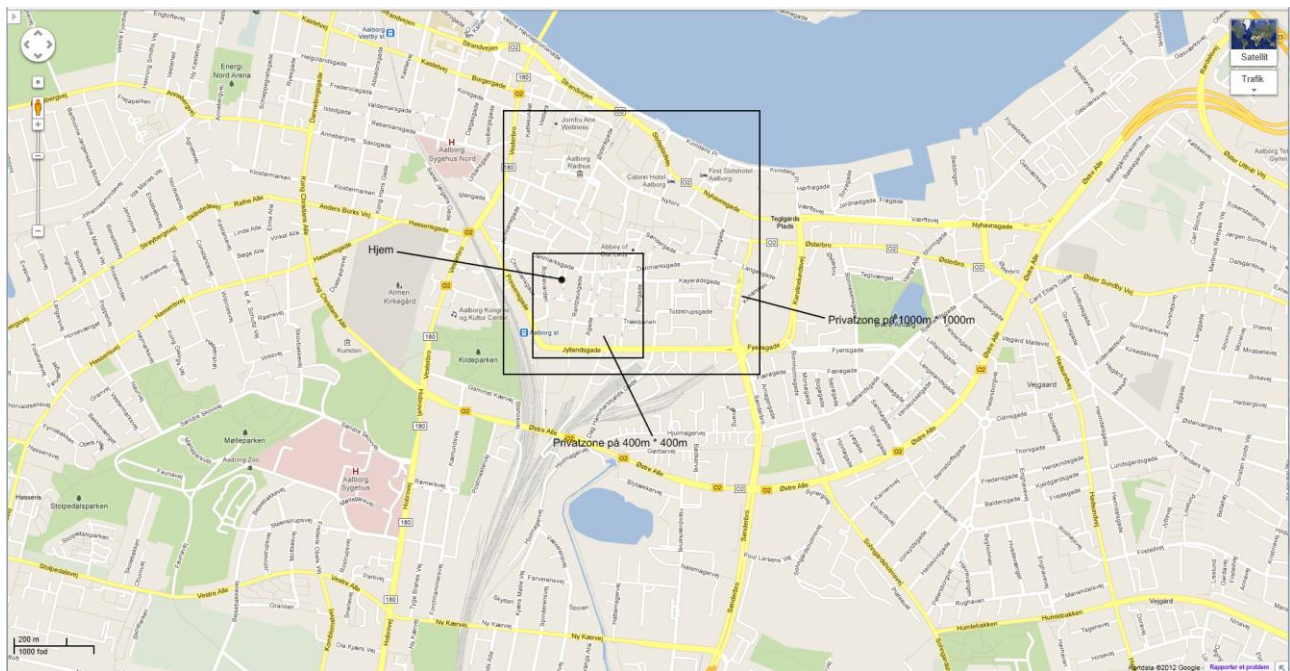
Ved 50 km er størrelsen af DIZ'en 277,8 meter se figur 4, så vil den have en størrelse på 416,70 meter, for motorvej gælder det for 130 km er størrelsen 722,2 meter se figur 6, ved 50 % forøgelse være 1083,3 meter.

Når der ligger en DIZ over et vejkryds eller et stykke vej, kan det opleves at den skal være meget stor i forhold til krydset. Da DIZ'en følger breddegrad og længdegrad gitteret, men det gør krydset eller vejen ikke. Dette kan afhjælpes ved at opdele DIZ'en i flere minder DIZ'er, se figur 6 hvor en motorvejs DIZ er opdelt i minder DIZ'er. Når en DIZ opdeles i minder DIZ'er kan der være overlap mellem dem, dette er tilladt men det skal så vidt muligt undgås.

Størrelse på en privatlivszone

Størrelsen på en PZ afhænger meget af hvor den er opsat, om det er i tæt bymæssigt bebyggelse eller om det er uden for byen, hvor bebyggelsen er mere spredt.

I tæt bymæssigt bebyggelse vil PZ være opsat som et rektangel, med en side længde på 400 m. Hvor bilisten bor et sted i zonen dog ikke centrum.

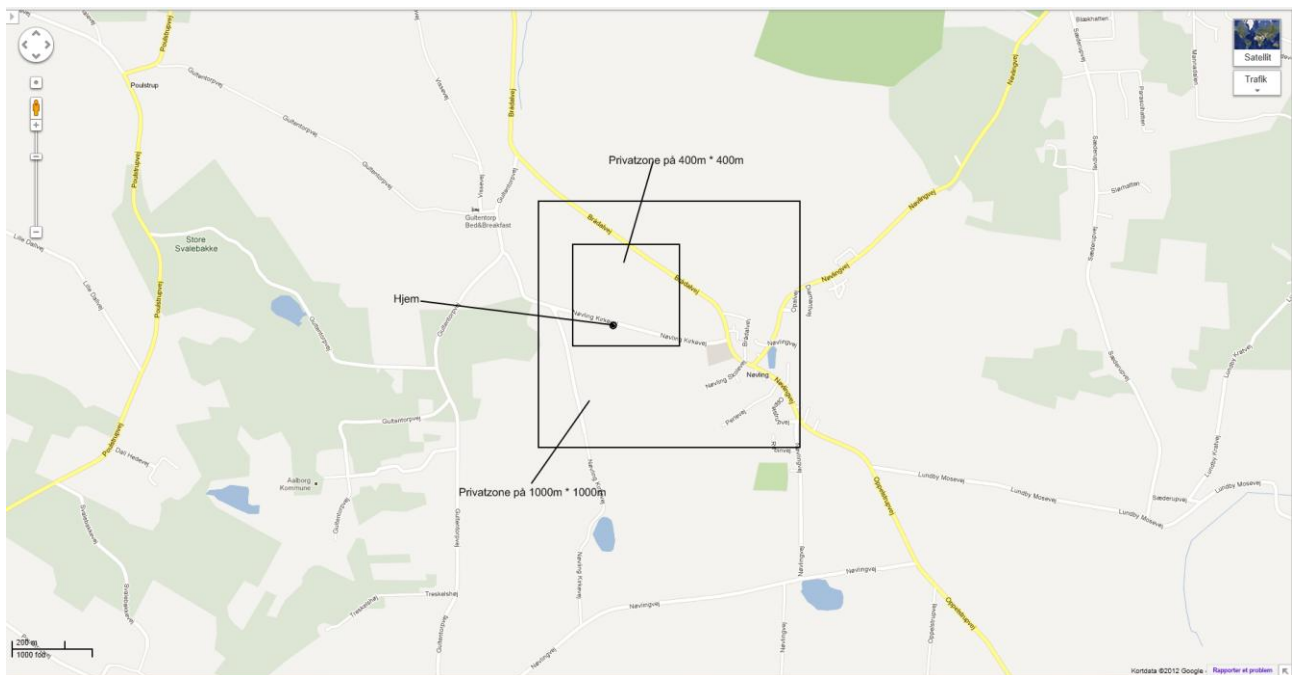


Figur 8: målestoks forhold 1:10000, hvor der er op sat to PZ en på 400 m og en på 1000 m.

Ved den valgt størrelse på en PZ er der mulighed for at dække over flere gader / boligblokke / ejendomme i byen, se figur 8. det kan konkluderes at en PZ i tæt bymæssigt bebyggelse skal have en maksimal størrelse på 400 m ellers vil den dække over for mange kryds. Hvor det vil kun resultere i ikke brugbare data. Desuden skal der være en grænse på hvor mange PZ en bilist kan opsætte, da der ellers vil være mulighed for at opsætte PZ'er så de dækker en hel by, denne grænse er sat til ti PZ'er. Ved at kunne opsætte op til ti PZ'er vil der være mulighed for at kunne opsætte PZ'er ved hjemmet, samt andre steder hvor mange har et ønske om at kunne være anonym.

Er PZ opsat uden for tæt bymæssig bebyggelse skal den sandsynligvis være større, for at den vil dække et stort nok areal til at give bilisten en sådan anonymitet, at det ikke kan gættes hvor han bor.

I figur 9 er der opsat to PZ'er med samme størrelse som dem der er opsat i tæt bymæssig bebyggelse, her ses det at ved at opsætte en PZ med den størrelse som giver en høj grad af anonymitet i byen ikke give det på landet. Ved at opsætte en PZ af denne størrelse vil da være forhold vis nemt at gætte hvor bilisten bor. Det ses at den PZ der var for stor til at bruge i byen, giver en høj grad af anonymitet på landet. Det kan konkluderes at en PZ i på landet skal have en størrelse på 1000 m. antallet af hvor mange PZ'er der kan oprettes er ens for både by og land.



Figur 9: målestoks forhold 1:10000, hvor der er op sat to PZ en på 400 m og en på 1000 m.

Ud fra figur 8 og 9 ses det at størrelse for en PZ er forskellig fra bymæssig bebyggelse til ikke bymæssig bebyggelse.

Hvilket giver at de størrelser for en PZ på 400 m i bymæssig bebyggelse og 1000 m på landet vil være en maksimalt øvre grænse. Men man skal lige tænke på hvad man får ud af at gøre den meget stor og hvilke dataindsamlingszoner den kan komme til at overlappe. Hvis man gør en PZ alt for stor burde man måske overveje om hvorvidt om man ønsker at afgive data.

Opsætning af Zoner

Hvor der opsættes enten DIZ'er eller PZ'er er det hurtigst at op sætte dem ud fra det øverste venstre hjørne og nederst hjørne, det giver et rektangel der følger gitteret for bredde- og længdegraderne. Ved at brug et sådan rektangel fås der problemer at ved vejkryds der ikke pænt følger gitteret for breddegraderne og længdegraderne, vil det være nødvendigt at gøre DIZ større nødvendigt for at dække det ønskede stykke vej igennem vejkrydset. For at løse det problem at et vejkryds ikke pænt følger gitteret for breddegraderne og længdegraderne, så kan der bruges et polygon til at definere dataindsamlingszonen så den er tilpasset til vejkrydset udformning. Der kan være opsat regler om hvornår igennem dagen der skal indsamles data. På kortet vil den ses som en blå transparent rektangel.

Dataindsamlingsalgoritmen

I dette afsnit beskrives dataindsamlingsalgoritmen, dens formål og virkemåde.

Dataindsamlingsalgoritmen har til formål at håndtere hvordan indsamlingen af data skal forgå. Det vil sige at den undersøger om en bilist er i en DIZ, er der overlap mellem en DIZ og en PZ, samt beregne afstanden til en DIZ. Derudover styre den også de forskellige DIZ prioriteter samt hvordan de har indflydelse på hinanden.

Generel data indsamling i en DIZ

Der kan indsamles data på følgende to måder. Der vil blive indsamlet data ud fra at en bilist kommer ind i en prioritet et eller to zone. Den anden måde er at afstanden til den nærmeste DIZ er for lang, er dette tilfældet vil der blive indsamlet data uden for en DIZ, det vil sige at der vil blive indsamlet data i en prioritet tre zone.

Det vil sige at de indsamlede data kan indeles i følgende tre dataindsamlings zoner prioriteter.

- Zoner med prioritet et, højeste prioritet data indsamlet i disse zoner skal altid gemme.
- Zoner med prioritet to, næst højeste prioritet, kan slettes til fordel for prioritet et data.
- Zoner med prioritet tre, laveste prioritet, kan slettes til fordel for prioritet et og to data.

For en Prioritet et og to zoner gælder det at de er opsat af trafikanalytikeren. Prioritet tre zoner bruges til at indsamle data fra bilister som har givet tilsagn om at de vil levere data til systemet, men de kommer ikke igennem dataindsamlingszoner. En prioritet tre zoner bliver beregnet at systemet. Se figur 11, hvor de forskellige zoner er opsat.

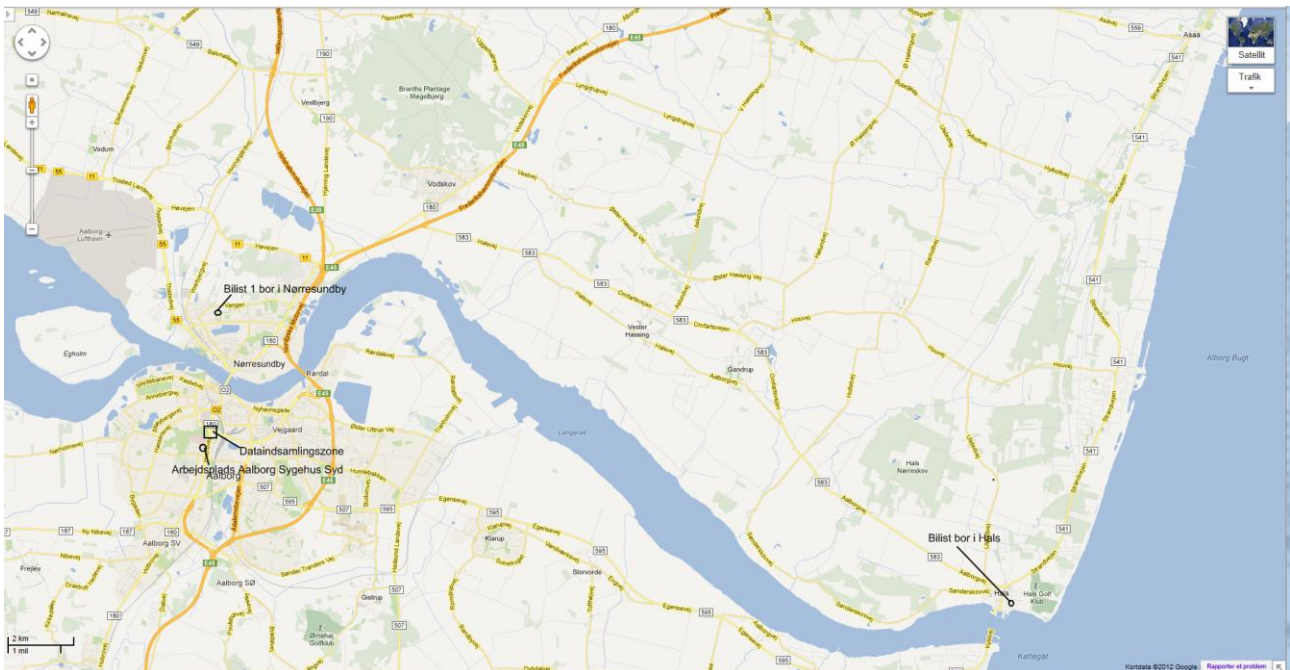
Algoritmen skal sikre at PZ og regler for disse altid overholdes, en PZ har den højeste prioritet i systemet. Det vil sige at hvis bilisten kommer ind i en PZ så stopper indsamlingen af data. Desuden overtrumfer en PZ altid en dataindsamlingszone, det vil sige at hvis en PZ overlapper et stykke af en PZ så vil der ikke blive indsamlet data fra denne PZ. Når bilisten kommer ud af en PZ da beregnes til næste dataindsamlingszone

Dataindsamlingen begynder ved at algoritmen beregner om nogle af DIZ er aktiv, da der er mulighed for at opsætte regler for hvornår en DIZ skal være aktiv, hvis der ikke er nogle aktive så beregnes hvor lang tid der er til den første bliver aktiv. Den tid der er beregnet til at en af DIZ er aktiv kaldes sovetid, algoritmen beregner flere sovetider.

Er der nogle aktive DIZ da beregnes afstanden til disse i tid, den beregnet tid kaldes også sovetid, tiden er beregnet ud fra hvor langt der er og bilistens hastighed, hvis hastigheden er under 10 km så sættes den til 50 km. Dette er gjort da de første dataindsamlinger er fortaget mens bilisten er ved at starte, køre ud fra en parkeringsplads. Hvis afstand er beregnet ud fra at hastigheden var under 10 km på beregnings tidspunktet, så vil algoritmen fortage en ny beregning efter halvdelen af den beregnet tid, dette gøres for at sikre at afstanden til DIZ'en er beregnet så korrekt som muligt.

Hvis afstanden fra hvor en bilist starter og til den nærmeste DIZ er over 20 minutter, vil der blive indsamlet data uden for en DIZ. De 20 minutter er valgt ud fra den betragtning at et en hastighed på 50 km kan der køres næsten 17 km. Ud fra dette er det kan det antages at hvis en bilist er mere end 17 km fra den nærmeste DIZ, er det ikke givet at bilisten køre til denne DIZ, og derfor vil der blive indsamlet data uden for

DIZ for at sikre at bilisten levere 20 rækker data, som er et af kravende se figur 10. Der vil blive opsamlet data med en opsamlings frekvens på 1 række pr. sekund.



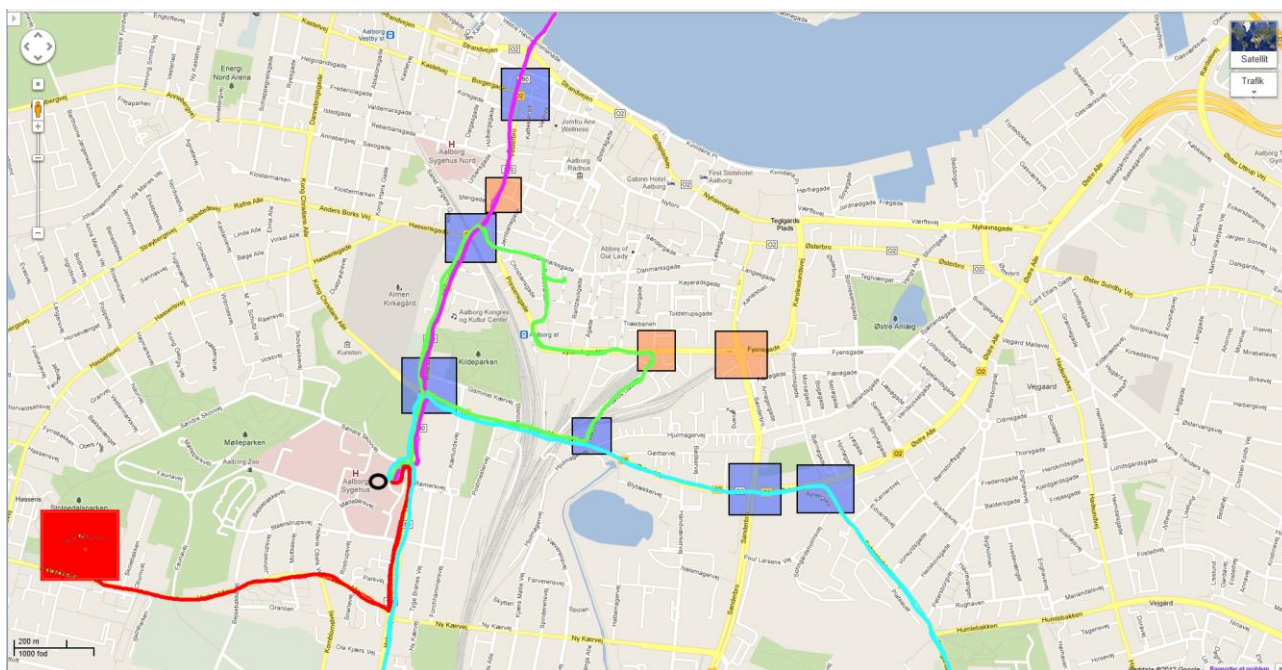
Figur 10: målestoks forhold 1:800000, hvor der ses at bilist der bor i Nørresundby kan nå sin arbejdsplads på under tyve minutter, men bilist 2 der bor i Hals kan ikke nå indtil sin arbejdsplads på under tyve minutter.

Dataindsamling i en DIZ

Når en bilist begynder at køre inden for det tidsinterval som bilisten har opsat for hvornår han ønsker at indsamle data. Så begynder systemet at beregne afstand til nærmeste DIZ for at begynde indsamlingen af data.

Når bilisten når frem til en enten prioritets et eller to DIZ og det er sikret at der ingen overlap er med en PZ, så begynder der at indsamle data.

Når bilisten forlader DIZ'en igen da beregnes afstanden til næste DIZ i minutter, den beregnet tid er sovetid.



Figur 11: målestoks forhold 1:10000, her ses hvordan der er opsat DIZ med prioritet 1 og 2 zoner. Lyseblå er prioritet 1 og orange prioritet 2 zoner. Der er også opsat en PZ, der er den rød massive rektangel, har en bilist opsat den for at beskytter hvor dennes hjem er. Figuren vise også bilisters vej til arbejde, hvor nogle kommer igennem flere forskellig DIZ'er, samt det at der en bilist som ikke kommer igennem en eneste på sin vej til arbejde. De gade- og vej navne som zonerne dækker over kan ses i appendiks A3.

Dataindsamling uden for DIZ

Hvis en bilist har meldt sig til systemet og vil lever data, men ikke kommer igennem en DIZ i løbet af en arbejdes dag da skal dette håndteres. Det er gjort ved at så snart klient applikationen detekter at den er uden for en PZ, og system tiden er i et af de tids intervaller hvor der indsamles data.

Afstanden beregnes til den nærmeste DIZ hvis den beregne tid er mere end 20 minutter, da vil applikationen begynde at indsamle data se figur 10. Indsamling af data vil forgå på følgende måde.

De 20 minutter er valgt ud fra, ved en hastighed på 50 km i timen vil kunne køre 16 km på 20 minutter. Hvis en bilist er mere end 16 km fra den nærmeste DIZ, er det ikke sikre at bilisten kører i retningen af DIZ'en. Så for at sikre at der indsamles data, begynder der at indsamles data efter reglerne for dataindsamling uden for en DIZ.

For at sikker at der vil blive indsamlet nok data fra en bilist, er det bestemt at dataindsamlingen skal forgå på følgende måde. Der bruges fire intervaller af længden 20 minutter se figur 12, det giver i alt et indsamlingsinterval på 80 minutter, ved at have et så langt interval kan der dække lange ture over et stort område.

De fire intervaller hvor i hvert interval er inddelt i fire underintervaller. Hvor de fire overordnede intervaller er af længden 20 minutter. Hvert af de fire under intervaller har længden 5 minutter.

20 minutter længde	20 minutter længde	20 minutter længde	20 minutter længde
--------------------	--------------------	--------------------	--------------------

Figur 12: den samlede interval længde $4 * 20 = 80$ minutter.

Først indsamles der data i fire underintervaller, hvor hvert enkelt interval er på fem minutter se figur 13.

Når der er data i alle fire underintervaller, i det første overordnede interval

1	2	3	4			
---	---	---	---	--	--	--

Figur 13: de fire underintervaller i det første overordnede interval er der data i.

Når der er data i alle fire underintervaller, i det første overordnede interval, så vil der beregnes en sovetid, denne sove tid er fra sluttidspunktet af det første overordnede interval, plus et tidsinterval på 0 – 15 minutter. Sove tiden er beregnet random, dette gøres for at sikre en spredning af underintervallerne. Systemet vil sove den beregnede tid, derefter vil et af de fire intervaller udvælges tilfældigt. Det valgte interval vil blive slettet, når dette er sket så vil der blive indsamlet data i dette interval. Det udvalgte underinterval vil nu være i det andet overordnede interval.

Der vil blive indsamlet data i dette underinterval se figur 14.

1	2		4		3			
---	---	--	---	--	---	--	--	--

Figur 14: Underinterval 3 er blevet slettet i det første overordnede interval, hvorefter der er blevet indsamlet data i det, i det andet overordnede interval.

Når der er blevet indsamlet data i det underinterval der er blevet flyttet til andet overordnede interval, så vil systemet beregne hvor meget tid der er mellem sluttidspunktet for underintervallet og til skellet mellem det andet og tredje overordnede interval, plus et tidsinterval på 0 – 15 minutter. Systemet vil sove den beregnede tid, derefter vil et af de tre underintervaller i det første overordnede interval udvælges tilfældigt. Det valgte interval vil blive slettet, når dette er sket så vil der blive indsamlet data i dette interval. Det udvalgte underinterval vil nu være i det tredje overordnede interval. Der vil blive indsamlet data i dette underinterval se figur 15.

	2		4		3			1		
--	---	--	---	--	---	--	--	---	--	--

Figur 15: Underinterval 1 er blevet slettet i det første overordnede interval, hvor efter der er blevet indsamlet data i det, i det tredje overordnede interval.

Når der er blevet indsamlet data i det underinterval der er blevet flyttet til andet overordnede interval, så vil systemet beregne hvor meget tid der er mellem sluttidspunktet for underintervallet og til skellet mellem det tredje og fjerde overordnede interval, plus et tidsinterval på 0 – 15 minutter. Systemet vil sove den beregnede tid, derefter vil et af de tre underintervaller i det første overordnede interval udvælges tilfældigt. Det valgte interval vil blive slettet, når dette er sket så vil der blive indsamlet data i dette interval. Det udvalgte underinterval vil nu være i det fjerde overordnede interval. Der vil blive indsamlet data i dette underinterval se figur 16.

	2			3			1			4	
--	---	--	--	---	--	--	---	--	--	---	--

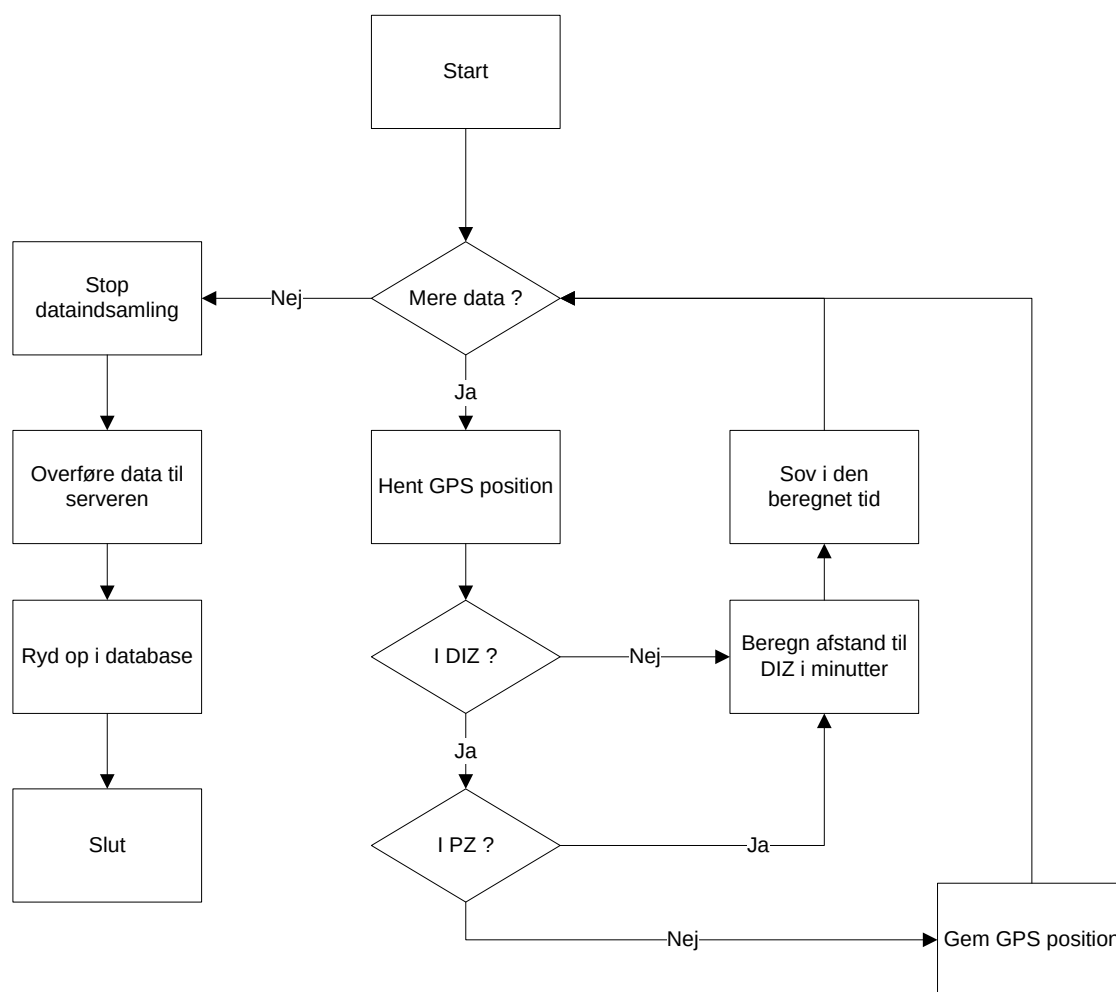
Figur 16: Underinterval 4 er blevet slettet i det første overordnede interval, hvor efter der er blevet indsamlet data i det, i det tredje overordnede interval.

Ved at indsamle data på denne måde har vi fået indsamlet data i fire intervaller af en længde på fem minutter. Det vil sige at vi har fået indsamlet data fra en bilist som har givet tilsagn om at ville levere data men ikke kommer igennem en DIZ. Samt ved at indsamle data sådan har vi fået data over et tidsinterval på 80 minutter.

Ved at samle data ind fra bilister som både køre igennem en DIZ og fra bilister der ikke køre igennem en DIZ får vi en så bred repræsentation af data som muligt.

Flowchart Diagram

I dette afsnit beskrives hovedalgoritmens virkemåde ud fra et flowchart diagram.



Figur 14: flowchar, beskrivelse af dataindsamlings algoritmens virkemåde. I figuren er der ikke vist hvilke prioritet zonen har hvor fra der bliver indsamlede data, grunden er at når det afgøres om den indsamles data er i en DIZ, så undersøges det også hvilke prioritet zonen har.

Pseudo kode

I dette afsnit opstilles pseudo kode der beskriver algoritmens virkemåde. Pseudo koden er opdelt i tre afsnit. Afsnit et er hvordan reglerne for DIZ'erne undersøges, det vil sige hvornår skal der indsamles data i hver enkelt DIZ, så beregnes afstanden fra det punkt bilisten er i og til alle DIZ'er for at finde ud af om der skal indsamles data i en prioritet et og to zoner eller i en prioritet tre zone, til sidst undersøges det er om

der er sat en begrænsning på hvor meget data der må indsamles pr. døgn. Afsnit to er hvordan der indsamles data i en prioritet et og to zone. Afsnit tre er hvordan der indsamles data i en prioritet tre zone.

Hent systemtid.

```
for alle DIZ hent regler
    undersøg start tid for DIZ

    if systemtid < start tid for DIZ
        beregn sovetid til starttid for DIZ og vent

    if starttid for DIZ < systemtid < stoptid for DIZ
        start dataindsamling
    if stoptid < systemtid
        beregn sovetid til næste starttid for DIZ og vent.

    hent min GPS position
    for alle DIZ breddegraders øverste venstre hjørne
        beregn afstand fra min GPS position og til DIZ'ernes øverste
        venstre hjørner

        Sæt bilens hastighed = 50 km.

        if Er afstanden < 20 minutter
            Beregn sovetid og vent, start derefter dataindsamlingen I zonen

        if er afstanden > 20 minutter

    if er der sat en øver grænse for hvor meget data der må sendes pr. døgn
        if er den øver grænse nået
            hvis ja stop dataindsamling
            hvis nej for med dataindsamling
```

Del et af dataindsamlingsalgoritmen håndtere regler for DIZ, afstand fra DIZ til bilisten. Der undersøges også om der er sat en øver grænse for meget data der må indsamles pr. døgn.

```
if skal der indsamles data fra en prioritet et eller to zone
    for alle DIZ breddegrader
        if er min GPS position i en DIZ
            for alle PZ breddegrader
                if er der over lap mellem DIZ'en og PZ'en
                    hvis ingen overlap mellem DIZ og PZ
                        gem da punktet
                    if skal der indsamles mere data
                        hvis ja forsæt dataindsamling
                    else
                        stop dataindsamling
```

Del to af dataindsamlingsalgoritmen håndter dataindsamling I en prioritets et eller to zone.

```
if skal der indsamles data fra en prioritet tre zone
```

```

for alle DIZ breddegrader
    if er min GPS position i en DIZ
        for alle PZ breddegrader
            if er der overlap mellem DIZ'en og PZ'en
                hvis ingen overlap mellem DIZ og PZ
                    gem da punktet
                if er der indsamlede 300 punkter
                    hvis nej forsæt indsamling
                else
                    skift til næste interval
                if er alle fire intervaller udfyldt
                    hvis nej forsæt indsamling
                else
                    if er det et fyldt under interval i hver af
                        de overordne intervaller
                            hvis nej forsæt indsamling
                    else
                        der er indsamlet den ønske mængde data,
                        og indsamlingen stoppes

```

Del tre af dataindsamlingsalgoritmen håndter dataindsamling i en prioritets tre zone.

Implementering

Implementeringen af applikationen er forgået ved brug af Eclipse, Java, og Google's egen Android API.

Eftersom applikationen gør brug af Google's kort, til at marker hvor der er DIZ og PZ, skal applikationen have adgang til kortet. For at få adgang skal der generes en nøgle igennem Google. Denne nøgle generes ud fra en md5 streng, hvor md5 strengen bliver generet på udviklerens computer ud fra en nøgle streng, der bliver installeret ved installationen af Android API'et. Når nøglen er kompillet med i kilde kode så har applikationen adgang til at bruge kortet.

Når applikationen skal installeres på en smartphone, er der flere forskellige muligheder. En måde er at bruge Android marked, eller kopiere apk filen (android installations filen) over på sd kortet. For at kunne installere applikationen fra sd kortet skal der bruges en applikations installer, jeg har brugt Astro app installer, den er downloadet fra Android marked. Med Astro app installer er der mulighed for at læse og installere filer fra sd kortet, skal applikationen slettes igen kan dette gøres igennem telefonens egen applikations manager.

Applikationen gemmer de indsamlede data på sd kortet, dette er gjort for at lette tilgangen til de indsamlede data, for videre arbejde igennem projektet. Når systemet er i drift vil data sendes fra telefonen til serverne igennem netværkskomponenten, hvor netværkskomponenten i test ikke er i brug.

Når applikationen ikke bruges til test vil de indsamlede data gemmes i den SQLite data der er på smartphonen. Breddegrader og længdegrader er i database gemt som to enkelte celler, og ikke et Geo punkt. Grunden til dette er at SQLite i Android ikke understøtter brugen Geo punkter.

I dette projekt er der gjort brug af en HTC Wildfire s smartphone, klient applikationen vil kunne køre følgende versioner af Android, 2.3.3 og fremefter se Appendiks A2 for kode beskrivelse.

Tabellerne til database på server siden er lavet i Toad, hvor det derigennem er autogeneret scripts.

Database design

I systemet bruges der følgende database tabeller:

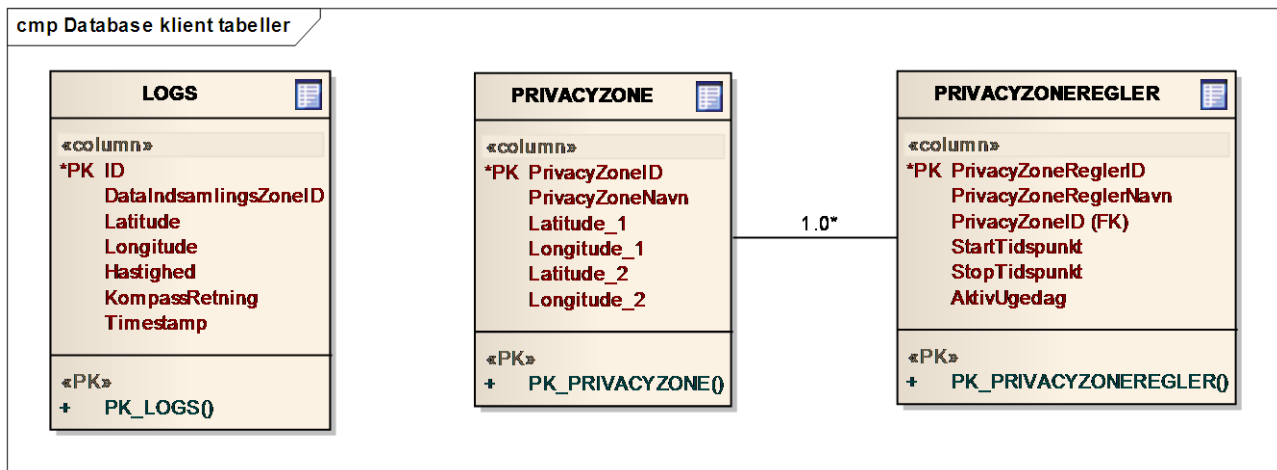
- LOGS
- PRIVACYZONE
- PRIVACYZONEREGLER
- DATAINDSAMLINGSZONE
- DATAINDSAMLINGSZONEREGLER
- INDBAKKETABEL

LOGS tabellen bruges til at gemme de opsamlede data på klienten før de sendes til serveren. På serveren er der en indbakketablel der modtager data fra de forskellige kliner, det vil sige at alt data kommer ind i indbakketablellen hvor fra det bagved liggende system læser disse data for vider behandling. Det vil sige at indbakke tabellens funktion er at den modtager data fra klienterne som skal vider bearbejdes i systemet

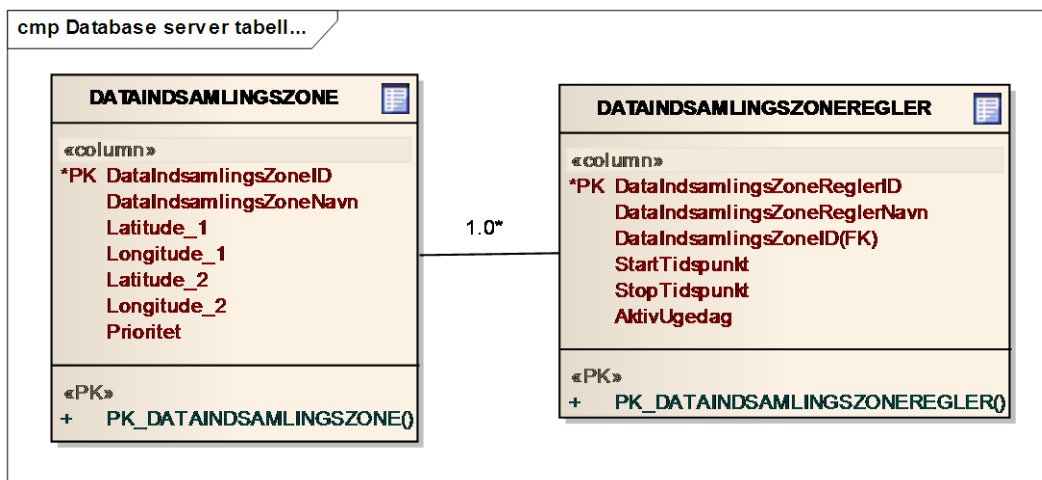
Privacyzone og privacyzoneregler tabellerne bruges kun på klienten, de bruges til at gemme opsætning for en privacyzone. Det vil sige hvordan den er defineret ud fra det øverste venstre hjørne (latitude_1, longitude_1) og nederste højre hjørne (latitude_2, longitude_2). Samt hvornår der gives tilladelse til at der må sendes data fra en privacyzone.

Dataindsamlingszone og Dataindsamlingszoneregler tabellerne bruges på både server og klienten. De bruges til at gemme opsætninger for dataindsamlingszone. Det vil sige hvordan den er defineret ud fra det øverste venstre hjørne (latitude_1, longitude_1) og nederste højre hjørne (latitude_2, longitude_2). Samt hvornår der skal indsamles data fra en zone. Dataindsamlingszone tabellen indeholder også en prioritet, hvor denne prioritet bruges til at udvælge hvilke dataindsamlingszoner der skal sendes data fra i de tilfælde hvor der er sat en begrænsning på hvor meget data en bilist vil sende om dagen. Prioriteten er fra 1 til 3 hvor 1 er den højeste og 3 den laveste.

På serveren er der en INDBAKKETABEL, denne tabel bruges til at modtage de data der sendes fra klienterne til serveren. Det vil sige at Indbakketablellen er en replikation af Logs tabellen fra klienten. På figur 15 ses tabellerne på klienten og figur 16 tabellerne på serveren.



Figur 15: Databasetabeller på klienten



Figur 16: Databasetabeller på serveren.

Udover de viste tabeller på klienten indeholder den også Dataindsamlingszone og Dataindsamlingszoneregler tabellerne. Men da der er en 1 til 1 mapning af tabellerne på serveren er de kun medtaget på tabeldiagrammet for serveren.

Det samme gør sig gældende for Logs tabellen da denne også er en 1 til 1 mapning af tabellen på klienten, er den kun medtaget på klient tabeldiagrammet.

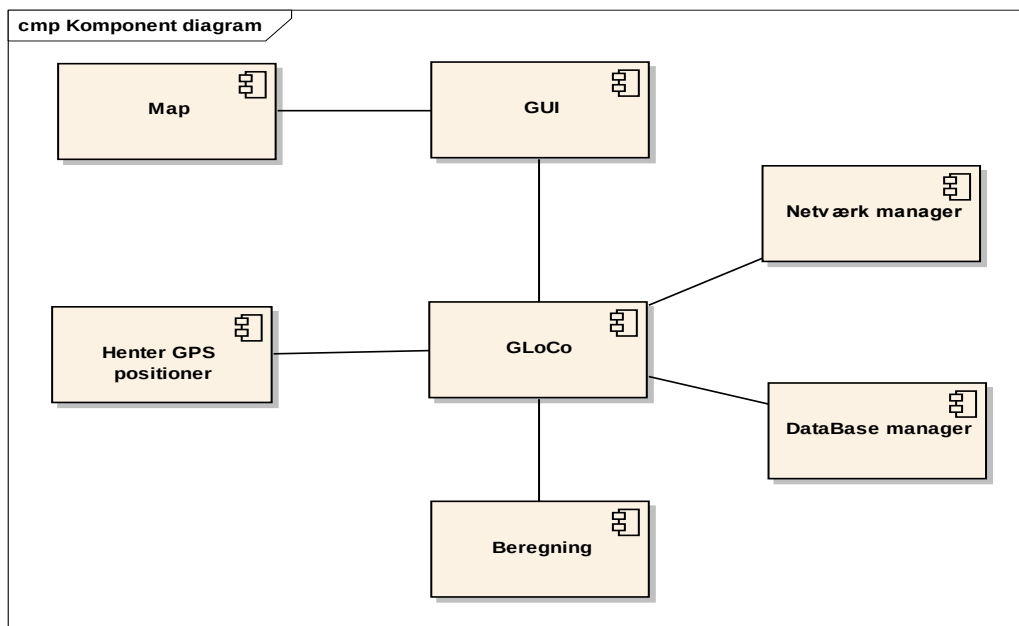
Komponent diagram

I dette afsnit beskrives hvilke komponenter klient applikationen består af samt deres virkemåde.

Applikationen består af følgende komponenter

- Map komponenten står for at vise kortet, i dette tilfælde bruges Google maps, hvor der er mulighed for at swipe, markere positioner, og zoome ud og ind.
- GUI komponenten binder de forskellige grafiske bruger flader sammen, som f.eks. map, konfigurations side og listerne over alle PZ og dataindsamlingszoner.
- GLoCo (navnet på app'en) dette er hoved motoren, den som holder styr på at kalde en GPS position, kalde de forskellige beregninger når der er brug for dem. F.eks. kald beregningerne for at beregne hvilken dataindsamlingszone bilisten kører imod samt beregne sovetid. Sæt applikationen til at sove i den beregne sovetid. Kalde database manageren for at enten læse / skrive data til databasen. Kalde netværks manageren for at sende de indsamlede data til serveren, modtage opdateringer om dataindsamlingszoner.
- Henter GPS position denne komponent står for at hente en GPS position og send den til GLoCo komponenten.
- Beregnings komponenten indeholder alle de beregnings algoritmer som applikationen gør brug af. F.eks. beregning af sovetid, hastighed, kompasretningen, hvilke datasamlingszone bilisten kører mod, hvordan udvælges de data som sendes til serveren når der er sat en begrænsning på hvor meget data der må sendes i løbet af en dag etc.
- Netværk manager står for alt kommunikation med serveren, send de indsamle data, modtage opdateringer på de forskellige dataindsamlingszoner.
- Database manager står for at oprette database første gang applikationen startes, og ellers står den for at gemme alle de indsamlede GPS positioner og hente dem igen når de skal sendes til serveren. Den håndter også opsætning af dataene for dataindsamlingszonerne og deres regler.

Komponentdiagrammet er vist på figur 17, hvor det ses at hovedkomponenten er GLoCo komponenten, det er den står for at indhente GPS positioner, få dem beregnet. Fortage alle beregner af afstande til DIZ'er og PZ'er, kommunikation op mod gui'en samt kommunikation op mod netværk- og database managerne.



Figur 17: Komponent diagram for applikationen (GLOCo)

Relateret arbejde

Et projekt der ligner meget det som der er arbejdet med igennem denne rapport er : Evaluation of Traffic Data Obtained via GPS-Enabled Mobile Phones: the Mobile Century Field Experiment[2]. Hvor der i denne artikel er et forsøg det er gennemført i San Francisco Bay, hvor der er brugt en Nokia N95 smartphone. Har de haft 165 studerede til at køre rundt med i alt 100 N95'er, for en bestemt dag at indsamle data i et bestemt tidsrum på 3 timer.

Igennem forsøgt sendes der live data til systemet der så beregner en real tids kørselstid, samt at der undersøges muligheden for at kunne opretholde en bestemt dækningsgrad af bilister igennem forsøget. Der undersøges også om bilisterne har kørt i ring.

Set i forehold til det forsøg der ligger til grunde for denne rapport, er der ingen anonymitet, eller bestemte dataindsamlingszoner. Det vil sige at ingen har anonymitet og der indsamles data hele tiden. Men ved at modtage data i real tid beregnes køretiden rundt i San Francisco Bay og kan give et øjebliksbillede af trafik situationen. Samt at det søges at opretholde en bestemt dækningsgrad af bilister, så der vides hvor mange der minimum skal være med til at levere data til systemet.

Feldtforsøg

Til at vise at algoritmen og klienten virker er der opstillet følgende tre mål:

- Mål 1: eftervise at der er muligt at indsamle brugbar GPS data, ved hjælp af en Android smartphone.
- Mål 2: eftervise at der kan opsamles data, og kun, i de opsatte dataindsamlingszoner. Derudover skal der eftervise at når en bilist køre inde i en dataopsamlingszone så vil der blive opsamlet data for denne bilist. Samt mulighed for at have zonerne aktive på forskellige tidspunkter.
- Mål 3: at det er muligt at sikre bilisterne en så høj grad af anonymitet som muligt, det vil sige at hvis en bilist er i en privatzone så må der ikke opsamles data om bilisten i denne zone. Desuden gælder det at hvis der er et overlap mellem en privatzone og en dataindsamlingszone så må der ikke opsamles data fra den berørte dataindsamlingszone.

For at eftervise at de tre mål er blevet opfyldt, har jeg på min vej til og fra arbejde indsamlede data. Data hvor der er indsamlet breddegrader, længdegrader, hastighed, tid i millisekunder og kompasretning, desuden er der indsamlet data er i KML formatet. Med KML data er det i Google Earth muligt at se den vej jeg har kørt til og fra arbejde, samt hvor der er PZ'er.

Jeg kører fra Sulsted nord for Aalborg, og til Nykredit i Aalborg øst. Hvor der er en PZ der dækker Sulsted hvor jeg bor, min arbejdsplads i Aalborg øst, derudover er der en PZ omkring Vestbjerg en by jeg kommer igennem på vej til arbejde.

Resultat

Jeg har kørt fra Sulsted og til Nykredit i Aalborg øst hvor der er indsamlede data. På figur 18 ses den kørt rute, det ses at der først begyndes at opsamle data når jeg er ude af den PZ der beskytter hvor jeg bor. Når jeg kommer til Vestbjerg er der igen en PZ, og der bliver ikke opsamlede data her. Når jeg kommer frem til min arbejdsplads stopper indsamlingen før jeg kommer helt frem. Det vil sige at mit hjem og min arbejdsplads begge ligger i en PZ, og herigennem er deres præcise adresse anonymiseret. Det ses også at når jeg kommer igennem Vestbjerg så er min færden her igen anonymiseret. Se appendiks A2 for data. Mens disse data er blevet indsamlet har applikationen været sat i baggrunden, så den er har kørt som en baggrunds applikation.



Figur 18, fra en øjenhøjde på 19 km: viser hele ruten fra Sulsted som ligger i en PZ igennem Vestbjerg som er en PZ, og frem til min arbejdsplads der ligger i en PZ.



Figur 19, fra en øjenhøjde på 3.30 km: Her ses det klar hvor PZ'en sluttet og data indsamlingen begynder.

På figur 20 ses det hvordan dataindsamle opføre sig når en bilist kommer ind i en PZ. I KML dataene kan ses at der er et spring fra 57,14072875 og til 57,12449189, henholdsvis nord og syd for Vestbjerg.

```
<coordinates>9.96861387,57.14094219</coordinates>
```

```
<coordinates>9.96864316,57.14072875</coordinates>
```

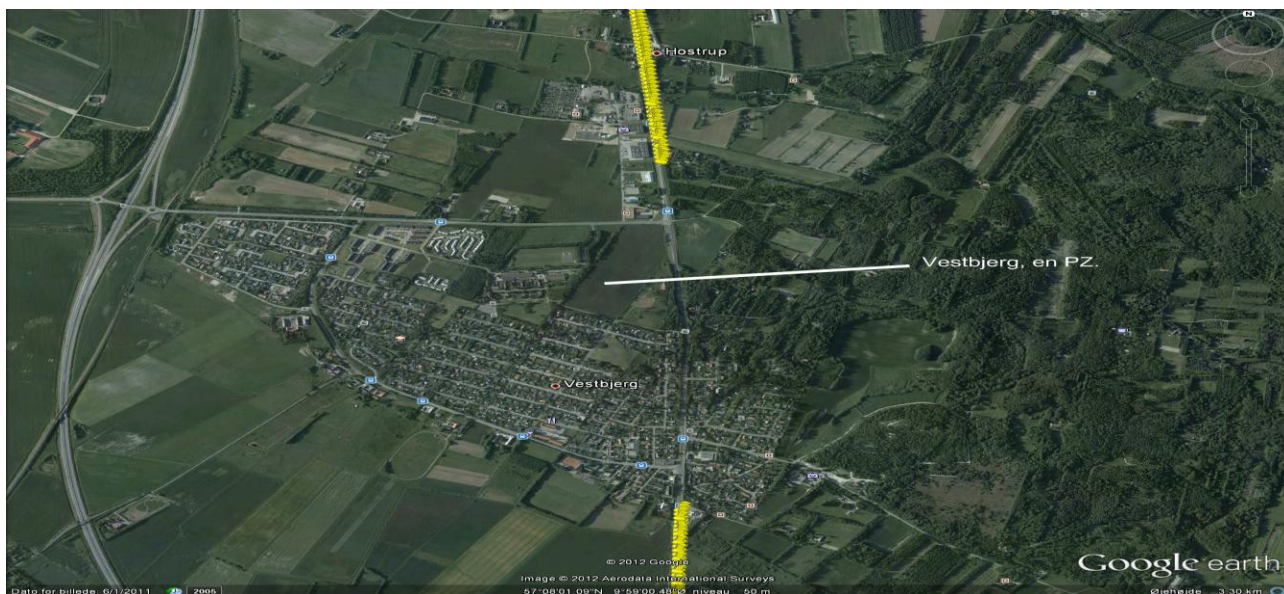
Her er jeg i en PZ

```
<coordinates>9.96955139,57.12449189</coordinates>
```

```
<coordinates>9.96953259,57.12432257</coordinates>
```

```
<coordinates>9.96953303,57.12415076</coordinates>
```

Det vil sige at når bilisten køre ind i en PZ så stopper dataindsamlingen og begynder igen når bilisten er kommet igennem.



Figur 20, fra en øjenhøjde på 3.30 km: her ses det at dataindsamlingen stopper ved den PZ der beskytter Vestbjerg. Samt at dataindsamlingen begynder igen efter PZ'en.



Figur 21, fra en øjenhøjde på 3.30 km: her ses det at dataindsamlingen stopper ved grænsen til den PZ der beskytter min arbejdsplads, Nykredit i Aalborg øst.

Ved at køre turen fra mit hjem til mit arbejde og tilbage, er den målt til at tage 32:18 minutter. Hvor der er indsamle data i to filer, i den ene GLoCo.gtx er det følgende data indsamlede for hver indsamlingspunkt.

my latitude = 57.14800615 my longitude = 9.9673053 Timestamp = 55 : 27 myTime = 1.338710127147E12 mySpeed = 22.5 myBearing = 353.70001220703125

my latitude = 57.14820606 my longitude = 9.96726962 Timestamp = 55 : 28 myTime = 1.338710128138E12 mySpeed = 22.0 myBearing = 354.0

hvor der for hvert punkt er indsamle følgende data :

- Breddegrad
- Længdegrad
- Et tidspunkt i millisekunder
- En hastighed
- En kompasret

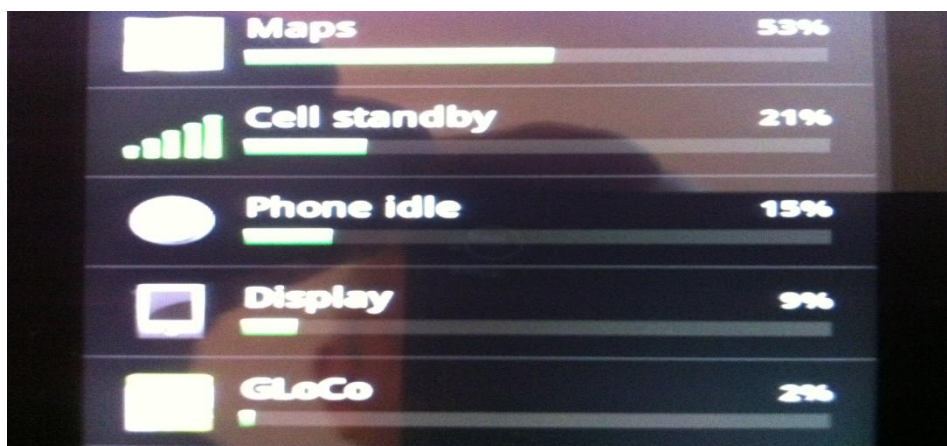
Disse data skal sendes til serveren for vider databearbejdning.

I GLoCo.kml filen er de indsamlede data gemt som KML data, for derefter at være loadet ind i Google earth. I Google earth kan den kørte rute ses.

	Ikke pakket	Pakket med zip
GLoCo.gtx	211643 kb	32525 kb
GLoCo.kml	143137 kb	14032 kb

Tabel 2: den indsamlede datamængde pakket og ikke pakket. Pakket med zip.

Ifølge tabel 2, skal der sendes 46557 kb pakket data fra klient til server over netværket for en tur på 32:18 minutter. Ved denne test har strømforbruget været at GLoCo applikationen har forbrugt 2%, de figur 22.



Figur 22: billede fra batteristatus på den bruge smartphone, det ses at når applikationen(GLoCo) er kørende så bruger den 2% af det samlede forbrug.

Ud fra ovenstående forsøg ses det at klient applikationen indsamler data som det er ønsket. Når bilisten er i en PZ så bliver der ikke indsamlede data, det opfylder kravet om at give bilisten en så høj grad af anonymitet som muligt, samt af det bestemmes hvor og hvornår der skal indsamles data. Samt at der er blevet indsamlede 46557 kb pakket data på 32 minutter, opfylder kravende om at klient skal have et minimalt memory forbrug, samt det skal belaste netværk så lidt som muligt når de indsamlede data sendes fra klienten til serverne. På figur 22, som er et billede taget af smartphonen's batteristatus se at applikationen (GLoCo) bruger 2% af det samlede strømforbrug, dette gør at den ikke dræner batteriet.

Videre arbejde

Det videre arbejde vil være tilføjelser af flere og ny funktionalitet til applikationen.

En ny funktionalitet vil være mulighed for at kommunikere med bilen CANBus protokol, for at trække informationer fra bilen. Disse informationer kan være bilens hastighed, motorens omdrejningstal, hvilket gear bruges der, temperaturen, oplysninger om benzin forbrug hvis der er muligt.

Brug systemet i real tid. Det vil sige at der kan ændres i dataindsamlingszonerne hele tiden og ændringerne træder i kraft med det samme. Så hvis der sker ting i trafikken da vil trafikanalytikeren oprette eller flytte en zone, hvorefter systemet vil få data fra denne zone. Ved at bruge real tid i systemet skal der tages højde for batteri forbruget, det vil gøre at bilisten sandsynligvis skal have et car kit installeret for ikke at dræne batteriet på smartphonen.

Hvis der i applikationen er et punkt hvor man kan angive graden af sin anonymitet, samt har mulighed for at angive hvad man er interesseret i, kan bilisten modtage målrette reklame.

Mulighed for at man kan oprette sig med et af systemet kendt id, man skal da gøres opmærksom på og acceptere at man så ingen anonymitet har, ved at gøre dette kan systemet analysere de indsamlede data for disse brugere. Ved at gøre brug af disse analyser, kan systemet i real tid give bilisten informationer hvordan vej situationen er for den vej som bilisten normalt benytter til og fra arbejde.

Konklusion

Resultatet af dette projekt, er udover denne rapport, en kørende klient applikation på en Android smartphone. Hvor klienten er brugt i et feltforsøg. I feltforsøget er der indsamlede data for at eftervis at der kan indsamles data, der overholdt de opstillede krav i Hypotesen.

Først skulle der indsamles data som er brugbare og giver mening at indsamle, derefter skal det sikres at der ikke indsamles data fra en privatzone, for at sikre bilisten en så høj grad af anonymitet som muligt. Det ses i feltforsøget at der blev indsamlet gode brugbare data der følger bilisten vej, når denne kommer ud af en privatzone, samt stopper igen når bilisten kommer ind i en privatzone. Dette viser at klienten sikre at bilisten får en så høj grad af anonymitet som muligt, da der ikke indsamles data for en bilist i en privatzone.

Under dataindsamlingen er det eftervist at klienten har virke som en baggrunds applikation. Klienten har uden problemer kunne opsamle data mens den har været sat i baggrunden. En af de applikationer som der har været sat i forgrunden har været en indbygget navigations applikation. Hvor det ses hvor bilisten er, denne applikation gør også brug af GPS. Men der har ikke været problemer ved at have begge applikationer kørende samtidig, der er blevet opsamlet data uden problemer.

For at sikre at klienten ikke dræner batteriet ved brug, er der igennem feltforsøget undersøgt hvor stort et strømforbrug klient har. Igennem hele feltforsøget har klient haft et batteri forbrug på 2%. Det vil sige at det ikke er klient der dræner batteriet.

Udover det meget begrænse strømforbrug, har klienten også et meget begrænset data forbrug. Ved en køre tur på 32 minutter er der indsamlet 46557 kb pakket data. Dette er med til at sikre at klienten vil belaste memory mindst muligt, samt når der skal sendes data over nettet, vil det være et meget lille dataforbrug.

Overordnet set er jeg tilfreds med den udviklede klient applikation. Den har vist at der kan indsamles brugbare data, med en smartphone. Derudover har den vist at klienten kan ligge i baggrunden, og den bruger en minimal mængde ressourcer.

Appendiks

A1. Beregning af længde for dataindsamlingszone

Til at beregne længden af en dataindsamlingszone ved en given hastighed og et bestemt antal rækker er brugt følgende formler. Hvor n er antallet af indsamlede datarækker i en dataindsamlingszone, x er hastigheden i km og t er tiden i sekunder.

$$(f1) \frac{x * 1000}{3600} = \text{meter pr sekund, ved hastigheden } x \text{ i km.}$$

$$(f2) n * \frac{x * 1000}{3600} = \text{længden af dataindsamlingszonen i meter.}$$

$$(f3) t * \frac{x * 1000}{3600} = \text{længden af en dataindsamlingszone i forhold til tid}$$

Der er lavet følgende beregninger ved de mest almindelige hastigheder i byer uden for byer og motorveje, ved tyve rækker.

Kilometer i timen	Meter pr sekund	Antal rækker	Længden
30	8,33	20	166,6
40	11,11	20	222,2
50	13,89	20	277,8
60	16,67	20	333,4
70	19,44	20	388,8
80	22,22	20	444,4
90	25	20	500
110	30,56	20	611,2
130	36,11	20	722,2

A2. Indsamlede data

De indsamlede data er vedlagt på cd-rom, hvor KML dataene ligger i KML folderen, og gtx ligger i gtx folderen. Klient koden der er skrevet i java, og går under navnet GLoCo ligger i GLoCo folderen. Koden er fordelt i flere filer, hvor de vigtigste er

- GLoCoActivity.java
- CollectALGO.java
- Tools.java

- GLoCoItemizedOverlay.java
- DatabaseManager.java.

GLoCoActivity: er en system fil hvor applikationen opsættes ud fra nogle Android specifikationer.

CollectALGO: er den fil styre indsamling af data, undersøger om bilisten er i en DIZ, eller en PZ, er der overlap mellem den DIZ bilisten er i om en PZ, beregner sovetid, afstand til nærmeste DIZ. Henter GPS positionen.

Tools: er hjælpe fil til CollectALGO, den består af hjælpe metoder til beregninger.

GLoColtemizedOverlay: er den fil der styre at tegne de forskellige zoner oven på kortet.

DatabaseManager: er den fil der styre al kommunikation med database, oprettes af tabeller, indsæt nye data, hente gamle data.

Under script folderen ligge de Toad generede script til at lave tabellerne på server siden. Disse script er Oracle script, og ligger under schema jspe.

I apk folderen ligger den klient installations file som feltdforsøget er udført med.

A3.Dataindsamlingszone

Kryds nummer	1
Kryds navn	Hotel hvide hus
prioritet	1
	Kong Christians Alle
	Vesterbro
	Østre Alle
	Hobrovej

Kryds nummer	2
Kryds navn	Dag Hammarskjølds syd
prioritet	1
	Østre Alle
	Dag Hammarskjølds Gade

Kryds nummer	3
Kryds navn	Sønderbro
prioritet	1
	Østre Alle
	Sønderbro

Kryds nummer	4
Kryds navn	Sohngårdsholm
prioritet	1

	Østre Alle
	Sohngårdsholmsvej

Kryds nummer	5
Kryds navn	Brandstationen
prioritet	2
	Jyllandsgade
	Kjellerupsgade
	Fyensgade
	Sønderbro
Kryds nummer	6
Kryds navn	Politigården
Prioritet	2
	Jyllandsgade
	Dag Hammarskjølds Gade

Kryds nummer	7
Kryds navn	Metropol
Prioritet	1
	Hasserisgade
	Vesterbro
	Prinsensgade

Kryds nummer	8
Kryds navn	Gåsepigen
Prioritet	2
	Vesterbro
	Vingårdsgade

Kryds nummer	9
Kryds navn	Broen
Prioritet	1
	Borgergade
	Vesterbro

Litteraturliste

[1] *Effective shadow detection in traffic monitoring applications*

Alessandro Bevilacqua

WSCG'2003, February 3-7, 2003, Plzen, Czech Republic

[2] *Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile centrury field experiment*

Juan C. Herrera, Daniel B. Work, Ryan Herring, Xuegang (Jeff) Ban and Alexandre M. Bayen.

UC Berkeley center for future urban transport 2009

[3] *Hierarchical traffic control and management with intelligent vehicles*

L.D. Baskar, B. De Schutter and H. Hellendoorn

IV'07, Istanbul, Turkey June 2007

[4] *Impacts of the Mobile internet on transportation cyberphysical systems: Traffic monitoring using smartphones*

Daniel B. Work and Alexandre M. Bayen

Washington DC. 2008

[5] *Intelligent traffic control system using RFID*

A, Chattaraj, S. Chakrabati, S. Bansal, S. Halder and A. Chandra

Device, Intelligent system and communication & networking 2008

[6] *Intelligent Vehicle control based on predictive urban traffic information*

Schuricht, Philipp and Backer, Bernard

Germany 2010

[7] *Floating car data for traffic monitoring*

Kristian Torp and Harry Lahrman

Aalborg, Denmark 2003

[8] *Location Privacy Techniques in Client-Server Architectures*

Christian S. Jensen, Hua Lu and Man Lung Yiu

2009

[9] *Metode til beregning af køretider, trængsel og forsinkelser I kryds vha. GPS data*

Kristian Torp, Harry Lahrman

Aalborg, Denmark 2009

[10] *Traffic control and intelligent vehicle highway systems: a survey*

L.D Basker B. De Schutter, J. Hellendoorn, Z. Papp

2009

[11] *Traffic Monitoring with TerraSAR-X*

h. Runge, S. Suchandt, A. Kotenkov, G. Palubinskas, U. Steinbrecher, D. Weiing
Germany 2008

[12] *Travel time collection and traffic monitoring via gps technologies*

Young-Ji Byon, Amer Shalaby and Baher Abdulhai

Intelligent Transportation Systems Conference, Toronto Canada, September 17-20, 2006

[13] <http://www.gartner.com/it/page.jsp?id=1622614>