KU LEUVEN

FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN

TECHNOLOGIECAMPUS GENT

Time Is Running Out

Assessing Temporal Privacy of Privacy Zones in Fitness Tracking Social Networks

Wout DELEU

Promotor: Prof. dr. ir. Stijn Volckaert

Begeleiders: Ing. Karel Dhondt,
Ing. Alicia Andries
Ing. Jonas Vinck

Masterproef ingediend tot het behalen van de graad van master of Science in de industriële wetenschappen: Elektronica/ICT Optie Smart Applications

Academiejaar 2022 - 2023

Voorwoord

Ik had graag eerst en vooral mijn ouders bedankt voor het financieren van mijn studies, en de ondersteuning gekregen in de periode. Daarnaast had ik graag Karel Dhondt, Stijn Volckaert, Alicia Andries en Jonas Vinck bedankt voor hun hulp en ondersteuning tijdens het schrijven van deze scriptie. Daarnaast in het bijzonder had ik ook graag Thomas Gruyaert bedankt, die tijdens het werken aan zijn eigen thesis ook een enorm grote hulp was. Als laatste had ik ook graag enkele van mijn kotgenoten bedankt voor de nodige afleiding tijdens de stressvolle perioden gedurende het academiejaar. In het bijzonder Angelo Pattyn en Jakob Sabbe, die zelf ook aan hun thesis werkten! Ook Sam Boeve, voor de erg hulpvolle adviezen gedurende het proces.

Samenvatting

In een maatschappij waar sociale media alom aanwezig is, zijn de privacybezorgdheden hierrond evenzeer erg actueel. Bij het ontwikkelingen van applicaties moeten privacywetgevingen en -bezorgdheden in acht genomen worden. Maar dit neemt niet weg dat in heel wat applicaties nog gaten te vinden zijn in het privacybeleid. In deze scriptie wordt de focus gelegd op het beleid binnen de fitness-trackers. Dit zijn platformen met als doel gegevens (die betrekking hebben op sportactiviteiten) op te slaan en weer te geven voor andere gebruikers. Dit zijn gegevens zoals hartslag, gps-locaties, ... Heel wat van deze platformen gaan op gelijkaardige manieren te werk om de privacy van de gebruiker proberen te garanderen. In de meeste gevallen gaat dit over het verbergen van een stuk van de activiteit, en zo de start- en/of eindpositie niet weer te geven op de kaart. Het achterhouden van dit routesegment gebeurd door het opstellen van een *Endpoint Privacy Zone*. Hierbij wordt een cirkel opgesteld waarbinnen de afgelegde weg wordt verborgen. In het verleden werd reeds aangetoond dat dit zeker geen waterdicht systeem is. Een tekortkoming van dit systeem, is het feit dat gebruik gemaakt wordt van gegeven afstanden die terug te vinden zijn in metadata die het desbetreffende platform vrijgeeft. Hiermee wordt de *Endpoint Privacy Zone* **EPZ** bepaalt, en zo kan uiteindelijk ook de effectieve startpositie van de activiteit achterhaald worden en zo deze gevoelige informatie bloot leggen [4].

Gedurende deze thesis wordt onderzoek gedaan naar manieren om deze EPZ te omzeilen, via metadata die terug te vinden is in de activiteiten in kwestie. De focus gedurende deze thesis ligt op snelheidsdata. Er wordt voornamelijk gezocht naar de omstandigheden waaronder deze aanval succesvol zou kunnen zijn, bij welke eigenschappen van de activiteit. En er wordt ook gezocht naar manieren om deze aanval in de best mogelijke omstandigheden uit te voeren. De studie focust zich voornamelijk op de berekeningen van afstanden en snelheden a.d.h.v. coördinaten. Het doel is om de EPZ te omzeilen, ervan uitgaande dat de desbetreffende fitnesstrackers afstandsinformatie achterhouden. Een grote focus ligt dus op het berekenen van afstanden via gps punten met een zo hoog mogelijke precisie. Een belangrijke techniek hierbij is gps-smoothing.

Keywords: fitness-trackers, privacy, endpoint privacy zone, gps-locaties, inference attack, snelheid

Abstract

Het extended abstract of de wetenschappelijke samenvatting wordt in het Engels geschreven en bevat **500 tot 1.500 woorden**. Dit abstract moet **niet** in KU Loket opgeladen worden (vanwege de beperkte beschikbare ruimte daar).

Keywords: fitness-trackers, privacy, endpoint privacy zone, gps-locations, inference attack, Speed

Inhoudsopgave

Voorwoord iii

Samenvatting iv

Abstract v

Inhoud vii

Figurenlijst viii

Tabellenlijst ix

1 Inleiding 1

1.1 Situering 1

1.2 Doelstelling 2

2 Achtergrond 4

2.1 Fitnesstrackers 4

2.1.1 Activiteiten 4

2.1.2 Berekening Afstanden 6

2.1.3 Algemeen Privacybeleid 9

2.2 Endpoint Privacy Zones 9

2.3 Gerelateerd werken 11

3 Setting aanval **13**

4 Resultaten **14**

5 Conclusies **16**

A Uitleg over de appendices **19**

Lijst van figuren

1.1 Voorbeeldactiviteit Strava 2

2.1 Data van een activiteit 7

2.2 Voorbeeld van de werking van een EPZ 8

2.3 Voorbeeld Data smoothing with moving average 9

2.4 Voorbeeld van de werking van een EPZ 10

2.5 Voorbeeld translatie EPZ 11

2.6 Voorbeeld filtering van punten binnen EPZ 11

2.7 Mechanisme EPZ beschreven door Wajih Ul Hassan 12

Lijst van tabellen

4.1 First run with standard gps points 14

4.2 Attack with first implementation of smoothening 14

4.3 Attack with smoothening window 10 14

4.4 Attack with smoothening window 5 14

4.5 Attack with smoothening window 15 14

4.6 Attack with smoothening window 20 14

4.7 Attack with startingpoints 15

4.8 Attack with n=25 15

4.9 Attack with n=50 15

4.10 Attack with n=25 15

4.11 Attack with n=100 15

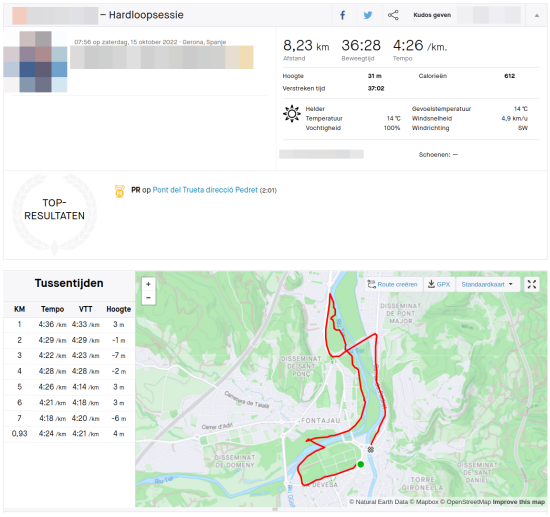
Hoofdstuk 1

Inleiding

1.1 Situering

Sociale media is zo goed als niet meer weg te denken uit het huidige moderne leven. Over de jaren heen zijn er verschillende definities gegeven. in In het werk van Howard en Park wordt sociale media gedefinieerd als de infrastructuur en tools om content te maken en te verspreiden[8]. Deze definitie is erg ruim, en vertakt zich dus in heel wat facetten, waaronder sociale netwerken, media sharing networks, etc. Maar ook de fitnesstrackers. Deze opkomst van nieuwe media brengt echter ook onbedoelde maar significante privacy bezorgdheden met zich mee.

De focus in deze dissertatie ligt op privacy binnen fitnesstrackers, meer specifiek platformen die gps-locaties gebruiken, zoals Strava, Nike Run Club, etc. Dit zijn platformen waar personen sportactiviteiten zoals lopen, fietsen, wandelen,... kunnen delen met elkaar. Het algemene concept is hierbij dat wanneer je een sportactiviteit uitvoert, je deze voor je volgers en vrienden beschikbaar maakt. De sportactiviteiten zullen natuurlijk bepaalde gegevens bevatten die zichtbaar zijn voor die andere gebruikers, Figuur 1.1 geeft bijvoorbeeld weer hoe Strava de afstand, bewegingstijd, en natuurlijk de gps-locaties deelt. Vele van deze gegevens hebben direct of indirect een negatieve impact op de privacy van de user. Deze negatieve gevolgen komen dan vooral in de vorm van het onbedoeld vrijgeven van *gevoelige locaties*. Onder het concept van een gevoelige locatie vallen heel wat beschrijvingen. Een algemene beschrijving kan zijn, een locatie die geografische informatie deelt die negatieve gevolgen kan hebben, en die je dus liever niet deel. In het kader van dit onderzoek, zal dit dan gaan over start en eindlocaties van activiteiten. Dit kan gaan over woonplaatsen, wat kan leiden tot o.a. stalking. Alsook locaties waar sportmateriaal wordt opgeborgen. Er zijn gevallen bekend van fietsdieven die Strava gebruiken om fietsen te kunnen lokaliseren[19][1]. Grootschaligere voorbeelden die zeker het vermelden waard zijn, zijn de gevallen waarbij geheime militaire basissen ontdekt worden door het bestuderen van de heatmap[7].



Figuur 1.1: Voorbeeldactiviteit Strava

Deze platformen implementeren elk manieren om de privacy van de users te verbeteren. De meest eenvoudige te bedenken is misschien wel de mogelijkheid om activiteiten te verbergen voor een selectie van personen (bv. iedereen die geen volger is). Zo kunnen enkel de mensen die de gebruiker expliciet toelaat activiteiten bekijken. Een complexer alternatief is het gebruik van *endpoint privacy zones* (EPZ). Hierbij wordt de weergegeven route voor de persoon die meekijkt gedeeltelijk verborgen. Er wordt als het ware een deel van de route afgekapt. De echte begin- en eindpunten zullen binnenin het afgekapte deel liggen. Er zullen nieuwe punten worden gegenereerd worden, op de rand van de cirkel, die voor de externe waarnemer het begin en einde zullen voorstellen. Het begin- en eind-deel van de route wordt dus onzichtbaar voor de andere gebruikers. Door de aanwezigheid van al deze pogingen tot privacyverbeteringen valt op dat de ontwikkelaars van de platformen erg bewust zijn van de mogelijke gevaren. Echter is er een afweging te maken bij de implementatie tussen de bruikbaarheid van het platform, en de privacy van de eindgebruiker. Hoe meer data wordt vrijgegeven, hoe groter de kans om mogelijk gevoelige info wordt meegegeven. Aan de andere kant, bij het weglaten van informatie gaat de gebruiksvriendelijkheid en de aanwezigheid van nuttige data van het platform serieus achteruit.

1.2 Doelstelling

In dit onderzoek bekijken we of er een mogelijkheid bestaat om private locaties (verborgen start- en eindlocaties) van een activiteiten te achterhalen, ondanks het gebruik van de EPZ 2.2 als privacy beveiligingsmechanisme. In het verleden werden enkele manieren beschreven om a.d.h.v. andere metadata zoals hoogtedata en afstanden de EPZ te omzeilen (*[4],[20]*). Gedurende deze thesis wordt meer in detail gegaan op het gebruik van snelheidsdata. Als basis voor deze aanval wordt de inferentie aanval op de EPZ van Dhondt et al. genomen. Er wordt dan onderzocht of deze

aanval nog steeds mogelijk is bij het weglaten van bepaalde gegevens, en dus door het gebruik van andere gegevens. De focus ligt in deze studie voornamelijk op snelheidsdata.

Om deze doelstelling te bekomen is eerst een berekeningsmechanisme nodig voor de afstanden die nodig zijn om de inferentie-aanval te kunnen uitvoeren. Daarna moet een analyse uitgevoerd worden tussen de berekende afstanden, en de waarden afgeleid volgens de berekeningen van Dhondt et al.. Zo kan de effectiviteit van de aanval a priori worden geschat. Er is een analyse van de beschikbare data, en een bespreking en reflectie over de resultaten van de aanval.

Hoofdstuk 2

Achtergrond

2.1 Fitnesstrackers

Zoals al enkele malen werd aangehaald, ligt de focus van deze scriptie op mogelijke tekortkomingen/vulnerabilities betreffende privacybeleid in fitnesstrackers. Maar voordat een aanval op basis van deze kwetsbaarheden kan opgezet worden, is het noodzakelijk om een te vat te krijgen op welke manier een fitnesstracker info verzamelt en weergeeft. En meer precies, hoe de mechanismen die de privacy voorzien voor de gebruikers in detail werken.

De data waarmee de aanval wordt opgezet en waarmee wordt geëxperimenteerd, is afkomstig van de populaire fitnesstracker *Strava*[1]. Dit is een sociaal netwerk, waarbij alle soorten sporters hun activiteiten kunnen delen. Dit gaat over lopen, wandelen, fietsen, zwemmen, . . . , maar ook sporten als fitnessen, voetballen, . . . De verzamelde data wordt volgens het perspectief van een mogelijke aanval gefilterd. Enkel data die gevoelige informatie met betrekking tot woonplaats zou kunnen vrijgeven wordt behouden. Dit zal er dus op neerkomen dat enkel activiteiten die relevante gps-informatie bevatten in beschouwing worden genomen. Dit gaat dan meer specifiek over *runs, hikes, walks, and rides*.

2.1.1 Activiteiten

Een Strava activiteit bevat erg veel informatie. Echter is niet alles even bruikbaar. Een correcte abstractie van de onnodige data is dus nodig. Figuur 2.1 geeft een voorbeeld van een gedetailleerde activiteit weer. Een gebruiker is in staat om de activiteit een titel te geven, en er een korte

[1] `https://www.strava.com/`

beschrijving aan toe te voegen. Ook een foto kan optioneel toegevoegd worden. De exacte datum en tijd van de start van de activiteit wordt hierbij ook weergegeven.

Rechts daarvan zijn de algemene basisstatistieken te zien. Deze zijn de totale afgelegde afstand, de totale bewegingstijd, de gemiddelde snelheid, het totale hoogteverschil, de totale verstreken tijden, en het aantal calorieën verbrand. Als extra kunnen hier enkele statistieken m.b.t. het gebruikte materiaal, zoals type fiets, loopschoenen, hartslagmeter, enzovoort worden weergegeven. Een belangrijk onderscheid in deze context is het verschil tussen de beweegtijd en de verstreken tijd. Deze twee lijken in definitie gelijk, maar dit zijn ze niet. Strava, en vaak fitnessplatformen in het algemeen werken met twee verschillende soorten tijdsberekeningen voor het bekomen van een accuratere gemiddelde snelheid. De verstreken tijd is simpelweg het tijdsinterval tussen het vertrek van de activiteit en de aankomsttijd ervan. De bewegingstijd is de tijd waarbij de gebruiker zich effectief bewoog. Met andere woorden worden de tijden waarbij de gebruiker stilstond uit de verstreken tijd gefilterd. Dit kan gaan over bijvoorbeeld een pauze, of het wachten voor een verkeerslicht. De snelheid wordt berekend aan de hand van de bewegingstijd. Dit kan simpel worden geverifieerd via een manuele berekening volgens de formule voor het berekenen van gemiddelde snelheid[2], met de data die terug te vinden is op Figuur 2.1 ($\frac{(39:17)\,\min}{7.44km} = 5:16\frac{\min}{km}$). Een kanttekening hierbij is dat dit enkel geldt voor activiteiten die niet gelabeld zijn als *race*, dan wordt de snelheid berekend in functie van de totaal verstreken tijd [15].

Onder de basisstatistieken zijn de *Strava-segmenten* te zien. Een Strava-segment is een specifiek deel van een bepaalde route dat door gebruikers van de sport-app kan worden gemarkeerd, gedeeld en vergeleken met andere gebruikers. Het segment is een bepaalde afstand en route, bijvoorbeeld een klim of afdaling, die vaak wordt beschouwd als een uitdagende of iconische sectie van een bepaalde fiets- of hardlooproute. Gebruikers van Strava kunnen een segment maken door de begin- en eindpunten op een kaart aan te geven en een naam en beschrijving toe te voegen. Zodra het segment is gemaakt, kunnen andere gebruikers het segment vinden en deelnemen aan een leaderboard, waarop de snelste tijden worden bijgehouden en vergeleken met andere gebruikers. Segmenten worden vaak gebruikt om prestaties te meten en te vergelijken.

Centraal op de figuur is ook de kaart duidelijk zichtbaar. Daarbij horen ook de tussentijden en de grafiek van snelheid. Optioneel kan hierbij ook nog een visualisatie van de afgelegde hoogte en de hartslag worden weergegeven, indien de gebruiker hiervoor met de juiste meetinstrumenten zijn sportactiviteit opneemt. De tussentijden en de grafiek van snelheid zijn qua inhoud gelijkaardig, met als verschil dat deze erg precies kan worden bestudeerd. Op de grafiek is voor elk afstandspunt de ogenblikkelijke snelheid zichtbaar. Bij de tussentijden wordt de gemiddelde snelheid over een kilometer weergegeven. De kaart die de route weergeeft is zeker ook belangrijk om even te bestuderen. Deze bevat namelijk alle gps-geregistreerde punten, en verbindt deze ook om zo één aaneensluitende route te vormen. Wanneer deze echter in detail bestudeerd wordt, samen met de legende die aanwezig is, is te zien dat de route uit twee delen bestaat, een zichtbaar deel en

[2] $v(\frac{min}{km}) = \frac{t(\min)}{d(km)}$

een onzichtbaar deel. Een andere gebruiker zal enkel zicht hebben op de het zichtbare deel, het onzichtbare deel zal dus voor een andere gebruiker niet zichtbaar zijn. Anders geformuleerd, de activiteit zal voor deze persoon dus als het ware afgekapt zijn, en zal in zijn zichtbare versie op een andere plek starten en eindigen. In de volgende Secties 2.1.3 & 2.2 wordt meer in detail ingegaan op de werking van deze methodiek.

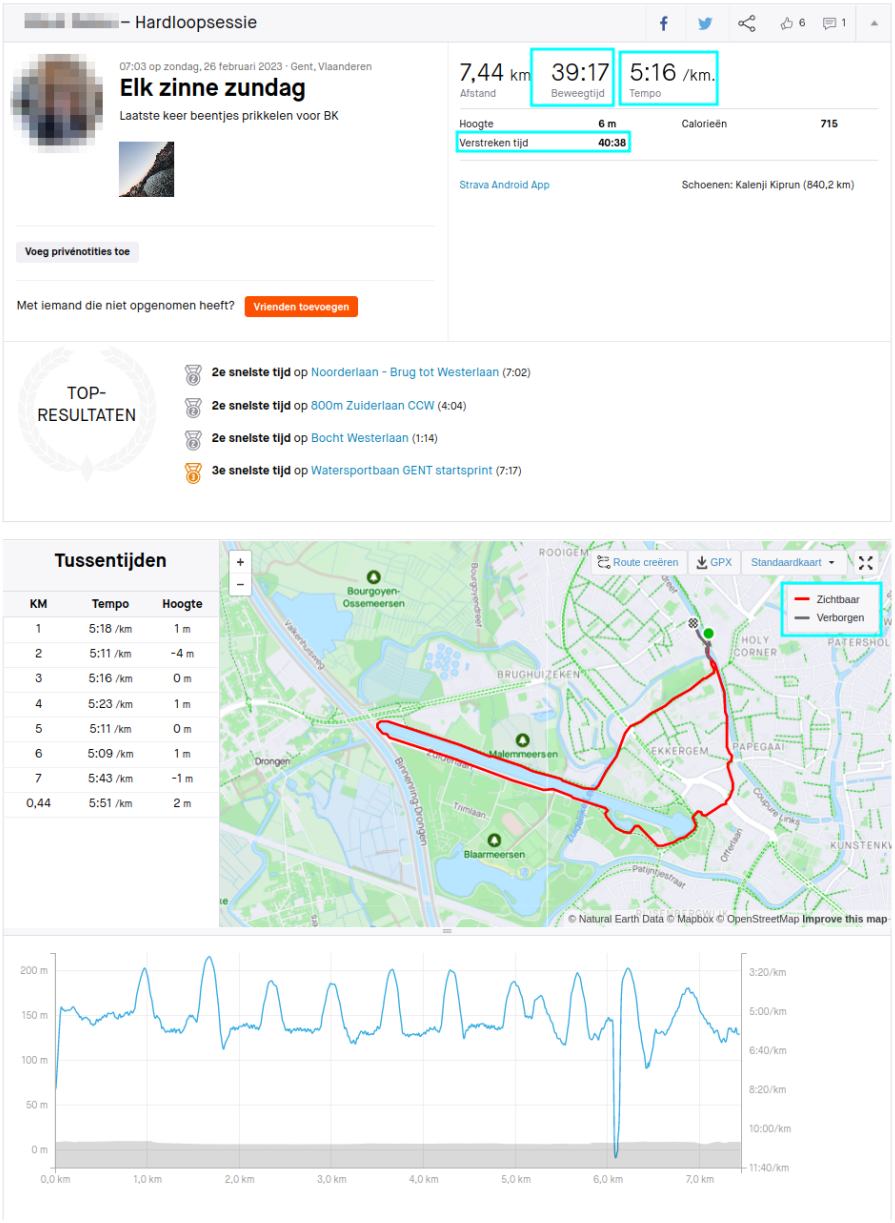
Een laatste kanttekening die hierbij gemaakt moet worden, is dat voor een gebruiker verschillende eenheden mogelijk zijn om uit te kiezen. Er is keuze mogelijk tussen de mijl en pond, en kilometer en kilogram. Gebruikers kiezen in welke eenheid ze de applicatie wensen te gebruiken. Voor de gebruiker in kwestie zal dus de volledige applicatie worden weergegeven in de gekozen eenheden.

2.1.2 Berekening Afstanden

Fitnesstrackers krijgen vanuit de buitenwereld ruwe data binnen. Deze data moet dus verwerkt worden vooraleer ze bruikbaar is voor de gebruiker. Er werd al kort ingegaan in Sectie 2.1.1 op de berekening die Strava gebruikt voor de snelheid. Echter is het ook interessant om de berekening van Strava eens onder de loep te nemen voor de afgelegde afstand. Strava maakt gebruik van twee verschillende methodieken voor het berekenen van deze afstand. De eerste is de *GPS-calculated Distance*. Dit bestaat eruit om de afstand tussen opeenvolgende gps-punten te berekenen, en deze op te tellen. Precisie is hier afhankelijk van de precisie van de gps-punten, aangezien de afstand wordt berekend door de punten met rechte lijnen te verbinden. Dit kan gebeuren in real time, via de gsm, smartwatch of ander toestel die gebruikt wordt om de activiteit op te nemen. Er zal dan ook mogelijkheid zijn om real time info te zien. Op elk punt zal de afstand vanaf het startpunt gekend zijn, en het is deze afstand die gedeeld zal worden op het platform. Het grote nadeel hierbij is het real-time aspect. Fouten kunnen moeilijker on the fly worden gecorrigeerd. Een tweede aanpak is om gps-data pas bij het uploaden te verwerken. De gps-data wordt dan geanalyseerd, en de nodige berekeningen worden uitgevoerd.

Het alternatief voor de GPS-calculated distance is de *Ground Speed Distance* methodiek. Deze afstand kan enkel worden bepaald in het geval van een fietsactiviteit. Deze afstand wordt berekend door het aantal omwentelingen te vermenigvuldigen met de omtrek van het fietswiel [18].

De bovenstaande afstandsberekeningen zijn de 2 technieken die officiële support documentatie van Strava beschrijft [18]. Echter blijkt wanneer de afstand op deze manier manueel berekent worden, afwijkende resultaten bekomen worden. Dit is zeer waarschijnlijk te wijten aan de pre-processing van de data die gebeurd bij het uploaden van een activiteit. Alhoewel dit niet expliciet gedocumenteerd staat doen de resultaten dit wel sterk vermoeden. De hypothese is dat tijdens het uploaden, de afstand herberekend wordt. De gps-punten zullen worden geanalyseerd, en er zullen



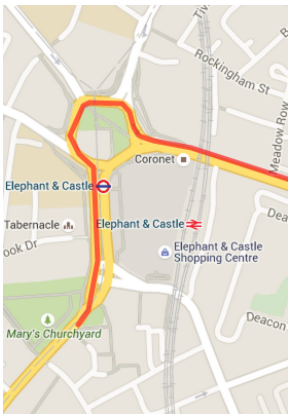
Figuur 2.1: Data van een activiteit

technieken worden gebruikt om de resultaten hiervan te verbeteren. De twee meest waarschijnlijke technieken zijn *Map Snapping* en *Smoothing*.

Map Snapping of Snap to Roads is een techniek waarbij gps-punten worden verschoven naar de dichtstbijzijnde weg. Per gps-punt wordt dan gezocht naar de dichtste node op de desbetreffende *roadgraph*[3] (op Figuur 2.2 is de werking ervan te zien)[3].

(a) Voorbeeldroute zonder map snapping

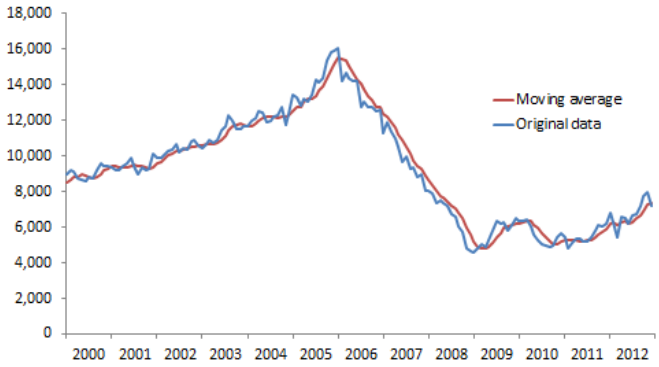
(b) Voorbeeldroute met map snapping



Figuur 2.2: Voorbeeld van de werking van een EPZ

Daarnaast bestaat de kans dat er gebruik gemaakt wordt van smoothening. Smoothening is een proces dat ruwe gps-punten (of datapunten in het algemeen) op een traject probeert te optimaliseren opdat ze een vloeiend 'curve' vormen. Dit wordt bekomen door ruis, schommelingen en onnauwkeurigheden te filteren uit het traject. Hiervoor bestaan verschillende implementaties. Aangezien Strava geen openbare informatie verstrekt over het gebruik van GPS-smoothing, is het niet bekend of ze deze techniek effectief toepassen. Het is dus gissen naar, indien ze deze zouden gebruiken, welke implementatie dan wel gebruikt wordt. De makkelijkste en meest modulaire methode om aan smoothening te doen, is *Smoothing met Moving Average*. Deze methode bestaat eruit om van een aantal punten in een bepaalde range (ook 'window' genoemd) het gemiddelde te nemen, en vervolgens op te schuiven. Het gemiddelde wordt berekend met volgende formule: $\overline{y_x} = \frac{y_x + y_{x+1} + ... + y_n}{x+n}$, voor punt x, met n als window-grootte. Zo kan voor elk punt een evenwichtige waarde op de nieuwe grafiek bekomen worden, en krijgt de grafiek een meer vloeiende vorm. Merk wel op dat de precisie van de route daalt wordt op deze manier. Bij het smoothen van een traject wordt het aantal gebruikte punten namelijk vermindert volgens de grote van de window. Afhankelijk van de grote, worden meer (resp. minder) punten samengenomen, en zo minder/meer punten weergegeven op de grafiek. Een voorbeeld is terug te vinden op figuur 2.3 [5][6][11].

[3]De roadgraph is afhankelijk van welke implementatie gebruikt wordt voor het snappen. Het is een wegennetwerk, omgezet in een graaf, bestaande uit edges en nodes. Elke weg of pad, bevat een of meerdere nodes, zodat een skeletstructuur ontstaat, die een abstractie van het wegennetwerk voorstelt [12].



Figuur 2.3: Voorbeeld Data smoothing with moving average

2.1.3 Algemeen Privacybeleid

Het delen van alle data die vervat zit in zo'n activitei met alle andere gebruikers op het platform, is zeker niet altijd wenselijk. De ontwikkelaars kiezen er dan ook voor om gebruikers de mogelijkheid te geven om hun privacy te bewaren. In deze sectie wordt de focus gelegd op de mechanismen gebruikt door *Strava*. Als opmerking valt te melden dat in heel wat andere sport-applicaties worden vergelijkbare, zo niet dezelfde methodieken gebruikt. Een eerste algemeen mechanisme bestaat eruit om de gebruiker de keuze te geven om alle activiteiten en alle gegevens over het profiel heen te laten voldoen aan bepaalde privacy regels. Deze regels kunnen ook per activiteit worden ingesteld. Onder de keuzes staan meestal drie opties, *zichtbaar voor iedereen*, *zichtbaar voor volgers* en *zichtbaar voor niemand*. Er kan ook zelf een keuze gemaakt worden om specifieke elementen van een activiteit niet te delen met de buitenwereld, zoals bijvoorbeeld de zichtbaarheid van de kaart die de route weergeeft.[16]

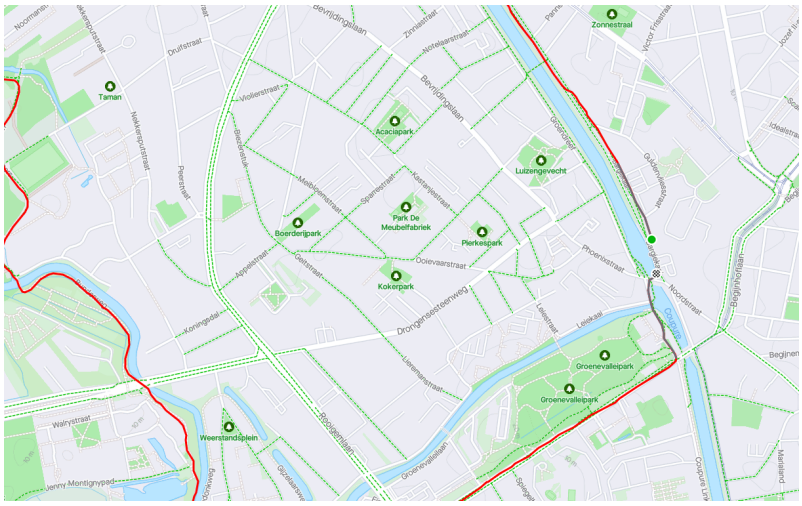
2.2 Endpoint Privacy Zones

Een tweede belangrijke maatregel is het gebruik van de de **Strava Endpoint Privacy Zones (EPZ)**. Een EPZ is een cirkelzone met een bepaalde straal rond een gps-punt. Het punt in kwestie zal dus de betreffende *gevoelige locatie* zijn. De straal van deze cirkel[4] kan worden gekozen door de gebruiker, en in het geval van Strava hebben gebruikers keuze uit waarden van 0 tot 1600m, in stappen van 200m. Wanneer een gebruiker binnen deze zone zijn activiteit beëindigt of begint, dan zal dat deel van de route binnen de EPZ niet zichtbaar zijn voor anderen. Vanuit het perspectief van een andere gebruiker zal de activiteit dus starten/eindigen op de rand van deze cirkel (die natuurlijk niet zichtbaar is). Merk op dat een sporter ook andere gevoelige locaties kan verbergen op de kaart. Bijvoorbeeld een frequent bezocht café, of een huis van een partner waar regelmatig een

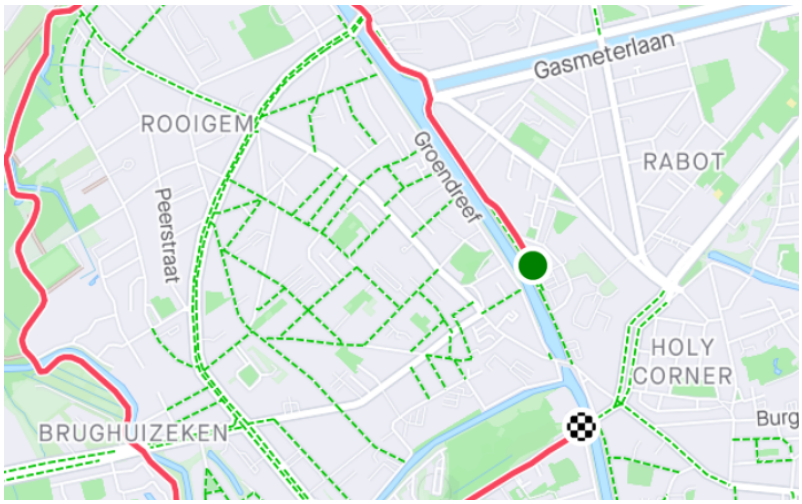
[4]Op Strava heeft de EPZ de vorm van een cirkel, maar op andere platformen kunnen andere vormen de norm zijn, bv. polygonen.

tussenstop plaatsvindt. Een tweede opmerking is dat wanneer een gebruiker de EPZ doorkruist, maar er niet in stopt, de route onaangepast blijft. Op Figuur 2.4 zijn de verschillende perspectieven te zien, hoe het er als uploader uitziet, en hoe het eruit ziet voor een andere gebruiker. Het traject die de buitenstaander te zien krijgt, zijn alle punten die zich buiten de EPZ bevinden. Merk ook op dat de eigenaar van de activiteit zicht heeft op de EPZ, en wat zal verborgen worden die zich buiten de EPZ bevinden. Dit onderscheid wordt gemaakt door het verschil in kleur (oranje voor de publiek zichtbare punten en grijs voor de onzichtbare).

(a) Perspectief eigenaar



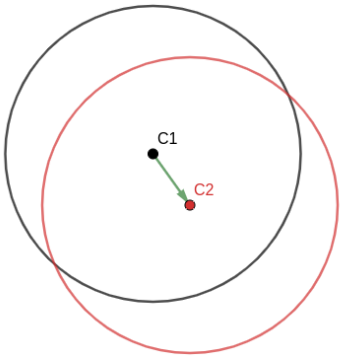
(b) Perspectief externe gebruiker



Figuur 2.4: Voorbeeld van de werking van een EPZ

Het opzetten van een EPZ is dus een belangrijk onderdeel bij het blootleggen van mogelijke zwakheden van dit systeem. Bij dit proces zal de gevoelige locatie worden genomen als beginlocatie. Hieruit zal a.d.h.v. de op voorhand vastgelegde EPZ-straal een cirkel worden opgesteld. Het centrum van deze cirkel zal hierna een translatie ondervinden in een willekeurige richting. Dit kan

een verschuiving zijn met een afstand die maximaal 70% van de straal van de EPZ bedraagt (Figuur 2.5). Het transleren van deze cirkel wordt ook *spatial cloaking* genoemd.



Figuur 2.5: Voorbeeld translatie EPZ

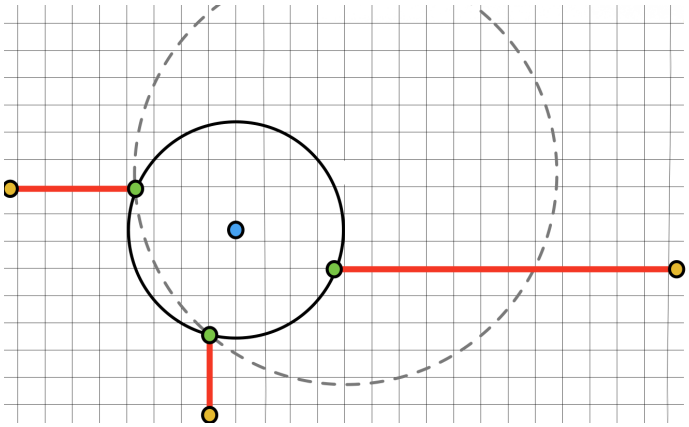
Daarna worden alle punten vertrekkende vanaf de gevoelige locatie tot aan de rand van de EPZ, en vanaf de rand van de EPZ tot aan de gevoelige locatie verwijdert van het zichtbare traject. Merk op dat punten die de EPZ doorkruisen, maar niet vertrekken/aankomen bij de gevoelige locatie niet worden gefilterd (Figuur 2.6).

Figuur 2.6: Voorbeeld filtering van punten binnen EPZ

2.3 Gerelateerd werken

In het verleden is al wat onderzoek verricht in de richting van de doeltreffendheid van EPZ's bij fitnesstrackers. Wajih Ul Hassan beschreef een implementatie van EPZ's waarbij het centrum van de zone de gevoelige locatie is. M.a.w. het identificeren van deze zone is dus voldoende om de gevoelige locatie te achterhalen[21]. In tegenstelling tot dit onderzoek, wordt ervan uitgegaan dat het centrum geen translatie ondervindt, en er dus geen spatial cloaking wordt toegepast. In deze paper

wordt gefocust op de reconstructie van de cirkel op basis van 3 punten op de rand (Figuur 2.7). Deze 3 randpunten worden dus bekomen door begin/eindpunten te nemen van activiteiten, volgens het perspectief van gebruiker die geen eigenaar is. Deze begin/eindpunten zullen zich altijd op de rand van de cirkel begeven. In deze paper wordt spatial cloaking wel aangehaald als mogelijke countermeasure tegen dit soort aanvallen.



Figuur 2.7: Mechanisme EPZ beschreven door Wajih Ul Hassan

Een onderzoek door Mink et al. toonde ook aan dat intuïtief heel wat mensen in staat zijn om de gevoelige locatie te achterhalen. Dit gebeurde op basis van enquêtes die werden afgenomen bij gebruikers van het platform. Uit het onderzoek bleek dat 68% van de ondervraagden bij een EPZ-radius van 200m de beschermde locatie tot op 50m nauwkeurig konden voorspellen. Deze resultaten op zich zijn alarmerend, en tonen aan dat EPZ's verre van perfect zijn.

Dhondt et al. voerde ook een studie naar de lekken aanwezig in het principe van EPZ's. Er wordt in deze paper een nadruk gelegd op de translatie van de EPZ, en hoe deze de privacy van een gebruiker verhoogt. Een inferentie aanval wordt er beschreven die gebruikmaakt van de totale afstand, terug te vinden bij de activiteit. Aan de hand van deze totale afstand in combinatie met het wegennetwerk, wordt een poging gedaan om alle mogelijke routes die de sporter binnenin de EPZ zou kunnen afgelegd hebben, voor elk traject te reconstrueren. Wanneer dit gedaan wordt voor verschillende trajecten, kan een locatie voorspeld worden die het meest waarschijnlijk wordt geacht om de gevoelige locatie te zijn.

Een laatste onderzoek die zeker ook het vermelden waard is, is de thesis van Verdonck et al. Deze thesis bouwt in grote mate verder op de paper van Dhondt et al., maar er wordt alternatieve data gebruikt. Er wordt gewerkt met hoogtedata i.p.v. totale afstanden, en zo wordt ook een inferentie aanval geconstrueerd.

Hoofdstuk 3

Setting aanval

Gedurende dit hoofdstuk wordt de bespreking

Hoofdstuk 4

Resultaten

Tabel 4.1 First run with standard gps points

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	72.06	59.92	21	81.89	473.05	2.22	1.01	33.43

Tabel 4.2 Attack with first implementation of smoothening

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	73.98	60.35	22	82.41	450.52	2.21	1.03	32.92

Tabel 4.3 Attack with smoothening window 10

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	70.59	69.52	22	81.14	480.38	2.2	1.06	33.34

Tabel 4.4 Attack with smoothening window 5

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	72.13	60.57	21	82.19	464.37	2.18	1.03	32.92

Tabel 4.5 Attack with smoothening window 15

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	71.67	61.49	22	82.75	480.13	2.17	1.01	32.96

Tabel 4.6 Attack with smoothening window 20

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	70.94	61.21	21	82.76	458.57	2.21	1.03	33.35

Tabel 4.7 Attack with startingpoints

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	81.43	35.96	15	86.01	322.32	1.91	0.68	28.33

Tabel 4.8 Attack with n=25

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	72.17	60.44	22	83.07	464.94	2.17	1.02	32.89

Tabel 4.9 Attack with n=50

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	72.12	60.67	20	82.4	451.9	2.15	1.03	32.28

Tabel 4.10 Attack with n=25

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	72.17	60.44	22	83.07	464.94	2.17	1.02	32.89

Tabel 4.11 Attack with n=100

Radius (m)	Success Rate (%)	Correctness (m)	Accuracy	Reduction (%)	Uncertainty Region (m^2)	Certainty	Spatial Certainty	Degree of Anonymity (%)
200	75.0	61.37	20	82.22	450.15	2.15	1.04	32.57

Hoofdstuk 5

Conclusies

Bibliografie

[1] Bowden, A. (2018). Cyclist who had five bikes stolen says thieves are looking for quick times on strava to try and find high-end bikes – warns other users to check their privacy settings — road.cc. https://road.cc/content/news/248798-cyclist-who-had-five-bikes-stolen-says-thieves-are-looking-quick-times-strava. (Accessed on 02/20/2023).

[2] Carr, C. T. and Hayes, R. A. (2015). Social media: Defining, developing, and divining. *Atlantic Journal of Communication*, 23(1):46–65.

[3] Croft, J. (2015). Snapping gps tracks to roads. https://www.jamesrcroft.com/2015/06/snapping-gps-tracks-to-roads/. (Accessed on 04/07/2023).

[4] Dhondt, K., Le Pochat, V., Voulimeneas, A., Joosen, W., and Volckaert, S. (2022). A run a day won't keep the hacker away: Inference attacks on endpoint privacy zones in fitness tracking social networks. osf.io/3m5ut.

[5] Early, J. (2020). Smoothing and interpolating noisy gps data. https://jeffreyearly.com/smoothing-and-interpolating-noisy-gps-data/. (Accessed on 04/07/2023).

[6] Early, J. J. and Sykulski, A. M. (2020). Smoothing and interpolating noisy gps data with smoothing splines. *Journal of Atmospheric and Oceanic Technology*, 37(3):449 – 465.

[7] Hern, A. (2018). Fitness tracking app strava gives away location of secret us army bases — gps — the guardian. https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases. (Accessed on 02/20/2023).

[8] Howard, P. and Parks, M. (2012). Social media and political change: Capacity, constraint, and consequence. *Journal of Communication*, 62.

[9] Ladetto, Q., Gabaglio, V., and Merminod, B. (2001). Combining gyroscopes, magnetic compass and gps for pedestrian navigation. *Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics, and Navigation*.

[10] Mink, J., Yuile, A. R., Pal, U., Aviv, A. J., and Bates, A. (2022). Users can deduce sensitive locations protected by privacy zones on fitness tracking apps. In *Proceedings of the 2022 CHI*

Conference on Human Factors in Computing Systems, CHI '22, New York, NY, USA. Association for Computing Machinery.

[11] of Dallas, F. R. B. (n.d.). Smoothing data with moving averages - dallasfed.org. https://www.dallasfed.org/research/basics/moving#:~:text=A%20moving%20average%20smoothes%20a,the%20variable's%20timeliness%20is%20lost. (Accessed on 04/13/2023).

[12] Seiler, K. M. (2022). Haul road mapping from gps traces.

[13] Strava, I. (2021a). Strava-privacybeleid. https://www.strava.com/legal/privacy. (Accessed on 02/20/2023).

[14] Strava, I. (2021b). Strava's year in sport 2021 charts trajectory of ongoing sports boom. https://blog.strava.com/nl/press/yis2021/. (Accessed on 02/26/2023).

[15] Strava, I. (2022). Moving time, speed, and pace calculations – strava support. https://support.strava.com/hc/en-us/articles/115001188684-Moving-Time-Speed-and-Pace-Calculations. (Accessed on 02/26/2023).

[16] Strava, I. (2023a). Activity privacy controls – strava support. https://support.strava.com/hc/en-us/articles/216919377-Activity-Privacy-Controls. (Accessed on 02/27/2023).

[17] Strava, I. (2023b). Bad gps data – strava support. https://support.strava.com/hc/en-us/articles/216917707-Bad-GPS-Data. (Accessed on 03/01/2023).

[18] Strava, I. (2023c). How distance is calculated – strava support. https://support.strava.com/hc/en-us/articles/216919487-How-Distance-is-Calculated. (Accessed on 03/01/2023).

[19] Vanmeldert, D. (2022). Sportapp strava laat fietsdieven of stalkers nog altijd meekijken — vrt nws: nieuws. https://www.vrt.be/vrtnws/nl/2022/10/28/strava-kul/. (Accessed on 02/20/2023).

[20] Verdonck, T. (2022). Inferentie-aanvallen met hoogteprofielen tegen (endpoint) privacy zones in fitness tracking sociale netwerken. Master's thesis, KU Leuven. Faculteit Industriële Ingenieurswetenschappen, Leuven. Book Title: Inferentie-aanvallen met hoogteprofielen tegen (endpoint) privacy zones in fitness tracking sociale netwerken.

[21] Wajih Ul Hassan, Saad Hussain, A. B. (2018). Analysis of privacy protections in fitness tracking social networks -or- you can run, but can you hide?

Bijlage A

Uitleg over de appendices

Bijlagen worden bij voorkeur enkel elektronisch ter beschikking gesteld. Indien essentieel kunnen in overleg met de promotor bijlagen in de scriptie opgenomen worden of als apart boekdeel voorzien worden.

Er wordt wel steeds een lijst met vermelding van alle bijlagen opgenomen in de scriptie. Bijlagen worden genummerd het een drukletter A, B, C,...

Voorbeelden van bijlagen:
Bijlage A: Detailtekeningen van de proefopstelling
Bijlage B: Meetgegevens (op USB)

FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN
TECHNOLOGIECAMPUS GENT
Gebroeders De Smetstraat 1
8200 GENT, België
tel. + 32 50 66 48 00
iiw.gent@kuleuven.be
www.iiw.kuleuven.be