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. ing. 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.

Samenvatting

0.1 Situering

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. Gedurende deze thesis worden gekende fitness-trackers onder de loep genomen, waaronder Strava. Er is op te merken dat heel wat van deze platformen op gelijkaardige manieren proberen privacy 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 verbergen van deze activiteit gebeurd door het opstellen van een *Endpoint Privacy Zone*. Hierbij wordt een cirkel opgesteld waarbinnen de afgelegde weg zal worden verborgen. In de paper van Dhondt et al. is een manier terug te vinden om door het combineren van verschillende gegeven afstanden (bijvoorbeeld totale afstand, afstand tussen 2 punten, . . .) en het bepalen van deze EPZ, de effectieve startpositie van de activiteit te achterhalen en zo gevoelige informatie bloot te leggen.

0.2 Doel

In deze thesis zal getracht worden om alternatieve manieren te vinden om deze gevoelige locatie te bepalen aan de hand van andere metadata zoals snelheid. Hierbij zal ook de effectiviteit ervan geanalyseerd worden, en zal getracht worden om enkele manieren te vinden om de privacy van het platform te verhogen. Er zal getracht worden deze werkwijze toe te passen aan de hand van verscheidene beschikbare metadata.

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: Voeg een vijftal keywords in (bv: Latex-template, thesis, ...)

Inhoudsopgave

Lijst van figuren

Lijst van tabellen

Hoofdstuk 1

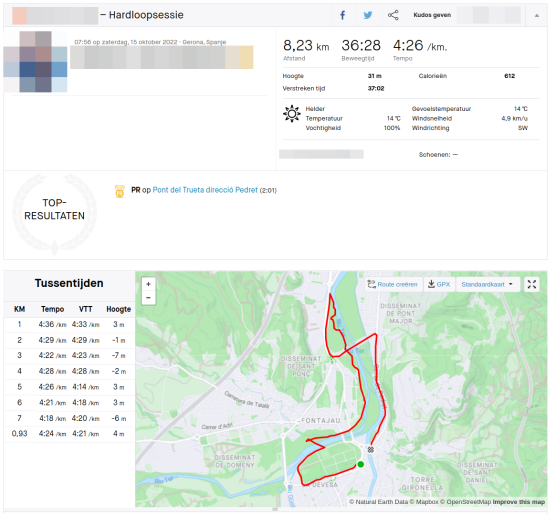
Inleiding

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 dan ook verschillende definities gegeven. Howard and Parks definiëren sociale media als de infrastructuur en tools om content te maken en te verspreiden[5]. Deze definitie is erg ruim, en vertakt zich dus in heel wat facetten, waaronder sociale netwerken, media sharing networks,... Maar ook de fitnesstrackers. Deze opkomst van nieuwe media brengen echter ook vaak 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 sportactiviteiten zoals lopen, fietsen, wandelen,... kunnen worden gedeeld met andere personen. Het algemene concept is hierbij dat wanneer je een sportactiviteit uitvoert, je deze voor je volgers en vrienden beschikbaar maakt. De sportactiviteit zal dan natuurlijk ook bepaalde gegevens bevatten die zichtbaar zijn voor die volgers, zoals tijdstippen, hartslagen, bewegingstijd, en vaak ook gps-locaties 1.1. 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 gevoelige locaties. 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[9][1]. Grootschaligere voorbeelden die zeker het vermelden waard zijn de gevallen waarbij geheime militaire basissen ontdekt worden door het bestuderen van de heatmap.

Deze platformen implementeren elk manieren om de privacy van de users te verbeteren. Hiervoor zijn verschillende manieren mogelijk. De simpelste is misschien wel de mogelijkheid om activiteiten



Figuur 1.1: Voorbeeldactiviteit Strava

te verbergen voor anderen. Zo kunnen enkel de mensen die de gebruiker expliciet toelaat activiteiten bekijken. Een complexere 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, en het eind- en startpunt worden verschoven. Het begin -en eind-deel van de route wordt dus onzichtbaar enkel voor de andere gebruikers. Het valt op dat deze profielen erg bewust zijn van de mogelijke gevaren van hun platformen. Echter is er een afweging te maken bij de implementatie tussen de bruikbaarheid van het platform, en de privacy van de eindgebruiker. Hoe meer info wordt vrijgegeven, hoe groter de kans om mogelijk schadelijke private info wordt meegegeven. Aan de andere kant, bij het weglaten van informatie gaat de gebruiksvriendelijkheid en de aanwezigheid van nuttige info van het platform serieus achteruit gaat.

1.2 Doelstelling

Het doel van deze scriptie is om private locaties (verborgen start- en eindlocaties) van een activiteiten te achterhalen, ondanks het gebruik van de EPZ 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 (*Dhondt et al., Verdonck*). Hier 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 succesvol kan worden uitgevoerd bij het weglaten van gegevens, en dus door het gebruik van andere gegevens, voornamelijk snelheid.

Om deze doelstelling te bekomen zal eerst een analyse op de afwijkingen van tussen de berekende afstanden nodig om de inferentie aanval uit te voeren, en de waarden afgeleid volgens de

berekeningen van Dhondt et al.. Er zal een analyse gebeuren over de alle users. Er zal ook een studie gebeuren van de effectiviteit van deze aanval op basis van de nieuw bekomen afstanden.

De doelstellingen zijn in eerste instantie vooral vanuit het oogpunt van een aanvaller, een user met slechte bedoelingen. Echter zal als laatste ook gereflecteerd worden over de gevallen waarin de aanval effectief blijft, bijvoorbeeld onder welke afrondingen. Bijgevolg volgt hieruit enkele mogelijke countermeasures, die dus zullen aangeven welke manieren fitnesstrackers zichzelf zouden kunnen toepassen om zichzelf te wapenen tegen de aanval die hier beschreven werd.

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 kan worden opgezet, 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. Dit is essentiële informatie om een aanval te kunnen opzetten.

De gebruikte data waarop de aanval wordt opgezet en waar op wordt geëxperimenteerd, is afkomstig van de populaire fitnesstracker *Strava*[1]. Dit is een sociaal netwerk, waarbij een alle soorten sporters hun activiteiten kunnen delen. Dit gaat over lopen, wandelen, fietsen, zwemmen, ..., maar ook sporten als fitnessen, voetballen, ... (Deze laatste zijn wel veruit in de minderheid bij de overschouwing van de totale verdeling van alle activiteiten[7]). De betreffende data wordt in perspectief van een mogelijke aanval gefilterd, opdat 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 zal dus moeten gemaakt worden van de onnodige data. Afbeelding 2.1 geeft een voorbeeld van een gedetailleerde activiteit weer. Een gebruiker is in staat om de activiteit een titel te

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

geven, en er een korte 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 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 2 lijken in definitie gelijk, maar dit zijn ze niet! Strava werkt met 2 verschillende soorten tijdsberekeningen voor een iets accuratere snelheidsberekening. 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. M.a.w. worden de tijden waarbij de gebruiker stilstond uit de verstreken tijd gefilterd. Dit kan gaan over bijvoorbeeld een pauze, of een verkeerslicht. De snelheid wordt berekend aan de hand van de bewegingstijd. Dit kan simpel worden geverifieerd worden via een manuele berekening ter bevestiging(2.1). Een kanttekening hierbij is dat dit enkel geldt voor activiteiten die niet gelabeld zijn als *race*. In dit geval wordt de snelheid berekend in functie van de totaal verstreken tijd.[8]

$$\frac{(39:17)\,\mathrm{min}}{7.44km} = 5:16\,\mathrm{min} \tag{2.1}$$

Daaronder zien we de *Strava-segmenten*. 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.

Daarnaast is ook de kaart duidelijk zichtbaar. Daarbij horen ook de tussentijden en de grafiek van snelheid. Optioneel kan hierbij ook nog een grafiek van de afgelegde hoogte en de hartslag worden weergegeven, indien de gebruiker hiervoor met het juiste materieel zijn sportactiviteit opneemt. De tussentijden en de grafiek van snelheid zijn qua inhoud gelijkaardig, met als verschil dat de grafiek erg precies kan worden bestudeerd worden. Op de grafiek is voor elk afstandspunt de ogenblikkelijke snelheid zichtbaar. Bij de tussentijden is 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 een onzichtbaar deel. Dit heeft betrekking tot wat zichtbaar is voor een andere gebruiker die deze

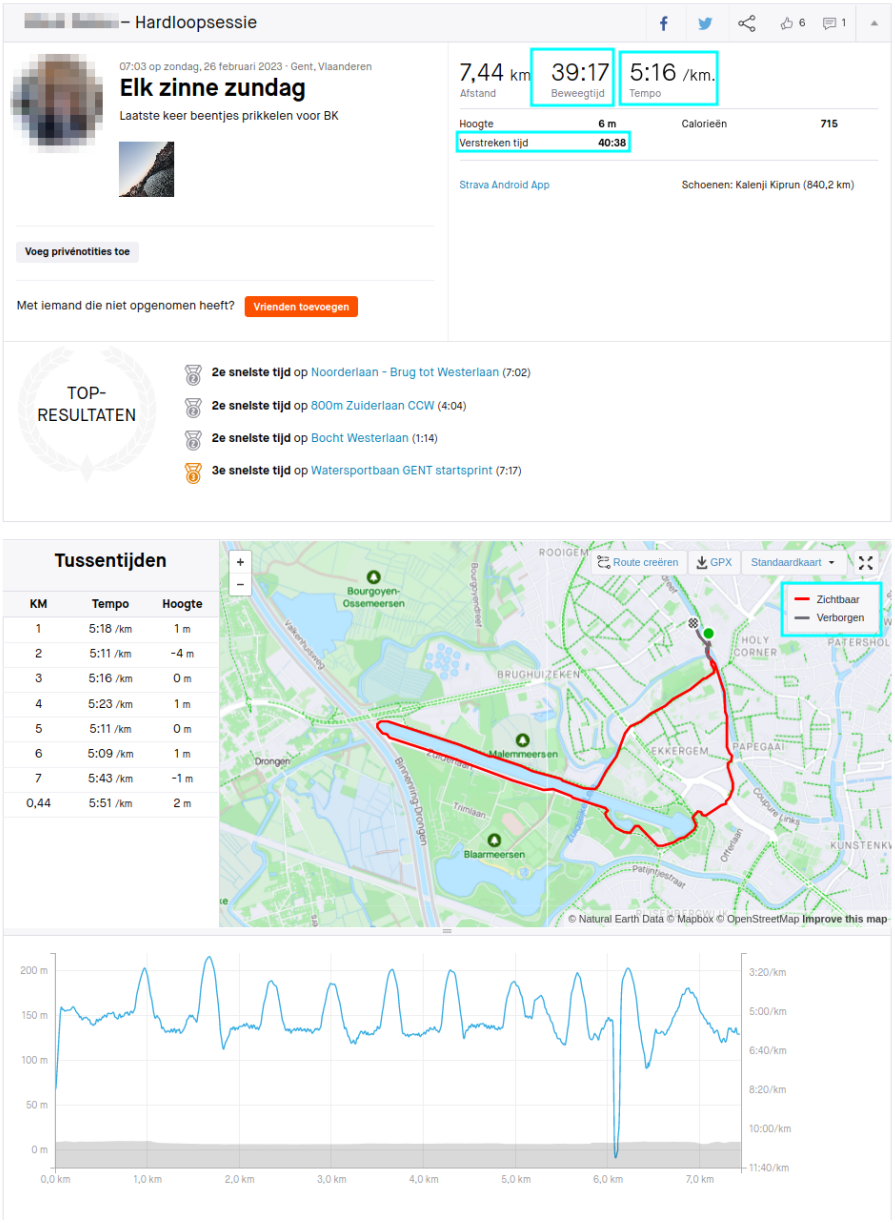
activiteit bekijkt. Deze gebruiker zal enkel zicht hebben tot de het zichtbare deel, het onzichtbare deel zal dus voor een andere gebruiker niet zichtbaar zijn. De activiteit zal voor deze persoon dus als het ware afgekapt zijn, en zal in zijn zichtbare versie op een andere plek starten en eindigen. In de volgende sectie 2.1.2 & 2.1.3 wordt op de betekenis 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. Een gebruiker kiest in welke eenheid hij/zij de applicatie wenst te gebruiken. Voor de gebruiker in kwestie zal dus alles worden weergegeven in de gekozen eenheid.

2.1.2 Algemene Privacy

Het delen van alle data die vervat zit in zo'n een activiteit zit met alle andere gebruikers op het platform, is zeker niet altijd wenselijk. Kiezen er dan ook voor om de gebruiker de mogelijkheid te geven om zijn privacy te verbeteren. In deze sectie wordt de focus gelegd op de mechanismen in gebruikt in *Strava*, maar in heel wat andere sport-applicaties worden vergelijkbare, zo niet dezelfde methodieken gebruikt.

2.1.3 Endpoint Privacy Zones



Figuur 2.1: Data van een activiteit

Hoofdstuk 3

Setting aanval

Hoofdstuk 4

Richtlijnen voor formules

Er zijn twee manieren om formules in LaTeX in te voeren:

- Inline: $a^2+b^2=c^2$ (`$a^2+b^2 = c^2$`)
- In een equation omgeving (`\begin{equation} a^2+b^2 = c^2 \end{equation}`):

$$a^2+b^2=c^2 \tag{4.1}$$

Griekse letters geef je in d.m.b. het backslash commando. Bijvoorbeeld de letter sigma σ verkrijg je door `$\sigma$` inline in te geven. Dit is analoog voor griekse letters in de equation omgeving. Een beknopte lijst van symbolen vind je op de Wikibooks pagina voor LaTeX (link). Alle andere nuttige informatie omtrent het gebruik van LaTeX voor formules vind je hier ook terug.

Hoofdstuk 5

Richtlijnen voor referenties

5.1 Inleiding

De referentielijst bevat de volledige lijst van literatuur en bronnen waarnaar in de tekst wordt verwezen. Door systematisch de referentielijst aan te vullen bij het schrijven van het literatuuroverzicht gaat er achteraf geen tijd verloren aan het opnieuw opzoeken van referenties.

5.2 Referentiestijl

Voor het verwijzen naar informatiebronnen wordt gebruik gemaakt van het numerisch systeem of van het auteur-jaar systeem. Dit kies je door volgend commando in het latex bronbestand aan te passen:

- numerisch (IEEE) : `\bibliographystyle{ieee}`
- alfabetisch (APA) : `\bibliographystyle{apalike}`

Plaats je bronnen in een *bibtex* bestand (evt. via software zoals bv. Jabref Endnote of Mendeley), waarnaar je verwijst vanuit je thesis text a.d.h.v. het commando `\cite`. Enkele links naar nuttige software in deze context:

- JabRef (Open Source)

- Mendeley (Freeware)
- EndNote (Paid license)

Indien je zelf een .bibtex bestand wil aanleggen dien je volgende syntax te volgen voor een tijdschriftartikel:

```
@article{hughes2005,
title={Isogeometric analysis: CAD, finite elements, NURBS, exact geometry
and mesh refinement},
author={Hughes, Thomas JR and Cottrell, John A and Bazilevs, Yuri},
journal={Computer methods in applied mechanics and engineering},
volume={194},
number={39},
pages={4135--4195},
year={2005},
publisher={Elsevier}
}
```

Enkele voorbeelden van het gebruik van bronnen in een tekst (in APA stijl):

Recent werd het Higgs boson experimenteel vastgesteld door Aad et al. [**?**] (syntax: `\cite{aad2012}`).

Als alternatief voor het discretiseren van een CAD model vooraleer een eindige elementenanalyse te kunnen toepassen, stellen Hughes et al. voor om de nodige elementenformulering rechtstreeks uit de NURBS beschrijving van de CAD geometrie te halen [**?**] (syntax: `\cite{hughes2005}`). Daarnaast introduceren ze tevens een k-iteratieve procedure als een verfijning van de geldende p- en h-iteratieve procedures in eindige elementen methoden [**?**] (syntax: `\cite{cottrell2009}`).

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] Dhondt, K., Pochat, V. L., 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.

[4] 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).

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

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

[7] 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).

[8] 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).

[9] 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).

[10] 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.

[11] 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