KØBENHAVNS UNIVERSITET

Veje og vildveje til øget monitorering af det mobile datamarked

Rapport til Dataetisk Råd leveret af Kristian Sick Svendsen, Sofie Flensburg og Stine Lomborg, Center for Tracking & Society, KU, Januar 2023

Vi bruger vores mobil hele tiden og til alt muligt – til at finde vej, koordinere aftaler med familie og venner, indkøb, underholdning og informationssøgning, arbejde på farten, og håndtere kontakten med det offentlige. Mange danskere vil måske ligefrem sige at mobilen er en nødvendighed for at få hverdagens ender til at mødes. Det betyder også at mobilen er en guldgrube af data – for os selv og for andre. Disse andre inkluderer appudbyderne selv og de tredjeparter, der leverer software-baserede services til os alle. Men vi ved ganske lidt om, hvordan personlige data fra vores mobilbrug rejser rundt i det mobile økosystem, og om de markedsstrukturer, der omgiver og påvirker vores stadig mere dataficerede liv.

I samarbejde med Dataetisk Råd har Center for Tracking & Society ved Københavns Universitet i efteråret 2022 afprøvet og afklaret relevante metoder til at kortlægge og monitorere brugen af tredjepartstjenester i apps i Danmark med henblik på at skabe mere transparens omkring mobile datastrømme. Fokus har været på at afsøge potentialer og begrænsninger ved forskellige tekniske metoder til at analysere dataflows i app-økosystemet. Arbejdet lægger vigtige brikker i et fundament for en mere langsigtet indsats, der kan styrke indsigt i, oplysning om og reguleringen af mobile datastrømme. Den viden, samarbejdet har tilvejebragt om metodiske strategier til at få indsigt i mobil tracking, kan dermed være et vigtigt skridt henimod øget transparens omkring mobile apps som en del af en stadig mere samfundskritisk datainfrastruktur, borgeren møder i hverdagen.

Hvorfor er det vigtigt at vide noget om mobil tracking og tredjepartstjenester?
Vi ved fra den eksisterende forskning, at de fleste mennesker synes, det er ret abstrakt, ja det kan ligefrem være en byrde at skulle bekymre sig om, hvem der indsamler og bruger ens data. Vi ved også, at der i et velfærdssamfund som det danske er en udbredt forventning om, at andre – politikere, myndigheder og de organisationer, der bruger data – skal dæmme op for de negative konsekvenser af dataindsamling og sikre ansvarlig dataanvendelse. Hvis vi som samfund skal løse den opgave, kræver det viden om de teknologiske og økonomiske vilkår, der rammesætter udbuddet og brugen af mobile apps og disses behandling af brugerdata. Mere specifikt har vi brug for 1) at afdække, hvad der er teknisk og juridisk muligt, ift. at indsamle og formidle informationer om mobil dataindsamling og 2) at udvikle redskaber til systematisk at undersøge, monitorere og ultimativt regulere de virksomheder, der lever af at indsamle, behandle og distribuere data. En øget indsigt i det mobile datamarked og de infrastrukturer, det er baseret på, vil endvidere skabe et stærkere fundament for at

håndhæve GDPR – også hvad angår tredjepartstjenester, der befinder sig langt ude i datakæden.

Hvad er digital tracking?
Indsamling og analyse af brugerdata er i dag en indgroet del af stort set alle former for digital kommunikation, og danskerne er i årtier blevet bedt om at godkende (eller fravælge) såkaldte 'cookies', når de logger ind på en hjemmeside. Disse cookies har været hjørnesten i tredjepartsbaseret tracking af vores digitale adfærd, når den er foregået via en webbrowser som eksempelvis Chrome eller Safari. Cookies er en teknologi, der tillader virksomheder og andre aktører løbende at indsamle informationer om eksempelvis den enkelte internetbrugers søgehistorik og andre former for brugsmønstre. I de senere år er der i imidlertid blevet udviklet en stribe nye teknologier til at indsamle data, og særligt udbredelsen af smartphones og den deraf følgende vækst i brugen af mobile apps har skabt nye muligheder og vilkår for dataindsamling særligt igennem mobile tredjepartstjenester.

Hvad er mobile tredjepartstjenester?
Når vi bruger mobile apps, indsamles ofte store mængder af brugerdata typisk ved hjælp af såkaldte Software Development Kits (SDK), der kan beskrives som pakker af kode, der installeres i appen og understøtter forskellige (tekniske såvel som kommercielle) funktioner. Disse pakker gør det muligt for appudviklere at udvikle applikationer til en bestemt platform som eksempelvis Android eller iOS og lancere en tjeneste uden at skulle 'bygge' hele infrastrukturen fra bunden. Mange SDK'er tilbyder eksempelvis færdige løsninger til at udføre brugeranalyse, identificere fejl i appen igennem logging, tilpasse indholdskvaliteten afhængigt af netværksforbindelsen, præsentere relevante annoncer, implementere sociale medie-funktioner og meget mere. SDK'er er ofte udbudt af eksterne virksomheder, hvilket konkret betyder, at brugernes data ikke kun indsamles og behandles af appudbyderne, men at der også kan forekomme en dataleverance af brugerdata mellem en applikation og de tredjeparter, der hjælper appen med både at fungere og tjene penge (f.eks. via videresalg af data til målrettet annoncering). Dette gør SDK'er særligt interessante i diskussionen om hvordan services og brugernes data udveksles på det mobile datamarked.

Analysetyper og deres fordele og begrænsninger
Sammenlignet med cookiebaseret tracking via hjemmesider, har mobil dataindsamling via SDK'er været genstand for begrænset forskningsmæssig såvel som politisk og reguleringsmæssig opmærksomhed. I vores arbejde med at afdække metoder til at skabe mere transparens og systematisk viden om det mobile datamarked har vi primært trukket på analyser og metoder udviklet på Oxford University (Kollnig et al. 2022), tidligere forskningsaktiviteter i Center for Tracking & Society på Københavns Universitet, samt på vidensdeling i relevante dataaktivistiske grupper på blandt andet Reddit, Discord og GitHub.

Analyser af mobil dataindsamling kan groft inddeles i to overordnede kategorier: 1) studier, der fokuserer på hvilke *typer af data,* der indsamles (fx når en app beder om tilladelse til at få adgang til en brugers kamera, kontakter, webhistorik, beskeder eller andet); og 2) studier, der fokuserer på hvilke *markedsaktører*, der er involveret i indsamlingen og distributionen af data. Førstnævnte kan set give en indsigt i, hvilken viden enkelte apps må formodes at ligge inde med om sine brugere, men er begrænset i forhold til at forstå, hvordan og hvorvidt disse data deles med andre aktører. Studier, der fokuserer på tredjepartstjenester, giver derimod

på nuværende tidspunkt begrænset viden om, hvilke informationer der indsamles, og hvad de bruges til, men kan derimod give indsigt i de kommercielle interesser og magtforhold, der påvirker mobil dataindsamling og selve udviklingen af de mobile infrastrukturer. En øget forståelse af tredjepartstjenesternes aktiviteter og forretningsmodeller vil desuden på længere sigt kunne skabe en kobling mellem de to perspektiver, idet brugen af specifikke SDK'er ofte vil medføre, at app'en skal have adgang til data, som ikke er åbenlyst relateret til appens virke. Dette afspejles eksempelvis i hvilke tilladelser appen kræver af brugeren for at kunne indsamle lokationsdata, tilgå kontakter eller læse beskeder.

I denne afrapportering fokuserer vi derfor på de metodiske muligheder og begrænsninger inden for analyser af mobile tredjepartstjenester. Inden for denne gren af den mobile dataforskning findes yderligere to forgreninger: 1) *dynamiske* studier, der analyserer netværkstrafik fra en konkret enhed (typisk en mobiltelefon) og, ved at lede netværkstrafikken igennem en proxy, identificerer de tredjepartsdomæner, som denne kalder op til; og 2) *statiske* analyser, der gennemgår dele af programkoden i specifikke apps med henblik på at identificere installerede SDK'er. Dynamiske analyser kan bruges til at kortlægge den enkelte mobiltelefons faktiske datastrømme og identificere, hvilke aktører der kaldes op til (og dermed kan sendes data til), men er vanskelige at generalisere ud over den specifikke kontekst. Statiske analyser er derimod mere velegnede til at give en generel viden om de mobile datainfrastrukturer og egner sig dermed bedre til bredere analyser af mobile datamarkeder. Dette projekt har derfor primært fokuseret på at udforske mulighederne for at foretage statiske analyser, dog med det væsentlige forbehold, at statiske analyser, som det uddybes nedenfor, er afhængige af eksisterende databaser over tredjepartstjenester, og må derfor anses som mindre fuldstændige end de dynamiske.

Fælles for bestræbelserne på at udvikle valide analysemetoder til undersøgelser af mobil dataindsamling er, at de kompliceres af, at de mobile infrastrukturer varierer markant afhængigt styresystemer og app stores. Der er således stor forskel på, om man har en iPhone og henter sine apps via Apple's App Store, eller om man bruger en smartphone med Androids styresystem og downloader apps via Google Play Store. Tillige varierer mulighederne for at indsamle viden om den dataindsamling, der foregår, alt efter hvilket økosystem, man fokuserer på. Endelig forandrer vilkårene for at analysere disse økosystemer sig løbende sig i takt med, at systemerne opdateres, nye versioner lanceres, virksomhedernes politikker og praksisser ændrer sig med mere. For at skabe et mere eksakt billede af metodernes nøjagtighed og potentialer for skalering, er der derfor brug for løbende efterprøvning og konsolidering af dem.

Statiske analyser af tredjepartstjenester i Android og iOS apps

I samarbejdet med Dataetisk Råd har vi anlagt et generelt perspektiv og fokuseret på mulighederne for at indhente sammenlignelige informationer om tilstedeværelsen af tredjepartstjenester i et sample af Android og iOS apps. Android apps er som udgangspunkt lettere tilgængelige for analyse på grund af styresystemets open source-arkitektur samt det grundlæggende design af programmeringssproget, der også betyder, at den har været genstand for flere tidligere analyser. For eksempel giver den fransk-udviklede Exodus Privacy database mulighed for at slå Android apps op i en online tjeneste og identificere indlejrede tredjepartstjenester. Apple's iOS styresystem, og apps som er udviklet til platformen er sværere at få indblik i og kræver mere 'benarbejde', herunder såkaldt 'jailbreaking' af Apple-

telefoner. Det primære fokus for vores arbejde har været at udvikle strategier til at foretage opslag af apps på iOS platformen, der kan sammenlignes med de informationer, man kan hente om Android apps via Exodus Privacy-platformen.

Metodiske skridt til statisk analyse:
Android og iOS-platformene har bestemte konventioner for programmeringssprog og appdesign, samtidig med at SDK'er er universelle og ikke app-specifikke. Derudover stiller platformsindehaverne krav til, hvordan apps på deres platform skal udvikles og deklareres, hvilket medfører en grad af uniformitet blandt apps. Den overordnede process for den statiske analyse er ens på tværs af Android og iOS, hvor målet er at finde bestemte signaturer for SDK'er igennem programmets kode. På Android gøres dette uden at åbne applikationen, da signaturerne kan aflæses i dele af programkoden ved hjælp af værktøjer fra Google. For iOS er der som tidligere nævnt flere forhindringer, som kræver, at applikationen bliver kørt på en enhed, før signaturerne kan læses. Ved at holde signaturnavnene fra analysen op mod de kendte signaturer brugt af en SDK, kan man indentificere hvilke SDK'er der er i applikationen.

Resultater: veje og vildveje
Det vigtigste resultat af vores undersøgelse er, at det er praktisk muligt – omend besværligt – at foretage statiske analyser af iOS-apps, der i nogen udstrækning kan sammenlignes med de informationer, vi kan få om Android-apps via Exodus. Det er et vigtigt fund, fordi det åbner en vej for monitorering af det mobile tredjepartsmarked og for sammenligninger af de forskellige mobile økosystemer, som danskernes data bevæger sig rundt i afhængigt af, hvilke type enheder og styresystemer, de bruger.
Samtidig har vores arbejde klarlagt, at statiske app-analyser af især iOS besværliggøres af flere forhold: For det første er det nødvendigt at jailbreake en iOS-enhed for at få adgang til systemets kode. Jailbreaking af iOS-enheder kræver tekniske færdigheder og er en eksperimentel metode til at analysere mobile apps. Der findes altså ikke officielle eller videnskabeligt etablerede fremgangsmåder til at foretage denne type interventioner. For det andet modarbejdes jailbreaking løbende af Apples opdateringer og ændringer af services brugt af tredjeparter. Det betyder konkret, at mange nyere versioner af apps ikke kan analyseres på denne måde, da de kræver et opdateret styresystem, som ikke muliggør jailbreaking. Endelig er det en udfordring, at den nuværende state-of-the-art database over iOS tredjepartstjenester, som de identificerede SDK'er slås op i, er temmelig begrænset (95 tredjepartstjenester er identificeret i databasen, mod ca. 500 i Exodus's database). Denne database bliver desuden ikke udviklet eller velligeholdt, modsat Exodus, hvis database aktivt udbygges og velligeholdes. Denne udfordring vil kunne overkommes ved at udbygge den eksisterende iOS-trackerdatabase gennem manuel inspektion af dataudtræk fra apps, hvilket dog er en omfattende opgavem idet en enkelt app typisk vil resultere i ca. 60.000 linjers dataudtræk, som skal screenes for tredjepartstjenester. Den statiske analyse er derfor på nuværende tidspunkt i stor grad begrænset af, at de specifikke SDK'er først skal identificeres, før de kan listes i en database til fremtidige analyser.

Anbefalinger til videre arbejde
De ovenfor skitserede udfordringer peger overordnet på, at statiske analyser og udviklingen af metoder, som dem vi har afprøvet, ikke bør anses som en 'engangsopgave'. Det kræver løbende vedligeholdelse, drift og tekniske ressourcer for at imødegå systemopdateringer og

udviklinger i markedet for tredjepartstjenester. Det er vores vurdering, at det er afgørende for en fremtidig systematisk monitorering af det mobile datamarked, at der investeres i teknisk ekspertise, der løbende kan udvikle og opdatere analysemetoderne. Derudover er udbygning og vedligholdelse af databaser over tredjepartstjenester og deres funktioner essentielt for løbende at overvåge det mobile appmarked, og dette vurderer vi bedst lader sig gøre, ved at sikre, at værktøjer og databasen over iOS, ligesom Exodus Privacy, er offentligt tilgængelige. Generelt er en systematisk og løbende monitorering af det mobile tredjepartsmarked udfordret af de mobile infrastrukturer, der kontrolleres af henholdsvis Alphabet (Google) og Apple gennem deres kontrol over de mobile styresystemer og appstores. De hyppige systemopdateringer og tekniske ændringer i, hvilke informationer der er tilgængelige om de enkelte apps, gør det vanskeligt at udarbejde automatiserede metoder til at identificere og overvåge det mobile datamarked, selv hvis gode og pålidelige databaser over tredjepartstjenester bliver tilvejebragt.

En anden udfordring for bestræbelserne for øget transparens om mobil dataindsamling er, at de SDK'er, der typisk danner baggrund for statiske analyser ofte har en række forskellige formål for den specifikke app og at det blandt andet er vanskeligt at adskille rent teknisk-funktionelle eller sikkerhedsmæssige funktioner og mere kommercielle funktioner fra hinanden. Det vil kræve dybere undersøgelser af de enkelte SDK'ers virkemåde.

Samlet set har vores arbejde vist, at der er et påtrængende behov for at udvikle analysemetoder, der kan skabe større transparens omkring det mobile datamarked og dermed muliggøre de ambitioner, som er fastslået i såvel Databeskyttelsesforordningen (GDPR) og senest i den nyligt vedtagne Digital Services Act (DSA). Vi ønsker med denne afrapportering at gøre opmærksom på, at der eksisterer muligheder for at styrke monitoreringen af mobil dataindsamling, men at det kræver vedvarende ressourcer og udvikling af teknologisk ekspertise. Det mobile datamarked er præget af omfattende kommercielle interesser, der former, hvilke informationer vi har adgang til og dermed vores muligheder for at have kvalificerede, kritiske og demokratiske diskussioner om, hvordan vores stadig mere dataficerede samfund skal udvikle sig. En styrket indsigt i og monitorering af de mobile datamarkeder vil ydermere gøre det muligt at stille krav til eksempelvis offentlige apps om indrapportering af anvendte SDK'er og deres formål, samt at indsamle og offentliggøre statistisk materiale om markedsudviklingen.

Referencer

Kollnig, K., Shuba, A., Binns, R., Van Kleek, M., & Shadbolt, N. (2022). Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps. Proceedings on Privacy Enhancing Technologies.