Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

Sofie Hagen, Susanne Kolberg, Oda Motrøen-Sevilhaug og Elida Krogstad Thorsrud

From Advertising to Intelligence: Investigating Tracking through Mobile Devices

Bachelor's thesis in Digital Infrastructure og Cyber Security
Supervisor: Arvind Sharma og Jan William Johnsen
May 2024

NTNU
Norwegian University of
Science and Technology

Sofie Hagen, Susanne Kolberg, Oda Motrøen-Sevilhaug og Elida Krogstad Thorsrud

From Advertising to Intelligence: Investigating Tracking through Mobile Devices

Bachelor’s thesis in Digital Infrastructure og Cyber Security
Supervisor: Arvind Sharma og Jan William Johnsen
May 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Abstract

Mobile devices have become daily extensions of ourselves, offering numerous functionalities that make our lives easier and more entertaining. However, this convenience has a darker side; mobile devices constantly gather personally identifiable information, presumably without our full awareness. This data feeds into a complex system that creates detailed user profiles for targeted advertising. Primarily designed for product marketing, the system is vulnerable and can be exploited to harvest intelligence about individuals, a practice known as Advertising Intelligence (ADINT). This is the advertising ecosystem: a complex web where our data is the currency fueling an industry that thrives on our digital footprints.

This thesis delves into the complexities of the advertising ecosystem, focusing on how personally identifiable information, with an emphasis on the Mobile Advertising ID, is collected by mobile applications and subsequently shared with third-party trackers. Through an investigation of iOS applications available in Norway, we examined how data is collected by mobile devices and traced the flow of this data to various advertising networks and third parties. Our study uncovers the collection of several user properties, such as location and gender, along with significant gaps in tracking transparency.

Sammendrag

Mobiltelefoner har blitt en forlengelse av oss selv i vårt daglige liv, og tilbyr en rekke funksjoner som gjør livene våre enklere og mer underholdende. Det finnes imidlertid en bakside; telefonene samler kontinuerlig inn informasjon som er personlig identifiserbar, antaglig uten vår viten. Informasjonen lagres i et system der det opprettes detaljerte brukerprofiler, som videre brukes til målrettet reklame. Systemet er laget for markedsføring, men er sårbart for utnyttelse til et annet formål, nærmere sagt annonsebasert etterretning. Dette er den digitale annonseindustrien; et komplekst økosystem hvor brukernes data er valutaen som får det hele til å gå rundt.

Oppgaven utforsker den komplekse annonseindustrien, med fokus på hvordan personlig identifiserbar informasjon, spesielt reklameidentifikatoren, samles inn av mobilapplikasjoner og videre distribueres til tredjeparter. Vi undersøker et utvalg av iOS-applikasjoner tilgjengelige i Norge, for å finne ut hvordan data samles inn av mobile enheter, før den sendes videre til forskjellige reklamenettverk og tredjeparter. Undersøkelsen avdekker at det samles inn en rekke personlig identifiserbar informasjon, slik som lokasjon og kjønn, uten at det gis tilstrekkelig informasjon om sporing til brukerne.

Preface

This thesis concludes our bachelor's degree in Digital Infrastructure and Cybersecurity at NTNU Gjøvik. It was written during our last semester, from January 1, 2024, to May 21, 2024, at the Department of Information Security and Communication Technology. Our thesis supervisors were researcher Jan William Johnsen and associate professor Arvind Sharma, both from NTNU. The research project is provided by Kripos, the Norwegian National Unit for Combating Organized and Other Serious Crime.

The topic was selected by proactively contacting Kripos and requesting a project. Kripos expressed an interest in exploring Advertising Intelligence (ADINT), particularly focusing on the collection of the Mobile Advertising ID (MAID) and other Personally Identifiable Information (PII) from mobile applications. The problem statement was not predefined, allowing freedom in selecting our own. This resulted in three research questions that could yield valuable insights for both Kripos and the public. Driven by curiosity and the wish to maximize our learning outcomes, we embraced the challenge of exploring the unknown field of the mobile advertising ecosystem and its intelligence opportunities.

Previous work and background theory are included to enable the readers to understand the context of our research area and follow through with our thesis. However, the readers will benefit from basic knowledge of networking concepts and cybersecurity.

We hope you enjoy reading our thesis.

Acknowledgement

Thank you to our supervisors, Arvind Sharma and Jan William Johnsen, for valuable guidance and feedback throughout the project. Their support helped keep us on track and within scope, as well as prompting us to think critically by asking insightful questions.

We would also like to thank our classmate Karl Henrik Horve for his technical support and valuable discussions in our project. Additionally, we extend our gratitude to Group 118 for taking the time to conduct peer reviews and provide helpful feedback on our thesis. We would also like to thank Alexander Roaas for his motivational support and Anders Rindal Loeng for his great advice and feedback on our thesis.

Finally, thank you to Kripos for contracting the project and giving us valuable insight into the topic.

Contents

Figures

Tables

Acronyms

ADINT Advertising Intelligence. iii, vii, 1, 2, 4, 9–11, 18–20, 23, 25, 33, 69–71, 74

AI Artificial Intelligence. 5, 65, 70, 75

ATS Advertisement and Tracking Services. 21

ATT App Tracking Transparency. 13, 48

CA Certificate Authority. 43

DSP Demand-Side Platforms. 2, 3, 7–9, 57

ECID Exclusive Chip Identification. 54

ePD ePrivacy Directive. 21, 23, 24

GDPR General Data Protection Regulation. 13, 21–24, 66, 67, 69

GUI Graphical User Interface. 27, 40

GUID Global Unique Identifier. 11, 12

HTTP Hypertext Transfer Protocol. 36, 40, 41, 48

HTTPS Hypertext Transfer Protocol Secure. 36, 40, 41, 48

HUMINT Human Intelligence. 9

ID Identifier. 12, 29, 42, 43, 62, 63, 67

IDFA Identifier for Advertisers. xiii, 1, 3, 4, 12, 13, 23, 25, 27, 33–35, 37, 39, 45, 51–58, 61–63, 66, 67, 72, 73

IMEI International Mobile Equipment Identity. 54

IP Internet Protocol. 13, 36, 48, 54, 55, 66, 67, 72

MAID Mobile Advertising ID. vii, 1–3, 9, 11, 12, 18, 19, 22, 71, 74

MiTM Man-in-The-Middle. 25, 27, 39, 43, 66, 72

NRK Norsk rikskringkasting. 10, 19, 20, 34, 36, 56

NTNU Norges teknisk-naturvitenskapelige universitet. vii, 18

OSINT Open-Source Intelligence. 9

PII Personally Identifiable Information. vii, 2, 3, 10, 12, 19, 20, 23–25, 27, 33, 34, 36–38, 41, 48, 51–59, 66, 68, 71–74

SDK Software Development Kit. 2, 8, 9, 15

SoC System-on-a-Chip. 40, 65

SPL Splunk Search Processing Language. 34, 45, 51–53, 59

SSL Secure Socket Layer. 43, 44

SSP Supply-Side Platform. 2, 7, 8, 19

TLS Transport Layer Security. 43

UDID Unique Device Identifier. 34, 52, 54, 56, 72

UN United Nations. 5, 65, 69, 70

VM Virtual Machine. 40, 44, 46

Chapter 1

Introduction

The following chapter provides an overview of our thesis. It introduces the thesis' topics, problem description, and scope. Additionally, it presents our research questions, project goals, and target audience. Finally, it outlines the structure of our thesis, presenting each chapter and what they will cover.

1.1 Topics covered by the thesis

The online advertising ecosystem is based on advertising networks' ability to know properties about users, e.g., their interests or location [1]. The properties are used to build detailed user profiles, which can be used for targeted advertising. Behind the targeted ads, an advertiser determines which audience the ads should serve. After the campaign is deployed, the advertiser receives metrics and reports on user responses, including impressions, clicks, and the demographics of the people who clicked [2]. While targeted advertising might seem straightforward, the ecosystem is complex and possible to exploit [1]. By posing as an advertiser, targeted ads can be deployed for purposes other than product marketing. Through the standard ad display and reporting processes, intelligence about other individuals can be harvested [1]. This concept is known as Advertising Intelligence, or ADINT for short, and it is one of the primary topics covered by our thesis.

When discussing ADINT, the Mobile Advertising ID, hereinafter referred to as MAID, plays a vital role. The MAID is a numeric unique identifier on a mobile device and is commonly communicated to application servers to identify the device for advertising purposes [3]. The MAID can be utilized to analyze user interactions within the app; however, if a threat actor obtains a MAID, it can be exploited for ADINT purposes. Apple's version of MAID is called Identifier for Advertisers (IDFA). As the IDFA uniquely identifies users across different applications, Apple requires explicit user consent when it is collected, due to its role in tracking activities [4].

To understand how ADINT becomes possible, it is essential to identify the entry points where the user's properties are extracted and where they are sent. While websites typically utilize cookies to gather such properties, mobile apps integrate advertising libraries, which distribute the MAID alongside other Personally Identifiable Information (PII) to trackers [1]. The ad libraries are available through Software Development Kits (SDKs), code that enables mobile apps to connect to third parties for services such as analytics and advertising. Trackers gather and exchange large amounts of properties across different platforms, a practice known as cross-platform tracking [5]. Cross-platform tracking enables more detailed profiling, which enhances the effectiveness of targeted advertising.

One method used to inform users about the types of data collected and used for tracking purposes is by implementing privacy labels. This aspect is particularly noteworthy given that these privacy labels are self-declared and lack verification by Apple [4]. By examining whether the data collected correlates with what applications declare in their privacy labels, one can gain insight into how transparent they truly are about the data they collect.

Keywords: Advertising Intelligence (ADINT), Personally Identifiable Information (PII), third-party tracking, privacy labels, and tracking transparency.

1.2 Problem description

The mobile phone has evolved into an indispensable tool that most individuals carry almost everywhere. Through a variety of applications, users can access journalistic content through newspapers, stay updated on others' lives through social media, and engage in diverse activities. Many of these apps rely on advertisements for revenue, making them part of the intricate advertising ecosystem.

The online advertising industry heavily relies on user data to target advertisements effectively. However, this industry has proven exploitable for purposes beyond product marketing, namely intelligence. Previous instances have demonstrated that ADINT is feasible through acquiring MAIDs and posing as legitimate advertisers. The advertising ecosystem is intricate, consisting of numerous components, including Demand-Side Platforms (DSP)s, Supply-Side Platform (SSP)s, data brokers, and user devices. Each component exchanges data with one another, creating a complex network that is challenging to decode. This makes it difficult to trace the flow of information and understand its destination.

This study will focus on a specific aspect of the advertising ecosystem to understand the exploitation of user data for targeted advertising and ADINT. The central element facilitating this is the MAID, collected by mobile applications offering advertisements. Our thesis aims to investigate precisely what information, particularly focusing on MAIDs and other PII, is collected by mobile phones and to whom it is sent. The thesis will also investigate how the collected data is stated in the mobile apps' privacy labels.

1.3 Scope

Our thesis deep-dives into one of many components of the complex advertising ecosystem. The research is scoped to investigate the source of user data. This means we will investigate the traffic sent from mobile apps before it arrives at data brokers, third-party trackers, and DSPs.

We have chosen to scope the thesis to investigate Apple mobile devices, as previous research primarily focuses on Android. This consequently scopes our thesis to iOS applications and Apple's version of MAID, known as IDFA.

1.4 Research questions

This thesis is highly motivated by increasing the transparency of app tracking. Our thesis seeks to investigate which PII is collected, where this data is sent, and whether privacy labels transparently declare this to the user for a set of popular apps. To do so, this thesis will answer three research questions as follows:

1. Which PII is collected by iOS applications during 15 minutes of use?
2. During the 15 minutes of app usage, which tracking domains are most frequently contacted?
3. To what extent are the privacy labels transparent about the data collected for tracking purposes?

1.5 Project goals

For this thesis, we have set goals to guide our performance and results. Performance goals highlight the reasons and basis for the project, while result goals specify the outcomes we want to achieve.

Performance goals:

- Enhance understanding of applications' tracking practices and how this can be exploited for intelligence
- Increase knowledge and insight in capturing the IDFA through network traffic interception
- Enhance user awareness regarding tracking transparency

Result goals:

- Identify the most commonly collected PII

- Map out frequently contacted tracking domains
- Determine tracking transparency in privacy labels

1.6 Target audience

This thesis's target audience is divided into two categories: those who wish to track others and those who wish not to be tracked.

For the audience that wishes to track others, techniques disclosed by this thesis regarding capturing the IDFA and the principles of how ADINT works might be of particular interest. In addition, specific apps that have been disclosed to collect the IDFA might also be of interest. This audience includes official government entities such as law enforcement and intelligence services.

On the other hand, the audience who wishes not to be tracked can consider this thesis a contribution to awareness. We assume the majority of the readers will be from this category, which is why we concentrate our thesis experiment on the user's perspective. This audience includes classmates, lecturers, researchers, family, and others interested in digital exposure, tracking, and privacy.

1.7 Thesis outline

This chapter has briefly introduced the topics covered by our thesis. Additionally, it has defined our scope, research questions, goals, and target audience. Further, we will guide you through the complexities of the advertising ecosystem and the concept of advertising intelligence. We will discuss previous work and outline our chosen methodologies before conducting an experiment investigating tracking through mobile apps. Lastly, we will discuss and reflect on our thesis before concluding our research and proposing further work. The structure of our thesis is stated below.

- **Chapter 2** presents the background theory necessary to understand our thesis.
- **Chapter 3** reviews prior research relevant to our topics, discussing how these studies contribute to our thesis and experimental work.
- **Chapter 4** outlines the various methods employed in our work, including how we compiled our datasets, selected applications, and preprocessed our datasets. Additionally, it outlines the methodology used to address the defined research questions.
- **Chapter 5** details the experimental design and environmental setup. It provides a detailed description of the steps conducted to answer our research questions, followed by a presentation and discussion of their results.
- **Chapter 6** discusses the practical recommendations, legal and ethical aspects and provides reflections on our thesis, highlighting its contribution to

the United Nations's sustainability goals and the use of Artificial Intelligence (AI).

- **Chapter 7** concludes our research and proposes further work.

Chapter 2

Background theory

The following chapter provides an overview of the theoretical framework relevant to our thesis, enhancing understanding of subsequent chapters and the experiment's components. It begins by outlining the online advertising ecosystem and setting the context for the following discussions. We will then delve into the specifics of advertising intelligence operations in this ecosystem. Furthermore, the chapter examines the advertising identifier, in addition to other key identifiers that are central to the topics. The subsequent section introduces a policy introduced by Apple, marking a shift in approaches to tracking and targeted advertising. Finally, the structure and implications of Apple's privacy labels are detailed.

2.1 The online advertising ecosystem

The online advertising ecosystem represents a complex and rapidly evolving field, significantly influenced by technological advancements and the increasing ability to collect and analyze great amounts of data [6]. This evolution has completely affected how advertisers reach consumers, leveraging detailed personal information to create comprehensive profiles. The greater the detail in the profiles, the higher their market value [6]. The correlation between the amount of data and its value explains how the market is configured to achieve favorable outcomes.

In Figure 2.1, we have created an illustration to clarify the complexities of the online advertising ecosystem. The figure illustrates that when a user opens an application, the app sends information to a Supply-Side Platform (SSP) indicating that it has an available ad space to sell, along with user data. Furthermore, the SSP handles the sale of these ad spaces to advertisers [1]. However, the advertisers cannot directly purchase ad spaces from the SSP; they must select a Demand-Side Platforms (DSP) that serves ads to the desired target audience. According to Arora [7], a DSP can target audiences in numerous ways, including:

Behavioural targeting: Targets users based on their prior activities, including ads they have clicked on or specific products they have viewed.

Lifestyle targeting: Targets users based on their specific interests or lifestyle behaviors.

Interest based targeting: Targets users according to their interests in particular topics or categories.

Demographic targeting: Targets users based on their location, gender, age, income, etc.

Device targeting: Targets users based on their device characteristics, including the operating system, device type, and other specifications.

Furthermore, advertisers provide their specific targeting criteria to the DSP [1]. The DSPs' algorithms then manage the bidding process on behalf of the advertisers [6]. When an SSP presents a user matching the specified criteria, the DSP evaluate the users' value and places a bid on behalf of the advertiser [6]. This bid is calculated based on the amount the advertiser initially agreed to pay for targeting those specific users [8]. The advertiser with the highest bid wins the advertising space [6].

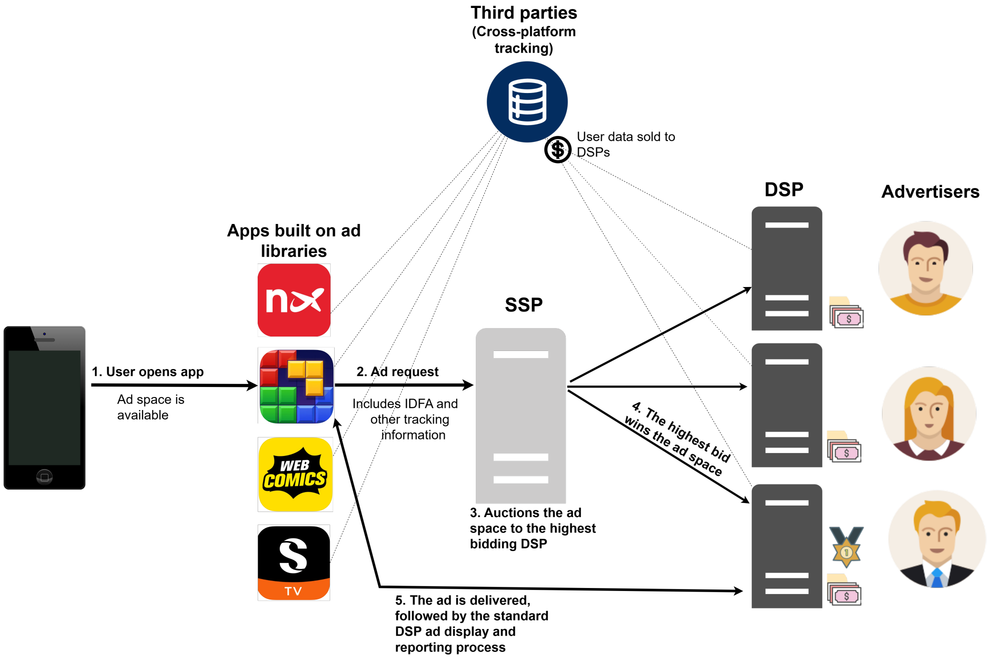

Third parties
(Cross-platform tracking)
User data sold to DSPs
DSP
Advertisers
Apps built on ad libraries
SSP
1. User opens app
Ad space is available
2. Ad request
Includes IDFA and other tracking information
4. The highest bid wins the ad space
3. Auctions the ad space to the highest bidding DSP
WEB COMICS
TV
5. The ad is delivered, followed by the standard DSP ad display and reporting process

Figure 2.1: The online advertising ecosystem

Developers integrate Software Development Kit (SDK)s and ad libraries into their application's code to facilitate the advertising process [1]. SDKs are collections of various software tools and libraries that allow developers to seamlessly integrate ad networks into their apps [9]. These SDKs collect user data, such as user interactions and location data, to deliver targeted ads [9]. Multiple applications may incorporate the same ad library, enabling the collection of user information across various platforms [5]. This capability facilitates the creation of detailed user profiles, as the aggregated data from different applications provides a comprehensive view of user behavior and preferences [5].

2.2 Advertising intelligence

The Norwegian intelligence service defines the term intelligence as follows: "Intelligence is the result of state-sanctioned collection, analysis, and evaluation of data and information that is generated openly or covertly and compiled to provide an advantage in decision-making processes" [10]. Intelligence can be categorized into various disciplines, including Human Intelligence (HUMINT), which involves human-based collection of information, and Open-Source Intelligence (OSINT), which involves gathering information from publicly available sources. Additionally, there is Advertising Intelligence (ADINT), which focuses on intelligence derived from online advertising [1].

There are several ways to exploit the online advertising ecosystem to harvest intelligence about an individual [1], [11], [12]. We have illustrated a sample scenario of such exploitation in Figure 2.2, where an ADINT attacker has registered as an advertiser to a DSP to participate in the ad bidding process and target ads to an individual they wish to track.

The study by Vines *et al.* [1] details the same scenario, describing how Alice can track Bob. The process was rather simple: First, they identified the target's MAID by sniffing network traffic. Next, they selected a DSP capable of serving ads to an app with a large user base, further allowing location tracking. They then placed requests to their DSP, including criteria that matched the target's residential area. Upon securing the bids, the ads were delivered, and they used the initially captured MAID to identify the target user among those who received the ads. In this manner, they harvested intelligence on the target, including its location-based movements.

In a report by Benjakob [11], he informs how such exploitation can be used to distribute spyware. Furthermore, he explains how Israeli cyber companies deploy such technology to gather data and monitor civilians by infecting ads before the bidding begins and delivering them to the target audience. Alarmingly, the spyware has been sold to a non-democratic country [11], highlighting the potential for misuse in the ADINT field. As of September 2023, there are no effective defense mechanisms against this spyware, which neither Google nor Microsoft can block [11].

Another way to utilize ADINT is by exploiting user data collected by apps and distributed to third parties [12]. These third-party companies gather vast amounts of data on various users and sell it to others, which can then be used for tracking individuals. Although the data is stated as anonymized, NRK's article *Exposed by the mobile* demonstrated that identifying individuals from these datasets can easily be done without extensive resources [12].

These types of exploitation indicate that the distribution of PII may contribute to unwanted and undisclosed tracking, negatively impacting individuals. However, it can also facilitate tracking for legitimate purposes, such as aiding law enforcement in criminal investigations. Dishonest individuals might use ADINT techniques to locate and harm specific people or groups, while law enforcement can use it to track suspects' locations and movement patterns. In this context, while controversial, the collection of PII could be essential for identifying and locating individuals.

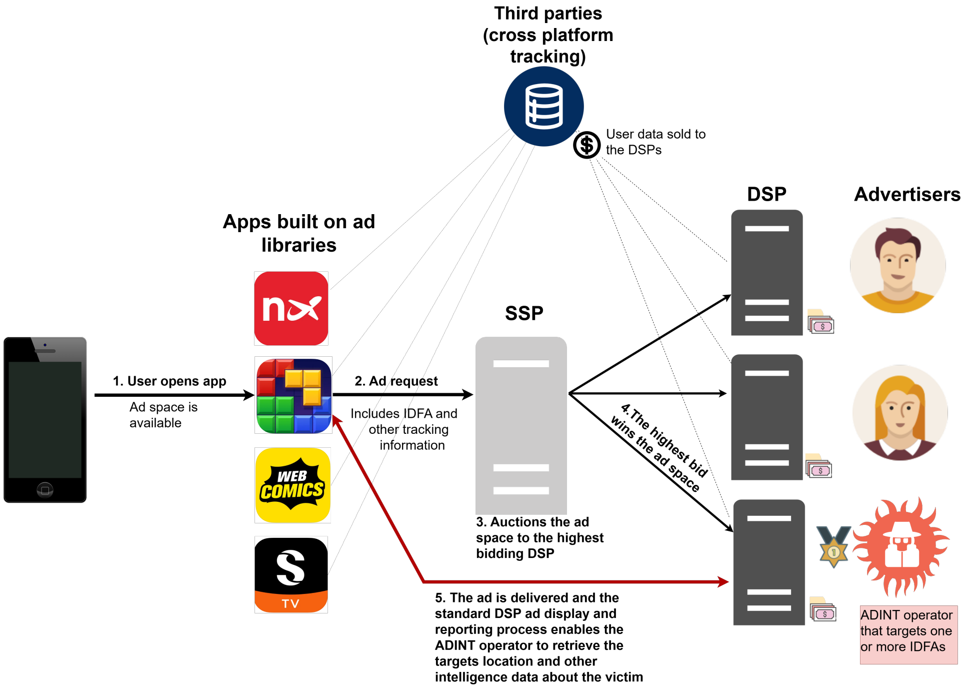

Third parties (cross platform tracking)
User data sold to the DSPs
DSP
Advertisers
Apps built on ad libraries
SSP
WEB COMICS
TV
1. User opens app
Ad space is available
2. Ad request
Includes IDFA and other tracking information
3. Auctions the ad space to the highest bidding DSP
4.The highest bid wins the ad space
5. The ad is delivered and the standard DSP ad display and reporting process enables the ADINT operator to retrieve the targets location and other intelligence data about the victim
ADINT operator that targets one or more IDFAs

Figure 2.2: ADINT attack example

2.3 Personally identifiable information

Personally Identifiable Information, or PII, refers to any data that can enable the identification of an individual, either alone or in combination with other information associated with a specific person [13]. This means that the information is directly connected to or can be associated with, an individual. This section will describe the PII relevant to our thesis.

2.3.1 The Mobile Advertising ID

The Mobile Advertising ID (MAID) is a unique identifier associated with a mobile device, provided by the device's operating system, which is specifically used for advertising [3]. Once a user consents to its use, the MAID is distributed and shared across the various applications the user engages with, allowing these applications to collect data on the user's preferences, activities, and interactions within each app [3]. This data is then utilized to tailor advertising campaigns, significantly enhancing their effectiveness by aligning ads with individual user behavior and preferences [3].

The MAID comes in a Global Unique Identifier (GUID) format consisting of groups of 8-4-4-4-12 characters, separated by four hyphens ('-'), as illustrated below. Please note that this is a randomly generated GUID for demonstration purposes, as we aim not to disclose real MAIDs in our thesis for privacy reasons.

5E41D290-9F36-4196-80D1-2657859104A5

The MAID is often transmitted unencrypted to ad exchanges via network traffic, making it possible to capture through network interception [1]. This vulnerability allows an attacker to capture an individual's MAID without physical access to the device simply by intercepting the device's network traffic. For example, attackers could be within the WiFi range of the target on an unsecured network, intercept cellular traffic, or access the WiFi router used by the target [1]. Notably, the MAID only needs to be captured once before ADINT attacks become possible.

An experiment conducted by Vines *et al.* [1] confirmed that capturing the MAID is feasible through these methods. Additionally, they found that the MAID can be obtained if the target interacts with an attacker's ad or through JavaScript in ads from major ad libraries. In addition, purchasing a target's MAID and other data online is also possible [12]. These findings underscore the ease with which attackers can exploit vulnerabilities in the digital advertising ecosystem for ADINT purposes.

Mobile Advertising ID vs. Cookie

MAIDs are compared to cookies in how they track user activity [14], although they operate differently, as outlined in Table 2.1. MAIDs, which require only the user's initial consent, track consistently across all applications on a device, making them highly effective for comprehensive advertising strategies. In contrast, cookies - small data files stored on a computer or mobile device by a website - demand more frequent user interaction [15]. First-party cookies are browser-specific and vary with each website, while third-party cookies, set by domains other than the visited one, track users across multiple websites often without explicit consent [16].

The consistency of MAID across apps within the same device allows for the aggregation of extensive user data, such as device details, location, and app usage patterns, enhancing targeted advertising effectiveness [3]. Conversely, first-party cookies gather data specific to website visits and user preferences, whereas third-party cookies broaden this tracking across sites.

Table 2.1: MAID vs. Cookies

Feature	MAID	First-Party Cookie	Third-Party Cookie
Association	Device-specific	Browser-specific	Browser-specific
Data collected	Device details, carrier, location, etc.	Site visit details, user preferences	Site visit details, tracking across multiple websites
Scope of tracking	Cross-application on the device	Limited to a single website per browser	Multiple websites across the internet
Access and manage	At the system level	Through the browser	Through the browser
Consistency	Consistent across all applications on the device	Unique to each browser and website; does not share data across different browsers or websites on the same device	Unique to each browser but shared across multiple websites within the same browser that uses the same third-party services
Average lifespan	7-8 months	24 hours	Variable, often longer than first-party cookies
User control	Reset or disable MAID	Delete individual cookies	Delete individual cookies

As detailed in Table 2.1, inspired by Equifax [14], additional differences include the lifespan of how long each identifier persists before it automatically resets. Furthermore, how the user can manage them; MAIDs can be reset or disabled, while cookies can be deleted or blocked through browser settings, plugins, or extensions.

Apple's Identifier for Advertisers

The identifier provided by Apple for uniquely identifying iOS devices is known as the Identifier for Advertisers (IDFA). This ID is used to track user activities and events related to advertising campaigns and marketing channels without revealing PII [17]. When tracking is enabled, the IDFA appears as an alphanumeric string; if disabled, it returns a GUID sequence only containing zeroes [18].

2.3.2 Additional identifiers

In the realm of digital and personal identification, the concept of an *identifier* extends far beyond traditional forms of identification like name or phone number. According to the University of Virginia, an identifier is defined as "any data that can either directly identify an individual or link an individual to their identity" [19]. Identifiers encompass a broad spectrum of data, from standard personal information to advanced digital markers like IP address and unique device identifiers. As technology evolves, so does the nature of identifiers. An IP address, for instance, can be traced back to a specific location or device, while unique device identifiers globally distinguish one electronic device from another. Additionally, geographical data from mobile devices can potentially reveal patterns about an individual's routine and location. With digital technology's advancement, identifiers now include various data points that can accurately track and identify an individual, highlighting the importance of understanding these identifiers for maintaining privacy according to GDPR and security in today's digital landscape.

2.4 App Tracking Transparency policy

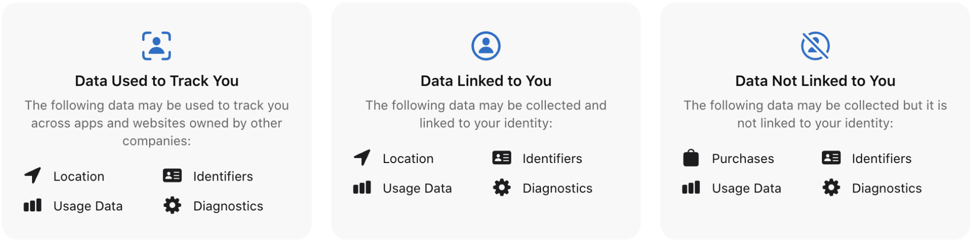
The App Tracking Transparency (ATT) policy, introduced by Apple with the iOS 14.5 update in April 2021 [20], represents a significant shift in the mobile advertising industry. This policy mandates that apps must obtain explicit permission from users before using the IDFA to track their activities across different apps [20]. Before the implementation, users were automatically opted into data tracking, but now, the policy restricts the amount of user data that app developers can share with external companies. This change has led to a significant reassessment of user privacy and data usage practices among app developers [21].

The ATT policy primarily addresses the use of third-party data, which is information collected through interactions between companies, rather than first-party data gathered directly by the app's provider. Due to the ATT policy, apps must now request permission to use third-party data for targeted advertising, share identifiers and location data with advertising networks, or merge third-party data with their own for ad targeting and efficient analysis. However, using first-party data within the same company's applications for advertising does not count as tracking under ATT policy and does not require user consent [22].

The primary aim of the ATT policy was to give iOS users more control over how third-party data is used in advertising, addressing growing concerns about privacy and data governance in the digital era. By requiring an opt-in for tracking, Apple emphasizes user privacy and sets a new industry standard, ensuring users can decide if and how their data is used for advertising purposes [20]. This initiative highlights a broader move towards enhanced user privacy and control across the tech industry.

2.5 Privacy labels

In the Apple App Store, privacy labels serve as concise summaries of privacy policies designed to inform users about an app's data collection and handling practices. Unlike lengthy privacy policies, which detail practices and procedures extensively, privacy labels focus on providing clear and accessible information about data handling without being overwhelming. An example is outlined in Figure 2.3.



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Location, Identifiers, Usage Data, Diagnostics

Data Linked to You

The following data may be collected and linked to your identity:

Location, Identifiers, Usage Data, Diagnostics

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

Purchases, Identifiers, Usage Data, Diagnostics

Figure 2.3: Example of privacy labels

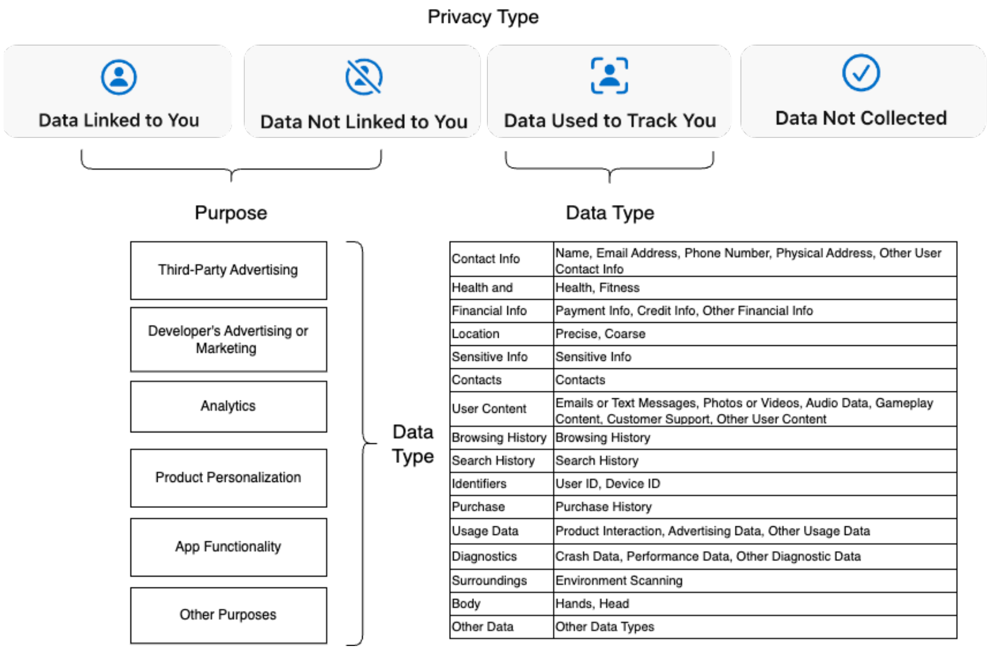
While privacy policies share similar objectives, they are legally binding agreements that may be difficult for users to fully comprehend due to their formal language and complexity [23]. In contrast, privacy labels are designed to be user-friendly, offering transparency and insight into an app's data practices in a clear and accessible manner [24]. It is essential to note that the information presented in privacy labels is provided by the app's developer, highlighting the distinction in the source of information between privacy policies and privacy labels.

The manual for privacy labels instructs developers to identify all data collected by either the specific application or their third parties [24]. Privacy labels are constructed to include various levels, including privacy type, data use, and data types. To illustrate their structure, we have created an overview in Figure 2.4.

At the topmost level is the *Privacy Type*, which entails four different categories. Three of these categories describe ways in which data is utilized: *Data Linked to You*, *Data Not Linked to You*, and *Data Used to Track You*. All of these three can coexist within the same application. The fourth category is *Data not Collected*, which simply indicates that no data is collected and does not add any further details to the label.

Furthermore, the privacy labels suggest six purposes for how collected data is utilized. These purposes are only applicable for the privacy types *Data Linked to You* and *Data Not Linked to You*, as shown in Figure 2.4. The six purposes of data collection are: *Third-Party Advertising*, *Developer's Advertising or Marketing*, *Analytics*, *Product Personalization*, *App Functionality*, and *Other Purposes*.

Apple has categorized various data types to detail the specific kinds of data that applications might collect. These data types further specify the particular data



Contact Info	Name, Email Address, Phone Number, Physical Address, Other User Contact Info
Health and	Health, Fitness
Financial Info	Payment Info, Credit Info, Other Financial Info
Location	Precise, Coarse
Sensitive Info	Sensitive Info
Contacts	Contacts
User Content	Emails or Text Messages, Photos or Videos, Audio Data, Gameplay Content, Customer Support, Other User Content
Browsing History	Browsing History
Search History	Search History
Identifiers	User ID, Device ID
Purchase	Purchase History
Usage Data	Product Interaction, Advertising Data, Other Usage Data
Diagnostics	Crash Data, Performance Data, Other Diagnostic Data
Surroundings	Environment Scanning
Body	Hands, Head
Other Data	Other Data Types

Figure 2.4: Overview of the privacy label structure

included, e.g. the data type *Identifiers* encompasses *Device ID* and *User ID* each accompanied by its own explanation clarifying the typical information it represents. These data types must be specified along with the purposes of data collection or under the *Data Used to Track You* label. This ensures that users are aware of what kinds of data are being collected and for what purposes, enhancing transparency and user awareness.

Apple's definition of tracking

According to Apple's definition, tracking means linking data collected from an app about specific users or devices with third-party data for targeted advertising or ad measurement [24]. Examples include using data from other companies for ads, sharing information with data brokers, or incorporating third-party SDKs that merge user data for ad targeting. Activities not considered tracking include processing data solely on the device, using data exclusively for fraud detection, or sharing it with consumer reporting agencies for credit assessments.

Data linked to the user

Apple considers data linked to the user if it can directly identify an individual via account, device, or personal details, either collected directly or through third parties [24]. To ensure privacy, such data must be anonymized before collection, and efforts to re-link it to the user or merge it with identifiable datasets after collection are prohibited. Under privacy laws, personal information and personal data are always linked to the user [25].

Chapter 3

Previous work

This chapter outlines the previous work on the topics covered by our thesis, found through our literature research. This includes previous work on mobile advertising and advertising intelligence, third-party trackers, and privacy labels. Finally, the chapter summarizes the previous work according to each topic and emphasizes their contribution to our thesis.

3.1 Literature research

We conducted literature research to establish the thesis's theoretical foundation and get an overview of previous work in the field. The purpose was to outline current knowledge and identify areas for further investigation. To accomplish this, we utilized various search terms, as outlined in Table 3.1. The search terms consist of various topics addressed in this thesis, guiding us to relevant studies. These keywords were utilized to conduct searches across five online academic databases, as listed in Table 3.2. The selection of these databases was strategic, as they offer high-quality studies to form a solid foundation for this thesis.

Table 3.1: Search keywords

Advertising Intelligence / ADINT	Identifier for Advertisers / IDFA
Advertising ID / Ad-ID	Mobile Advertising ID / MAID
Mobile Ad-ID	Advertising ecosystem
Phone tracking	Advertising tracking
Privacy labels	Third-Party tracker domains

Table 3.2: Online academic databases

Google Scholar	Semantic Scholar
HAL Open Science	ACM Digital Library
IEEE Xplore	

To filter through the previous work and identify relevant articles, we established the following selection and exclusion criteria. The selection and exclusion criteria were as follows:

Selection criteria:

a) Time published: Literature published in 2014 or later, which scopes the research to the last ten years.

b) Relevance to the research questions: It must cover research on the MAID, tracking of individuals, or specifically address ADINT.

Exclusion criteria:

a) Language: Literature is written in a different language than Norwegian or English.

b) Accessibility: The literature is not openly accessible nor through the services provided by NTNU.

3.2 Advertising Intelligence

Advertising Intelligence (ADINT) is defined as the exploitation of the online advertising network to gather sensitive information about individuals by purchasing advertisements [1]. Vines *et al.* [1] investigated the potential, capabilities, and operational aspects of ADINT, exploring the types of information that can be obtained about individuals through its use.

Govindaraj *et al.* [26] tested both iOS and Android systems, jailbreaking the iOS devices to access the Safari history database. They initiated their study by simulating the advertisement ecosystem, subsequently extracting and analyzing various ads to identify users. The results demonstrate how this information can be used to identify a user, regardless of whether they use the same device, multiple devices, various networks, or exhibit diverse usage patterns. Several studies have been using various tools regarding this matter [1], [26], [12].

Numerous mobile applications incorporate third-party libraries, which are utilized to extract sensitive real-time geographical data from users for the purpose of location-based targeted advertising [5]. Hu *et al.* [5] employed a tool that traversed various geographical areas to gather data to characterize the severity and significance of location-based private data collection in mobile ad networks. The instrument they used to collect data provides a clear indication of the extensive gathering of private information about users from mobile ad networks across various applications. This is facilitated by the presence of the same ad library in multiple applications [5].

A major driving force behind cross-site targeting from Supply-Side Platforms (SSPs) is that the more information they can provide about a user, such as PII to the bidders, the higher bids they will receive [5]. This type of privacy leakage brings significant concerns about users' privacy across various applications. Vines *et al.* [1] assessed the necessary resources for executing a successful ADINT campaign. The findings showed that the attacker must be able to serve ads and obtain the target's MAID, which could be achieved by intercepting network traffic as it is often sent unencrypted to the ad exchanges [1]. Furthermore, Vines *et al.* show that the ability to acquire information through the ads could identify which applications the user has installed and how frequently they have been accessed.

Identifying specific apps used by an individual could potentially reveal sensitive details, particularly when it comes to apps where users register their religious beliefs or sexual orientation, as this is considered sensitive personal information [27]. Religious apps can pose a threat to their users when advertising is involved, as different ADINT actors may seek to identify users based on such information with the intent of stalking or surveilling them without their consent or knowledge.

Specifically, this means that if the MAID is known, ads can be served to their device, and the attackers can retrieve the victim's precise location before carrying out other malicious objectives. Vines *et al.* [1] describe this scenario with gay apps, where the threat actor is an ideological vigilante "with the objective of enforcing cultural norms or ideological positions on members in their community".

The fact that mobile ads can reveal the application name and version, a list of device capabilities, user-provided age and gender, user emotional state, user location, system language, and other valuable data about users [26], highlights how much information that could be exposed through mobile ads and potentially exploited by dishonest advertisers.

The potential for data to be exploited was proven by NRK when they purchased data from a British data broker, Tamoco [12]. Despite their purchased dataset lacking names and phone numbers, NRK managed to trace the movements of 140,000 Norwegians throughout 2019. Through a dialogue with one of the authors behind NRK's report, Martin Gundersen, we were informed that it took approximately six months to track a single individual (see full dialogue in Appendix A). NRK demonstrated the invasive nature of their method by identifying a Norwegian citizen by obtaining details about his home address, workplace, job changes, the timing of an interview, multiple hospital visits - which was disclosed on social media as the birth of his first child - and visits at a Zoo [12].

This was followed by tracking individuals working within no-drone zones, such as restricted and military areas, including several officers and an individual within the secured special forces area [12]. This method of identifying individuals demonstrates that it has the potential to pose a risk to national security.

Additionally, NRK identified 8,243 unique mobile devices, all of which were located at Norwegian hospitals [12]. Devices were also pinpointed at psychiatric institutions and crisis centers, highlighting the sensitive nature of the locations involved and the potential for severe privacy violations.

Furthermore, NRK successfully identified a Member of the Norwegian Parliament by analyzing location data from a makeup application. This data revealed frequent visits to a particular office at the Parliament and to an address used by several representatives as a commuter residence, various trips abroad, and weekends spent at a private residence. By combining this information with open-source searches and social media information, they were able to successfully identify the individual.

Through the dialogue with Gundersen, we were further informed that a particularly significant finding was how data from mobile apps circulated among various actors and ultimately ended up with a company that included U.S. authorities on its client list. This transformed what began as an innocent marketing data collection into national security surveillance. This illustrates how data collected from mobile devices facilitates advertising intelligence.

NRK demonstrates the extensive personal information one can leave behind through app usage and the implications of sending PII to third-party companies. Ads tailored for intelligence gathering can be activated within minutes and the information acquired through them can likewise be obtained within a short time frame [1], demonstrating the speed and effectiveness of ADINT attacks.

3.3 Third-party tracking

A study by Vallina-Rodriguez *et al*. [28] highlights that numerous applications depend on third-party providers for app functionality, with a substantial portion of these providers gathering user information. There is an issue with the insufficient clarity provided to the users regarding these third-party providers and the specific data that they are collecting [28], [29], [30].

Vallina-Rodriguez *et al*. [28] used the ICSI Haystack app, which can monitor domains different applications connect to, identifying domains used for mobile advertisement and tracking purposes. Furthermore, they received data from 690 Haystack users and flows generated from 1,732 applications.

Another approach was done by Klais [29], where *Record App Activity* was used, a feature on iPhone that came with the iOS 15.2 release. This research tested 200 applications spanning 20 app categories and kept the permission for apps to request tracking turned off. Each app was downloaded and opened only once, with the only interaction being to decline notifications.

The findings from different studies show that many applications contact third-party domains that are used for advertising and tracking [28], [29], [30]. In the research by Vallina-Rodriguez *et al.* [28], 446 third-party domains were identified, with 60% of them linked to at least one Advertisement and Tracking Service (ATS), and 20% connected to at least five ATS services. Categories such as social media, news, and games were most frequently linked to ATS services [28].

Additionally, Klais [29] found that iOS applications contacted 1,100 domains, with each app typically reaching out to 15 domains on average, 12 of which were unfamiliar third-party domains. One challenge we face today when it comes to third-party domains, specifically ATS services, is the lack of a comprehensive overview and classification of tracking domains, making it challenging to categorize domains that serve ATS purposes [28].

An intriguing discovery was made in a study performed by Paci *et al.* [30] consisting of 400 popular applications, evenly divided with 200 applications tailored for iOS and the remaining 200 optimized for Android. Findings showed that 50% of the examined applications communicate with third-party tracking domains despite the users not consenting to be tracked. This is a violation of the General Data Protection Regulation (GDPR) and ePrivacy Directive (ePD) requirements regarding valid consent. These privacy regulations stipulate that "consent should be obtained prior to collecting and processing data" [30]. Furthermore, Paci *et al.* also indicate a higher non-compliance rate among iOS applications than their Android counterparts.

Paci *et al.* [30] reveal that the most frequented third-party tracker domain for iOS was *inappcheck.itunes.apple.com*, accessed by 35,5% of the 200 iOS applications. Closely following behind, they found *app.measurement.com* (alphabet analytics services), whereas *graph.facebook.com* (Facebook social network) was in third place.

Cross-site platform tracking [5], [28] is also a method used by ATS services. This is highlighted in the findings from the studies by Vallina-Rodriguez *et al.* [28]. They revealed that 68,5% of the analyzed applications were linked to at least one website listed among the Alexa top 1000 most visited websites. This provides insight into how much user information is shared with other applications and shows the relevance of this thesis's research on tracking domains.

3.4 Privacy labels

Privacy labels, also known as nutrition labels, were introduced by Apple in 2020 [31]. This initiative aimed to enhance user awareness regarding application privacy practices. However, these privacy labels are self-declared by the given app's developer and lack Apple's and other verification [4], [31]. Consequently, developers could claim their app doesn't collect user data, even if it does.

Koch *et al.* [4] conducted a study examining 1,687 iOS applications from the German App Store to assess their compliance with stated privacy labels and adherence to GDPR. The findings from this study align closely with the study conducted by Ali *et al.* [31], where they utilized a privacy label tool to cross-reference 515,920 privacy policies with their corresponding privacy labels, to evaluate their consistency. Both of these studies discovered divergences regarding privacy labels. Koch *et al.* discovered that 48,87% of all apps on the German App Store were missing privacy labels [4]. Furthermore, the findings indicate that Games was the category most frequently collecting data for tracking purposes, which aligns with the study that Vallina-Rodriguez *et al.* [28] conducted. Additionally, Koch *et al.* found that the category Photo and Video was prone to collecting the MAID [4].

One significant discovery that Ali *et al.* [31] found was the frequent mismatch between the labels and the policies, particularly concerning data linked to users. Another notable finding was the disparity between app claims of not collecting user data and their actual practices. Furthermore, the researchers analyzed the network traffic of 30 apps to determine if the information gathered differed from what was declared in the privacy policies and labels, and the answer was yes.

The previous work on privacy labels indicates that Apple needs to revise its procedures concerning privacy labels. Both studies, Koch *et al.* [4] and Ali *et al.* [31], consider the potential of deploying an automated tool to review each privacy label within the App Store, thus assisting developers crafting precise labels. While this could potentially improve the situation, it has been observed through various studies that numerous applications fail to disclose their use of data for user tracking, and some even falsely claim not to collect any data although they do.

Several studies on privacy labels paint a clear picture; while Apple may need to enhance its standards in this regard, developers also bear a responsibility to inform users accurately. It's alarming that many applications intentionally omit this crucial information, raising concerns for users. Consequently, the privacy label may provide a false sense of security regarding the type of data collected by these applications.

3.5 Summary

Throughout this chapter, we have detailed the previous work related to the topics covered by the thesis: advertising intelligence, third-party tracking domains, and privacy labels. Notably, most of these studies have been conducted on Android. This section gives a summary of the previous work related to these topics. A summary of which literature aligns with which topic is illustrated in Table 3.3.

Table 3.3: Summary of articles used in our literature research

Research subjects	Research articles
Advertising intelligence	[1][12][26][5]
Third-party tracking	[28][29][30]
Privacy labels	[28][4][31]

Advertising intelligence

ADINT leverages the online advertising network to gather sensitive information by purchasing advertisements. Studies by Vines *et al.* [1], Govindaraj *et al.* [26], and others show that ADINT can reveal various forms of personal information, including location, app usage, and details about religious beliefs or sexual orientation. This can lead to significant privacy concerns, which could be exploited by people with malicious intent.

Third-party tracking

Many applications rely on third-party providers, and significant portions of these providers gather user information, often without clear consent or transparency. Studies by Vallina-Rodriguez *et al.* [28] and others have identified several tracking domains and found that a significant percentage of apps communicate with them, violating GDPR and ePD requirements. This is especially prevalent in categories like social media, news, and games, highlighting the need for stricter privacy measures. Additionally, Koch *et al.* [4] found that the category Photo and Video collected the MAID most frequently.

Privacy labels

Privacy labels, introduced by Apple in 2020, aim to increase user awareness about app privacy practices. However, studies by Koch *et al.* [4] and Ali *et al.* [31] reveal inconsistencies between app privacy labels and actual practices, with many apps failing to disclose their tracking activities accurately. This suggests a need for both Apple and app developers to improve transparency, ensuring users clearly understand how their data is used.

Contributions and inspiration

Our thesis is inspired by the categories and types of PII defined by Hu *et al.* [5] in their research, where they conducted a similar experiment to ours but on Android phones. Vines *et al.* [1] assessed the necessary resources for executing a successful ADINT campaign, which was relevant to our research, especially in terms of obtaining IDFA by intercepting network traffic.

Our thesis also drew inspiration from the study performed by Koch *et al.* [4] who, like Govindaraj *et al.* [26], performed jailbreaking on iPhones. This gave us an understanding that jailbreaking an iPhone was essential to circumvent Apple's restrictions and security protocols. Additionally, Koch *et al*. provided us with the insight that mitmproxy was an effective tool for intercepting traffic. This study also offered valuable insights into privacy labels and their functionality.

Similar to Klais [29], Koch *et al*. [4] used a no-touch method when analyzing the apps. Additionally, their methodology involved running each app for only one minute without running the apps in the background. Our thesis seeks to fill these gaps by investigating PII transmission during 15 minutes of use.

In addition to Koch *et al*. [4], our work was influenced by Ali *et al*. [31], who performed an experiment similar to ours. Their research not only examined the alignment between privacy policies and privacy labels but also analyzed network traffic from various applications to compare it with the information provided in the privacy policies and privacy labels.

The findings from Paci *et al*. [30] indicated a higher non-compliance rate among iOS applications than their Android counterparts. This is particularly relevant since we focus exclusively on iOS applications. Although they specifically examine compliance with ePD and GDPR, this is relevant to our work as it relates to the transparency of various applications regarding tracking practices.

All of these studies have inspired our work to achieve the desired results and provided a valuable framework for our investigation.

Chapter 4

Methodology

This chapter outlines the methodology used to address our research questions, briefly describing the experimental concept and justifying the chosen approaches. First, we outline our experimental setup, before detailing the application selection method and the dataset preparation steps. Lastly, each research question's experimental design is further explained, including the methods used, their limitations, and expected results.

4.1 Experiment outline

Depending on the specific targets that ADINT operators aim to identify, the targeting criteria offered by advertising networks can directly enable the identification of these targets [1]. We aim to investigate how this information is gathered and distributed to various tracking domains in order to understand how data collected by mobile apps facilitate ADINT activities.

We conducted an experiment to investigate the collection of PII, including the IDFA, and determine where this traffic goes. Although our experiment involved setting up a proxy to monitor traffic from an Apple mobile device, it functionally resembled a MiTM attack, where typically attackers intercept communication between two legitimate parties [32].

In our experiment, this translates to capturing and analyzing the requests made by different applications and the responses they are getting back. As illustrated in Figure 4.1, our conceptual representation of a MiTM attack, we generated traffic from an Apple mobile device (target) and tunneled it through an experimental laptop (MiTM adversary). Based on our observations regarding PII collection and tracking domains in the captured traffic, we examined the apps' privacy labels to assess whether they accurately disclosed their tracking practices. Despite the differences from a traditional MiTM attack, we will continue to use the term to describe our methodology due to its similarities.

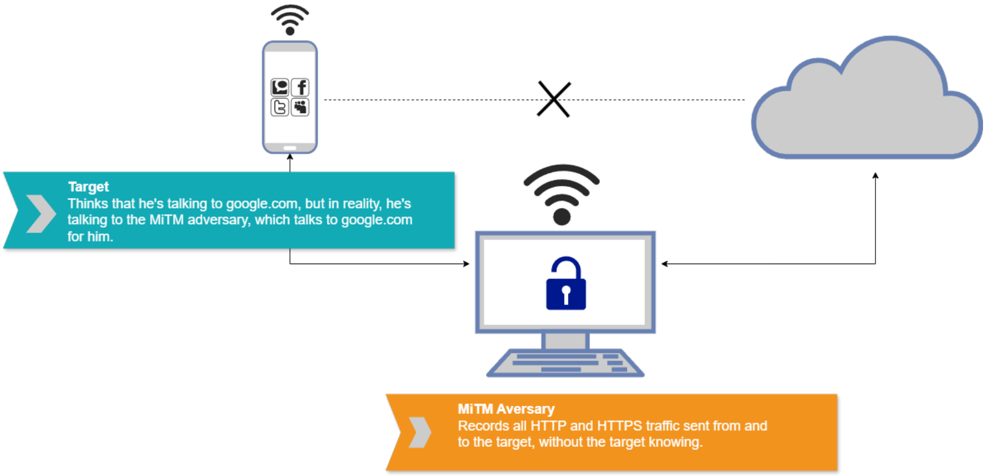

Target
Thinks that he's talking to google.com, but in reality, he's talking to the MiTM adversary, which talks to google.com for him.
MiTM Aversary
Records all HTTP and HTTPS traffic sent from and to the target, without the target knowing.

Figure 4.1: Conceptual representation of a MiTM attack

Before settling on our experimental approach, we evaluated three methods: static reverse engineering, dynamic reverse engineering, and network traffic inspection. Static reverse engineering involves analyzing the application's code without executing it, providing basic insights into its functionality and sometimes network signatures [33]. In contrast, dynamic reverse engineering involves observing the app's behavior during execution, usually through the help of a debugger, and is most useful when trying to obtain information that is difficult to gather with static techniques [33]. In general, when reverse engineering software, it is necessary to use a variety of techniques and tools to see the entire picture [33]. Additionally, obfuscation techniques in an app binary could challenge the work and impact the results. This complicates the process and is not particularly efficient when investigating a range of apps.

With this in mind, we found inspecting network traffic the most suitable method for our experiment. This approach has several advantages, such as capturing user-specific input and dynamically downloaded code in real time, which are crucial for a comprehensive analysis. Inspecting the network traffic directly from the iPhone precisely depicts where the traffic goes without further complication. Furthermore, using specialized tools to intercept network traffic is an effective strategy for our objectives, allowing us to analyze more apps than what would be feasible through reverse engineering alone.

To build the experiments dataset, we selected a set of applications to investigate. To ensure compatibility with the experiment iPhones and a broad selection of apps across categories, we defined a method for application selection, as described in Section 4.2. The applications derived from this method are listed in Table 5.4 and discussed in Section 5.1.1.

Once the applications were selected, we ran them on the iPhone and intercepted the traffic through a specialized tool. The tool used in our experiment was mitmweb, a tool from the mitmproxy suite that works as a web-based GUI that can intercept, decrypt, and display proxied traffic from other devices. In our experiment, mitmweb facilitates the MiTM attack, enabling us to capture the IDFA and get insight into collected PII and contacted domains.

To conduct the analysis, we used Splunk, a tool that specializes in log analysis. Splunk enables the execution of queries that can identify desired data in the dataset. After analyzing the traffic in Splunk, we could check whether the privacy labels of the relevant apps were transparently declared.

The above briefly outlines the experiment conducted and how it intends to reach our main objectives. A further description of how the experiment will answer each research question is detailed in Section 4.5. Additionally, the details of the experiment execution, its results, and result discussions are described in Chapter 5.

4.2 Application selection

This section describes the method used to determine which applications to analyze. The method was used to navigate the complexity of the Apple App Store and filter through the multitude of available applications while ensuring a diverse selection of *popular* applications. By *popular*, we refer to applications ranked on the top list within their category, indicating a high number of downloads and positive user reviews in Norway at the specific time period of the experiment.

The method consisted of three steps: 1) category grouping into thematic groups, 2) applying selection and exclusion criteria, and 3) ensuring equal distribution within each thematic group. These steps are further explained throughout this section.

1. Category grouping into thematic groups

The App Store organizes applications into 28 distinct categories, which we further categorized into thematic groups. The thematic grouping included a systematic organization of the 28 available categories into broader thematic groups, as demonstrated in Table 4.1. The reclassification optimized the selection by grouping related categories into broader themes and ensuring wider category representation.

The Apple Watch category was excluded due to its specific device and platform requirements. AR-Apps and Developer Tools were also excluded, as they lack top lists in their categories, which is essential for this method. This results in 25 categories divided into five thematic groups.

Table 4.1: Thematic grouping of App Store categories

Thematic group	**AppStore categories**
Life and wellness	Health and Fitness Medical Food and Drink Lifestyle
Education and information	Education News Books Magazines and Newspapers References
Entertainment and Creativity	Music Entertainment Photo and Video Sports Kids Games
Productivity and Business	Business Tools Productivity Finance Graphics and Design Safari Extensions Shopping
Social and Mobility	Social Networks Travel Navigation

2. Applying selection and exclusion criteria

Next, selection and exclusion criteria were applied to the apps in each thematic group. The application selection criteria focused on compatibility and availability, while the exclusion criteria removed any apps requiring additional subscriptions or products. This was necessary because apps that demand legitimate customer verification, such as bank apps that require signing in with BankID to confirm customer relationships, restrict our ability to interact with the app without sharing personal sensitive information. This would significantly impact our experiment in terms of hindering interactive use, which is a key aspect of our research. The selection and exclusion criteria are described below.

Selection criteria:

a) High ranking, indicating popularity: Ranked among the top 15 free applications within its category.

b) Compatibility: Require a version compatible with iOS 12.5.7. This ensures all selected apps can be installed and run on the experiment device(s).

Exclusion criteria:

a) External device dependency: The application functionality is contingent upon connecting to an external device.

b) Requires legitimate customer relationship: The application's primary functionality depends on a legitimate customer relationship, e.g. BankID.

3. Ensuring equal distribution within each thematic group

We selected five applications from each thematic group to ensure an equal distribution of apps across the groups, resulting in 25 applications in total. An equal distribution within each thematic group was desirable as it allows for more representative comparisons between the results from each thematic group. Comparing three apps from one thematic group against seven in another would skew the results, making it difficult to draw fair and accurate conclusions between the groups.

To manage this, the applications were selected based on the highest ranking within their respective category in the App Store. This means that if ten apps were initially included in the thematic group, based on the selection and exclusion criteria, the five apps with the overall highest ranking would be selected, whereas the lowest would be excluded. This resulted in the applications described in Section 5.1.1.

4.3 Dataset

The experiment is based on two datasets: one containing the network traffic from the 25 selected applications and one containing tracking domains. This section describes how the datasets were generated, detailing their size, composition, and sources.

Dataset containing network traffic

The dataset containing the network traffic from the 25 applications was generated by tunneling the network traffic through mitmweb, as described in Chapter 5. Each application was run separately and for an equal amount of time, in addition to being the only active app on the device, ensuring equal treatment and minimal traffic from other applications.

We determined the duration of the traffic capture to be 15 minutes per app. The initial ten minutes included active user interaction, whereas the app was run in the background in the five remaining minutes. This allowed for continuous engagement with the app during the interactive phase and enough time to see whether the network activity from the app continued to flow in the background.

Based on this, the network traffic dataset amounted to six hours and 25 minutes. It comprised 448,5 MB of network traffic in total, averaging 17,94 MB per app. When uploaded to Splunk, a total of 11,877 network events are displayed, averaging 475 events per app. A summary of the key aspects of the dataset is listed in Table 4.2.

Table 4.2: Key aspects of network traffic dataset

Key aspect	Value
Number of apps	25
Time on each app	15 minutes
Total duration	Six hours, 25 minutes
Avg. size of capture per app	17,94 MB
Total size	448,5 MB
Avg. number of events per app	475
Total events	11877

Dataset containing tracking domains

For the dataset containing traffic domains, we searched through literature to compile a list of previously found top third-party tracking domains. The articles containing tracking domains included *A Comprehensive Study on Third-Party User Tracking in Mobile Applications* [30], *Keeping privacy labels honest* [4], *New Research Across 200 iOS Apps Hints that Surveillance Marketing is Still Going Strong* [29] and *Tracking the Trackers: Towards Understanding the Mobile Advertising and Tracking Ecosystem* [28].

From these articles, we were able to define a comprehensive list of tracking domains. When summarized, the list contained 181 tracking domains when compiled. A summary of each article's contribution to our compiled list is given in Table 4.3.

Table 4.3: Articles and their corresponding number of tracking domains

Article	**Count**
New Research Across 200 iOS Apps Hints that Surveillance Marketing is Still Going Strong	100
Keeping Privacy Labels Honest	40
Tracking the Trackers: Towards Understanding the Mobile Advertising and Tracking Ecosystem	25
A Comprehensive Study on Third-Party User Tracking in Mobile Applications	16
Total:	181

4.4 Preprocessing

Both datasets had to be preprocessed before the analysis could be conducted. This section provides a detailed description of the steps taken to prepare them for further examination.

Preprocessing the network traffic

Although mitmweb excels in intercepting traffic, its analysis capabilities are somewhat limited. For instance, mitmweb lacks statistical visualization and PDF export features, which are necessary to provide results that are easy to interpret and can be documented. Therefore, we chose to use Splunk as an alternative analysis tool. In general, capturing traffic in one tool and analyzing it in another can be challenging, mainly due to variations in data formats and tool capabilities. For this reason, the data derived from mitmweb had to be preprocessed to facilitate the transfer from mitmweb to Splunk.

When capturing traffic with mitmweb and saving the output, the data is in a non-human-readable format by default. This makes analysis challenging, as the analysts won't be able to interpret the results, necessitating translation into a human-readable format. To do so, we used mitmdump, another tool from the mitmproxy-suite, setting the parameter *flow_detail* to level three.

In addition, in the traffic flow files, some events lacked timestamps entirely or had inconsistent timestamp formatting, a crucial element for proper indexing in Splunk. Although Splunk can read various formats, it specializes in log files, where timestamps help track incident timelines and index events correctly. To address this, we utilized a Python-based mitmproxy addon script obtained from *Use mitmdump to Capture Refinitiv Real-Time - Optimized Content* by J. Phuriphanvichai [34], as documented in Appendix B. This add-on was utilized to append relative timestamps to each event based on the script's execution time. Precise timestamps were unnecessary for our experiment as we focused on packet content, not sequence.

The translation to human-readable format and appending the relative timestamps were executed in a single command. The command used was as follows:

```
 mitmdump -nr ~/Downloads/flows-file --set flow_detail=3
| python3 py_timestamp.py file_timestamp.txt
```

Preprosessing the tracking domains

Some duplicates occurred as the list containing tracking domains was compiled from four different sources, as previously described in Table 4.3. The command below removed these, leaving 154 domains.

```
sort -u raw_domains.txt > unique_domains.txt
```

After executing the command, we observed that the dataset still included domains that looked similar, such as *amazon-adsystem.com* and *amazon-adsystem.com*, along with *google-analytics.com* and *google-analytics.com*. Despite appearing similar, these domains differed in their use of Unicode characters. Specifically, *amazon-adsystem.com* used a hyphen (Unicode character U+2010), while *amazon-adsystem.com* used a minus (Unicode character U+2212). Since only the hyphen is permitted within domain names [35], we excluded the domains containing a minus. This refinement resulted in 152 unique domains.

Next, we filtered out the domains we knew served broader purposes beyond tracking, analytics, or advertising. This step aimed to give an overview of the dataset and bring us closer to pinpointing specific tracking domains rather than just third-party domains. For instance, the domain *google.com* primarily serves as a search engine, which is not specifically used for facilitating advertising, analytics, or tracking and is, therefore, too broad to be classified as a tracking domain in our research. However, Alphabet, the company owning Google, has many other domains, like *googleadservices.com*, specifically used for advertising. The same applies to domains like *yahoo.com* and *amazonaws.com*. As their primary functions are broader than specifically facilitating tracking, analytics, or advertising, these domains were filtered out. A total of six domains were filtered out in this step, leaving us with 146 potential tracking domains that needed further verification.

The remaining domains had to be verified through further research. To do so, we checked all domains using VirusTotal, an online service that provides category tags such as tracking, advertising, and analytics based on aggregated data from various antivirus engines and website scanners [36]. Domains tagged as tracking, advertising, or analytics were deemed verified and automatically became part of the final list. After checking each domain in VirusTotal, the list comprised 70 verified domains, leaving 75 domains needing even further analysis.

Finally, we searched the internet to analyze the remaining domains further. This included using search engines, verdict websites like AlienVault and Netify, and the tracking domains' web pages, if available. Through this process, we verified an additional 40 domains as tracking, analytics, or advertising facilitators, filtering out 35 domains from our initial list.

After removing duplicates and verifying that the domains were associated with tracking, advertising, or analytics, the final list comprised 111 tracking domains. Summarized, 76% of the initial domains were proved to be valid tracking domains through our preprocessing steps. The process and the specific domains relevant to each preprocessing step, including the final list, are documented in Appendix C.

4.5 Experimental design

The following section describes how the thesis will answer the defined research questions.

4.5.1 Personally Identifiable Information collected by iOS applications

This research question investigates what PII is collected by the selected applications during 15 minutes of use. The 15 minutes are divided into ten minutes of interactive use, followed by five minutes of background running. The extent of the PII collection, especially alongside the IDFA, will indicate how the application contributes to user tracking, and thereby enabling ADINT.

The 10 minutes of active use aim to trigger and disclose as much data collection as possible. Using the app interactively for a certain period of time helps portray a realistic view of a person's potential use and, thereby, a realistic amount of collected data. Interactive use includes signing up for accounts, if applicable, and utilizing core application functionality like querying searches, scrolling through content, and pushing different buttons.

The five minutes of background running aim to check whether the traffic from the app continues to flow even though the app is running in the background. This aims to portray how aggressive the tracking is. To answer this, we press the home button on the mobile device, and carefully monitor the traffic in mitmweb for five minutes directly after completing the ten initial minutes.

The data will be collected by tunneling network traffic from an iPhone, where the app is run, to a MacBook, which will decrypt and display it in mitmweb. The traffic will be actively monitored in mitmweb while the capture is ongoing before exporting it to Splunk for further analysis. The analysis looks for the specific PII-values outlined in Table 4.2. The PII values will be identified through Splunk Search Processing Language (SPL) queries that search for a specific PII value, e.g., the IDFA, and highlight the string if present. By doing so, the findings disclose whether or not a certain PII has been collected.

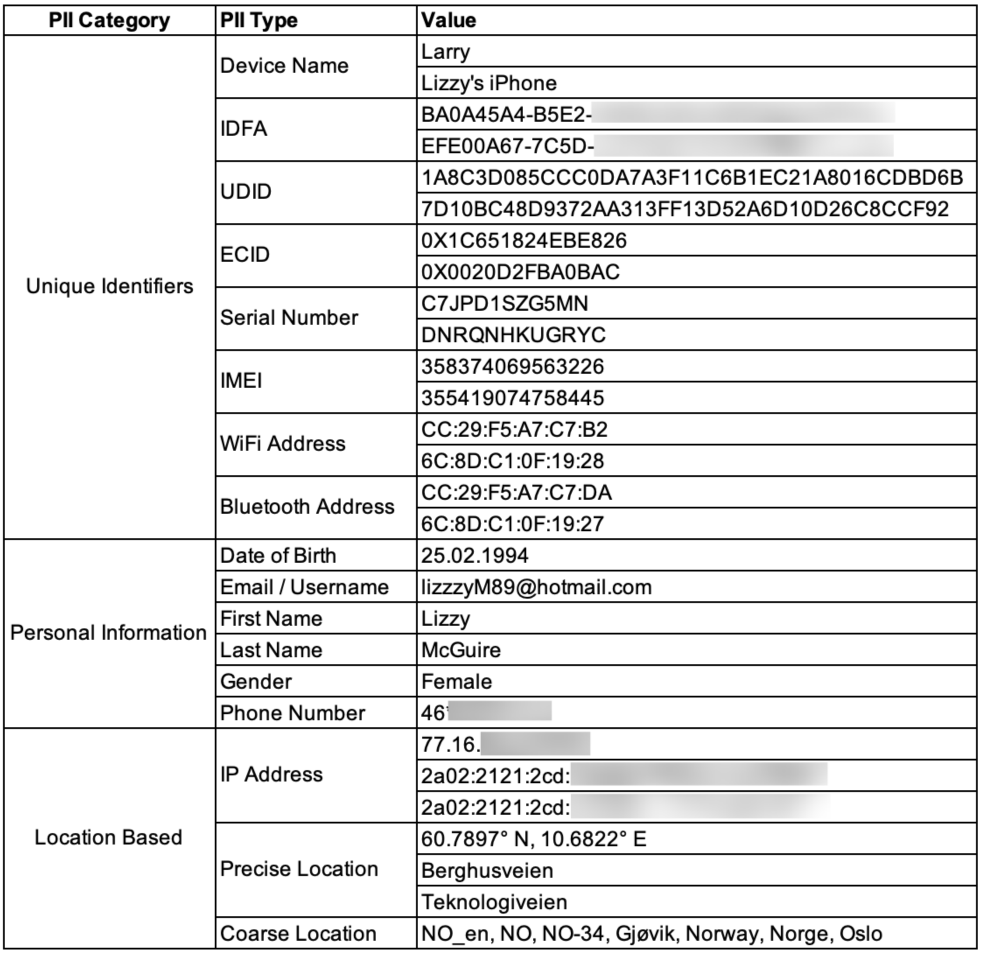
Limitations

This research question investigates a limited amount of PII and may, therefore, not fully depict all collected PII. For instance, age is a key parameter in targeted advertising; however, it is not included in our research. The main reason for this is that it is difficult to identify a string, e.g., "30", particularly in the context of age, when the dataset contains several strings consisting of these two digits in several different contexts. Such strings can be session tokens, Unix timestamps, dates, and connection details, just to mention a few. In addition, we have excluded personal numbers and financial information, as this would expose the team members' personal information in the experiment, which is undesirable due to privacy concerns.

In the case of encrypted data in the traffic flow, we will not attempt further decryption than the capabilities available through mitmweb. Some app developers implement security and confidentiality measures, like encryption, to make attacks like ours challenging. Mitmweb has some built-in decryption capabilities, but it's not given that all selected apps only utilize these. Therefore, our analysis is limited to decryption capabilities only available through mitmweb.

Expected results

First, we expect unique identifiers, such as UDID, serial number, and IDFA, to be frequently collected as these can uniquely identify a device. This expectation is based on NRKs investigation [12], in which we found that the word *device_id* was used to identify the devices in their purchased dataset. NRK describes this field as a unique identifier used to identify the device that had been tracked. However, the contents in this column were hashed or encrypted, meaning we could not identify



PII Category	PII Type	Value
Unique Identifiers	Device Name	Larry
		Lizzy's iPhone
	IDFA	BA0A45A4-B5E2-[redacted]
		EFE00A67-7C5D-[redacted]
	UDID	1A8C3D085CCC0DA7A3F11C6B1EC21A8016CDBD6B
		7D10BC48D9372AA313FF13D52A6D10D26C8CCF92
	ECID	0X1C651824EBE826
		0X0020D2FBA0BAC
	Serial Number	C7JPD1SZG5MN
		DNRQNHKUGRYC
	IMEI	358374069563226
		355419074758445
	WiFi Address	CC:29:F5:A7:C7:B2
		6C:8D:C1:0F:19:28
	Bluetooth Address	CC:29:F5:A7:C7:DA
		6C:8D:C1:0F:19:27
Personal Information	Date of Birth	25.02.1994
	Email / Username	lizzzyM89@hotmail.com
	First Name	Lizzy
	Last Name	McGuire
	Gender	Female
	Phone Number	46[redacted]
Location Based	IP Address	77.16.[redacted]
		2a02:2121:2cd:[redacted]
		2a02:2121:2cd:[redacted]
	Precise Location	60.7897° N, 10.6822° E
		Berghusveien
		Teknologiveien
	Coarse Location	NO_en, NO, NO-34, Gjøvik, Norway, Norge, Oslo

Figure 4.2: Categories, types, and values of PII

which type of unique identifier this specifically was. The term *device_id* is broad and can include different types of unique identifiers, which is why we expect a combination of these identifiers to appear in our collected traffic.

We expect to find IDFA in apps that, as far as we know, don't have a primary income besides the app. App developers range from private individuals to nation-states, with different needs regarding monetizing the app through advertising. Unlike private developers who might include ad libraries in their apps to generate additional revenue through ads, nation-state-operated apps, such as apps for emergency services, do not typically require ad monetization due to their funding through the government.

From the personal information category, we expect gender and email to appear in the traffic. Gender helps with targeted advertising, as many products are specific to men or women. Email is relevant as it contains a value that can be used for cross-platform tracking.

Furthermore, we anticipate finding location-based data among the most frequently sent PII, also based on NRKs findings [12]. Along with the *device_id*, *longitude* and *latitude* also appeared in the NRKs dataset. We therefore expect to find precise locations in our collected traffic. We also expect the coarse location to be found in numerous events, as the language code is included as part of the coarse location PII type. This is because the *NO* in the language code *en_NO* is derived from a user's country. As the language code is often used to provide core app functionality, it might be sent numerous times, contributing to many hits in contexts besides tracking. We also expect the IP addresses to be found, as this also discloses a user's location.

Lastly, we anticipate finding PII transmission mainly generated by the users themselves during the interactive phase and that the traffic does not continue flowing after the home button is pressed and the app runs in the background. Minimal background traffic is expected due to background updates being turned off in settings, as described in Section 5.1, in addition to the responses the apps could receive when requesting access, being somewhat restricted, as described in Section 5.1.2.

4.5.2 Frequently contacted tracking domains

To answer the research question regarding which tracking domains are frequently contacted, we continue analyzing the network traffic using Splunk. Essentially, the difference between this question and the previous one is that the queries executed in Splunk look for different strings. This means that while the previous queries executed in the previous question search for our defined PII values, the queries in this question examine the URLs in the HTTP/HTTPS header of the requests and responses.

Based on our list of 111 tracking domains, as listed in Appendix C, we develop and execute a query that searches through the traffic, looking for matches on the given domains. If a match is made, the count on the given domain increases by one, ultimately finding the most frequently contacted tracking domains.

Limitations

Our list of tracking domains is based on the top tracking domains identified in four research articles, which restricts our analysis to 111 domains after preprocessing. Running queries in Splunk based on this predefined list might result in false negatives, as other tracking domains may occur undetected. Thus, our results are limited to only those domains included in our list.

Expected results

We expect Alphabet-owned domains such as *googleadservices.com* and *google-analytics.com* to dominate our results of the most frequently contacted tracking domains. This is because companies like Alphabet have several roles in the advertising ecosystem, such as buyers and sellers of ad spaces and providers of data analysis. This makes Google the most powerful player in the advertising ecosystem [6]. Previous research by Paci *et al.* [30], which found Alphabet analytics services and Facebook social network to have a frequent presence and are among the top domains in their studies, supports our expectation that these platforms will feature prominently in our findings.

Additionally, we anticipate encountering Meta-owned domains such as *graph.facebook.com*, reflecting Facebook's extensive data collection practices. This domain was also identified as one of the most accessed by the iOS applications in the study by Paci *et al.* [30]. The Cambridge Analytica scandal illustrated the significance of Facebook's data practices in 2016, where user data was exploited to create voter profiles for the Trump campaign [37]. Given Facebook's proven capability to influence user behavior significantly, it is clear that they manage a detailed data repository. Following Facebook, we anticipate that Appsflyer domains will be prominently featured, with additional notable traffic to Branch and smaller advertising specialists, as previously documented in research by Vallina-Rodriguez *et al.* [28] and Paci *et al.* [30].

4.5.3 Privacy label tracking transparency

When an application collects PII and transmits this data to third parties, this practice must be clearly disclosed in Apple App Store privacy labels. Our research aims to assess how this data collection and sharing process is transparently declared to users, specifically regarding the types of data collected for tracking purposes.

To assess the transparency of tracking in privacy labels, we examined apps collecting IDFA, which serves as the most obvious indicator of third-party tracking, as it is directly linked to advertising. Our evaluation is based on findings from research question one, identifying which apps collect IDFA.

While reviewing the privacy labels of these applications, we specifically focused on how tracking activities were declared and the collected data types. Subsequently, we compared this information against Apple's definitions and standards for app developers submitting or managing apps in the App Store. This comparison helps us assess the accuracy of privacy labels concerning their tracking declarations and determine whether these labels are incomplete or misleading about the purposes for which the data is used.

Limitations

When examining the privacy labels, the privacy type *Data Not Collected* is excluded as it explicitly indicates that no data is gathered, leaving no further details to investigate within the privacy label. Concerning the purposes for which data is utilized, our analysis is specifically limited to *Third-Party Advertising*, as this category is uniquely associated with tracking activities. Furthermore, our examination is restricted to the data type *Device ID*, as it is the sole data type definitively implicated in tracking operations.

Furthermore, our research emphasizes privacy labels and does not extend to privacy policies. As a result, we may overlook additional contexts and details provided in the full privacy declarations.

Expected results

The outcomes of this research question depend on the results from research question one. Based on our expected findings regarding the previous questions, we anticipate the results of this experiment will show that privacy labels may not fully reflect the actual practices of collecting and sharing PII for tracking purposes. Based on the data identified through the initial research questions, this suggests that we might uncover privacy labels that are not sufficiently transparent.

Existing research on privacy labels suggests that not all applications provide complete transparency [31]. Prior studies also indicate the possibility of misleading labels [4]. This suggests that while some applications may provide clear and accurate information, others could obscure or distort the true extent of data usage and privacy impacts.

Chapter 5

Experiment

The following chapter provides a detailed description of the experimental work conducted as part of this study. It includes a description of the environmental setup and experimental design, followed by a detailed step-by-step description of the experimental procedures. In addition, the chapter presents the execution and results, followed by a discussion of the findings regarding each research question. The experiment is described in such detail to ensure that it can be replicated by other researchers in the future.

5.1 Environment setup and experimental design

The experiment was structured into three phases, each requiring a unique environmental setup, as illustrated in Figure 5.1. The figure provides a visual representation of the devices used and the software installed on them, respectively. As illustrated, the first phase of the experiment included capturing the network traffic by conducting the MiTM attack. The second phase included analyzing the captured traffic to find the answers to research questions one and two. As part of the results from research question one, we will find which apps have collected the IDFA. These apps will serve as the data foundation in the third phase, which examines the privacy labels of the respective apps.

As outlined in Section 4.1, the equipment utilized for the first phase of the experiment - the MiTM attack - consisted of an Apple mobile device and a laptop. Specifically, we used a jailbroken iPhone 6 and a MacBook Pro. Additionally, we provided a jailbroken iPhone 6s as a backup device in case the primary device encountered technical issues such as unexpected crashes or network delays. The backup device turned out useful, as seven out of the 25 apps encountered such issues during the experiment execution.

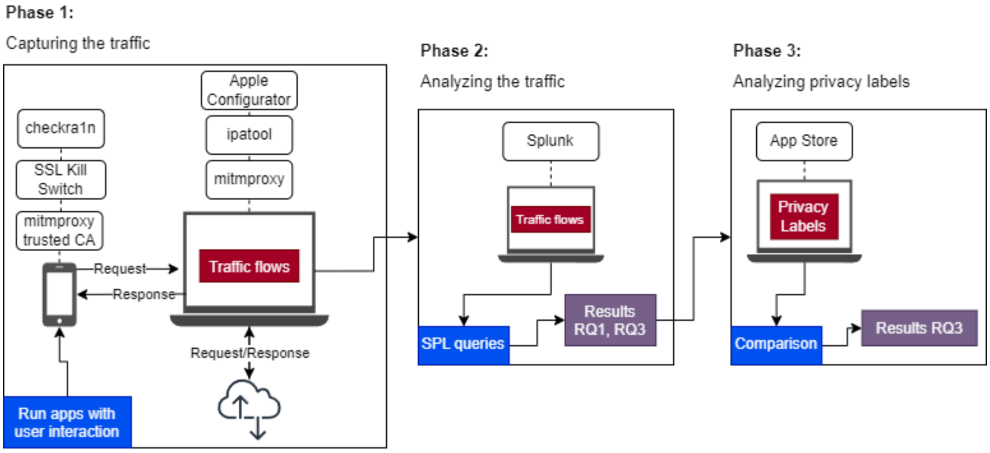


Figure 5.1: Environment setup and experiment design

The traffic analysis, i.e., the second phase, was conducted in Splunk on a Windows 10 Virtual Machine (VM). For the final phase of the experiment, we used the MacBook to assess the apps' privacy labels in the App Store. The devices and their specifications are summarized in Table 5.1 and 5.2.

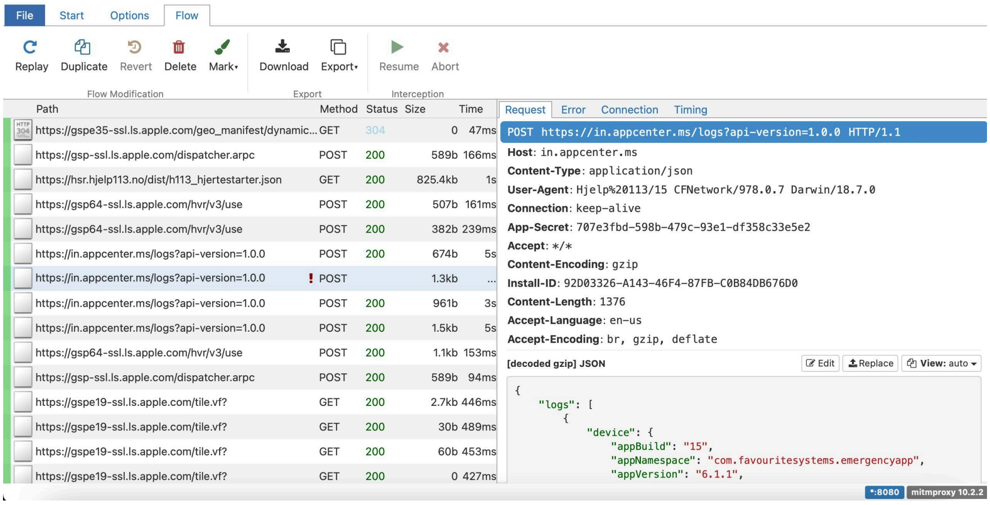
Table 5.1: Experiment laptops specifications

	MacBook Pro (Retina, early 2015)	Windows 10 VM
Operating system	macOS Monterey 12.7.4	Windows 10
Processor	2,7 GHz Intel Core i5	5 CPUs
Memory	8GB 1867 MHz DDR3	2048 MB
Graphics	Intel Iris Graphics 6100 1536 MB	VBoxSVGA

Table 5.2: Experiment iPhones specifications

	iPhone 6	iPhone 6s (backup)
Operating system	iOS 12.5.7	iOS 15.8.1
Processor SoC	Apple A8	Apple A9
Memory	1GB	2GB
Graphics	PowerVR GX6450 GPU	PowerVR GT7600 GPU
Capacity	16GB	64GB
Jailbreak	checkra1n	palera1n

To prepare the MacBook for phase one, we first installed mitmproxy, a comprehensive suite for HTTP and HTTPS proxying that includes the mitmweb, mitmdump, and mitmproxy tools. The main advantage of using mitmproxy was that the web-based GUI provided by mitmweb was intuitive to use and displayed data such that it was easy to interpret. An example of what mitmweb looks like when



Path	Method	Status	Size	Time
https://gspe35-ssl.ls.apple.com/geo_manifest/dynamic...	GET	304	0	47ms
https://gsp-ssl.ls.apple.com/dispatcher.arpc	POST	200	589b	166ms
https://hsr.hjelp113.no/dist/h113_hjertestarter.json	GET	200	825.4kb	1s
https://gsp64-ssl.ls.apple.com/hvr/v3/use	POST	200	507b	161ms
https://gsp64-ssl.ls.apple.com/hvr/v3/use	POST	200	382b	239ms
https://in.appcenter.ms/logs?api-version=1.0.0	POST	200	674b	5s
https://in.appcenter.ms/logs?api-version=1.0.0 !	POST		1.3kb	...
https://in.appcenter.ms/logs?api-version=1.0.0	POST	200	961b	3s
https://in.appcenter.ms/logs?api-version=1.0.0	POST	200	1.5kb	5s
https://gsp64-ssl.ls.apple.com/hvr/v3/use	POST	200	1.1kb	153ms
https://gsp-ssl.ls.apple.com/dispatcher.arpc	POST	200	589b	94ms
https://gspe19-ssl.ls.apple.com/tile.vf?	GET	200	2.7kb	446ms
https://gspe19-ssl.ls.apple.com/tile.vf?	GET	200	30b	489ms
https://gspe19-ssl.ls.apple.com/tile.vf?	GET	200	60b	453ms
https://gspe19-ssl.ls.apple.com/tile.vf?	GET	200	0	427ms

```
POST https://in.appcenter.ms/logs?api-version=1.0.0 HTTP/1.1
Host: in.appcenter.ms
Content-Type: application/json
User-Agent: Hjelp%20113/15 CFNetwork/978.0.7 Darwin/18.7.0
Connection: keep-alive
App-Secret: 707e3fbd-598b-479c-93e1-df358c33e5e2
Accept: */*
Content-Encoding: gzip
Install-ID: 92D03326-A143-46F4-87FB-C0B84DB676D0
Content-Length: 1376
Accept-Language: en-us
Accept-Encoding: br, gzip, deflate
```

[decoded gzip] JSON

```
{
    "logs": [
        {
            "device": {
                "appBuild": "15",
                "appNamespace": "com.favouritesystems.emergencyapp",
                "appVersion": "6.1.1",
```

Figure 5.2: Mitmweb when capturing traffic

capturing an app's traffic is provided in Figure 5.2. As the image depicts, the sequential flow of the HTTP/HTTPS requests is given in the left pane. These requests can be individually inspected by clicking on the desired event. This opens a right pane containing information about the request, including the type of request, the contacted domain, and the decrypted data contained in the event. As illustrated, mitmweb, along with the other tools from the mitmproxy suite, is specifically designed for intercepting and inspecting HTTP/HTTPS traffic, which are the protocols we must investigate to get insight into the PII collected by the applications' and the contacted domains.

As illustrated in the figure, mitmweb decrypts HTTPS traffic straightforwardly and presents it in an easily interpretable output, unlike other network inspection tools like Wireshark. Although Wireshark is powerful, in our experiment, we target the content in the HTTP and HTTPS traffic only, and using Wireshark for this purpose would lead to unnecessary large amounts of detailed traffic, making the analysis more difficult and time-consuming. Furthermore, we have seen mitmproxy utilized in previous works by Koch. *et al*. [4] and Hu *et al*. [5], indicating a reliable tool safe to adopt in our research.

Secondly, we installed Apple Configurator, a tool exclusively available on MacOS that enables effective device management on Apple devices. Apple Configurator was used to install the apps on the mobile device efficiently.

Without Apple Configurator, apps have to be searched for, downloaded, and installed manually through the App Store on the iPhone each time. This is time-consuming in comparison to using a specialized tool. Additionally, given the experiment's duration over several days and the frequency of app updates, we would risk selected apps becoming incompatible before downloading and running them, necessitating the search for alternatives and unnecessarily increasing our workload. For this reason, we chose to download all applications beforehand, as described in the following paragraph.

To download the applications beforehand, we used ipatool, an open-source command line tool that can download apps in .ipa format from the App Store [38]. In combination with using Apple Configurator, a more efficient process for installing and removing apps from the experiment devices was facilitated. Four commands were utilized to retrieve the .ipa file of the desired app, repeated for each application.

The first command authenticates with the App Store using an Apple ID. The second command searches for the desired app based on the app name given as a parameter. This command is similar to searching directly for an app in the App Store, as it, without the option *–limit*, will list several apps that relate to the search. Therefore, we searched for the specific app name and limited the search results to only one hit. The app's bundle ID, which is a string that uniquely identifies an application within Apple's ecosystem, will be displayed as a result of the search command [39]. This string is used as a parameter for the following commands to specify the exact app to be purchased and downloaded. All apps used in our research were free; however, it is still necessary to execute the purchase command in case it has an initial cost. The last command downloads the app specified by the bundle ID, which results in the desired .ipa file. The commands executed are listed below.

Ipatool commands:

```
ipatool auth login --email "appleID" --password "password"
```

```
ipatool search --limit 1 "app name"
```

```
ipatool purchase -b "app.bundleID"
```

```
ipatool download -b "app.bundleID"
```

To protect team members' private information and prevent software licensing issues, we created a dummy Apple ID using fabricated personal details such as name, email address, and date of birth, as listed under the category *personal information* in Table 4.2. Additionally, we linked the dummy account to a team member's phone number, as this was a requirement for account activation. As described in Section 6.2, we have carefully considered privacy regarding this decision and obfuscated and redacted all sensitive information before including it in our thesis.

The Apple ID was used to sign in on the iPhones and authenticate with the App Store, both on the devices and through ipatool. It is crucial to use the same Apple ID when downloading the apps as when signing into the device they will be running on, as app licenses are linked to the Apple ID [40]. While downloading directly from the iPhone's App Store automatically ensures this alignment, our use of ipatool for pre-downloading the apps required careful coordination of the Apple ID. Additionally, as the Apple ID was provided with fabricated personal details, we used it to create accounts on the selected apps during our interactive phase of the traffic capture.

Regarding the iPhones, both experiment iPhones were jailbroken so that we could access and install the necessary tweaks for our experiment iPhones. Tweaks can be viewed as additional settings that can be downloaded and, once enabled, change how the device looks or behaves [41]. To access these tweaks, we need a package manager that can provide them, such as Cydia or Sileo, meaning the device must be jailbroken.

The tweaks necessary for our research included SSL Kill Switch 2 and SSL Kill Switch 3, along with their dependencies, including Debian Packager, Cydia Substrate, and PreferenceLoader. SSL Kill Switch was essential as apps affected by SSL pinning cannot be intercepted and decrypted by mitmweb. SSL pinning is a technique used by developers to mitigate MiTM attacks like ours by associating a server with a specific SSL/TLS certificate [42].

SSL Kill Switch disables SSL pinning on these apps, allowing the iPhone to recognize the MacBook as a Certificate Authority (CA). This facilitates our experiment by enabling even more extensive traffic capture. As we have not investigated whether any of the selected apps are subject to SSL pinning in advance, this step is deemed necessary in our research. However, if it were guaranteed that none of the apps were subject to the pinning, jailbreaking the device and installing SSL Kill Switch would not be necessary.

Given the differences in iOS versions and hardware between the primary iPhone and the backup iPhone, we adapted the software accordingly, using different jailbreak types and SSL Kill Switch versions for each device. For instance, the iPhone 6 is jailbroken with checkra1n, whereas the iPhone 6s uses palera1n, as listed in Table 5.2. This is due to jailbreaks essentially being vulnerability exploits, which implies that they depend on the vulnerabilities present in the device to be compatible [41]. Naturally, Apple will patch vulnerabilities between releases to improve security, making previous jailbreaks ineffective. Accordingly, new vulnerabilities appear, enabling new exploits to be developed and new jailbreaks to evolve. As different jailbreak types include different package managers, the available tweaks vary. For this reason, different versions of SSL Kill Switch were used on each device; SSL Kill Switch 2 was installed on the iPhone 6, whereas SSL Kill Switch 3 was installed on the iPhone 6s.

Since checkra1n is rootful and palera1n is not, we decided to use the iPhone 6 as the primary experiment device, substituting it with the iPhone 6s when technical issues occurred. Although the iPhone 6 had lower technical specifications than the iPhone 6s, rootful jailbreaks leverage more privileges, increasing the chance of bypassing potential SSL pinning on the app [43].

Furthermore, we deleted all unnecessary apps on the iPhones, as they could generate background traffic and poison the capture. However, some apps are unremovable, such as the Apple App Store. To minimize the chance of getting potential App Store traffic included, we took precautionary measures described in Section 5.1.2 and configured the device's settings to turn background app refresh and automatic downloads to off. Regarding the settings, we also allowed as much tracking as possible by ensuring *Limit Ad Tracking* was off and *Location Services* was on. The specific settings enabled and disabled, along with their default value, are summarized below.

Device settings:

1. **Privacy:**
   - Location Services → toggle Location Services to ON (default is ON)
   - Advertising → toggle Limit Ad Tracking to OFF (default is OFF)
   - Analytics → toggle Share iPhone Analytics to ON (default is ON)
   - Analytics → Share With App Developers to ON (default is ON)
2. **General:**
   - Background app refresh → OFF (default is ON)
3. **iTunes & App Store**
   - App store → toggle automatic downloads for apps to OFF (default is ON)
4. **Bluetooth:**
   - Toggle Bluetooth to ON (default is ON)
5. **Display and brightness**
   - Auto-lock → select never (default is one minute)

For the second phase of our experiment, which involved traffic analysis, we set up Splunk Enterprise on our Windows 10 VM. Splunk Enterprise is a comprehensive analytical platform known for its ability to search, analyze, and visualize data. Among its key advantages is the ability to efficiently upload large amounts of data in various formats and generate statistics for data visualization. Additionally, Splunk empowers users to create custom searches using Splunk Search Processing Language (SPL) to extract and present specific data according to their needs [44].

Next, we uploaded all traffic flow files to Splunk using the same index. An index is a repository for data, which means that by uploading all files to the same index, we can create searches for them collectively [45].

During the preprocessing stage, as outlined in Section 4.4, we added relative timestamps to ensure consistent timestamps for each event. However, the dataset contained some existing timestamps, which interfered with the appended timestamps and led to inaccurate indexing of events. To address this, we explicitly configured Splunk with a custom source type. A source type determines how data is formatted during the indexing process [46], which includes defining where each event starts and ends. We tailored this custom source type by creating a regex pattern that specifically matched the relative timestamps, thus ensuring consistent indexing of events. Additionally, we optimized the truncate setting to handle the maximum line length of our dataset. This adjustment was crucial for ensuring the complete retrieval of all event lines during the data upload process in Splunk.

The custom source type, incorporating both the regex pattern for the timestamp and the truncate setting for maximum line length, proved essential for our data analysis process. It ensured the accuracy and flexibility needed to manage the complexities of our data files in Splunk. The configuration of the custom source type was set as follows:

Custom source type settings:

1. **Event breaks:**
   - Regex Pattern: ([\r\n]+)\d{4}-\d{2}-\d{2}\d{2}:\d{2}:\d{2}.\d{3}
2. **Advanced:**
   - New setting: TRUNCATE: 9999999

In phase three, we built upon the findings from research question one, which identified the specific apps that collected the IDFA. We then accessed the web-based App Store to retrieve the privacy labels for the respective apps. The web-based App Store offers a preview of the apps, showcasing images, reviews, and privacy labels. Once we collected the privacy labels, we compared these details with Apple's definitions and guidelines to analyze the consistency and accuracy of the privacy disclosures.

This section has covered the experimental design in terms of the three phases and the environment setup in each of them. It covered which devices were used in the experiment and which software was installed respectively. A summary of the experiment devices, the software used, and their versions are given in Table 5.3.

Table 5.3: Summary of devices and software used in the experiment

Device	Software	Version
MacBook Pro	mitmproxy	10.2.2
	ipatool	2.1.4
	Apple Configurator	2.12.1
iPhone 6 (primary)	checkra1n	N/A
	SSL Kill Switch 2	0.14c
iPhone 6s (backup)	palera1n	N/A
	SSL Kill Switch 3	1.5.1
Windows 10 VM	Splunk Enterprise	9.2.1

5.1.1 Selected applications

The selected applications, determined by the methodology outlined in Section 4.2, are listed in Table 5.4. It is important to note that these apps are frequently updated and may no longer be available, compatible, or in the same state as when analyzed.

In addition, the top lists in the App Store are highly dynamic, implying that the apps may no longer be present in the top lists and, thereby, no longer be considered popular. This is because the App Store's top list is determined by its number of downloads within a limited time frame, particularly four to seven days [47], along with several other metrics, which means that the apps deemed popular at the time of our study were affected by their popularity at the specific time rather than their enduring popularity.

For this reason, apps like Facebook, YouTube, Spotify, Messenger, and Instagram, which might be considered *popular*, are not on the list. Although these apps have many downloads in general, they did not have many downloads at the time of the experiment. We can see this time sensitivity influencing the selected apps through the presence of the Islam-related apps, namely Muslim: Ramadan 2024, Azkar, and Namaz, which could be reasoned by our experiment period aligning with the beginning of Ramadan.

Table 5.4: Selected applications

Thematic group	**Application**
Life and wellness	Hjelp 113 Headache Calendar Pizzabakeren Norge Brain Twin Espresso House
Education and information	Muslim: Ramadan 2024, Al Quran Azkar: Athan & Prayer Namaz App: Learn Salah Prayer Web comics - Webtoon, Manga Jotun Colourpin
Entertainment and creativity	CapCut – Photo & Video editor ShortTV – Watch dramas & Shows BlockBlast MinFotball Toca Life World: Build a story
Productivity and business	Temu: Shop like a billionaire DaVinci – AI image generator Zoom – One platform to connect QR-Reader for mobile Nordnet: Stocks and funds
Social and mobility	Autopay – Park & Charge SAS WhatsApp Messenger Norwegian Telegram

5.1.2 Traffic capture

Capturing the traffic consisted of five steps. These steps are described in detail in the following section. The steps were repeated for each selected application, which produced one network traffic flow file per app. Summarized, the five steps were as follows:

1. Initialize the capture (10 minutes)
2. Allow requested access
3. Utilize application functionality
4. Initialize background traffic capture (five minutes)
5. Finish capture

Initialize the capture

To initialize the capture, we connected the iPhone and the MacBook to the same network and configured their proxy settings. On the iPhone, this involved specifying the MacBook's IP address as the proxy address. Meanwhile, on the MacBook, we activated the HTTP/HTTPS proxy settings to facilitate the traffic capture.

Next, the application's .ipa file was installed onto the iPhone via Apple Configurator, and all background processes were cleared to minimize other traffic as detailed in Section 5.1. This step is crucial to prevent traffic from external sources from poisoning the traffic. We then initialized mitmweb from the MacBook's terminal to display the incoming traffic.

Once mitmweb was operational, we launched the application and initiated a ten-minute timer. It was first at this stage we could verify the proper functioning of the application and switch from the primary iPhone to the backup iPhone in case of technical issues.

Allow requested access

When prompted, the app was granted permission to access data such as location, camera, and microphone to provoke as much PII collection and domain contact as possible. The permission options available were *Allow While in Use*, *Allow Once*, or *Don't Allow*. To grant the most privileged access possible, we selected the *Allow While in Use* option in all popup dialogues appearing.

Allowing requested access through the pop-up dialogues is essential for facilitating accurate results in our experiment. Due to the ATT policy, apps must list tracking domains in their privacy manifest and seek permission from the users to track them [20]. The operating system blocks network traffic to the listed domains if the permission is denied, which consequently impacts the amount of traffic. Therefore, granting the most privileged access possible is essential to get accurate results regarding tracking.

Utilize application functionality

To utilize the application's functionality during the interactive phase, we conducted several steps. As each application varied in functionality, we have generalized the typical interactions and presented them in the flowchart in Figure 5.3, which is provided at the end of this section. As illustrated by the figure, some examples of core features include signing up for an account, sharing preferences, and utilizing search functions. This figure illustrates how this step proceeded, from when the app was opened to when the ten minutes were up.

In our research, we aimed to portray a realistic usage pattern, which is why we opted for manual interaction. Hu *et al*. [5] utilized the Android application exerciser "Monkey," a script that automates user interaction on Android devices. Although a similar tool exists for iOS, e.g., "UI AutoMonkey", and using it could increase efficiency, such tools generate random events that do not necessarily reflect typical user behavior [48]. These tools are primarily designed for stress-testing apps by varying event frequencies and intervals, which is more suitable for testing an app's robustness rather than portraying realistic usage patterns. Furthermore, the use of such automation tools could be more justified if we were analyzing a larger set of applications; with only 25 apps in our study, manual interactions are manageable and more aligned with our objectives.

Initialize background traffic capture

As illustrated in Figure 5.3, we clicked the home button to exit the app when the initial ten minutes were up. We then started a five-minute timer while running the app in the background. The setting to disable auto-lock, as detailed in Section 5.1, was crucial for maintaining the device's active screen without requiring any user interaction. Throughout these five minutes, we closely observed the device's network traffic through mitmweb, tracking any new requests that would be sent from the device.

Finish capture

When the five-minute timer was up, the capture was finished, and we saved the traffic flow from mitmweb. This ensured no delayed traffic would appear in the dataset beyond the determined time frame. The application was then uninstalled from the iPhone, and the traffic flow file was preprocessed, as previously described in Section 4.4. When all 25 apps were completed, the dataset was validated, as we will cover in the following section.

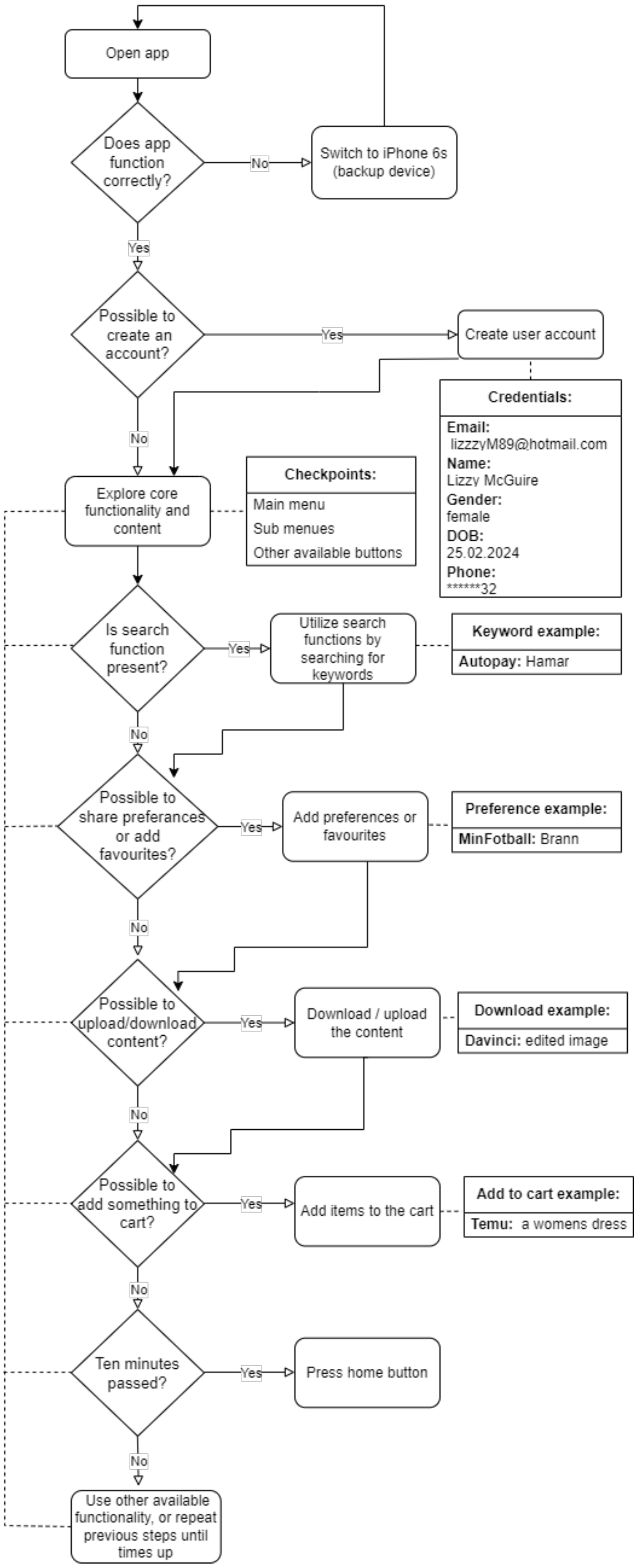

Open app
Does app function correctly?
No
Switch to iPhone 6s (backup device)
Yes
Possible to create an account?
Yes
Create user account
Credentials:
Email: lizzzyM89@hotmail.com
Name: Lizzy McGuire
Gender: female
DOB: 25.02.2024
Phone: ******32
No
Explore core functionality and content
Checkpoints:
Main menu
Sub menues
Other available buttons
Is search function present?
Yes
Utilize search functions by searching for keywords
Keyword example:
Autopay: Hamar
No
Possible to share preferances or add favourites?
Yes
Add preferences or favourites
Preference example:
MinFotball: Brann
No
Possible to upload/download content?
Yes
Download / upload the content
Download example:
Davinci: edited image
No
Possible to add something to cart?
Yes
Add items to the cart
Add to cart example:
Temu: a womens dress
No
Ten minutes passed?
Yes
Press home button
No
Use other available functionality, or repeat previous steps until times up

Figure 5.3: The user interaction process

5.1.3 Dataset validation

In this section, we describe the steps conducted to validate the network traffic dataset. Validating the dataset was done to ensure reliability, in terms of consistently producing the same results, and integrity, in terms of being in the correct format after preprocessing. Validation of the tracking domains was performed as part of their preprocessing, as detailed in Section 4.4. Therefore, we have not covered this dataset in the following section.

To validate the dataset we reproduced some of the traffic captures. As the requests appearing were approximately the same each time, the dataset was deemed valid. It is important to note that it is very difficult to reproduce the same traffic when applying user interaction as this implicitly will be unique for each user. In addition, the traffic may vary greatly despite being reproduced by the same user, as the delay between each event may still vary, and dynamic variables, such as tokens and timestamps, are difficult to reproduce. Furthermore, only a general usage pattern is described based on the flow chart for application usage in Figure 5.3. This means that the exact same sequence of events triggered in our research may be impossible to reproduce.

We also manually validated the content in each flow file to ensure the data was preprocessed correctly. This included being in a human-readable format and appended with the relative timestamps necessary for correct indexing in Splunk.

5.2 Collection of Personally Identifiable Information

This section presents the execution, results, and discussion regarding the first research question. This question investigates which PII is collected by the selected applications, with an emphasis on IDFA. In the execution section, we detail the SPL queries executed to retrieve the results, whereas the result section includes the specific findings. Lastly, in the discussion section, we discuss the expected results against the actual results reflecting on the underlying reasons.

Research question one:

Which PII is collected by iOS applications during 15 minutes of use?

5.2.1 Execution

To investigate which PII is collected by the selected applications, we uploaded each application's traffic flow file to Splunk to the same index (*search_index*). This enabled us to execute specific search queries on all apps combined. We developed three queries to investigate this question. Initially, we formulated a query to identify which PII was collectively collected by the selected applications. Subsequently, we crafted another query to find which apps were collecting the IDFA. Lastly, we created a query determining which PII was sent alongside the IDFA.

In developing these queries, we had to consider the lack of uniform variable definitions across our dataset. This was due to the selected applications being developed by 25 different developers, each with unique approaches to naming variables. For instance, a field containing the IDFA might be named *advertiser_id* by one developer, and *adid* by another. Consequently, our Splunk queries had to be tailored to search for specific PII values rather than the variable names containing these values. This means we have to search for specific strings that theoretically could be part of other strings, as previously mentioned in Section 4.5.1.

Considering that Splunk searches are case-insensitive [49], it was sufficient to search for the values in their original format. However, we have supplied the IDFA with an additional string, a copy of itself without hyphens, to increase the chance of detecting all occurrences. This is due to uncertainties regarding how this value is extracted from the app.

Considering these factors, we conducted extensive trial and error to build the queries to give the most accurate results. Accurate results mean a reduced number of false positives and false negatives. While false positives give more hits than what is actually present, false negatives give fewer hits than what is present. To reduce the false negatives, we included additional search terms, i.e., possible variable names, whereas to reduce the false positives, we removed some PII that were impossible to detect, for instance, age, as outlined in Section 4.5.1.

For the first query, which aimed to find all PII across all apps combined, we aggregated searches for different PII values. For instance, as illustrated in the SPL query below, we first searched for events containing IDFA, before appending with an additional search for UDID. After each PII value was appended, we generated statistics by counting each hit. The full query is available in Appendix D.

Sample from Splunk query to find the PII collected by the 25 applications:

```
index="search_index"
| search
| stats count as TOTAL

| append [search index="search_index"
| search ("*BA0A45A4-B5E2-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*EFE00A67-7C5D-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*BA0A45A4B5E2XXXXXXXXXXXXXXXXXXXX*" OR "*EFE00A677C5DXXXXXXXXXXXXXXXXXXXX*")
| stats count as IDFA]

| append [search index="search_index"
| search ("*1A8C3D085CCC0DA7A3F11C6B1EC21A8016CDBD6B*" OR
↪  "*7D10BC48D9372AA313FF13D52A6D10D26C8CCF92*")
| stats count as UDID]
```

To identify which applications collected the IDFA, we searched for specific sources, i.e., traffic flow files, with events containing the IDFA. To do so, we executed the SPL query listed below. This query searches through all the traffic flow files in the index to locate events containing the IDFA. For clean representation, a regular expression is used to extract a portion of the source name, which is then used to rename the source to the corresponding app's name. Finally, the output is organized in descending order by the count of IDFA events. The query is cut short for demonstration purposes. The full query can be found in Appendix E.

Splunk query to find applications collecting the IDFA:

```
index="search_index" source=*
| search ("*BA0A45A4-B5E2-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*EFE00A67-7C5D-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*BA0A45A4B5E2XXXXXXXXXXXXXXXXXXXX*" OR "*EFE00A677C5DXXXXXXXXXXXXXXXXXXXX*")
| rex field = source "(?<source_short>.+)_time"
| eval Source = case(source_short="telegram", "Telegram",
↪  source_short="norwegian", "Norwegian")
| stats count as "IDFA" by Source
| sort - IDFA
```

To detect which PII was transmitted alongside the IDFA, we modified our initial query to search for events containing both IDFA and other PII. This modification involved using the 'AND' operator to ensure both IDFA and the additional PII were present in the same event, effectively pinpointing such occurrences. A snippet from the query is exemplified below, whereas the full query is detailed in Appendix F.

Splunk query to find the PII transmitted along with the IDFA:

```
index="search_index"
| search ("*BA0A45A4-B5E2-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*EFE00A67-7C5D-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*BA0A45A4B5E2XXXXXXXXXXXXXXXXXXXX*" OR "*EFE00A677C5DXXXXXXXXXXXXXXXXXXXX*")
| stats count as IDFA

| append [search index="search_index"
| search ("*BA0A45A4-B5E2-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*EFE00A67-7C5D-XXXX-XXXX-XXXXXXXXXXXX*" OR
↪  "*BA0A45A4B5E2XXXXXXXXXXXXXXXXXXXX*" OR "*EFE00A677C5DXXXXXXXXXXXXXXXXXXXX*")
↪  AND ("*1A8C3D085CCC0DA7A3F11C6B1EC21A8016CDBD6B*" OR
↪  "*7D10BC48D9372AA313FF13D52A6D10D26C8CCF92*"))
| stats count as UDID]
```

5.2.2 Results

Through our examination of traffic flows in search of PII, we discovered that 13 out of the 17 types of PII were collected. Figure 5.4 illustrates which PII were identified in the traffic and the number of events accordingly.

From the results, we find that coarse location emerged as the most frequently collected PII, appearing in 1,413 events, which accounts for 12% of all recorded events. Following behind, IP addresses were found in 976 events, representing 8% of the total. Email/username transmissions appeared in 672 events, equivalent to 6%.

Furthermore, sensitive PII, such as precise location, was recorded in 436 events, or 4% of the total, while the IDFA was found in 417 events, making up 3.5% of the dataset. Gender appeared slightly less frequently, found in 303 events, or 2.5%.

The less frequently collected types of PII were UDID with 56 occurrences (0.5%), followed by phone number with 29 hits (0.24%) and serial number with 26 (0.22%). Furthermore, first name was found 16 (0.13%) times, and device name and last name were each recorded five times (0.04%). Finally, IMEI was noted only twice (0.01%). Each of these types was identified in fewer than 60 events. Interestingly, the dataset contained no instances of date of birth, ECID, WiFi addresses, or Bluetooth addresses.

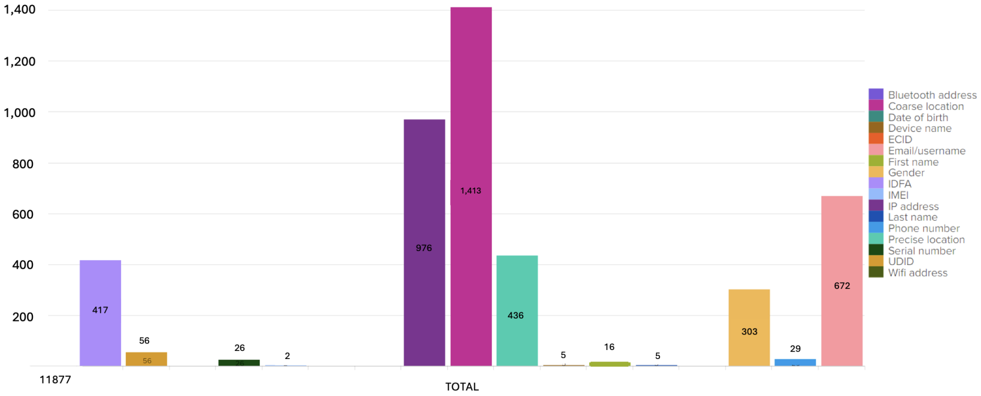


Figure 5.4: PII collected by the selected applications

Our investigation into which applications collected the IDFA revealed that nine out of the 25 analyzed apps were involved, as shown in Figure 5.5. This means that 36% of the examined apps collected the IDFA. ShortTV was the leading collector, transmitting the IDFA in 173 events, followed closely by Webcomics, which transmitted it in 127 events. QR-Reader, CapCut, and Muslim Ramadan each transmitted the IDFA in around 30 events. Davinci was found to transmit the IDFA 12 times, followed by Azkar with nine hits, BlockBlast with five, and Norwegian with the least of all with only three.

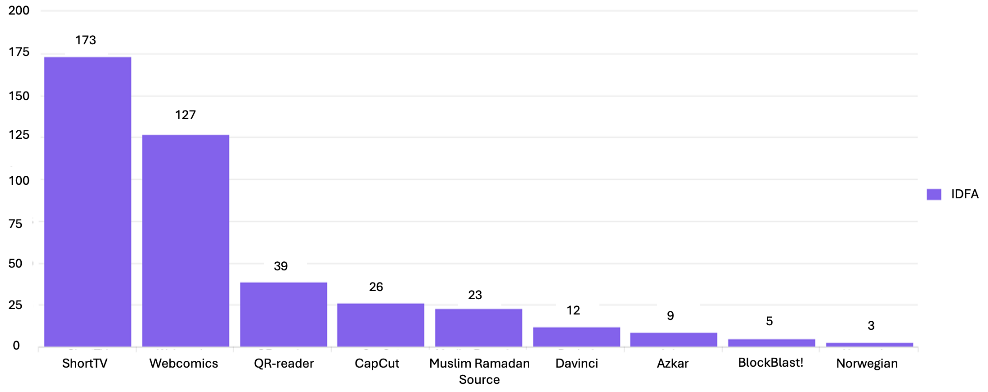


Figure 5.5: Applications collecting the IDFA and their frequency

By examining the events containing the IDFA, we discovered that only four out of the 17 types of PII were transmitted alongside it, as illustrated in Figure 5.6. The results show that coarse location was the most frequently transmitted PII type, occurring in 240 events, followed by gender, appearing in a total of 122 events. The email address only occurred in 25 events, whereas the IP address occurred only once.

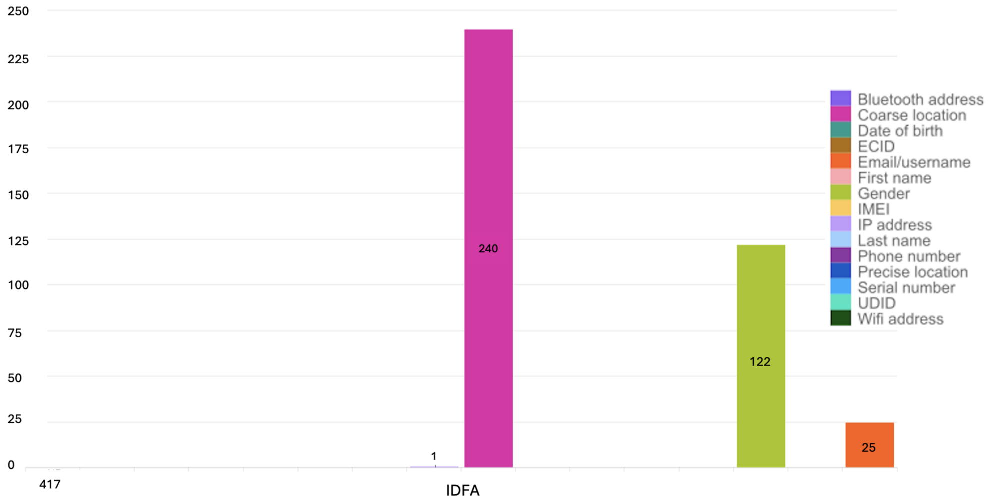


Figure 5.6: PII sent along the IDFA

Lastly, in our monitoring of the traffic using mitmweb, while the app ran in the background, we observed that none of the apps continued to transmit data. However, we did notice some requests being sent to domains owned by Apple. The specific purposes of these requests could not be determined, due to lack of information available online, but it is generally understood that such domains are used for various purposes [50]. These include checking for updates, syncing data, or other background processes. Based on this, we concluded that these requests are not a result of the app's tracking practices but rather infrastructure maintenance on the Apple device. The specific domains we observed being contacted in the background were as follows:

- *gsp-ssl.ls.apple.com*
- *gsp64.ssl.ls.apple.com*
- *gspe-11-ssl.ls.apple.com*
- *gsp10-ssl.apple.com*
- *gs-loc.apple.com*
- *gsp6464-ssl-ls.apple.com*

5.2.3 Discussion

As discussed in Section 4.5.1, we anticipated that unique identifiers such as UDID, serial number, and IDFA would be frequently collected due to their ability to uniquely identify a device. However, our findings are partly contradictory to these expectations. While IDFA was indeed frequently collected, appearing in 36% of the apps and emerging as the most collected type of PII from the *unique identifiers* category, UDID and serial number were among the least collected in the dataset. The low frequency of UDID and serial number transmission suggest these identifiers are most likely used for infrastructure maintenance rather than tracking purposes. This strongly indicates that the *device_id* field in the NRK database likely contains the IDFA, which signifies its powerful ability to track individuals when correlated to location data.

From the results, we find that the transmission of IDFA is primarily observed within the thematic groups *Entertainment and Creativity* and *Education and Information*, each with three apps represented on the list. *Entertainment and creativity* is represented by ShortTV with 173 events, CapCut with 26 events, and BlockBlast with five events. Education and information are represented through WebComics with 127 events, Muslim Ramadan with 23 events, and Azkar with nine events.

Interestingly, both Azkar and Muslim Ramadan are Islam-specific apps, both subject to revealing sensitive information about their users. As previously discussed in Section 3.2, sensitive apps may pose a threat to their users by including third-party advertising. Vines *et al.* [1] described this scenario using gay apps, and we see a similar link to our experiment's religious apps. Based on the NRK dataset, which included a column labeled *app_name* occasionally listing the app that shared the given data, we see how the apps Azkar and Muslim Ramadan can contribute to disclosing such sensitive information when an individual's IDFA is known. Our finding indicates that such apps, despite revealing sensitive information, are prone to collecting IDFA.

The research by Koch *et al.* [4] suggested that the Photo and Video category is particularly prone to collecting IDFA. This finding is consistent with our observations of the app CapCut, a ByteDance-owned app, which transmits the IDFA in 26 events. These findings indicate a persisting trend where IDFA is commonly found in video and editing apps, suggesting little change in how this app category handles IDFA transmission.

Given that CapCut is developed by ByteDance, the same company that owns TikTok, it is not surprising that it is on the list of apps collecting the IDFA, as ByteDance heavily relies on advertising on their digital platforms for income [51]. Based on what we know about TikTok's data collection and sharing practices [52], we have reason to assume that similar practices occur in other ByteDance-owned apps. If anything is surprising with our results, it would be that the IDFA is not sent more frequently than 26 times.

From the personal information category, we expected gender and email to frequently appear in the traffic. Overall, gender appeared to be the sixth most collected PII type, with 303 events in the dataset, of which 122 also contained the IDFA. This proves gender to be a key attribute in targeted advertising, as assumed.

On the other hand, the email address was the third most collected PII overall, appearing more than twice as often as gender. These results indicate that email addresses hold an even more significant role in the tracking industry than initially assumed. However, it is only collected along the IDFA 25 times.

One reason email is so frequently collected overall could be that we have merged email and username in the same variable. This means that the findings may not specifically represent email collection for third-party services solely; however, email may also be collected as a consequence of user interaction, e.g., signing up for user accounts. User interaction involving signing up for an account applied to 13 out of the 25 applications. However, we do not believe this has produced 672 events alone.

We believe some of these email events have been produced for tracking purposes, exemplified by the 25 events where email has been transmitted along the IDFA. Cross-referencing unique identifiers creates a comprehensive dataset about a user, which facilitates more detailed profiling. Emails are closely related to a user, as they often are utilized to sign up for accounts, subscriptions, or newsletters. This reveals the user's preferences, which can further be used to enhance targeted advertising. In addition, the information gathered about an individual is commonly shared with others, including DSPs [6]. For instance, the Norwegian beauty and wellness chain Vita disclosed their data processors on their website. The list, involving both Google and Meta accompanied by 31 other data processors, indicated a broad cooperation [53]. This practice enables even more detailed profiling and more enhanced targeted advertising.

We guess many people are unaware of how signing up for such accounts, subscriptions, and newsletters contributes to profiling and cross-platform tracking, as we generally understand this as a not well-communicated subject. Although the information about data collection and sharing practices might be hidden inside an *accept terms and conditions* check box, we do not believe that this is commonly read in-depth by the users, proposing a lack of knowledge and understanding to the users.

From the PII category *location based*, we expected location data to be among the most frequently sent PII, also found to be true. Coarse location was the PII type most frequently collected both overall and along with the IDFA. However, our expectation of finding the precise location was not met to the extent previously anticipated. While coarse location is collected in 1,413 events, precise location was only sent in 436 events. The reason for the difference could be, as we presumed, that language code is included, and this is likely collected in other contexts than tracking. However, we also believe that coarse location is an efficient way to target audiences. For instance, X Business states several examples of coarse location types that an advertiser can target their campaigns to, including countries, regions, metros, cities, postal codes, or a radius around a location [54].

The fact that precise location was not collected as frequently as we initially assumed can be attributed to our selection of applications. Tracking practices vary among apps, including how aggressively they track location, meaning not all apps consistently track precise location. As coarse location includes language code, all 25 apps have likely collected this PII type. However, we have only identified nine apps that, with certainty, track users by collecting the IDFA. We guess these nine apps collect precise locations due to their tracking practices, making comparing the number of events in the coarse and precise location unfair. This discrepancy might explain why precise location is not collected as frequently as coarse location. If we had investigated specific app categories prone to tracking, our results might have differed, potentially showing a more frequent collection of precise location.

Lastly, we anticipated finding minimal background traffic due to background updates and refreshes being turned off in the device's settings. The background update and refresh settings control how and when apps can update content in the background. Additionally, the strongest level of access granting was *allow while in use*, as described in Section 5.1.2, which restricts network traffic to only happen while the app is open and actively in use. The only domains contacted were Apple domains, which we concluded were likely related to Apple's services rather than tracking. Combined, we conclude these to be the reasons behind the lack of background traffic.

5.3 Frequently contacted tracking domains

This section presents the execution and findings related to research question two, which investigates frequently contacted tracking domains and the domains that collect the most PII. We will outline how the investigation was executed and present the results, followed by discussing these findings.

Research question two:

During the 15 minutes of app usage, which tracking domains are most frequently contacted?

5.3.1 Execution

To investigate the tracking domains that are most frequently contacted, we continue working on the same index in Splunk, containing the traffic flows from all 25 applications. Additionally, we utilized the compiled list of 111 tracking domains to compare with the domains contained in our traffic.

For the Splunk search query, we used a search macro, which is a reusable chunk of SPL that can be inserted into other searches [55]. The search macro increased readability by shortening the search query and reducing redundant code. We implemented a macro named ’tracking_domains’ containing the 111 tracking domains, as illustrated below. Please note that the query is cut short for demonstration purposes.

Search macro (*tracking_domains*):

```
rex field=_raw "(?i)(?<domains_tracking>(2mdn\.net|adcolony\.com \
|adjust\.com|adnxs\.com|adobedtm\.com|amazon-adsystem\.com))"
```

In the query outlined below, the macro *tracking_domains* is used to extract tracking domains from the dataset. It matches the strings contained in the macro and counts the number of occurrences of each identified domain. Next, it groups them by each domain and finally sorts these counts in descending order. This process reveals which tracking domains are most frequently contacted in the dataset.

Splunk search to find the total number of tracking domains:

```
index="search_files"
 | `tracking_domains`
 | stats count by domain_tracking
 | sort - count
```

5.3.2 Results

As illustrated in Figure 5.7, the analysis of frequently contacted tracking domains shows a significant dominance by Alphabet-owned domains, and Meta's *graph.facebook.com*. Other notable domains include *applovin.com* and *app-measurement.com*. We find that from the total 11,877 events analyzed, 2,006 (17%) were directed to the tracking domains contained in our list. Note that the figure only illustrates the 25 most contacted tracking domains and that the entire list of tracking domains is given in Appendix G.

The most frequently contacted tracking domain in our dataset is *googleads.g.doubleclick.net*, owned by Alphabet, with 403 interactions representing 3.4% of all events. Other prominent Alphabet-owned domains such as *googleapis.com*, with 324 hits (2.7%), *googlesyndication.com*, with 250 hits (2.1%), and *app-measurement.com* also rank highly, with 115 hits (1%), underscoring Google's significant presence in the tracking industry. Facebook's *graph.facebook.com* was the second most contacted tracking domain, with 363 interactions, making up 3.1% of the dataset. We also find *applovin.com* among the most frequently contacted domains with 150 hits, translating to 1.3% of all interactions.

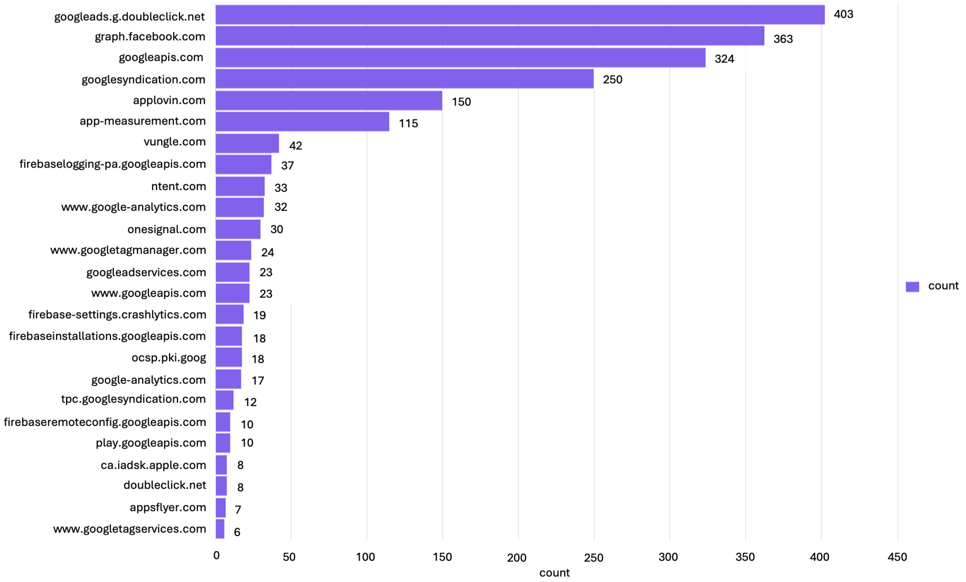


Figure 5.7: The 25 most frequently contacted tracking domains

Less frequently contacted domains included *vungle.com* with 42 hits and *ntent.com* with 33 hits. The least contacted domains include *flurry.com*, *launches.appsflyer.com*, *sessions.bugsnag.com*, *supericonads.com*, each with only two hits each. With only one hit each we find *adcolony.com*, *attr.appsflyer.com*, *conversions.appsflyer.com*, *gcdsdk.appsflyer.com*, and *sentry.io*.

5.3.3 Discussion

As previously discussed in Section 4.5.2, we expected Alphabet-owned domains like *googleadservices.com* and *google-analytics.com* to dominate our results of the most frequently contacted tracking domains. Our findings confirmed this, with Alphabet's presence particularly pronounced across multiple domains. Notably, the most frequently contacted tracking domain, *googleads.doubleclick.net*, owned by Alphabet, registered 403 interactions. According to Netify, this domain is a "catchall" for the various marketing platforms from Google [56], meaning it serves as a central hub or a common entry point for multiple marketing services that Google offers. This could include advertising, analytics, and tracking in one, all managed under this single domain. However, *googleadservices.com* was not that frequently contacted, based on its 23 hits overall in the traffic.

Additionally, we anticipated encountering Meta-owned domains such as *graph.facebook.com*, reflecting Facebook's extensive data collection practices. This hypothesis was confirmed, as illustrated in Figure 5.7, where *graph.facebook.com* lands in second place. Of the 2,006 events directed to tracking domains, 363 were sent to *graph.facebook.com*, translating to 18% of all events sent to tracking domains.

Following Facebook, we anticipated that AppsFlyer domains would be prominently featured, with additional notable traffic to Branch and smaller advertising specialists. The results show that although AppsFlyer is present with six different domains in total, they are not prominently featured in terms of events. As illustrated in Figure 5.7, the domain *appsflyer.com* is contacted in seven events only. On the full list, detailed in appendix G, the remaining AppsFlyer domains, such as *attr.appsflyer.com* and *launches.appsflyer.io* are listed. However, the total of AppsFlyer interactions is only 13 events, despite being represented by six distinct domains. Contrary to expected results, we do not see any traffic related to Branch-owned domains.

As previously discussed in section 5.2.3, the ByteDance-owned app CapCut was found to collect the IDFA, indicating its involvement in tracking activities. Based on ByteDance's growth in the advertising industry, with TikTok generating $16.1 billion in revenue in 2023 [57], we would expect ByteDance domains to be represented in our results. However, our list of tracking domains does not include any ByteDance-owned domains, implying that such domains will remain undetected. If these domains were present, on the other hand, we would expect them to appear closely behind Google and Meta-owned domains.

5.4 Privacy label tracking transparency

This section presents the execution and findings related to research question three, which focuses on the tracking transparency of the selected applications' privacy labels. We will outline the methodology applied to our research question, followed by presenting the results, before concluding with a discussion of these findings.

Research question three:

To what extent are the privacy labels transparent about the data collected for tracking purposes?

5.4.1 Execution

To determine the transparency of the selected applications' privacy labels, we employed the methodology outlined in Section 4.5.3. Based on the results from research question one, we identified nine apps that collect IDFA, which will be further examined in this research question.

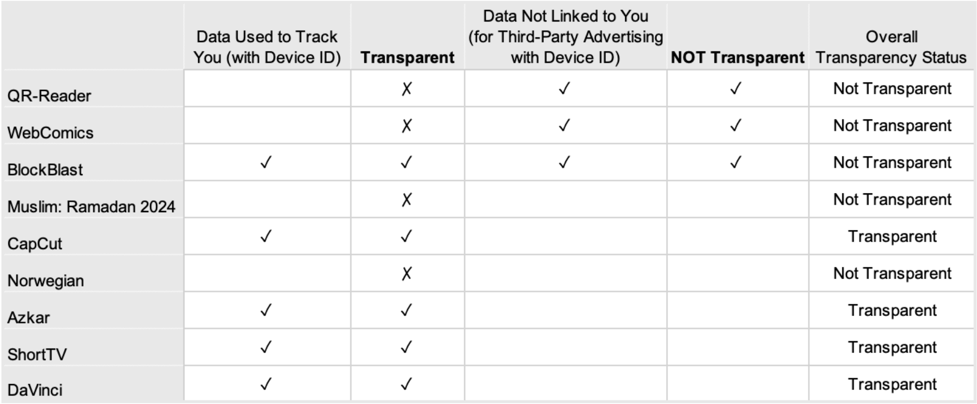
Initially, we have a set of nine applications that all engage in tracking activities using IDFA, or *Device ID*, as per Apple's privacy label guidelines [24]. Consequently, the nine apps are expected to present the privacy label *Data Used to Track You*, accompanied by the data type *Identifiers* with *Device ID* inside. However, if the label is present but the data types are not, the declaration is ambiguous, revealing tracking but not clearly specifying which data types are used.

Next, we assessed how data usage is outlined in the privacy label stated *Data Not Linked to You*. A significant contradiction arises when applications state that certain data is *not linked* to the user, while also acknowledging that the data is used for *Third-Party Advertising*. Despite Apple's definition of this purpose, as previously described in Section 2.5, it involves displaying third-party ads within the app or sharing data with entities exhibiting such ads. Listing the *Device ID*, namely the IDFA, as a data type collected for this purpose while claiming it is *not linked* to the user is, therefore, contradictory.

Furthermore, if the label *Data not Collected* appears on any of these applications, it implies a significant lack of transparency, eliminating the need for further investigation as it offers no additional information.

5.4.2 Results

From our investigation, we identify the apps DaVinci, ShortTV, Azkar, and CapCut as transparently disclosing their tracking activities. In contrast, Norwegian, QR-Reader, WebComics, BlockBlast, and Muslim Ramadan were found *not* transparent. Figure 5.8, which is explained below, illustrates these findings, along with the privacy label attributes that were evaluated.



	Data Used to Track You (with Device ID)	**Transparent**	Data Not Linked to You (for Third-Party Advertising with Device ID)	**NOT Transparent**	Overall Transparency Status
QR-Reader		✗	✓	✓	Not Transparent
WebComics		✗	✓	✓	Not Transparent
BlockBlast	✓	✓	✓	✓	Not Transparent
Muslim: Ramadan 2024		✗			Not Transparent
CapCut	✓	✓			Transparent
Norwegian		✗			Not Transparent
Azkar	✓	✓			Transparent
ShortTV	✓	✓			Transparent
DaVinci	✓	✓			Transparent

Figure 5.8: Tracking transparency in privacy labels

Explanation of Figure 5.8: For applications with a privacy label that specifies *Data Used to Track You* along with the *Device ID*, a checkmark is placed in the first column, indicating transparency, as noted in the following column. Conversely, if the privacy label specifies *Data Not Linked to You* for *Third-Party Advertising* with the *Device ID*, it is marked with a checkmark indicating a lack of transparency and is also checked as not transparent. Empty columns indicate that these details are not specified. The final column combines these results to determine if the app is overall transparent or not. An app is considered transparent only if the *Transparent* column is checked and the *Not-transparent* column is empty.

Initially, we found that four out of nine apps disclose their tracking activities insufficiently. Although QR-Reader, WebComics, and Muslim Ramadan specify *Data Used to Track You* in their privacy label, they fail to be fully transparent as they do not specify the *Device ID*, namely the IDFA, as collected data. Furthermore, Norwegian fails to declare any tracking activities at all. Consequently, these four apps lack full transparency regarding their tracking practices, as outlined in the two first columns of Figure 5.8.

Additionally, three applications, namely QR-Reader, WebComics, and BlockBlast, collect Device ID for third-party advertising purposes but claim it is not linked to the user. This indicates a lack of transparency and adds BlockBlast to the list of non-transparent apps. Consequently, five out of the nine apps are not transparent.

In conclusion, the apps CapCut, Azkar, ShortTV and DaVinci are considered transparent as their privacy labels indicate tracking activities along with the data type *Device ID*. The labels for these four apps are not misleading regarding tracking activities. All nine applications' privacy labels are documented in Appendix H.

5.4.3 Discussion

Our analysis reveals shortcomings regarding the transparency and accuracy of privacy labels within five of the selected applications. These all fail to be transparent regarding their tracking activities, especially the app Norwegian, which fails to acknowledge any tracking whatsoever. This demonstrates an obvious lack of transparency regarding its data collection practices. Given that all other applications in some form disclose tracking, Norwegian emerges as the least transparent among them.

Furthermore, our findings reveal contradictions in the stated purposes for data collection, particularly concerning third-party advertising. Data that clearly suggests a direct link to the user was categorized as *not linked* in three applications. This is concerning because data used for third-party advertising inherently involves user-specific targeting. According to Apple's definition, as previously outlined in Section 2.5, third-party advertising fundamentally involves leveraging user data to deliver targeted advertisements. This requires a link to user identities to ensure that ads are relevant. As previously described in Section 2.4, when first-party data is connected with third-party data, this action is recognized as tracking. Whether used alone or combined with other data, data involved in tracking can be used to identify specific users. Previous research outlined in Chapter 3, has proven that tracking individuals is feasible even with supposedly anonymized data.

Given that privacy labels are more accessible than privacy policies, it is reasonable to assume that users may be misinformed. For users to exercise their privacy rights, it is essential that data usage is accurately presented. It is vital that users easily understand how their data is used and who possesses it. When information is inaccurately disclosed, it undermines the system's transparency, suggesting that similar issues may apply to other applications beyond those specifically analyzed in this thesis.

Our findings are consistent with our anticipated results, which suggest that not all apps are expected to be transparent regarding their tracking activities. This suggests that privacy labels may not always fully reflect the true extent of data usage and privacy impacts, as indicated by existing research on privacy labels.

Chapter 6

Discussion

This chapter serves as a discussion of our thesis. It presents practical recommendations regarding the experiment, followed by legal and ethical aspects. Furthermore, we reflect on our thesis, including the challenges we encountered and potential improvements. The chapter concludes by discussing the project's contribution to the UN's sustainability goals and our use of AI.

6.1 Practical recommendations

The following section covers practical recommendations regarding the experiment conducted in Chapter 5, which future researchers would benefit from knowing. The experiment should be easy to reproduce as the tools are open-source and available online.

First, we recommend having one or two backup devices available during the experiment. We encountered technical issues, including applications failing to open and unexpectedly crashing on our primary device. Luckily, we had an alternative backup option, which was useful in such cases. A backup device is particularly recommended when handling jailbroken devices, as their tampered operating systems can decrease reliability and performance. The backup device will ensure the continuity of the experimental method. We also advise having devices running different iOS as problems could arise from a specific and outdated iOS. Furthermore, we advise using iPhones with sufficient memory, capacity, and SoC specifications. We believe this was the cause of the technical issues we encountered on the iPhone 6, as no technical issues occurred on the iPhone 6s.

An additional recommendation is to utilize an appropriate log analysis tool. We opted to use Splunk Enterprise, primarily due to its effectiveness in log analysis, as well as its cost-effectiveness, with a trial subscription. However, using Splunk required several preprocessing steps, as outlined in Section 4.4, due to the data being imported from a different software. This included adding relative timestamps and creating a custom source type for correct indexing. Moreover, the dataset was unstructured, consisting of arrays in various data formats, reflecting the diverse nature of web traffic captured. Due to the unstructured nature, Splunk's built-in *key-value* pair recognition, known as *field discovery* [58], did not work optimally, leading to multiple fields needing to be extracted manually with regex. This resulted in complex search queries for data extraction and result visualization. Therefore, we suggest exploring alternatives to Splunk for data analysis. It is also possible that finding an alternative to mitmproxy, such as Fiddler or Charles, for intercepting the traffic could solve this challenge as well.

6.2 Legal and ethical aspects

In our experiment, we intercepted network traffic from a mobile device, simulating a MiTM attack, to get insight into PII, such as the IDFA. While this technique can be unethical and illegal if done with malicious intent, it proved effective in retrieving IDFA, demonstrating a method that threat actors might use. Recognizing that unauthorized packet sniffing is illegal and can lead to criminal charges, we took several measures to ensure that our research was conducted legally and ethically.

The first part of our experiment, i.e., the MiTM attack, was carried out exclusively on personally owned equipment, specifically old iPhones that were no longer in active use by our team members. Before the experiment, these devices were reset to factory settings to eliminate any previously stored personal data. Our objective was solely to analyze network traffic for research purposes, not to invade the privacy of other individuals or gain undue advantages. Unauthorized packet sniffing with malicious intent is illegal and subject to criminal charges under § 205 of the Norwegian Penal Code, which safeguards private communication [59]. By experimenting on our own devices, we ensured compliance with this legal standard.

To preserve the personal privacy of the team members during the experiment, essential precautions were adhered to by GDPR, which mandates the anonymization of personal information such as full name, email address, phone number, and IP address. We established a dummy account on iTunes using a fictitious full name, a new email address, and a fabricated date of birth. As the registration process required a valid phone number and home address, we provided a team member's details, ensuring privacy by only displaying the first two digits of the number in our table of PII and excluding the address entirely.

Given that our experiment investigated PII transmission and relied on queries looking for the specific PII in the network traffic, of which we have provided examples, we implemented a policy to redact any personal data disclosed before including it in our thesis. For instance, we have redacted portions of the IDFA due to the uncertainty regarding the information it might reveal following our experiment. Given that we used personal phone numbers in our research, we aim to avoid the risk of exposing such sensitive information through the IDFA. This aligns with Article 6(a) of the GDPR, which requires explicit consent for personal data processing [60]. In addition, as our research involved inspecting traffic from applications, potentially revealing vulnerabilities or more information than what was intended for regular user access, traffic flow files will not be made available online.

According to GDPR, IP addresses are classified as personal information [25], due to their potential to identify an individual. We have, therefore, partly redacted this as well. As noted in Chapter 4.2, we also chose to exclude applications requiring BankID in our experiment to mitigate the need to share personal financial information during the traffic capture.

Although using a dummy account could aid in preserving the personal privacy of the team members, it has to be carefully considered. As outlined in § 202 of the Norwegian Penal Code, any action that involves using another's identity, or one easily confused with another's, with the intent to deceive, gain an unjust advantage, or cause harm, is punishable [61]. If Lizzy McGuire was associated with a real individual or closely resembled a real identity, using it without consent could be considered a violation. For this reason, our dummy account was not linked to any real individual.

To ensure compliance with the Norwegian Copyright Act, §37 "Reproduction of artworks and photographic works in critical and scientific discourse and biographies" [62], we chose to create our own illustrations. This approach granted greater flexibility to tailor the illustrations to our needs and academic content and enhanced our understanding of the topics.

6.3 Reflections on our thesis

During our bachelor's thesis, we have identified multiple areas that could have been addressed differently. This section will discuss these areas and reflect on alternative approaches.

In developing our theoretical foundation, we conducted literature research rather than a literature review. This means that we have not done a comprehensive evaluation of all relevant research available, which influenced the scope of our research. A literature review is more comprehensive and relies on a systematic approach that results in a broader scope of literature than we examined. While our literature research included several articles, we believe that many more are yet undiscovered. Undertaking a full-scale literature review could have provided a richer informational base, thereby facilitating more informed decisions in formulating our research questions and guiding our overall study.

We recognize some limitations regarding the number of apps we analyzed. Due to our small sample size of selected apps, it is challenging to draw definitive conclusions and present statistically accurate comparisons between the thematic groups. We started with 25 apps, of which 9 were examined further. We understand that this dataset is not comprehensive enough to generalize how all apps collect user data and adhere to privacy labels. However, the apps we analyzed were among the most popular and widely used by the Norwegian population at the time of our writing. In addition, the applications were selected with diversity in mind, such that the findings would represent a broad specter of apps. Thus, it is likely that our findings apply to other applications in the App Store as well.

Furthermore, the applications we selected do not necessarily align with the conventional definition of popularity, such as globally recognized apps like Facebook, YouTube, or Spotify. These well-known platforms do not always appear at the top of the Apple App Store charts due to the App Store ranking algorithm favoring recent downloads and user reviews, as detailed in Section 5.1.1. Additionally, compatibility issues with an older operating system restricted our ability to include some of the more popular apps. Despite these challenges, each app we analyzed ranked within the top 15 in its category at the time of selecting them, affirming their relevance for iOS users in Norway. A different approach could have been targeting the persistently popular apps instead.

We believe that the dataset likely contains more PII than we identified, due to the limited scope of our search. As discussed in Section 4.5.1, some types of PII, such as search history, contacts, and age, are challenging to detect through our method and were not included in our study. A manual search through the dataset before analysis could aid in identifying variable names for these properties, which further could be used to build the queries. With this approach, PII values like age, which is impossible to detect in the traffic with our method, could then be identified. However, this approach would require considerably more time and might still not capture all attributes, given the data's complexity and varied formatting, which could make the data difficult to interpret.

Of the 111 tracking domains initially identified, only 39 were found in our traffic, representing a 35% match rate. By effectively only comparing our traffic against 39 tracking domains, there is a high likelihood that additional tracking domains occurred undetected. Our results regarding the most frequently contacted tracking domains may, therefore, not depict the entire realistic picture.

Additionally, our preprocessing stage regarding the tracking domains might have been too strict, potentially filtering out some valid tracking domains. If we had not reduced our list of tracking domains from 153 to 111, more domains might have appeared in the traffic. We prioritized caution and excluded any uncertain domains, though some unverified ones might be tracking domains. As stated in the previous work by Vallina-Rodriquez *et al*. [28], verifying the true purpose of different domains is difficult. Although using tools such as VirusTotal gives insight based on category tags, validating their exact purpose remains challenging. A different approach could have been to utilize web scrapers or simply use a trusted and verified blocking list, enabling a more efficient process in validating the domains and a more comprehensive list to compare against the traffic.

In our research, we chose to analyze privacy labels rather than privacy policies, due to the user most likely reading the privacy labels rather than the privacy policies. However, analyzing the privacy labels is insufficient in determining whether apps violate rules and regulations, as these labels are not legally binding agreements between the user and the app. Privacy labels offer a simplified overview of an app's practices, which might not capture the full extent of data handling and compliance. A more thorough examination of the privacy policies themselves would likely provide a stronger basis for assessing compliance with regulations like GDPR. By delving into the detailed statements within privacy policies, we could more accurately evaluate whether the apps truly uphold the legal privacy and data protection standards.

Despite facing challenges during the work on our thesis, we also gained a sense of achievement and knowledge by overcoming them. Through careful planning, guidance from our supervisors, and clear result goals, we managed to answer all our research questions to the best of our abilities.

6.4 Sustainability

Research and technology are closely linked to sustainability, as they are essential for innovating solutions that can address today's global challenges. This section addresses the United Nations (UN) Sustainable Development Goals relevant to our thesis, highlighting how our work can contribute to a sustainable future.

"The UN Sustainable Development Goals are the world's shared action plan to eradicate poverty, combat inequality, and stop climate change by 2030" [63]. These goals consist of 17 objectives, each containing multiple sub-goals. Collectively, these objectives aim to promote sustainability across land, business, and civil society [63]. While sustainability might be associated only with large entities, the increasingly digital world means that all stakeholders in the tech industry should keep this in mind when developing and improving new technologies.

Sustainable Development Goal 16, sub-goal 16.10 states: "Ensure public access to information and protect fundamental freedom, in accordance with national legislation and international agreements" [64]. In our bachelor's thesis, we demonstrate how technology can influence society by raising awareness. We explore the field of ADINT and discuss how certain apps fail to disclose the user data they collect for tracking purposes. This highlights potential breaches of Norway's Personal Data Act and GDPR. Such practices directly violate Sustainable Development Goal 16.10, as these entities fail to ensure public access to information and do not protect users' fundamental rights concerning their personal data.

On a broader scale, Sustainable Development Goal 16 focuses on peace, justice, and strong institutions. Our thesis highlights the lack of transparency within the advertising ecosystem and exposes the use of ADINT. By addressing these issues, we aim to promote human rights by empowering citizens with knowledge of their inherent rights. Additionally, we seek to influence authorities, companies, and stakeholders to uphold these laws and regulations [65].

By highlighting the lack of transparency within the advertising ecosystem, we demonstrate the need for change to ensure honesty towards its users. This dishonesty is exploited by actors seeking to track individuals. By raising awareness of this issue, we emphasize the urgency of making the system more transparent and sustainable. This aligns with sub-goal 9.5 of the UN Sustainable Development Goals, which states: "Enhance scientific research, upgrade the technological capabilities of industrial sectors in all countries..." [66]. Our thesis helps to strengthen scientific research and demonstrate that today's technology should be upgraded.

6.5 Artificial Intelligence

The team has utilized ChatGPT and Grammarly throughout our work. These tools have been valuable in enhancing our writing and spelling. Given that Norwegian is our primary language and our English writing experience varies, we found it valuable to utilize AI to refine our sentences and propose alternative ways to express ourselves. This involved constructing the sentences ourselves but receiving assistance from AI to express them more effectively.

Chapter 7

Conclusion and further work

In this final chapter, we will conclude our thesis by reviewing our experiment and summarizing the results of each research question. Finally, we suggest further work that can expand upon our thesis and address the remaining gaps.

7.1 Conclusion

Our thesis has focused on a specific aspect of the advertising ecosystem to understand the exploitation of user data for targeted advertising and ADINT. We have explored a central element, the MAID, and its key role in mobile advertising and, thereby, mobile tracking. Our thesis has investigated precisely what information, particularly focusing on MAIDs and other PII, was collected by 25 iOS applications available in Norway and to whom it was sent. Additionally, our thesis investigated how the collected data was stated in the mobile apps' privacy labels to determine whether the apps transparently declare tracking to their users.

To investigate these areas, we developed three research questions, which we answered through an experiment that inspected network traffic from a mobile device. This section concludes these questions sequentially, emphasizing the methods employed and the results found from each investigation.

Research question one:

Which PII is collected by iOS applications during 15 minutes of use?

Our first research question delves into the intricate world of online advertising, which heavily relies on the collection and distribution of PII. We started our work by researching literature to gather insight into the utilization of PII and ADINT in the advertising industry, as outlined in Chapter 3. This gave us a comprehensive understanding of how the collection of PII is leveraged for tracking purposes, thereby facilitating ADINT.

We set up our experiment environment, simulating a MiTM attack, to intercept the network traffic from a mobile device. We downloaded, installed, and ran a set of applications for 15 minutes each.

As the App Store contains a large number of applications, there was a need for a thorough selection process, as described in Section 4.2. Therefore, we grouped the various categories into thematic groups, providing a feasible amount of applications across diverse App Store categories to investigate. Specifically, we ended up with a total of 25 applications, as shown in Table 5.4, equally distributed across our five thematic groups. Further, we defined a list of PII to look for, which based on four categories, comprised 17 types of PII.

As outlined in Section 5.2.2, 13 of the 17 types of PII were discovered in the traffic. Coarse location was the most collected, followed by the IP address, both indicating location tracking. Additionally, we found email address to be the third most collected PII, suggesting they also play a key role in the tracking industry. Furthermore, we found sensitive information such as precise location, IDFA, and gender, proving these to be key attributes in tracking as well. UDID, phone number, and name were among the least collected types, suggesting that these are less frequently employed for tracking purposes compared to the ones previously mentioned.

The IDFA appeared as the fifth most collected PII, originating from nine out of the 25 apps, indicating a high likelihood of tracking activities. Four types of PII, namely coarse location, gender, email address, and IP address, were found transmitted along the IDFA. Especially the transmission of email came to our attention, suggesting cross-referencing between the two identifiers, facilitating more detailed user profiling.

The apps collecting the IDFA primarily belong to the thematic groups *Education and Information* and *Entertainment and Creativity*, with two applications being considered sensitive due to their association with religious beliefs. Such sensitive apps may pose a threat to their users if the IDFA is known and correlated with these apps, risking sensitive information being disclosed about the users.

Research question two:

During the 15 minutes of app usage, which tracking domains are most frequently contacted?

Furthermore, our study investigated where the PII was sent, specifically focusing on third-party trackers. Based on the findings from four previous studies, as outlined in Section 4.3, we compiled a list of tracking domains utilized in our experiment.

A challenge we became aware of during the literature research, which also applied to our research, was the difficulty in verifying a domain's purpose. Information about tracking domains is poorly described online, despite several verdict websites, such as VirusTotal [67], being available. However, we managed to verify 111 tracking domains through online searches, as described in Section 4.4, which we used in our analysis to identify the most frequently contacted tracking domains during the 15 minutes of running the application.

Of the 111 tracking domains initially identified, our analysis identified 39 domains in the traffic, representing a 35% match rate. Alphabet owned the most frequently contacted domain, namely *googleads.g.doubleclick.net*, presumably serving as a common entry point for multiple marketing services that Google offers. Additionally, other Alphabet-owned domains were represented in the traffic, such as *googleapis.com*, confirming Google as one of the most powerful players in the advertising ecosystem [6]. Furthermore, we found that Meta's *graph.facebook.com* was the second most frequently contacted domain, reflecting Facebook's extensive data collection practices, as previously disclosed through the Cambridge Analytica scandal [37].

Included in our list of selected applications was the Byte-Dance-owned app CapCut, which was found to collect the IDFA. Consequently, we would expect to find tracking domains associated with ByteDance in our results. Regrettably, our list of tracking domains did not include any ByteDance-specific domains, resulting in these occurring undetected.

Research question three:

To what extent are the privacy labels transparent about the data collected for tracking purposes?

Following our second research question, we aimed to assess the transparency of privacy labels in disclosing tracking practices. This investigation is built upon research question one, where we found that nine apps collected the IDFA. In this part of our research, these apps were further examined, as they proved to track their users.

To investigate the respective apps' privacy labels, we examined the content in the labels *Data Used to Track You* and *Data Not Linked to You*. As the IDFA clearly is used for tracking, it should be declared in a label named *Data Used to Track You*. Similarly, as the IDFA is linked to a device, a contradiction arises if it is declared as *Data not linked to you* while being used for *Third-Party Advertising*.

Our investigation uncovers that privacy labels are unreliable. Only four out of the nine applications that collected the IDFA, transparently disclosed their tracking activity, as described in Section 5.4.2. Interestingly, one of the applications, namely Norwegian, failed to acknowledge any tracking whatsoever, despite transmitting the IDFA in three events.

As the privacy label intends to provide users with insight into how their PII is being used, it is important that these are honestly declared. When the privacy label fails to transparently disclose its data collection practices, the intended purpose of the privacy label vanishes, leaving users misinformed about the use of their personal data.

7.2 Further work

Our research has identified different tracking elements in our dataset on a network level and compared our findings to the privacy labels. In addition, we have discussed how this facilitates ADINT and exposure of sensitive information. Consequently, the research comprises several areas that could benefit from further work.

Expand datasets, time frames, and compare iOS against Android

Our research analyzed traffic from 25 applications, running for 15 minutes each, totaling six hours of usage. The number of selected applications, tracking domains identified, and time duration of the app usage significantly impacted our results. Expanding these parameters could enhance the data, facilitate better comparison across categories, and improve statistical accuracy. To address the challenges of manual data collection, a tool that automates installation, interaction, and analysis on the apps, similar to Koch *et al.* [4], could be developed.

Comparative studies between iPhone and Android devices regarding their handling of MAIDs and privacy implications would be beneficial. As our thesis focuses solely on iPhones, such an analysis could highlight platform-specific privacy approaches and MAID transmission. The expanded dataset and time duration would provide a comprehensive view of mobile privacy on both platforms, enriching the understanding of platform differences and informing users and policymakers about privacy standards and practices in the broader mobile ecosystem.

Target specific demographics

The six-hour period of traffic collection closely aligns with the 2022 Media Barometer finding that young Norwegians aged 16 to 24 use their mobile phones for about four and a half hours on average every day [68]. However, our results do not fully represent the usage patterns of this demographic or other age groups, as different demographics likely exhibit significant variations in their mobile usage. For instance, children likely have more games on their devices than social media apps due to age restrictions. Investigating PII collection and tracking domain contact based on apps typical for specific demographics could provide more targeted insights. Tailoring the study to include diverse apps popular among various age groups and interests might reveal nuanced differences in privacy exposure and data handling practices. Additionally, a comparison of app usage across different demographics could offer further insights into these variations.

Use AI for tracking domain verification

During our preprocessing stage regarding the tracking domains, we faced challenges in verifying the domains' purposes. We believe that developing more sophisticated tools or algorithms could enhance the efficiency and accuracy of this process. Existing methods, such as web scrapers and specialized apps like Lumen and ICSI Haystack [28], have laid a foundation, yet there is room for improvement. Advancements could include creating AI-driven models trained on extensive datasets of known tracking domains, which could automate and refine the verification process. Such tools would enhance domain validation and the reliability of privacy assessments, as conducted in our thesis.

Privacy labels and policies

As our study only examined privacy labels and not privacy policies, we suggest further work by examining whether the respective applications' privacy policies align with the data that is collected. This would determine whether the applications comply with their legal responsibilities to users.

Additionally, Koch *et al.* [4] recommend that Apple automate the creation of privacy labels to ensure they accurately reflect privacy policies without relying solely on developers. Moving from a single point of responsibility to incorporating Apple's quality assurance could significantly improve the integrity and reliability of privacy labels. Our results illustrate that little improvement in this area has been made since the investigation by Koch *et al*.

Bibliography

[1] P. Vines, F. Roesner and T. Kohno, ‘Exploring ADINT: Using ad targeting for surveillance on a budget - or - how Alice can buy ads to track Bob,’ in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017.

[2] C. Smales. ‘An overview of Google ads reports and metrics.’ (n.d.), [Online]. Available: `https://www.datafeedwatch.com/blog/google-ads-reports-metrics` (visited on 04/04/2024).

[3] OnAudience. ‘Mobile Ad ID: User identification for mobile ad campaign.’ (n.d.), [Online]. Available: `https://onaudience.com/mobile-ad-id-user-identification-for-mobile-ad-campaign` (visited on 02/03/2024).

[4] S. Koch, M. Wessels, B. Altpeter, M. Olvermann and M. Johns, ‘Keeping privacy labels honest,’ *Proc. Priv. Enhancing Technol.*, 2022.

[5] B. Hu, Q. Lin, Y. Zheng, Q. Yan, M. Troglia and Q. Wang, ‘Characterizing location-based mobile tracking in mobile ad networks,’ in *2019 IEEE Conference on Communications and Network Security (CNS)*, 2019.

[6] Datatilsynet. ‘Det store datakappløpet.’ (2015), [Online]. Available: `https://www.datatilsynet.no/globalassets/global/dokumenter-pdfer-skjema-ol/rettigheter-og-plikter/rapporter/kommersialisering-norsk-endelig.pdf` (visited on 01/02/2024).

[7] A. Arora. ‘Amazon DSP targeting: 11 targeting options explained.’ (2023), [Online]. Available: `https://adbrew.io/blog/amazon-dsp-targeting/` (visited on 20/05/2024).

[8] N. Figas. ‘How to build a Demand-Side Platform (DSP).’ (2024), [Online]. Available: `https://clearcode.cc/blog/how-to-build-a-demand-side-platform-dsp/#toc-label-1` (visited on 20/05/2024).

[9] SpotSense. ‘What is SDK mobile advertising? A comprehensive guide.’ (n.d.), [Online]. Available: `https://spotsense.io/what-is-sdk-mobile-advertising-a-comprehensive-guide/#` (visited on 18/05/2024).

[10] Etterretningstjenesten. ‘Begreper og definisjoner.’ (2024), [Online]. Available: `https://www.etterretningstjenesten.no/om-etterretning/begreper-og-definisjoner` (visited on 06/04/2024).

[11] O. Benjakob, ‘Revealed: Israeli cyber firms have developed an ’insane’ new spyware tool. No defense exists,’ *Haaretz*, 2023.

[12] H. Lied, M. Gundersen, T. Furuly, Ø. B. Skille, H. K. Jansson, M. Grafsrønningen, S. Grut and M. Arnesen, 'Avslørt av mobilen,' *Norsk rikskringkasting AS*, 2020.

[13] National Institute of Standards and Technology, *Personally identifiable information*, n.d. [Online]. Available: https://csrc.nist.gov/glossary/term/personally_identifiable_information (visited on 07/05/2024).

[14] Equifax. 'MAIDS: A useful alternative to third-party cookies.' (2024), [Online]. Available: https://www.equifax.com.au/knowledge-hub/marketing-services/maids-useful-alternative-third-party-cookies (visited on 07/05/2024).

[15] S. M. Kerner. 'What is a cookie?' (2021), [Online]. Available: https://www.techtarget.com/searchsoftwarequality/definition/cookie (visited on 07/05/2024).

[16] K. Yasar. 'Third-party cookie.' (2023), [Online]. Available: https://www.techtarget.com/whatis/definition/third-party-cookie (visited on 07/05/2024).

[17] Adjust. 'What is the identifier for advertisers (IDFA)?' (n.d.), [Online]. Available: https://www.adjust.com/glossary/idfa/ (visited on 07/05/2024).

[18] 'Advertisingidentifier,' Apple Developer, Tech. Rep., n.d. [Online]. Available: https://developer.apple.com/documentation/adsupport/asidentifiermanager/advertisingidentifier (visited on 01/03/2024).

[19] University of Virginia. 'Identifiers.' (n.d.), [Online]. Available: https://research.virginia.edu/irb-sbs/identifiers (visited on 20/03/2024).

[20] Adjust. 'What is App Tracking Transparency (ATT)?' (n.d.), [Online]. Available: https://www.adjust.com/glossary/app-tracking-transparency/ (visited on 03/03/2024).

[21] Appsflyer. 'App tracking transparency (ATT).' (n.d.), [Online]. Available: https://www.appsflyer.com/glossary/app-tracking-transparency/ (visited on 18/05/2024).

[22] K. Jerath, 'Mobile advertising and the impact of Apple's App Tracking Transparency policy,' *Apple Inc.*, 2022. [Online]. Available: https://www.apple.com/privacy/docs/Mobile_Advertising_and_the_Impact_of_Apples_App_Tracking_Transparency_Policy_April_2022.pdf (visited on 10/05/2024).

[23] Ironclad. 'What is a Privacy Policy? Everything you need to know.' (n.d.), [Online]. Available: https://ironcladapp.com/journal/contracts/how-to-create-the-best-privacy-policy-for-your-business/ (visited on 05/04/2024).

[24] 'App privacy details on the App Store,' Apple Inc., Tech. Rep., n.d. [Online]. Available: https:/developer.apple.com/app-store/app-privacy-details/#data-type-usage (visited on 20/04/2024).

[25] European Parliament and Council of the European Union, *GDPR Personal Data*, 2016. [Online]. Available: `https://gdpr-info.eu/issues/personal-data/` (visited on 09/05/2024).

[26] J. Govindaraj, R. Verma and G. Gupta, 'Analyzing mobile device ads to identify users,' in *12th IFIP International Conference on Digital Forensics (DF)*, 2016.

[27] Datatilsynet. 'Hva er en personopplysning?' (2023), [Online]. Available: `https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/` (visited on 02/03/2024).

[28] N. Vallina-Rodriguez, S. Sundaresan, A. Razaghpanah, R. Nithyanand, M. Allman, C. Kreibich and P. Gill, 'Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem,' *ArXiv*, 2016.

[29] B. Klais, 'New research across 200 ios apps hints that surveillance marketing is still going strong,' *URL Genius https://app. urlgeni. us/blog/new-research-across-200-ios-apps-hints-surveillance-marketing-maystill-be-going-strong*, 2022.

[30] F. Paci, J. Pizzoli and N. Zannone, 'A comprehensice study on third-party user tracking in mobile applications,' in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, 2023.

[31] M. M. Ali, D. G. Balash, C. Kanich and A. J. Aviv, 'Honesty is the best policy: On the accuracy of apple privacy labels compared to apps' privacy policies,' *ArXiv*, 2023.

[32] Rapid7. 'Man in the middle (mitm) attacks.' (n.d.), [Online]. Available: `https://www.rapid7.com/fundamentals/man-in-the-middle-attacks/` (visited on 08/05/2024).

[33] M. Sikorski and A. Honig, *Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software*. 38 Ringold Street, San Francisco, CA 94103: No Starch Press, 2012.

[34] J. Phuriphanvichai, *Use mitmdump to capture refinitiv real-time - Optimized content*, 2019. [Online]. Available: `https://developers.lseg.com/en/article-catalog/article/use-mitmdump-to-capture-rt-content` (visited on 20/04/2024).

[35] K. Collins. 'Should you put hyphens in domain names?' (n.d.), [Online]. Available: `https://www.fasthosts.co.uk/blog/hyphens-in-domain-names/` (visited on 09/05/2024).

[36] VirusTotal, *How it works*, 2023. [Online]. Available: `https://docs.virustotal.com/docs/how-it-works` (visited on 01/05/2024).

[37] N. Confessore. 'Cambridge Analytica and Facebook: The scandal and the fallout so far.' (2018), [Online]. Available: `https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html` (visited on 08/05/2024).

[38] M. Alfhaily, 'IPATool,' Tech. Rep., 2024, README.md. [Online]. Available: `https://github.com/majd/ipatool` (visited on 18/04/2024).

[39] 'Bundle ids,' Apple Developer, Tech. Rep., n.d. [Online]. Available: https://developer.apple.com/documentation/appstoreconnectapi/bundle_ids (visited on 10/05/2024).

[40] M. Black, *Keeping purchased apps between transfers of phones and apple ids*, 2021. [Online]. Available: https://discussions.apple.com/thread/253392767 (visited on 10/05/2024).

[41] Kaspersky. 'What is jailbreaking – Definition and explanation.' (n.d.), [Online]. Available: https://www.kaspersky.com/resource-center/definitions/what-is-jailbreaking (visited on 29/02/2024).

[42] 'What is SSL Pinning? – A quick walk through,' Indusface, Tech. Rep., n.d. [Online]. Available: https://www.indusface.com/learning/what-is-ssl-pinning-a-quick-walk-through/ (visited on 09/05/2024).

[43] LRVT. 'Rootful or Rootless: The current state of iOS jailbreaking for pentesters.' (2024), [Online]. Available: https://blog.lrvt.de/rootful-or-rootless-jailbreaks-for-pentesting (visited on 10/05/2024).

[44] C. Kidd. 'What is Splunk & what does it do? A Splunk intro.' (2024), [Online]. Available: https://www.splunk.com/en_us/blog/learn/what-splunk-does.html (visited on 22/04/2024).

[45] 'Managing indexers and clusters of indexers,' Splunk, Tech. Rep., n.d. [Online]. Available: https://docs.splunk.com/Documentation/Splunk/9.2.1/Indexer/Aboutindexesandindexers (visited on 13/05/2024).

[46] 'Source type,' Splunk, Tech. Rep., n.d. [Online]. Available: https://docs.splunk.com/Splexicon:Sourcetype (visited on 13/05/2024).

[47] A. Blacker. 'App store rank explained: What is an app's Apple Google store ranking and what impacts it?' (n.d.), [Online]. Available: https://apptopia.com/blog/app-store-rank-explained-what-is-an-apps-apple-google-store-ranking-and-what-impacts-it/ (visited on 04/05/2024).

[48] J. Penn, *'UI AutoMonkey'*, n.d. [Online]. Available: https://github.com/jonathanpenn/ui-auto-monkey (visited on 11/05/2024).

[49] 'Source type,' Splunk, Tech. Rep., 2023. [Online]. Available: https://docs.splunk.com/Documentation/Splunk/latest/Search/UseCASEandTERMtomatchphrases (visited on 04/05/2024).

[50] Apple Inc., *Gspe21-ssl.ls.apple.com*, 2024. [Online]. Available: https://discussions.apple.com/thread/255512236?sortBy=best (visited on 12/05/2024).

[51] R. Larsen. 'Demystifying TikTok's business and revenue model: An in-depth explanation.' (2024), [Online]. Available: https://www.untaylored.com/post/demystifying-tiktok-s-business-and-revenue-model-an-in-depth-explanation (visited on 12/05/2024).

[52] S. Tallaksrud. 'TikTok sporer finger-bevegelsene dine: – Innsamlingen er den mest omfattende jeg har sett.' (2023), [Online]. Available: https://www.tekna.no/magasinet/dette-gir-du-TikTok-tilgang-til/ (visited on 27/04/2024).

[53] Vita, *Databehandlere*. [Online]. Available: https://www.vita.no/databehandlere (visited on 14/05/2024).

[54] X. Business. 'Geo, gender, language, and age targeting.' (n.d.), [Online]. Available: https://business.x.com/en/help/campaign-setup/campaign-targeting/geo-gender-and-language-targeting.html (visited on 12/05/2024).

[55] 'Knowledge manager manual,' Tech. Rep., 2024. [Online]. Available: https://docs.splunk.com/Documentation/SplunkCloud/9.1.2312/Knowledge/Usesearchmacros (visited on 05/04/2024).

[56] 'Doubleclick.net - Domain info,' Netify, Tech. Rep., n.d. [Online]. Available: https://www.netify.ai/resources/domains/doubleclick.net (visited on 30/04/2024).

[57] M. Iqbal. 'TikTok revenue and usage statistics (2024).' (2024), [Online]. Available: https://www.businessofapps.com/data/tik-tok-statistics/ (visited on 15/05/2024).

[58] 'About fields,' Splunk, Tech. Rep., 2017. [Online]. Available: https://docs.splunk.com/Documentation/Splunk/latest/Knowledge/Aboutfields (visited on 18/05/2024).

[59] Straffeloven, *Lov om straff (straffeloven)*, § 205. Krenkelse av retten til privat kommunikasjon, 2009. [Online]. Available: https://lovdata.no/lov/2005-05-20-28/%C2%A7205 (visited on 07/05/2024).

[60] European Parliament and Council of the European Union, *Art. 6 GDPR Lawfulness of processing*, Article 6(1)(a), 2016. [Online]. Available: https://gdpr-info.eu/art-6-gdpr/ (visited on 05/05/2024).

[61] Straffeloven, *Lov om straff (straffeloven)*, § 202. Identitetskrenkelse, 2009. [Online]. Available: https://lovdata.no/dokument/NL/lov/2005-05-20-28/KAPITTEL_2-6#%C2%A7202 (visited on 07/05/2024).

[62] Åndsverkloven, *Lov om opphavsrett til åndsverk mv. (åndsverkloven)*, § 37. Gjengivelse av kunstverk og fotografiske verk i kritisk og vitenskapelig fremstilling og biografier, 2018. [Online]. Available: https://lovdata.no/lov/2018-06-15-40/%C2%A737 (visited on 05/05/2024).

[63] United Nations. 'FNs bærekraftsmål.' (2024), [Online]. Available: https://fn.no/om-fn/fns-baerekraftsmaal (visited on 08/05/2024).

[64] United Nations. 'Fred, rettferdighet og velfungerende institusjoner.' (2023), [Online]. Available: https://fn.no/om-fn/fns-baerekraftsmaal/fred-rettferdighet-og-velfungerende-institusjoner (visited on 08/05/2024).

[65] J. Fulwood, J. Morrison, D. J. Boorman and P. C. Rudoplh, *Cybersecurity and sustainable development and intersectional analysis*, 2022.

[66] United Nations. 'Industri, innovasjon og infrastruktur.' (2023), [Online]. Available: https://fn.no/om-fn/fns-baerekraftsmaal/industri-innovasjon-og-infrastruktur (visited on 09/05/2024).

[67] 'Virus Total (Preview),' Tech. Rep., 2024. [Online]. Available: https://learn.microsoft.com/en-us/connectors/virustotal/.

[68] E. Schiro, 'Norsk mediebarometer 2022,' *Artikler om norsk mediebarometer*, 2024.

Appendix A

Dialogue with Martin Gundersen regarding NRK's SKUP report

Samtale om SKUP-rapport om mobilsporing av NRK
Dato: 13.02.2024

Samtale med Martin Gundersen, NRK

Martin Gundersen er journalist fra NRK og skriver artikler om personvern, sikkerhet og teknologi. Han har vært en sentral bidragsyter til NRKs graveprosjekt om mobilsporing. Vi tok kontakt med Martin gjennom Signal den 5. februar, der vi raskt fikk respons samme dag, for å høre mer om prosjektet og oppklare noen detaljer gruppen var nysgjerrige på.

Dato: 13.02.2024

Tid: kl. 16:28 – 16:43

Sted: Digitalt over Signal

Tilstede: Martin Gundersen, Sofie Hagen, Oda Motrøen-Sevilhaug og Susanne Kolberg

Samtale:

Sofie: Hvor stoppet NRKs prosjektet opp?

Martin: Dette var et graveprosjekt som startet med å kjøpe lokasjonsdata, litt fordi flere snakket om hvor verdifull denne dataen var, og hvor detaljert den var.

Det var ikke så store tanker om hva det skulle bli til til slutt. Og det tok omtrent 6 måneder før vi fikk sporet en enkeltperson i stavanger. Deretter fulgte vi opp mot Forsvaret og nasjonal sikkerhet. Den tredje store saken var hvordan data fra mobilapper fløt mellom ulike aktører og til slutt havnet hos et selskap som hadde amerikanske myndigheter på kundelisten. Da gikk det fra å være uskyldig markedsføringsbehov til nasjonal sikkerhets-overvåkning.

Videre har vi fulgt journalistikk fra andre, eksempelvis litt om real time bidding der det deles mye data om brukerne. Videre hadde vi ingen gode spor å jobbe videre på som følge av mangel på kilder å ta det videre med. Deretter fulgte vi opp med noen saker om hvordan hack industries brukte statisk analyse av mobilapper til å finne SDK og omtalte noen selskaper som hadde dette, der noen tok det ut fra appen etterpå fordi NRK tok kontakt. Det ble ikke mer etter dette. Men vi følger fortsatt med på feltet.

Teknologien endrer seg veldig fortløpende. Blant annet har Android har blitt strengere. Levekårene for selskaper som hadde SDK, altså en kodesnutt i appene for å hente ut data, ble vanskeligere. I tillegg er real time bidding prosessen veldig komplisert og vanskelig å gå inn i uten et godt utgangspunkt. Dermed førte dette til at NRK ikke fant et videre naturlig neste sted å fortsette.

Sofie: Er det noen gjøremål dere ikke fikk gjennomført?

Martin: Ja og nei. Vi fikk aldri fullt kartlagt nasjonal sikkerhets-sporet, noe vi gjerne skulle gjort. Men dette gjorde Haaretz og Washington journal. Det er ofte ting i journalist-prosjekt man ikke blir helt ferdig med. Nasjonal sikkerhet er det som er interessant å jobbe videre med, samt real time bidding, og hvordan det blir misbrukt.

Samtale om SKUP-rapport om mobilsporing av NRK
Dato: 13.02.2024

Sofie: Hva har dere sett for dere at blir veien videre?

Martin: Det er andre temaer som følges tettere. Dunhammer, kabelovervåkningsprosjekt med USA fra Danmark, og AI. Personvern er også interessant. Det er ikke noe videre arbeid relatert til skup-rapporten, men som journalist kan en heller ikke fortelle om hva som jobbes med videre hvis dette er upublisert journalistikk. Men vi har antenner ute og feltet følges med på. For å bli en god nyhetssak må det likevel kunne bli konkret nok, dokumenteres og oppleves relevant for norske borgere.

Det siste spørsmålet som stilles skal egentlig være «Side 16 punkt 10 - her lurer vi på om SDK (Software Development Kit) er det som beskrives?», men Martin har ikke rapporten foran seg. Vi spør derfor om han kan forklare hva en SDK er.

Sofie: Hva er en SDK?

Martin: Dette er et kodebibliotek, litt som på GitHub når du installerer et bibliotek som du kan bygge på, og er noe en app-utvikler kan legge inn i appen som den utvikler. Du kan eksempelvis ha en for google maps, og en for å sende data ut fra appen. Det er ulike aktører som har dette, og leverandørene betalte gjerne mobilutviklerne for å legge dette inn i appen for å få ut data. Det kunne være at man fikk betalt en viss sum, eksempelvis per 1000 brukere. Dette er den mest brukte metoden for å hente ut lokasjonsdata.

Appendix B

Python addon script: to append relative timestamp

```
###/
#  This python script adds "relative" timestamp to files.
#  The code is reused from the article:
#  "Use mitmdump to capture Refinitiv Real-Time - Optimized content"
#  Source: https://developers.lseg.com/en/article-catalog/article/use-mitmdump-to-capture-rt-content
#  @file: py_timestamp.py
#  @author: Jirapongse Phuriphanvichai
###/

# Importing modules
import sys
import re
from datetime import datetime

# Main program
fileObject = None
try:
   if len(sys.argv) == 2:
      fileObject=open(sys.argv[1],"w")
   for line in sys.stdin:
      if line.strip() != "":
        if re.match("^(?:[0-9]{1,3}\.){3}[0-9]{1,3}:[0-9]*", line.strip()):
            line = "\n"+datetime.now().strftime('%Y-%m-%d %H:%M:%S.%f')[:-3] + ": " + line
        print(line)
        if fileObject:
            fileObject.write(line)
            fileObject.flush()
except:
   if fileObject:
      fileObject.close()
```

Appendix C

Preprocessing of domains

DOMAIN PREPROSESSING DOCUMENTATION

1. REMOVING DUPLICATE DOMAINS

Unique domains after sort -u command:	
2mdn.net	1
adcolony.com	1
adjust.com	1
adnxs.com	1
adobedtm.com	1
amazon-adsystem.com	1
amazonaws.com	1
amazon–adsystem.com	1
amp-api.apps.apple.com	1
amplitude.com	1
analytics.localytics.com	1
api-adservices.apple.com	1
api.apptentive.com	1
api.mixpanel.com	1
api.segment.io	1
api.snapkit.com	1
api2.branch.io	1
app-measurement.com	1
app.adjust.com	1
appboy-images.com	1
applovin.com	1
appsflyer.com	1
arcus-uswest.amazon.com	1
assets.adobedtm.com	1
attr.appsflyer.com	1
azureedge.net	1
braze-images.com	1
bugsnag.com	1
c.amazon-adsystem.com	1
c00.adobe.com	1
ca.iadsdk.apple.com	1
cdn-settings.segment.com	1
cdn.branch.io	1
cdn.cookielaw.org	1
cdn.optimizely.com	1
chartboost.com	1
clients3.google.com	1
combine.urbanairship.com	1
config.emb-api.com	1
config2.mparticle.com	1
control.kochava.com	1
conversions.appsflyer.com	1
crashlytics.com	1
criteo.com	1
crittercism.com	1
d-xxxxxxxxxx.cloudfront.net	1
data.emb-api.com	1
demdex.net	1
device-api.urbanairship.com	1
device-metrics-us-2.amazon.com	1
device-provisioning.googleapis.com	1
doubleclick.net	1
dpm.demdex.net	1
facebook.com	1
fcmtoken.googleapis.com	1
firebase-settings.crashlytics.com	1
firebasedynamiclinks.googleapis.com	1
firebaseinappmessaging.googleapis.com	1
firebaseinstallations.googleapis.com	1
firebaselogging-pa.googleapis.com	1
firebaseremoteconfig.googleapis.com	1
fls-na.amazon.com	1
flurry.com	1

Removed by this step:
unknown
27

fonts.googleapis.com	1
fonts.gstatic.com	1
ft.com	1
gcdsdk.appsflyer.com	1
google-analytics.com	1
google.com	1
googleads.g.doubleclick.net	1
googleadservices.com	1
googleapis.com	1
googlesyndication.com	1
googletagmanager.com	1
googletagservices.com	1
googleusercontent.com	1
google−analytics.com	1
graph.facebook.com	1
gsp-ssl.ls.apple.com	1
gstatic.com	1
identity.mparticle.com	1
ild.googleapis.com	1
images-na.ssl-images-amazon.com	1
in.com	1
inappcheck.itunes.apple.com	1
inapps.appsflyer.com	1
inmobi.com	1
ioam.de	1
itunes.apple.com	1
jpush.cn	1
kvinit-prod.api.kochava.com	1
launches.appsflyer.com	1
lh3.googleusercontent.com	1
localytics.com	1
logx.optimizely.com	1
m.media-amazon.com	1
mads.amazon-adsystem.com	1
manifest.localytics.com	1
mcias-va7.cloud.adobe.io	1
mesu.apple.com	1
mixpanel.com	1
moatads.com	1
mobile-collector.newrelic.com	1
mobile-data.onetrust.io	1
mobileapptracking.com	1
mopub.com	1
msh.amazon.com	1
nativesdks.mparticle.com	1
newrelic.com	1
ntent.com	1
oauthaccountmanager.googleapis.com	1
ocsp.digicert.com	1
ocsp.pki.goog	1
ocsp.sectigo.com	1
omtrdc.net	1
onesignal.com	1
optanon.blob.core.windows.net	1
paypalobjects.com	1
play.googleapis.com	1
profile.localytics.com	1
pubads.g.doubleclick.net	1
px-conf.perimeterx.net	1
r3.0.lencr.org	1
remote-data.urbanairship.com	1
s.amazon-adsystem.com	1
s3.amazonaws.com	1
sb.scorecardresearch.com	1
scontent-ort2-1.xx.fbcdn.net	1
scorecardresearch.com	1
sdk-assets.localytics.com	1
sdk.iad-01.braze.com	1
sdk.iad-03.braze.com	1
sentry.io	1

sessions.bugsnag.com	1
skadsdk.appsflyer.com	1
skadsdkless.appsflyer.com	1
sp.auth.adobe.com	1
ssl.google-analytics.com	1
supersonicads.com	1
tpc.googlesyndication.com	1
unagi.amazon.com	1
urbanairship.com	1
vungle.com	1
web.facebook.com	1
www.google-analytics.com	1
www.google.com	1
www.googleapis.com	1
www.googletagmanager.com	1
www.googletagservices.com	1
www.gstatic.com	1
www.paypalobjects.com	1
yahoo.com	1
youtube.com	1
z.moatads.com	1
154	

2. REMOVING MISSPELLED DOMAINS

2. Unique domains (after domains with minus are removed) :	
2mdn.net	1
adcolony.com	1
adjust.com	1
adnxs.com	1
adobedtm.com	1
amazon-adsystem.com	1
amazonaws.com	1
amp-api.apps.apple.com	1
amplitude.com	1
analytics.localytics.com	1
api-adservices.apple.com	1
api.apptentive.com	1
api.mixpanel.com	1
api.segment.io	1
api.snapkit.com	1
api2.branch.io	1
app-measurement.com	1
app.adjust.com	1
appboy-images.com	1
applovin.com	1
appsflyer.com	1
arcus-uswest.amazon.com	1
assets.adobedtm.com	1
attr.appsflyer.com	1
azureedge.net	1
braze-images.com	1
bugsnag.com	1
c.amazon-adsystem.com	1
c00.adobe.com	1
ca.iadsdk.apple.com	1
cdn-settings.segment.com	1
cdn.branch.io	1
cdn.cookielaw.org	1
cdn.optimizely.com	1
chartboost.com	1
clients3.google.com	1
combine.urbanairship.com	1
config.emb-api.com	1
config2.mparticle.com	1
control.kochava.com	1
conversions.appsflyer.com	1

Removed by this step:
google−analytics.com
amazon−adsystem.com
2

crashlytics.com	1
criteo.com	1
crittercism.com	1
d-xxxxxxxxxx.cloudfront.net	1
data.emb-api.com	1
demdex.net	1
device-api.urbanairship.com	1
device-metrics-us-2.amazon.com	1
device-provisioning.googleapis.com	1
doubleclick.net	1
dpm.demdex.net	1
facebook.com	1
fcmtoken.googleapis.com	1
firebase-settings.crashlytics.com	1
firebasedynamiclinks.googleapis.com	1
firebaseinappmessaging.googleapis.com	1
firebaseinstallations.googleapis.com	1
firebaselogging-pa.googleapis.com	1
firebaseremoteconfig.googleapis.com	1
fls-na.amazon.com	1
flurry.com	1
fonts.googleapis.com	1
fonts.gstatic.com	1
ft.com	1
gcdsdk.appsflyer.com	1
google-analytics.com	1
google.com	1
googleads.g.doubleclick.net	1
googleadservices.com	1
googleapis.com	1
googlesyndication.com	1
googletagmanager.com	1
googletagservices.com	1
googleusercontent.com	1
graph.facebook.com	1
gsp-ssl.ls.apple.com	1
gstatic.com	1
identity.mparticle.com	1
ild.googleapis.com	1
images-na.ssl-images-amazon.com	1
in.com	1
inappcheck.itunes.apple.com	1
inapps.appsflyer.com	1
inmobi.com	1
ioam.de	1
itunes.apple.com	1
jpush.cn	1
kvinit-prod.api.kochava.com	1
launches.appsflyer.com	1
lh3.googleusercontent.com	1
localytics.com	1
logx.optimizely.com	1
m.media-amazon.com	1
mads.amazon-adsystem.com	1
manifest.localytics.com	1
mcias-va7.cloud.adobe.io	1
mesu.apple.com	1
mixpanel.com	1
moatads.com	1
mobile-collector.newrelic.com	1
mobile-data.onetrust.io	1
mobileapptracking.com	1
mopub.com	1
msh.amazon.com	1
nativesdks.mparticle.com	1
newrelic.com	1
ntent.com	1
oauthaccountmanager.googleapis.com	1
ocsp.digicert.com	1
ocsp.pki.goog	1

ocsp.sectigo.com	1
omtrdc.net	1
onesignal.com	1
optanon.blob.core.windows.net	1
paypalobjects.com	1
play.googleapis.com	1
profile.localytics.com	1
pubads.g.doubleclick.net	1
px-conf.perimeterx.net	1
r3.0.lencr.org	1
remote-data.urbanairship.com	1
s.amazon-adsystem.com	1
s3.amazonaws.com	1
sb.scorecardresearch.com	1
scontent-ort2-1.xx.fbcdn.net	1
scorecardresearch.com	1
sdk-assets.localytics.com	1
sdk.iad-01.braze.com	1
sdk.iad-03.braze.com	1
sentry.io	1
sessions.bugsnag.com	1
skadsdk.appsflyer.com	1
skadsdkless.appsflyer.com	1
sp.auth.adobe.com	1
ssl.google-analytics.com	1
supersonicads.com	1
tpc.googlesyndication.com	1
unagi.amazon.com	1
urbanairship.com	1
vungle.com	1
web.facebook.com	1
www.google-analytics.com	1
www.google.com	1
www.googleapis.com	1
www.googletagmanager.com	1
www.googletagservices.com	1
www.gstatic.com	1
www.paypalobjects.com	1
yahoo.com	1
youtube.com	1
z.moatads.com	1
152	

REMOVING DOMAINS WITH BROAD PURPOSES

Domains after broad purpose domains are removed:		**Removed by this step:**	
2mdn.net	1	amazonaws.com	1
adcolony.com	1	www.google.com	1
adjust.com	1	yahoo.com	1
adnxs.com	1	youtube.com	1
adobedtm.com	1	google.com	1
amazon-adsystem.com	1	facebook.com	1
amp-api.apps.apple.com	1	**6**	
amplitude.com	1		
analytics.localytics.com	1		
api-adservices.apple.com	1		
api.apptentive.com	1		
api.mixpanel.com	1		
api.segment.io	1		
api.snapkit.com	1		
api2.branch.io	1		
app-measurement.com	1		
app.adjust.com	1		
appboy-images.com	1		
applovin.com	1		
appsflyer.com	1		
arcus-uswest.amazon.com	1		

assets.adobedtm.com	1
attr.appsflyer.com	1
azureedge.net	1
braze-images.com	1
bugsnag.com	1
c.amazon-adsystem.com	1
c00.adobe.com	1
ca.iadsdk.apple.com	1
cdn-settings.segment.com	1
cdn.branch.io	1
cdn.cookielaw.org	1
cdn.optimizely.com	1
chartboost.com	1
clients3.google.com	1
combine.urbanairship.com	1
config.emb-api.com	1
config2.mparticle.com	1
control.kochava.com	1
conversions.appsflyer.com	1
crashlytics.com	1
criteo.com	1
crittercism.com	1
d-xxxxxxxxxx.cloudfront.net	1
data.emb-api.com	1
demdex.net	1
device-api.urbanairship.com	1
device-metrics-us-2.amazon.com	1
device-provisioning.googleapis.com	1
doubleclick.net	1
dpm.demdex.net	1
fcmtoken.googleapis.com	1
firebase-settings.crashlytics.com	1
firebasedynamiclinks.googleapis.com	1
firebaseinappmessaging.googleapis.com	1
firebaseinstallations.googleapis.com	1
firebaselogging-pa.googleapis.com	1
firebaseremoteconfig.googleapis.com	1
fls-na.amazon.com	1
flurry.com	1
fonts.googleapis.com	1
fonts.gstatic.com	1
ft.com	1
gcdsdk.appsflyer.com	1
google-analytics.com	1
googleads.g.doubleclick.net	1
googleadservices.com	1
googleapis.com	1
googlesyndication.com	1
googletagmanager.com	1
googletagservices.com	1
googleusercontent.com	1
graph.facebook.com	1
gsp-ssl.ls.apple.com	1
gstatic.com	1
identity.mparticle.com	1
ild.googleapis.com	1
images-na.ssl-images-amazon.com	1
in.com	1
inappcheck.itunes.apple.com	1
inapps.appsflyer.com	1
inmobi.com	1
ioam.de	1
itunes.apple.com	1
jpush.cn	1
kvinit-prod.api.kochava.com	1
launches.appsflyer.com	1
lh3.googleusercontent.com	1
localytics.com	1
logx.optimizely.com	1
m.media-amazon.com	1

mads.amazon-adsystem.com	1
manifest.localytics.com	1
mcias-va7.cloud.adobe.io	1
mesu.apple.com	1
mixpanel.com	1
moatads.com	1
mobile-collector.newrelic.com	1
mobile-data.onetrust.io	1
mobileapptracking.com	1
mopub.com	1
msh.amazon.com	1
nativesdks.mparticle.com	1
newrelic.com	1
ntent.com	1
oauthaccountmanager.googleapis.com	1
ocsp.digicert.com	1
ocsp.pki.goog	1
ocsp.sectigo.com	1
omtrdc.net	1
onesignal.com	1
optanon.blob.core.windows.net	1
paypalobjects.com	1
play.googleapis.com	1
profile.localytics.com	1
pubads.g.doubleclick.net	1
px-conf.perimeterx.net	1
r3.0.lencr.org	1
remote-data.urbanairship.com	1
s.amazon-adsystem.com	1
s3.amazonaws.com	1
sb.scorecardresearch.com	1
scontent-ort2-1.xx.fbcdn.net	1
scorecardresearch.com	1
sdk-assets.localytics.com	1
sdk.iad-01.braze.com	1
sdk.iad-03.braze.com	1
sentry.io	1
sessions.bugsnag.com	1
skadsdk.appsflyer.com	1
skadsdkless.appsflyer.com	1
sp.auth.adobe.com	1
ssl.google-analytics.com	1
supersonicads.com	1
tpc.googlesyndication.com	1
unagi.amazon.com	1
urbanairship.com	1
vungle.com	1
web.facebook.com	1
www.google-analytics.com	1
www.googleapis.com	1
www.googletagmanager.com	1
www.googletagservices.com	1
www.gstatic.com	1
www.paypalobjects.com	1
z.moatads.com	1
146	

3 AND 4: DOMAIN VERIFICATION THROUGH VIRUSTOTAL AND OSINT

Domains verified through VirusTotal:		Domains verified through OSINT:		Removed by these steps:	
2mdn.net	1	2mdn.net		amp-api.apps.apple.com	1
adcolony.com	1	adcolony.com		arcus-uswest.amazon.com	1
adjust.com		adjust.com	1	azureedge.net	1
adnxs.com	1	adnxs.com		cdn.cookielaw.org	1
adobedtm.com	1	adobedtm.com		clients3.google.com	1
amazon-adsystem.com	1	amazon-adsystem.com		config.emb-api.com	1
amp-api.apps.apple.com		amp-api.apps.apple.com		d-xxxxxxxxxx.cloudfront.net	1
amplitude.com	1	amplitude.com		device-provisioning.googleapis.com	1
analytics.localytics.com	1	analytics.localytics.com		fcmtoken.googleapis.com	1

api-adservices.apple.com	1	api-adservices.apple.com		fonts.googleapis.com	1
api.apptentive.com	1	api.apptentive.com		fonts.gstatic.com	1
api.mixpanel.com	1	api.mixpanel.com		ft.com	1
api.segment.io	1	api.segment.io		googleusercontent.com	1
api.snapkit.com		api.snapkit.com	1	gsp-ssl.ls.apple.com	1
api2.branch.io	1	api2.branch.io		gstatic.com	1
app-measurement.com	1	app-measurement.com		images-na.ssl-images-amazon.com	1
app.adjust.com	1	app.adjust.com		in.com	1
appboy-images.com	1	appboy-images.com		inappcheck.itunes.apple.com	1
applovin.com	1	applovin.com		itunes.apple.com	1
appsflyer.com	1	appsflyer.com		lh3.googleusercontent.com	1
arcus-uswest.amazon.com		arcus-uswest.amazon.com		mcias-va7.cloud.adobe.io	1
assets.adobedtm.com	1	assets.adobedtm.com		mesu.apple.com	1
attr.appsflyer.com	1	attr.appsflyer.com		mobile-data.onetrust.io	1
azureedge.net		azureedge.net		msh.amazon.com	1
braze-images.com		braze-images.com	1	oauthaccountmanager.googleapis.com	1
bugsnag.com		bugsnag.com	1	ocsp.digicert.com	1
c.amazon-adsystem.com	1	c.amazon-adsystem.com		ocsp.sectigo.com	1
c00.adobe.com		c00.adobe.com	1	optanon.blob.core.windows.net	1
ca.iadsdk.apple.com	1	ca.iadsdk.apple.com		paypalobjects.com	1
cdn-settings.segment.com	1	cdn-settings.segment.com		px-conf.perimeterx.net	1
cdn.branch.io	1	cdn.branch.io		s3.amazonaws.com	1
cdn.cookielaw.org		cdn.cookielaw.org		unagi.amazon.com	1
cdn.optimizely.com	1	cdn.optimizely.com		web.facebook.com	1
chartboost.com	1	chartboost.com		www.gstatic.com	1
clients3.google.com		clients3.google.com		www.paypalobjects.com	1
combine.urbanairship.com	1	combine.urbanairship.com		**35**	
config.emb-api.com		config.emb-api.com			
config2.mparticle.com	1	config2.mparticle.com			
control.kochava.com	1	control.kochava.com			
conversions.appsflyer.com		conversions.appsflyer.com	1		
crashlytics.com		crashlytics.com	1		
criteo.com	1	criteo.com			
crittercism.com		crittercism.com	1		
d-xxxxxxxxxx.cloudfront.net		d-xxxxxxxxxx.cloudfront.net			
data.emb-api.com		data.emb-api.com	1		
demdex.net		demdex.net	1		
device-api.urbanairship.com	1	device-api.urbanairship.com			
device-metrics-us-2.amazon.com	1	device-metrics-us-2.amazon.com			
device-provisioning.googleapis.com		device-provisioning.googleapis.com			
doubleclick.net	1	doubleclick.net			
dpm.demdex.net		dpm.demdex.net	1		
fcmtoken.googleapis.com		fcmtoken.googleapis.com			
firebase-settings.crashlytics.com	1	firebase-settings.crashlytics.com			
firebasedynamiclinks.googleapis.com		firebasedynamiclinks.googleapis.com	1		
firebaseinappmessaging.googleapis.com		firebaseinappmessaging.googleapis.com	1		
firebaseinstallations.googleapis.com		firebaseinstallations.googleapis.com	1		
firebaselogging-pa.googleapis.com	1	firebaselogging-pa.googleapis.com			
firebaseremoteconfig.googleapis.com		firebaseremoteconfig.googleapis.com	1		
fls-na.amazon.com	1	fls-na.amazon.com			
flurry.com	1	flurry.com			
fonts.googleapis.com		fonts.googleapis.com			
fonts.gstatic.com		fonts.gstatic.com			
ft.com		ft.com			
gcdsdk.appsflyer.com	1	gcdsdk.appsflyer.com			
google-analytics.com	1	google-analytics.com			
googleads.g.doubleclick.net	1	googleads.g.doubleclick.net			
googleadservices.com	1	googleadservices.com			
googleapis.com		googleapis.com	1		
googlesyndication.com	1	googlesyndication.com			
googletagmanager.com		googletagmanager.com	1		
googletagservices.com	1	googletagservices.com			
googleusercontent.com		googleusercontent.com			
graph.facebook.com		graph.facebook.com	1		
gsp-ssl.ls.apple.com		gsp-ssl.ls.apple.com			
gstatic.com		gstatic.com			
identity.mparticle.com	1	identity.mparticle.com			
ild.googleapis.com		ild.googleapis.com	1		
images-na.ssl-images-amazon.com		images-na.ssl-images-amazon.com			
in.com		in.com			

inappcheck.itunes.apple.com		inappcheck.itunes.apple.com	
inapps.appsflyer.com	1	inapps.appsflyer.com	
inmobi.com	1	inmobi.com	
ioam.de		ioam.de	1
itunes.apple.com		itunes.apple.com	
jpush.cn	1	jpush.cn	
kvinit-prod.api.kochava.com	1	kvinit-prod.api.kochava.com	
launches.appsflyer.com		launches.appsflyer.com	1
lh3.googleusercontent.com		lh3.googleusercontent.com	
localytics.com	1	localytics.com	
logx.optimizely.com	1	logx.optimizely.com	
m.media-amazon.com		m.media-amazon.com	1
mads.amazon-adsystem.com	1	mads.amazon-adsystem.com	
manifest.localytics.com		manifest.localytics.com	1
mcias-va7.cloud.adobe.io		mcias-va7.cloud.adobe.io	
mesu.apple.com		mesu.apple.com	
mixpanel.com	1	mixpanel.com	
moatads.com	1	moatads.com	
mobile-collector.newrelic.com	1	mobile-collector.newrelic.com	
mobile-data.onetrust.io		mobile-data.onetrust.io	
mobileapptracking.com	1	mobileapptracking.com	
mopub.com		mopub.com	1
msh.amazon.com		msh.amazon.com	
nativesdks.mparticle.com	1	nativesdks.mparticle.com	
newrelic.com	1	newrelic.com	
ntent.com		ntent.com	1
oauthaccountmanager.googleapis.com		oauthaccountmanager.googleapis.com	
ocsp.digicert.com		ocsp.digicert.com	
ocsp.pki.goog		ocsp.pki.goog	1
ocsp.sectigo.com		ocsp.sectigo.com	
omtrdc.net		omtrdc.net	1
onesignal.com		onesignal.com	1
optanon.blob.core.windows.net		optanon.blob.core.windows.net	
paypalobjects.com		paypalobjects.com	
play.googleapis.com		play.googleapis.com	1
profile.localytics.com		profile.localytics.com	1
pubads.g.doubleclick.net	1	pubads.g.doubleclick.net	
px-conf.perimeterx.net		px-conf.perimeterx.net	
r3.0.lencr.org		r3.0.lencr.org	1
remote-data.urbanairship.com	1	remote-data.urbanairship.com	
s.amazon-adsystem.com	1	s.amazon-adsystem.com	
s3.amazonaws.com		s3.amazonaws.com	
sb.scorecardresearch.com		sb.scorecardresearch.com	1
scontent-ort2-1.xx.fbcdn.net		scontent-ort2-1.xx.fbcdn.net	1
scorecardresearch.com		scorecardresearch.com	1
sdk-assets.localytics.com		sdk-assets.localytics.com	1
sdk.iad-01.braze.com	1	sdk.iad-01.braze.com	
sdk.iad-03.braze.com	1	sdk.iad-03.braze.com	
sentry.io		sentry.io	1
sessions.bugsnag.com	1	sessions.bugsnag.com	
skadsdk.appsflyer.com		skadsdk.appsflyer.com	1
skadsdkless.appsflyer.com	1	skadsdkless.appsflyer.com	
sp.auth.adobe.com		sp.auth.adobe.com	1
ssl.google-analytics.com	1	ssl.google-analytics.com	
supersonicads.com	1	supersonicads.com	
tpc.googlesyndication.com	1	tpc.googlesyndication.com	
unagi.amazon.com		unagi.amazon.com	
urbanairship.com	1	urbanairship.com	
vungle.com	1	vungle.com	
web.facebook.com		web.facebook.com	
www.google-analytics.com	1	www.google-analytics.com	
www.googleapis.com		www.googleapis.com	1
www.googletagmanager.com		www.googletagmanager.com	1
www.googletagservices.com	1	www.googletagservices.com	
www.gstatic.com		www.gstatic.com	
www.paypalobjects.com		www.paypalobjects.com	
z.moatads.com	1	z.moatads.com	
71		**40**	

FINAL LIST OF VERIFIED DOMAINS

Final list:	
2mdn.net	1
adcolony.com	1
adjust.com	1
adnxs.com	1
adobedtm.com	1
amazon-adsystem.com	1
amplitude.com	1
analytics.localytics.com	1
api-adservices.apple.com	1
api.apptentive.com	1
api.mixpanel.com	1
api.segment.io	1
api.snapkit.com	1
api2.branch.io	1
app-measurement.com	1
app.adjust.com	1
appboy-images.com	1
applovin.com	1
appsflyer.com	1
assets.adobedtm.com	1
attr.appsflyer.com	1
braze-images.com	1
bugsnag.com	1
c.amazon-adsystem.com	1
c00.adobe.com	1
ca.iadsdk.apple.com	1
cdn-settings.segment.com	1
cdn.branch.io	1
cdn.optimizely.com	1
chartboost.com	1
combine.urbanairship.com	1
config2.mparticle.com	1
control.kochava.com	1
conversions.appsflyer.com	1
crashlytics.com	1
criteo.com	1
crittercism.com	1
data.emb-api.com	1
demdex.net	1
device-api.urbanairship.com	1
device-metrics-us-2.amazon.com	1
doubleclick.net	1
dpm.demdex.net	1
firebase-settings.crashlytics.com	1
firebasedynamiclinks.googleapis.com	1
firebaseinappmessaging.googleapis.com	1
firebaseinstallations.googleapis.com	1
firebaselogging-pa.googleapis.com	1
firebaseremoteconfig.googleapis.com	1
fls-na.amazon.com	1
flurry.com	1
gcdsdk.appsflyer.com	1
google-analytics.com	1
googleads.g.doubleclick.net	1
googleadservices.com	1
googleapis.com	1
googlesyndication.com	1
googletagmanager.com	1
googletagservices.com	1
graph.facebook.com	1
identity.mparticle.com	1
ild.googleapis.com	1
inapps.appsflyer.com	1
inmobi.com	1
ioam.de	1
jpush.cn	1
kvinit-prod.api.kochava.com	1

launches.appsflyer.com	1
localytics.com	1
logx.optimizely.com	1
m.media-amazon.com	1
mads.amazon-adsystem.com	1
manifest.localytics.com	1
mixpanel.com	1
moatads.com	1
mobile-collector.newrelic.com	1
mobileapptracking.com	1
mopub.com	1
nativesdks.mparticle.com	1
newrelic.com	1
ntent.com	1
ocsp.pki.goog	1
omtrdc.net	1
onesignal.com	1
play.googleapis.com	1
profile.localytics.com	1
pubads.g.doubleclick.net	1
r3.0.lencr.org	1
remote-data.urbanairship.com	1
s.amazon-adsystem.com	1
sb.scorecardresearch.com	1
scontent-ort2-1.xx.fbcdn.net	1
scorecardresearch.com	1
sdk-assets.localytics.com	1
sdk.iad-01.braze.com	1
sdk.iad-03.braze.com	1
sentry.io	1
sessions.bugsnag.com	1
skadsdk.appsflyer.com	1
skadsdkless.appsflyer.com	1
sp.auth.adobe.com	1
ssl.google-analytics.com	1
supersonicads.com	1
tpc.googlesyndication.com	1
urbanairship.com	1
vungle.com	1
www.google-analytics.com	1
www.googleapis.com	1
www.googletagmanager.com	1
www.googletagservices.com	1
z.moatads.com	1
111	

Appendix D

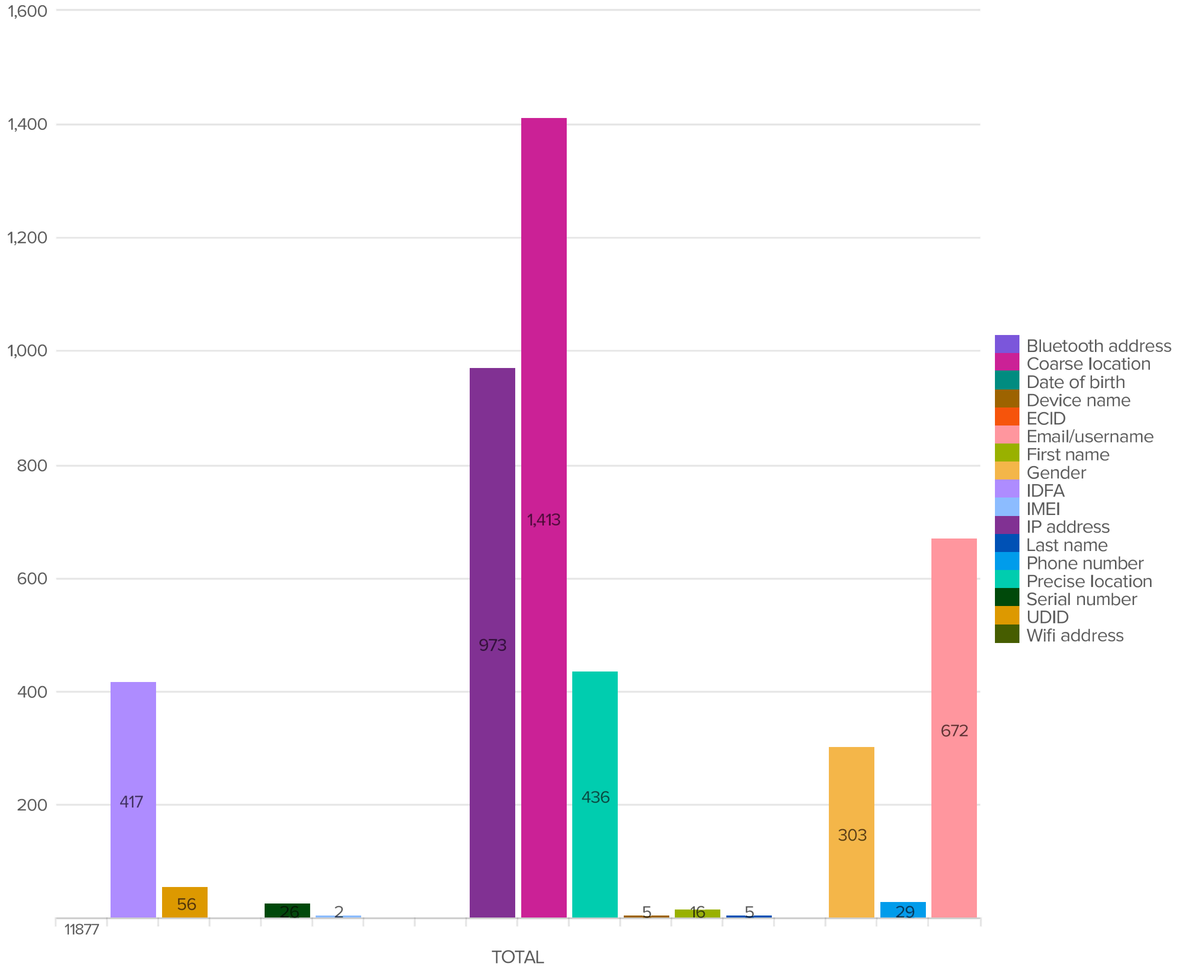
Splunk query: PII collected by the applications

```
index="search_index"
| search
| stats count as "TOTAL"

| append [search index="search_index"
| search ("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
   *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*")
| stats count as "IDFA"]

| append [search index="search_index"
| search ("*1A8C3D085CCC0DA7A3F11C6B1EC21A8016CDBD6B*" OR "*7D10BC48D9372AA313FF13D52A6D10D26C8CCF92*")
| stats count as "UDID"]

| append [search index="search_index"
| search  ("*0X0020D2FBA0BAC*" OR "*0X1C651824EBE826*")
| stats count as "ECID"]

| append[search index="search_index"
| search  ("*C7JPD1SZG5MN*" OR "*DNRQNHKUGRYC*")
| stats count as "Serial number"]

| append [search index="search_index"
| search  ("*358374069563226*" OR "*355419074758445*")
| stats count as "IMEI"]

| append [search index="search_index"
| search  ("*CC:29:F5:A7:C7:B2*" OR "*6C:8D:C1:0F:19:28*")
| stats count as "Wifi address"]

| append [search index="search_index"
| search   ("*CC:29:F5:A7:C7:DA*" OR "*6C:8D:C1:0F:19:27*")
| stats count as "Bluetooth address"]

| append [search index="search_index"
| search  ("[redacted]" OR "[redacted]' OR "[redacted]
   ')
| stats count as "IP address"]

| append [search index="search_index"
| search ("en_NO" OR"NO-34" OR "Gjøvik" OR "Norway" OR "Norge" OR "Oslo" OR "Innlandet")
| stats count as "Coarse location"]

| append [search index="search_index"
| search (("lon:*" OR "long:*" OR "longitude*") AND ("lat:*" OR "latitude*")) OR (("60.7*" AND "10.6*") OR
   "Berghusvegen*" OR "Teknologiveien*")
| stats count as "Precise location"]

| append [search index="search_index"
| search  (*Larry* OR "*Lizzy's iPhone*")
| stats count as "Device name"]

| append [search index="search_index"
| search  ("*Lizzy*")
| stats count as "First name"]

| append [search index="search_index"
| search  ("*McGuire*")
| stats count as "Last name"]
| append  [search index="search_index"
| search ("25.02.1994" OR "25-02-1994" OR "25/02/1994" OR "1994-02-25" OR "February 25, 1994" OR "25
  February 1994" OR "25th of February 1994")
| stats count as "Date of birth"]

| append [search index="search_index"
| search ("*gender*" OR "*female*" OR "*male*" OR "*sex*")
| stats count as "Gender"]

| append [search index="search_index"
| search  ("*[redacted]*" OR "*[redacted]*")
| stats count as "Phone number"]

| append [search index="search_index"
| search  ("*lizzzyM89@hotmail.com*" OR "email")
| stats count as "Email/username"]
```

All time

✓ **11,877 events** (before 5/13/24 2:19:48.000 PM) No Event Sampling

Visualization



Appendix E

Splunk query: IDFA collected by the applications

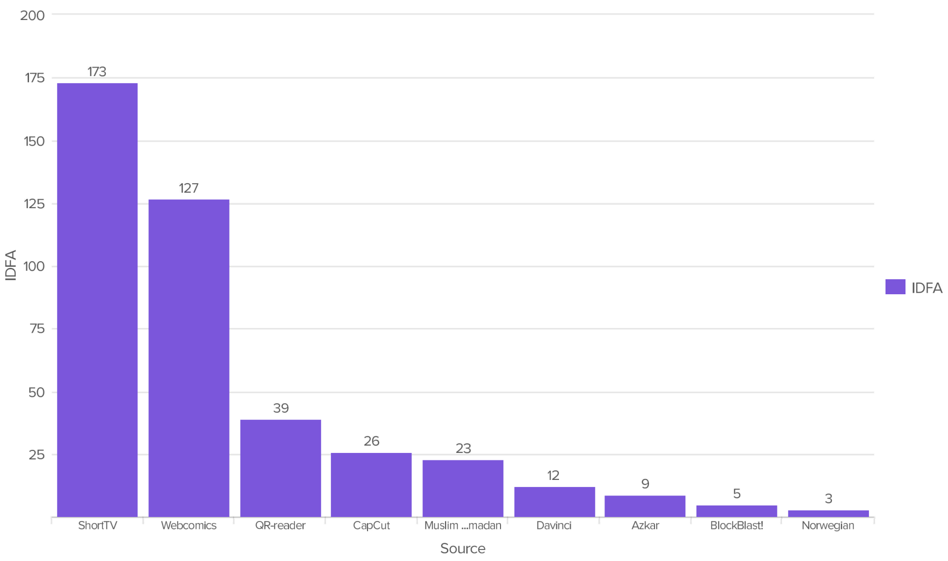
New Search

```
index="search_index" source=*
| search ("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*")
| rex field=source "(?<source_short>.+)_time"
| eval Source = case(source_short="telegram", "Telegram", source_short="norwegian", "Norwegian",
    source_short="whatsapp", "WhatsApp", source_short="sas", SAS, source_short="autopay", "Autopay",
    source_short="nordnet", "Nordnet", source_short="qr-reader", "QR-reader", source_short="zoom", "Zoom",
    source_short="davinci", "Davinci", source_short="temu", "Temu", source_short="toca-world-life", "Toca
    Life World", source_short="minfotball", "MinFotball", source_short="blockblast", "BlockBlast!",
    source_short="shortTV", "ShortTV", source_short="capcut", "CapCut", source_short="jotun", "Jotun",
    source_short="webcomics", "Webcomics", source_short="namazapp", "Namaz", source_short="azkar", "Azkar",
    source_short="muslim-Ramadam2024", "Muslim Ramadan", source_short="espresso-house", "Espresso house",
    source_short="brain-twin", "Brain Twin", source_short="pizzabakeren-norge", "Pizzabakeren",
    source_short="headache-calendar", "Headache calender", source_short="hjelp113", "Hjelp113")
| stats count as "IDFA" by Source
| sort - IDFA
```

All time

✓ **417 events** (before 5/14/24 4:08:43.000 PM) No Event Sampling

Visualization



Source	IDFA
ShortTV	173
Webcomics	127
QR-reader	39
CapCut	26
Muslim Ramadan	23
Davinci	12
Azkar	9
BlockBlast!	5
Norwegian	3

Appendix F

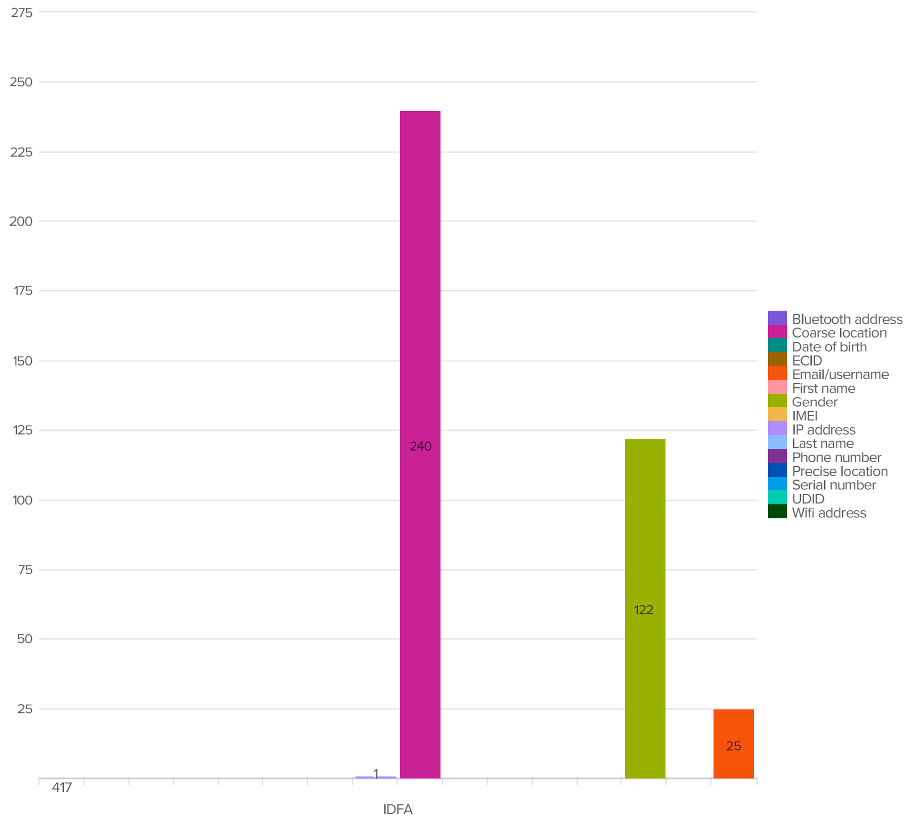
Splunk query: PII collected along with the IDFA

```
index="search_index"
| search ("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*")
| stats count as IDFA

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("
    *1A8C3D085CCC0DA7A3F11C6B1EC21A8016CDBD6B*" OR "*7D10BC48D9372AA313FF13D52A6D10D26C8CCF92*"))
| stats count as UDID]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*0X0020D2FBA0BAC*"
    OR "*0X1C651824EBE826*"))
| stats count as ECID]

| append[search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*C7JPD1SZG5MN*" OR "
    *DNRQNHKUGRYC*"))
| stats count as "Serial number"]

| append [search index="search_index"  source="webcomics_time.txt"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*358374069563226*"
    OR "*355419074758445*"))
| stats count as IMEI]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*CC:29:F5:A7:C7:B2*"
    OR "*6C:8D:C1:0F:19:28*"))
| stats count as "Wifi address"]

| append [search index="search_index"
| search (("BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]" OR "EFE00A67-7C5D-[redacted]-[redacted]-[redacted]" OR
    "BA0A45A4B5E2[redacted]" OR "EFE00A677C5D[redacted]") AND ("*CC:29:F5:A7:C7:DA*" OR
    "*6C:8D:C1:0F:19:27*"))
| stats count as "Bluetooth address"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("[redacted]" OR
    "[redacted]" OR "[redacted]"))
| stats count as "IP address"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("en_NO" OR"NO-34" OR
    "Gjøvik" OR "Norway" OR "Norge" OR "Oslo" OR "Innlandet"))
| stats count as "Coarse location"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND (("lon:*" OR "long:*"
    OR "longitude:*") AND ("lat:*" OR "latitude:*"))  (("60.7*" AND "10.6*") OR "Berghusvegen*" OR
    "Teknologiveien*"))
| stats count as "Precise location"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*Lizzy*"))
| stats count as "First name"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND "*McGuire*")
| stats count as "Last name"]


| append  [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
    *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("25.02.1994" OR "25
    -02-1994" OR "25/02/1994" OR "1994-02-25" OR "February 25, 1994" OR "25 February 1994" OR "25th of
    February 1994"))
| stats count as "Date of birth"]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
  *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*gender*" OR "
  *female*" OR "*male*" OR "*sex*"))
| stats count as Gender]

| append [search index="search_index"
| search (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
  *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*[redacted]*" OR "
  *[redacted]*"))
| stats count as "Phone number"]

| append [search index="search_index"
| search  (("*BA0A45A4-B5E2-[redacted]-[redacted]-[redacted]*" OR "*EFE00A67-7C5D-[redacted]-[redacted]-[redacted]*" OR "
  *BA0A45A4B5E2[redacted]*" OR "*EFE00A677C5D[redacted]*") AND ("*lizzzyM89@hotmail
  .com*" OR "email"))
| stats count as Email/username]
```

All time

✓ **417 events** (before 5/13/24 2:22:33.000 PM)   No Event Sampling

Visualization



IDFA	Bluetooth address	Coarse location	Date of birth	ECID	Email/username	First name	Gender	IMEI	IP address	Last name	Phone number	Precise location	Serial number	UDID	Wifi address
417				0				0					0	0	0

Appendix G

Splunk query: Total tracking domains

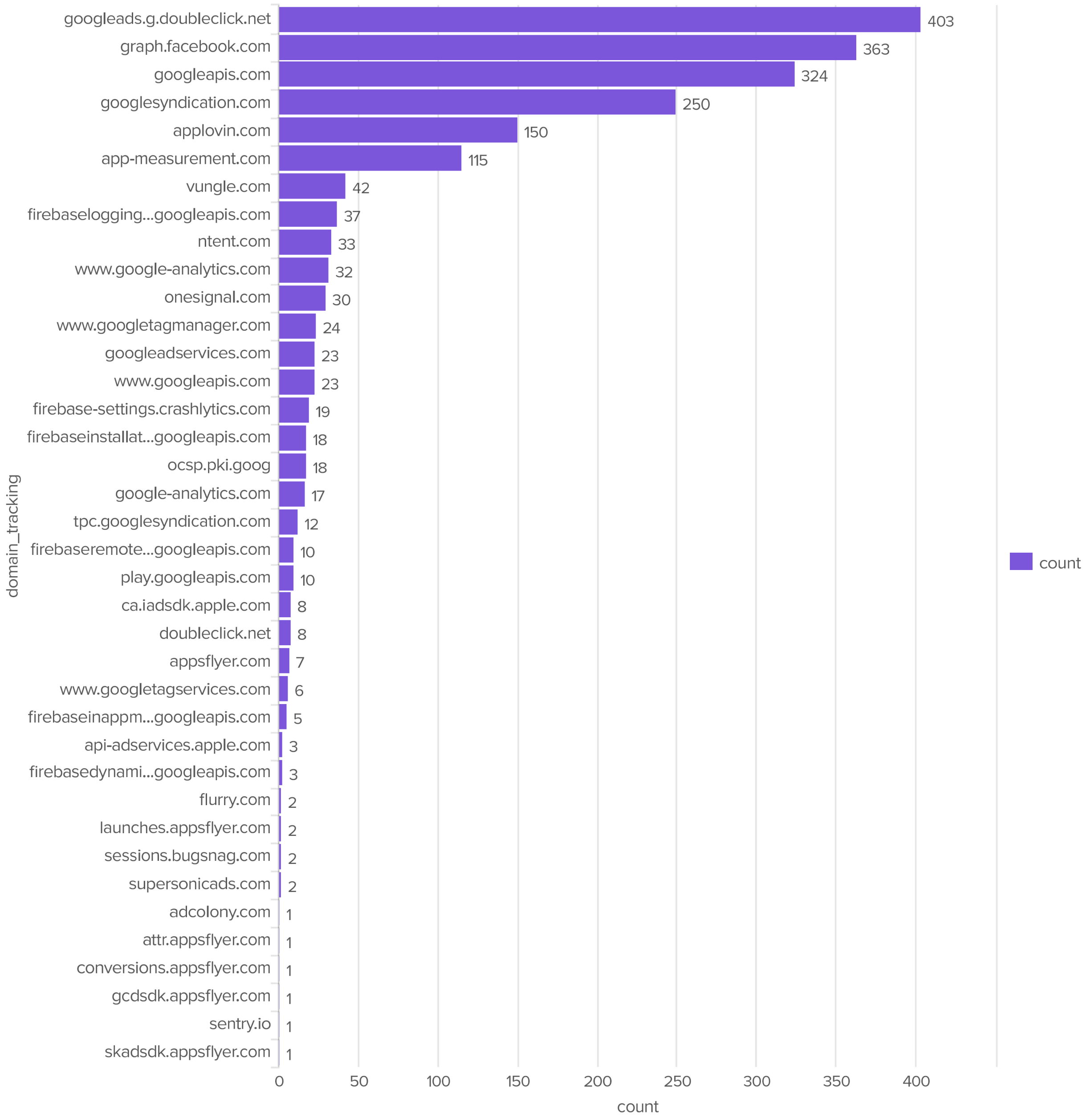
New Search

```
index="search_index"
| `tracking_domains`
| stats count by domain_tracking
| sort - count
```

All time

✓ **11,877 events** (before 5/6/24 6:48:10.000 PM) No Event Sampling

Visualization



domain_tracking	count
googleads.g.doubleclick.net	403

domain_tracking	count
graph.facebook.com	363
googleapis.com	324
googlesyndication.com	250
applovin.com	150
app-measurement.com	115
vungle.com	42
firebaselogging-pa.googleapis.com	37
ntent.com	33
www.google-analytics.com	32
onesignal.com	30
www.googletagmanager.com	24
googleadservices.com	23
www.googleapis.com	23
firebase-settings.crashlytics.com	19
firebaseinstallations.googleapis.com	18
ocsp.pki.goog	18
google-analytics.com	17
tpc.googlesyndication.com	12
firebaseremoteconfig.googleapis.com	10
play.googleapis.com	10
ca.iadsdk.apple.com	8
doubleclick.net	8
appsflyer.com	7
www.googletagservices.com	6
firebaseinappmessaging.googleapis.com	5
api-adservices.apple.com	3
firebasedynamiclinks.googleapis.com	3
flurry.com	2
launches.appsflyer.com	2
sessions.bugsnag.com	2
supersonicads.com	2
adcolony.com	1
attr.appsflyer.com	1
conversions.appsflyer.com	1
gcdsdk.appsflyer.com	1

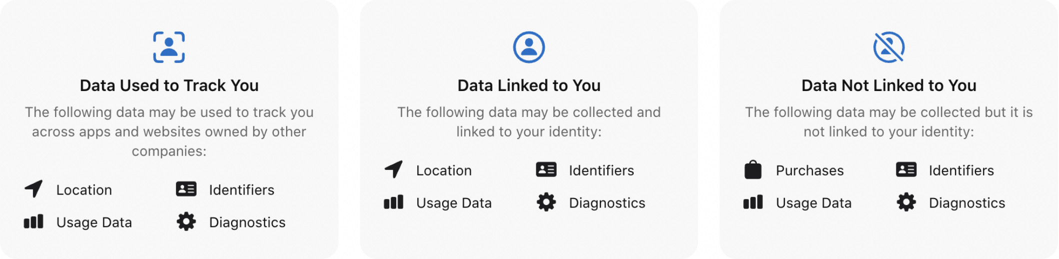
domain_tracking	count
sentry.io	1
skadsdk.appsflyer.com	1

Appendix H

Privacy Labels for Tracking Applications

Privacy Labels for Tracking Applications

Azkar • اذكار : Athan & Prayer



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Location | Identifiers | Usage Data | Diagnostics

Data Linked to You

The following data may be collected and linked to your identity:

Location | Identifiers | Usage Data | Diagnostics

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

Purchases | Identifiers | Usage Data | Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Location

Coarse Location

Identifiers

Device ID

Usage Data

Product Interaction
Advertising Data

Diagnostics

Crash Data
Performance Data

Data Linked to You

The following data, which may be collected and linked to your identity, may be used for the following purposes:

Third-Party Advertising

Location

Coarse Location

Identifiers

Device ID

Usage Data

Product Interaction
Advertising Data

Diagnostics

Performance Data

Analytics

Location

Coarse Location

Identifiers

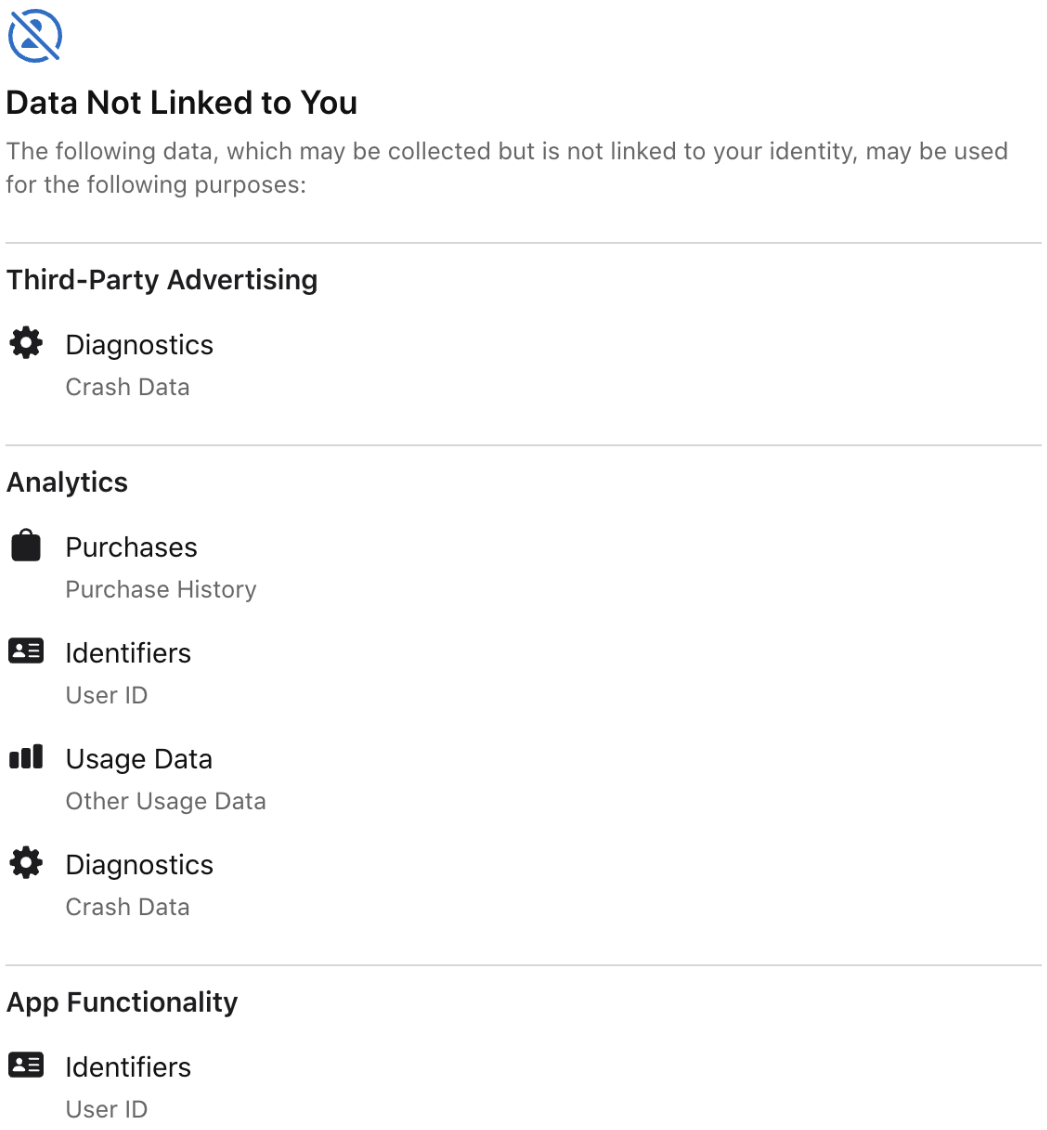
Device ID

Usage Data

Product Interaction
Advertising Data

Diagnostics

Performance Data



Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Third-Party Advertising

Diagnostics
Crash Data

Analytics

Purchases
Purchase History

Identifiers
User ID

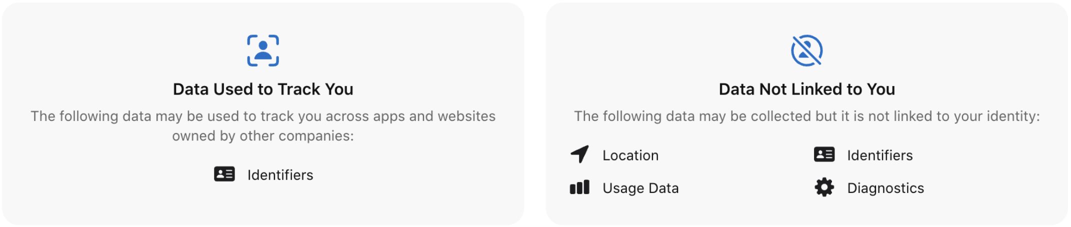
Usage Data
Other Usage Data

Diagnostics
Crash Data

App Functionality

Identifiers
User ID

QR-Reader



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

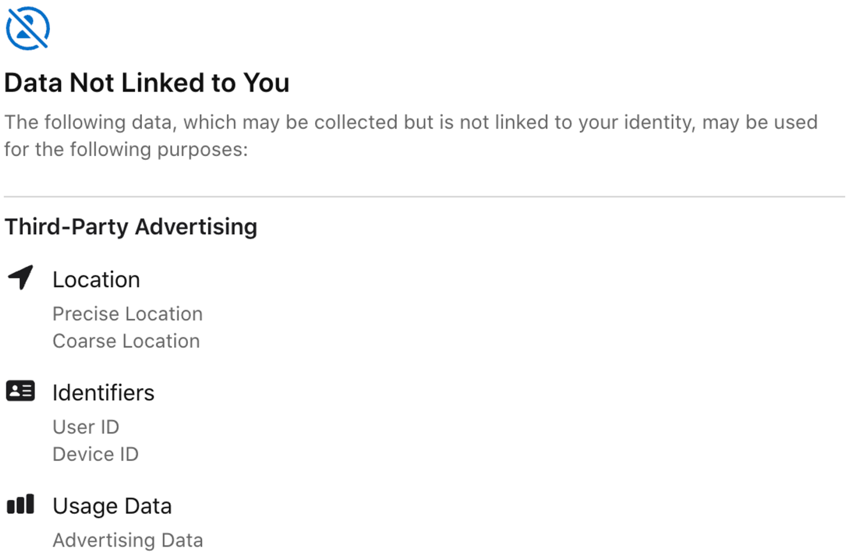
Location
Identifiers
Usage Data
Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

User ID



Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Third-Party Advertising

Location

Precise Location
Coarse Location

Identifiers

User ID
Device ID

Usage Data

Advertising Data

Product Personalization

Location

Precise Location
Coarse Location

App Functionality

Identifiers

User ID

Diagnostics

Crash Data

WebComics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

- Usage Data

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

- Purchases
- User Content
- Usage Data
- Contact Info
- Identifiers
- Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Usage Data

Advertising Data

Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Third-Party Advertising

Identifiers

User ID
Device ID

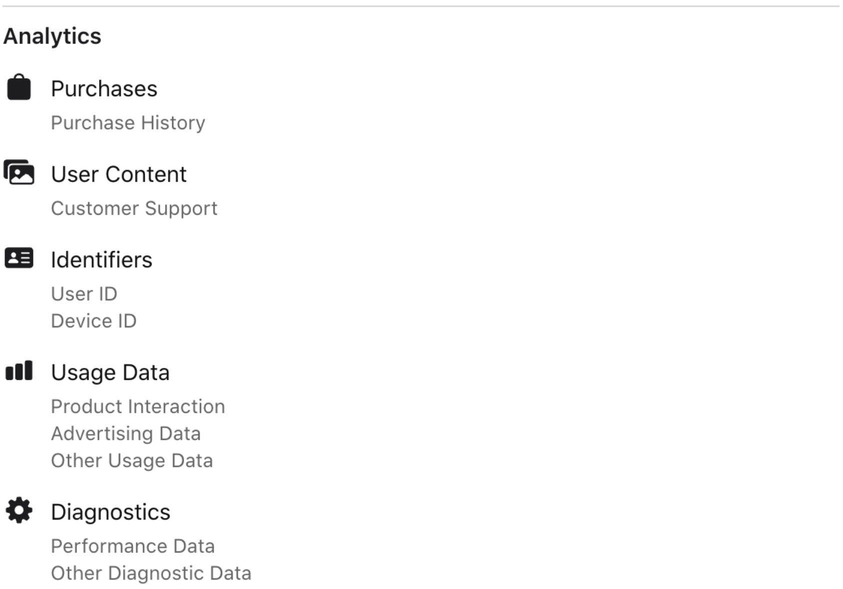
Usage Data

Advertising Data

Developer’s Advertising or Marketing

Usage Data

Advertising Data



Analytics

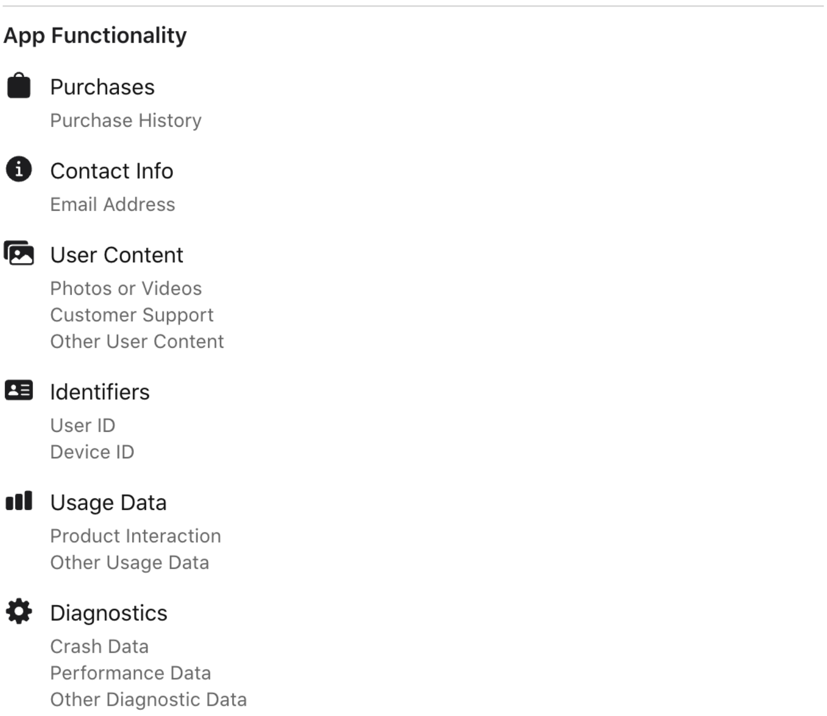
Purchases
- Purchase History

User Content
- Customer Support

Identifiers
- User ID
- Device ID

Usage Data
- Product Interaction
- Advertising Data
- Other Usage Data

Diagnostics
- Performance Data
- Other Diagnostic Data



App Functionality

Purchases
- Purchase History

Contact Info
- Email Address

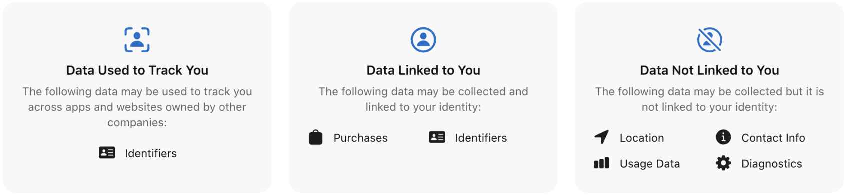
User Content
- Photos or Videos
- Customer Support
- Other User Content

Identifiers
- User ID
- Device ID

Usage Data
- Product Interaction
- Other Usage Data

Diagnostics
- Crash Data
- Performance Data
- Other Diagnostic Data

Muslim: Ramadan 2024, Iftar



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Data Linked to You

The following data may be collected and linked to your identity:

Purchases

Identifiers

Data Not Linked to You

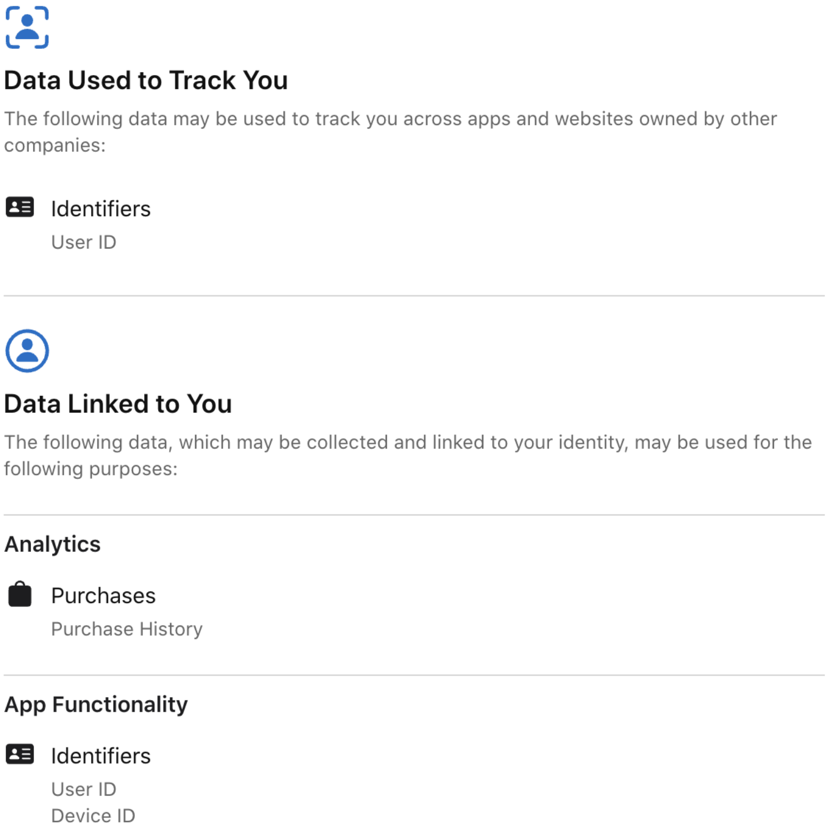
The following data may be collected but it is not linked to your identity:

Location

Contact Info

Usage Data

Diagnostics



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers
User ID

Data Linked to You

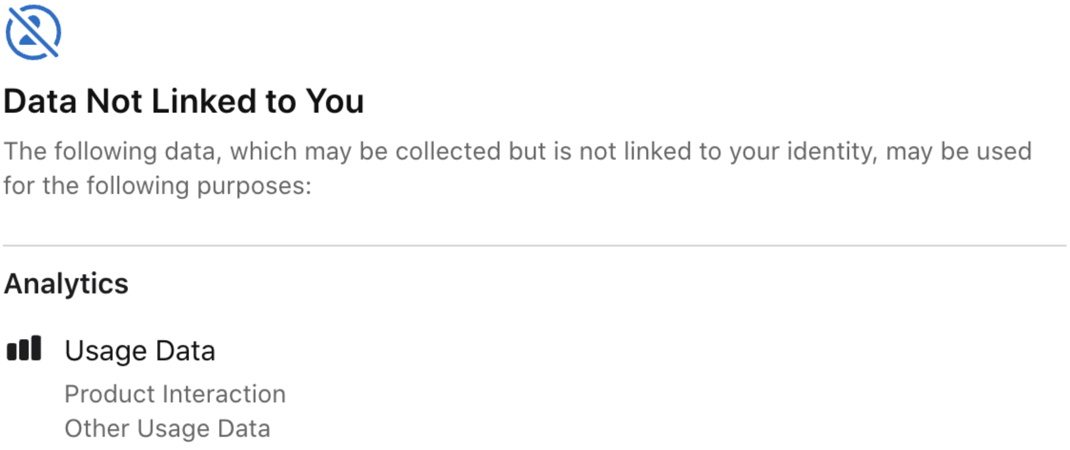
The following data, which may be collected and linked to your identity, may be used for the following purposes:

Analytics

Purchases
Purchase History

App Functionality

Identifiers
User ID
Device ID

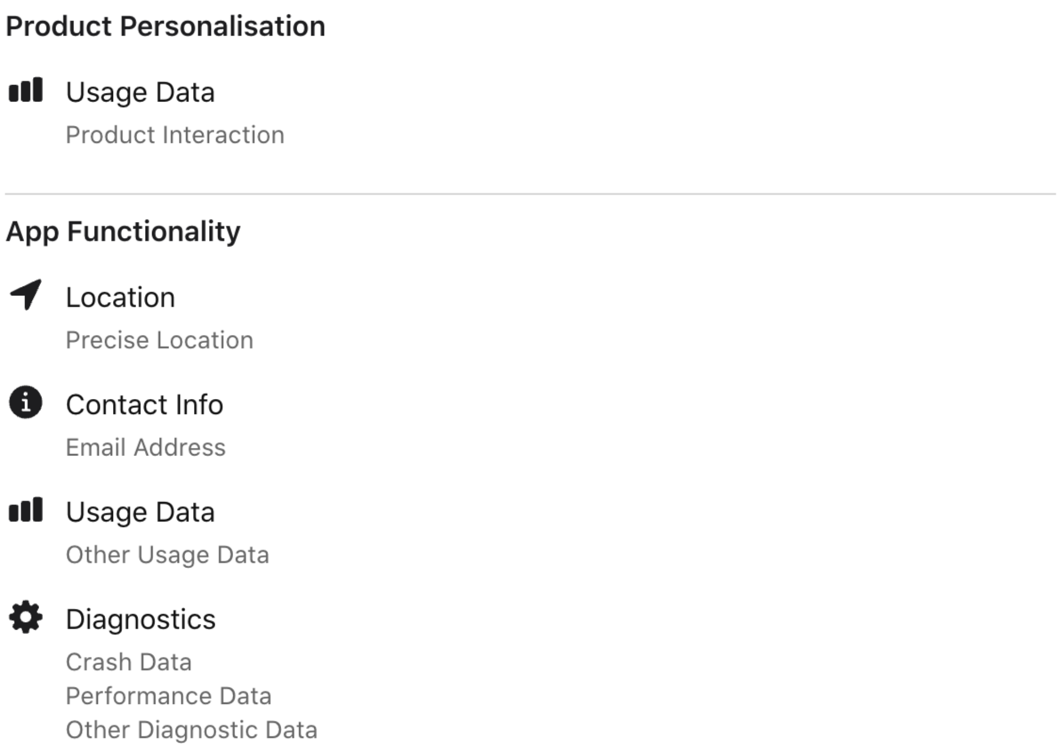


Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Analytics

Usage Data
Product Interaction
Other Usage Data



Product Personalisation

Usage Data
Product Interaction

App Functionality

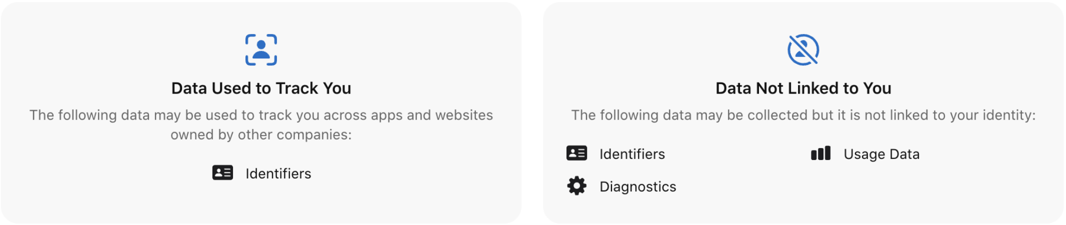
Location
Precise Location

Contact Info
Email Address

Usage Data
Other Usage Data

Diagnostics
Crash Data
Performance Data
Other Diagnostic Data

BlockBlast!



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

Identifiers
Usage Data
Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Device ID

Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Third-Party Advertising

Identifiers

Device ID

Usage Data

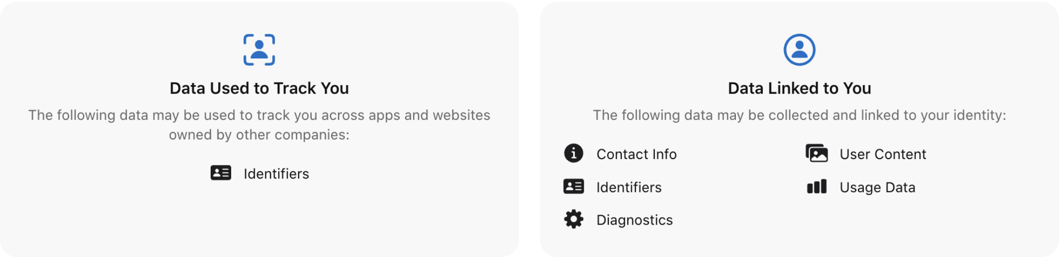
Advertising Data

Diagnostics

Crash Data

Performance Data

CapCut - Video Editor



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Data Linked to You

The following data may be collected and linked to your identity:

Contact Info

User Content

Identifiers

Usage Data

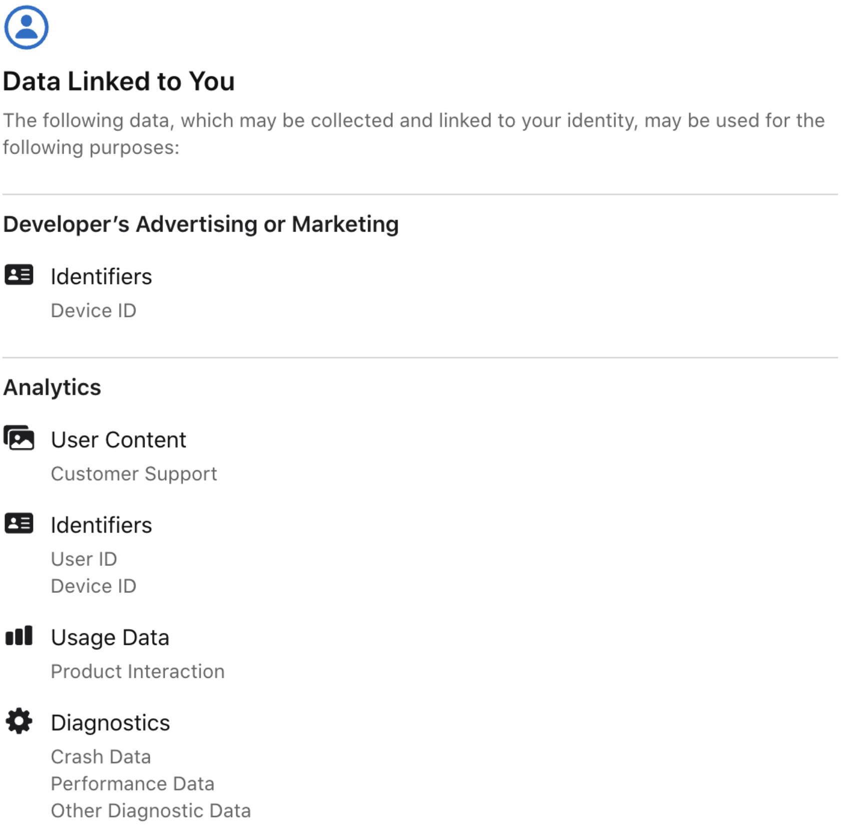
Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Device ID



Data Linked to You

The following data, which may be collected and linked to your identity, may be used for the following purposes:

Developer's Advertising or Marketing

Identifiers

Device ID

Analytics

User Content

Customer Support

Identifiers

User ID
Device ID

Usage Data

Product Interaction

Diagnostics

Crash Data
Performance Data
Other Diagnostic Data

App Functionality

Contact Info

Other User Contact Info

User Content

Photos or Videos
Audio Data
Customer Support
Other User Content

Identifiers

User ID
Device ID

Usage Data

Product Interaction

Diagnostics

Crash Data
Performance Data
Other Diagnostic Data

Other Purposes

User Content

Customer Support

ShortTV - Watch Dramas & Show

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

Data Linked to You

The following data may be collected and linked to your identity:

Identifiers

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

Location
Search History
Usage Data
Diagnostics

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Identifiers

User ID
Device ID

Data Linked to You

The following data, which may be collected and linked to your identity, may be used for the following purposes:

Third-Party Advertising

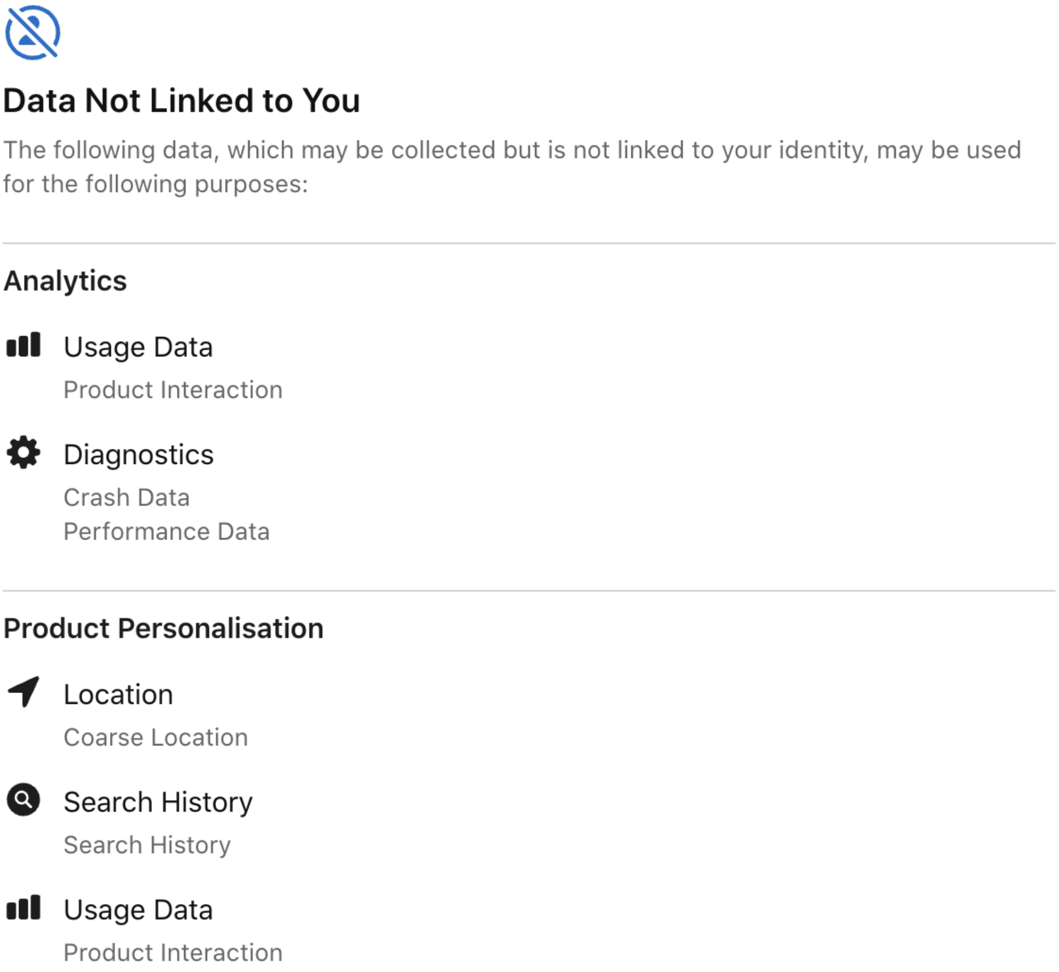
Identifiers

User ID
Device ID

Product Personalisation

Identifiers

User ID
Device ID



Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Analytics

Usage Data
Product Interaction

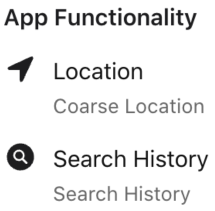
Diagnostics
Crash Data
Performance Data

Product Personalisation

Location
Coarse Location

Search History
Search History

Usage Data
Product Interaction

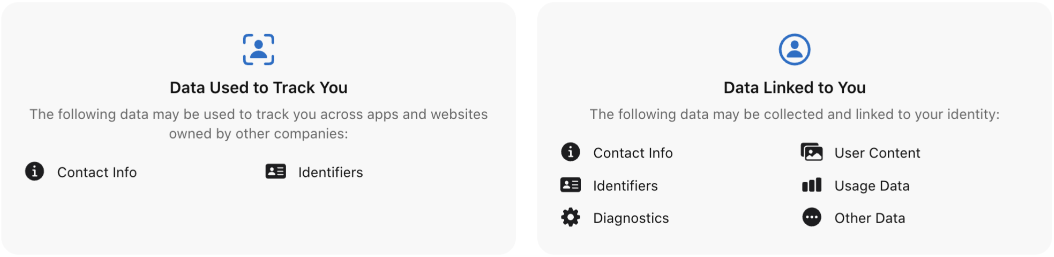


App Functionality

Location
Coarse Location

Search History
Search History

DaVinci - AI Image Generator



Data Used to Track You

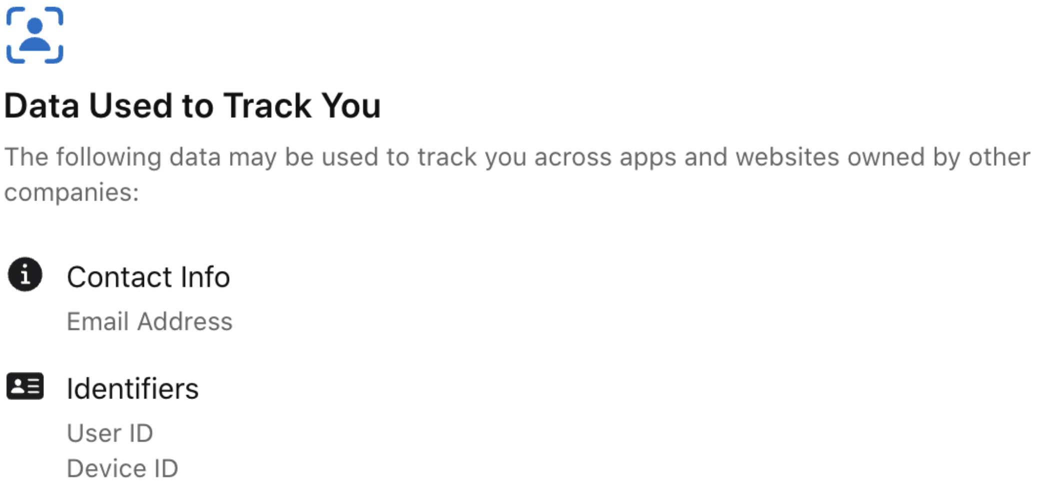
The following data may be used to track you across apps and websites owned by other companies:

- Contact Info
- Identifiers

Data Linked to You

The following data may be collected and linked to your identity:

- Contact Info
- Identifiers
- Diagnostics
- User Content
- Usage Data
- Other Data



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

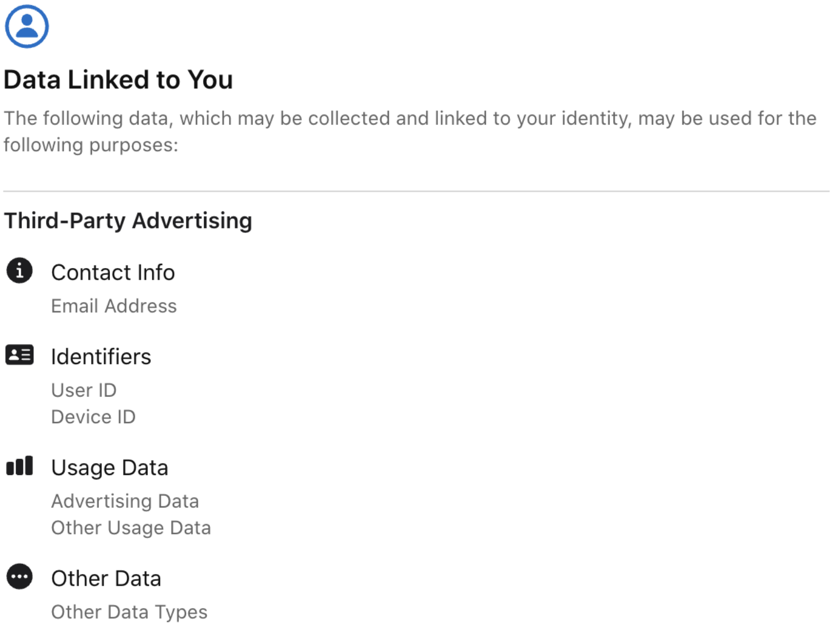
Contact Info

Email Address

Identifiers

User ID

Device ID



Data Linked to You

The following data, which may be collected and linked to your identity, may be used for the following purposes:

Third-Party Advertising

Contact Info

Email Address

Identifiers

User ID

Device ID

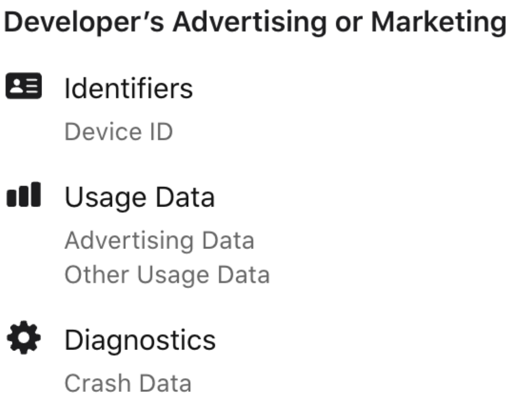
Usage Data

Advertising Data

Other Usage Data

Other Data

Other Data Types



Developer’s Advertising or Marketing

Identifiers

Device ID

Usage Data

Advertising Data

Other Usage Data

Diagnostics

Crash Data

Analytics

Contact Info
Email Address

Identifiers
User ID
Device ID

Usage Data
Product Interaction
Advertising Data
Other Usage Data

Diagnostics
Crash Data
Performance Data
Other Diagnostic Data

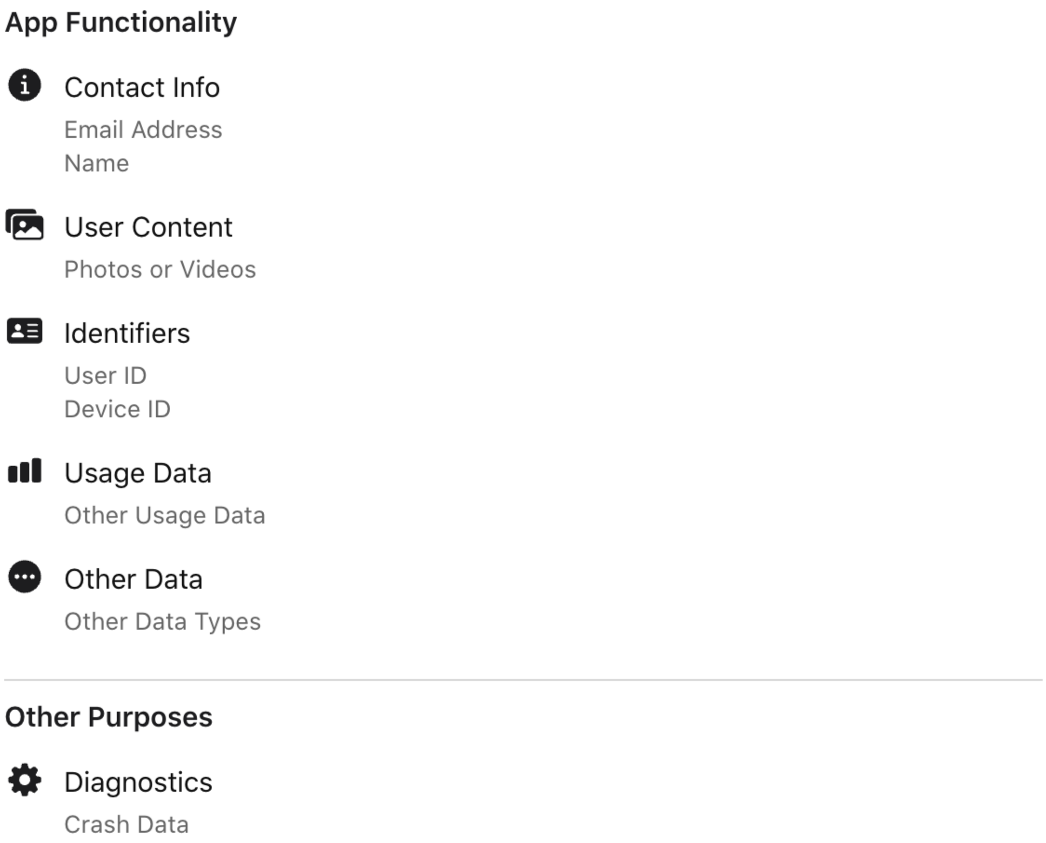
Other Data
Other Data Types

Product Personalization

Contact Info
Email Address

Identifiers
User ID

Other Data
Other Data Types



App Functionality

Contact Info

Email Address
Name

User Content

Photos or Videos

Identifiers

User ID
Device ID

Usage Data

Other Usage Data

Other Data

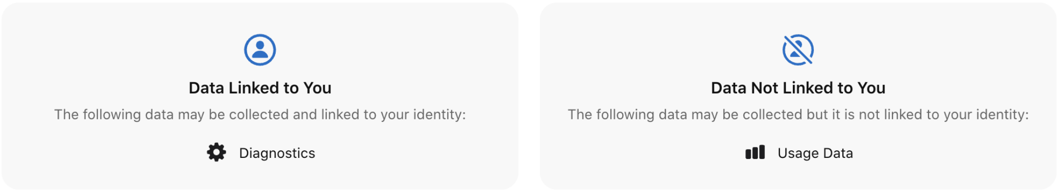
Other Data Types

Other Purposes

Diagnostics

Crash Data

Norwegian Travel Assistant



Data Linked to You

The following data may be collected and linked to your identity:

Diagnostics

Data Not Linked to You

The following data may be collected but it is not linked to your identity:

Usage Data

Data Linked to You

The following data, which may be collected and linked to your identity, may be used for the following purposes:

Analytics

Diagnostics

Crash Data
Performance Data

App Functionality

Diagnostics

Crash Data
Performance Data

Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Analytics

Usage Data

Product Interaction

App Functionality

Usage Data

Product Interaction

NTNU

Norwegian University of
Science and Technology