KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Analysis and Design of Privacy-Enhancing Information Sharing Systems

Iraklis Symeonidis

Supervisor:
Prof. dr. ir. Bart Preneel

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

May 2018

Analysis and Design of Privacy-Enhancing Information Sharing Systems

Iraklis SYMEONIDIS

Examination committee:
Prof. dr. ir. Omer Van der Biest, chair
Prof. dr. ir. Bart Preneel, supervisor
Prof. dr. ir. Vincent Rijmen
Prof. dr. ir. Frank Piessens
Prof. dr. Claudia Diaz
Prof. dr. Evangelos Markatos
(University of Crete)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

May 2018

Uitgegeven in eigen beheer, Iraklis Symeonidis, Kasteelpark Arenberg 10 Bus 2452, B-3001 Leuven, Belgium, B-3001 Leuven (Belgium)

Dedicated to my beloved parents and sisters!

Acknowledgements

Look again at that dot. That's here. That's home. That's us. On it everyone you love, everyone you know, everyone you ever heard of, every human being who ever was, lived out their lives.

CARL SAGAN, *Pale Blue Dot: A Vision of the Human Future in Space*

This dissertation marks the conclusion of a journey and the beginning of another one. It is undoubtedly the longest and most interesting one which it makes it very difficult for me to compile it in a few lines of text. Nonetheless, I will try my best to summarise and provide an accurate approximation of the long list of people that stood by me all these years!

First and foremost, I owe this memorable journey to my promotor prof. Bart Preneel. Thank you for offering me a PhD position at COSIC, believed in me and trusted me throughout these years. I would also like to thank you for your guidance, constructive critique, motivation and for the freedom and flexibility you provided me to conduct my research. All these conditions offered the necessary ground this dissertation to bring it to a closure!

My gratitude also extends to my supervisory committee prof. Frank Piessens and prof. Claudia Diaz for their insightful comments on my research. I would also like to thank the additional members of the jury prof. Vincent Rijmen and prof. Evangelos Markatos for their time and effort invested in this dissertation, and prof. Omer Van der Biest for chairing the jury. I am honoured for having you members of my committee!

I would notably want to thank all of the people that I collaborated with, for the many fruitful discussions and the valuable feedback and comments that I received throughout these years: Dalal Al-Azizy, Gergely Biczók, Mustafa A. Mustafa, Fatemeh Shirazi,

Cristina Pérez-Solà, Jessica Schroers, Bart Mennink, Pagona Tsormpatzoudi and Siemen Dhooghe. Gergely, thank you for answering my email with my questions about your work on *interdependent privacy*. It opened new dimensions for my research while the collaboration resulted in an important friendship! I would like to acknowledge FWO for funding my research visit at CrySyS lab in Budapest. I will never forget the long stays at the office, the research discussions, but also the long walks alongside the beautiful riverbank and the night lights of Budapest. Our efforts paid off and our work is now published which I am glad and proud for that! I would also like to sincerely thank my colleague and friend Mustafa for his generous guidance, assistance and support! I hope that the future will bring further research results and collaboration. My sincere thoughts also goes to Dalal for her motivation and endless encouragement on my first steps as a researcher. Thank you Dalal for being always present for me! A thank you also to Lorenzo Dalla Corte for our discussions on privacy risk scoring and Data Protection Impact Assessment (DPIA).

COSIC is a research group and a hub of many talented and motivated researchers. Being part of it, it was greatly interesting experience and a life's adventure! Thank you all for the fruitful discussions, the COSIC seminars and the group meetings. Moreover, thank you for the Friday beers, the movie nights and other spontaneous COSIC events!

First and foremost within COSIC, I would like to thank Péla, Wim, and Elsy who helped me through the non-research aspects of my PhD. Thank you, Péla for being there in the good and most importantly in the challenging times!

I am grateful to my master thesis supervisor at COSIC Andreas Pashalidis for his support and motivation! He was always there for me when I had questions answering even the not so intelligent ones. I would also like to thank Markus Huber and SBA Research Centre for providing us with the necessary material for our study, and prof. Frederik Vercauteren and Seda Gürses for their feedback. A thank you also to Michael Kyriakopoulos for introducing me to the realm of aircraft and aviation security research, a passion that I have developed at the early stages of my life!

I want to cardinally thank my " PhD siblings" Marc Juárez, Gunes Acar, Ren Zhang, Sara Cleemput and Yunwen Liu for a great time, especially for the nice Spanish–Turkish–Chinese–Greek dinners!

Furthermore, I was lucky and fortunate to share an office with several brilliant researchers visiting and working at COSIC. The office B01.05 with the narrow-view window had a magnified vision effect in my research that brought tremendous opportunities for scientific discussions and friendship: Faruk Gölogluu, Wei Li, Wenying Zhang, Dusan Bozilov and Pinar Çomak were only a few of many. I remember discussing with Faruk about history, music and food that both our countries are sharing with, including also topics related to this dissertation! Thank you, Wei and Wenying for learning me how to pronounce some basic Chinese words (I am still puzzled)!

Charlotte Bonte and Ward Beullens were very kind to help me with the Dutch translation of the abstract and Alan Szepieniec for translating in Dutch the title of this dissertation. Moreover, I am grateful to many of my colleagues and especially to Atul Luykx, Michael Herrmann, Nikos Mavrogiannopoulos, Rafael Galvez, Dragos Rotaru, Danilo Sijacic, Cyprien Delpech de Saint Guilhem, Bohan Yang, Tomer Ashur, Lennert Wouters, Christina-Angeliki Toli and Vladimir Rozic for their discussions and feedback. A special thank you to my friend and colleague Eleftheria Makri for our "Greek discussions" and for helping me with the organisation of my PhD reception. Thank you Eleftheria!

My thoughts also go to my friends that became the very definition of what Aristotle's call goodwill friendship: Stefanos T., Natasha S., Enzo De S. and Stefano L.. I met you during my academic adventure, and that have been one of the best things that recently happened in my life! Moreover, I will always be looking forward to visiting my hometown and meet my very best friends Dimitris P., Marianna P. and Vaggelis N., with whom I shared my very first thoughts about this journey! Dimitris T. and Theofanis K., friends since my childhood, were very kind and the first ones to support me, accommodating me in Athens and Brussels for several weeks! A special thank you to my friends in Leuven Georgia A., Dimitris T. and Marilena S. for the enjoyable moments and my friend Niki L. for announcing me as the "typical Greek guy" (I am still figuring out why and how)! Moreover, I would like to extend my gratitude to all of my friends for their support and amazing times we have spent together: Ioannis A., Athina A., Maria A., Dovile B., Christos C., Morfo I., Haris K., Giorgos K., Marianthi K., Stefania K., Ourania K., Giorgos M., Veggelis M., Theofanis O., Christos P., Aggelos P., Dimitris R., Eleni R., Aminata S., Konstantina T. and Ioannis V.. It has been an honour sharing my life with all of you!

I am grateful to my flatmates Andrea Di Maria and Myriam Verdonk Gallego for their support, patience and the interesting discussions we have shared. We created an important friendship that resulted in four peaceful and memorable years in Leuven!

Above all, this thesis is dedicated to my parents Zinovia and Konstantinos, and to my sisters, Eleni and Danai for their love, unconditionally support, patience and encouragement throughout these years. Finally, I would like to remember my grandmother Eleni for her endless love and care and my grandfathers Irakli and Panagioti which are continuously showing me the way of virtue, ethics and brevity in challenging times.

Iraklis Symeonidis
Leuven, May 2018

Abstract

Recent technological advancements have enabled the collection of large amounts of personal data of individuals at an ever-increasing rate. Service providers, organisations and governments can collect or otherwise acquire rich information about individuals' everyday lives and habits from big data-silos, enabling profiling and micro-targeting such as in political elections. Therefore, it is important to analyse systems that allow the collection and information sharing between users and to design secure and privacy enhancing solutions.

This thesis contains two parts. The aim of the first part is to investigate in detail the effects of the *collateral information collection* of third-party applications on Facebook. The aim of the second part is to analyse in detail the security and privacy issues of car sharing systems and to design a secure and privacy-preserving solution.

In the first part, we present a detailed multi-faceted study on the *collateral information collection* privacy issues of Facebook applications; providers of third-party applications on Facebook exploit the interdependency between users and their friends. The goal is to (i) study the existence of the problem, (ii) investigate whether Facebook users are concerned about the issue, quantify its (iii) likelihood and (iv) impact of *collateral information collection* affecting users, (v) identify whether collateral information collection is an issue for the protection of the personal data of Facebook users under the legal framework, and (vi) we propose solutions that aim to solve the problem of collateral information collection. In order to investigate the views of the users, we designed a questionnaire and collected the responses of participants. Employing real data from the Facebook third-party applications ecosystem, we compute the likelihood of collateral information collection affecting users and quantify its significance evaluating the amount of attributes collected by such applications. To investigate whether collateral information collection is an issue in terms of users' privacy we analysed the legal framework in light of the General Data Protection Regulation. To provide countermeasures, we propose a privacy dashboard extension that implements privacy scoring computations to enhance transparency towards collateral information collection.

In the second part, we investigate physical-keyless car sharing systems that allow users to share their cars with other users without the need to exchange physical keys. The goal is to (i) elicit the security and privacy requirements, and (ii) design a secure and privacy-enhancing protocol for car access provision. First, we propose a novel physical-keyless car sharing system. We then provide comprehensive security and privacy analysis and elicit the requirements for a secure and privacy-enhancing system. Finally, we design a secure and privacy-enhancing protocol for car access provision, named SePCAR. The protocol is fully decentralised and allows users to share their cars conveniently without sacrificing their security and privacy. It provides generation, update, revocation, and distribution mechanisms for access tokens to shared cars, as well as procedures to solve disputes and to deal with law enforcement requests, for instance in the case of car incidents.

Owing to the massive amounts of personal information handled by social networks such as Facebook and by car sharing systems, privacy plays a central role to the users of such systems. In this thesis, we provide a comprehensive analysis of the privacy issues for these applications and provide solutions based on transparency and privacy-enhancing technologies.

Beknopte samenvatting

Recente technologische ontwikkelingen hebben het verzamelen van grote hoeveelheden persoonlijke informatie zeer sterk doen toenemen. Dienstverleners, organisaties en overheden verzamelen massale hoeveelheden informatie en delen deze via databanken; dit stelt hen in staat om op grote schaal mensen en hun gedrag te karakteriseren en zelfs te beïnvloeden. Zo kent iedereen het voorbeeld waarin micro-targetting gebruikt werd om verkiezingen te beïnvloeden. Het is dus belangrijk om systemen, die het verzamelen en het delen van deze informatie mogelijk maken, grondig te analyseren; daarnaast is het belangrijk om nieuwe systemen te ontwikkelen die de veiligheid van de data en privacy van de gebruikers waarborgen.

Deze thesis bestaat uit twee delen. Het eerste deel bevat een diepgaande analyse van de effecten van het doorgeven van gegevens aan externe toepassingen op Facebook. Het tweede deel focust op systemen voor autodelen; het maakt een diepgaande analyse van de risico's op vlak van beveiliging en privacy en biedt ook een oplossing voor deze problemen in de vorm van een cryptografisch protocol.

In het eerste deel bestuderen we in detail de privacy implicaties die ontstaan uit het verzamelen van informatie door toepassingen gelinkt aan Facebook, de zogenaamde derden ("third-party") toepassingen op Facebook die de afhankelijkheidsrelatie exploiteren tussen gebruikers en hun vrienden. Dit onderzoek bestudeert de volgende aspecten: (i) onderzoek naar het bestaan van het probleem, (ii) onderzoeken of Facebook gebruikers bezorgd zijn over het feit dat hun persoonlijke gegevens beschikbaar kunnen zijn voor externe applicaties en hoe dit mogelijk gemaakt wordt, (iii) bepalen hoe waarschijnlijk het is dat hun data wordt doorgegeven, (iv) bepalen welke invloed het doorgeven van persoonlijke gegevens heeft op de gebruikers, (v) nagaan of het verzamelen van informatie door derden legaal gezien valt onder de bescherming van persoonlijke data van Facebook gebruikers en tot slot (vi) bieden we oplossingen voor dit probleem. Om de mening van de gebruikers te achterhalen hebben we een enquête afgenomen. Steunend op het ecosysteem van externe applicaties van Facebook hebben we berekend hoe waarschijnlijk het is dat de gegevens van gebruikers via Facebook beschikbaar worden voor externe applicaties en bepalen

we het belang van de informatie waar deze toepassingen toegang tot krijgen. Om te onderzoeken of het verzamelen van informatie door externe applicaties op Facebook een aantasting vormt van het recht op privacy van de gebruikers, bekijken we het juridisch kader gecreëerd door de Algemene Verordening Gegevensbescherming (General Data Protection Regulation). Om deze transfer van informatie in kaart te brengen, stellen we een systeem voor dat een privacy score berekend zodat er meer transparantie ontstaat rond welke informatie beschikbaar wordt.

In het tweede deel onderzoeken we systemen om auto's te delen zonder fysieke sleutels zodat gebruikers hun auto kunnen delen met anderen zonder dat ze de sleutel aan elkaar moeten doorgeven. We starten met een analyse die de veiligheidsvoorschriften en privacy noden aan het licht brengt. Daarna ontwikkelen we een veilig protocol dat een gebruiker zelfs zonder sleutel toegang kan geven tot een auto en bovendien de privacy van de gebruikers beschermt. Ten eerste stellen we een nieuw systeem voor autodelen zonder fysieke sleutels voor. Daarna voorzien we een diepgaande veiligheids- en privacy analyse en stellen we de vereisten voor een veilig, privacy-vriendelijk systeem op onder de naam SePCAR. Tot slot modelleren we het SePCAR protocol. Dit protocol is volledig gedecentraliseerd en maakt het mogelijk voor gebruikers om hun auto op een comfortabele manier te delen zonder hun privacy in het gedrang te brengen. Het systeem voorziet een mechanisme om tokens voor het autodelen te genereren, te updaten, te herroepen en te verdelen, alsook een procedure om geschillen op te lossen en om te gaan met situaties waarbij de politie betrokken is, bijvoorbeeld in het geval van een ongeval.

Omdat zowel sociale netwerken zoals Facebook en toepassingen voor autodelen over een enorme hoeveelheid persoonlijke informatie bezitten, speelt privacy een belangrijke rol voor de gebruikers van deze systemen. Deze thesis biedt een uitgebreide analyse van de mogelijke privacy problemen in deze toepassingen en stelt oplossingen voor die de gebruikers meer transparantie bieden en hun privacy beter beschermen.

Contents

List of Figures

List of Tables

List of Abbreviations

AES Advanced Encryption Standard

ACL Access-Control List

API Application Programming Interface

app third-party application

app provider third-party application provider

BGW Ben-Or, Goldwasser and Wigderson

CA Cambridge Analytica

CAS Central Authentication Service

CBC Cipher Block Chaining

CIA Central Intelligence Agency

CIC Collateral Information Collection

CM Car Manufacturer

CTR CounTeR mode

DB Data Base

DoS Denial-of-service attack

DPIA Data Protection Impact Assessment

FR Functional Requirement

FTC Federal Trade Commission

GDPR General Data Protection Regulation

GMW Goldreich, Micali and Wigderson

HMAC Hashed-based Message Authentication Code

ICTs Information and Communication Technologies

IdP Identity Provider

IOI Items of Interest

IRT Item Response Theory

KSApp Keyless car Sharing Application

KSMS Keyless car Sharing Management System

KSS physical-Keyless car Sharing System

LTE Long-Term Evolution

MAC Message Authentication Code

MAU Monthly Active Users

MD Mobile Device

MPC Secure Multiparty Computation

NFC Near-Field Communication

NSA National Security Agency

OBU On-Board Unit

OS Operating System

OSN Online Social Network

PETs Privacy Enhancing Technologies

PHP PHP: Hypertext Preprocessor

PID Personal Information Dashboard

PKI Public-Key Infrastructure

PL Public Ledger

PR Privacy Requirement

PRF Pseudo-Random Function

PS Privacy Score

RBAC Role-Based Access Control

RSA Rivest–Shamir–Adleman

RSA-KEM RSA-Key Encapsulation Mechanism

SAML Security Assertion Markup Language

SARL SAML Revocation List

S-SDLC Secure-System Development Life Cycle

SePCAR Secure and Privacy-enhancing protocol for Car Access Provision

SP Service Provider

SQL Structured Query Language

SR Security Requirement

SSO Single Sign-On

TET Transparency Enhancing Technology

TETs Transparency Enhancing Technologies

TPs Third Parties

TPM Trusted Platform Module

TTP Trusted Third Party

VPKI Vehicular Public-Key Infrastructure

VSS Verifiable Secret Sharing

Chapter 1

Introduction

Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say.

EDWARD SNOWDEN, *Computer professional, former Central Intelligence Agency (CIA) employee and whistle blower*

1.1 Motivation

The advancement of Information and Communication Technologies (ICTs) has eased the exchange of information between individuals, while enabling the collection of large amounts of their personal data at an ever-increasing rate [302]. Individuals are extensively using ICTs such as the Internet, mobile networks, wireless networks and online platforms for several purposes: from searching the web, throughout engaging in online communities and performing online transactions. Service providers, organisations and governments can collect individuals' browsing information, monitor their daily activities in online communities and fuse massive amounts of information for transactions and payments. The former Central Intelligence Agency (CIA) employee Edward Snowden showed the extensive capabilities of the U.S. and U.K. intelligence services on tracking individuals' everyday activities [301]. They can monitor web

searches, collect posts in online communities and wiretap file transfers of cloud infrastructure [275, 299, 304]. Moreover, data analytic companies such as Cambridge Analytica (CA) [49] and Palantir [238] can acquire and collect rich information about individuals' everyday lives and habits from big data-silos, enabling profiling and micro-targeting of individuals such as in political elections [221, 302].

In line with the vast impact of the ICTs and the implications of information sharing, online communities such as online social networks and systems for sharing physical assets have populated the world while threatening the privacy of individuals [302]. For example, individuals can express their opinions using Twitter [3], update their professional profiles using LinkedIn [2], share their photos with their friends using Facebook [117] and book their accommodation for their next trip using Airbnb [9]. While individuals perform such online activities, the collection and storage of their personal information is enabled. Thus, privacy threats emerge with service providers that can compile their activities into detailed dossiers often without their permission or even knowledge. Personal information can encompass not only data that identifies individuals such as their identity, birthday and home address but it can also reveal their interests, behaviours [101], political and religious beliefs [255] and even their health status [137].

All is not lost. While ICTs bring danger of diminishing the privacy of individuals, the research community is actively identifying, extensively analysing and thoroughly mitigating privacy problems from different perspectives. For identifying and analysing privacy issues there are multiple tools including network simulations (which measure the impact of information dissemination), questionnaires (which study the opinion of individuals) and mathematical models (which quantify the amount of disseminated information). Moreover, there is an extensive literature on privacy and transparency enhancing technologies for designing and implementing a wide range of solutions such as technologies for preserving anonymity, unlinkability, and confidentiality properties. We refer the interested reader to the taxonomy of privacy by Solove [274], the terminology and definitions of privacy properties by Pfitzmann and Hansen [245] and to the methodology for applying these properties by Deng et al. [84].

In this thesis, we focus on ICTs and systems that allow the collection and information sharing between users and more specifically on Online Social Networks (OSNs) and on physical assets sharing systems. In short, OSNs offer a global platform enabling a wide range of human activities and social interactions resulting in an extensive amount of users. In the U.S., 65 percent of the Internet-using adults in 2010 are members of OSNs such as Twitter, LinkedIn and Facebook [14] while more than 2 Billion users per month are actively using Facebook (in 2018) [277]. We focus on Facebook, a key OSN, and the privacy issue of *interdependent privacy*; providers of third-party applications on Facebook exploit the interdependency between users and their friends. Moreover, we examine systems that enable individuals to share physical assets – Uber, a flagship company for peer-to-peer car sharing is used by more than 200 Million users. Owing

to the massive amounts of personal information handled by OSNs and car sharing systems, privacy plays a central role to the users of such systems.

1.2 Problem statement

This thesis deals with two main privacy problems related to information sharing. First, we detail the privacy implications of *interdependent privacy*, an issue enabled by third-party applications affecting Facebook users. Second, we highlight the privacy threats that online platforms and systems that utilise ICTs enable for sharing physical assets, with an example car sharing systems.

As a concept, privacy is widely used, and it is dependent on social, academic and legal contexts, as pointed out by Gürses [147]. Thus, we initially provide the definitions of *privacy* and *data protection* that we will use throughout this thesis.

1.2.1 Definition of privacy and data protection

Historically, privacy was defined as the "right to be left alone" by Warren and Brandeis [327]. In an interconnected world of online platforms and systems that utilise ICTs, privacy is often referred as the ability of individuals to control the dissemination of their personal information. In other words, privacy is defined by Westin and Ruebhausen [332], as:

Definition. *The right of the individual to decide what information about himself should be communicated to others and under what circumstances.*

Note that if a system respects the privacy of their users, that does not necessarily mean that their personal information are never disclosed to anyone. It rather means that a system allows their users to control what information is revealed, to whom and under which conditions. Personal information has a very broad meaning and includes photos, likes, location, IP addresses, behavioural information, and of course identity.

There are scenarios where the privacy of individuals is bound to be affected by the decisions of others [35]. For example, the genome of an individual contains sensitive information about a person but also correlated information about the person's family members such as ethnicity, kinship, and tendency to diseases [167]. In OSNs such as on Facebook, individuals are connected with others with friendship relationships, enabling privacy interdependence on data sharing.

Definition. *We define interdependent privacy as the scenario when the privacy of individuals is affected by decisions taken and actions initiated by anyone but the individuals themselves.*

Note that, in scenarios where privacy interdependence is enabled, individuals are not able to control the dissemination of their personal information.

Definition. *We define collateral information collection as the acquisition of users' personal data through any means or tool initiated by anyone but the users themselves.*

The right to privacy is a fundamental right and defined in Article 7 – "*Respect for private and family life*" of the Charter of Fundamental Rights of the European Union (the 'Charter') [114] as:

Definition. *Everyone has the right to respect for his or her private and family life, home and communications.*

The right for the protection of personal data of individuals is also a fundamental right, defined in Article 8(1) – "*Protection of personal data*" of the Charter of Fundamental Rights of the European Union (the 'Charter') [114]:

Definition. *Everyone has the right to the protection of personal data concerning him or her.*

Safeguarding the privacy of individuals is related to the protection of their electronic communications [112] and of their personal data while stored and processed by online providers [113]. Considering the data protection of individuals, it is dependent on the legal framework defined by the General Data Protection Regulation (GDPR) [113]: [1]

Definition. *The principles of and rules on the protection of natural persons about the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular, their right to the protection of personal data.*

In a nutshell, the GDPR regulates the protection of personal data of individuals while being processed (e.g., collected, recorded and stored) by online providers (i.e., data controllers and processors) against automated-decision making, profiling and discriminatory practices. Compliance should be transparent and monitored by organisations and regulatory authorities; the lack of it will invoke heavy fines (Article 83 – "*General conditions for imposing administrative fines*" of GDPR [113]). Hence, OSNs and car sharing providers should manage the personal data of individuals rightfully and be compliant with the *data protection principles* of the GDPR.

To safeguard the aforementioned rights, online platforms and systems that utilise ICTs should by design and default protect the personal data of individuals, as it is stated in Article 25 (1) and Article 25 (2) of the GDPR respectively.

[1] The *General Data Protection Regulation* (GDPR) becomes applicable as of May 25, 2018.

1.2.2 Facebook third-party application providers and privacy interdependence

From a private network of students at Harvard University, Facebook became the most dominant OSN with more than 2 Billion users in 2018 [277]. By design and popularity, Facebook has morphed into an immense information repository, storing individuals' personal information; logging interactions between users and their friends, groups, events and pages. Users are willing to provide large amounts of their personal information on Facebook taking no action to limit their information visibility, as pointed out by Gross and Acquisti [145].

Facebook also offers third-party applications (apps) developed by other vendors, i.e., third-party application providers (app providers). While installing an app on Facebook, it enables access to the profile information of a user. Accepting the permissions, the app collects personal and often sensitive information of the user, such as profile images, dating preferences and political interests [145, 42, 99] (see Chapter 2). Interestingly, Golbeck and Mauriello [139] identified that a large percentage of users on Facebook were under-informed and did not believe that apps could access their data, having an incorrect perception on the Facebook data app access capabilities [139].

Beside such information collection, the profile information of a user can also be acquired when a friend of a user installs an app [291] entailing *privacy interdependence* issues on Facebook [35, 290]. A user who shares personal information with his/her friends on Facebook has no idea whether a friend has installed an app that also accesses the shared content. In a nutshell, when a friend of a user install an app, the app can request and grant access to the profile information of a user such as the birthday, current location, and history [291]. Such access takes place outside the Facebook ecosystem with the user not being aware whether a friend has installed an app collecting his/her information; this *collateral information collection* is enabled only with the friends' consent and not with consent of the user. On Facebook, apps privacy settings allow by default the profile information of a user to be accessed by their friends' apps, unless they manually uncheck the relevant boxes "Apps other use". Note that, in some cases, one or more app providers may cluster several apps and thus gain access to a collection of personal information.

Such *collateral information collection* goes far beyond the legitimate expectations of users and their friends, as there is lack of transparency and consent. In the U.S., the Federal Trade Commission (FTC) stressed that such apps cannot imply consent, which should be affirmatively expressed by users [125, 126]. In other words, consent should be provided only by the user, whose data will be collected by the apps, and not by his or her friend.

1.2.3 Security and privacy threats of online platforms: the case of car sharing

Car sharing services have shown considerable growth [34, 276, 276]. As a byproduct, an immense amount of data, generated by its users, is transferred to these systems and stored in their online platforms [272]. Although car sharing systems provide users with an essential functionality, a platform that helps them to share cars conveniently and on demand, they also introduce several issues concerning security and privacy.

First, in scenarios where users share their cars, cyber-physical security is of utmost importance. Car sharing systems utilise online platforms to facilitate services, and the platform itself should have secure mechanisms. Most platforms run on a public cloud infrastructure, inheriting its security issues. Also, if the platform handles payments, it usually utilises a third-party financial provider to process transactions hence it is vulnerable to attacks on the financial provider. For example, in a breach of security, the information of 57 million customers and drivers was compromised at Uber [222]. In another incident, several users found their accounts illegally accessed, used and charged for rides performed by fraudulent activities [303]. In short, an adversary can try to eavesdrop and collect information exchanged within a car sharing system, tamper with the car sharing details, extract the key of a car stored in un-trusted devices, generate a rogue access token to access a car or deny having accessed it maliciously.

Another major concern is the amount of personal and potentially sensitive data collected by these systems [272]. For instance, Uber allegedly used a program called “Hell” to spy on their rival company drivers [137, 300]. In another case, a program called “God view” or “Heaven” was used to track *always* the location of their users [78, 97]. In short, an adversary can try to correlate and link two car sharing requests of the same user or the car, identify car usage patterns and deduce the sharing preferences of users. These preferences can be established by collecting information about sharing patterns such as rental time, duration, pickup location, when, where and with whom someone is sharing a car [101].

Owing to the massive amounts of personal data handled, compliance with *data protection law* plays a central role for car sharing systems. The GDPR [113] establishes stringent requirements on online providers, and more specifically it imposes the obligation to implement organisational and technical measures to process personal data legally and further safeguard data subjects’ rights, i.e., users.

1.3 Research challenges

1.3.1 Collateral information collection of Facebook third-party application providers

Collateral information collection is the manifestation of *interdependent privacy*; the scenario where the privacy of an individual user is affected by the decisions of other users [35]. Prior research in the domain of Facebook apps is mainly focused on the topics of information collection [15, 165], tracking [5, 4, 165], discrepancy between the amount of permissions and the nature of apps [133, 59, 323] and game theory [35, 249, 250]. However, there is currently no prior work that extensively analyses how the friends of a user in the Facebook ecosystem directly or indirectly affect their privacy through third-party app and app providers.

To investigate in detail the effects of the *interdependent privacy* of the third-party apps and app providers on Facebook and the *collateral information collection* that entails to users, the following challenges need to be addressed:

- **Users' opinion on and concern about the collateral information collection of third-party apps on Facebook.** Interdependent privacy has been identified to be a privacy issue for Facebook users by both the research community [326, 35, 250] and organisations, i.e., the FTC [125, 126]. Complementary to that, in our research, we want to investigate the opinion of Facebook users, and whether they are concerned about the collateral information collection.
- **The likelihood of the collateral information collection to occur.** Assuming that users are concerned, in our research, we want to identify whether such an issue is highly likely to happen and from what it is dependent on. For instance, what about popular apps such as TripAdvisor [310]?
- **The significance of the collateral information collection.** Assuming that popular apps have a high probability to enable the collateral information collection, our research intends to identify the amount of personal information collected by third-party apps and app providers (i.e., data fusion of multiple apps by an app provider).
- **The issues of collateral information collection of Facebook third-party app providers under the prism of the data protection regulation.** Our research intends to investigate the issues of collateral information collection under the prism of the GDPR [113] for the third-party app providers of Facebook.
- **Solutions to collateral information collection on Facebook.** Implementing privacy enhancing solutions on Facebook such as cryptographic countermeasures

can be interesting but challenging to maintain [3, 2, 1] as they are frequently detected and blocked [196, 26, 68]. It is challenging to propose solutions aiming to help users making informed decisions and enhance transparency [130].

1.3.2 Secure and privacy-enhancing car sharing systems

Prior research in the domain of car sharing is mainly focused on topics of relocation strategies [330, 341], economic models [271, 270, 270, 128, 129], secure authentication [88], PKI infrastructure [254, 182], tamper-proof devices [312, 259], location privacy [311, 21, 21] and data mining [205, 239]. However, none of the prior work provides an analysis of the security and privacy threats of car sharing systems, elicits the security and privacy requirements and designs fully-fledged solutions for secure and privacy-enhancing car sharing systems.

In order to investigate in detail the security and privacy issues for car sharing systems and design solutions, the following challenges need to be addressed:

- **Security and privacy threats of car sharing systems and the elicitation of requirements.** No prior work describes systems for sharing cars using ICTs such as mobile phones, wireless communication and online platforms. Our research intends to design a high-level system model, specify the entities and operations of such a system, the security and privacy threats and the requirements for a secure and privacy preserving system design.
- **Design and develop a full-fledged secure and privacy-enhancing car access provision protocol.** Unlike other researchers assuming centralised and fully trusted car sharing systems [88], our research intends at designing solutions for secure and privacy-enhancing car access provision. Note that forensic evidence also needs to be provided when an unlawful action occurs as users have physical access to the shared car.

1.4 Aim of this thesis

This thesis is organised into two parts and it has the following objectives.

Collateral information collection. The aim of the first part (see Part II) is to investigate in detail the effects of the *interdependent privacy* of the third-party apps and app providers and the *collateral information collection* on Facebook users. The following objectives support the aim of this part:

- To collect the opinion of individuals; whether they are concerned about the collateral information collection of Facebook third-party apps.
- To formulate, compute and evaluate the probability that an installed app enables the collateral information collection.
- To evaluate the significance of the collateral information collection by apps and app providers.
- To analyse and clarify who the data controllers and data processors are, whether collateral information collection is an adequate practice from a data protection point of view and to identify who is accountable.
- To propose transparency enhancing solutions for raising user awareness and helping them to make informed decisions. Moreover, to analyse other alternatives to provide transparency.

Physical keyless car sharing systems. The aim of the second part (see Part III) is to analyse in detail the security and privacy issues of car sharing systems, to elicit the requirements and design a secure and privacy-preserving solution. The following objectives support the aim of this part:

- To design a high-level system model, specify its main entities and the necessary functional requirements.
- To perform a threat analysis identifying the security and privacy threats for car sharing systems in a systematic way.
- To elicit the requirements for a secure and privacy-preserving solution design of car sharing systems.
- To design and implement a secure and privacy-enhancing car access provision protocol considering forensic evidence provision when an unlawful action occurs.
- To prove that the protocol fulfils the desired security and privacy requirements.
- To evaluate the theoretical complexity and practical efficiency of the designed protocol.

1.5 Research methodology

In this section, we describe our research methodology. For the analysis and the solution design, we use the system engineering approach.

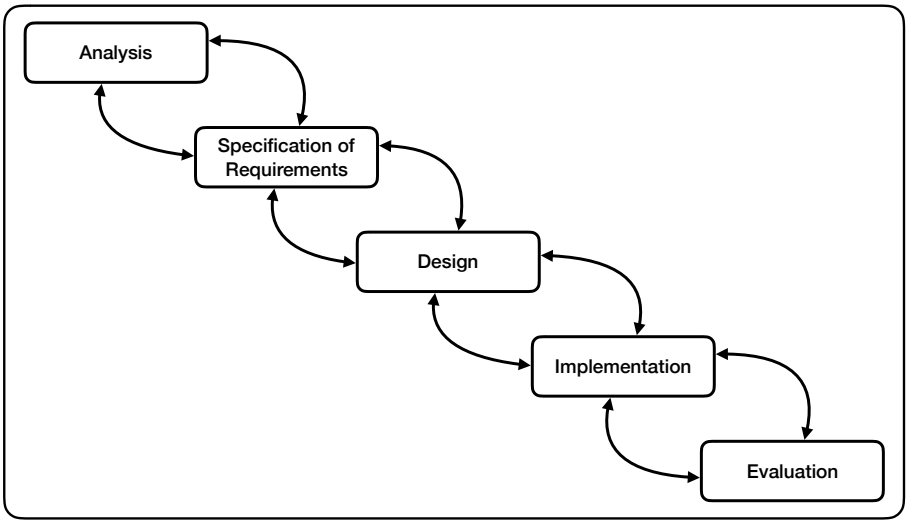


Figure 1.1: Secure-System Design Life Cycle (S-SDLC).

System engineering aims at finding solutions to a problem in a methodological and structural way [148], using the Secure-System Development Life Cycle (S-SDLC) methodology [214]. It focuses on the design and management of complex systems as a whole and from its components, meaning that distinct elements together can produce results not obtainable by the elements alone. Using a holistic life-cycle approach, system engineering aims to capture and describe the problems (i.e., *analysis*), transform them into a description of requirements (i.e., *elicitation of requirements*), provide a solution design (i.e., *design*), develop a proof of the design or a full fledged implementation (i.e., *implementation*) and evaluate as well as identify probable failures that can occur (i.e., *verification*). S-SDLC inherits its logic and structure from the waterfall model [261, 79], enabling a top-down synthesis and iterative processes of life-cycle stages (see Fig. 1.1) [37, 20, 261, 190].

I. Analysis

- We analysed the privacy issues of the collateral information collection of Facebook apps examining the app permissions system, the users' privacy settings and a real-world snapshot of the apps on Facebook [15]. It resulted in four parts for our collateral information collection study: (i) whether individuals are concerned, (ii) what is the probability of such information collection to happen, (iii) whether it is significant, and finally (iv) which are the legal implications and whether it represents a risk for the privacy of users.

- We analysed the security and privacy issues for car sharing systems. This analysis resulted in four parts for our study: to identify (i) the system model such as the functionalities, entities and operations involved, (ii) the security and finally (iii) the privacy issues that such a system should be able to deal with.

II. Specification of requirements

- We specified the functional requirements for a privacy dashboard extension aiming at enhancing transparency towards collateral information collection.
- We specified three main directions that should be considered for car sharing systems: (i) the functional issues such as forensic evidence provision, (ii) the security classification of threats and the corresponding requirements, and (iii) the privacy classification of threats and the corresponding requirements

III. Design

- We designed a questionnaire to investigate the individuals' opinion on, and their concerns about: i) the *collateral information collection*, ii) the attributes of their profile information they are more concerned about, and iii) whether they want to be notified and which actions they prefer to take in such cases.
- We developed a mathematical formula to estimate the likelihood of (i) a uniform app adoption model, which takes into account the popularity of an app and the Facebook users, and (ii) a non-uniform app adoption model, that takes into account different network topologies and app adoption models tailored to the Facebook ecosystem.
- We developed a mathematical formula to measure the number of the user's attributes (i.e., significance) when collected (i) by apps and (ii) by app providers as an effect of data fusion. Moreover, we developed different cases of data acquisition: (i) direct collection from the users themselves, (ii) indirect collection through the friends of a user, and (iii) exclusive indirect collection through the only friends of a user.
- We proposed a privacy dashboard extension that implements privacy scoring computations for enhancing transparency, as a solution-design for the collateral information collection of Facebook third-party apps and app providers. Moreover, we developed a privacy scoring formula for evaluating the collateral information collection over the apps and app providers.

- We designed a novel high-level system model for a car sharing system to demonstrate the necessary functionalities, entities involved and operations performed.
- We designed a decentralised secure and privacy-enhancing protocol for car access provision, named SePCAR, based on the security and privacy requirements identified in the requirements elicitation phase.

V. Implementation

- We implemented and distributed a questionnaire using online web forms.
- We implemented our computations to measure the significance using the PHP: Hypertext Preprocessor (PHP), JavaScript and Structured Query Language (SQL) programming languages.
- We implemented our simulations for using various network topologies [11, 329, 102] and app adoption models [283] tailored to the Facebook ecosystem using Python libraries.
- We reviewed the literature to design our system model for car sharing and to identify the security and privacy threats.
- We implemented a protocol for car access provision namely SePCAR as a proof-of-concept in C and evaluated its theoretical complexity and practical efficiency.

IV. Evaluation

- We investigated the views of individuals by collecting responses from 114 participants.
- We performed simulations to evaluate the likelihood that an installed app enables collateral information collection for networks with a size of one million nodes.
- We measured the significance of collateral information collection based on a real-world snapshot of 16,808 apps on Facebook [15].
- We clarified through the prism of GDPR the legal implications and the risks to users' privacy for the collateral information collection.
- We compiled the list of security and privacy classes of threats and their corresponding requirements using the STRIDE framework [163] and the LINDDUN framework [84] respectively.
- We proved that SePCAR satisfies the security and privacy requirements, provided that its underlying cryptographic primitives are sufficiently secure.

1.6 Summary of contributions and thesis outline

In this section, we outline the structure of this thesis. For each chapter, we also provide the summary of contributions, list the co-authors and mention the related publications. Our goal is to facilitate the evaluation of all contributions in this thesis.

PART I: PRELIMINARIES

Chapter 2 – Background and related work. In this chapter, we provide the necessary background, and we give an overview of the related work.

Chapter 3 – Tools for analysis, specification of requirements and design of privacy-enhancing systems. In this chapter, we provide the necessary tools we use throughout this thesis for analysis, elicitation of requirements and design of privacy-enhancing systems.

PART II: COLLATERAL INFORMATION COLLECTION OF THIRD-PARTY APPLICATIONS

Chapter 4 – A comprehensive analysis of Facebook third-party applications (apps): Friends, providers, and privacy interdependence. Third-party apps on Facebook can collect personal data of the users when their friends install them. That raises serious privacy concerns as the users are not notified by the apps nor by Facebook, and they have not given their consent. In this chapter we present a detailed multi-faceted study on the *collateral information collection* of the apps on Facebook. To investigate the views of the users, we designed a questionnaire and collected the responses of 114 participants. The results show that participants are concerned about the collateral information collection and in particular about the lack of notification and of mechanisms to control the data collection. Based on real data, we compute the likelihood of collateral information collection affecting users: we show that the probability is *significant* and greater than 80% for popular apps such as TripAdvisor. We also demonstrate that a substantial amount of profile data can be collected by apps, which enables third-party application providers (app providers) to *profile* users. To investigate whether collateral information collection is an issue of the users' privacy we analysed the legal framework in light of the General Data Protection Regulation (GDPR). We provide a detailed analysis of the entities involved and investigate which entity is accountable for the collateral information collection. To provide countermeasures, we propose a privacy dashboard extension that implements privacy scoring computations to

enhance transparency towards such collateral information collection. Furthermore, we discuss alternative solutions, highlighting other countermeasures such as notification mechanisms, access control solutions, cryptographic systems and app auditing. To the best of our knowledge, this is the first work that provides a detailed multi-faceted study of this problem and that analyses the threat of user *profiling* by app providers. It can serve as a guide for future third-party apps and app providers aiming at identifying interdependent aspects of privacy and helping design apps considering privacy [149, 76].

This is a joint work with Gergely Biczók, Fatemeh Shirazi, Cristina Pérez-Solà, Jessica Schroers, Pagona Tsormpatzoudi and Bart Preneel and published in [290, 291] and at Computers & Security Journal [287]. I am the main author except for the network topologies and application adoption simulations, and the legal analysis.

PART III: PHYSICAL-KEYLESS CAR SHARING SYSTEMS

Chapter 5 – A security and privacy analysis and elicitation of requirements of physical-keyless car sharing systems. In this chapter we propose a novel physical-keyless car sharing system, we provide a systematic security and privacy analysis and elicit the requirements for a secure and privacy-enhancing system. We first present a high-level model describing its main entities and specifying the necessary functional requirements, allowing users to share their cars (with other users) without the need to exchange physical keys. Based on the proposed model and functional requirements, we present a detailed threat analysis of the system. We focus on the threats that affect the system and its users regarding security and privacy threats. The analysis results in a specification of an extensive set of security and privacy requirements for the system. This work can be used as a guide for a future keyless car sharing system designs and as a mean to assess the security and privacy risks imposed on users by such systems.

It is joint work with Mustafa A. Mustafa, and Bart Preneel and published in [288]. I am the main author of this work.

Chapter 6 – SePCAR: A secure and privacy-enhancing protocol for car access provision. In this chapter, we present an efficient secure and privacy-enhancing protocol for car access provision, named SePCAR. The protocol is fully decentralised and allows users to share their cars conveniently without sacrificing their security and privacy. It provides generation, update, revocation, and distribution mechanisms for access tokens to shared cars, as well as procedures to solve disputes and to deal with law enforcement requests, for instance in the case of car incidents. We prove that SePCAR meets its appropriate security and privacy requirements and that it

is efficient: our practical efficiency analysis through a proof-of-concept implementation shows that SePCAR takes only 1.55 seconds for a car access provision.

It is joint work with Abdelrahaman Aly, Mustafa A. Mustafa, Bart Mennink, Siemen Dhooghe and Bart Preneel and published in [285]. I am the main author except for the security proofs and the performance evaluation.

Chapter 7 – Conclusions and future work. In this chapter, we conclude this dissertation, and we present open questions pointing to the future directions of our work.

1.7 Further contributions

In this section, we shortly summarise other findings that played an important role in developing the results presented in this thesis, in one way or another. Therefore, we summarise the additional contributions, presenting them in inverse chronological order.

Security analysis of the drone communication protocol: Fuzzing the MAVLink protocol. The MAVLink protocol, used for bidirectional communication between a drone and a ground control station, will soon become a worldwide standard. The protocol has been the subject of research many times before. Through this work, we introduce the method of fuzzing as a complementing technique to the other research, aiming to find vulnerabilities that have not been found until now. The goal is to identify possible vulnerabilities in the protocol implementation aiming to make it more secure.

It is joint work with Karel Domin (main author) and Eduard Marin and published in [91]. I helped on the security analysis of the MAVlink protocol.

A literature survey and classifications on data deanonymisation. The problem of disclosing private anonymous data has become increasingly serious particularly with the possibility of carrying out deanonymisation attacks on publishing data. The related work available in the literature is inadequate regarding the number of techniques analysed and is limited to certain contexts such as Online Social Networks. We survey a large number of state-of-the-art techniques of deanonymisation achieved in various methods and on different types of data. We aim to build a comprehensive understanding of the problem. For this survey, we propose a framework to guide a thorough analysis and classifications. We are interested in classifying deanonymisation approaches based on type and source of auxiliary information and the structure of the target datasets. Moreover, potential attacks, threats and some suggested assistive

techniques are identified. This work can help the research community in gaining an understanding of the deanonymisation problem and assist in the advancement of privacy protection.

It is joint work with Dalal Al-Azizy (main author), David Millard, Kieron O'Hara and Nigel Shadbolt and published in [10]. I helped on the research methodology and provided feedback for the deanonymisation framework.

Collateral damage of Facebook apps: an enhanced privacy risk scoring model. Establishing friendship relationships on Facebook often entails information sharing between users and their friends. In this context, Facebook offers apps developed by app providers, which may grant access to the personal data of users via apps installed by their friends. Such access takes place outside of the Facebook ecosystem with the user not being aware whether a friend has installed an app collecting her data. In some cases, one or more app providers may cluster several apps and thus gain access to a collection of personal data. As a consequence privacy risks emerge. Previous research has mentioned the need to quantify privacy risks in Online Social Networks (OSNs). Nevertheless, most of the existing work do not focus on the personal data disclosure via apps. Moreover, the problem of personal data clustering from app providers has not been studied. In this work, we perform a general analysis of the privacy threats stemming from the personal data requested by apps installed by the friends of a user from a technical and legal point of view. To assist users, we propose a privacy risk scoring formula to calculate the amount of personal data that are exposed to app providers. Moreover, we propose algorithms that based on clustering, computes the visibility of each personal data to the app providers.

This is COSIC technical report [286] and is joint work with Pagona Tsormpatzoudi and Bart Preneel. I am the main author.

SAML Revocation List (SARL): A revocation mechanism for long-lived assertions on Shibboleth. SARL is the abbreviation of the Security Assertion Markup Language (SAML) Revocation List. It provides the revocation functionality in Single Sign-On (SSO) systems for long-lived assertions, i.e., authentication tokens. SSO systems use SAML assertions to allow users to log in from different devices such as mobile, desktop, tablet and laptop to multiple services. Usually, authentication tokens such as SAML assertions expire in a short period. However, there are cases that a long-lived SAML assertion is necessary such as when identification and authentication should be performed when there is a lack of network connectivity. The SARL project drives a solution for long-lived SAML assertions as a requirement for managing assertions in cases where these should be revoked, denying access to a user for a specific service. For instance, an authentication token can have a validation period of a week, a month or more. Unlikely, a device, for which an authentication process

occurred, can get stolen allowing anyone holding the authenticated device to have access to the services. With a proof of concept, we demonstrated that a revocation process is possible as a solution for denying access to a specific service and (stolen) device. There are various SSO systems such as Shibboleth, Central Authentication Service (CAS) and OAuth for which the revocation process for long-lived authentication tokens can be implemented. Most of them use two main components. A Service Provider (SP) which provides a service to the users and an Identity Provider (IdP) which is responsible for authenticating a user and generating the authentication token. Each authentication token will be sent to the SP each time a user requests to access a service. The current implementation can be deployed in various SSO systems. However, our proof of concept was deployed and tested for the Shibboleth SSO system, a well known open source SSO platform actively used in many real-world authentication services. For Shibboleth, the authentication token, that the IdP generates and the SP consumes, is called SAML assertion. The applicability of our current SARL implementation in various SSO systems is derived by the use of simple and essential tools such as the Operating System (OS) commands and Web technologies (e.g., PHP, MySQL and browser cookies). For instance, the collection of the required information in SAML assertions, i.e., attribute values, is extracted using a parser script while the logout functionality is implemented using cookies. The parser script uses Linux commands to extract the necessary information from the Shibboleth log files. Furthermore, when a revocation process takes place, the SARL application set cookies validation time to expire and forces the service to deny access to a specific user.

This is COSIC technical report [289] for an iMinds project. I am the main author.

Part I

Preliminaries

Chapter 2

Background and related work

Information is power.

PIERS ANTHONY, *American author*

In this chapter, we give an overview of the background of Online Social Networks (OSNs) and physical asset sharing systems, and we provide the related work for the *interdependent privacy* issue on Facebook and the privacy threats of information dissemination in car sharing systems.

2.1 The past

The second half of the twentieth-century has played a crucial role in establishing and advancing the Information and Communication Technologies (ICTs). The early development of computer networks is often attributed to the first artificial satellite, Sputnik, in 1957. In the late 1960s, ARPANET was developed as a computer-mediated communication network for exchanging of text messages, calculations, accessing data and, soon afterwards, electronic mail and information exchange [204]. In the mid-1980s, the computers—ARPANET—based network evolved into the *Internet*, an inter-networking of computer networks. The Internet or "network of networks", allowed the sharing of resources and information globally. In the early 1990s, the Internet evolved through the *World Wide Web*, originally developed as a "Web" of hyperlinked documents for scientific publishing by Tim Berners-Lee and his colleagues at the European Laboratory for Particle Physics (CERN) in Switzerland. The Internet and the Web allowed the interplay between people, resources, information, services, and

technologies. It enabled multiple online providers to support many services intended for various users [150].

ICTs and the Internet became an essential infrastructure across the globe transforming many aspects of the society for a wide range of human activities and social interactions. People use the Internet for a variety of purposes in their everyday lives and work to sharing information and goods; applications include OSNs, blogs, wikis and booking cars and apartments for their next trip.

OSNs represent the most popular online communities that enable users to discuss, co-create, modify and share user-generated content. Although online communities of this nature emerged in the late 1990s, they rose significantly in early 2000, attracting large amounts of users and becoming integrated into their daily routines. Early chat-rooms and bulletin boards introduced the notion of the profile, where users could link their screen name to their personal information. In the UNIX-based operating system, users could store their profile information in .project and .plan files and display them with the "finger" command. With the rise of the Web, homepages became an essential site of profile information where users could store their photos, biography and other personal information [92].

Early OSNs such as Friendster and the Whole Earth 'Lectronic Link (WELL) were profile-centric representing individuals within a system resembling typically for online dating and social events [42]. They were designed explicitly around the profile of users, enabling them to upload photos and filling fields with self-descriptive text. Later on, OSNs provided two notable possibilities to users. First, they allowed them to connect with others such as their friends. Second, they enabled a public space for user-supplied comments where individuals can navigate through other profiles such as traverse through the list of their friends (early definition of OSN [42]). Such possibilities, enabled an enormous growth for OSNs; 65% of the Internet-using adults in the United States (U.S.) were reported to use OSNs such as Twitter, LinkedIn and Facebook in 2010 [14].

In late 2010 until now, information sharing between users has been ever more extended with the use of new information and communication technologies such as personal computers, mobile devices, wireless networks and online platforms [150]. Online platforms such as cloud infrastructure allowed users to store information online in significant volumes and share it on demand with other users. Such advance possibilities in data storage and network capacity enabled a wide range of applications and services such as online marketplaces, payment systems, and platforms for the collaborative economy [152]. Users are storing their photos online using Google Drive, iCloud and Dropbox, book their trips from their phones using Booking.com and share properties using AirBnB and DriveNow.

2.2 The present

2.2.1 Online social networks in a nutshell

Nowadays, OSNs have become a popular online platform for people to express themselves, interact with each other and get their daily entertainment. OSNs allow users to generate content, connect with other users and share information [42]. Currently, there exists a rich and diverse ecosystem of OSNs enabling users to post instant news, activities and professional achievements such as using Twitter [3], LinkedIn [2] and Facebook [121]. For instance, LinkedIn is focused on professional networks where users can post professional achievements. Twitter is an online news OSN site, where users can post and receive real-time messages, i.e., "tweets". Facebook, a flagship OSN, connect users and provide easy–to–use tools for posting updates and sharing multimedia content.

Following the early definition by Boyd and Elison [42], an OSN is as an online platform that offers users the possibility to build their online profiles in a bounded system (i.e., online platform), connect with others such as friends, i.e., "latent ties" [157], view their profiles and share information with them. Their profile can contain information about the age, location, interests and the "about me" of users. Moreover, OSNs allow users to upload and share multimedia content. Extending the initial definition of OSNs [42], OSNs provide open Application Programming Interfaces (APIs) allowing other third-party application providers (app providers) to develop third-party applications (apps) (see OSN definition 2.0 by Boyd and Elison [99, 94]). That enables third-party apps to gain access at the OSN social graph allowing information to transfer from the OSNs into online platforms of third-party app providers. For example, Facebook allows users to add modules, i.e., third-parties apps, which can be developed by either Facebook or third-parties app providers, offering games, lifestyle and entertainment possibilities such as Candy Crush Saga [119], Angry Birds [260] and TripAdvisor [310].

2.2.2 Interdependent privacy and the app-related privacy issues in online social networks

We describe the related work on privacy issues that can arise from the use of third-party apps in the Facebook ecosystem. Chaabane et al. [5] showed that apps can have tracking capabilities and disseminate the collected information to "fourth party" app provider [4]. Similarly, Huber et al. [165] developed an automated evaluation tool, AppInspect [15], and demonstrated that personally identifiable information of Facebook users was leaked by a large set of Facebook third-party apps to analytic and advertisement companies [165]. To the best of our knowledge, Weng et al. [323] were the first to report the fact that the Facebook API allows third-party apps to

collect the profile of users through their friends, and Biczók and Chia [35] were the first to introduce the notion of *interdependent privacy*. From their work, Biczók and Chia modelled the impact of the interdependent privacy problem performing a game theoretic study (2-player, 1-app). Sharing a user's information without his/her direct consent can lead to the emergence of externalities; positive externalities, e.g., personalised experience for social-networking apps and negative externalities, e.g., exposed profile items. Pu and Grossklags [249] extended the work of Biczók and Chia [35] and developed a formula to simulate the app adoption behaviour of users and their friends and the monetary impact of the interdependent privacy problem [250]. Currently, Harkous and Aberer [156] extended the *interdependent privacy* problem to cloud services ecosystems such as Dropbox and Google Drive. They studied the privacy loss of a user when the user and his/her collaborators grant access to a shared file stored by a cloud vendor.

Adding to these concerns, Thomas et al. [305] examined the lack of privacy controls over the shared content of friends on Facebook. Alice, Bob's friend, can share information and unintentionally violate Bob's privacy leading to *privacy conflicts*. Using a formal representation of the privacy conflict concept, they estimated the aggregated information of Bob under the presence of privacy conflicts with Alice which, can lead to uncovering sensitive information about Bob. Frank et al. [133] showed that low-reputation apps often deviate from the permission request patterns, while Chia et al. [59] showed that certain apps collect more information than necessary. Moreover, Wang et al. [323] identified third-party app bad practices for privacy notice and consent on Facebook, studying $1,800$ most popular apps. For instance, using the Facebook API, an app can overwrite the users' and their friends' privacy settings in a calendar, displaying the birthdays of the user and her friends while the privacy settings are set to "Only me" on both sides.

2.2.3 Physical asset sharing and ICTs: the case of car sharing

Services for sharing physical assets are many-fold: 17 sectors have been identified from transport through accommodation rental to utilities [242]. These services can be defined as peer-to-peer based activities between users for dynamic share, access and use of goods such as apartments or cars. Sharing possibilities are enabled by using ICTs such as mobile networks, apps, and online platforms [23, 62]. For instance, consider the case of car sharing systems. Portable devices, in-vehicle telematics and online platforms allow users to share cars in a convenient way [215]. There are many examples of successful and well-regarded physical asset sharing systems such as Uber, Airbnb, Zipcar, TaskRabbit and eBay. The success of such systems is best demonstrated by Uber ($69 billion) and Airbnb ($31 billion), the two flagship companies that have attained astronomic valuations.

Even though the concept of car ownership may not cease, the idea of car sharing is gaining popularity. Nowadays, several car companies consider car sharing systems as a prominent solution and have started investing in them. It includes car manufacturers and suppliers such as Volvo [320], Daimler [53], BMW [39], Toyota [316] and Continental [64], rental companies such as Europcar, Hertz, Sixt and railway operators such as Deutsche Bahn [269], existing car sharing companies such as Zipcar [342] and Lyft [198] and big technological firms such as Apple [315]. Within Europe, car sharing is considered by many analysts as an increasing vector alongside with other connectivity-based sharing systems such as accommodation rental, i.e., Airbnb [6].

Car sharing also offers a high potential for the smart cities concept [61, 226, 160]. It can allow cities to manage their available resources and assets in a more effective, efficient and sustainable manner [226]. One of the major challenge in modern cities is managing efficiently the usage of transportation assets (i.e., *Smart Mobility*) such as cars, by better utilising the already available ones [281, 162]. Note that, on average cars are being used 5% of their time [93]. Their optimal usage can contribute to a decrease in the number of cars, effectively reducing the need for parking space [226]. Moreover and unlike traditional car ownership, it can also provide a relatively inexpensive alternative to users who need a car occasionally and on-demand [83, 333].

2.2.4 Secure and privacy preserving vehicle information sharing protocols

Prior research in the domain of information sharing considering cars is mainly focused on topics of (i) driving-behaviour monitoring, location-tracking and privacy of data mining and (ii) security analysis, access control and secure entity authentication. Note that in the domain of car sharing systems, existing work focuses on the realm of car relocation strategies [330, 341].

Considering relocation strategies, Weikl and Bogenberger [330] analysed and evaluated different relocation algorithms for free-floating car sharing systems. They proposed a model for optimal vehicle positioning and relocation, having an off-line and on-line demand module. The offline demand module calculates the optimal car pick-up location based on yearly data, whereas the on-line demand module performs the calculations several times per day based on real-time data. For the station-based car sharing systems, Zhu et al. [341] proposed an optimisation approach for determining the depots location using deep learning.

Considering privacy, Enev et al. [101, 306] demonstrated high identification rates of individuals (15 drivers), from 87% to 99% accuracy, based on data collected by the sensors of a car from 15 minutes of open-road driving. Martínez-Ballesté et al. [205] introduced the concept of citizens' privacy in *smart cities* by distinguishing the

following five dimensions: identity privacy, query privacy, location privacy, footprint privacy and owner privacy. Li et al. [191] analysed the information collection by mobile devices of users in *smart cities* and raised the alarm for data over-collection and the risks to privacy. Pan et al. [239] proposed an analysis of traces of moving objects in *smart cities* such as cars, aiming to depict semantics concerning mobility patterns and city dynamics.

Mustafa et al. [224] performed a security analysis on smart electric vehicle charging systems. Troncoso et al. [311] proposed a pay-as-you-drive scheme to enhance the location privacy of drivers by sending aggregated data to insurance companies. Balasch et al. [21] proposed an electronic toll pricing protocol where the onboard unit of a car calculates locally the driver's annual toll fee while disclosing a minimum amount of location information. For colluding (dishonest) users [21], Kerschbaum et al. [180] proposed a privacy-preserving spot checking protocol that allows observations in public spaces. Mustafa et al. [225] proposed an anonymous electric vehicle charging protocol with billing support. EVITA [116, 336] and PRESERVE [248, 243] are designated projects on the analysis and specification of the secure architecture of onboard units. Raya et al. [254] described the need for a Vehicular Public-Key Infrastructure (VPKI), and Khodaei et al. [182] proposed a generic pseudonymization approach to preserve the unlinkability of messages exchange between vehicles and VPKI servers. Our work is closely related to the protocol proposed by Dmitrienko and Plappert [88]. They designed a centralised and secure free-floating car sharing system that uses two-factor authentication including mobile devices and RFID tags, e.g., smart-cards. However, in contrast to our research, their protocol assumes a fully trusted car sharing provider who has access to the master key of smart-cards and also collects and stores all the information exchanged between the car provider and their users for every car access provision.

2.3 The future

It is hard to predict the future. However, the use online platforms and systems that utilise ICTs will continue to expand resulting in an ever-increasing rate of information collection from users. In May 2018 Facebook announced a new dating service that could result to collect detailed information about partnership preferences of individuals [297]. Moreover, there are reports that car manufacturers are investigating new streams of revenues aiming to utilise the features that connected cars provide [138, 83]. One promising direction for car manufacturers is to expand and become car-sharing platform operators, acquiring information about the car condition, the car settings of the driver and passengers, and also the infotainment information of what the passengers watch and listen. Analysing and designing solutions aiming to safeguard the privacy of

individuals be default and by design will continue to be an essential and necessary direction for research in the near future.

Chapter 3

Tools for analysis, elicitation of requirements and design of privacy-enhancing systems

There are three principal means of acquiring knowledge. observation of nature, reflection and experimentation. Observation collects facts; reflection combines them; experimentation verifies the result of that combination.

DENIS DIDEROT, *French philosopher*

In this chapter, we describe the tools we use throughout this thesis. We map these tools to the specific stages of Secure-System Development Life Cycle (S-SDLC) and, thus, we structure this chapter (see Sect. 1.5). These stages are: *analysis* (see Sect. 3.1), *compilation of requirements* (see Sect. 3.2) and *system design* (see Sect. 3.3).

3.1 Analysis tools

In this section, we describe the tools which are necessary for the analysis of this thesis.

3.1.1 Online questionnaire: a tool for survey studies

An online questionnaire is a tool for collecting information about the opinions of individuals [317]. Using ICTs and online platforms, online questionnaires provide the ability of large-scale sampling, large geographical distribution, convenience in time of answering questions, respondents' anonymity and low cost of distribution [115]. Questionnaires as a tool can be used for full-fledged survey studies or pilot studies. A pilot study is a short version of full-scale study that can be used as a pre-testing instrument [246]. It offers the advantage of testing a research question and it helps to identify the main directions of the study [80]), providing valuable insights [317]. Pilot studies can be qualitative and used as a primary instrument to collect and analyse data in a relatively unexplored topic and as a mean to design a subsequent quantitative full-scale study [294].

In OSNs, studies that utilise questionnaires are extensively used for problem analysis [145, 326, 7, 44]. On Facebook, Liu and Gummadi [195] studied the discrepancy between the desired privacy in contrast to reality. Through a Facebook third-party app, they collected responses of 200 participants and compared them to the participants' actual privacy settings. They identified that almost 36% of the information was shared with the default privacy settings while 37% matched their actual desired setting; the default privacy settings exposed the participants' information to more Facebook users than expected. Wang et al. [325] identified bad practices of third-party apps for privacy notice and consent, and they proposed enhanced app authorisation dialogues. To validate their authorisation designs, they performed a survey study. Interestingly, a small set of participants reported the unfair privacy practices of Facebook third-party apps that can gather information about one's friends, i.e., "user does not own [his or her] friends' information". That information collection is happening without their notice or consent, i.e., "[the user's] friends never download or agreed to the application's term". Such unfair privacy practise initially investigated by Bloemendaal et al. [38]; whether users were *aware* of this problem. They collected responses of 250 participants and identified that 70% of them were unaware of its existence, i.e., the *interdependent privacy* issue. In other words, 70% of the participants shared all types of information of their friends while installing third-party apps on Facebook, without being aware of it. However, they did not investigate whether participants were *concerned* as a fact of users' lacking notification and consent of the interdependent privacy problem.

3.1.2 Simulation tools: network topologies and application adoption models

OSNs can be represented as graphs, consisting of nodes as users and edges demonstrating the connections among them. The characteristics of graphs can be described by the diameter, the clustering coefficient and the degree distribution. The *diameter of a graph* is the maximum distance between any pair of nodes, i.e., the maximum number of edges along the shortest path connecting an arbitrary pair of nodes in the graph. For example, William et al. [309] identified that there is "six degrees of separation" between any two people in the United States, while Lederer et al. [11] identified that distances are even shorter on Facebook corresponding to a diameter of four. *Clustering* in OSNs is related to users and their circles of friends or acquaintances. The *clustering coefficient* quantifies how well connected the neighbours of a node in a graph are [273]. For instance, Mislove et al. [218] found that the periphery of OSNs consists of strongly clustered, low-degree nodes connected by a core of high-degree nodes. The degree of a node in a graph is the number of connections it has to other nodes; the *degree distribution* is the probability distribution of these degrees over the whole graph. [329]. For example, Mislove et al. [218] found that the degree distribution of OSNs follows a power law.

There are several relevant models for generating OSN topologies. The Barabási-Albert model [11] generates networks with a power-law degree distribution using preferential attachment; several studies about real-world OSNs confirm the existence of such a power-law [127, 218]. The Watts-Strogatz model generates small-world networks, i.e., networks with high clustering and small diameter, properties that have also been found in many OSNs [334, 218]. Note that the Watts-Strogatz model can generate graphs with high clustering without exhibiting a power-law degree distribution, whereas the Barabási-Albert model creates a power-law degree distribution without high clustering. The "small-world phenomenon" was also studied by Kleinberg [183]. The main purpose of the Kleinberg model is to study the search for short paths within graphs in a decentralised way, a task that is not related to our evaluations. The Erdős-Rényi model [102] generates a random graph with uniform degree distribution and, therefore, can be used as a baseline in network simulations.

Considering the *adoption models* of apps in OSNs, a uniform model can be used for baseline analysis. In the uniform app adoption model, all users install an app with the same probability, and each installation is independent of other installations and the underlying network topology. However, owing to local network effects [283] prevalent in OSNs, a preferential model is more realistic (see Fig. 3.1). The probability of a friend of a user installing an app is proportional to the number of friends who have already installed the app. Even if Alice (node 2) is not directly connected to Bob (node 19), and does not benefit directly from Bob's application adoption, Alice and Bob may still affect each other through the *local network effect*, as a friend of Alice (node 1) may

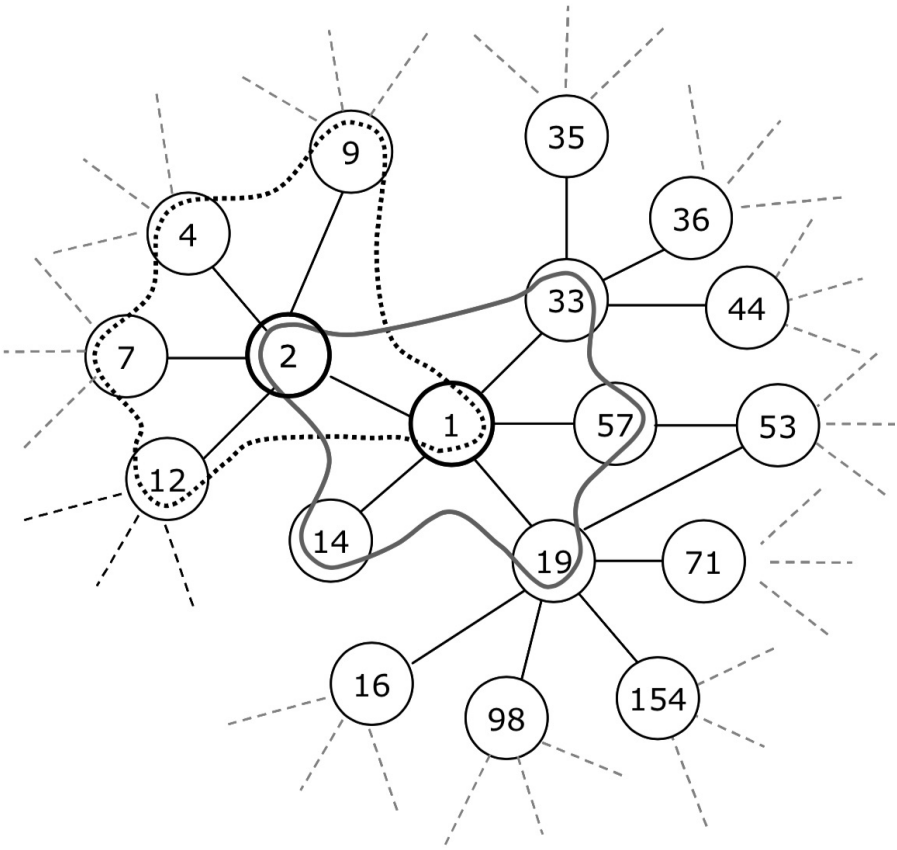


Figure 3.1: Local network effect and its effect to application adoption [283].

be a friend of Bob as well. In other words, strongly clustered *local networks* are often interconnected to each other, i.e., the whole network, via high-degree nodes (hubs).

3.1.3 Quantification metrics: privacy score in online social networks

Quantification metrics such as privacy scores can measure the level of information disclosure of a user towards an entity such as a friend or an app provider [228]. For privacy savvy users, a privacy score can inform and better support decisions about exposing or hiding their personal information towards a given entity. Tools for enhancing transparency and awareness, is in line with the requirements of data

protection by design [85, 108, 209] of the General Data Protection Regulation (GDPR) [113], as they can strengthen users' control on the disclosure of personal data.

Considering quantification metrics, Maximilien et al. [208] proposed a formula to estimate the privacy score for a user, evaluating the number of profile items the user has disclosed to their friends on Facebook. They computed the privacy score as the product of *sensitivity* and *visibility* of personal data. Liu and Terzi [194] extended this work and proposed a framework for computing the privacy score using a probabilistic model based on Item Response Theory (IRT). Although IRT presents an interesting approach for computing the *sensitivity* of the user's personal data, *visibility* is not properly addressed. Moreover, Sánchez and Viejo [265] developed a formula to assess the sensitivity of unstructured textual data, such as wall posts in OSNs . Their model aims to control the dissemination of the user's data to different recipients in OSNs [319]. Minkus et al. [217] estimated the *sensitivity* and *visibility* of the privacy settings based on a survey of 189 participants. Finally, Nepali and Wang [229] proposed a privacy index to evaluate the inference attacks as described by Sweeney [284], while Ngoc et al. [230] introduced a metric to estimate the potential leakage of private information from public posts in OSNs. To the best of our knowledge, there is no related work quantifying the collateral information collection of third-party apps and app providers in OSNs.

3.2 Security and privacy threat analysis: the elicitation of requirements

For a systematic compilation of the security and privacy requirements, a threat analysis approach is required [169, 279, 135]. Threat modelling aims at identifying (all) possible entities and actions that can harm the system as a whole or its distinct parts [90, 236]. Threats and adversarial actions can be realised by an arsenal of attacks. To identify the attack surface of a system we refer the reader to two well-known threat modelling frameworks: STRIDE [163, 213, 235] and LINDDUN [84]. STRIDE mainly covers security threats while LINDDUN focuses on the privacy threats. Based on the capabilities of an adversary and the threats identified during the system analysis, the requirements for system design can be defined aiming at providing security and privacy guarantees (see Table 3.1).

Aside from the STRIDE framework, there are other security threat analysis frameworks. *Process for Attack Simulation and Threat Analysis* [313] is a risk-centric framework, *Trike* [264] is a conceptual framework for security auditing, and *Visual, Agile, and Simple Threat modelling* [8] is an Agile S-SDLC framework intended to support various stakeholders of a business life-cycle such as app architects, developers, cybersecurity

Table 3.1: Security and privacy threats and the corresponding properties.

	Threats	Properties (i.e., countermeasures)
Security	Spoofing	Entity authentication
	Tampering with data	Data authentication
	Repudiation	Non-repudiation
	Information disclosure	Confidentiality
	Denial-of-Service (DoS)	Availability
	Elevation of privilege	Authorisation
Privacy	Linkability	Unlinkability / Pseudonymity
	Identifiability	Anonymity
	Non-repudiation	Plausible deniability
	Detectability	Undetectability
	Disclosure of Information	Confidentiality
	Unawareness	Content awareness
	Non-compliance	Policy and consent compliance

personnel, and senior executives. STRIDE is the most suitable framework for security threat analysis, as it emphasises security threats and the specification of requirements which is in line with the focus of this work. Analogous to STRIDE, LINDDUN provides a systematic analysis of privacy requirements. For our threat analysis, we chose STRIDE and LINDDUN as both frameworks are used by both industry [235, 213] and the research community [84, 318].

3.3 Secure and privacy-preserving technologies in a nutshell

In this section, we describe the countermeasures we use for our designs of secure and privacy-enhancing car sharing protocol and the *collateral information collection* of Facebook third-party apps and app providers.

3.3.1 Symmetric key cryptosystem

A symmetric key cryptosystem is a cryptosystem that allows two parties, i.e., the sender and receiver, to confidentially communicate over an insecure channel. They use the same key to perform cryptographic operations, concealing the content of the messages exchanged towards any other party that might be observing and listening to the channel. A party can be any physical entity or a computer terminal. A channel can be any

communication link between the two parties such as the Internet, wireless LAN and mobile networks. For instance, Alice wants to send the details of the location and the access code of her car to Bob without Oscar (i.e., an opponent) to be able to understand and retrieve the content of the exchanged message [211, 237].

Symmetric key cryptography can provide powerful solutions and can be used to offer *message confidentiality*. To transmit a message m confidentially, the sender uses an encryption function denoted as $\mathsf{E}()$ to generate a ciphertext c using as inputs the message m and a secret key (or symmetric key) K such that:

$$c \leftarrow \mathsf{E}(K, m) \ .$$

The ciphertext is transmitted to the receiver through an insecure communication channel. The receiver uses a decryption function denoted as $\mathsf{D}(K, c)$ to decrypt the message m using the same secret key as the sender K [223] such that:

$$m \leftarrow \mathsf{D}(K, c) \ .$$

According to *Kerckhoffs' principles* on the cryptographic functions of encryption and decryption utilised in a system should be "if not theoretically unbreakable, unbreakable in practice", "compromise of the system details should not inconvenience the correspondents", the "key should be rememberable without notes and easily changed", the "cryptogram should be transmissible". the "encryption apparatus should be portable and operable by a single person" and the "system should be easy, requiring neither the knowledge of a long list of rules nor mental strain" [244, 211]. The benefit of using a strong symmetric encryption function is that Oscar will learn nothing from the encrypted message as the ciphertext will be indistinguishable from random data. A well-known symmetric key cryptosystem is the Advanced Encryption Standard (AES) [95, 71].

3.3.2 Asymmetric key cryptosystems

An asymmetric key cryptosystem or public-key cryptosystem is a cryptosystem which parties, such as the sender and receiver, can communicate over an insecure channel securely. They are using a pair of keys which are distinct but mathematically linked to perform cryptographic operations. One of these keys namely the public key, i.e., known to everyone, is used to encrypt a message, whereas a matching private key, i.e., only known to its owner, is used to decrypt a message. In short, any party can generate a key pair (Pk, Sk) and make public the encryption key Pk, i.e., public key. It allows any other party to send him/her encrypted messages. As long as the decryption key Sk, i.e.,

secret key, is kept secret, only its owner can decrypt the encrypted messages. For Alice to send a message to Bob securely, she uses the public key of Bob, and the message m as inputs in a function denoted as $\mathsf{enc}()$ to encrypt and generate the ciphertext c, i.e.,

$$c \leftarrow \mathsf{enc}(Pk, m) \ .$$

Over any (insecure) communication channel, Alice sends the ciphertext to Bob. Upon Bob receiving the ciphertext, he applies an inverse transformation function denoted as $\mathsf{dec}()$ using his private key Sk and the ciphertext c as inputs to decrypt and obtain the original message such that:

$$m \leftarrow \mathsf{dec}(Sk, c) \ .$$

Public-key cryptography is used to offer information security services such as *message confidentiality*, provided by the encryption and decryption functions of a message over an insecure channel. Public-key encryption can be understood as a mail-box which everyone can place a letter in a box, but only the person who holds the key can unlock and retrieve the letters. The primary advantage of public-key cryptography is the key distribution for encryption in comparison to the symmetric counterparts.

However, public-key cryptography is computationally intensive and substantially slower than symmetric-key encryption algorithms. In practice, public-key encryption is used to establish symmetric keys, i.e., key establishment, whereas symmetric encryption is used to encrypt and decrypt messages. Hybrid protocols that incorporate symmetric and public-key algorithms are used in real-world applications such as the SSL/TLS protocol for secure Web connections [86] and the IPsec protocol for network layer security [237]. Well-known asymmetric-key cryptosystems are the RSA [256] and ElGamal [98]. Moreover, public-key cryptography offers security services such as message authenticity and non-repudiation using digital signatures and digital certificates as described below.

3.3.3 Hash functions

A hash function is a function that receives as inputs a bit-string of arbitrary length and outputs a bit-string of some fixed length, namely hash-value or message digest. A hash-value, denoted by z, is a unique representation of a message m generated by a function denoted as $\mathsf{hash}()$ such that:

$$z \leftarrow \mathsf{hash}(m) \ .$$

Unlike other cryptographic functions, hash functions do not need a key. For a hash function to be secure, the following properties need to be guaranteed [237]:

- **Collision resistance**: It is computationally infeasible to find, for two distinctive inputs m and m', i.e., $m \neq m'$, the same output, i.e., $\mathsf{hash}(m) = \mathsf{hash}(m')$.
- **Preimage resistance ("one-wayness")**: Given an output z, it is computationally infeasible to find any input m such that $\mathsf{hash}(m) \leftarrow z$, i.e., $\mathsf{hash}()$ one way.
- **Second preimage resistance**: For an input m, and thus $\mathsf{hash}(m)$, it is computationally infeasible to find any $m \neq m'$ such that $\mathsf{hash}(m) = \mathsf{hash}(m')$.

Well-known hash functions are RIPEMD-160 [89] and the SHA-2 [77] and SHA-3 [220] family of functions.

3.3.4 Message authentication code algorithms

A Message Authentication Code (MAC) algorithm is a cryptographic check-sum that combines a shared symmetric-key and a message of arbitrary length as inputs and outputs a fixed length bit-string. Given a message m and a symmetric secret key K, a sender generates an authentication tag denoted as $\mathsf{mac}(K,m)$ such that:

$$v \leftarrow \mathsf{mac}(K,m) \ .$$

The sender appends the authentication tag v to the message m and sends both to the receiver through an insecure communication channel. For instance, Alice sends to Bob the availability period of her car, i.e., the pair (m,v). Upon receiving the pair, the receiver verifies the authenticity of the message, since only the sender holds an identical copy of the symmetric secret key. Moreover, the integrity of the message can also be verified using the authentication tag. Bob recomputes the MAC-value using the message received such that $v' \leftarrow \mathsf{MAC}(m)$ and compares it for equality to the original MAC-value, i.e., $v' \stackrel{?}{=} v$. If the equality holds then the message has not been altered during transit. The receiver can verify that the originator of the message was Alice, and can detect whether the message was altered in transit; Oscar cannot forge the message and compute a valid MAC since he lacks knowledge of the symmetric secret key.

As a keyed-dependent function, MAC provides *data origin authentication* and *data integrity* of a message. Data origin authentication guarantees that the source of the message is the claimed sender, whereas data integrity guarantees that the message has not been altered during transmission. There are several well-known MAC algorithms such as CBC-MAC based on a block cipher [30] and Hashed-based Message Authentication Code (HMAC) such as HMAC-SHA-2 family of functions [179, 186].

3.3.5 Pseudo-random Functions

A Pseudo-Random Function (PRF) is a function that generates unpredictable random bit-strings [96]. A PRF should offer strong statistical guarantees: an adversary cannot tell whether the output is produced by the PRF or a truly random function for any given input [211]. Given an input K and a seed s, the PRF function generates output indistinguishable from randomness, i.e., z, such that:

$$z \leftarrow \mathsf{prf}(K, s) \ .$$

For instance, computing the n-bit output for an arbitrary input, the probability to guess the result is extremely small and is equal to $1/2^n$ [237]. PRF functions can be implemented using AES in CounTeR mode (CTR) if $\ll 2^{n/2}$ outputs are considered.

3.3.6 Digital signatures

A digital signature of a message is a cryptographic primitive that provides means for an entity, i.e., sender, to bind its identity into a message such that it allows another entity, i.e., receiver, to verify the authenticity of the message. A digital signature uses a pair of keys, according to the principles of asymmetric-key cryptography. A private key, which is known only to the signer, is used for signing a message and a matching public key, which is known to everyone, is used for verifying that message. Digital signatures can be used to resolve disputes between parties because they can be verified by a third party without access to the private key.

Digital signatures can be considered as the digital counterpart to handwritten signatures. In detail, the use of digital signature schemes requires three functions: key generation, signature generation and signature verification. The key generation produces a public/private key pair for an entity, i.e., (Pk, Sk). To digitally sign a message, an entity (signer) can create an authentication tag for that message by utilising a signature function denoted as $\mathsf{sign}()$ and provide as inputs the message m and the generated private key Sk. Essentially, it hashes the message $z \leftarrow \mathsf{hash}(\mathsf{m})$ and digitally signs it using his/her private key Sk, such that:

$$\sigma \leftarrow \mathsf{sign}(Sk, z) \ .$$

After signing, the signature is appended to the message and the pair (m, σ) is transmitted to the receiver over an insecure communication channel. The receiver can verify the signature σ of a message m, by obtaining the public key of the sender and utilising the verification function, denoted by $\mathsf{verify}()$, i.e.,

$$true/false \leftarrow \mathsf{verify}(Pk, z', \sigma) \ .$$

The receiver (verifier) accepts the signature if the signature verification is successful, i.e., the algorithm outputs "true". The verification fails otherwise (i.e., outputs false). The private key Sk is known only to its owner, and it should be computationally infeasible for any entity other than the owner of the private key to find m such that $\mathsf{verify}(Pk', z, \sigma)$ holds.

Until now we assumed that an adversary, Oscar, can be an external entity. However, Alice and Bob may attempt to cheat each other. Digital signatures can prove to a mutually agreed Trusted Third Party (TTP) that one of two parties generated the message. For instance, if Alice (signer) denies she signed the message m held by Bob (receiver), then Bob can present the signature σ for m to a TTP along with m. The TTP can rule in favour of Bob if $true \leftarrow \mathsf{verify}(Pk, z, \sigma)$ or in favour of Alice otherwise.

Digital signatures have many applications in information security such as *message authentication* meaning that the sender of a message is authentic, *message integrity* that a message was not modified in transit and *non-repudiation* of origin (receipt) that the sender (recipient) of a message cannot deny the creation (receipt) of the message. Well-known digital signature schemes are RSA [256] and Digital Signature Algorithm (DSA) [51].

3.3.7 Public-key certificates

A public-key certificate is a means to bind a public key to an identity or to one or more attributes of an entity. It consists of the data part and signature part. The data part contains the name and attributes of an identity, the public key corresponding to that identity, the validity period of the certificate and other relevant information such as the certificate generation statement and the revocation information. The digital signature part consists of the signature of the data part and protects the public-key certificate against manipulation [66].

The public-key certificate can be self-signed when two parties have an authenticated certificate of the other's party's public key or signed by an intermediary and mutually TTP namely a *certification authority* or trusted authority. If Alice and Bob are unknown to each other and do not trust each others' public keys directly, they can trust a certification authority and establish a chain of trust. They need an authentic certificate of the public key of the certification authority. Then, the *certification authority* can generate and issue certificates for all entities in a system establishing trust between entities unknown to each other.

Upon verification of the identity of an entity, the certification authority issues a certificate containing the entity's information and public key while the generated certificate is digitally signed by the private key of the certification authority. Any entity including Bob can verify the authenticity of the certificate and the corresponding public

key of Alice. Using the public key of the certification authority, Bob can verify the authority's signature on Alice's certificate. If this signature verifies correctly, then Bob can accept the certificate and the attached public key in the certificate of Alice as authentic.

3.3.8 Secure multiparty computation: a brief introduction

A Secure Multiparty Computation (MPC) is an interactive protocol, that allows a set of n parties to jointly and correctly evaluate a function f from a given set of private inputs $\{x_1,\ldots,x_n\}$ such that,

$$y_1,\ldots,y_n = f(x_1,\ldots,x_n) \ .$$

The i-th party learns y_i and nothing else. MPC is useful when evaluations have to be performed by a set of untrusted parties. Let's assume that Alice needs to give her secret value x to a set of n parties such that, none of n parties will learn x. A simple way of doing so, is to split up x into n random "shares" $\{x_1,\ldots,x_n\}$ such that, $x = x_1 + \cdots + x_n$. If the i-th party receives only the x_i share, it cannot learn anything about x individually and Alice can secretly share x. To the contrary, if n parties collaborate they can reconstruct x. MPC protocols can evaluate any computable function [18] based on secret sharing, providing ways of evaluating on secret shared values [178]. In practice, MPC protocols can be applied in several apps such as auctions [40], voting [68], authorisation [16] and access provision protocols [285].

For instance, in sugar beet auctions in Denmark, buyers and sellers (i.e., farmers) submit their bids to an auctioneer. The auctioneer aims to find the market clearing price. Both the buyers and sellers want to pay and being paid an optimal price respectively, meaning that a buyer aims to spend less than a maximum price and a seller to get paid more than a minimum bid. If an auctioneer knows such true minimums and maximums, it can collaborate with a bidder and abuse the process. Since the auctioneer receives a percentage of the selling price, it has the incentive to increase the amount of a selling item. Such a problem, introduced by a single auctioneer party, can be solved using a set of multiple parties. Both buyers and sellers can submit their true maximums and minimums to several parties without any of the external parties learn anything but the output of the bidding process, thus keeping the input-bids private.

The security of an MPC protocol can be specified in terms of the number of corrupted parties that it can tolerate and how the corrupted parties may behave. Considering the number of corrupted parties, there is a threshold parameter t: an MPC protocol remains secure as long as no more than t out of n parties are corrupted or collude with each other [178]. An adversary can have corruption capabilities over multiple parties such that cheating parties can work together to learn information or violate the MPC

protocol logic and specifications. An adversary may aim to corrupt the majority of parties, i.e., dishonest majority, or less than half of the parties, i.e., honest majority. In short, an MPC protocol can be secure towards an honest majority of malicious parties, i.e., $t < n/2$, or a dishonest majority that can be even $t < n - 1$ number of malicious parties. For instance, Ben-Or, Goldwasser and Wigderson (BGW) [33] and Araki et al. [16] protocols are secure over honest majority of malicious parties while Goldreich, Micali and Wigderson (GMW) [141], SPDZ [75] and MASCOT [178] are capable of dealing with $n - 1$ malicious parties.

There are a number of ways to model the adversarial behaviour of corrupted parties. Corrupted parties are usually modelled considering a central adversary with a cheating strategy that aims to corrupt a set of parties [207]. In its simplest form, a *passive security* (or "honest-but-curious" or semi-honest"), an adversary aims to determine information on honest parties' secret input by analysing protocol messages while executing the protocol honestly. For instance, in an election protocol, an adversary wants to discover who voted for whom yet without corrupting the tally. A more realistic, and much stronger adversarial model is *active security* (or malicious security), where an adversary can send arbitrary messages to other parties aiming to make the protocol deviate arbitrarily and thus, force parties to reveal their inputs unintentionally [178].

The security guarantees of an MPC protocol can be further categorised into information-theoretic (or unconditionally) secure or computationally (or simulation-based) secure. *Information-theoretic* guarantees "perfect" security and is obtained unconditionally. An adversary can have unrestricted power and capabilities yet having zero or negligible probability to violate the security of the protocol successfully. *Computational* security is obtained in the presence of an adversary whose computational power is restricted [178] and on the cryptographic hardness assumptions of computation problems such as factoring of products of large prime numbers [256]. In practice, any computable function can be evaluated securely with information-theoretic security for both, passive security with no more than $t < n/2$ corruptions and active security with $t < n/3$ corrupted parties. Examples of information-theoretic MPC protocols are by Chaum et al. [57] and BGW [33] and computationally secure are by Yao [339] and GMW [141].

An implementation of any algorithm in MPC can be specified as a Boolean or arithmetic circuit using multiplications and additions over the (shared) private inputs. For the special case of Boolean circuits, such operations can be realised by XOR gates for additions and AND gates for multiplications. In practice, there are two main approaches for MPC protocols: *secret sharing* [27, 57, 141] and *Garbled-Circuits* [27, 339]. For secret sharing, parties interact for every multiplicative gate of the circuit (addition is free) while for Garbled-Circuits parties construct a garbled version of the circuit and send it at once. Secret-sharing sends small messages per (multiplication) gate, and can perform better for low-bandwidth networks achieving high-throughput comparing to garbled-circuit which are large and costly in bandwidth. For low-latency networks and high-depth circuits, Garbled-Circuits outperform secret sharing most of the time due to

constant-round of communications as secret sharing circuits grow linearly in depth of the circuit that is computed.

3.4 Transparency enhancing technologies and privacy dashboards

A wide selection of technologies enhancing transparency are listed in the surveys of Hedbom [158] and Janic et al. [175]. Along with privacy icons [161] and privacy nudges [325], privacy dashboards [293, 46] are a well-studied concept for notification and enhancing awareness of users [130]. A privacy dashboard as a transparency-enhancing tool provides summaries of the personal data of users that can be collected by other entities such as friends, apps and app providers. In short, a dashboard aims to raise users' awareness and answer the common user question: "how much does an entity know about me?", and it does so in a way the user can understand and take appropriate actions if necessary. A dashboard uses visualisations and quantification metrics in a unified, structured and declarative manner that aims to inform users better about the type and amount of disseminated data. Being declarative, dashboard designs should present information in a fair and comprehensive way, and should ensure they not mislead users [76, 189, 108]. For instance, the Data Track privacy dashboard [328] is a design of the successful European projects PRIME (FP6) and PrimeLife (FP7) [131].

Concerning privacy dashboards, several designs have been proposed to enhance transparency in OSNs [326, 59]. Bier et al. [36] proposed and implemented PrivacyInsight, which was designed under legal and usability requirements defined by the GDPR [113] and the ISO standard 9241-11 [170]. Talukder et al. [293] proposed Privometer, merely aiming at measuring the amount of sensitive information that can be leaked from a single profile of a user on Facebook. However, the authors do not investigate a scenario with either joint control of personal data or multi-app *data fusion*. (see Sect. 4.5). Buchmann et al. [46] designed the Personal Information Dashboard (PID), aiming to provide transparency representing the information that users disclose across multiple OSN domains. Although neither one of the solutions treats the analysed problem directly, their designs can be extended to implement an enhanced privacy dashboard, and include the collateral information collection case coupled with multi-app *data fusion*. Such an information collection of apps can be given by the users or through their friends, while a dashboard can foster better notification and consent by both the users and their friends; a friend might be willing to uninstall an app if it enables collateral information collection (see Sect. 4.2).

3.4.1 Alternative countermeasures to privacy dashboards

Wang et al. [325] proposed techniques to nudge users helping them to avoid online disclosure on Facebook that they may regret later. To investigate the efficiency of privacy nudges, they developed a Facebook app and complemented the study with a survey questionnaire of 21 participants. They identified that "Stop and Think" nudges are better at helping the participants to avoid regrettable postings. The "Pay attention to the audience" nudge helped them to identify the audience and be more cautious about their posts. Whereas, the "Content feedback" nudge was perceived as needless and not a very positive nudge as participants were feeling being judged by the tool. Paul et al. [241] studied an enhanced interface for the Facebook privacy settings. They proposed a coloured representation of the privacy settings (*C4PS - Colours for Privacy Settings*), to demonstrate the visibility of the profile items of users, such as photos. To verify their assumptions about the effectiveness of such a design, they created a mock-up of the Facebook privacy settings and performed a survey were they gathered responses of 40 participants (students). They identified that such an enhancement can help users to better employ the desired privacy settings.

Moreover, FSEO [25], FaceCloak [197] and NOYB [146] are privacy schemes focused on OSNs and particularly on Facebook. Their main goal is to achieve privacy by providing fake information, considering both the app provider and the OSN user as adversaries. Scramble! [26] proposes an *access control mechanism* over a user's data, making the use of encryption techniques. According to this model, authorised users have partial access to the data, depending on the access control lists. flyByNight [196] is another privacy solution for OSNs, that makes use of symmetric-key cryptosystems. This approach tries to overcome the limitations of Facebook by introducing a privacy platform through a proxy server. FaceVPSN [63] introduces a distributed platform for storing information, providing fake information to the OSN. Furthermore, there exist other solutions that propose privacy-friendly architectures such as Safebook [70], EASiER [172] and more [173, 321, 69].

3.5 Summary

In this part, we provide the necessary background, and we give an overview of the related work of this thesis. Moreover, we describe the necessary tools that we use throughout this thesis. In the following chapters, we utilise these tools to analyse the problem of 1) collateral information collection on Facebook and car sharing systems and 2) to design solutions for them.

Part II

Collateral information collection of third-party applications

Chapter 4

A comprehensive analysis of Facebook third-party applications: Friends, providers, and privacy interdependence

What Kogan offered us was something that was way cheaper, way faster and of a quality that nothing matched.

CHRISTOPHER WYLIE, *Former data scientist at Cambridge Analytica and whistle blower*

Publication Data:
I. Symeonidis G. Biczók, F. Shirazi, C. Pérez-Solà, J. Schroers and B. Preneel, Collateral damage of Facebook third-party applications: a comprehensive study, Computers & Security Journal Volume 77, Elsevier, pp. 179–208, 2018.

Contributions: Main author except for the network topologies and application adoption simulations (Sect. 4.3.3) and the legal analysis (Sect. 4.5).

4.1 Introduction

Online Social Networks (OSNs) have become a dominant platform for people to express themselves, share information and interact with each other. By design and popularity, Facebook has morphed into an immense information repository [277], storing users' personal data and logging their interactions with friends, groups, events and pages [302]. The sheer amount and potentially sensitive nature of such data have raised a plethora of privacy issues for the Facebook users including the lack of awareness from users [7], the cumbersome privacy controls [195], the accidental information disclosure [326] and the reconstruction of the identity of users [323]. In this chapter we focus on the interdependent information collection [35, 290]: third-party applications (apps) installed by users collect data about users but also about their friends, which brings significant privacy implications.

Due to Facebook's popularity, third-party application providers (app providers) use the Facebook's Developers Platform [122] to launch their apps, benefiting from the existing massive user base [277]. Currently, Facebook offers a set of more than 25 K apps [15], such as Criminal Case [120], Candy Crush Saga [119], Angry Birds [260] and TripAdvisor [310]. When a user installs an app from the Facebook app store, the app may collect their information on Facebook. For instance, Candy Crush Saga collects the name, profile picture, country, friend list and email address of a user. Other apps may collect other types of information.

When a friend of a user installs an app on Facebook, this app not only can collect the friend's information but, it may also collect information about the user herself. For instance, if a friend of a user installs travelling apps such as *TripAdvisor*, these may collect the user's current location, hometown, work location and likes. That allows for sharing travel preferences [202] aiming to notify both the user and their friends whether they have visited the same point of interest. If a friend of a user installs a horoscope app such as *Horoscopes*, the app may collect the friend's and the user's birthday.

App installation and information flow. From a technical standpoint, Facebook relies on permission-based platform security and applies the least privilege principle to third-party apps [122]. For installation and operation, each app requests from the user a set of *permissions*, granting the app the right to access and collect information such as the profile name (steps 1 to 3 in Fig. 4.1). After the user's and Facebook approval, apps can collect the profile information of the user (i.e., personal data). Hence, they can store it on servers outside Facebook's ecosystem and out of the user's control (steps 7 in Fig. 4.1) [118, 155].

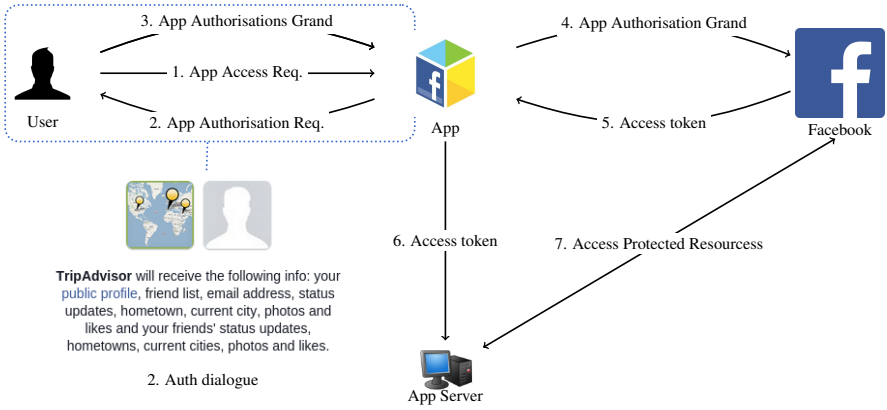


Figure 4.1: Facebook Developers app architecture.

Facebook API and friend permissions. The API v1.x of Facebook (April 2010 – April 2015) provided a set of permissions to the apps, i.e., *friends_xxx*, such as *friends_birthday* and *friends_location* [122, 291]. These permissions gave the apps the right to access and collect personal data of users via their friends, such as their birthdays and locations. Facebook, updated the API to version v2.x (April 2014), and replaced the *friends_xxx* permissions with the single *user_friends* permission, stating that since 2015 (API v2.0) this problem has been mitigated; it required mutual consent and mandated app reviews [123]. Interestingly, in our previous work we identified that by default the *user_friends* permission could retrieve information of up to fourteen user profile attributes via the apps installed by the friends of a user up until the v2.3 of the API (April 2014 – July 2017) [292]. To analyse the number of friend permission requests among popular apps [290], we used the publicly available AppInspect dataset [15] provided by Huber et al. [165] (2012 – 2014) and identified a proportion of more than 10% of such collateral permission requests.

The Cambridge Analytica case. Data analytics companies such as Cambridge Analytica (CA) can collect rich information about individuals' everyday lives and habits from big data-silos, enabling profiling and micro-targeting such as in the case of political elections [221]. As it has been reported at several major news outlets, in 2015 approximately 87 million profiles of users on Facebook have been harvested by Aleksandr Kogan's app "thisisyourdigitallife" in collaboration with CA [302, 298, 227]. They exploited the collateral information collection mechanism of third-party apps on Facebook, where the "thisisyourdigitallife" app was installed by 270,000 users reaching tens of millions of friends [210, 335]. The collected data have been used to draw a detailed psychological profile for every person affected [314], which in turn enabled CA to target them with personalised political advertisement potentially

affecting the outcome of the 2016 U.S. presidential elections [221, 262, 185]. Both Kogan and CA have denied allegations and said they have complied with regulations and acted in good faith [48, 47].

Privacy interdependence and collateral information collection. The sub-optimal privacy controls and the server-to-server (and potentially offline) communication between Facebook and app providers make any privacy protection mechanism hard to apply [100]. As a result, the user's profile attributes can be arbitrarily retrieved by an app provider without automatic notification or on-demand approval by the user through their friends.

Collateral information collection may inflict a privacy loss due to the lack of transparency and consent given by the users for the case of Facebook third-party apps. Fundamentally speaking, *collateral information collection* is the manifestation of *interdependent privacy*; the scenario when the privacy of an individual user is affected by the decisions of other users [35]. From an economic point of view, sharing a user's information without their direct consent can lead to the emergence of externalities, i.e., unwanted side-effects. While sharing someone else's information may yield benefits for them (positive externalities, such as personalised experience in apps), it is also almost certain to cause a decrease in their utility (negative externality, e.g., exposed profile items). Existing research is limited to pointing out the existence of and risks stemming from such negative externalities on the Facebook app ecosystem [35], and its potential impact on app adoption [249, 251].

App providers, data fusion and profiling. Third-party app providers can be owners of several apps (see Fig. 4.2): the app provider$_1$ offers the apps A_1, A_2, A_3 and A_4. For instance, the app providers Vipo Komunikacijos and Telaxo offer 163 and 130

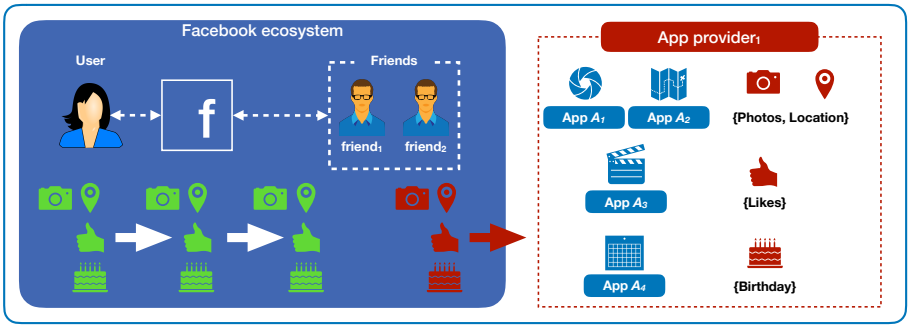


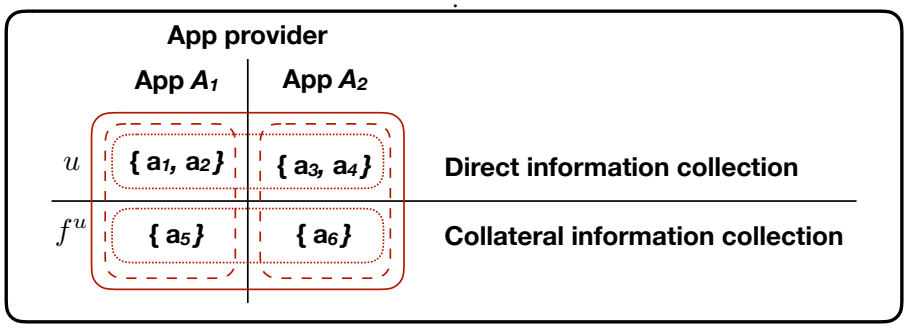
Figure 4.2: *Collateral information collection* scheme on Facebook.

apps; among those, 99 and 118 have more than 10 K monthly active users, respectively (extracted from the AppInspect dataset [15]). As a consequence, an app provider may cluster several apps and thus may collect more personal data from the profile of the users. Moreover, every app retrieves the Facebook user's ID, which uniquely identifies a user over apps. Hence, the app provider could utilise a type of data fusion over all apps offered [151], and construct a relatively complete representation of the profiles of the users. Such data fusion partly constitutes and enables *profiling* as defined in Article 4 of the General Data Protection Regulation (GDPR) [113]. Note that this definition is analogous to its common meaning in marketing in the context of consumer behaviour [176].

Table 4.1 illustrates the mutually amplifying interaction between *collateral information collection* and multi-app data fusion. It is clear that collateral information collection allows an increasing amount of data collection through vertical coupling for example, by adding $\{a_5\}$ to the directly collected $\{a_1, a_2\}$. On the other hand, data fusion over multiple apps alone allows for horizontal coupling such that combining $\{a_1, a_2\}$ and $\{a_3, a_4\}$ into a consistent $\{a_1, a_2, a_3, a_4\}$. With both mechanisms in full effect, the app provider is able to compile an extended attribute set of $\{a_1, a_2, a_3, a_4, a_5, a_6\}$. Therefore, with the help of multiple apps installed by the friends of a user, an app provider could profile a user partly or entirely without her consent, which constitutes a privacy breach and has legal implications [113].

Contribution. In this thesis, we investigate the collateral information collection through Facebook third-party apps installed by one or more friends of a user, taking into account the fact that an app provider may own multiple apps. Specifically, we identified five research questions to advance our understanding of indirect and collateral information collection in the case of Facebook apps.

Table 4.1: The mutually amplifying effect of *collateral information collection* and multi-app data fusion



- *Are users concerned that apps installed by their friends can collect their profile data on Facebook?* To identify whether users were concerned, we designed a questionnaire and distributed it among 114 participants. We aimed at identifying their concerns on *collateral information collection, lack of transparency (notification) and not being asked for their approval (consent)*. Our user opinion study serves as evidence that the majority of participants are indeed concerned. On top of that, their concern is bidirectional: they would like to both notify their friends and be notified by their friends when installing apps enabling collateral information collection. They would also like to be able to restrict which attributes are shared that way.

- *What is the likelihood that an installed app enables collateral information collection?* To answer this question, we estimated the probability of that event to occur in the Facebook ecosystem. We show that the likelihood of collateral information collection for a user depends on the number of friends a user has and the popularity of apps (number of active users). We run simulations with a network size of 1 million users and investigated various network topologies and app adoption models tailored to the Facebook ecosystem. Based on the results obtained, we demonstrate that for an average user (≈ 200 friends) this likelihood is greater than 80% for popular apps such as TripAdvisor.

- *How significant is the collateral information collection?* To answer how much information is collected, we quantified the amount of attributes collected by apps that enable collateral information collection. The quantification depends on popular apps available on the Facebook ecosystem, and we estimated the number of attributes that each app is collecting. We also considered that several apps belong to the same app providers. To investigate that, we developed a mathematical model that also takes into account the access control mechanisms that Facebook provides to its users. For our calculations, we used the Appinspect dataset [15] which is a real world snapshot of the apps on Facebook.

- *Under the data protection legislation, is collateral information collection considered a risk for the protection of the personal data of Facebook users?* We investigated this research question under the prism of the GDPR [113]. First, our analysis clarifies who the data controllers and data processors are. Second, it scrutinises whether collateral information collection is an adequate practice from a data protection point of view. Finally, it identifies who is merely accountable.

- *How can we mitigate collateral information collection?* For this end, we analysed various countermeasures as follows. First, we propose the use of Transparency Enhancing Technologies (TETs) aiming at raising user awareness and helping the users make informed decisions. For this purpose, we outline a privacy dashboard extension that implements privacy scoring computations for enhancing transparency towards collateral information collection. Furthermore, we discuss

alternative solutions highlighting other countermeasures such as notification and access control solutions, cryptographic tools and a legal application assessment framework.

The rest of this chapter is organised as follows (see Fig. 4.3). Sect. 4.2 characterises collateral information collection and presents our user opinion study. Sect. 4.3 constructs a mathematical model of collateral information collection and estimates the likelihood of a user being affected. Sect. 4.4 extends the model and quantifies the significance of collateral information collection illustrated by a case study of popular apps. Sect. 4.5 conducts a legal analysis focusing on the GDPR. Sect. 4.6 outlines potential solutions to mitigate collateral information collection including the high-level design for a customised privacy dashboard. Finally, Sect. 4.7 provides the summary of this chapter.

4.2 Relevance of collateral information collection on Facebook: evidence of user concerns

In this section we investigate the research question: *are users concerned that apps installed by their friends can collect their profile data on Facebook?* To answer this question, we designed a questionnaire and distributed it to 114 participants. We first characterise collateral information collection with three shortcomings, which we refer to as the three pillars of collateral information collection. These three pillars establish the relevance of collateral information collection as a key challenge on Facebook privacy. Next we present the users' concerns that serve as evidence of the relevance of collateral information collection. Malhotra et al. [203] identified three essential dimensions of concern: "inappropriate collection of personal information", "lack of awareness of privacy practices" and "lack of control over personal information". As a followup, our questionnaire covers concerns about the *collateral information collection, lack of transparency (notification) and not being asked for their approval (consent).*

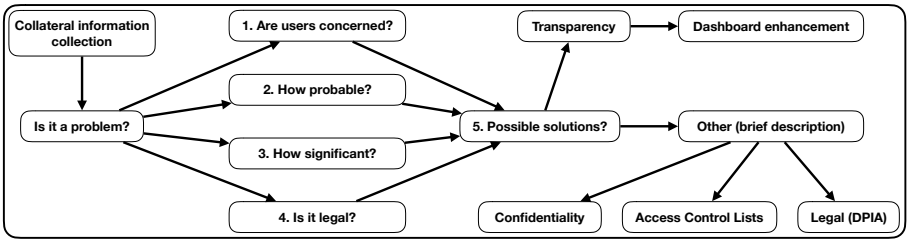


Figure 4.3: Paper contribution to collateral information collection on Facebook.

4.2.1 The three pillars of collateral information collection on Facebook

Lack of transparency, false sense of privacy and over-sharing. The primary problem of collateral information collection is that users are not aware of that their data is being collected [290, 38]. Awareness has been reported as a key issue for Facebook users [35, 290, 251, 125, 45]. If a friend of a user installs an application, it is unclear whether the friend is aware of the data collection by this application. This lack of transparency makes it impossible for the user and the friends to give informed consent or to influence the decision of use of the personal data of users in any way.

Due to the lack of transparency, users may assume that their data are only visible to their friends. In such cases, users might be less restrictive with information sharing [184]. Hence, in an environment where apps installed by the user's friends and can collect information from the user themselves, the average user may experience a false sense of privacy. Such a gap between a Facebook user's general privacy expectation and reality has been well-investigated by Liu et al. [195] and Madejsk et al. [200]. The added potential collateral information collection stemming from interdependent privacy can further widen this gap.

Lack of control by Facebook users: (un)informed consent and its (in)transitivity. The lack of control is one of the fundamental privacy problems in OSNs [7, 326, 290, 45, 322]. However, in the case of collateral information collection the lack of transparency and the inability to prevent such information collection without un-friending the Facebook friend, who has installed the app, limits the user even further in having control. For some users, the price for protecting their privacy might become too high.

In particular, the lack of *informed consent* is often an integral part of privacy loss on the web, the OSNs [187] and on Facebook considering the collateral information collection. Due to lack of transparency, informed consent cannot be given by a user when one of their friends installs an app that can collect information about the user themselves. One can argue that the friends that installed the app might be informed about such data collection. Hence they can decide for the user. In this case, even assuming that the friend is aware and privacy savvy, the question arises whether consent is transitive? Transitivity of consent and preferences has been the subject of debate in the literature [232]. One can assume that the user is giving indirect consent to their friends to share their data by making their personal profile attributes accessible to them. This stance can be criticised on (at least) two grounds. First, the user does not know whether their friend installed an app collecting such data. Hence, their assumed indirect consent cannot be informed. Second, the default settings on Facebook are clearly pro-sharing

when it comes to the collateral information collection; also, Facebook users often leave the default settings unchanged [195, 43].

Facebook friends are not (always) real friends. Even if we assume that privacy decision-making can be transitive in some cases among friends, there have been reports in the literature that the relationship between Facebook friends cannot be equated to real-life friendships, i.e., "latent ties" [157, 81]. Therefore, it is debatable whether the user trusts their Facebook friends to make the right (privacy) decision on their behalf.

4.2.2 User opinion study

To investigate the research question: "are users concerned about collateral information collection?", we conducted an online questionnaire. We explored the users' concerns about the disclosure of personal data by Facebook apps installed by the friends of a user, and we investigated the users' concerns about un-consented information collection on Facebook. We observed that the participants' opinion about collateral information collection, can be characterised as remarkably concerned in general and in particular when the information collection is un-consented. Furthermore, the majority of users prefer to take action to prevent the collateral information collection.

Methodology. After an introduction, our questionnaire consisted of four main parts (see A.1). First, we asked users about their concerns on *collateral information collection*. We assessed users' standpoints and concerns about default privacy settings and the lack of notification for indirect and un-consented information collection. This assessment is necessary to be able to differentiate users who are concerned independent of their intentions to take actions against such practices. The second part explores which attributes users were more concerned about. We investigated the type of personal data on Facebook users find most sensitive. The third part is twofold: 1) whether users want to be notified when their friends' apps can collect their personal data or when their installed apps can collect personal data from their friends; 2) which actions users prefer to take in such cases. Users replied the questions by marking their responses on a scale of 1 to 5 where 1 stands for "not concerned at" all and 5 stands for "extremely concerned" [217, 206]; we also provided a text field where necessary. The fourth part of the questionnaire collects demographics and information regarding the participants' and the use of Facebook apps.

Participants. Our response pool consisted of 114 participants. Participants were recruited from the authors' direct and extended friend circles (including mostly, but not

only, Facebook friends). A large proportion of participants are aged between 20 and 35 and are well educated.

Survey design limitations. Our response pool is a convenience sample of the Facebook users. As such a sample is usually limited in size and diversity, we do not extrapolate our findings to the general population of Facebook and do not lean on quantitative results from the user opinion study. While a full-scale quantitative survey might constitute important future work (see Sect. 7.1), we would like to point out that notable previous research works on Facebook users have been limited to student populations. Acquisti and Grossklags [7] used a sample of 294 were more than 89% were undergraduate and graduate college students; whereas the 89 respondents of Boyd and Hargittai [43] were fist-year university students.

4.2.3 Results

For the first part, we observe that for all four statements users show concern (see Fig. 4.4). For instance, 66% (i.e., $37+38/114 \approx 0.66$) of users are at least very concerned about the default privacy setting of Facebook that allows the collateral information collection. Similarly, 77% (i.e., $38+49/114 \approx 0.77$) of users are at least very concerned about not being notified when their friends enable collateral information collection and 67% (i.e., $43+33/114 \approx 0.67$) for not being notified when one of the user's own apps can collect their friends' information. Finally, 81% (i.e., $32+60/114 \approx 0.81$) of users are at least very concerned about the collateral information collection and the lack of their approval. Note that Golbeck and Mauriello [139] have investigated how informed users are regarding the privacy risks of using Facebook apps. Their findings show that users do not always comprehend what type of data is collected by apps even when they have installed the app themselves. Therefore, it is reasonable to assume an incomplete understanding of apps installed by their friends, which is in line with our results.

For the second part of our questionnaire (see Fig. 4.5), we found that although users are concerned about a number of attributes, their concern is relatively subjective and differs between users. However, it is noteworthy that certain attributes clearly stand out and have been marked as more concerned about than others by a large proportion of the participants. For example, most of the users identify photos (84% are at least very concerned), videos (79%), their current location (76%), and family and relationships (54%) as profile attributes that participants are concerned the most about. The profile attributes that participants are least concerned about are proved to be birthday and relationship status. Note that the level of concern about the attributes is likely to depend on the context. For example, although a birthday attribute might seem harmless on

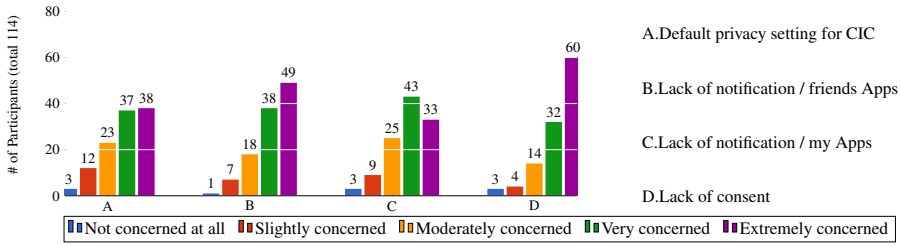


Figure 4.4: Results for the first part of the questionnaire where we asked participants about their opinions on four statements regarding default settings, lack of notification (for friends and for the user themselves), and lack of consent for the Collateral Information Collection (CIC).

its own, participants might feel different if a weather app would be collecting this information, or when a flashlight app requests the location of users [124].

For the third part of our questionnaire, we asked participants whether they want to be notified and also take action for the collateral information collection. Concerning notification (see Fig. 4.6), we identify that 77% of users always want to be notified when friends' apps can collect their personal data, 22% only want to be notified in particular cases, while only about 1% do not want to be notified at all. Moreover, 69% of users always want to be notified when their apps are collecting information from their friends, 27% in particular cases and only about 1% not at all. We observe that users are

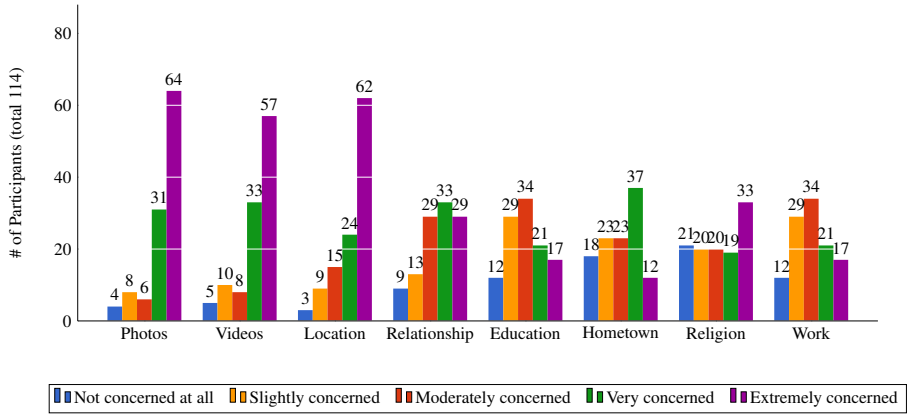


Figure 4.5: Total number of apps and app providers requesting collateral information collection of sensitive attributes (per attribute).

also seriously concerned about harming their friends' privacy. This corroborates the finding on attested shreds of evidence concerning other-regarding preferences [252, 65]. Concerning notification, there exist tools that can be very useful to enhance privacy awareness for un-consented data collection. Note that Golbeck et al. [139] have shown that the privacy awareness of users can be changed significantly through educational methods.

When participants were asked which actions (see Fig. 4.7) they would take if they are

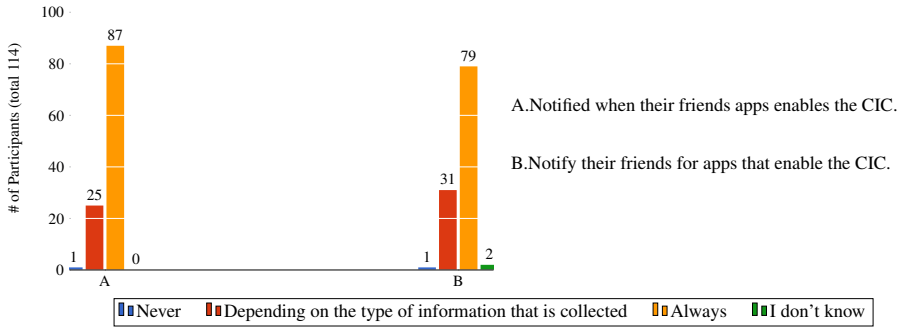


Figure 4.6: Number of participants who want to a) be notified by their friends b) notify their friends when installing apps enabling CIC.

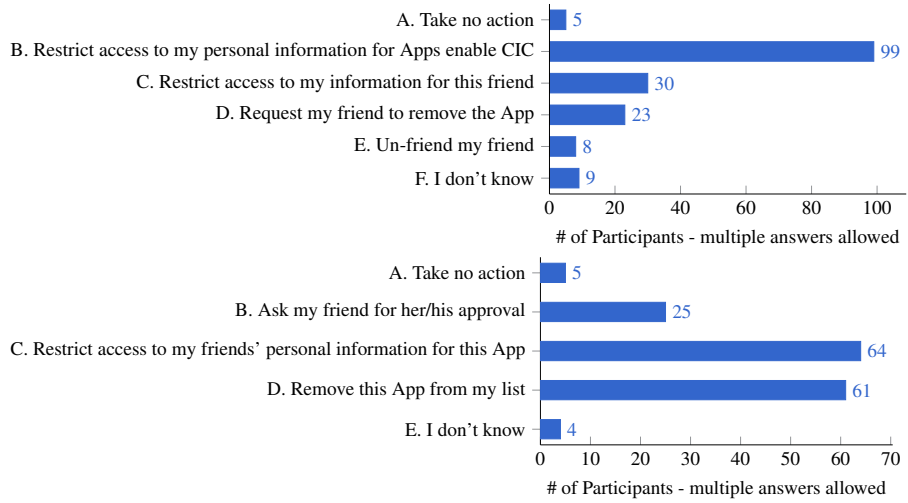


Figure 4.7: Number of participants preferring to take a given action when a) their friends install an app, b) they install an app enabling CIC.

notified that their friends' apps are about to collect their information (multiple answers allowed), 99 out of 114 participants answered that they would restrict access to their personal data while 8 participants answered that they would un-friend their Facebook friend. Only 5 participants answered that they would take no action. We emphasise that the reaction of a user may strongly depend on the relationship between the user and their friends. When participants were asked what action they would take if they are notified that one of their apps is about to collect their friends' information (multiple answers allowed), 64 out of 114 replied that they would restrict access to their friends' personal information for this app. Only 5 out of 114 answered that they would take no action. The answers to the questions in the third part help to confirm that the answers of our participants in the first part were not due to salience bias; participants who were concerned in the first part about not being notified for the *collateral information collection* replied that they also want to take an action in the third part.

The last part of our questionnaire collected demographics and statistics about Facebook and app usage. Participants were between 16 and 53 years old with an average age of 29 years. They have had their Facebook accounts for between 6 months and 10 years. Moreover, 69% of our participants have installed an app at least once, and among those 87% have installed 1 or 2 apps in the last six months. 54% of the participants were female, 42% male while 4% preferred not to disclose their gender. Participants varied greatly in their number of friends, from 10 to 1,000. The 51% of participants changed their privacy settings on Facebook; 79% restricted who could see their profile information, 41% who could see them in searches, and 35% who can collect their information through friends' apps (multiple answers were allowed). Considering the participants that have had not change their privacy settings, one explanation could be that privacy settings on Facebook are constantly changing and following these changes might be cumbersome [323]. Interestingly, among users who already took an action by restricting their permissions to their friends' apps, 90% choose to be notified too. Furthermore, 82% of our participants is pursuing or has obtained a higher education degree, where 55% had an IT background based on personal interest and 44% through higher education.

We conclude from our questionnaire that there is indeed evidence that users are concerned about collateral information collection. Our participants' concern is bidirectional, meaning that the large majority of them prefer to be notified and even take actions, whether this occurs from their friends or their own side, to prevent collateral information collection. While we do not want to generalise our findings to all Facebook users, we argue that these findings justify that collateral information collection is a privacy issue on Facebook, and thus, it merits a deeper investigation.[1]

[1] Results of our Facebook questionnaire study: `http://iraklissymeonidis.info/fbapps/Survey/survey.html`

4.3 Likelihood of collateral information collection

In this section, we investigate the research question: *what is the likelihood that an installed app enables collateral information collection?* To answer this question, we estimate the likelihood of at least one Facebook friend of a user installing an app that enables collateral information collection. In order to estimate this likelihood, we build a model incorporating Facebook friends, the application space and the probabilities of installing apps. Then, we conduct two case studies to estimate this likelihood. Our first case study assumes a uniform app adoption model, which takes into account the popularity of an app and the Facebook users. Our second case considers a more realistic, non-uniform app adoption model, alongside different network topologies tailored to the Facebook ecosystem. We have instantiated our numerical estimations with the AppInspect dataset [15].

4.3.1 Basic likelihood model

Let an OSN with k users and the corresponding set be denoted by $\mathcal{U}$, i.e., $\mathcal{U} = \{u_1, \ldots, u_k\}$. The user is denoted by u, with $u \in \mathcal{U}$. Let $f \in \mathsf{F}^u$ be a friend of u and $\mathsf{F}^u \subseteq \mathcal{U}$ the set of u's friends. Moreover, let A_j be an app and $\mathcal{L}$ the set of all A_js that are offered by the OSN to every u_i, and s the size of the set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Moreover, let AU_j be the number of users who have installed A_j. For our likelihood estimation, we consider the number of Monthly Active Users (MAU) to represent the number of active users. For instance, when we conducted this research Facebook had $k \approx 2 \times 10^9$ users (i.e., MAU) [277] and more than $s > 25{,}000$ apps [15] (January 2017).

To estimate the likelihood that u's personal data can be collected by A_j, installed by f, we compute the probability of at least one arbitrary f installing any available A_j. Let Q^f be the probability of f installing A_j that enables collateral information collection. For all the friends of u (i.e., F^u) the probability of not installing any A_j is the product of probabilities for each f (assuming that these probabilities are independent). Let Ω be the probability of at least one of u's friends installing A_j (regardless if u has installed A_j), i.e.,

$$\Omega = 1 - \prod_{f \in \mathsf{F}^u} (1 - Q^f) \ . \tag{4.1}$$

To estimate the likelihood Ω, we compute the probability of a friend of a user installing an app using two different app adoption models (uniform and a non-uniform) as follows.

4.3.2 Case study 1 – uniform distribution

Each f decides whether to install A_j without considering any app adoption signals from other users. The probability of at least one friend of u installing A_j is uniformly distributed among u's friends and is equal to $1-Q$. Note that $Q = Q^{f_1} = \cdots = Q^{f_{k'}}$ for uniform distribution with $1 \leq k' \leq k$. The probability Q, is then computed as the number of all users who installed the app, i.e., AU_j, divided by the number of users of the OSN, i.e., the cardinality of $\mathcal{U}$ (with regard to active users):

$$Q = \frac{AU_j}{|\mathcal{U}|} \quad . \tag{4.2}$$

We used the publicly available dataset provided by Huber et al. [15, 165] to extract the range of MAU for apps that enable collateral information collection. The dataset consists of 16,808 Facebook apps from the period between 2012 and 2014. It contains the application name, ID, number of active users (daily, weekly and monthly) and the requested permissions. To illustrate the influence of different values of MAUs on Ω, we only consider the upper tier of apps, i.e., over 500,000 MAU; while the most popular app that collects friends' data has 10,000,000 MAU, therefore $5 \cdot 10^5 \leq AU_j \leq 1 \cdot 10^7$. To cover most users, we assume the number of friends for a given u (i.e., $|\mathsf{F}^u|$) to be between 0 and 1,000. Finally, we estimate the population of Facebook to be $1.1 \cdot 10^9$ MAU for the period of 2012 to 2014 [277].

For A_js with $AU_j \geq 5 \cdot 10^6$, the probability Ω grows steeply with the average number of friends (see Fig. 4.8). For a median of 200 friends the probability Ω is larger than 0.6. For a user with 300 friends and more, the probability Ω exceeds 0.8. Note that most of Facebook users have more than 200 friends [314]. From Eqns. (4.1) and (4.2) it is clear that Ω depends strongly on AU_j. For instance, our most popular app TripAdvisor [310] has approximately $1 \cdot 10^7$ MAU (i.e., $AU_j \approx 1 \cdot 10^7$). Assuming that on average a user has 200 friends [314] (i.e., $|\mathsf{F}^u| \approx 200$), and considering $\mathcal{U} = 1.1 \cdot 10^9$ (i.e., the population of Facebook), we estimate that the probability of at least one of u's friends installing TripAdvisor is larger than 78% ($\Omega \geq 0.78$).

To further elaborate our results, we empirically computed the probability Ω on a synthetically generated OSN network graph. We used the Erdős-Rényi [102] model (ER), to create a graph with $1 \cdot 10^6$ nodes and a mean node degree of $d = 200$. To emulate the theoretical results depicted by Eqn. (4.2), the simulation assigned app installations randomly following a uniform distribution. Therefore, each node in the graph has the same probability to install the app, i.e., $Q = \frac{AU_j}{|\mathcal{U}|}$.

The Fig. 4.9a shows the probabilities Ω observed in the simulation. Note that, the results are consistent with Fig. 4.8, with Ω increasing with the degree of the nodes. The plot is noisy for degrees less than 170 and higher than 230, since there are just a

few nodes with these degrees in the simulated graph, as expected for an ER graph (see Fig. 4.9b). Indeed, ER graphs are created by including each edge uniformly at random with a certain probability Q. Therefore, the resulting degree distribution is binomial with an expected value of 200. Fig. 4.9b shows the theoretical degree distribution of an ER graph with $k = 1 \cdot 10^6$ and $d = 200$. The standard deviation is very low (14.14), thus most of the nodes have a degree close to 200.

The probability of a user having at least one friend with the app installed is computed assuming uniformity. Both Eqn. (4.2) and the simulation (see Fig. 4.9a) are based on the assumption that the probability of a friend installing A_j is equal to the mean app adoption rate of the network, i.e., $Q^f = Q$ for all friends of u. Case study 2 deals with the analysis of Ω when the uniformity assumption is relaxed.

4.3.3 Case study 2 – non-uniform distribution

Realistic OSN do not usually conform to the uniformity assumption, thus assuming $Q = \frac{AU_j}{|\mathcal{U}|} = Q^{f_1} = \cdots = Q^{f_{k'}}$ where $1 \leq k' \leq k$ may not be realistic. App adoption has been proclaimed to be affected by different signals [249], which in turn may be affected by the underlying network structure. Research in OSNs has reported node degree distributions following a power law and clustering coefficients much higher than in a random network [127, 334, 218]. We have resorted to simulations in order to introduce all these factors into the estimations of the probability Ω.

Each simulation uses two models: one to generate the synthetic network in which the probability under study is computed (i.e., network topology), and the other to decide which users of the network install the application (i.e., app adoption model). Regarding the *network topology*, we have considered two different models: Barabási-Albert [11] (BA); that generates networks with a power-law degree distribution; and Watts-Strogatz [329] (WS), that is able to generate small-world networks with high clustering and small diameter. Regarding *app adoption*, two different models have been implemented: uniform (unif), where all users install an app with the same probability (and each installation is independent from other installations and the underlying network topology); and preferential (prop), where the probability of a user installing an app is proportional to the number of friends that have already installed the app owing to local network effects [283].

Therefore, the *configuration* of a simulation is defined by the network model, the app adoption model, and a set of parameters that configure both models: the number of nodes denoted by k, expected mean node degree d, and a fraction of users installing the app denoted by e. The number of nodes k determines the network size for every model. The expected mean node degree d (and its relation to k) is used to adjust the additional specific parameters of the underlying network models. Finally, the fraction of users

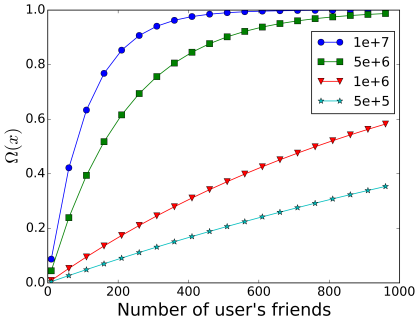
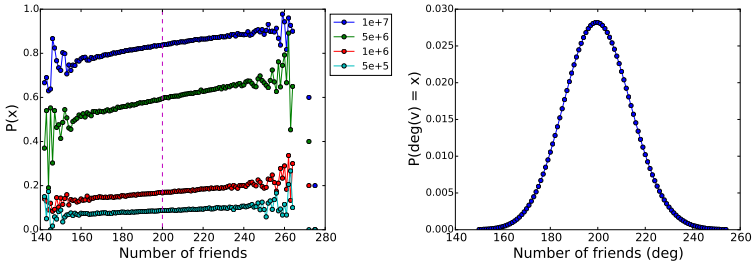


Figure 4.8: Likelihood of collateral information collection based on real data [15] (per MAU).



(a) Probability of being affected

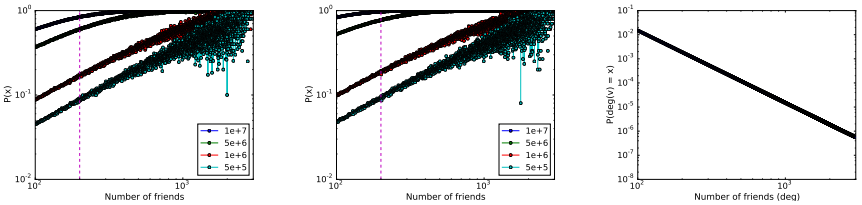
(b) Degree distribution

Figure 4.9: Simulation results on an ER graph with $k = 1{,}000{,}000$ and $d = 200$.

in the network installing the app e is used by the app adoption model to compute the actual user base.

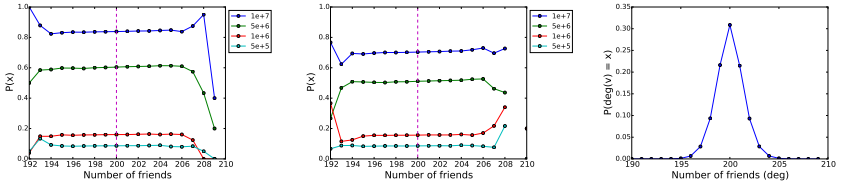
We performed simulations with *configurations* using all possible pairs of network and app adoption models, for network sizes $k \in [10^4, 10^5, 10^6]$, mean degree $d \in [200]$ and $e = [5 \cdot 10^5/1.1 \cdot 10^9, 10^6/1.1 \cdot 10^9, 5 \cdot 10^6/1.1 \cdot 10^9, \cdot 10^7/1.1 \cdot 10^9]$. The Fig. 4.10 and Fig. 4.11 present the results for $k = 10^6$ and all tested e values. Owing to computational limitations, the largest simulated network size is $k = 1 \cdot 10^6$; however, the trends are consistent across all network sizes. We have omitted the results for smaller k for the sake of brevity.

The Fig. 4.10 shows the results of the simulation using the BA model to generate the graph, together with the theoretical degree distribution of those graphs. The BA model generates graphs with a few very high degree nodes (known as hubs) and lots of



(a) Uniform app adoption (b) Preferential app adoption (c) Degree distribution

Figure 4.10: Simulation results on a BA graph with $k = 1{,}000{,}000$ and $d = 200$.



(a) Uniform app adoption (b) Preferential app adoption (c) Degree distribution

Figure 4.11: Simulation results on a WS graph with $k = 1{,}000{,}000$ and $d = 200$.

low degree nodes, as shown in the theoretical degree distribution of a BA graph (see Fig. 4.10c). Regarding the simulation results for the two app adoption models, there are small differences between the observed Ω values. When combining networks generated with the BA model with a uniform app adoption model (see Fig. 4.10a), the probability for a hub to install an app is the same as for any other node. To the contrary, when combining BA with the prop app adoption model (see Fig. 4.10b), hubs have a higher probability of installing the app than non-hubs, since having a higher degree makes them more likely to have (more) friends with the app installed. As a consequence, each installation affects more users on average, and thus, Ω increases.

The Fig. 4.11 shows the results of the simulation using the WS model, together with the theoretical degree distribution of those graphs. The WS model generates highly clustered networks[2] with very low variability in node degrees. Indeed and as it is illustrated in Fig. 4.11c, the vast majority of nodes will have a degree between 195 and 205. As a consequence, Fig. 4.11a and Fig. 4.11b are noisy for degrees outside this interval. The simulation results show, on the one hand, that Ω is about the same for all nodes in the graph, which is consistent with the fact that the degrees of the nodes are similar. On the other hand, the results also show a significant difference between using the unif (see Fig. 4.11a) and the prop (see Fig. 4.11b) app adoption models.

[2]The expected clustering coefficient can be adjusted with the rewiring probability parameter.

This is caused by the strong clustering coefficient exhibited by the network. With `unif` app adoption, each node installs the app with the same probability. On the contrary, with `prop` app adoption, when an app is installed by a member of a community, it propagates to be adopted also by other members. However, each new installation inside the same community implies only a small increase in the overall Ω. That is because most of the users affected by the installation were already affected by installations from other members of the community.

In summary, irrespective of app adoption models, the likelihood that a friend of a given user installs an app enabling collateral information collection is significant. If the app is popular, i.e., has millions of users, it is *highly likely* that a user is affected by collateral information collection from one of her friends.

4.4 Significance of collateral information collection

In this section, we answer the research question: *how significant is the collateral information collection?* To estimate how much information is collected, we build a model including friends, profile attributes, and Facebook access control, i.e., privacy settings and third-party application (app) permissions. This model can be used to compute the amount of the user's attributes that can be collected by apps (and thus third-party application providers (app providers)) when installed by the friends of a user. We investigate different ways of acquiring data: direct collection from the users themselves, indirect and exclusive collection through the friends of a user. To instantiate our model, we use several snapshots of the most popular apps on Facebook using the AppInspect dataset [15].

4.4.1 Basic collateral information collection model

Users and users' friends. Each user u_i in an OSN (i.e., $u_i \in \mathcal{U}$) has a personal profile where they can store, update, delete and administer their personal data [42]. A given u_i's profile consists of attributes a_i such as name, email, birthday and hometown. We denote the set of attributes of a u_i's profile as $\mathcal{T}$ and n as the size of $\mathcal{T}$, i.e., $\mathcal{T} = \{a_1, \ldots, a_n\}$. For instance, Facebook currently operates with a set of $n = 25$ profile attributes. Note that we use the term attribute and profile item interchangeably.

For any user u_i in $\mathcal{U}$, we consider u to be a user under consideration and $f_i \in \mathsf{F}^u$ one of their friends. Let F^{u*} be the union of u's friends and the u itself, or $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$, and let f^* be an element of F^{u*}, i.e., $f^* \in \mathsf{F}^{u*}$. Clearly $\mathsf{F}^u \cap \{u\} = \emptyset$, as u is not a friend of u. For instance, $\mathsf{F}^{u*} = \{u, f_1, \ldots, f_{k'}\}$ describes a user u and her k' friends, where $1 \leq k' \leq k$. Note that Table 4.2 list the notations used in this chapter.

Applications and application providers. Let $\mathcal{L}$ be the set of apps an app provider can offer to every u_i in an OSN and s the size of this set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Let A_j, for $1 \leq j \leq s$, be the set of attributes that each A_j can collect, i.e., $\mathsf{A}_j \subseteq \mathcal{T}$. Each A_j is owned and managed by an app provider denoted by P_j. The set of A_js that belong to P_j

Table 4.2: Notation for the collateral information collection model.

Notation	Description
$\mathcal{U} = \{u_1, \ldots, u_k\}$	Set of k users u_i in an Online Social Network (OSN).
$\mathsf{F}^{u*} = \{u, f_1, \ldots, f_{k'}\}$	Set of k' friends (i.e., $f_i \in \mathsf{F}^u$) and u themselves (i.e., the user under consideration), where $k' \leq k$, $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$ and $\{u\} \cap \mathsf{F}^u = \emptyset$.
$\mathcal{T} = \{a_1, \ldots, a_n\}$	Set of n attributes a_i of u's profile.
A_j / $A_j^{u,f}$ / A_j^{u,F^u}	An app j / an app j installed by: a user f / all users in F^u which can collect u's profile attributes.
$\mathsf{A}_j^{u,f}$ / $\mathsf{A}_j^{u,\mathsf{F}^u}$	Set of a_is for each A_j installed by: a user f / all users in F^u which can collect attributes of u's profile.
$\mathcal{L} = \{A_1, \ldots, A_s\}$	Set of s apps A_j hosted by an OSN.
$\mathsf{L}^{u,f}$ / $\mathsf{L}^{u,\mathsf{F}^u}$	Set of A_js installed by: a user f / all users in F^u, that can collect attributes of u's profile.
AU_j	The number of Monthly Active Users (MAU) of an A_j.
P_j / $P_j^{u,f}$ / P_j^{u,F^u}	An app provider P_j offering a set of A_js / a set of A_js installed by user f / a set of A_js installed by all users in F^u offered by P_j that can collect u's profile attributes.
$\mathsf{P}_j^{u,f}$ / $\mathsf{P}_j^{u,\mathsf{F}^u}$	Set of a_is all A_js installed by, a user f / all users in F^u and belong to the P_j that can collect u's profile attributes.
$\mathcal{AP} = \{P_1 \ldots P_m\}$	Set of m app providers P_js hosted by an OSN.
$\mathsf{AP}^{u,f}$ /$\mathsf{AP}^{u,\mathsf{F}^u}$	Set of P_js whose A_js: installed by a user f / all users in F^u which can collect u's profile attributes.
$\mathsf{C}_{PM,A_j}^{u,f}$ / $\mathsf{C}_{PM,A_j}^{u,\mathsf{F}^{u*}}$	Set of permissions r_i an A_j can request and, a user f (i.e., $f \in \mathsf{F}^{u*}$) / all users in F^{u*} (i.e., $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$), accept(s) to collect u's profile attributes.
$\mathsf{C}_{PV}^{u,f}$ / $\mathsf{C}_{PV}^{u,\mathsf{F}^u}$	Set of privacy settings that an OSN offers and allows: a user f / all users in F^u, to access u's profile attributes.
$\mathsf{C}_{A_j}^{u,f}$ / $\mathsf{C}_{A_j}^{u,\mathsf{F}^u}$	Set of access settings (i.e., a combination of permissions and privacy settings) granted and accepted by u and: a user f / / all users in F^u to access and collect u's profile attributes by an A_j.
$\Pi_{A_j}^{u,\mathsf{F}^u}$ / $\Pi_{P_j}^{u,\mathsf{F}^u}$	Set of a_is for each, A_j / P_j, installed by all users in F^u that can collect u's profile attributes.
$\Delta^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ / $\Delta^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$	Set of a_is for each, A_j / P_j, *exclusively* installed by all users in F^u (and not through the user themselves, i.e., $\mathsf{A}_j'^u \cap \mathsf{A}_j^{u,\mathsf{F}^u}$) that can collect u's profile attributes.
d / e	The mean node degree / the fraction of users that installed an app, i.e., $\frac{AU}{\lvert\mathcal{U}\rvert}$

is denoted by P_j, i.e., $\mathsf{P}_j \subseteq \mathcal{L}$. The set of all P_js is denoted by $\mathcal{AP}$ and m the size of the set, i.e., $\mathcal{AP} = \{P_1, \ldots, P_m\}$. From our analysis, we identified $s = 16,808$ apps and $m = 2,055$ app providers on Facebook indicating that a P_j can have more than one A_j, i.e., $\mathsf{P}_j = \{A_1, \ldots, A_{s\prime}\}$ with $1 \leq s\prime \leq 160$ [15].

Collateral information collection by an application A_j. When A_j is installed by f (i.e., $f \in \mathsf{F}^u$), a set of attributes a_i can be collected from u's profile. We denote by A_j^{u,F^u} an A_j that users in F^u installed and as $\mathsf{A}_j^{u,\mathsf{F}^u}$ the set of attributes a_i that A_j^{u,F^u} can collect from u's profile. Clearly, $\mathsf{A}_j^{u,\mathsf{F}^u} \subseteq \mathsf{A}_j \subseteq \mathcal{T}$. The set of all A_j^{u,F^u}s installed by the users in F^u is denoted by $\mathsf{L}^{u,\mathsf{F}^u}$. Clearly, $\mathsf{L}^{u,\mathsf{F}^u} \subseteq \mathcal{L}$.

We denote by $\vec{a}_i$ a vector of size $n \times 1$ which corresponds to a_i, i.e.,

$$\vec{a}_i = [\overset{1}{0} \ldots \overset{i}{0}1\overset{}{0} \ldots \overset{n}{0}] \ .$$

Moreover, we consider $\vec{A}_j^{u,\mathsf{F}^u}$ a vector of length n, which corresponds to $\mathsf{A}_j^{u,\mathsf{F}^u}$, i.e.,

$$\vec{A}_j^{u,\mathsf{F}^u} = \bigvee_{a_i \in A_j^{u,\mathsf{F}^u}} \vec{a}_i \ := \vec{A}_j^{u,\mathsf{F}^u}[i] \ = \begin{cases} 1 & \text{if } a_i \in \mathsf{A}_j^{u,\mathsf{F}^u} \\ 0 & \text{if } a_i \notin \mathsf{A}_j^{u,\mathsf{F}^u} \end{cases}, \tag{4.3}$$

for $1 \leq i \leq n$ and $1 \leq j \leq s$.

Note that:

- $x \cup y = \begin{cases} z = 0 & \text{if } x = y = 0, \\ z = 1 & \text{otherwise.} \end{cases}$
- and $\vec{x} \vee \vec{y} = \vec{z}$ where $\vec{x}[i] \vee \vec{y}[i] = \vec{z}[i]$.

For instance, an $\mathsf{A}_j^{u,\mathsf{F}^u} = \{a_1, a_i, a_n\}$ is represented as $\vec{A}_j = \vec{a}_1 \vee \vec{a}_i \vee \vec{a}_n = [\overset{1}{1}0 \ldots \overset{i}{0}1\overset{}{0} \ldots \overset{n}{0}1]$. It represents the attributes that can be collected by A_j when is installed by f (i.e., the user's friend).

Collateral information collection by application provider P_j. We denote by $\mathsf{AP}^{u,\mathsf{F}^u}$ the set of app providers whose apps A_j^{u,F^u}s are installed by users in F^u and who

can collect attributes of user u's profile. Hence,

$$\mathsf{AP}^{u,\mathsf{F}^u} = \bigcup_{f \in \mathsf{F}^u} \mathsf{AP}^{u,f} \quad . \tag{4.4}$$

Each P_j^{u,F^u} consists of a set of A_j^{u,F^u}s denoted by $\mathsf{P}_j^{u,\mathsf{F}^u}$ which users in F^u installed. Each $\mathsf{P}_j^{u,\mathsf{F}^u}$ can collect attributes of u's profile. To identify which a_is can be collected by P_j we consider $\vec{P}_j^{u,\mathsf{F}^u}$ a vector of length n (i.e., $n \in \mathcal{T}$), which corresponds to $\mathsf{P}_j^{u,\mathsf{F}^u}$, i.e.,

$$\vec{P}_j^{u,\mathsf{F}^u} = \bigvee_{\substack{A \in \mathsf{P}_j^{u,f} \\ f \in \mathsf{F}^u}} \vec{A}^{u,f} = \bigvee_{A \in \mathsf{P}_j^{u,\mathsf{F}^u}} \vec{A}^{u,\mathsf{F}^u} \quad . \tag{4.5}$$

Note that: $\vec{P}_j^{u,\mathsf{F}^u} = \bigvee_{f \in \mathsf{F}^u} \vec{P}_j^{u,f} = (\vec{P}_j^{u,f_1} \vee \cdots \vee \vec{P}_j^{u,f_i})$.

The complexity of evaluating Eqn. (4.5) for all f in F^u is $O(n \times |\mathsf{P}_j^{u,\mathsf{F}^u}|)$.

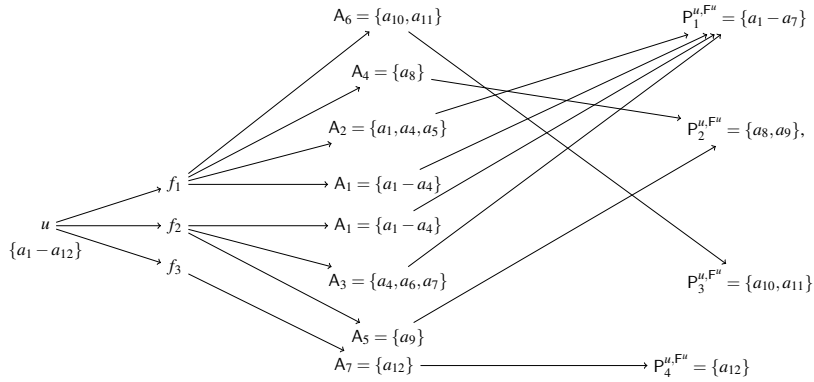


Figure 4.12: Example for collateral information collection while enabling profiling to P_j.

Example (see Fig. 6.2). Let $\mathsf{F}^u = \{f_1, f_2, f_3\}$ friends of u. The set of A_js that all $f \in \mathsf{F}^u$ installed is $\mathsf{L}^{u,\mathsf{F}^u} = \{A_1^{u,\mathsf{F}^u} \ldots A_7^{u,\mathsf{F}^u}\}$. The set of P_js for all A_js installed is described as $\mathsf{AP}^{u,\mathsf{F}^u} = \mathsf{AP}^{u,f_1} \cup \mathsf{AP}^{u,f_2} \cup \mathsf{AP}^{u,f_3} = \{\mathsf{P}_1^{u,f_1}, \mathsf{P}_2^{u,f_1}, \mathsf{P}_3^{u,f_1}\} \cup \{\mathsf{P}_1^{u,f_2}, \mathsf{P}_2^{u,f_2}\} \cup \{\mathsf{P}_4^{u,f_3}\} = \{(\mathsf{P}_1^{u,f_1} \cup \mathsf{P}_1^{u,f_2}), (\mathsf{P}_2^{u,f_1} \cup \mathsf{P}_2^{u,f_2}), \mathsf{P}_3^{u,f_1}, \mathsf{P}_4^{u,f_3}\}$. Each $\mathsf{P}_1^{u,\mathsf{F}^u} = \mathsf{P}_1^{u,f_1} \cup \mathsf{P}_1^{u,f_2} = \{(\mathsf{A}_1^{u,f_1} \cup \mathsf{A}_2^{u,f_1}) \cup (\mathsf{A}_1^{u,f_2} \cup \mathsf{A}_3^{u,f_2})\}$, $\mathsf{P}_2^{u,\mathsf{F}^u} = \mathsf{P}_2^{u,f_1} \cup \mathsf{P}_2^{u,f_2} = \{\mathsf{A}_4^{u,f_1} \cup \mathsf{A}_5^{u,f_2}\}$, $\mathsf{P}_3^{u,\mathsf{F}^u} = \{\mathsf{A}_6^{u,f_1}\}$ and $\mathsf{P}_4^{u,\mathsf{F}^u} = \{\mathsf{A}_7^{u,f_3}\}$. Each A_j installed by u's friends can collect a set of

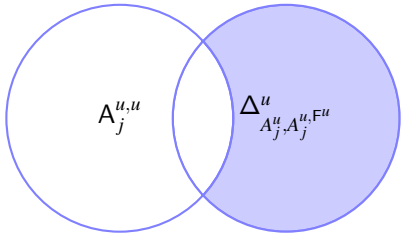


Figure 4.13: Exclusive collateral information collection of attributes by an application A_j.

a_i attributes from u's profile such that, $\mathsf{A}_1 = \{a_1, a_2, a_3, a_4\}$, $\mathsf{A}_2 = \{a_1, a_4, a_5\}$, $\mathsf{A}_3 = \{a_4, a_6, a_7\}$, $\mathsf{A}_4 = \{a_8\}$, $\mathsf{A}_5 = \{a_9\}$, $\mathsf{A}_6 = \{a_{10}, a_{11}\}$, $\mathsf{A}_7 = \{a_{12}\}$. The total collection of a_is for each P_j is $\mathsf{P}_1^{u,\mathsf{F}^u} = (\mathsf{A}_1^{u,f_1} \cup \mathsf{A}_2^{u,f_1}) \cup (\mathsf{A}_1^{u,f_2} \cup \mathsf{A}_3^{u,f_2}) = (\{a_1, a_2, a_3, a_4\} \cup \{a_1, a_4, a_5\}) \cup (\{a_1, a_2, a_3, a_4\} \cup \{a_4, a_6, a_7\}\}) = \{a_1 - a_7\}$, $\mathsf{P}_2^{u,\mathsf{F}^u} = \mathsf{A}_4^{u,f_1} \cup \mathsf{A}_5^{u,f_2} = \{a_8\} \cup \{a_9\} = \{a_8, a_9\}$, $\mathsf{P}_3^{u,\mathsf{F}^u} = \mathsf{A}_6^{u,f_1} = \{a_{10}, a_{11}\}$ and $\mathsf{P}_4^{u,\mathsf{F}^u} = \mathsf{A}_7^{u,f_1} = \{a_{12}\}$.

Exclusive collateral information collection by application A_j and application provider P_j. We denote by $\Delta^u_{A_j^u,A_j^{u,\mathsf{F}^u}}$ the set of a_i attributes that can be collected by A_j exclusively from u's friends (and not through the user themselves), i.e., $\mathsf{A}_j'^u \cap \mathsf{A}_j^{u,\mathsf{F}^u}$.

Let $\vec{\Delta}^u_{A_j^u,A_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\Delta^u_{A_j^u,A_j^{u,\mathsf{F}^u}}$ provides, where $n = |\mathcal{T}|$, i.e.,

$$\vec{\Delta}^u_{A_j^u,A_j^{u,\mathsf{F}^u}} = \vec{A}_j'^u \bigwedge \vec{A}_j^{u,\mathsf{F}^u} \quad . \tag{4.6}$$

Note that $\vec{x}' \wedge \vec{x} = [\overset{1}{0} \ldots \overset{n}{0}]$ and $\vec{x}' \vee \vec{x} = [\overset{1}{1} \ldots \overset{n}{1}]$. The complexity of evaluating Eqn. (4.6) for all $f \in \mathsf{F}^u$ is $O(n^4)$.

Similarly, we denote by $\Delta^u_{P_j^u,P_j^{u,\mathsf{F}^u}}$ the set of a_is that can be collected by A_js exclusively from u's friends and belong to P_j^{u,F^u}, and $\vec{\Delta}^u_{P_j^u,P_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\Delta^u_{P_j^u,P_j^{u,\mathsf{F}^u}}$ provides, i.e.,

$$\vec{\Delta}^u_{P_j^u,P_j^{u,\mathsf{F}^u}} = \vec{P}_j'^u \bigwedge \vec{P}_j^{u,\mathsf{F}^u} \quad . \tag{4.7}$$

The complexity of evaluating Eqn. (4.7) for all $f \in \mathsf{F}^u$ is $O(n^4 \times |\mathsf{P}_j^u| \times |\mathsf{P}_j^{u,\mathsf{F}^u}|)$.

4.4.2 Application permissions and user's privacy settings

To control the collateral information collection of A_j from a u's profile, *access control settings* are provided by the OSN. On Facebook the *access control settings* consist of permissions and privacy settings [122]. Permissions depend on the friend f, where $f \in \{u\} \cup \mathsf{F}^u$ and on the application A_j, as each f should accept the permissions that each A_j is requesting. Privacy settings also depend on the user u, as each u chooses with whom their profile information is shared.

Permissions and application A_j. Each A_j can request a set of permissions r_i to be accepted from a user f such as *user_emails*, *user_friends* and *friends_birthday*. We denote by $\mathcal{AC}$ the set of permissions an application A_j can request from a friend f (with cardinality n), i.e., $\mathcal{AC} = \{r_1, \ldots, r_n\}$, where $f \in \mathsf{F}^{u*}$. Moreover, we denote by $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$ the set of permissions an A_j can request, and F^{u*} accepts; thus A_j can collect attributes of u's profile, where $1 \leq j \leq s$. Clearly $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j} \subseteq \mathcal{AC}$.

We consider $\vec{r}_i$ a vector of length n which corresponds to r_i, for $1 \leq i \leq n$, i.e.,

$$\vec{r}_i = [\overset{1}{0} \ldots \overset{i}{0} 1 \overset{}{0} \ldots \overset{n}{0}] \ .$$

Moreover, we consider $\vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$ a vector of length n, which corresponds to $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$, i.e.,

$$\vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j} = \bigvee_{r \in \mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}} \vec{r} := \vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}[i] = \begin{cases} 1 & \text{if access is provided by } \mathsf{F}^{u*} \\ & \text{to } u\text{'s profile for a given } A_j, \\ 0 & \text{otherwise,} \end{cases} \tag{4.8}$$

for $1 \leq i \leq n$ and for $1 \leq j \leq s$.

Privacy settings and user u. Each u can allow a user f (i.e., $f \in \mathsf{F}^u$) such as a friend, friend of a friend or any user to access the attributes of u's profile. We denote by $\mathsf{C}^{u,\mathsf{F}^u}_{PV}$ the set of attributes of u's profile that u allows to access for F^u using the privacy settings of Facebook.

We consider $\vec{C}^{u,\mathsf{F}^u}_{PV}$ a vector of length n, which corresponds to $\mathsf{C}^{u,\mathsf{F}^u}_{PV}$, i.e.,

$$\vec{C}^{u,\mathsf{F}^u}_{PV} = \begin{cases} [\overset{1}{1} \ldots \overset{i}{1} \ldots \overset{n}{1}] & \text{if an access to } u\text{'s profile is provided by } u \text{ to } \mathsf{F}^u, \\ [\overset{1}{0} \ldots \overset{i}{0} \ldots \overset{n}{0}] & \text{otherwise} \ , \end{cases} \tag{4.9}$$

for $1 \leq i \leq n$.

Permissions vs. privacy settings. We denote as $\mathsf{C}_{A_j}^{u,\mathsf{F}^u}$ the set of access settings provided by u and F^u to u's profile for an A_j, as a combination of permissions (i.e., $\mathsf{C}_{PM,A_j}^{u,\mathsf{F}^{u*}}$) and privacy settings (i.e., $\mathsf{C}_{PV}^{u,\mathsf{F}^u}$).

We consider $\vec{C}_{A_j}^{u,\mathsf{F}^u}$ a vector of length n which correspond to $\mathsf{C}_{A_j}^{u,\mathsf{F}^u}$, i.e.,

$$\vec{C}_{A_j}^{u,\mathsf{F}^u} = \vec{C}_{PV}^{u,\mathsf{F}^u} \wedge \vec{C}_{PM,A_j}^{u,\mathsf{F}^{u*}} \ , \tag{4.10}$$

for $1 \leq j \leq s$.

- Remark: $a \wedge b = \begin{cases} 1 & if\ a = b = 1, \\ 0 & \text{otherwise}, \end{cases}$
- Extension: $\vec{a} \wedge \vec{b} = \vec{c}$ where $\vec{a}[i] \wedge \vec{b}[i] = \vec{c}[i]$.

The complexity of evaluating Eqn. (4.10) for all f in F^u is $\mathsf{O}(n^2)$.

Collateral information collection with permissions. We denote by $\Pi_{A_j}^{u,\mathsf{F}^u}$ and $\Pi_{P_j}^{u,\mathsf{F}^u}$ the set of attributes that can be collected by A_j and P_j respectively, for the accepted access settings to u's profile by u and F^u.

Let $\vec{\Pi}_{A_j}^{u,\mathsf{F}^u}$ be a vector of length n which $\Pi_{A_j}^{u,\mathsf{F}^u}$ provide, i.e.,

$$\vec{\Pi}_{A_j}^{u,\mathsf{F}^u} = \vec{A}_j^{u,\mathsf{F}^u} \wedge \vec{C}_{A_j}^{u,\mathsf{F}^u} \ . \tag{4.11}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (4.11) for all f in F^u is $O(n^3)$.

Let $\vec{\Pi}_{P_j}^{u,\mathsf{F}^u}$ be a vector of length n which $\Pi_{P_j}^{u,\mathsf{F}^u}$ provides, i.e.,

$$\vec{\Pi}_{P_j}^{u,\mathsf{F}^u} = \bigvee_{A_j \in \mathsf{P}_j^{u,\mathsf{F}^u}} (\vec{A}_j^{u,\mathsf{F}^u} \wedge \vec{C}_{A_j}^{u,\mathsf{F}^u}) \ . \tag{4.12}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (4.12) for all f in F^u is $O(n^3 \times |\mathsf{P}_j^{u,\mathsf{F}^u}|)$.

Exclusive collateral information collection with permissions. We denote by $\Delta^{u}_{A^{u}_{j},A^{u,\mathsf{F}^{u}}_{j}}$ and $\Delta^{u}_{P^{u}_{j},P^{u,\mathsf{F}^{u}}_{j}}$ the set of attributes that can be collected by A_j and P_j (respectively) exclusively from u's friends (and not through the user themselves, i.e., $\mathsf{A}'^{u}_{j} \cap \mathsf{A}^{u,\mathsf{F}^{u}}_{j}$), for the accepted access settings to u's profile by u and F^{u}.

Let $\vec{\Delta}^{u}_{A^{u}_{j},A^{u,\mathsf{F}^{u}}_{j}}$ be a vector of length n which $\Delta^{u}_{A^{u}_{j},A^{u,\mathsf{F}^{u}}_{j}}$ provides, i.e.,

$$\vec{\Delta}^{u}_{A^{u}_{j},A^{u,\mathsf{F}^{u}}_{j}} = (\vec{A}^{u}_{j} \wedge \vec{C}^{u}_{A_j})' \wedge (\vec{A}^{u,\mathsf{F}^{u}}_{j} \wedge \vec{C}_{A^{u,\mathsf{F}^{u}}_{j}}) \ , \tag{4.13}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (4.13) for all f in F^{u} is $O(n^{12})$ and $O(n^{12} \times |\mathsf{P}^{u}_{j}| \times |\mathsf{P}^{u,\mathsf{F}^{u}}_{j}|)$

Let $\vec{\Delta}^{u}_{P^{u}_{j},P^{u,\mathsf{F}^{u}}_{j}}$ be a vector of length n which $\Delta^{u}_{P^{u}_{j},P^{u,\mathsf{F}^{u}}_{j}}$ provides, i.e.,

$$\vec{\Delta}^{u}_{P^{u}_{j},P^{u,\mathsf{F}^{u}}_{j}} = \vec{\Pi}'^{u}_{j} \bigwedge \vec{\Pi}^{u,\mathsf{F}^{u}}_{j} \ . \tag{4.14}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (4.14) for all f in F^{u} is $O(n^{12} \times |\mathsf{P}^{u}_{j}| \times |\mathsf{P}^{u,\mathsf{F}^{u}}_{j}|)$.

4.4.3 Numerical study of collateral information collection: the case of Facebook apps

To illustrate the significance of collateral information collection, we extended our analysis to Facebook apps (i.e., A_js) and app providers (i.e., P_js) using the AppInspect dataset [15, 165]. For each A_j, apart from the application name and ID, the dataset provides us with the number of monthly active users, the requested permissions and the A_js each P_j owns. We compute and compare the proportion of attributes A_js and their respective P_js can collect: 1) through the user themselves (i.e., direct collection by apps and potential data fusion by app providers), 2) through the user and the user's friends combined (i.e., collateral information collection) and 3) exclusively through the user's friends (i.e., exclusive collateral information collection). Out of the $16,808$ apps in the dataset, $1,202$ enable *collateral information collection* corresponding to 7.15%. Out of these $1,202$, our analysis focuses on A_js and P_js that have more than $AU \geq 10,000$ MAU; there are 207 and 88, respectively.[3]

[3]Results of our numerical study on Facebook: `http://iraklissymeonidis.info/fbapps/fb_apps_statistics/index.html`.

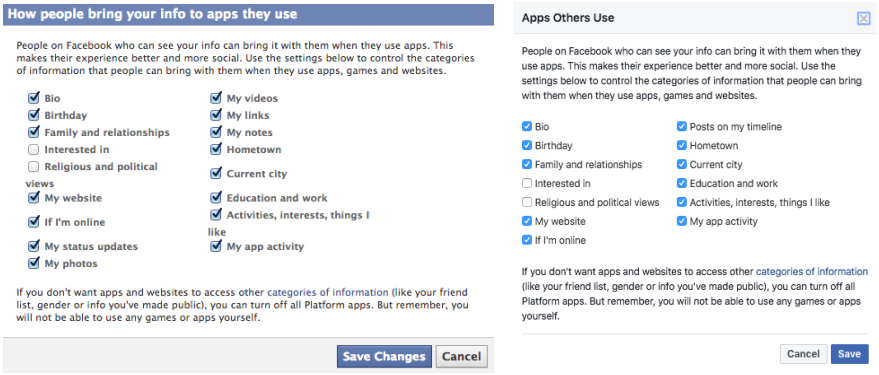


Figure 4.14: Default permissions settings for apps on (a) 2012 (left) and (b) 2017 (right) on Facebook.

Before delving into the details, we point our two limitations of our numerical study. First, the dataset does not contain information about the individual users of the listed apps, but only their aggregate numbers. Therefore, it is not possible to perform a per-user evaluation of the privacy loss due to collateral information collection. Instead, we had to resort to per app and per app provider analysis as it would impact an average user and an average friend of the user.

Second, the dataset does not contain information about the privacy settings of individual users. Therefore we are not able to factor in those when numerically evaluating the extent of the collateral information collection; we assume that all permission requests are granted. This is not a far-fetched assumption, as default privacy settings for apps have been and still are very much pro-sharing (see Fig. 4.14). In a nutshell, most attributes are accessible by default for apps installed by the friends of a user. Furthermore, the more recent "Apps Others use" dialogue (see Fig. 4.14 (b)), does not list photos, potentially indicating that these might not be accessible for apps through Facebook friends.

Direct information collection. First, we aimed to investigate the number of attributes that apps can collect: from users giving explicit consent when installing an app. Figure 4.15 (a) shows that more than half of the 207 apps collect a single attribute from the users. Other characteristic number of attributes are 2 (14%) and 7 (12%), respectively. Furthermore, there was a single app collecting 9 attributes. Note that 10% of the apps did not ask for any extra attribute (i.e., *user_xxx*) outside of the basic information granted to all apps.

Collateral information collection. Second, and more importantly, we were interested in the extent of collateral information collection, when apps and app providers become entitled to collect attributes of a user without direct consent, through their friend installing the app. Performing the analysis over the dataset with regard to apps, using Eqn. (4.3) we found that 72.4% of apps can collect exactly one attribute from F^u (see Fig. 4.15 (b) for details). There are some apps that collect between 2 and 7 attributes in a collateral fashion, with a peak at 6 attributes. Furthermore, there were several apps that collected 11 attributes from the friends of the user.

Exclusive collateral information collection. With regard to attributes that can be collected *exclusively* from the friends of a user (i.e., collected from friends but not from the user themselves, see Fig. 4.13), using Eqn. (4.6), we found that more than 30% of the apps collect at least one attribute. Specifically, 28.9% of the apps under consideration can collect exactly one attribute, 1.45% as much as 10, and one app 11 attributes (see Fig. 4.15 (c) for details).

Potential data fusion by app providers offering multiple apps. Considering direct information collection and turning towards the 88 app providers, Fig. 4.15 (d) shows that providers predominantly collect 1 or 2 attributes per user. The slight increase in collected attributes indicates that app providers offering multiple apps are able to gather an enriched set of personal information from a single user. Extremes also do exist: one app provider collect as much as 17 attributes from a single user via a combined offering of multiple apps. It is interesting to see that are 4 app providers collecting at least 7 attributes, while there are almost 30 apps collecting the same amount; this indicates that data-ravenous apps are concentrated in a few app providers.

Data fusion also strengthens the extent of collateral information collection, as seen in Fig. 4.15 (e) compared to Fig. 4.15 (b): there is a slight but visible increase towards collecting more attributes compared to the single app scenario throughout the whole range of the histogram. Furthermore, some app providers collect as much as 14 and 18 attributes, exhibiting strong potential for data fusion and thus *profiling*. A similar effect of data fusion can be seen with regard to *exclusive* collateral information collection in Fig. 4.15 (f). A surprising nuance is the disappearance of the 11-attribute app: the *exclusive* collateralness of a given attribute vanishes, if another app from the same app provider collects the same one directly.

The most important observation here is that collateral information collection and *data fusion* are orthogonal mechanisms, which, surprisingly, amplify each other's effects, resulting in more pronounced and obscure information gathering. As a countermeasure, bringing transparency to such a scenario should provide straightforward benefits to Facebook users.

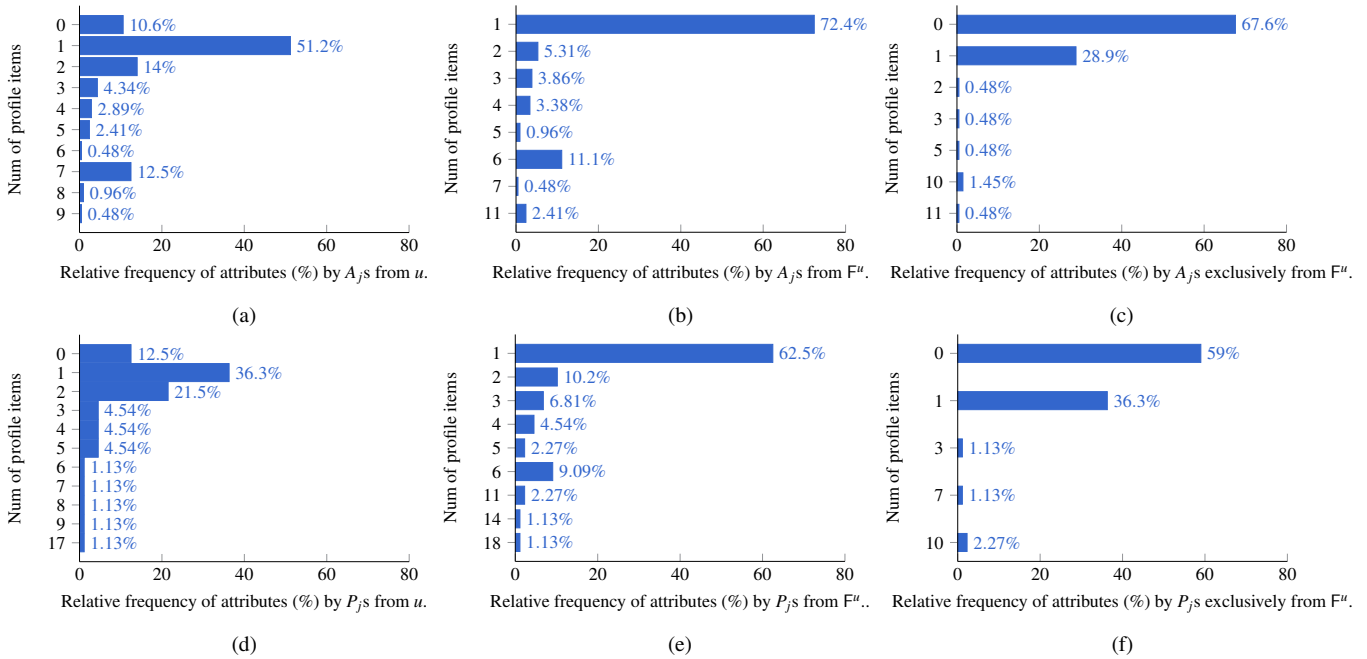


Figure 4.15: Number of attributes gathered via direct, collateral and exclusive *collateral information collection* wrt apps (A_j, top) and app providers (P_j, bottom).

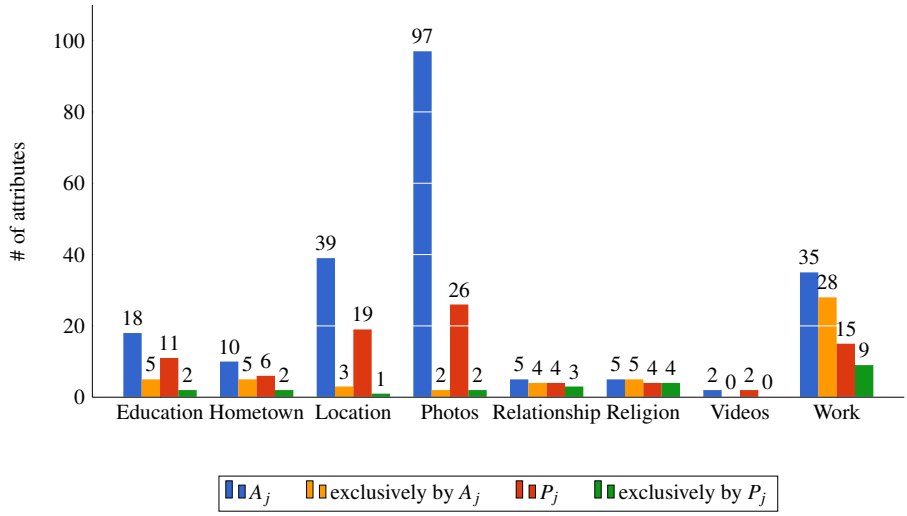


Figure 4.16: Total number of apps and app providers requesting collateral information collection of sensitive attributes (per attribute).

Potentially sensitive profile items. The collection of profile items deemed potentially sensitive by users deserve a more detailed look. Both related literature and our user study (see Sect. 4.2) found that attributes with connection to location, relationships, religious/ political views and photos/videos are particularly guarded more closely by users (i.e., where participants were at least very concerned). Analysing the AppInspect dataset with an eye on permissions granting access to such information, we derived the following results concerning collateral and exclusive collateral information collection and data fusion (see Fig. 4.16). The two most popular groups affected by collateral information collection are location-related attributes (including work history, education, hometown and location) and photos. Almost, half of the 207 apps collected the photos of the user's friends (97), with location (39) and work history (35) being the second and third most popular attributes. It is also easy to see that apps requesting these three attributes are clustered to a few app providers (between 2 to 4 app per app provider, mean value over all 3 attributes). In case of less popular collaterally collected attributes, the requesting apps are more uniformly distributed across app providers (between 1 to 1.6 app per app provider, mean value over the remaining 5 attributes). Note that exclusive collateral information collection also happens as evidenced by the yellow and green bars in Fig. 4.16.

4.5 Legal analysis

In this section, we investigate the research question: *Under the data protection legislation, is collateral information collection considered a risk for the protection of the personal data of Facebook users?* To answer this question, we use the General Data Protection Regulation (GDPR) [113] as a point of reference. We investigate the responsibility of collateral information collection, meaning who the data controller and data processor are. We discuss the relevance of transparency and consent for data protection. Furthermore, we review the necessity of implementing concepts of data protection by design and default. Based on our legal analysis, we identify who is responsible for such information collection, and argue that the lack of transparency and consent makes collateral information collection dubious from a data protection point of view.

4.5.1 General Data Protection Regulation

After 20 years and many technological innovations, the Data Protection Directive (Directive 95/45/EC) [111] was due for an update. A Regulation was chosen over a Directive, as a Regulation applies directly and does not need to be transposed into national law. After long negotiations on 24 May 2016 the European GDPR [113] entered into force and applied from 25 May 2018. The GDPR constitutes a single set of rules for all Member States that regulates the processing of personal data if it is done in the context of an establishment of an actor (controller/processor) within the EU, or if personal data of people who are in the EU is processed in the context of offering goods or services or the monitoring of their behaviour. Besides this extended territorial scope, the GDPR also brought some additional changes. In line with the research question of this section is the specification of (joint) controllership and the introduction of data protection by design and by default.

4.5.2 Controllers and processors

The GDPR, as the Directive 95/46/EC before, still uses the distinction between controllers and processors to associate certain duties on actors in the case of data processing. A controller is anybody who, alone or with others, determines the purposes and means of the processing of personal data [113]. A processor is any party that processes the personal data on behalf of the controller. This distinction is important, as the controller is the main responsible party for implementing and complying with the provisions of the GDPR.

The controller should normally be, as the name already says, the entity who has the control. The control can be stemming from 1) explicit or implicit legal competence (the controller might be appointed by law, or the capacity to determine is not explicitly laid down by law but stems from common legal provisions or established legal practice) or b) factual influence: in this case an assessment of the factual circumstances is made [105]. It should be noted that therefore, the determination of who is controller and processor is a factual one and cannot be contractually allocated to another person/entity which, considering the factual circumstances, has neither legal nor factual influence on how the personal data are processed. However, the content of contracts can be useful to assess who factually has the control. The control relates to the determination of the purposes and means of the data processing. Essentially, this means what level of influence the entity has on the "why" and the "how" of the processing [105].

4.5.3 Joint control

It is possible that for a single data processing several controllers are involved. The Directive was not explicit on this, and the concept of joint controllers was mainly provided by the Article 29 Working Party, however, the GDPR now includes explicitly the concept of joint controllers in Article 26 GDPR. The Regulation specifies that where two or more controllers jointly determine the purposes and means of processing, they are joint controllers and they need to determine their respective responsibilities for compliance with the Regulation in a transparent manner. This includes an arrangement the essence of which shall be made available to the data subject, and which may designate a contact point for data subjects. Nonetheless, the data subject may exercise his or her rights in respect of and against each of the controllers.

Just the fact that different parties co-operate when processing personal data does not mean that they are joint controllers. If they do not share common purpose or means with regard to the specific processing, then it might only be a transfer of data between separate controllers. It will often not be clear-cut and depend on the factual circumstances whether several controllers are considered to be separate controllers or in case they are considered joint controllers, to which extent they share the control. The same data set can be processed by several separate controllers or by joint controllers, e.g. when a travel agency books a journey and hotel every controller (travel agency, airline, and hotel) is a separate controller, however, if the controllers set up a common platform for the booking and thereby determine the essential elements of the means, they are joint controllers for this data processing [105].

4.5.4 Collateral information collection: who is controller?

The Article 29 Working Party in their opinion 5/2009 on online social networking specified that social networking service providers (i.e., Online Social Networks (OSNs)), such as in the case of Facebook, are data controllers under the Data Protection Directive [13]. As the concept of controller and processor has not much changed under the Regulation, it can be understood that Facebook is still a controller under the Regulation. Furthermore, third-party application providers (app providers) might be data controllers, if they develop applications which run in addition to the ones from social networking service providers, and users decide to use such a third-party application (app) [106]. Note that, we will use the terms social networking service provider and OSN interchangibly.

Whether Facebook and the app provider are separate or joint controllers will depend on in how far they share the purposes and means of the data processing. In general, the app provider will be a separate controller when it determines its own purposes and means for the processing of personal data. With regards to the processing activities of the app provider, the social networking service provider will generally not be a controller as it only provides access to the data (and thereby transfers it to the next controller), except of course the app provider is acting on behalf of the social networking service provider [13]. In that case, it could be either that the app provider is a processor whereby the purpose is fully decided upon by the social network service provider, or, if they determine the means, e.g. the platform together, there could be a case of joint control.

Is your Facebook friend a controller? According to Helberger et al. [159], "users are in many cases the primary perpetrators of privacy infringements on social networking sites". The Regulation, like the Directive, does not apply in certain cases. One of these exemptions is referred to as the household exemption. Article 2 of GDPR, provides that the Regulation does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity. In their opinion on social networking service providers, the Article 29 Working Party explained that the users of OSNs will be data subjects in most cases, or might fall under the household exemption [106]. However, there can be cases in which the user might not be covered by the household exemption. This is especially the case if the user is acting on behalf of a company or association, or uses Facebook for commercial, political or charitable goals [106]. Also, the exemption does not apply if the information is made accessible to an indefinite number of people [67]. This is, for instance, the case when the profile information is accessible beyond self-selected contacts or is index-able by search engines [106]. In such cases, it is beyond the personal or household sphere, and the user might be a controller [106].

If users qualify as controllers, they have to comply in principle with the data protection

obligations. However, the European data protection law was not intended for *amateur controllers* [159]. Some provisions can be applied without many problems. For instance, amateur controllers should also ask for their friends' consent before they process their friends' data and in principle provide information to their friends such as why they publish their data and where they do it [159]. However, in many cases "compliance with many provisions in data protection law probably exceeds the knowledge and experience of most users" [159]. Therefore, some provisions are not possible to be complied with by amateur controllers [159].

4.5.5 Transparency and information

While the principle of transparency was not as such mentioned in Directive 95/46/EC, it can be found very explicitly in the GDPR [113]. Article 5 specifically requires that data shall be processed in a transparent manner in relation to the data subject. Moreover, in Section 1 of GDPR, the rights of the data subject includes specific requirements on transparent information provision. Even Article 26 GDPR on joint controllers includes the requirement that the determination of the respective responsibilities should be done in a transparent manner. The Recitals 39 and 58 of the GDPR clarify that the principle of transparency requires that any information and communication relating to the processing of personal data need to be concise, easily accessible, easily understandable, in clear and plain language, and that visualisation can be used. In particular, information should be provided on the identity of the controller, the purposes of the processing and "further information to ensure fair and transparent processing in respect of the natural persons concerned and their right to obtain confirmation and communication of personal data concerning them which are being processed" [113]. However, users do not only need information when they are the data subjects: if they are controllers, they will need information to comply with their duties as data controllers [159]. To be more specific, they will need practical information about the potential security and privacy risks on the social networking services and they will need legal information [159]. The Article 29 Working Party recommends social network providers to provide adequate warnings to users about the privacy risks related to themselves and others when they upload information to the social networking service.

Even though in principle every controller must ensure the transparency of its processing, the social networking service provider faces higher expectations regarding information provisioning [159].

4.5.6 Consent and information

From a legal perspective, one of the main challenges of data protection attached to app permissions, as described above, is the fact that personal data processing may lack

legitimacy. Article 6 of the GDPR [113] provides a limited number of legal grounds legitimising personal data processing, such as the consent of the data subject. Consent, as stated in Article 4 (11) GDPR, is defined as "any freely given, specific, informed and unambiguous indication for the data subject's wishes". As enshrined in Article 6 (1) (a), the data controller, i.e., Facebook or apps, may collect, store, use and further disclose the data if the user has given her consent. For the consent to be valid, it has to fulfil certain criteria, which are specified in Article 7 GDPR, including that the controller needs to be able to demonstrate that consent has been given, the request for consent needs to be clearly distinguishable from other matters, understandable for the user and the user can withdraw consent at any time.

4.5.7 Data protection by design and default

An important change by the GDPR that is relevant to this chapter is the introduction of data protection by design and default in Article 25 of the GDPR. This provision obliges the controller to take appropriate technical and organisational measures to implement data protection principles to ensure that by default only personal data that are necessary for each specific purpose of the processing are processed. Such measures should ensure that by default personal data are not made accessible to an indefinite number of persons without the individual's intervention. This provision is important, as the main problem in the Facebook case described in this chapter is the possibility to access user A's data via user B since the default settings allow this data sharing while user A is not aware and never actively consented to it. Article 25 GDPR will require that the default settings are set in the most data protection friendly way, and therefore the user typically will have to actively opt-in if she wants to give apps access to data through her friends.

Furthermore, the GDPR stipulates that appropriate technical and organisational measures should be implemented which meet in particular the principles of data protection by design and by default [113]. These measures could provide "transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features" [113].

4.5.8 Profiling

Profiling is defined in Article 4 GDPR as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" [113]. As stated in Sect. 4.1,

app providers offering multiple apps could utilise *data fusion* to construct a more complete representation of users. Hence, data fusion implements certain sub-cases of profiling (e.g., directly extracting personal preferences and interests from Facebook attributes) while enabling other sub-cases of profiling by virtue of expanding the knowledge of the app provider (i.e., the data controller) on a given data subject. Regulating profiling plays a significant role in the GDPR, as it is mentioned in several articles and recitals. Article 22 GDPR declares that data subjects shall have the right not to be profiled if this profiling results in legal or other significant effects detrimental to them. The only exception relevant to our case is when the data subjects give explicitly consent. As per Recital 71 GDPR, in any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Also, such measure should not concern a child.

Article 13 GDPR moreover states that in the case of profiling, the controller is obliged to provide the data subject with information necessary to ensure fair and transparent processing, including the profiling logic and the significance and the envisaged consequences of such processing for the data subject. On top of this, Article 15 GDPR states that the data subjects have the right to access their personal data and the above information anytime (with a reasonable frequency) after the profiling took place. Article 21 GDPR also defines the right to object if profiling is done to enable direct marketing at any time and free of charge. Last, Article 35 GDPR makes a Data Protection Impact Assessment (DPIA) mandatory, in particular, if the processing is performed in order to take decisions relating to natural persons, and includes a systematic evaluation with respect to natural persons based on automated processing, in general, and profiling, in particular. Another focal case for a DPIA is if the processing is on a large-scale and involves special categories of data (as defined in Article 9 GDPR).

4.5.9 Is collateral information collection a risk for the protection of the personal data?

It is clear from the above that the GDPR poses some well-defined requirements to the social networking service providers, i.e., Facebook in our case, with regards to the protection of users' personal data. Several points, mostly concerning the legal obligation of users as amateur data controllers, on the one hand are murky at best. On the other hand the problem clearly converges when the personal data of users can be transferred from one controller to the other, as from Facebook to an app provider, without *notifying* the users (i.e., Article 5 GDPR) and without obtaining *consent* from the user (i.e., Articles 6, 7 GDPR).

With respect to the obligations of the data controller and processor to transparency, app providers can become data controllers and processors of a user's personal data, without the user becoming aware of such data transfer (i.e., *collateral information collection*). It should be noted that a Facebook user and their friends have insufficient information on both the amount of data that will be collected and the purposes their data will be used for by an app and its provider [291]. In other words, data collection and processing go far beyond the user's and their friends' legitimate expectations and interferes with the principle of *transparency* as per Article 5 GDPR.

Furthermore, considering the obligations of data controller and processor with regard to consent, third-party apps on Facebook may collect and proceed to process a user's personal data without prior, informed and direct consent from the user themselves, that is operating exclusively based on the consent provided by one of her friends. In other words, consent can be given only by the user who installed the app (i.e., friend) and not by the data subject (i.e., user) whose data is collected and processed. One might say that Facebook app settings give the users control over their personal data to be handed to apps by their friends via ticking the appropriate checkboxes in the sub-menu "Apps Others Use". Therefore, one could claim that consent has been theoretically given. However, it should not be considered as valid as it is not informed. In fact, owing to the default, pro-sharing privacy setting on Facebook (see Fig. 4.14), users are generally unaware of the fact that they have to uncheck the boxes (actively opt out) in order to prevent such data processing [126, 38]. This also goes against the concept of privacy by default [85]. It is worth mentioning that in a relevant case in the U.S., the Federal Trade Commission (FTC) required that such apps cannot imply consent, but this should be rather affirmatively expressed by users [125].

Finally, *profiling* is a separate concern on its own behalf; nevertheless, data fusion (a technique partly constituting and greatly enabling profiling) has an amplifying effect on collateral information collection (see Tab. 4.1). Moreover, this amplifying effect is mutual. Therefore, data protection obligations of app providers offering multiple apps are even more pronounced. As for all automated processing, especially if the outcome of automated processing are decisions relating to data subjects (natural persons) by the data processor, details about how and why the processing is done, and its potential effects on the data subject should be given to the data subject. We have no examples of this happening in the case of app providers or Facebook: we are not aware of any systematic means (potentially with the technical help of the Facebook Developers Platform[122]) where this information flow can be realised. To sum up, data fusion potentially resulting in profiling may happen without any transparency and consent: the data subject is simply not aware of such processing taking place. Another, intriguing aspect of profiling is the obligation for the profiler to conduct a DPIA if the decision based on profiling affect data subjects significantly and/or profiling is large-scale and involves special categories of data. Given the popularity of certain apps and the type of attributes collected a DPIA could be obligatory. Finally, the Article 29 Working Party

identifies any "datasets that have been matched or combined, for example originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject" likely to result in a high risk; therefore any operations resulting in such datasets should be covered by a DPIA [104].

4.6 Solutions to collateral information collection

In this section, we investigate the research question: *how can we mitigate collateral information collection?* To answer this question, we propose a privacy dashboard extension that aims at improving transparency and helping users to make informed decisions about such information collection, while introducing a scoring computation for evaluating the collateral information collection over the third-party applications (apps) and third-party application providers (app providers). Moreover, we discuss alternative solutions focusing on notification, access control mechanisms, cryptographic solutions, on app auditing, and in particular on a Data Protection Impact Assessment (DPIA) driven by the General Data Protection Regulation (GDPR).

4.6.1 Enhancing transparency with dashboard

Transparency Enhancing Technologies (TETs) [263, 331, 131], can support decision-making processes by helping users to understand the problem and foster users' control by assisting them to make informed decisions [154, 209]. For instance, to take action about the amount of information that may be affected by collateral information collection. The increase of awareness on personal data collection is in line with the legal principle of data protection by design and default (Articles 39 and 78 of the GDPR [113]). Furthermore, driven by our questionnaire responses, there are evidences that participants are not only concerned about the collateral information collection, but they also want to be notified and restrict access to their personal data on Facebook. They also consider removing the apps that can cause the collateral information collection.

In order to design a Transparency Enhancing Technology (TET) with regard to collateral information collection, we need two main ingredients: 1) a quantitative measure characterising the added exposure of profile attributes, and 2) a usable way to present this information to the user. Fortunately, established methods for constructing both ingredients are available in the literature [194, 293]. However, both the quantitative metrics such as a privacy score and usable presentation schemes such as dashboards are not yet tailored to the specifics of collateral information collection. In the following section we demonstrate how to integrate our quantification mechanisms for collateral information collection into these frameworks.

Privacy Score

In a nutshell, we describe the premises to compute the Privacy Score (PS) for an app (i.e., A_j) and an app provider (i.e., P_j), as an indicator of the level of collateral information collection. The higher the PS the more significant the threat for a user, meaning that more personal data of the profile of a user (i.e., u) can be collected by an A_j and P_j. To compute the PS, Liu and Terzi [194] proposed a formula consisting of the product of the visibility of a user's profile attribute (i.e., a_i) in an Online Social Network (OSN) graph with its sensitivity that represents its perceived importance (weight). Our PS computation formula represents the case of collateral information collection from A_js and P_js in an OSN.

Sensitivity. We denote by $\mathcal{S}$ the set of different attribute weights in u's profile, i.e., $\mathcal{S} = \{\sigma_1, \ldots, \sigma_n\}$. We consider σ_i for $1 \leq i \leq n$ the number of different attribute weights in u's profile where $\{\sigma_i \in \mathbb{Q} | 0 \leq \sigma_i \leq 1]\}$, i.e., a_i is more sensitive than $a_{i'}$ iff $\sigma_i > \sigma_{i'}$, therefore,

$$\sigma_i = \begin{cases} 1 & \text{if } \textit{very sensitive}, \\ 0 & \text{if } \textit{not sensitive}, \\ 0 < \sigma_i < 1 & \text{if } \textit{in between}, \end{cases} \tag{4.15}$$

for $1 \leq i \leq n$.

Privacy Score of u for A_js. We denote by $\mathtt{PS}_A^{u,\mathsf{F}^u}$ the PS of u for all A_js, where $\{\mathtt{PS}_A^{u,\mathsf{F}^u} \in \mathbb{Q} |\ 0 \leq \mathtt{PS}_A^{u,\mathsf{F}^u} \leq 1\}$, when A_js can collect attributes from u's profile in F^u, i.e.,

$$\mathtt{PS}_A^{u,\mathsf{F}^u} = \frac{\sum_{i=1}^{n} (\sigma_i \times \vec{\Pi}_{A_j}^{u,\mathsf{F}^u})}{|\mathcal{T}|} = \frac{\sum_{i=1}^{n} (\sigma_i \times \bigvee_{A \in \mathsf{L}^{u,\mathsf{F}^u}} (\vec{A}_j^{u,\mathsf{F}^u}[i] \wedge \vec{C}_{A_j}^{u,\mathsf{F}^u}[i]))}{|\mathcal{T}|} \ . \tag{4.16}$$

The complexity of evaluating Eqn. (4.16) for all f in F^u is $O(n^5 \times |\mathsf{L}^{u,\mathsf{F}^u}|)$.

Privacy Score of u for all P_j. We denote by $\mathtt{PS}_P^{u,\mathsf{F}^u}$ the PS of u for all P_js, where $\{\mathtt{PS}_P^{u,\mathsf{F}^u} \in \mathbb{Q} | 0 \leq \mathtt{PS}_P^{u,\mathsf{F}^u} \leq 1]\}$, when P_js can collect attributes from the u's profile in F^u, i.e.,

$$\mathrm{PS}_P^u = \frac{\sum_{i=1}^{n}\left(\sigma_i \times \vec{\Pi}_{P_j}^{u,\mathsf{F}^u}\right)}{|\mathcal{T}|} = \frac{\sum_{i=1}^{n}\left(\sigma_i \times \bigvee_{P\in\mathcal{AP}^{u,\mathsf{F}^u}}\left(\bigvee_{A_j\in\mathsf{P}_j^{u,\mathsf{F}^u}}\left(\vec{A}_j^{u,\mathsf{F}^u}[i]\wedge\vec{C}_{A_j}^{u,\mathsf{F}^u}[i]\right)\right)\right)}{|\mathcal{T}|}\,. \tag{4.17}$$

The complexity of evaluating Eqn. (4.17) for all f in F^u is $O(n^5 \times |\mathsf{P}_j^{u,\mathsf{F}^u}| \times |\mathcal{AP}^{u,\mathsf{F}^u}|)$.

Quantifying sensitivity. One way to measure and quantify the perceived sensitivity of users is by running survey studies. Minkus et al. [217] measured the sensitivity attributed to different privacy settings by Facebook users from a survey of 189 participants. They estimated how users perceive the importance of each privacy setting. They identified variations of how sensitive each privacy permission is. For instance, the permission "What personal information goes into apps others use?" is considered more sensitive (i.e., $\sigma_i = 2.82$) than "Who can send you friend requests?" (i.e., $\sigma_i = 1.09$).[4] Their sensitivity estimations, however, is slightly different from quantifying the sensitivity of the attributes in our case. Although there is a relation between privacy settings and attributes, there are no studies to identify the correlation between the perceived sensitivity from a privacy setting to the set of attributes. Moreover, there are no studies yet for measuring the sensitivity of the attributes collected by app / app provider and neither for *collateral information collection*. Nevertheless, Minkus et al. [217] provide us with sensible default sensitivity values; these can be then updated by running a new survey (this is left for future work). Privacy-conscious users can then adjust the sensitivity values on demand and dynamically.

A different promising direction for quantifying the sensitivity of attributes is by demonstrating that attributes an app can collect before, during or after the installation. This can be achieved through a privacy dashboard demonstrating all possible attributes an app can collect during app installation with the help of authorisation dialogues [322].

4.6.2 Damage control: Privacy dashboard

To enhance transparency, several tools have been suggested [158, 175]; among these privacy dashboard is a well established instrument [328]. For privacy dashboards, several designs have been proposed [36, 130], with some of them specifically tailored for Facebook [293, 46]. Within our work, we propose components that can be used in existing dashboard designs [293, 36, 46], aiming to extend and enhance the minimisation of data disclosure by the users for the *collateral information collection* (see Fig. 4.17 for an initial user interface design).

[4] Note that, our proposed sensitivity values (i.e. σ_i) should be from $0 \leq \sigma_i \leq 1$ and other evaluations of σ_i can be scaled accordingly.

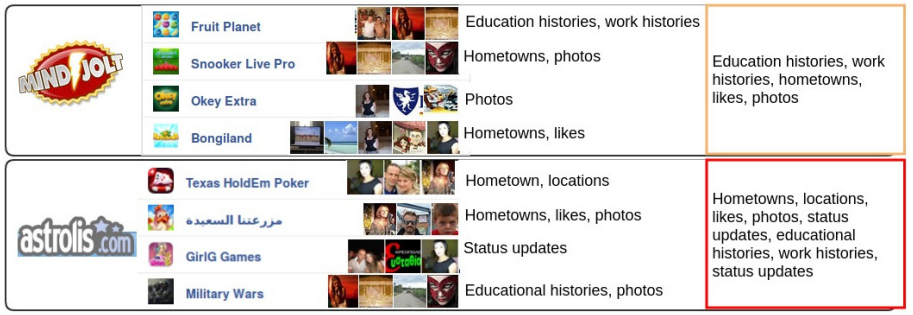


Figure 4.17: Privacy dashboard: user interface concept.

Functional requirements. Transparency should be reflected by visualising the personal data that can be disclosed to apps and app providers by the friends of a user on Facebook. Specifically, the dashboard extension should demonstrate which app is collecting which of the personal data of a user and through which friend. Moreover, the dashboard extension should demonstrate the impact of *data fusion* via multiple apps offered by a single app provider. The dashboard extension should provide all the quantitative representations in a readable and understandable format. Orthogonal to these issues, a user should be able to adjust the weights of sensitivity for her personal profile attributes, as this should be reflected in the computation of the privacy score. However, she should also be able to use the dashboard right away with reasonable default weights.

Proposed design. Technically speaking, the proposed dashboard tools illustrate how the data disclosure of a user takes place through the acquisition of the user's personal data via apps (and respective app providers) when installed by their Facebook friends. It displays the nature and proportion of the user's personal data that can be collected by apps and, more importantly, by app providers. To better help users to understand, the privacy metrics are visualised and the level of potential information dissemination is represented by colour scale to indicate the level of collateral information collection [153].

From our user opinion study, we have identified that Facebook users are more concerned about certain types of personal information such as photos, videos, location, and relationships [217]. Our dashboard can accommodate the visualisation of *data fusion* and the degree of collateral information collection by assigning different weights (sensitivity) to each attribute in the user's profile. These weights can then be manually fine-tuned by the user for an app and app provider.

Dashboard technical limitations on Facebook. Implementing a privacy dashboard on Facebook has several limitations. To identify which of the installed apps for a user enable the collateral information collection, the list of apps needs to be retrieved from Facebook. Currently, as the Facebook Application Programming Interface (API) does not allow for inspecting the installed apps of a user, one way is by scraping the app centre page. However, Facebook has designed their web interface to resist high volume content retrieval. It is possible for a skilled developer to circumvent this protection and to collect and identify which installed apps enable the collateral information collection for a user. That can be automated but the task is far from trivial for a non technically-savvy user (e.g., using Selenium [268]). The list of apps can be retrieved while a user is logged in to her Facebook account. To identify which friends have also installed similar apps, a scraping operation needs to be performed for each page of an app. To identify apps that enable collateral information collection and have been installed only by friends of a user, it needs all possible pages of apps to be scraped and downloaded from the Facebook app centre; assuming that a developer has a compilation of the majority of popular apps that enable collateral information collection. Making the problem more complex, Facebook is susceptible to constant changes: it updates the interface, the permissions, and the privacy settings regularly, making the development of a dashboard cumbersome and with a need for continuous modification.

Detailed design and implementation of the dashboard remain the next step in our future work. Additional information such as claimed purpose of collection by the apps can be added to the dashboard. Moreover, further functionality can be added to the dashboard such as leading the users from the dashboard to uninstall the app, which would strengthen compliance with the GDPR [113].

4.6.3 Alternative countermeasures

In this chapter, we discussed that TETs can support notification and facilitate direct consent in order to comply with the GDPR [113]. Notification can be enhanced using *privacy nudges* [326, 59] while direct consent with *privacy authorisation dialogues* for apps [322, 338], and *access control mechanisms* [324, 164, 240] tailored to collateral information collection. Considering *privacy nudges*, such an enhanced notification can be provided in the following manner: a notice about the additional attributes that the app provider will collect, a timer interface before the Facebook users clicks the "app installation", or even a "sentimental nudge" about collateral information collection as bad practice. Furthermore, as standalone consent mechanisms, *access controls* and *authentication dialogues* should be designed in the following manner: to provide permissions for collateral information collection for both the friends and the users, and to explicitly highlight such permissions while Facebook users install an app that enables such information collection. Moreover, permissions should have a direct, one-to-one

mapping to the attributes of the profile of a Facebook user, which is not the case for the *user_friends* permission.

Additional to transparency solutions, other countermeasures to collateral information collection can use cryptographic primitives to minimise or mitigate such indirect information collection. Such solutions can help to better control the dissemination of the information of Facebook users, providing strong privacy guarantees to the users. For instance, flyByNight [196] and Scramble! [26] are utilising encryption schemes to ensure confidentiality and integrity of messages exchanged among Facebook users. However, encryption solutions on Facebook and other OSNs and are commonly detected and blocked from their systems.

Considering Facebook, v2.x of the API has the potential to decrease both the likelihood and the impact of collateral information collection. Apps using these API versions can only see a shortened friend list for the installing user: those who have also installed the app *and* granted the *u*ser_*f*riends permission. The app can gather profile attributes only from users appearing in this friend list. While this is an API change we acknowledge, it does not constitute a total solution. First, although it may considerably reduce the risk of getting exposed to collateral information collection, those affected may still suffer from *exclusive* collateral information collection: there could be some attributes exclusively collected through friends and not directly. Second, since there are a plethora of affected apps, the beneficial effect of the API change is not so pronounced for a user: the probability of installing at least one app enabling exclusive collateral information collection is still high. Third, this API change does not have any effect on multi-app data fusion. Illustrating this through Table 4.1, the new API makes sure that the top cells are nonempty sets, should the lower cells be nonempty sets as well.

Finally, a DPIA can be viewed more than just an obligation for high-risk data processing operation: it can be also viewed as a legal countermeasure against privacy threats with regard to collateral information collection. As DPIA is a process for building and demonstrating compliance with GDPR, it can provide meaningful information both to data protection authorities and concerned data subjects, thereby promoting transparency and also trust. Moreover, for Facebook apps and app providers, a qualitative DPIA can be augmented by a more quantitative privacy risk assessment based on our proposed *Privacy Score*, which captures all traditional aspects of risk computation (see Sect. 4.6.1). Such an enhanced DPIA promotes comparability over different apps and app providers, and conveys more tangible information towards users.

4.7 Chapter summary

We conclude that *collateral information collection* poses a privacy threat for Facebook users. In a nutshell, we presented a multi-faceted study concerning the collateral information collection caused by third-party applications (apps) when installed by the friends of a user on Facebook. Our main findings were the following. Using a questionnaire, we showed that the vast majority of our participants were very concerned and would like proper notification and control mechanisms regarding collateral information collection. Also, they would like to restrict the apps accessing the profile data of both the user and their friends, when their friends' enable the collateral information collection and vice versa. Running simulations for various network topologies and app adoption models, we quantified the probability that a Facebook user can be affected by the collateral information collection. Assuming an app with more than 1 M users (such as TripAdvisor), there is an 80% probability for this problem to appear. Employing real data from the Facebook third-party apps ecosystem, we quantified the significance of collateral information collection by computing the proportion of attributes collected by apps installed by the friends of a user. Based on popular apps, we identified that almost half of the 207 apps that enable the collateral information collection collected the photos of the friends of a user, with location and work history being the second and third most popular attributes. Through the prism of the General Data Protection Regulation (GDPR), we investigate and conclude that collateral information collection is likely to result in a risk for the protection of the personal data of the Facebook users. The cause is the lack of transparency and consent, the non-existence of privacy by default in Facebook's privacy settings, and the amplifying effect of *data fusion* and, potentially, profiling.

To mitigate the collateral information collection, we proposed solutions aiming at enhancing transparency and increasing control of the information dissemination of Facebook users through third-party apps. In particular and for enhancing transparency, we proposed a *privacy dashboard extension* to enhance the existing privacy dashboard designs when collateral information collection needs to be instantiated and quantified. Such an enhancement can also help to empower the users' decisions and enforce restrictive actions if necessary. Moreover, we discuss alternative solutions, focusing on notification and access control mechanisms, cryptographic countermeasures and app auditing and in particular on Data Protection Impact Assessment (DPIA) driven by the GDPR.

To the best of our knowledge, our work is the first one to report the potential user *profiling* threat that could be posed by third-party application providers (app providers). They can gain access to complementary subsets of personal data from user profiles by offering multiple apps. Moreover, our study can serve as a guide for future apps, software platforms, or permission systems such that, the interdependent aspects of privacy can be directly taken into consideration in the design phase.

Part III

Physical-keyless car sharing systems

Chapter 5

A security and privacy analysis and elicitation of requirements for physical-keyless car sharing systems

Auto makers are evolving from being 'just' producers of motor vehicles to networked mobility service providers.

ERIK JONNAERT, *ACEA Secretary General*

Publication Data:
I. Symeonidis, M. A. Mustafa, and B. Preneel, "Keyless Car Sharing System: A Security and Privacy Analysis," In IEEE International Smart Cities Conference, (ISC2 2016), Trento, Italy, September 12-15, 2016, pp. 1–7.

Contributions: Main author.

5.1 Introduction

Car sharing demonstrates a high potential for users with the short-term, occasional and dynamic use of cars [215]. As a combination of car rental and car ownership, cars can be owned by private individuals or companies and shared to different users, potentially unknown to car owners [6]. Traditional car sharing systems, even if they are extensively used today [6, 215], may not be flexible enough regarding users' convenience. For a car owner to share a car, she (or another designated person) has to transfer the physical key to the one who is going to use the car. Handing physical keys may not even be feasible, in some cases.

In this work, we consider physical-Keyless car Sharing System (KSS), where a user can use a car without the need of analogue physical keys. Such physical keys can be replaced by digital keys on mobile devices such as smart phones, or tablets. Using a mobile device as a key and in-vehicle telematics a user can access (i.e., lock and unlock) and drive a car [234, 132, 52]. Moreover, with the use of mobile devices, users can share their cars with others such as family members, friends [315, 320, 181, 64] or even with people unknown to them, complicating the landscape of car sharing.

5.2 Security and privacy threats and functional challenges

Despite the advantages, security and privacy play a crucial role in car sharing systems. Security-wise, car sharing systems face physical objects being shared (cyber-physical security issues [24]), digital platforms running on public cloud resources being utilised (inheriting infrastructure security issues) and third-party financial providers being integrated (e.g., attack on the financial provider of Uber [222]), among others. Concerning the privacy angle, car sharing systems generate, control, process and transfer large amounts of personal data [300, 78, 97]. Thus, new threats have emerged: providers are able to track users via their transactions (e.g., in car sharing [101, 272]), poorly managing privacy (see Fig. 5.1) and inferring additional sensitive information such as religious beliefs [255] or even health-related information [137]. Sensitive personal data are related to *fundamental rights and freedoms*, and merit protection regarding the collection and processing by the *data controllers*, as it is articulated in the General Data Protection Regulation (GDPR) [113].

Also, a car sharing system can introduce various other concerns with respect to connectivity issues [88], car key revocations when a mobile device is stolen [143], and the fact that malicious users can attempt to manipulate or even destroy potential forensic evidence on a car or their devices. In short, the shared car is a mobile entity,

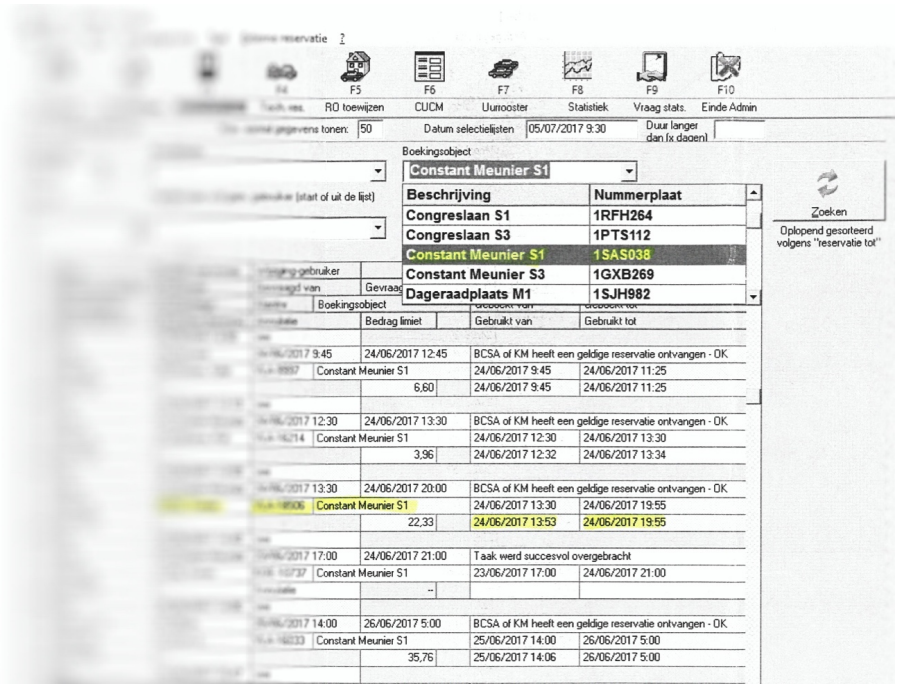


Figure 5.1: Example of bad privacy practices of a car sharing company: Screen-shot used as evidence for a speeding ticket report. The highlighted record corresponds to the user and the period of car use, while several other users are visibly exposed. Together with the name and client-ID (i.e., blurred left part), additional information is included in the report such as the pick-up location, the type, the usage-cost and the period of each listed user used a car.

and network connectivity with a physical-keyless car sharing system may not always be established. Hence, sharing and accessing a car can become problematic for locations with less reliable, or even limited network connectivity. Moreover, to access (unlock) a car, the digital key may need to be stored in a mobile device of the owner. However, the annual theft rate of mobile devices (e.g., smart phones), only in the U.S. exceeds 220 M with a rising tendency nowadays (in 2018) [278, 143]. Considering the fact that a revocation and update operation of the digital key of cars should be performed on a stolen or damaged device, this can be problematic for car owners. Such operations are expensive in time and cost, as they are mainly performed on stations authorised by the car manufacturer. Furthermore, as users have physical access to the shared car, they may act maliciously and attempt to manipulate or even destroy the evidence stored in the car or on their devices. However, in the case of disputes of car incidents, law enforcement should be able to retrieve the information related to the shared car without

violating the privacy of other users [113].

5.3 Contribution

To the best of our knowledge, there is no prior work that proposes a detailed design of KSS and that extensively specifies the security and privacy threats of such a system. Therefore, the main contributions of this chapter are the following:

- We propose a novel high level system model for a KSS that allows car owners to use and share their car with others, unknown to them, by generating and distributing access rights (e.g., authorisation access tokens). It uses digital keys of cars and a fully-fledged system to provide such sharing possibilities. For the analysis, we describe the entities involved, we define the functional requirements for such a system and specify the required operations.
- We investigate the threats of such a system by performing an initial adversarial analysis for each entity that is participating in the KSS. In order to comprehensively examine the threats that can jeopardise the security of the system and the privacy of its users, we use two frameworks namely STRIDE and LINDDUN, respectively.
- Based on the threat analysis, we specify the security and privacy requirements that need to be fulfilled, aiming to allow owners to use their cars as well as others to book and share cars in a secure and privacy-preserving manner.

The remainder of this chapter is organised as follows: Sect. 5.4 describes the aim and the methodology of the analysis of a KSS. Sect. 5.5 proposes a novel KSS, Sect. 5.6 analyses the security and privacy threats and Sect. 5.7 specifies countermeasures as a set of security and privacy requirements tailored to the proposed system. Sect. 5.8 describes the key points of this chapter while Sect. 5.9 provides the summary of this work and proposes the follow-up directions of the chapter.

5.4 Methodology for secure and privacy-preserving system analysis

The aim of this work is to systematically perform an *analysis* of the KSS and to *specify the requirements* that such a system needs to fulfil (see Fig. 5.2). The requirements can be functional, to define how the system should operate, security related, to define how

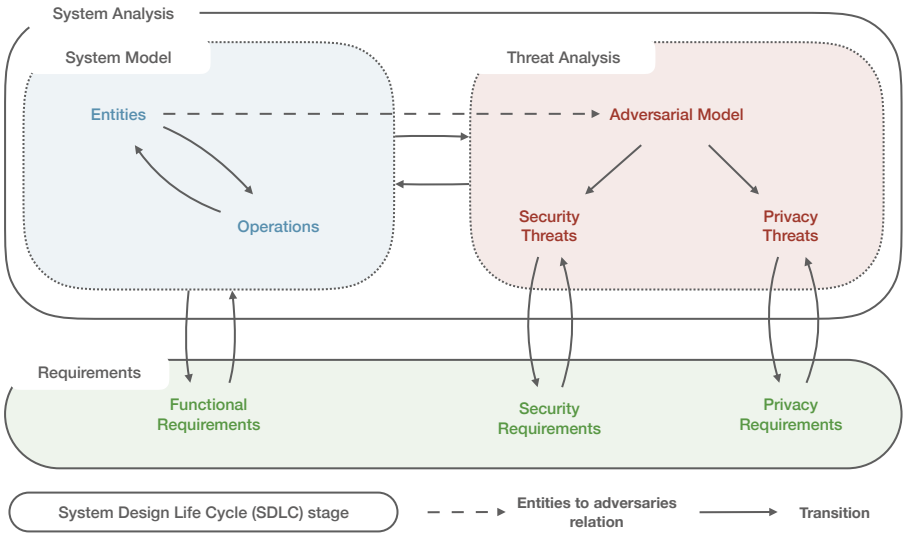


Figure 5.2: System *analysis* and *specification of requirements* corresponding to the Secure-System Development Life Cycle (S-SDLC) stages.

the KSS should be designed to restrict adversaries aiming to exploit the operations of the system and privacy related to define how the KSS should be designed to restrict adversaries aiming to violate the privacy of the KSS users.

To *analyse* a KSS holistically and in a structured way, that is to specify the system model (i.e., entities and operations) and identify the threats, we used the system engineering approach (i.e., S-SDLC), as described in Sect. 1.5. In particular, we *analyse* the system model and identify the entities involved in a KSS, meaning all the players in the system. Moreover, by defining which functional requirements should be put in place, we specify the minimal operations of such a system. Note that, since there was no physical-keyless car sharing system described in the literature, several iterations of the system model, the requirements and the threat analysis were performed until finalisation.

The entities of the KSS can threaten the system, holistically or its parts, by becoming *adversaries*. The adversaries can, deliberately, try to exploit the vulnerabilities of the KSS *operations* and harm the system. To identify such vulnerabilities, we perform a comprehensive analysis of the threats (i.e., security and privacy) taking into consideration the capabilities of the adversaries. For such analysis, we used relevant threat-analysis frameworks (i.e., STRIDE and LINDDUN), as we describe in Sect. 5.6. To mitigate the threats, it is essential that the outcome of the *analysis* results into *requirements* (i.e., security and privacy) aiming to harden the KSS operations in a secure and privacy-preserving manner, restricting the opportunities for any potential

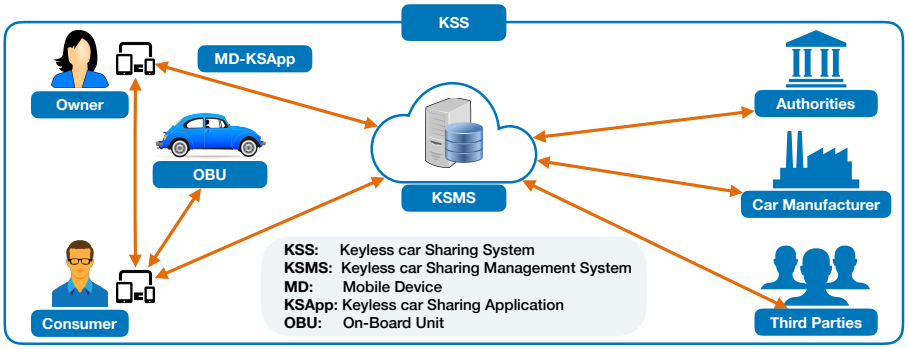


Figure 5.3: System model of physical-Keyless car Sharing System (KSS).

adversary to exploit the system.

5.5 High-level system model analysis

This section details the high-level system model and specifies the functional requirements that the KSS should fulfil.

5.5.1 Entities involved

Our high-level system model for a KSS consists of the following entities (see Fig. 5.3).

Users. Users are individuals or companies, who are willing to share their cars (i.e., *owners*) or use cars that are available for sharing (i.e., *consumers*). Car owners can provide *consumers* with permanent or temporary (on-demand) access to their cars. A *consumer* can be anyone such as a family member, friend or even unknown to the *owner* of a car.

Keyless car Sharing Management System (KSMS). The Keyless car Sharing Management System (KSMS) is a server (or a set of servers) that manages the operations of the KSS. It aims to provide administrative support such as (i) *profile management*, i.e., cars and users registration, (ii) *booking management*, i.e., post/search offers and requests, and car bookings, (iii) *access provision management*, i.e., assign, update and revoke access rights for users to access and use cars, (iv) *payment*

management, i.e., billing calculation and fee payment. Note that we will use physical-keyless and keyless interchangeably throughout the text.

On-Board Unit (OBU). The On-Board Unit (OBU) is an embedded or standalone hardware/software component. It is used for storing the digital key of the car and performing critical operations related to sharing the car such as the authorisation of users accessing the car. It is equipped with a wireless interface such as WiFi, Bluetooth, Near-Field Communication (NFC) and Long-Term Evolution (LTE) communication.

Car Manufacturer (CM). The Car Manufacturer (CM) is a company that manufactures cars. It provides the car and its owners with static physical and digital keys (bound to the physical key) for their cars. It is also responsible for the generation, distribution and revocation of the keys which are stored within the OBU of the car.

Keyless car Sharing Application (KSApp). The Keyless car Sharing Application (KSApp) is a software application developed for mobile devices such as smart phones. Users can interact with the KSMS via the KSApp application.

Mobile Device (MD). The Mobile Device (MD) is any mobile device such as smart phone or tablet that can run the KSApp. It is equipped with a wireless interface such as WiFi, Bluetooth, NFC and LTE. It is noteworthy to stress that while MDs are becoming powerful, none of them still have unlimited memory, processing power and battery capacity.

Authorities. Authorities are government agencies that are responsible for managing information about cars and their owners. Their aim is to administrate records for car ownership, blacklists of stolen cars, safety certificates of cars, driving licence certificates of users and to perform forensic investigations of car incidents.

Third Parties (TPs). Third Parties (TPs) are any organisations, institutions or companies that may be interested in the data generated by the KSMS, such as insurance companies.

5.5.2 Functional requirements

In order to be acceptable to the KSS users, our proposed system model should satisfy the following Functional Requirements (FRs).

- **FR1.** Users should be able to share their cars with other users. In detail, a car *owner* and a *consumer* should be able to (i) create (register), update and withdraw a profile, (ii) post car sharing offers and requests, (iii) search for offers and requests, (iv) accept (i.e., booking) or rejects requests, (v) use the KSApp on her MD to access a car, (vi) pay (be paid) for using (sharing) a car and (viii) assign/receive reputation scores to/from other users.
- **FR2.** An owner should be able to (i) generate, update, distribute and revoke (temporary) authorisation access tokens to the selected consumers, (ii) update and revoke (temporary) digital keys to lock/unlock and drive her car, (iii) retrieve the drop-off location of the car, and (v) report to KSMS when the car is not returned and/or stolen, and determine the location of the car with the help of he KSMS and *authorities*.
- **FR3.** A consumer should be able to (i) book a car, (ii) receive an authorisation access tokens (temporary) for the car, and (iii) retrieve the pickup location of the car.
- **FR4.** The KSMS should be able to (i) validate the identity of users and their driving licences, (ii) validate the profile information of cars, ownership and safety conditions, (iii) manage the users' profile access, search, update and withdrawal, (iii) post and inquire car sharing offers, requests and bookings, (iv) generate, distribute, update and revoke digital keys to cars and users, and (v) generate, distribute, update and revoke authorisation access tokens to the selected consumers.
- **FR5.** The MD should not perform computationally intensive operations as most of MDs are resource-constrained devices.
- **FR6.** The OBU should perform all the necessary operations (e.g., cryptographic) to authenticate and allow the selected users (i.e., consumers) to access the reserved cars.

5.5.3 System model specification of operations

For the FRs of the system, the specifications of all operations that a KSS should provide are described next (see Fig. 5.4).

A. KSMS operational management. Operations that the KSMS needs to perform.

A.1 System Setup. In technical terms, the KSMS performs all the necessary steps to offer the car sharing service such as system setup and network configuration. From an administration perspective, it specifies the servers involved and enables the interaction with the *authorities*, the CMs and TPs.

A.2 KSMS management. KSMS manages all or a part of the operations of the KSS, including management of the profile of users and cars, the *booking management*, the *car access provision management*, the *billing-payment management* and the *reputation-scores management*.

B. Users and cars profile management. Operations of the profile management of users and cars such as the registration and account management.

B.1 Registration of users and cars. A user provides to the KSMS all of the necessary information for the service registration such as an email address, a proof of her identity (e.g., passport, identity card), and her driving licence. Moreover, an owner of a car provides the KSMS with all the necessary information for registering her car such as the type, model, colour, engine power and certificates. The KSMS can also communicate with the authorities and the insurance companies (i.e., third parties) for completing and verifying the validity of the information provided by the owner and her car.

B.2 Users and cars operational management. A user can access, store, update or delete her profile information on the KSMS via the KSApp on her MD such as her username, age, contact details and friends. For stolen cars, the car owner can communicate with the KSMS to revoke the digital key of her car, stored in OBU, using her MD-KSApp. The KSMS then is responsible for revoking (and re-issuing) such a digital key for cars by communicating with the CM.

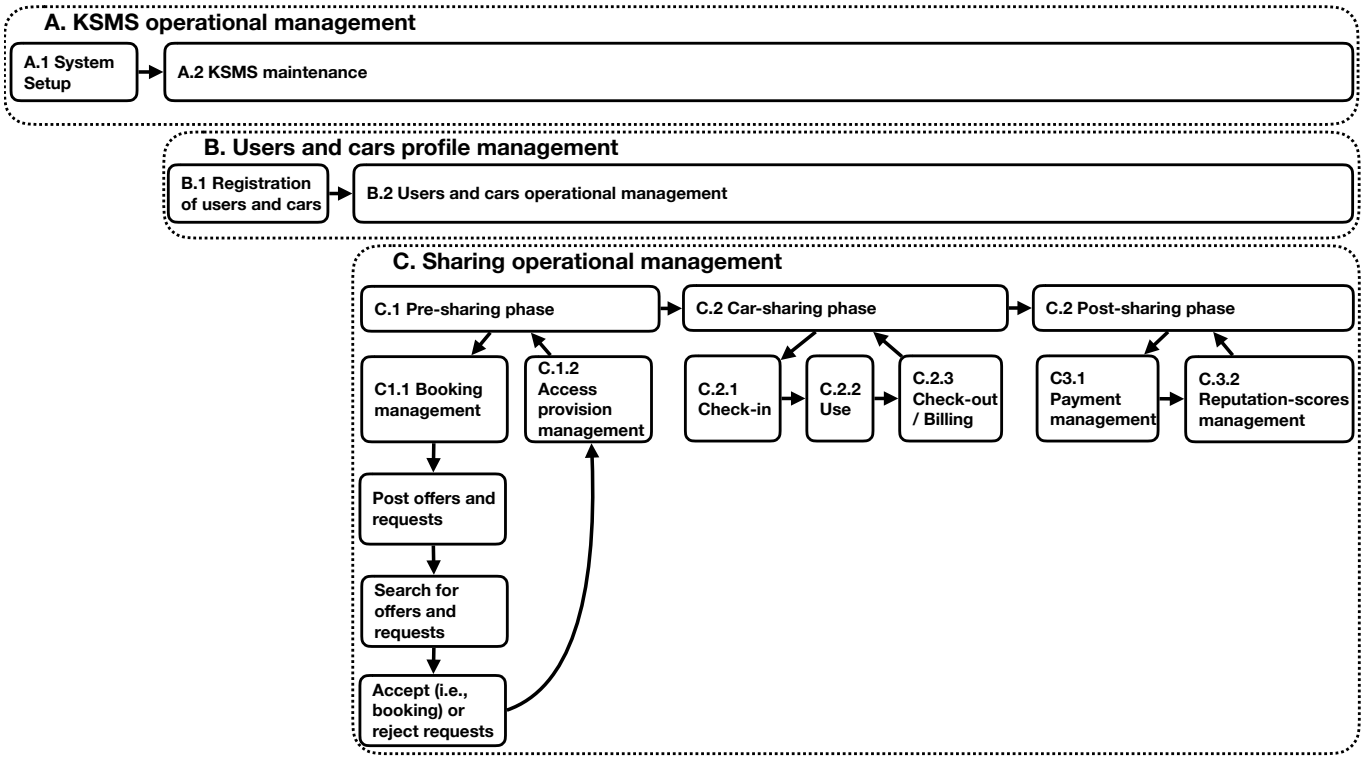


Figure 5.4: Specification of operations for a KSS.

C. Sharing operational management. Operations related to car sharing such as booking, using and paying for a car, are performed in distinct phases i.e., pre-sharing, sharing, post-sharing respectively.

C.1 Pre-sharing phase. Takes place before a car is shared.

C.1.1 Booking management. Users communicate with the KSMS using their KSApps on their MDs for the operations of posting, searching and booking a car.

- *Post car sharing offers and requests* An owner (consumer) post a car sharing offer (request) which includes: her profile data, the requesting (offering) price of sharing, the (preferred) pickup and drop-off location, the availability (preferred) period, and the (preferred) profile data of her (a) car.
- *Search for offers and requests* Users send inquiries for the available offers or requests to KSMS.
- *Accept (i.e.,* booking*) or rejects requests* If a user is interested in booking a car, the following operations are performed: (i) a consumer submits a request to book the car, (ii) the owner receives one or more requests for the car and (iii) the owner accepts one of the requests notifying the selected consumer or rejects all of the requests.

C.1.2 Access provision management. The KSMS provides all of the necessary actions for *car access provision*. In short, the MD-KSApp of an owner assigns access rights to the selected consumer and the car tailored to the pre-agreed *booking* period. Such car access provision can be realised by an *authorisation access token* and cryptographic operations over such a token. Both the MD-KSApp of the consumer and the OBU receive the authorisation access token which can be used to (temporarily) access the car. For "misbehaving" consumers, the car owner can update and revoke on demand the access token using the entities of the KSS such as her MD-KSApp and the KSMS.

C.2 Car sharing phase. Takes place while a car is shared.

C.2.1 Check-in. The selected consumer uses the received authorisation token to access the car using her MD-KSApp.

C.2.2 Using the car. The OBU of the car monitors and (continuously) verifies whether the shared-car satisfies the pre-agreed booking conditions. For instance, it verifies whether the car is within the pre-agreed booking period, distance covered and region of travelling.

C.2.3 Check-out and billing. After car use by a consumer, the car can be used again by another user such as the owner or another consumer. The OBU of the car and/or the consumer via her MD-KSApp, notifies the KSMS and/or the MD-KSApp of the owner for the time and drop-off location of the car. Then, the KSMS computes the cost of using the car considering the initial booking information and the actual sharing period.

C.3 Post-sharing phase. Takes place after a car is shared.

C.3.1 Payment management. The consumer pays the owner the fee for using the car with the assistance of the KSMS.

C.3.2 Reputation scores management. The owner (consumer) of the car assigns/receives a reputation score for sharing (using) a car with the assistance of the KSMS.

5.6 Threat analysis

This section provides an extensive and structured analysis of the security and privacy threats of the proposed KSS model. It describes the adversarial model for each entity of the KSS in detail. Based on the adversaries and the operations specified, we identify the threats that can harm the security of the system and privacy of users.

5.6.1 Adversaries

Every KSS entity can be realised into a broad spectrum of adversarial models [84, 55]. In a nutshell, adversaries can be *passive adversaries*, meaning that they aim to collect and extract information exchanged within the system without violating the protocol [55]. An entity can be an *active adversary* where it uses an arsenal of attacks and deliberately aims, by all possible means, to break or extract information from the system.

For each entity, we consider the following levels of adversarial activity that the KSS should be able to confront.

Users are passive or active adversaries. A malicious user can try to passively and/or actively manipulate the information exchanged with and stored within the KSS. For instance in an attempt to gain financial benefits or lower the credibility of the system, an adversary can try to extract, collect or alter the availability period, location and profile information (e.g., type, and the number of seats) of a car targeting the KSS users or the system itself. Such adversaries can vary from sophisticated hackers, to organised crime or even to governmental agencies. Depending on the resources and capabilities, an adversary can try to corrupt any set of users or other entities (i.e., KSMS) of the KSS. However, we assume that adversaries are not able to break the underlying cryptographic primitives.

OBU is tamper-proof and tamper-evident. An adversary can try to attack the OBU of a car, aiming to retrieve the cryptographic keys or to collect the personal data of passengers during sharing. For the personal data of passengers, it can try to collect the location history of cars and the driving behaviour of passengers [306]. We assume that an OBU is equipped with tamper-resistant mechanisms such as a Trusted Platform Module (TPM) [312, 259], that can safely store cryptographic keys, perform cryptographic operations and validate software updates [231]. Moreover, an OBU needs to be tamper-evident [282] aiming to detect and keep irrefutable evidence when an adversary attempts to break or alter the hardware and software components of the device. We also assume that the KSMS and the CM(s) patch the software bugs [296, 201] regularly in order to preclude intrusive attacks [82, 216], which can threaten the passengers' safety [295].

MD-KSApp is untrustworthy but tamper-proof and tamper-evident. We assume that only the legitimate user of a MD-KSApp can access the KSS through the device using entity authentication mechanisms. The MD-KSApp should be equipped with security mechanisms to provide protection against unauthorised access, data breaches and malware. For instance, a MD-KSApp should be equipped with a credential management mechanisms, such as keychain tools [174], that can be used to authenticate users, encrypt data and store private keys, passwords and certificates in a secure manner. The MD-KSApp should also record, in a tamper-evident way, attempts of adversaries to disturb the operations of the MD and the KSApp.

KSMS and CM can be passive or active adversaries. The KSMS and/or the CM can try to learn and extract information about the KSS acting as honest-but-curious

adversaries. They can try to extract information about the booking preferences of users, with whom an owner is sharing her car, and with which frequency. Advancing to active adversaries, the KSMS and/or the CM can try to disturb the functionality of the KSS. For instance, they can try to collect the information exchanged within the KSS or execute only a fraction of operations honestly. However, we assume that they are not able to break the underlying cryptographic primitives.

Authorities, third parties and any external entity can be passive or active adversaries. They can try to eavesdrop and collect information exchanged within the KSS. Their aim can be to gain access to, collect and/or modify information exchanged within the KSS in an attempt to disrupt and extract information from the users and the KSS. Adversaries can be sophisticated hackers, organised crime or even governmental agencies that can be capable of taking control of a fixed set of users or any user of a KSS. However, we assume that such adversaries are not able to break the underlying cryptographic primitives.

5.6.2 Security and privacy threat analysis

This section analyses the threats to a KSS. The threat analysis is based on the STRIDE and LINDDUN frameworks.

Security threats

The STRIDE framework describes the security threats in six main classes, i.e., *spoofing*, *tampering with data*, *repudiation*, *information disclosure*, *denial-of-service*, and *elevation of privilege*.

Spoofing. An adversary can attempt to illegally access a legitimate entity of a KSS such as the MD-KSApp or the KSMS. Spoofing attacks introduce functional and trust related issues and may have an economic impact on the KSS. For instance, an adversary can try to gain a competitive advantage for a booking request to be accepted by impersonating a trusted (for the car owner) consumer such as a family member, a friend or an acquaintance. Regarding the economic implications, an adversary can attempt to benefit from making an impersonated profile to pay for the car she booked and used. Therefore, it is essential to have thorough user registration procedures and strong entity authentication mechanisms.

Tampering with data. An adversary can attempt to modify the information provided by and exchanged within the KSS such as manipulating the booking's availability period, location and profile information (e.g., type, number of seats) of a car. For instance, an adversary can try to alter the travel duration and distance covered by the shared-car aiming to affect the cost. By stating inaccurate information, an adversary can attempt to gain financial advantages and lower the credibility of users and the KSMS. Therefore, the integrity and authenticity of the messages and stored information should be guaranteed.

Repudiation. Disputes can arise when entities (do not) perform an action while stating the contrary such as by claiming inaccurate information about the travelling period and location of the car. Hence, the non-repudiation of messages exchanged and actions performed must be guaranteed and disputes must be consistently resolved.

Information disclosure. An adversary can attempt to eavesdrop messages sent through the KSS. By eavesdropping messages exchanged among the KSMS, the MD-KSApps of users, and the OBUs of cars, an adversary can attempt to retrieve critical information about the system. For instance, it can try to retrieve the access token, the booking details and the location of a car. By collecting such information, an adversary may aim to reuse the valid messages (i.e., authorisation access token) and digital key obtained to access a car, impersonating a legitimate user. Hence, confidentiality of information must be guaranteed. Moreover, information disclosure also constitutes a privacy threat to users posing additional risks, such as profiling of a user.

Denial-of-Service (DoS). Denial-of-service attack (DoS) aims to make car sharing services unavailiable to a single or a set of multiple users. For the KSS, an adversary can attempt to make any KSS *operation* unavailable to its users such as to post offers/requests, and to generate, distribute and revoke digital keys to access a car, by blocking the MD-KSApp of users, the KSMS, and the OBU of cars. Moreover, an adversary can try to perform an attack manually targeted to specific users. For instance, an adversary can attempt to gain competitive advantages by blocking offers from other users. Therefore, the KSMS should be safeguarded with network security tools, while the MD-KSApp of users and OBU of cars should be protected from malware using software security tools.

Elevation of privilege. An adversary can attempt to gain elevated access to the resources of the KSS. For instance, an elevated access can imply that an adversary can attempt to elevate her profile privileges from (i) consumer to car owner gaining unlimited access to a car, and (ii) from passenger to car driver. Moreover, an escalated

privilege at KSMS incurs that an adversary can attempt to execute operations as a system administrator aiming to retrieve information of users, alter car sharing offers/requests and bookings. In order to mitigate privilege escalation attacks, authorisation and access control mechanisms that comply with the principle of least privilege for processes and accounts of users should be deployed.

Privacy threats.

The LINDDUN framework describes the privacy threats in seven main classes, i.e., *linkability*, *identifiability*, *non-repudiation*, *detectability*, *information disclosure*, *content unawareness* and *policy and consent noncompliance*.

Linkability. An adversary can attempt to distinguish whether two or more Items of Interest (IOI) such as messages, actions and subjects are related to the same user. For instance, an adversary can try to correlate and deduce whether a user posted a car sharing request, booked a car, and drove to a particular location. Hence, unlikability among IOI must be guaranteed.

Identifiability. An adversary can attempt to correlate and identify a user from messages exchanged and actions performed. For instance, an adversary can try to identify a user by analysing the messages the user exchanged with the OBU, KSMS and MD-KSApp to access a car. Thus, the anonymity and pseudonymity of users must be preserved.

Non-repudiation. In contrast to security, non-repudiation can be used against the privacy of users. An adversary can attempt to collect evidence provided and exchanged through the KSMS and the OBU of cars aiming to deduce information about a user. It may, for example, deduce whether a user drove to a particular location (e.g., clinic). Hence, plausible deniability must be guaranteed.

Detectability. An adversary can try to distinguish the type of IOI such as messages exchanged among the KSS entities from random noise. For instance, an adversary can attempt to identify when the MD-KSApp of a user communicates with a KSMS and the OBU of cars. Thus, undetectability and unobservability of IOI must be guaranteed.

Information disclosure. An adversary can attempt to eavesdrop and collect information exchanged within the KSS. Information disclosure can affect not only

the security of the system but can also lead to privacy issues of users through *profiling* [199, 109, 101]. For instance, an adversary can attempt to learn the location and availability of a car, whether a user is absent from home and with whom a user is travelling with. Moreover, the behaviour of users may be inferred by a systematic collection and automated processing of the personal data of users by an adversary [306, 101]. For instance, an adversary may infer the (i) sharing preferences of a car owner by collecting information about their sharing patterns such as rental time, duration, and car location, (ii) free time activities of consumers by analysing the history of the pickup, drop-off, and drive locations, and (iii) circles of friends by analysing with whom, when and how often they share their cars, such as to family members, friends and acquaintances. An adversary can even attempt to infer sensitive information about users such as their health condition, by identifying users who use cars for disabled people, or regular visits to hospitals. Profiling constitutes a high risk for the privacy of users [109]. Therefore, the confidentiality of information must be guaranteed.

Content Unawareness. An adversarial KSS can attempt to collect more information than necessary about their users, violating the principle of proportionality [113] and explicit consent by the users [110]. For instance, it can use the personal data of users for unauthorised purposes such as advertisement and monetising the collected data with data brokers [267]. In contrast, the KSMS may only need to know about the eligibility of a user to drive a car without the need to collect other personal information such as date of birth, gender and the issuing country of the driving licence. In another example, the location of the car can be revealed only when necessary such as upon a misbehaving consumer exceeding the geographical restrictions agreed during the booking or when a car is not returned on time. Hence, transparency and awareness on the information dissemination of users must be guaranteed.

Policy and Consent Noncompliance. An adversarial KSS can attempt to collect, store, and process personal information of users in contrast to the principles described in the GDPR [113, 103]. For instance, an adversarial KSMS can attempt to (i) collect sensitive information about users such as sexual orientation and religion beliefs [255], (ii) export information about their users to data brokers for revenue [49], (iii) read contacts of their friends from their MDs and their Online Social Network (OSN) profiles (e.g., Facebook or Google+) [42, 290] for advertisement reasons, and (iv) not allow users to opt-out from the KSS service [113, 103]. An adversarial KSS can also attempt not to comply with the privacy policy that it advertises.

5.7 Security and privacy requirements

Based on the threat analysis, this section specifies a set of Security Requirements (SRs) and Privacy Requirements (PRs) for the proposed KSS to mitigate the identified threats.

Security requirements

To mitigate the security threats of Sect. 5.6, security measures need to be applied aiming to protect the messages exchanged, actions performed and information stored within the KSS. The relevant countermeasures are *entity authentication*, *data authentication*, *non-repudiation*, *confidentiality*, *availability*, and *authorisation* defined by the STRIDE framework, as high-level solutions.

- **SR1 – Entity Authentication.** It assures to an entity (e.g., KSMS or OBU) that the identity of a second entity (e.g., owner or consumer) is the one that is claiming to be. It aims to mitigate spoofing attacks. Entity authentication is achieved when a user proves that (i) she knows something such as passwords, (ii) possesses something such as an access token, a ticket and a specific device, (iii) has specific properties (i.e., use of biometrics), or (iv) a combination of those. Regarding the use of passwords, it is important for the KSS to support strong password policies, and that passwords are sent in an encrypted form and stored always as salted hashes within the KSS databases.
- **SR2 – Data Authentication.** It ensures that the information stored and exchanged within the KSS is authentic. For instance, such information can the booking details, the (temporal) authorisation access tokens, the digital keys and the profile of the entities involved (e.g., users and cars). It aims to mitigate tampering with data attacks. Data authentication is achieved with the use of integrity tools such as Message Authentication Code (MAC) algorithms [247] and digital signatures [256].
- **SR3 – Non-repudiation.** It is achieved when an entity (e.g., owner or consumer) is technically not able to refuse an action or transaction performed. It provides protection against false denial of origin, receipt, submission and delivery of a message [340]. For instance, a user cannot deny that she posted an offer, agreed to book a car, accessed a car, drove to a particular location and paid for using a car. It aims to mitigate repudiation attacks (i.e., disputes). Non-repudiation can be achieved with the use of digital signatures and audit trails.
- **SR4 – Confidentiality.** It ensures that only the intended users will be able to read the information stored and transferred within the KSS. It aims to mitigate information disclosure attacks. For instance, confidentiality of the information

exchanged needs to be guaranteed while the MD-KSApp, the OBU and the KSMS are communicating interchangeably. Confidentiality can be achieved with cryptographic tools such as symmetric [71], asymmetric [188] and homomorphic encryption schemes [136]. Confidentiality can also be combined with MAC algorithms when authenticated encryption is used [247].

- **SR5 – Availability.** It ensures that the resources of the KSS are available to legitimate users. It aims to mitigate DoS attacks. To safeguard availability, network tools need to be applied such as firewalls, Intrusion Detection Systems and Intrusion Prevention Systems. To protect the MD-KSApp of users and the OBU of cars, software security tools are necessary such as anti-malware tools.
- **SR6 – Authorisation.** It ensures that an entity has the access rights to read, write, and execute a set of resources of the KSS such as files and operations. It aims to mitigate elevation of privilege attacks. For authorisation, access control mechanisms need to be used such as Access-Control Lists (ACLs), or Role-Based Access Control (RBAC) mechanisms [266]. Moreover, the access control policies should follow the principle of least privilege for user accounts and processes.

Privacy requirements

To mitigate the privacy threats listed in Sect. 5.6, Privacy Enhancing Technologies (PETs) and Transparency Enhancing Technologies (TETs) need to be deployed aiming to safeguard and protect the personal data of users which will be exchanged, processed and stored within the KSS. The countermeasures relevant to privacy-threats are *unlinkability*, *anonymity*, *pseudonymity*, *plausible deniability*, *undetectability and unobservability*, *confidentiality*, *content awareness*, and *policy and consent compliance* defined by the LINDDUN framework, as high-level solutions.

- **PR1 – Unlinkability.** It ensures that two or more IOI such as messages exchanged and actions performed within the KSS, cannot be linked to the same user [245]. It aims to mitigate linkability attacks. Unlinkability can be achieved with the use of pseudonyms [225], anonymous credentials [50] and private information retrieval [22].
- **PR2 – Anonymity.** It ensures that messages exchanged and actions performed cannot be correlated to the identity of a single user, within the KSS. It aims to mitigate identifiability attacks. Anonymity can be achieved using Mix-nets [58], anonymous communication channels [308] and Secure Multiparty Computation (MPC) [68].

- **PR3 – Pseudonymity.** It ensures that pseudonyms are used instead of the real identities of the KSS users. Considering anonymity, it aims to mitigate identifiability attacks. Pseudonymity can be achieved by using Pseudo-Random Function (PRF) [166] that can generate random identifiers [32].

- **PR4 – Plausible deniability.** Unlike *non-repudiation*, it ensures that an adversarial entity is not able to trace and prove that a user has performed a specific action or operation such as driving to a particular location or booking a car for a selected period. It aims to mitigate non-repudiation privacy threats. Plausible deniability is achieved with the use of off-the-record messaging [41]. In distinctive cases and in the contrary to *plausible deniability*, the security requirement of *non-repudiation* should be provided by the KSS such as upon legal requests for car incidents. In short, there should be mechanisms to identify an action performed, a message sent or a specific user of the KSS upon a specific (abusive) incident and a corresponding legal request. However, it should be performed without violating the data protection rights of other KSS users or even the same user and for other (sharing) operations [113].

- **PR5 – Undetectability and unobservability.** It ensures that the messages exchanged and the actions performed cannot be distinguished from others by an adversary. Essentially, an adversary observes only noise and cannot extract any meaningful information by monitoring the traffic of the KSS. It aims to mitigate detectability attacks. Undetectability and unobservability can be achieved with the use of Mix-nets [56] and dummy traffic [58].

- **PR6 – Confidentiality.** Apart from security, confidentiality is also an important privacy requirement for the information provided by and exchanged within the KSS. It can be achieved using authenticated encryption [31], MPC [68] and private information searches [60].

- **PR7 – Content Awareness.** It aims to raise the awareness of users by better informing them on the amount and quality of the submitted information within the KSS. It aims to mitigate the content unawareness privacy threats. Content awareness can be achieved with the use of TETs such as privacy nudges [325], privacy dashboards [290, 228] and privacy risk metrics [193].

- **PR8 – Policy and consent compliance.** It aims to ensure the compliance of the KSS under the existing data protection legislation such as the GDPR [113]. Compliance should be achieved in advance and before users access and use such a system. It aims to mitigate the *policy and consent non-compliance* privacy threat. Policy and consent compliance can be achieved with the use of Data Protection Impact Assessment (DPIA) [107, 337] of the KSS.

5.8 Observations and key points

Although a KSS can provide benefits for users, it can also introduce several security and privacy issues. Concerning security, one of the most challenging tasks is the *access provision management* such as generation, distribution and revocation of (temporary) digital keys and authorisation access tokens for a car. The desired properties of such operations should be user, car and period specific. In short, the (temporary) digital keys and access tokens should be valid for only (i) the selected users (i.e., owner and consumers), (ii) the selected car, and (iii) the agreed period of sharing the car (i.e., booking). There are threats that can be mitigated by using techniques such as end-to-end encryption, MAC algorithms and digital signatures. However, it is important to ensure that this solution is under control of multiple parties. The owner, consumers, the car and the KSMS should be all involved in the generation and distribution process of these digital keys. Thus, any single party should not be able to generate the access keys and potentially abuse the system. Advanced cryptographic technologies such as MPC, homomorphic encryption and distributed ledgers can be used to achieve these properties.

Regarding privacy issues, protecting the users of KSS against authorised insiders such as the KSMS, is probably the most challenging task. A curious KSMS may be able to infer personal data about users by analysing (i) the booking history of consumers such as the type of cars, manufacturers and engine power, (ii) the sharing history of car owners such as the rental time, duration and car location history, and the (iii) friends circles of users by analysing how often and with whom an owner (consumer) share (use) a car such as to (with) family members, friends, acquaintances or even with people unknown to them. These privacy concerns call for PETs solutions to be used in the KSS. An example of cost-effective and commonly used PETs are pseudonyms, MPC and anonymity systems such as Tor.

Considering the system's performance, a KSS should be able to execute the necessary operations aiming to provide users with a "good" service level. Among various cryptographic possibilities, designing such a system should take into consideration performance constraints such as the communication and computational costs of the existing mechanisms. In order to satisfy all the requirements, the protocol designers should find the "right" balance between the system's security and users' privacy in combination with the KSS functionality.

5.9 Chapter summary

In this chapter, we presented a novel physical-keyless car sharing system that allows users to share their cars with others more conveniently. First, we devised a high-level

model, we described the functional requirements and defined the operations that need to be provided by such a system. Based on this model and considering the STRIDE and LINDDUN frameworks as a reference, we performed a comprehensive security and privacy threat analysis, respectively. Finally, to mitigate the identified threats, we specified a set of security and privacy requirements for such systems.

Follow-up directions. As a follow-up work to KSS analysis, we aim to design, implement and evaluate a protocol for *car access provision*.

Chapter 6

SePCAR: A Secure and privacy-enhancing protocol for car access provision

Publication Data: I. Symeonidis, A. Aly, M. A. Mustafa, B. Mennink, S. Dhooghe, and B. Preneel, "SePCAR: A Secure and Privacy-Enhancing Protocol for Car Access Provision," In 22nd European Symposium on Research in Computer Security (ESORICS 2017), Lecture Notes in Computer Science 10493, S. N. Foley, D. Gollmann, and E. Snekkenes (eds.), Springer-Verlag, pp. 475–493, 2017.

Contributions: Main author except for the security proofs (Sect. 6.5) and the performance evaluation (Sect. 6.6).

6.1 Introduction

In this chapter, we present an efficient secure and privacy-enhancing protocol for car access provision, named SePCAR. The protocol is fully decentralised and allows users to share their cars conveniently without sacrificing their security and privacy. It provides a generation, update, revocation, and distribution mechanisms for access tokens to shared cars, as well as procedures to resolve disputes and to deal with law enforcement requests, for instance in the case of car incidents. We prove that SePCAR meets its appropriate security and privacy requirements and that it is efficient: our

practical efficiency analysis through a proof-of-concept implementation shows that SePCAR takes only 1.55 seconds for a car access provision.

6.1.1 Straw-man arguments

A way to mitigate the security and privacy concerns is to implement a peer-to-peer protocol between both users and cars. The *car owner* can generate a temporary access token for her car using the car key and distribute it to the other user, the *consumer*, who can use the token to access the car. This approach has two main limitations: (i) the owner and the consumer may not trust each other, thus affecting the accountability of the system, and (ii) the owner has to have a copy of the car key on her personal device which is prone to get lost or stolen. These limitations can be overcome by having a centralised entity to generate the access token on behalf of the car owner. However, such a centralised entity will have to be fully trusted by both users, which might not be realistic under real-world scenarios. It can jeopardise the users' privacy as it will have access to booking details of users and car keys.

6.1.2 Our contributions

We design a concrete and fully decentralised secure and privacy-enhancing protocol for car access provision, named SePCAR. The protocol provides generation and distribution of access tokens for car access provision, as well as update and revocation operations to facilitate mutually agreed modifications of the booking details and protecting against misbehaving consumers, respectively. It internally uses secure multiparty computation to facilitate forensic evidence provision in the case of car incidents or at the request of law enforcement. SePCAR is described in detail in Sect. 6.4.

We prove that the protocol fulfils the desired security and privacy requirements. First, departing from Chapter 5, we give a detailed list of security and privacy requirements in Sect. 6.2. Then, in Sect. 6.5, we prove that SePCAR meets its security and privacy requirements as long as its underlying cryptographic primitives (listed in Sect. 6.3) are secure. Our theoretical complexity and practical efficiency analysis in Sect. 6.6 demonstrates SePCAR's competitiveness. In particular, we implemented a prototype as a proof-of-concept in C and we achieved a car access provision in ≈ 1.55 seconds.

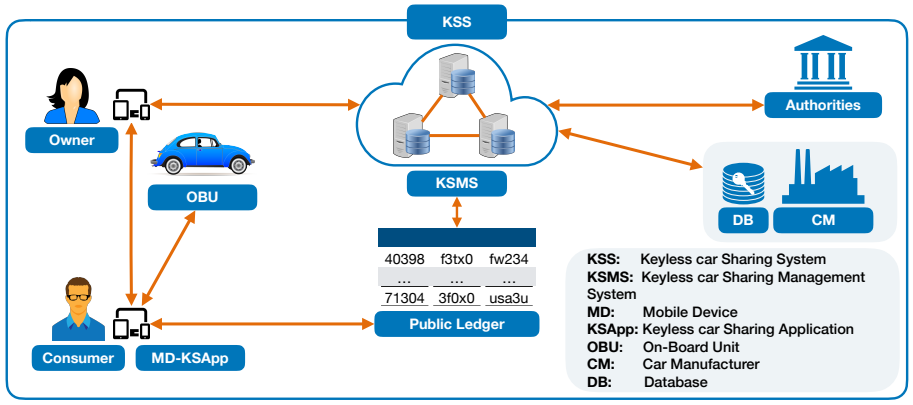


Figure 6.1: System model of a physical-Keyless car Sharing System (KSS) [288].

6.2 System model and requirements

We describe the system model and functionalities of a physical-Keyless car Sharing System (KSS). Moreover, we specify the threat model, the security, privacy and functional requirements which it needs to satisfy, and our assumptions about the system.

6.2.1 System model

We follow the KSS system model of Chapter 5 (see also Fig. 6.1). *Users* are individuals who are willing to share their cars, *owners* (u_o), and use cars which are available for sharing, *consumers* (u_c); both use of Mobile Devices (MDs) such as smart phones. An On-Board Unit (OBU) is an embedded or a standalone hardware/software component [171] that is part of the secure access management system of a *car*. It has a wireless interface such as Bluetooth, Near-Field Communication (NFC) or Long-Term Evolution (LTE) communication. The Car Manufacturer (CM) is responsible for generating and embedding a digital-key into each car. These keys are used for car sharing and are stored in the manufacturers' Data Base (DB). The Keyless car Sharing Management System (KSMS) is a set of Secure Multiparty Computation (MPC) *servers* that assists owners with car access token generation, distribution, update and revocation. Each *server* individually retrieves its share of the car key, K^{car}, and the servers jointly encrypt the booking details, M^B, to generate an access token, AT^{car}. The access token is published on a Public Ledger (PL), which serves as a public bulletin board that guarantees the integrity of the data [212]. While a bulletin board PL can be part of the KSMS, we will refer to them distinctively from one to another for improved readability

of the protocol description (SePCAR). The booking details are typically agreed upon by owner and consumer prior to the beginning of the protocol.

6.2.2 Threat model

Within the KSS, the KSMS, the CM and the PL are considered honest-but-curious entities. They will perform the protocol honestly, but they are curious to extract private information about users. Owners are passive adversaries while consumers and outsiders are malicious. The OBU of a car is trusted and equipped with a Trusted Platform Module (TPM) [312, 259] that supports secure key storage and cryptographic operations such as symmetric and public key encryption, following the EVITA [116, 336] and PRESERVE [248, 243] specifications. The MD of users are untrusted as they can get stolen, lost or broken.

6.2.3 Protocol design requirements

The keyless car sharing system should satisfy the following Security Requirements (SRs), Privacy Requirements (PRs) and Functional Requirements (FRs). Here, we recall that M^B refers to the booking details, AT^{car} the access token to the car and K^{car} the car key. Note that a high-level description of the requirements was presented at Chapter 5.

- **SR1 – Confidentiality of M^B.** No one but the shared car, u_o and u_c should have access to M^B.
- **SR2 – Authenticity of M^B.** The shared car should verify the origin and integrity of M^B from u_o.
- **SR3 – Confidentiality of AT^{car}.** No one but the shared car and u_c should have access to AT^{car}.
- **SR4 – Confidentiality of K^{car}.** No one but the shared car and the CM should have access to K^{car}.
- **SR5 – Backward and forward secrecy of AT^{car}.** Compromise of a key used to encrypt any AT^{car} should not compromise other tokens (future and past) published on the PL of any honest u_c.
- **SR6 – Non-repudiation of origin of AT^{car}.** The u_o should not be able to deny it has agreed to the terms of M^B, and participated in providing the respective AT^{car}.

- **SR7 – Non-repudiation of delivery of AT^{car}.** The u_c should not be able to deny it has obtained and used the AT^{car} to open the car (once she has done so).
- **PR1 – Unlinkability of u_c and the car.** No one but the shared car, u_o and u_c should be able to link two booking requests of the same u_c for the car.
- **PR2 – Anonymity of u_c and the car.** No one but the shared car, u_o and u_c should learn the identity of u_c and the car.
- **PR3 – Undetectability of AT^{car} operation.** No one but the shared car, u_o and u_c (if necessary) should be able to distinguish between AT^{car} generation, update and revocation.
- **PR4 – Forensic evidence provision.** The KSMS should be able to provide authorities with the transaction details of an access provision to a car at the request of law enforcement without violating the other users' privacy.
- **FR1 – Offline authentication.** Access provision should be provided for locations where cars have limited (or no) network connection.

6.2.4 Assumptions

For SePCAR, we assume that before every car access provision, the booking details are agreed upon by owner and consumer, but that both keep these booking details confidential against external parties. SePCAR relies on a Public-Key Infrastructure (PKI) [248], and we assume that each entity has her private/public key pair with their corresponding digital certificates. The communication channels are secure and authenticated among entities using SSL/TLS [86] and NFC [17]. The OBU is equipped with a TPM [312, 259], and it is designed to resist deliberate or accidental physical destruction (i.e., black box). The MPC servers are held by non-colluding organisations, i.e., organisations with conflicting interests such as authorities, car owner unions and car manufacturers.

6.3 Cryptographic building blocks

This section specifies, the cryptographic functionalities that are used across this chapter, as well as the MPC functionalities and cryptographic building blocks.

6.3.1 Cryptographic functionalities

SePCAR uses the following cryptographic building blocks. The suggested instantiations are the ones used in our proof-of-concept implementation.

- $\sigma \leftarrow \mathsf{sign}(Sk, m)$ and $\mathsf{true}/\mathsf{false} \leftarrow \mathsf{verify}(Pk, m, \sigma)$ are public-key operations for signing and verification respectively. These can be implemented using RSA as defined in the PKCS #1 v2.0 specifications [177].
- $z \leftarrow \mathsf{prf}(K, counter)$ is a Pseudo-Random Function (PRF) that uses as input a key and a counter. This function can be implemented using CounTeR mode (CTR) with AES (as the message input is small).
- $c \leftarrow \mathsf{enc}(Pk, m)$ and $m \leftarrow \mathsf{dec}(Sk, c)$ are public-key encryption and decryption functions. These can be implemented using RSA as defined in the RSA-KEM specifications [19, 219].
- $c \leftarrow \mathsf{E}(K, m)$ and $m \leftarrow \mathsf{D}(K, c)$ are symmetric key encryption and decryption functions. These can be implemented using CTR mode with AES.
- $v \leftarrow \mathsf{mac}(K, m)$ is a symmetric key MAC function. This function can be implemented using CBC-MAC with AES.[1]
- $z \leftarrow \mathsf{hash}(m)$ is a cryptographic hash function. This function can be implemented using SHA-512.

We will furthermore use the notation $z \leftarrow \mathsf{query}(x, y)$ to denote the retrieval of the xth value from the yth database *DB* (to be defined in Sect. 6.4), and $z \leftarrow \mathsf{query\_an}(y)$ to denote the retrieval of the yth value from the PL through an anonymous communication channel such as Tor [308], aiming to anonymously retrieve a published record submitted using the $\mathsf{publish}(y)$ function.

6.3.2 Secure multiparty computation

Ben-or et al. [33] (commonly referred to as BGW) proved that it is possible to calculate any function with perfect security in the presence of active and passive adversaries under the information-theoretic model, as long as there is an honest majority: $1/2$ for passive and $2/3$ for active adversaries. The former can be achieved by assuming the use of private channels among the servers and the latter using Verifiable Secret Sharing (VSS).

[1] CBC-MAC is proven to be secure as long as it is *only* evaluated on equal-size messages (or on prefix-free messages) [30], which is the case for SePCAR. For variable length messages, one should resort to *encrypted* CBC-MAC or replace the key for the last block CMAC [168].

Our protocol is *MPC-agnostic*, meaning that it does not depend on the solution that implements the MPC functionality; example protocols that could be executed within our protocol are SPDZ [75] or MASCOT [178]. However, the three-party protocol for Boolean circuits that was introduced by Araki et al. [16] is fairly suited for our current needs, given its performance and threshold properties. Hence, we use this protocol in our simulation. It can perform non-linear operations with relatively high throughput and somewhat low latency (when tested on 10 Gbps connections). The scheme provides threshold security against semi-honest and malicious parties. Note that Furukawa et al. [134] further adapt the protocol [16] to provide security against a malicious adversary.

On an incremental setup for KSMS. Our protocol can support an incremental setup and deployment where an $(l > 2)$-case of KSMS servers is trivial, e.g., using Ben-Or, Goldwasser and Wigderson (BGW) [33]. The 2-party case setting could also be achieved with MPC protocols such as SPDZ [75], however, the forensic properties would require that all the shares would be handed out to a single party (i.e., authorities).

6.3.3 Multiparty computation functionalities

SePCAR uses the following cryptographic functionalities for MPC:

- $[x] \leftarrow \mathsf{share}(\mathsf{x})$ is used to secretly share an input. This function can be instantiated using Araki et al.'s sharing functionality.
- $x \leftarrow \mathsf{open}([x])$ reconstructs the private input based on the secret shares.
- $[z] \leftarrow \mathsf{XOR}([x],[y])$ outputs a secret shared bit, representing the XOR of secret shared inputs $[x]$ and $[y]$. Note that for both arithmetic or Boolean circuits, such functionality could be implemented without requiring any communication cost.
- $[z] \leftarrow \mathsf{AND}([x],[y])$ outputs a secret shared bit, representing the AND of two secret shared inputs $[x]$ and $[y]$. This function can be instantiated using Araki et al.'s AND operation.
- $[z] \leftarrow \mathsf{eqz}([x],[y])$ outputs a secret shared bit, corresponding to an equality test of two secret shared inputs $[x]$ and $[y]$. This is equivalent to computing $[z] \leftarrow [x] \stackrel{?}{=} [y]$ where $z \in \{0,1\}$.
- $[C] \leftarrow \mathsf{E}([K],[M])$ secretly computes a symmetric encryption from a secret shared key $[K]$ and a secret shared message $[M]$. We include a succinct review on how to implement AES below.

- $[V] \leftarrow \mathsf{mac}([K],[M])$ secretly computes a MAC from a secret shared key $[K]$ and a secret shared message $[M]$.

On the secure equality test. Various protocols have been proposed to implement the equality tests (previously referred to an eqz functionality). Common approaches provide either constant rounds or a logarithmic number of them in the bit size of its inputs, which could be proven more efficient for sufficiently small sizes. Furthermore, they also offer different security levels, i.e., perfect or statistical security [54, 72, 192]. In this work we assume the use of any logarithmic depth construction, which matches the current state of the art.

On AES over MPC. The block-cipher AES has been the typical functionality for bench-marking MPC protocols during the last few years; this has resulted in faster and leaner MPC implementations of the cipher. As the MPC parties hold a secret shared key K and a secret shared message M, the outcome of the operation is a secretly shared AES encrypted ciphertext [12, 73, 74, 144]. Note that we assume the use of the methods proposed by Damgård and Keller [73] with some minor code optimisations.

6.4 SePCAR

This section provides a detailed description of SePCAR. For simplicity and without loss of generality, we consider a single owner, consumer and shared car. The description straightforwardly scales to a larger set of owners, consumers, and cars. Table 6.1 lists the notation used in this chapter and Fig. 6.2 illustrates the high-level overview of SePCAR.

SePCAR consists of four steps: *session keys generation and data distribution*, *access token generation*, *access token distribution and verification* and *car access*. We will discuss these steps in detail in the remainder of the section, with an overview picture given in Fig. 6.8. We first discuss a few *prerequisite* steps which have to be performed. After the discussion of the fourth (and last) step, we complete the section with an overview of the possible operations after SePCAR: *access token update and revocation*.

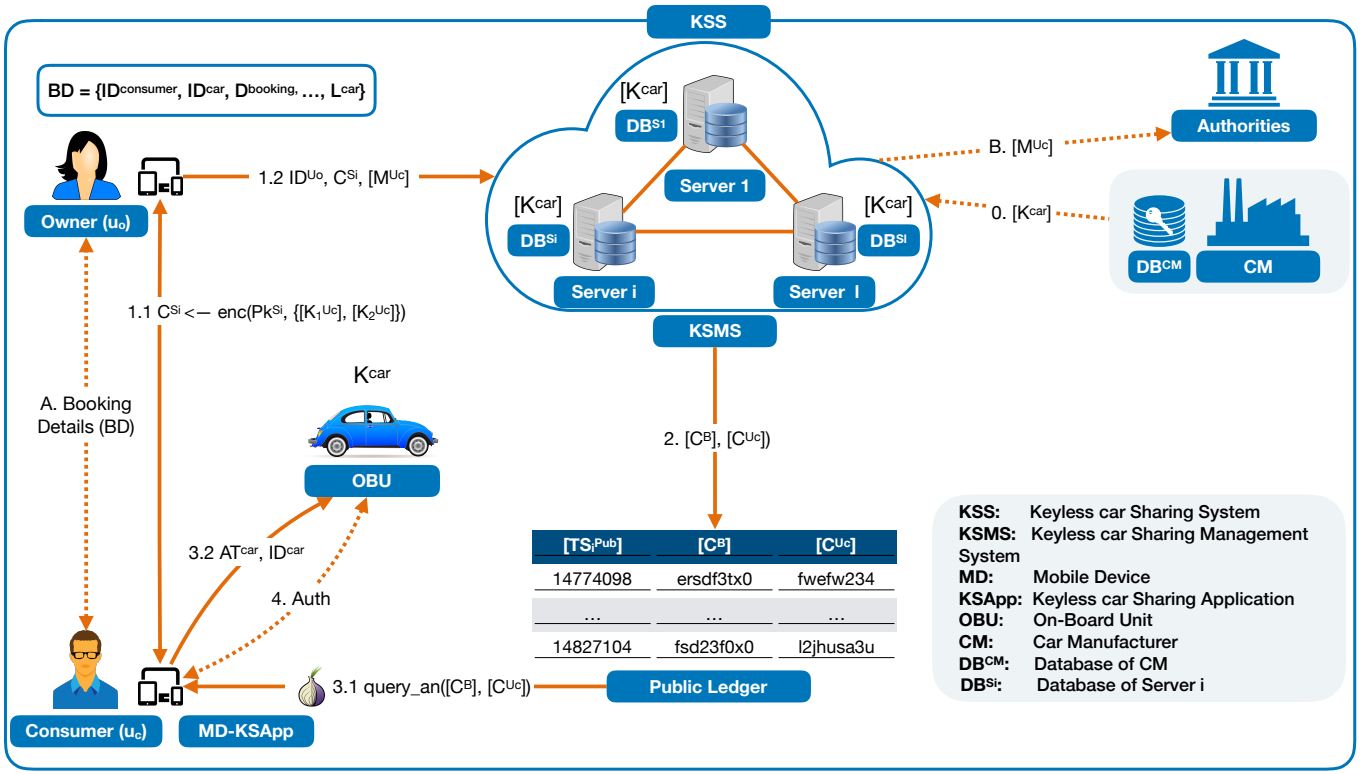


Figure 6.2: SePCAR high level overview.

$$DB^{\mathsf{CM}} = \begin{pmatrix} ID_1^{u_o} & ID_1^{car_{u_o}} & K_1^{car_{u_o}} \\ \vdots & \vdots & \vdots \\ ID_x^{u_o} & ID_y^{car_{u_o}} & K_y^{car_{u_o}} \\ \vdots & \vdots & \vdots \\ ID_m^{u_o} & ID_n^{car_{u_o}} & K_n^{car_{u_o}} \end{pmatrix} \qquad DB^{\mathsf{S}_i} = \begin{pmatrix} ID_1^{u_o} & [ID_1^{car_{u_o}}] & [K_1^{car_{u_o}}] \\ \vdots & \vdots & \vdots \\ ID_x^{u_o} & [ID_y^{car_{u_o}}] & [K_y^{car_{u_o}}] \\ \vdots & \vdots & \vdots \\ ID_m^{u_o} & [ID_n^{car_{u_o}}] & [K_n^{car_{u_o}}] \end{pmatrix}$$

Figure 6.3: The DB of CM (left) and the DB of the ith server S_i (right).

6.4.1 Prerequisites

Before SePCAR can commence, two prerequisite steps need to take place: *car key distribution* and setting the details for the *car booking*.

Car key distribution takes place immediately after the xth owner, $ID_x^{u_o}$, has registered her yth car, $ID_y^{car_{u_o}}$, with the KSMS. The KSMS forwards $ID_y^{car_{u_o}}$ to the CM to request the symmetric key $K_y^{car_{u_o}}$ of the car. The CM retrieves $K_y^{car_{u_o}}$ from its DB, DB^{CM} and generates ℓ secret shares of $K_y^{car_{u_o}}$ and $ID_y^{car_{u_o}}$, denoted by $[K_y^{car_{u_o}}]$ and $[ID_y^{car_{u_o}}]$, respectively. Then, it forwards each share to the corresponding KSMS server, i.e., S_i. Upon receipt of the shares, each S_i stores ID^{u_o} together with the shares $[ID_y^{car_{u_o}}]$ and $[K_y^{car_{u_o}}]$ in its local DB, DB^{S_i}. The representations of the DB of CM and S_i are shown in Fig. 6.3. For simplicity, in some parts of SePCAR we will use ID^{u_o}, ID^{car} and K^{car} instead of $ID_x^{u_o}$, $ID_y^{car_{u_o}}$ and $K_y^{car_{u_o}}$.

Car booking allows u_o and u_c to agree on the booking details, i.e., $M^B = \{\mathsf{hash}(Cert^{u_c}), ID^{car}, L^{car}, CD^{u_c}, AC^{u_c}, ID^B\}$, where $\mathsf{hash}(Cert^{u_c})$ is the hash of the digital certificate of u_c, L^{car} is the pick-up location of the car, CD^{u_c} is the set of conditions under which u_c is allowed to use the car (e.g., restrictions on locations, time period), AC^{u_c} are the access control rights under which u_c is allowed to access the car and ID^B is the booking identifier. Recall that it is assumed that an owner and a consumer agree on the booking details beforehand.

6.4.2 Step 1: Session keys generation and data distribution

The consumer u_c generates two symmetric session keys, $K_1^{u_c}$ and $K_2^{u_c}$. Key $K_1^{u_c}$ will be used by each S_i to encrypt the access token, such that only u_c has access to it. $K_2^{u_c}$ will be used to generate an authentication tag which will allow u_c to verify that the access token contains M^B which was agreed during the *car booking*. In addition, u_o sends the necessary data to each S_i, such that the access token can be generated. In detail, as shown in Fig. 6.4, u_o sends a session-keys-generation request, *SES_K_GEN_REQ*, along with ID^B to u_c. Upon receipt of the request, u_c generates $K_1^{u_c}$ and $K_2^{u_c}$ using the

Table 6.1: SePCAR notation.

Symbol	Description
KSMS, S_i	Set of KSMS servers, the ith server for $i \in \{1 \dots l\}$
PL, CM	Public Ledger, Car Manufacturer
u_o, u_c	owner, consumer
ID^B,ID^{u_o},ID^{u_c}, ID^{car}	ID of booking, u_o, u_c, car
CD^{u_c}/AC^{u_c}, L^{car}	Set of conditions/access rights under which u_c is allowed to access a car, car's location
DB^{CM} / DB^{S_i}	Database that CM holds with (ID^{u_o}, $ID^{car_{u_o}}$, $K^{car_{u_o}}$) / that S_i holds with (ID^{u_o}, $[ID^{car_{u_o}}]$, $[K^{car_{u_o}}]$) for all owners (u_o's) and their registered cars
$\vec{D}^{u_o}$	Car records ($ID_x^{u_o}$, $[ID_y^{car_{u_o}}]$, $[K_y^{car_{u_o}}]$) of the xth u_o for the yth car extracted (query) from DB^{S_i}, where $\lvert\vec{D}^{u_o}\rvert = n$
$\vec{D}^{car}$	The matched (eqz output) yth car key $(\overset{1}{[0]}\cdots[0]\overset{y}{[1]}[0]\cdots\overset{n}{[0]})$, where $\lvert\vec{D}^{car}\rvert = n$
Pk^x / Sk^x, $Cert^{u_c}$	Public/private key pair of the KSS entity x, certificate of u_c
M^B	Booking details, i.e, $\{\mathsf{hash}(Cert^{u_c}), ID^{car}, L^{car}, CD^{u_c}, AC^{u_c}, ID^B\}$
σ^{u_o}, σ^{car}_{Access}	Signature (sign output) of M^B with Sk^{u_o}, and $\{M^B, TS^{car}_{Access}\}$ with Sk^{car}
K^{car}, K^{u_c}, $K_1^{u_c}$/$K_2^{u_c}$	Symmetric key of the car, u_c's master key, u_c's session keys generated be (prf output) K^{u_c} and $counter$/$counter+1$
M^{u_c}, AT^{u_c}	Concatenation of M^B with σ^{u_o}, a secure access token as the encryption (E output) of M^{u_c} with K^{car}
C^{S_i}	Ciphertext (enc output) of session keys $\{[K_1^{u_c}], [K_2^{u_c}]\}$ with Pk^{S_i}
$[C^{u_c}]$	Ciphertext (E output) of $\{[AT^{u_c}], [ID^{car}]\}$ with $[K_1^{u_c}]$
C^B, $[C^B]$	Message digest (mac output) of M^B with $K_2^{u_c}$, and $[M^B]$ with $[K_2^{u_c}]$
TS_i^{Pub}, TS^{car}_{Access}	Time-stamp of u_c accessing the shared car, a record published (publish) on the PL submitted by S_i

$\mathsf{prf}()$ function instantiated by u_c's master key, i.e., K^{u_c} and $counter$ and $counter+1$. Then, u_c transforms these into ℓ secret shares, $[K_1^{u_c}]$ and $[K_2^{u_c}]$, one for each S_i in such a way that none of the servers will have access to the keys but that they can jointly evaluate functions using these keys securely. Then, it encrypts $[K_1^{u_c}]$ and $[K_2^{u_c}]$ with the public-key of each S_i, $C^{\mathsf{S}_i} = \mathsf{enc}(Pk^{\mathsf{S}_i}, \{[K_1^{u_c}], [K_2^{u_c}]\})$, such that only the corresponding S_i can access the corresponding shares. Finally, u_c forwards to u_o an acknowledgement message *SES_K_GEN_ACK* along with ID^B and $\{C^{\mathsf{S}_1}, \dots, C^{\mathsf{S}_l}\}$.

While waiting for the response of u_c, the owner u_o signs M^B with her private key, i.e., $\sigma^{u_o} = \mathsf{sign}(Sk^{u_o}, M^B)$. In a later stage, the car will use σ^{u_o} to verify that M^B has been

approved by u_o. Then u_o transforms $M^{u_c} = \{M^B, \sigma^{u_o}\}$ into ℓ secret shares, i.e., $[M^{u_c}]$. Upon receipt of the response of u_c, u_o forwards to each S_i an access-token-generation request, *AT_GEN_REQ*, along with ID^{u_o}, the corresponding C^{S_i} and $[M^{u_c}]$.

6.4.3 Step 2: Access token generation

The servers generate an access token and publish it on the PL. In detail, as shown in Fig. 6.5, upon receipt of *AT_GEN_REQ* from u_o, each S_i uses the ID^{u_o} to extract $[K^{car}]$ from DB^{S_i} as follows. Initially, each S_i uses ID^{u_o} to retrieve the list of identities of all cars and car key shares related to the set of records that correspond to u_o. The result is stored in a vector $\vec{D}^{u_o}$ of size $n \times 3$, i.e.,

$$\vec{D}^{u_o} = \begin{pmatrix} ID^{u_o} & [ID_1^{car_{u_o}}] & [K_1^{car}] \\ \vdots & \vdots & \vdots \\ ID^{u_o} & [ID_y^{car_{u_o}}] & [K_y^{car}] \\ \vdots & \vdots & \vdots \\ ID^{u_o} & [ID_n^{car_{u_o}}] & [K_n^{car}] \end{pmatrix},$$

where n is the number of cars which u_o has registered with the KSS.

To retrieve the record for the car to be shared, each S_i extracts $[ID^{car}]$ from $[M^{u_c}]$ and performs a comparison with each of the n records of $\vec{D}^{u_o}$ using the $\mathsf{eqz}()$ function. The comparison results in 0 for mismatch and 1 for identifying the car at position y. The result of each iteration is stored in a vector $\vec{D}^{car}$ of length n, i.e.,

$$\vec{D}^{car} = \left(\overset{1}{[0]} \cdots [0] \overset{y}{[1]} [0] \cdots \overset{n}{[0]} \right) .$$

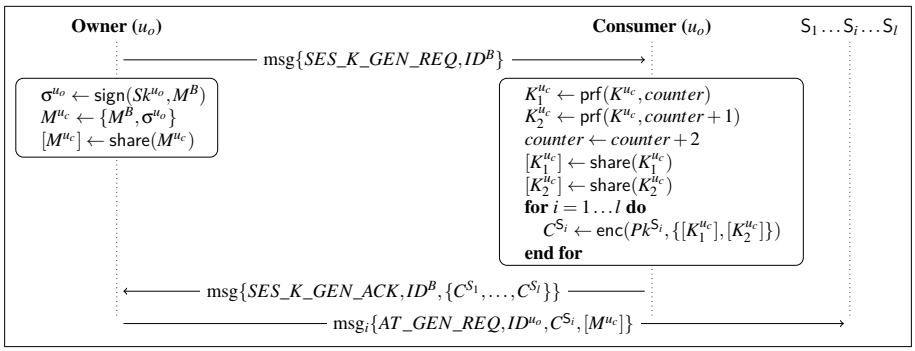


Figure 6.4: SePCAR step 1: Session keys generation and data distribution.

Each S_i then multiplies $\vec{D}^{car}$ and $\vec{D}^{u_o}$ to generate a third vector of length 3, i.e.,

$$\vec{D}^{car} \times \vec{D}^{u_o} = \left(ID^{u_o} \; [ID_y^{car_{u_o}}] \; [K_y^{car_{u_o}}] \right) \; ,$$

from which the share of the secret-key $[K^{car}]$ of a car can be retrieved. Then, the KSMS servers S_i collaboratively encrypt $[M^{u_c}]$ using the retrieved $[K^{car}]$ to generate an access token for the car in shared form, $[AT^{car}]$.

As AT^{car} and ID^{car} need to be available only to u_c, a second layer of encryption is performed using $K_1^{u_c}$. To retrieve the shares of the session keys, $\{[K_1^{u_c}], [K_2^{u_c}]\}$, each S_i decrypts C^{S_i} using its private key. Then, the servers encrypt $[AT^{car}]$ and $[ID^{car}]$ with $[K_1^{u_c}]$ to generate $[C^{u_c}]$. In addition, they generate an authentication tag, $[C^B]$, using the $\mathsf{mac}()$ function with $[K_2^{u_c}]$ and $[M^B]$ as inputs. Finally, each S_i sends to PL an access-token-publication request, *AT_PUB_REQ*, along with $[C^B]$ and $[C^{u_c}]$.

6.4.4 Step 3: Access token distribution and verification

The PL publishes the shares of the encrypted access token which are then retrieved by u_c. Once retrieved, u_c can obtain the access token and use it to access the car. In detail, as shown in Fig. 6.6, upon receipt of *AT_PUB_REQ*, PL publishes $[C^B]$, $[C^{u_c}]$ and TS^{Pub}, which is the time-stamp of the publication of the encrypted token. Then PL sends an acknowledgement of the publication, *AT_PUB_ACK*, along with TS_i^{Pub} to at least one S_i which forwards it to u_o who, in turn, forwards it to u_c.

Upon receipt of *AT_PUB_ACK*, u_c uses TS_i^{Pub} and the $\mathsf{query\_an}()$ function to anonymously retrieve $[C^{u_c}]$ and $[C^B]$ from PL, such that PL cannot identify u_c. Then, u_c uses the $\mathsf{open}()$ function to reconstruct C^B and C^{u_c} using the retrieved shares. Next, u_c verifies the authentication tag C^B locally using the $\mathsf{mac}()$ function with $K_2^{u_c}$ and M^B

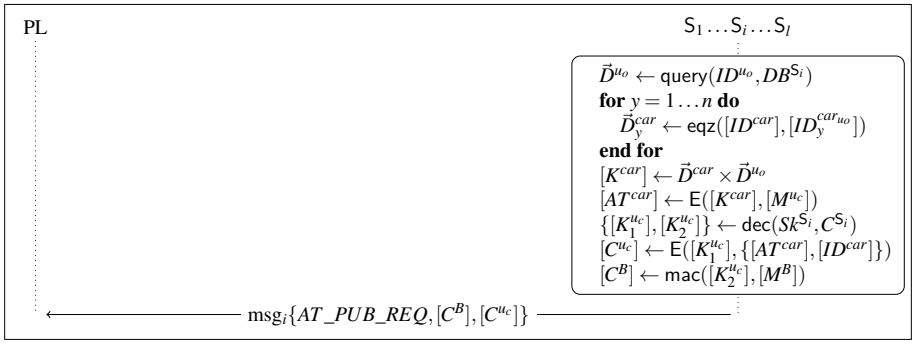


Figure 6.5: SePCAR step 2: Access token generation.

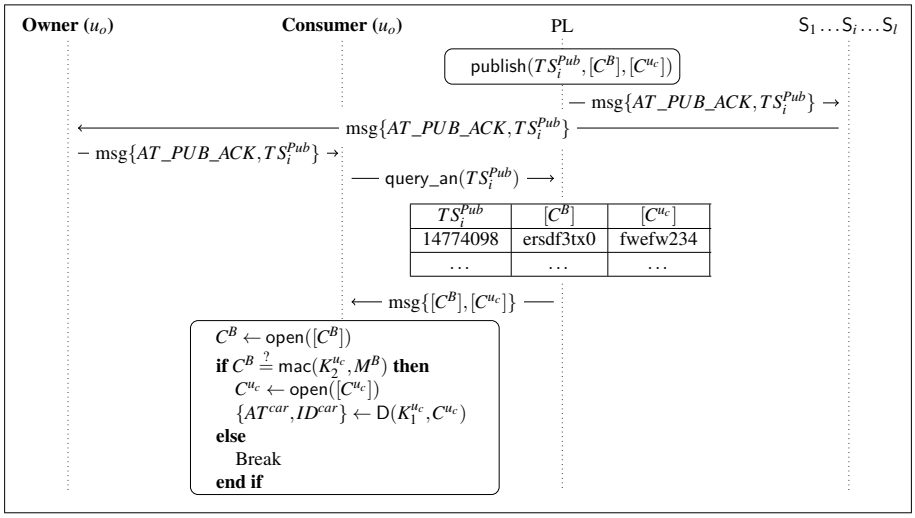


Figure 6.6: SePCAR step 3: access token distribution and verification.

as inputs. In the case of successful verification, u_c is assured that the token contains the same details as the ones agreed during *car booking*. Then, u_c decrypts C^{u_c} using $K_1^{u_c}$ to obtain the access token and the car identity, $\{AT^{car}, ID^{car}\}$.

6.4.5 Step 4: Car access

The consumer uses the access token to obtain access to the car. In detail, u_c sends $\{AT^{car}, ID^{car}, Cert^{u_c}\}$ to the car using a close range communication channel such as NFC or Bluetooth (see Fig. 6.7). Upon receipt, the OBU of the car obtains $M^{u_c} = \{M^B, \sigma^{u_o}\}$ by decrypting AT^{car} with K^{car}. It then performs three verifications. It checks if the access attempt satisfies the conditions specified in M^B. Then, it verifies σ^{u_o} to be assured that the booking details M^B have not been modified and have been indeed approved by the car owner. Finally, it verifies the identity of u_c. For the last verification, as the OBU receives $Cert^{u_c}$ (along with the hash($Cert^{u_c}$) in M^B), it can use any challenge-response protocol based on public/private key [87] and RFIDs [88]. If any of these verifications fails, the OBU terminates the car access process and denies access to the car. Otherwise, it grants u_c access to the car, signs $\{M^B, TS_{Access}^{car}\}$, where TS_{Access}^{car} is the time-stamp of granting the access and asynchronously sends msg$\{\sigma_{Access}^{car}, TS_{Access}^{car}\}$ to u_o.

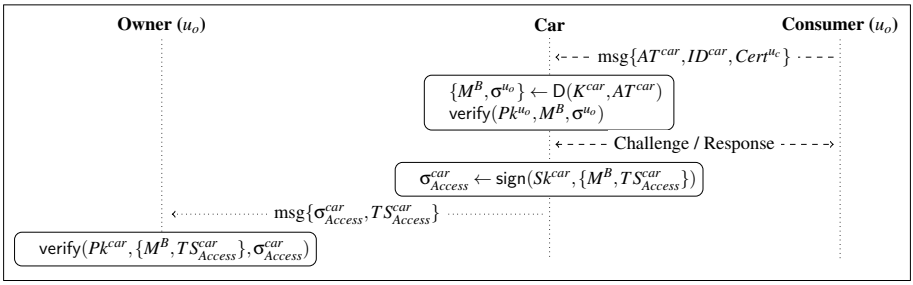


Figure 6.7: SePCAR step 4: car access. Dashed lines represent close range communication.

6.4.6 Access token update and revocation

Upon an agreement between u_o and u_c to update or revoke an access token, SePCAR can be performed as described in steps 1-3. The values of an update request can be changed according to new booking details, $\hat{M}^B$, whereas for revocation, each of the parameters in $\hat{M}^B$ can receive a predefined value indicating the revocation action. However, there are occasions when u_o may need to enforce an update or revocation of an access token. To prevent u_c from blocking such operations, SePCAR should be executed only by u_o, without the involvement of u_c. More specifically, u_o generates session keys, requests an access token, queries the PL, and sends the token to the car using long range asynchronous communication channel such as LTE.

6.5 Security and privacy analysis

We prove that SePCAR satisfies the security and privacy requirements of Sect. 6.2, provided that its underlying cryptographic primitives are sufficiently secure. The theorem statement and the proof given below are informal; a formal description of the security models and the stand-alone proof are given in Appendix A.2.

Theorem 1. *If communication takes place over private channels, the MPC is statistically secure,*

- *the signature scheme* sign *is multi-key existentially unforgeable [142],*
- *the pseudo-random function* prf *is multi-key secure [140],*
- *the public-key encryption scheme* enc *is multi-key semantically secure [28],*
- *the symmetric key encryption scheme* E *is multi-key chosen-plaintext secure [29],*

- *the MAC function* mac *is multi-key existentially unforgeable [142], and*
- *the hash function* hash *is collision resistant [258],*

then SePCAR fulfils the security and privacy requirements of Sect. 6.2.

Note that, indeed, for each of the keyed cryptographic primitives we require security in the *multi-key* setting, as these are evaluated under different keys. For example, sign is used by all owners, each with a different key; enc is used for different keys, each for a different party in the KSMS, and E and mac are used for independent keys for every fresh evaluation of the protocol. We refer to Bellare et al. [28] for a discussion on generalizing semantic security of public-key encryption to multi-key security; the adaptation straightforwardly generalizes to the other security models.

sketch. We describe the security and privacy requirements and analyse how these are achieved from the cryptographic primitives, separately. We recall that consumer and owner have agreed upon the booking details prior to the evaluation of SePCAR, hence they know each other.

SR1 – Confidentiality of M^B. In one evaluation of the protocol, u_c, u_o, and the shared car learn the booking details by default or design. The KSMS servers only learn shares of the booking data, and under the assumption that the MPC is statistically secure, nothing about the booking data is revealed during the MPC. The outcomes of the MPC are C^B and C^{u_c} satisfying

$$C^B = \mathsf{mac}(K_2^{u_c}, M^B), \tag{6.1}$$

$$C^{u_c} = \mathsf{E}(K_1^{u_c}, \{\mathsf{E}(K_y^{car_{u_o}}, \{M^B, \sigma^{u_o}\}), ID^{car}\}), \tag{6.2}$$

both of which reveal nothing about M^B to a malicious outsider due to the assumed security of mac, E, and the independent uniform drawing of the keys $K_1^{u_c}$ and $K_2^{u_c}$. The nested encryption E does not influence the analysis due to the mutual independence of the keys $K_1^{u_c}$ and $K_y^{car_{u_o}}$.

SR2 – Authenticity of M^B. An owner who initiates the access token generation and distribution, first signs the booking details using its private key before sending those to the KSMS in shares. Therefore, once the car receives the token and obtains the booking details, it can verify the signature of the owner on the booking details. In other words, the car can verify the source of the booking details, the owner and their integrity. Suppose, to the contrary, that a malicious consumer can get access to a car of an owner u_o. This particularly means that it created a tuple (M^B, σ^{u_o}) such that

$\mathsf{verify}(Pk^{u_o}, M^B, \sigma^{u_o})$ holds. If σ^{u_o} is new, this means that u_c forges a signature for the secret signing key Sk^{u_o}. This is impossible by assumption that the signature scheme is existentially unforgeable. On the other hand, if (M^B, σ^{u_o}) is old but the evaluation is fresh, this means a collision $\mathsf{hash}(Cert^{u_c}) = \mathsf{hash}(Cert^{u_c\prime})$, which is computationally infeasible as the hash function is collision resistant.

SR3 – Confidentiality of AT^{car}. The access token is generated by the KSMS servers obliviously (as the MPC is statistically secure) and only revealed to the public in encrypted form, through C^{u_c} of (A.5). Due to the uniform drawing of the key $K_1^{u_c}$ (and the security of the public-key encryption scheme used to transmit this key), only the legitimate user can decrypt and learn the access token. It shares it with the car over a secure and private channel.

SR4 – Confidentiality of K^{car}. Only the car manufacturer and the car itself hold copies of the car key. The KSMS servers learn these in shared form, hence learn nothing about it by virtue of the statistical security of the MPC. Retrieving a car key from encryptions made under this key constitutes a key recovery attack, which in turn allows to break the chosen-plaintext security of the symmetric key encryption scheme.

SR5 – Backward and forward secrecy of AT^{car}. The access token is published on the public ledger as C^{u_c} of (A.5), encrypted under symmetric key $K_1^{u_c}$. Every honest consumer generates a fresh key $K_1^{u_c}$ for every new evaluation, using a pseudo-random function prf that is secure, i.e., that is indistinguishable from a random function. This implies that all session keys are drawn independently and uniformly at random. In addition, the symmetric encryption scheme E is multi-key secure. Concluding, all encryptions C^{u_c} are independent and reveal nothing about each other. (Note that nothing can be said about access tokens for malicious users who may deviate from the protocol and reuse one-time keys.)

SR6 – Non-repudiation of origin of AT^{car}. The car, that is a trusted entity, verifies the origin through verification of the signature, $\mathsf{verify}(Pk^{u_o}, M^B, \sigma^{u_o})$. The consumer u_c verifies the origin through the verification of the Message Authentication Code (MAC) algorithm, $C^B \stackrel{?}{=} \mathsf{mac}(K_2^{u_c}, M^B)$. Note that the consumer does not effectively verify AT^{car}, but rather C^B, which suffices under the assumption that the MPC servers evaluate their protocol correctly. In either case, security fails only if the asymmetric signature scheme or the MAC function are forgeable.

SR7 – Non-repudiation of delivery of AT^{car}. The owner can verify correct delivery through the verification of the message sent by the car to the owner, $\mathsf{verify}(Pk^{car}, \{M^B, TS^{car}_{Access}\}, \sigma^{car}_{Access})$ at the end of the protocol. The security breaks only if the signature scheme is forgeable.

PR1 – Unlinkability of u_c and the car. The only consumer-identifiable data is in the consumer's certificate included in the booking details. Note that these are agreed upon between the consumer and the owner, so the owner learns the identity of the consumer by default. Beyond that, the consumer only communicates with the car, which is supposed to learn the identity of the consumer so that it can perform proper access control. The consumer consults the PL over an anonymous channel. The booking details are transferred to and from the KSMS, but these are secretly shared in M^{u_c} and encrypted with $K_1^{u_c}$, K^{car} using AES; hence, they do not leak by virtue of their confidentiality (security requirement SR1).

PR2 – Anonymity of u_c and the car. The reasoning is identical to that of PR1.

PR3 – Undetectability of AT^{car} operation. Access token generation, update, or revocation is performed using the same steps and the same type of messages sent to the KSMS and PL. Hence, outsiders and system entities cannot distinguish which operation has been requested.

PR4 – Forensic evidence provision. In the case of disputes, the information related to a specific transaction (and only this information) may need to be reconstructed. This reconstruction can be done only if the KSMS servers collude and reveal their shares. In our setting, these servers have competing interests, thus they would not collude unless law authorities force them to do so. Due to the properties of threshold secret sharing, the private inputs can be reconstructed by a majority coalition. That is, if the KSMS consists of three parties, it suffices two of such parties to reconstruct the secrets (for semi-honest and malicious cases).

FR1 – Offline authentication. Note that steps 1-3 of the protocol require a network connection, but step 4, car access, is performed using close range communication and with no need of a network connection. The decryption and verification of the access token can be performed by the car offline (it has its key K^{car} and the owner's public-key Pk^{u_o} stored). Sending the confirmation signature σ^{car}_{Access} can also be done offline. □

6.6 Performance evaluation

Below we analyse the theoretical complexity and practical efficiency of SePCAR.

6.6.1 Theoretical complexity

The complexity of MPC protocols is typically measured by the number of communication rounds produced by non-linear operations, as linear operations can usually be performed without any information exchange and are virtually free of charge. In one evaluation of SePCAR, the non-linear operations performed by the KSMS servers are (i) the retrieval of the car key through multiple calls of the eqz functionality using the ID^{car} and their counterparts in $\vec{D}^{car}$ as parameters, and (ii) two evaluations of the encryption scheme E and one evaluation of the mac algorithm.

For (i) the evaluations of the eqz functionality, the multiplicative depth in $\lceil \log(|ID^{car}|) \rceil + 1$, where $|ID^{car}|$ is the number of bits in ID^{car}. Note that we can parallelize the eqz call for all $\vec{D}^{car}$ entries. Therefore, the bulk of the overhead of extracting the car key comes from implementing the equality test in logarithmic depth [192]. Besides executing the eqz tests, we also have to perform an extra communication round since we need to multiply the result of each equality test with its corresponding car key. The total number of communication rounds for (i) is thus $\lceil \log(|ID^{car}|) \rceil + 1$.

For (ii) the two evaluations of the encryption scheme E and the single evaluation of mac we use, as mentioned in Sect. 6.3, CTR mode of AES and CBC-MAC with AES, respectively. Note that in a single AES evaluation the number of non-linear operations equals the number of S-Boxes evaluated in these functions, but many can be parallelized. Denote by ν the number of communication rounds needed to encrypt a single 128-bit block using AES. The two evaluations of CTR mode can be performed in parallel, and cost $2 \cdot \nu$ rounds. The evaluation of CBC-MAC is inherently sequential and costs $\left\lceil \frac{|M^B|}{128} \right\rceil \cdot \nu$ communication rounds. The total number of communication rounds can thus be expressed as:

$$\left(\lceil \log(|ID^{car}|) \rceil + 1 \right) + 2 \cdot \nu + \left\lceil \frac{|M^B|}{128} \right\rceil \cdot \nu \ . \tag{6.3}$$

6.6.2 Efficiency

Our protocol is agnostic towards the underlying multiparty protocol. In our experiments we have incorporated the 3-party semi-honest protocol by Araki et al. [16], given its relative efficiency of AES calls compared to alternatives such [75, 178]. The upshot of our experiments is that SePCAR needs only 1.55 seconds for a car access provision.

We elaborate on our simulation below, following the steps of Sect. 6.4. An allocation of the time on the different steps is provided in Table 6.2 where time is averaged over 1000 runs.

Step 1. Recall that step 1 handles the preparation and sharing of the booking details and generation of keys. For enc we use RSA defined in the PKCS #1 v2.0 specifications [177] with 2048-bit keys ($\approx$ 2 ms). For sign we use RSA with SHA-2 with a 512-bit output ($\approx$ 50 ms). The prf is implemented using AES in CTR mode ($\approx$ 2 ms). For all these functions we use OpenSSL [233]. The share function is implemented by the sharing primitive introduced by Araki et al. [16].

Step 2. In this step, the KSMS servers retrieve the car key and perform the corresponding encryption and other subroutines linked to generating the MAC. We consider the following message configuration size: $\mathsf{hash}(Cert^{u_c})$ of 512 bits string, ID^{car} of 32 bits string, L^{car} of 64 bits string, CD^{u_c} of 96 bits string, AC^{u_c} of 8 bits string, ID^{B} of 32 bits string and σ^{u_o} of 512 bits string. The booking details M^B are of size 768 bits (including padding) and the final access token AT^{u_c} is of size 1408 bits (including padding). For the dec function we use RSA with 2048-bit keys ($\approx$ 2 ms). The symmetric encryption E is implemented in CTR mode and the mac in CBC mode. As mentioned before, the functions E, mac, and eqz use the primitives proposed by Araki et al. [16], and we use the multiparty AES method of Damgård and Keller [73]. Using this method, a single S-Box evaluation takes 5 communication rounds. A single evaluation of AES consists of 20 sequential evaluations of an S-Box, where we included the key expansion and took into account that parallelizable S-Boxes do not add up to the number of communication rounds, hence encryption requires $\nu = 100$ communication rounds. From (6.3) we obtain that in our simulation the total number of communication rounds is

$$\Big(5+1\Big)+2\cdot 100+6\cdot 100=806\ .$$

Key expansion for different keys needs to be performed only once, and for multiple evaluations of SePCAR for the same car the round complexity reduces.

Step 3. In this step the consumer retrieves, reconstructs, and verifies the assigned access token. The PL is implemented using SQLite. The implementation of open() again follows the primitive of Araki et al. [16], and mac is implemented using AES in CBC mode ($\approx$ 13 ms).

Table 6.2: Performance of SePCAR.

Phase	Description	Time (in sec)
Step 1	Sharing the booking details and keys	0.220 ± 0.027
Step 2	Extracting car key and making access token	1.274 ± 0.032
Step 3	Verifying the access token	0.055 (+1 Tor [307])
Total		1.551 ± 0.043 (+1 Tor)

Step 4. The final step consists of a challenge-response protocol between u_c and the car, but it does not directly affect the performance of SePCAR hence we omit it from our implementation.

Environment Settings. We implemented our simulation for SePCAR in C programming language and evaluated it using a machine equipped with an Intel *i*7, 2.6 GHz CPU and 8 GB of RAM.[2] The communication within the KSMS was simulated using socket calls and latency parameters. We used the setting from Araki et al. [16] to simulate the LAN latency ($\approx$ 0.13 ms) and from Ramamurthy et al. [253] for WiFi ($\approx$ 0.50 ms). We did not assume any specific network configuration for our experimentation.

[2]The implementation can be obtained from `https://bitbucket.org/Siemen11/sepcar`.

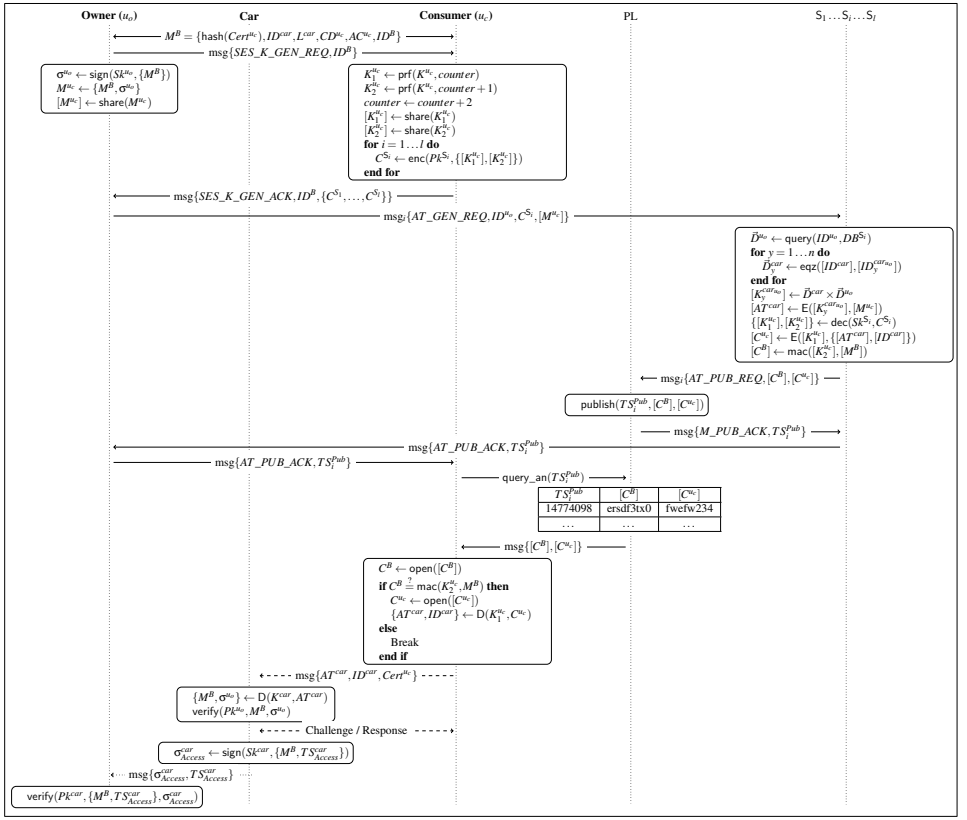


Figure 6.8: SePCAR complete representation.

6.7 Chapter summary

In this chapter, we presented a protocol for car access provision namely SePCAR. Driven by the security and privacy requirements identified in the previous chapter (Chapter 5), SePCAR provides a generation, update, revocation, and distribution mechanisms for access tokens to shared cars and also procedures to solve disputes and to deal with law enforcement requests, for instance in the case of car incidents. We proved the security and privacy properties and measured the efficiently of SePCAR for a car access provision.

Chapter 7

Conclusions and future work

Our personal information is being bought and sold without our knowledge and consent.

BRUCE SCHNEIER, *Cryptographer, computer security professional and privacy specialist*

We conclude this dissertation by summarising our main research results and contributions. Moreover, we discuss open problems and point out some future research directions.

7.1 Conclusions

The focus of this thesis was to analyse and design privacy-enhancing solutions for information sharing systems. More in particular:

- Chapter 4: We comprehensively analysed the interdependent privacy issue of Facebook third-party applications (apps) and third-party application providers (app providers), quantified the collateral information collection, and proposed solutions as countermeasures.
- Chapter 5: We methodologically analysed the security and privacy threats of physical-keyless car sharing systems, and we elicit the requirements as countermeasures.

- Chapter 6: We designed, developed and evaluated a protocol namely Secure and Privacy-enhancing protocol for Car Access Provision (SePCAR) for secure and privacy-enhancing car access provision.

Collateral information collection of third-party apps on Facebook. We conclude that collateral information collection is a privacy issue for Facebook users. Our main findings were the following. The vast majority of our participants are very concerned about the collateral information collection and would like proper notification and control mechanisms regarding collateral information collection, for themselves and also their friends. We identified that the likelihood of a Facebook user to be affected by the collateral information collection is mainly related to the popularity of an app (e.g., 80% probability for TripAdvisor). We quantified the significance of collateral information collection and identified that almost half of the 207 apps that enable the collateral information collection, collect the photos, the location and the work history of a user from their friends. We identified that the lack of transparency and consent is the main issue considering the General Data Protection Regulation (GDPR) and that poses a risk to the privacy of users. Finally, we proposed solutions to collateral information collection in the direction of enhancing transparency and increasing control with a particular focus on a *privacy dashboard extension*. Moreover, we discussed alternative solutions focusing on notification and access control mechanisms, cryptographic countermeasures and application auditing and in particular on Data Protection Impact Assessment (DPIA) driven by the GDPR. To the best of our knowledge, our work is the first to report the potential user *profiling* threat that could be posed by app providers.

A security and privacy analysis for keyless car sharing systems. We presented a novel keyless car sharing system that allows users to share their cars with others more conveniently. First, we devised a high-level model of our system and specify the functional requirements and interactions among system entities. Based on this model and taking the STRIDE and LINDDUN frameworks as a reference, we performed a comprehensive threat analysis. Finally, to mitigate the identified threats, we specified a set of security and privacy requirements for such systems. In a nutshell, this work can be used as a guide (i) to design secure and privacy-preserving protocols that support other physical asset sharing systems designs such as accommodation rental (e.g., Airbnb), and (ii) to assess the security and privacy risks imposed on users by such systems.

SePCAR: A secure and privacy-enhancing protocol for car access provision. We designed, developed and evaluated a protocol namely SePCAR for a car access provision. SePCAR is proven to be secure and privacy-enhancing, efficiently performing in ≈ 1.55 seconds for a car access provision. The security level of SePCAR relies on its underlying cryptographic primitives. The implementation was performed

in C and evaluated using an Intel *i*7, 2.6 GHz CPU and 8 GB of RAM. We presented a formal analysis of the security and privacy requirements of our protocol, and we designed a prototype as a proof-of-concept. SePCAR provides a complementary solution to physical keys, aiming for those that hold mobile devices and want a dynamic and efficient way to access a car.

7.2 Future work

There are several directions for future research that we can distil from our results, that worth further investigation and effort.

Interdependent privacy. We identified four directions for potential future work on collateral information collection. First, conducting a fully rounded quantitative survey would pay immediate dividends by (1) extending the sample size and demographic coverage compared to our questionnaire, (2) collecting answers from online social network sites other than Facebook and (3) quantifying the perceived sensitivity of users concerning various profile attributes.

Second, we are particularly interested in using DPIA as a countermeasure to collateral information collection and other social and mobile apps related to privacy threats, especially by augmenting the current DPIA state-of-the-art with a quantitative privacy risk assessment. It is likely that app providers (in particular those with multiple apps and/or many users) are obliged to perform a DPIA, both for individual apps and for their complete app suite, to demonstrate compliance to the GDPR.

Third, the *data fusion* aspect of app providers offering multiple apps may have another dimension: when app providers offer apps on multiple platforms, such as Facebook, Android or Google Drive. We would like to investigate, whether such cross-platform data fusion (and thus, profiling) is happening and/or at least feasible. For example, imagine if social profile items from Facebook could be combined with the precise location of users from a mobile app. Furthermore, Facebook Login, a widely used Single Sign-On (SSO) mechanism, may transfer the collateral information collection issue to other websites and services.

Fourth and last, an interesting but unexplored aspect of collateral information collection is whether such data gathering could harm the privacy of users, or can be beneficial to users under some special circumstances.

Design and develop secure and privacy-preserving physical asset sharing systems. We will design and implement fully-fledged secure and privacy-preserving

solutions for physical asset sharing systems. We are aiming at extending our work [285] and design a system that handles other operations such as bookings and payments offering the same security and privacy guarantees to [285]. The goal will be to identify all the non-functional requirements and design solutions that satisfy them. Physical asset sharing systems are many fold: in fact, 17 sectors have been identified from transport through accommodation rental to equipment. In our design solutions, we plan to combine various advanced technologies such as distributed ledgers, smart contracts, Secure Multiparty Computation (MPC) and zero-knowledge proofs to offer technical guarantees for satisfying the non-functional requirements. We will also pay attention to provide accountability, conditional privacy and forensic evidence provision for such systems. Moreover, being compliant with the GDPR and consider other non-functional requirements such as risk assessment and efficiency in the system design and development phase is also within the objectives that we are aiming to investigate.

Appendix A

Appendix

A.1 Questionnaire: Whether users are concerned about the collateral information collection

Facebook offers entertainment applications such as Crime scene, Candy Crash saga, and Angry birds. When you install an app from the Facebook app store, the app may collect your information on Facebook. For instance, Candy Crash saga collects your name, profile picture, country and email address. Other apps may collect other types of information.

When your friends install apps on Facebook, these apps not only collect their information but may also collect your information. The types of information that a friend's Facebook app can collect about you are listed in Fig. A.1.

Example 1. If your friends install Travelling apps, they may collect your current location to notify your friends if you are close by.

Example 2. If your friends install a Dating app (finding potential dating matches for you) the app may collect your birthday, gender, and sexual preferences to find out whether you are attracted by the same physical preferences (e.g., short, tall, brunette, blond) as your friend.

By default, Facebook allows apps that your friends install to collect information about you. Note that apps that your friends install collect your information without notifying

Type of Information	Description (of Informations from your Facebook profile)
Bio	Details you write in the "ABOUT ME" section.
Birthday	Date of birth you have added.
Family and relationships	Relationship status and family members you have added.
Interested in	Gender of interest you have added.
Religious and political views	Religious and political views you have added.
My website	Personal website (link) you have added.
If I'm online	Indicator of your online presence on Facebook.
My status updates	Status updates on your Facebook timeline excluding links, videos or photos.
My photos	Photos you have uploaded or have been tagged in.
My videos	Videos you have uploaded or have been tagged in.
My links	Links you have added.
My notes	Notes you have added.
Home Town	Home town you have added.
Current location	Current city you have added.
Education and work	Workplaces, professional skills and university studies you have added.
Activities, interests, things I like	List of activities in your profile, the pages you have liked and the particular interests those pages represent.
My app activity	App activities that are published in your timeline.

Figure A.1: The types of information that a Facebook app can collect

you in advance or asking for your approval. However, Facebook does have app settings to manually restrict the collection of your information.

Bellow you will find a questionnaire regarding user preferences about Facebook apps. Your answers will only be used for scientific purposes. No personal information will be used and responses will be aggregated if published.

On a scale from 1 to 5 where 1 means 'not concerned at all' and 5 means 'extremely concerned', how concerned are you about the following: (Not concerned at all, Slightly concerned, Moderately concerned, Very concerned, Extremely concerned).

- The default privacy settings on Facebook allow my friend's apps to collect my information.
- Facebook does not notify me in advance of the possibility that one of my friend's apps is going to collect information about me.
- Facebook does not notify me in advance of the possibility that one of my apps is going to collect information about my friends.
- Facebook does not ask for my approval in advance of the possibility that one of my friend's apps is going to collect information about me.

On a scale from 1 to 5 where 1 means 'not concerned at all' and 5 means 'extremely concerned', how concerned are you about the following pieces of your information: (Not concerned at all, Slightly concerned, Moderately concerned, Very concerned, Extremely concerned).

- Bio
- Birthday
- Family and relationships
- Interested in Religious and political views
- If I'm online
- My status updates
- My photos
- My videos
- My links
- My notes
- Home Town
- Current location
- Education and work
- Activities, interests, things I like
- My app activity

Would you like to be notified in the following cases? (Never, Depending on the type of information that is collected, Always, I don't know)

- When one of my friend's apps is going to collect pieces of my information.
- When one of my apps is going to collect pieces of my friend's information.

Can you please elaborate on your answer to the previous question? (open answer).

Which actions would you take if you are notified that one of your friend's apps is going to collect your information? (Check all that apply)

- I would take no action.
- I would restrict access to my personal information for apps that collect my information.
- I would restrict access to my information for this friend.
- I would request my friend to remove the App.
- I would un-friend my friend.
- I don't know.
- Other:

Which actions would you take if you are notified that one of your apps is going to collect your friends' information? (Check all that apply)

- I would take no action.
- I would ask my friend for her/his approval.
- I would restrict access to my friends' personal information for this App.
- I would remove this app from my list.
- I don't know.
- Other:

For how long have you had a Facebook account? (Approximately)

How many Facebook friends do you have? (Approximately)

Have you ever installed an app on Facebook?

- Yes
- No

If yes, how many Facebook apps have you installed in the past six months?

- 1 – 2
- 3 – 4
- 5 – 6
- 7+

What kind of apps do you often use on Facebook? (For instance games, lifestyle, navigation, weather, etc)

Have you ever changed the apps privacy settings?

- Yes
- No

If yes, do you remember which permission you changed? (Check all that apply)

- I restricted who can see my profile information.
- I restricted who can see me in searches.
- I restricted who can collect my information through my friends' apps.
- Other:

In which country were you born?

In which year were you born?

What is your gender?

- Male
- Female
- Prefer not to answer

What is the highest level of education you have completed?

- No degree or up to high school
- Bachelor's degree or equivalent
- Master's degree and above
- Other:

Do you have an IT background? (Check all that apply)

- Online courses or seminars
- Higher education
- Personal interest
- Other:

Remarks and Questions.

A.2 Extended Security and Privacy Analysis

We prove that SePCAR satisfies the security and privacy requirements of Section 6.2, provided that its underlying cryptographic primitives are sufficiently secure. In Section A.2.1 we describe the security models of the cryptographic primitives. Then, the formal reasoning is given in Section A.2.2.

A.2.1 Cryptographic Primitives

The security definitions for signature schemes and MAC functions are inspired by Goldwasser et al. [142], for pseudorandom functions by Goldreich et al. [140], for public key encryption by Bellare et al. [28], and for symmetric key encryption by Bellare et al. [29].

We will, in fact, need security of the cryptographic primitives in the *multi-key* setting, as these are evaluated under different keys. For example, sign is used by all owners u_o, each with a different key; enc is used for different keys, each for a different party in the KSMS, and E and mac are used for independent keys for every fresh evaluation of the protocol. We refer to Bellare et al. [28] for a discussion on generalising semantic security of public key encryption to multi-key security; the adaptation straightforwardly generalises to the other security models.

In below definitions, for a function f, we define by $\text{Func}(f)$ the set of all functions with the exact same interface as f_K. We denote a random drawing by $\xleftarrow{\$}$.

Definition. *Let $\mu \geq 1$. Consider a signature scheme* $\text{sign} = (\mathsf{keygen}, \mathsf{sign}, \mathsf{verify})$. *For any adversary $\mathcal{A}$, we define its advantage in breaking the μ-multikey existential unforgeability as*

$$\text{Adv}^{\mu\text{-euf}}_{\text{sign}}(\mathcal{A}) = \Pr\left((Pk^1, Sk^1), \ldots, (Pk^\mu, Sk^\mu) \xleftarrow{\$} \mathsf{keygen} \ : \ \mathcal{A}^{\mathsf{sign}(Sk^i,\cdot)}(Pk^i) \textit{ forges}\right),$$

where "forges" means that $\mathcal{A}$ outputs a tuple (i, M, σ) such that $\mathsf{verify}(Pk^i, M, \sigma) = 1$ *and M has never been queried to the i-th signing oracle. We define by* $\text{Adv}^{\mu\text{-euf}}_{\text{sign}}(q,t)$ *the supremum over all adversaries making at most q queries and running in time at most t.*

Definition. *Let $\mu \geq 1$. Consider a pseudorandom function* $\text{prf} = (\mathsf{keygen}, \mathsf{prf})$. *For any adversary $\mathcal{A}$, we define its advantage in breaking the μ-multikey pseudorandom function security as*

$$\text{Adv}^{\mu\text{-prf}}_{\text{prf}}(\mathcal{A}) = \left|\Pr\left(K^1, \ldots, K^\mu \xleftarrow{\$} \mathsf{keygen} \ : \ \mathcal{A}^{\mathsf{prf}(K^i,\cdot)} = 1\right) - \Pr\left(\$^1, \ldots, \$^\mu \xleftarrow{\$} \text{Func}(\mathsf{prf}) \ : \ \mathcal{A}^{\$^i} = 1\right)\right|.$$

We define by $\text{Adv}^{\mu\text{-prf}}_{\text{prf}}(q,t)$ *the supremum over all adversaries making at most q queries and running in time at most t.*

Definition. *Let $\mu \geq 1$. Consider a public-key encryption scheme* $\mathsf{enc} = (\mathsf{keygen}, \mathsf{enc}, \mathsf{dec})$. *For any adversary $\mathcal{A}$, we define its advantage in breaking the μ-multikey semantic security as*

$$\mathrm{Adv}^{\mu\text{-pke}}_{\mathsf{enc}}(\mathcal{A}) = \Big|\Pr\Big((Pk^1, Sk^1), \ldots, (Pk^\mu, Sk^\mu) \xleftarrow{\$} \mathsf{keygen} \;:\; \mathcal{A}^{O_0}(Pk^i) = 1\Big) - \Pr\Big((Pk^1, Sk^1), \ldots, (Pk^\mu, Sk^\mu) \xleftarrow{\$} \mathsf{keygen} \;:\; \mathcal{A}^{O_1}(Pk^i) = 1\Big)\Big|,$$

where O_b for $b \in \{0,1\}$ gets as input a tuple (i, m_0, m_1) with $i \in \{1, \ldots, \mu\}$ and $|m_0| = |m_1|$ and outputs $\mathsf{enc}_{Pk^i}(m_b)$. We define by $\mathrm{Adv}^{\mu\text{-pke}}_{\mathsf{enc}}(t)$ the supremum over all adversaries running in time at most t.

Definition. *Let $\mu \geq 1$. Consider a symmetric-key encryption scheme* $\mathrm{E} = (\mathsf{keygen}, \mathsf{E}, \mathsf{D})$. *For any adversary $\mathcal{A}$, we define its advantage in breaking the μ-multikey chosen-plaintext security as*

$$\mathrm{Adv}^{\mu\text{-ske}}_{\mathrm{E}}(\mathcal{A}) = \Big|\Pr\Big(K^1, \ldots, K^\mu \xleftarrow{\$} \mathsf{keygen} \;:\; \mathcal{A}^{\mathsf{E}(K^i, \cdot)} = 1\Big) - \Pr\Big(\$^1, \ldots, \$^\mu \xleftarrow{\$} \mathrm{Func}(\mathsf{E}) \;:\; \mathcal{A}^{\$^i} = 1\Big)\Big|.$$

We define by $\mathrm{Adv}^{\mu\text{-ske}}_{\mathrm{E}}(q,t)$ the supremum over all adversaries making at most q queries and running in time at most t.

Definition. *Let $\mu \geq 1$. Consider a MAC function* $\mathsf{mac} = (\mathsf{keygen}, \mathsf{mac})$. *For any adversary $\mathcal{A}$, we define its advantage in breaking the μ-multikey existential unforgeability as*

$$\mathrm{Adv}^{\mu\text{-mac}}_{\mathsf{mac}}(\mathcal{A}) = \Pr\Big(K^1, \ldots, K^\mu \xleftarrow{\$} \mathsf{keygen} \;:\; \mathcal{A}^{\mathsf{mac}(K^i, \cdot)} \textit{ forges}\Big),$$

where "forges" means that $\mathcal{A}$ outputs a tuple (i, M, σ) such that $\mathsf{mac}(K^i, M) = \sigma$ and M has never been queried to the i-th MAC function. We define by $\mathrm{Adv}^{\mu\text{-mac}}_{\mathsf{mac}}(q,t)$ the supremum over all adversaries making at most q queries and running in time at most t.

Finally, we consider the hash function hash to be collision-resistant. We denote the supremal probability of any adversary in finding a collision for hash in t time by $\mathrm{Adv}^{\mathrm{col}}_{\mathsf{hash}}(t)$. The definition is, acknowledgeably, debatable: for any hash function there exists an adversary that can output a collision in constant time (namely one that has a collision hardwired in its code). We ignore this technicality for simplicity and refer to [258, 280, 257] for further discussion.

A.2.2 Analysis

We prove that SePCAR satisfies the security and privacy requirements of Section 6.2 provided that its underlying cryptographic primitives are sufficiently secure.

Theorem 2. *Suppose that communication takes place over private channels, the MPC is statistically secure,* hash *is a random oracle, and*

$$\mathrm{Adv}_{\mathrm{sign}}^{\mu_o+\mu_{\mathrm{car}}\text{-euf}}(2q,t)+\mathrm{Adv}_{\mathrm{prf}}^{\mu_c\text{-prf}}(2q,t)+\mathrm{Adv}_{\mathrm{enc}}^{l\text{-pke}}(t)+$$
$$\mathrm{Adv}_{\mathrm{E}}^{q+\mu_{\mathrm{car}}\text{-ske}}(2q,t)+\mathrm{Adv}_{\mathrm{mac}}^{q\text{-mac}}(q,t)+\mathrm{Adv}_{\mathsf{hash}}^{\mathrm{col}}(t)\ll 1\,,$$

where μ_o denotes the maximum number of u_os, μ_c the maximum number of u_cs, μ_{car} the maximum number of cars, l the number of servers in the KSMS*, q the total times the protocol gets evaluated, and t the maximum time of any adversary.*

Then, SePCAR fulfills the security and privacy requirements of Section 6.2.

Proof. Recall from Section 6.2 that u_os and CM are honest-but-curious whereas u_cs and outsiders may be malicious and actively deviate from the protocol. Cars are trusted.

Via a hybrid argument, we replace the pseudorandom functions $\mathsf{prf}(K^{u_c},\cdot)$ by independent random functions $\$^{u_c}$. This step is performed at the cost of

$$\mathrm{Adv}_{\mathrm{prf}}^{\mu_c\text{-prf}}(2q,t)\,, \tag{A.1}$$

as in every of the q evaluations of SePCAR there are two evaluations of a function prf, and there are at most μ_c instances of these functions. As we assume that the MPC is performed statistically secure, we can replace the KSMS by a single trusted authority (with l interfaces) that is trusted, perfectly evaluates the protocol, and does not reveal/leak any information. Assuming that the public-key encryption reveals nothing, which can be done at the cost of

$$\mathrm{Adv}_{\mathrm{enc}}^{l\text{-pke}}(t)\,, \tag{A.2}$$

we can for simplicity replace it with a perfectly secure public-key encryption ρ^{KSMS} to the KSMS directly (an encryption does not reveal its origin and content, and only KSMS can magically decrypt), therewith eliminating the fact that KSMS has l interfaces and has to perform multiparty computation. Now, as the pseudorandom functions are replaced by random functions, the keys to the symmetric encryption scheme E are all independently and uniformly distributed, and as the public-key encryption scheme is secure, these keys never leak. Therefore, we can replace the symmetric encryption

functionalities by perfectly random invertible functions, $\pi^{car_{u_o}}$ for the cars and unique π^{u_c}'s for every new encryption under u_c's session keys, at the cost of

$$\mathrm{Adv}_{\mathrm{E}}^{q+\mu_{\text{car-ske}}}(2q,t)\,, \tag{A.3}$$

as there are $q+\mu_{\text{car}}$ different instances involved and at most $2q$ evaluations are made in total. Note that this means that, instead of randomly drawing $K_1^{u_c} \leftarrow \$^{u_c}$, we now randomly draw $\pi^{u_c} \xleftarrow{\$} \mathrm{Func}(\mathsf{E})$.

We are left with a simplified version of SePCAR, namely one where the KSMS is replaced by a single trusted authority, the pseudorandom functions are replaced by independent random drawings (u_c uses $\$^{u_c}$ which generates fresh outputs for every call), public-key encryptions are replaced with a perfectly secure public-key encryption function ρ^{KSMS}, and symmetric-key encryptions are replaced by perfectly random invertible functions $\pi^{car_{u_o}}$ and π^{u_c}. The simplified protocol is given in Figure A.2. Here, the derivation of the car key (or, formally, the random function corresponding to the encryption) from the database is abbreviated to $\pi^{car_{u_o}} \leftarrow \mathsf{query}(ID^{u_o}, DB^{\mathsf{KSMS}})$ for conciseness.

We will now treat the security and privacy requirements, and discuss how these are achieved from the cryptographic primitives, separately. We recall that u_c and u_o have agreed upon the booking details prior to the evaluation of SePCAR, hence they know each other by design.

SR1 – Confidentiality of M^B. In one evaluation of the protocol, u_c, u_o, the *trusted* KSMS, and the shared car learn the booking details by default or design. The booking details only become public through the values C^B and C^{u_c} satisfying

$$C^B = \mathsf{mac}(K_2^{u_c}, M^B)\,, \tag{A.4}$$

$$C^{u_c} = \pi^{u_c}(\{\pi^{car_{u_o}}(\{M^B, \sigma^{u_o}\}), ID^{car}\})\,. \tag{A.5}$$

The latter value reveals nothing about M^B as π^{u_c} is randomly generated for every evaluation, whereas the former value reveals nothing about M^B as $K_2^{u_c}$ is randomly generated for every evaluation. The nested encryption $\pi^{u_c} \circ \pi^{car_{u_o}}$ does not influence the analysis due to the mutual independence of the two functions.

SR2 – Authenticity of M^B. An owner who initiates the access token generation and distribution, first signs the booking details using its private key before sending those to the KSMS in shares. Therefore, once the car receives the token and obtains the booking details, it can verify u_o's signature on the booking details. In other words,

the car can verify the source of M^B, u_o, and its integrity. Suppose, to the contrary, that a malicious consumer can get access to a car of an u_o. This particularly means that it created a tuple (M^B, σ^{u_o}) such that $\mathsf{verify}(Pk^{u_o}, M^B, \sigma^{u_o})$ holds. If σ^{u_o} is new, this means that u_c forges a signature for the secret signing key Sk^{u_o}. Denote the event that this happens by

$$\mathsf{E}_1 \ : \ \mathcal{A} \text{ forges } \mathsf{sign}(Sk^{u_o}, \cdot) \text{ for some } Sk^{u_o}\,. \tag{A.6}$$

On the other hand, if (M^B, σ^{u_o}) is old but the evaluation is fresh, this means a collision $\mathsf{hash}(Cert^{u_c}) = \mathsf{hash}(Cert^{u_c\prime})$. Denote the event that this happens by

$$\mathsf{E}_2 \ : \ \mathcal{A} \text{ finds a collision for } \mathsf{hash}\,. \tag{A.7}$$

We thus obtain that a violation of SR2 implies $\mathsf{E}_1 \vee \mathsf{E}_2$.

SR3 – Confidentiality of AT^{car}. The access token is generated by the KSMS obliviously (as it is trusted), and only revealed to the public in encrypted form, through C^{u_c} of (A.5). Due to the uniform drawing of π^{u_c} (and the security of ρ^{KSMS} used to transmit this function), only the legitimate user can decrypt and learn the access token. It shares it with the car over a secure and private channel.

SR4 – Confidentiality of K^{car}. By virtue of our hybrid argument on the use of the symmetric-key encryption scheme, $\mathsf{E}_{K^{car}}$ got replaced with $\pi^{car_{u_o}}$, which itself is a keyless random encryption scheme. As the key is now absent, it cannot leak.

SR5 – Backward and forward secrecy of AT^{car}. The access token is published on PL as C^{u_c} of (A.5), encrypted using π^{u_c}. Every honest u_c generates a uniformly randomly drawn function π^{u_c} for every new evaluation. Therefore, all encryptions C^{u_c} are independent and reveal nothing of each other. (Note that nothing can be said about access tokens for malicious users who may deviate from the protocol and reuse one-time keys.)

SR6 – Non-repudiation of origin of AT^{car}. The car, who is a trusted identity, verifies the origin through verification of the signature, $\mathsf{verify}(Pk^{u_o}, M^B, \sigma^{u_o})$. The consumer u_c verifies the origin through the verification of the MAC function, $C^B \stackrel{?}{=} \mathsf{mac}(K_2^{u_c}, M^B)$. Note that u_c does not effectively verify AT^{car}, but rather C^B. In either case, security fails only if the asymmetric signature scheme or the MAC function are forgeable. The former is already captured by event E_1 in (A.6). For the latter, denote the event that this happens by

$$\mathsf{E}_3 \ : \ \mathcal{A} \text{ forges } \mathsf{mac}(K_2^{u_c}, \cdot) \text{ for some } K_2^{u_c}\,. \tag{A.8}$$

We thus obtain that a violation of SR6 implies $\mathsf{E}_1 \vee \mathsf{E}_3$.

SR7 – Non-repudiation of delivery of AT^{car}. u_o can verify correct delivery through the verification of the message sent by the car to the him/her, $\mathsf{verify}(Pk^{car}, \{M^B, TS^{car}_{Access}\}, \sigma^{car}_{Access})$ at the end of the protocol. Security breaks only if the signature scheme is forgeable. Denote the event that this happens by

$$\mathsf{E}_4 \ : \ \mathcal{A} \text{ forges } \mathsf{sign}(Sk^{car}, \cdot) \text{ for some } Sk^{car}. \tag{A.9}$$

We thus obtain that a violation of SR7 implies E_4.

PR1 – Unlinkability of u_c and the car. The only consumer-identifiable data is in u_c's certificate included in the booking details. Note that these are agreed upon between u_c and u_o, so u_o learns the identity of u_c by default. Beyond that, u_c only communicates with the car, which is supposed to learn u_c's identity so that it can perform proper access control. u_c consults PL over an anonymous channel. The booking details are transferred to and from the KSMS, but these are encrypted and do not leak by virtue of their confidentiality (security requirement SR1).

PR2 – Anonymity of u_c and the car . The reasoning is identical to that of PR1.

PR3 – Undetectability of AT^{car} operation. Access token generation, update, or revocation is performed using the same steps and the same type of messages sent to the KSMS and PL. Hence, outsiders and system entities can not distinguish which operation has been requested.

PR4 – Forensic evidence provision. In the case of disputes, the information related to a specific transaction (and only this information) may need to be reconstructed. This reconstruction can be done only if the KSMS servers collude and reveal their shares. In our setting, these servers have competing interests, thus they would not collude unless law authorities enforce them to do so. Due to the properties of threshold secret sharing, the private inputs can be reconstructed by a majority coalition. This is, if the KSMS consists of three parties, it suffices two of such parties to reconstruct the secrets (for semi-honest and malicious cases).

FR1 – Offline authentication. Note that steps 1-3 of the protocol require a network connection, but step 4, car access, is performed using close range communication and with no need of a network connection. The decryption and verification of the access token can be performed by the car offline (it has its $\pi^{car_{u_0}}$ and u_o's public key Pk^{u_o} stored). Sending the confirmation signature σ^{car}_{Access} can also be done offline.

Conclusion. In conclusion, SePCAR operates securely as long as the costs of (A.1-A.3), together with the probability that one of the events (A.6-A.9) occurs, are sufficiently small:

$$\mathrm{Adv}_{\mathrm{prf}}^{\mu_c\text{-prf}}(2q,t) + \mathrm{Adv}_{\mathrm{enc}}^{l\text{-pke}}(t) + \mathrm{Adv}_{\mathrm{E}}^{q+\mu_{\mathrm{car}}\text{-ske}}(2q,t) + \Pr(\mathsf{E}_1 \vee \mathsf{E}_2 \vee \mathsf{E}_3 \vee \mathsf{E}4) \ll 1\,.$$

By design, the probability that event $\mathsf{E}_1 \vee \mathsf{E}_4$ occurs is upper bounded by $\mathrm{Adv}_{\mathrm{sign}}^{\mu_o+\mu_{\mathrm{car}}\text{-euf}}(2q,t)$, the probability that event E_3 occurs is upper bounded by $\mathrm{Adv}_{\mathrm{mac}}^{q\text{-mac}}(q,t)$, and the probability that E_2 occurs is upper bounded by $\mathrm{Adv}_{\mathsf{hash}}^{\mathrm{col}}(t)$. We thus obtain

$$\Pr(\mathsf{E}_1 \vee \mathsf{E}_2 \vee \mathsf{E}_3 \vee \mathsf{E}4) \leq \mathrm{Adv}_{\mathrm{sign}}^{\mu_o+\mu_{\mathrm{car}}\text{-euf}}(2q,t) + \mathrm{Adv}_{\mathrm{mac}}^{q\text{-mac}}(q,t) + \mathrm{Adv}_{\mathsf{hash}}^{\mathrm{col}}(t)\,,$$

which completes the proof. □

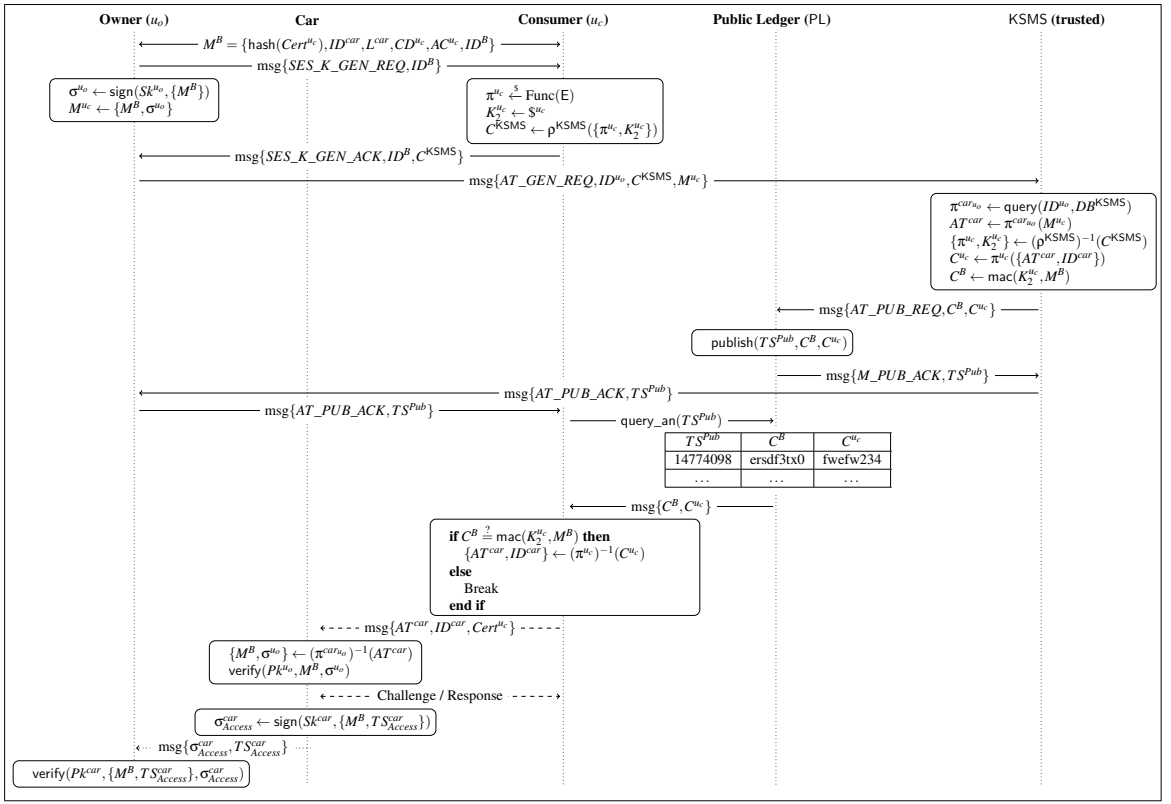


Figure A.2: Simplified representation of SePCAR for the proof of Theorem 2.

Bibliography

[1] Google plus. https://plus.google.com. Accessed Feb, 2018.

[2] Linkedin. https://www.linkedin.com/. Accessed Feb, 2018.

[3] Twitter. https://twitter.com/. Accessed Feb, 2018.

[4] ABDELBERI, C., DING, Y., DEY, R., KÂAFAR, M. A., AND ROSS, K. W. A closer look at third-party OSN applications: Are they leaking your personal information? In *Passive and Active Measurement - 15th International Conference, PAM 2014, Los Angeles, CA, USA, March 10-11, 2014, Proceedings* (2014), pp. 235–246.

[5] ABDELBERI, C., KÂAFAR, M. A., AND BORELI, R. Big friend is watching you: analyzing online social networks tracking capabilities. In *Proceedings of the 2012 ACM workshop on Workshop on Online Social Networks, WOSN 2012, Helsinki, Finland, August 17, 2012* (2012), pp. 7–12.

[6] ACEA. Carsharing: Evolution, challenges and opportunities. https://goo.gl/NTec4l. Accessed April, 2017.

[7] ACQUISTI, A., AND GROSS, R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy Enhancing Technologies, 6th International Workshop, PET 2006, Cambridge, UK, June 28-30, 2006, Revised Selected Papers* (2006), pp. 36–58.

[8] AGARWAL, A. Meaningful threat modeling for CISOs. https://www.peerlyst.com/posts/meaningful-threat-modeling-for-cisos-anurag-agarwal. Accessed Oct, 2017.

[9] AIRBNB. https://www.airbnb.com. Accessed Feb, 2018.

[10] AL-AZIZY, D., MILLARD, D. E., SYMEONIDIS, I., O'HARA, K., AND SHADBOLT, N. A literature survey and classifications on data deanonymisation. In *Risks and Security of Internet and Systems - 10th International Conference,*

CRiSIS 2015, Mytilene, Lesbos Island, Greece, July 20-22, 2015, Revised Selected Papers (2015), pp. 36–51.

[11] ALBERT, R., AND BARABÁSI, A. Statistical mechanics of complex networks. *CoRR cond-mat/0106096* (2001).

[12] ALBRECHT, M. R., RECHBERGER, C., SCHNEIDER, T., TIESSEN, T., AND ZOHNER, M. Ciphers for MPC and FHE. In *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I* (2015), pp. 430–454.

[13] ALSENOY, B. V. *Regulating data protection: the allocation of responsibility and risk among actors involved in personal data processing*. PhD thesis, Research Unit KU Leuven Centre for IT IP Law (CiTiP), 2016.

[14] ANDERSON, J., BERNOFF, J., REITSMA, R., AND SORENSEN, E. A global update of social technographics. *A Social Computing* (2010).

[15] APPINSPECT. Appinspect: A framework for automated security and privacy analysis of OSN application ecosystems. http://ai.sba-research.org/. Accessed Aug, 2017.

[16] ARAKI, T., FURUKAWA, J., LINDELL, Y., NOF, A., AND OHARA, K. High-throughput semi-honest secure three-party computation with an honest majority. In *Proceedings of the 2016 ACM SIGSAC CCS* (2016), pp. 805–817.

[17] ARNOSTI, C., GRUNTZ, D., AND HAURI, M. Secure physical access with NFC-enabled smartphones. In *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia, MoMM 2015, Brussels, Belgium, December 11-13, 2015* (2015), pp. 140–148.

[18] ARORA, S., AND BARAK, B. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.

[19] ATKINSON, R. J., KALISKI, B., BRAINARD, J. G., AND TURNER, S. Use of the RSA-KEM key transport algorithm in the cryptographic message syntax (CMS). *Internet Engineering Task Force (IETF), Request for Comments (RFC) 5990* (2010), 1–27.

[20] BALAJI, S., AND MURUGAIYAN, M. S. Waterfall vs. V-model vs. Agile: A comparative study on SDLC. *International Journal of Information Technology and Business Management 2*, 1 (2012), 26–30.

[21] BALASCH, J., RIAL, A., TRONCOSO, C., PRENEEL, B., VERBAUWHEDE, I., AND GEUENS, C. Pretp: Privacy-preserving electronic toll pricing. In

19th USENIX Security Symposium, Washington, DC, USA, August 11-13, 2010, Proceedings (2010), pp. 63–78.

[22] BANAWAN, K. A., AND ULUKUS, S. The capacity of private information retrieval from coded databases. *IEEE Trans. Information Theory 64*, 3 (2018), 1945–1956.

[23] BARDHI, F., AND ECKHARDT, G. M. Access-based consumption: The case of car sharing. *Journal of Consumer Research 39*, 4 (2012), 881–898.

[24] BBC. Italian police break up Fiat 500 car-share theft gang. http://www.bbc.com/news/world-europe-38314493. Accessed Jan, 2018.

[25] BEATO, F., ION, I., CAPKUN, S., PRENEEL, B., AND LANGHEINRICH, M. For some eyes only: protecting online information sharing. In *Third ACM Conference on Data and Application Security and Privacy, CODASPY'13, San Antonio, TX, USA, February 18-20, 2013* (2013), pp. 1–12.

[26] BEATO, F., KOHLWEISS, M., AND WOUTERS, K. Scramble! your social network data. In *Privacy Enhancing Technologies - 11th International Symposium, PETS 2011, Waterloo, ON, Canada, July 27-29, 2011. Proceedings* (2011), pp. 211–225.

[27] BEAVER, D., MICALI, S., AND ROGAWAY, P. The round complexity of secure protocols (extended abstract). In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA* (1990), pp. 503–513.

[28] BELLARE, M., BOLDYREVA, A., AND MICALI, S. Public-key encryption in a multi-user setting: Security proofs and improvements. In *Advances in Cryptology - EUROCRYPT* (2000), pp. 259–274.

[29] BELLARE, M., DESAI, A., JOKIPII, E., AND ROGAWAY, P. A concrete security treatment of symmetric encryption. In *IEEE FOCS* (1997), pp. 394–403.

[30] BELLARE, M., KILIAN, J., AND ROGAWAY, P. The security of the cipher block chaining message authentication code. *J. Comput. Syst. Sci. 61*, 3 (2000), 362–399.

[31] BELLARE, M., AND NAMPREMPRE, C. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. *J. Cryptology 21*, 4 (2008), 469–491.

[32] BELLARE, M., AND ROGAWAY, P. Random oracles are practical: A paradigm for designing efficient protocols. In *CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993.* (1993), pp. 62–73.

[33] BEN-OR, M., GOLDWASSER, S., AND WIGDERSON, A. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA* (1988), pp. 1–10.

[34] BERT, J., COLLIE, B., GERRITS, M., AND XU, G. What's ahead for car sharing?: The new mobility and its impact on vehicle sales. https://goo.gl/ZmPZ5t. Accessed June, 2017.

[35] BICZÓK, G., AND CHIA, P. H. Interdependent privacy: Let me share your data. In *Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers* (2013), pp. 338–353.

[36] BIER, C., KÜHNE, K., AND BEYERER, J. Privacyinsight: The next generation privacy dashboard. In *Privacy Technologies and Policy - 4th Annual Privacy Forum, APF 2016, Frankfurt/Main, Germany, September 7-8, 2016, Proceedings* (2016), pp. 135–152.

[37] BLANCHARD, B. S., FABRYCKY, W. J., AND FABRYCKY, W. J. *Systems engineering and analysis*, vol. 4. Prentice Hall Englewood Cliffs, NJ, 1990.

[38] BLOEMENDAALAND, E., AND ØVERÅSEN, I. M. H. Interdependent privacy on Facebook. https://goo.gl/6CNqsf. Accessed Feb, 2018.

[39] BMW. DriveNow car sharing. https://drive-now.com/. Accessed Nov, 2016.

[40] BOGETOFT, P., CHRISTENSEN, D. L., DAMGÅRD, I., GEISLER, M., JAKOBSEN, T. P., KRØIGAARD, M., NIELSEN, J. D., NIELSEN, J. B., NIELSEN, K., PAGTER, J., SCHWARTZBACH, M. I., AND TOFT, T. Secure multiparty computation goes live. In *Financial Cryptography and Data Security, 13th International Conference, FC 2009, Accra Beach, Barbados, February 23-26, 2009. Revised Selected Papers* (2009), pp. 325–343.

[41] BORISOV, N., GOLDBERG, I., AND BREWER, E. A. Off-the-record communication, or, why not to use PGP. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society, WPES 2004, Washington, DC, USA, October 28, 2004* (2004), pp. 77–84.

[42] BOYD, D., AND ELLISON, N. B. Social network sites: Definition, history, and scholarship. *J. Computer-Mediated Communication 13*, 1 (2007), 210–230.

[43] BOYD, D., AND HARGITTAI, E. Facebook privacy settings: Who cares? *First Monday 15*, 8 (2010).

[44] BRANDIMARTE, L., ACQUISTI, A., AND LOEWENSTEIN, G. Misplaced confidences: Privacy and the control paradox. In *9th Annual Workshop on the Economics of Information Security, WEIS 2010, Harvard University, Cambridge, MA, USA, June 7-8, 2010* (2010).

[45] BUCHMANN, J., CAPURRO, R., LÖW, M., MÜLLER, G., PRETSCHNER, A., ROSSNAGEL, A., WAIDNER, M., EIERMANN, K.-I., ELDRED, M., KELBERT, F., ET AL. *Internet Privacy: Options for adequate realisation*. Springer Science & Business Media, 2014.

[46] BUCHMANN, J., NEBEL, M., ROSSNAGEL, A., SHIRAZI, F., SIMO, H., AND WAIDNER, M. Personal information dashboard: Putting the individual back in control. In *Digital Enlightenment Yearbook 2013: The Value of Personal Data*, no. http://ebooks.iospress.nl/publication/35. IOS Press, 2013, pp. 139–164.

[47] CAMBRIDGE ANALYTICA. CA responds to announcement that GSR dataset potentially contained 87 million records. https://ca-commercial.com/news/ca-responds-announcement-gsr-dataset-potentially-contained-87-million-records. Accessed April, 2018.

[48] CAMBRIDGE ANALYTICA. Cambridge Analytica responds to false allegations in the media. https://ca-commercial.com/news/cambridge-analytica-responds-false-allegations-media. Accessed April, 2018.

[49] CAMBRIDGE ANALYTICA. Data drives all we do. https://cambridgeanalytica.org/. Accessed Jan, 2018.

[50] CAMENISCH, J., AND LYSYANSKAYA, A. Signature schemes and anonymous credentials from bilinear maps. In *Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings* (2004), pp. 56–72.

[51] CAMERON, K. FIPS PUB 186-4: Digital Signature Standard (DSS). *National Institute of Standards and Technology (NIST)* (2013), 1–130.

[52] CAR CONNECTIVITY CONSORTIUM® (CCC). Building Digital Key Solution for Automotive. http://carconnectivity.org/. Accessed Sept, 2017.

[53] CAR2GO. DriveNow Car Sharing. https://www.car2go.com. Accessed Oct, 2017.

[54] CATRINA, O., AND DE HOOGH, S. Improved primitives for secure multiparty integer computation. In *Security and Cryptography for Networks, 7th International Conference, SCN 2010, Amalfi, Italy, September 13-15, 2010. Proceedings* (2010), pp. 182–199.

[55] CHAI, Q., AND GONG, G. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers. In *IEEE ICC, Ottawa, ON, Canada, June 10-15, 2012* (2012), pp. 917–922.

[56] CHAUM, D. The dining cryptographers problem: Unconditional sender and recipient untraceability. *J. Cryptology 1*, 1 (1988), 65–75.

[57] CHAUM, D., CRÉPEAU, C., AND DAMGÅRD, I. Multiparty unconditionally secure protocols (abstract). In *Advances in Cryptology - CRYPTO '87, A Conference on the Theory and Applications of Cryptographic Techniques, Santa Barbara, California, USA, August 16-20, 1987, Proceedings* (1987), p. 462.

[58] CHAUM, D. L. Untraceable electronic mail, return addresses, and digital pseudonyms. *Com. of the ACM 24*, 2 (1981), 84–90.

[59] CHIA, P. H., YAMAMOTO, Y., AND ASOKAN, N. Is this app safe?: a large scale study on application permissions and risk signals. In *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012* (2012), pp. 311–320.

[60] CHOR, B., KUSHILEVITZ, E., GOLDREICH, O., AND SUDAN, M. Private information retrieval. *J. ACM 45*, 6 (1998), 965–981.

[61] CHOURABI, H., NAM, T., WALKER, S., GIL-GARCÍA, J. R., MELLOULI, S., NAHON, K., PARDO, T. A., AND SCHOLL, H. J. Understanding smart cities: An integrative framework. In *45th Hawaii International International Conference on Systems Science (HICSS-45 2012), Proceedings, 4-7 January 2012, Grand Wailea, Maui, HI, USA* (2012), pp. 2289–2297.

[62] CODAGNONE, C., AND MARTENS, B. Scoping the sharing economy: Origins, definitions, impact and regulatory issues.

[63] CONTI, M., HASANI, A., AND CRISPO, B. Virtual private social networks. In *First ACM Conference on Data and Application Security and Privacy, CODASPY 2011, San Antonio, TX, USA, February 21-23, 2011, Proceedings* (2011), pp. 39–50.

[64] CONTINENTAL. Digital car key: vehicle access with cellular phone and bluetooth. https://www.youtube.com/watch?v=vdnrr5i4naE. Accessed Oct, 2017.

[65] COOPER, D., AND KAGEL, J. H. Other regarding preferences: a selective survey of experimental results. *Handbook of Experimental Economics 2* (2009).

[66] COOPER, D., SANTESSON, S., FARRELL, S., BOEYEN, S., HOUSLEY, R., AND POLK, W. T. Internet X.509 public key infrastructure certificate and

certificate revocation list (CRL) profile. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 5280* (2008), 1–151.

[67] COURT OF JUSTICE OF THE EUROPEAN UNION. Case C-101/01, Bodil Lindqvist, OJ 2004 C7/3, ECLI:EU:C:2003:596. http://curia.europa.eu/juris/liste.jsf?num=C-101/01. Accessed Aug, 2017.

[68] CRAMER, R., DAMGÅRD, I., AND NIELSEN, J. B. *Secure Multiparty Computation and Secret Sharing*. Cambridge University Press, 2015.

[69] CRISTOFARO, E. D., SORIENTE, C., TSUDIK, G., AND WILLIAMS, A. Hummingbird: Privacy at the time of Twitter. In *IEEE Symposium on Security and Privacy, SP 2012, 21-23 May 2012, San Francisco, California, USA* (2012), pp. 285–299.

[70] CUTILLO, L. A., MOLVA, R., AND ÖNEN, M. Safebook: A distributed privacy preserving Online Social Network. In *12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WOWMOM 2011, Lucca, Italy, 20-24 June, 2011* (2011), pp. 1–3.

[71] DAEMEN, J., AND RIJMEN, V. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002.

[72] DAMGÅRD, I., FITZI, M., KILTZ, E., NIELSEN, J. B., AND TOFT, T. Unconditionally secure constant-rounds multi-party computation for equality, comparison, bits and exponentiation. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings* (2006), pp. 285–304.

[73] DAMGÅRD, I., AND KELLER, M. Secure multiparty AES. In *Financial Cryptography and Data Security, 14th International Conference, FC 2010, Tenerife, Canary Islands, January 25-28, 2010, Revised Selected Papers* (2010), pp. 367–374.

[74] DAMGÅRD, I., KELLER, M., LARRAIA, E., MILES, C., AND SMART, N. P. Implementing AES via an actively/covertly secure dishonest-majority MPC protocol. In *Security and Cryptography for Networks - 8th International Conference, SCN 2012, Amalfi, Italy, September 5-7, 2012. Proceedings* (2012), pp. 241–263.

[75] DAMGÅRD, I., PASTRO, V., SMART, N. P., AND ZAKARIAS, S. Multiparty computation from somewhat homomorphic encryption. In *Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings* (2012), pp. 643–662.

[76] DANEZIS, G., DOMINGO-FERRER, J., HANSEN, M., HOEPMAN, J., MÉTAYER, D. L., TIRTEA, R., AND SCHIFFNER, S. Privacy and data protection by design - from policy to engineering. *CoRR abs/1501.03726* (2015).

[77] DANG, Q. H. FIPS PUB 180-4: Secure Hash Standard (SHS). *National Institute of Standards and Technology (NIST)* (2015), 1–36.

[78] DARING FIREBALL. Regarding Uber's new "always" location tracking. https://daringfireball.net/2016/12/uber_location_privacy. Accessed Oct, 2017.

[79] DAVIS, A. M., BERSOFF, E. H., AND COMER, E. R. A strategy for comparing alternative software development life cycle models. *IEEE Trans. Software Eng. 14*, 10 (1988), 1453–1461.

[80] DE VAUS, D. A. *Surveys in social research*, 3rd ed. ed. UCL press, London, 1991.

[81] DEBATIN, B., LOVEJOY, J. P., HORN, A., AND HUGHES, B. N. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *J. Computer-Mediated Communication 15*, 1 (2009), 83–108.

[82] DEF CON 2013. How to hack a Tesla model S. http://tinyurl.com/ovve3wq. Accessed June, 2016.

[83] DELOITTE. Smart mobility: Reducing congestion and fostering faster, greener, and cheaper transportation options. https://www2.deloitte.com/insights/us/en/industry/public-sector/smart-mobility-trends.html. Accessed May, 2018.

[84] DENG, M., WUYTS, K., SCANDARIATO, R., PRENEEL, B., AND JOOSEN, W. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering 16*, 1 (2011), 3–32.

[85] DIAZ, C., AND GÜRSES, S. Understanding the landscape of privacy technologies. *Extended abstract of invited talk in proceedings of the Information Security Summit* (2012), 58–63.

[86] DIERKS, T., AND RESCORLA, E. The transport layer security (TLS) protocol version 1.2. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 5246* (2008), 1–104.

[87] DIFFIE, W., VAN OORSCHOT, P. C., AND WIENER, M. J. Authentication and authenticated key exchanges. *Des. Codes Cryptography 2*, 2 (1992), 107–125.

[88] DMITRIENKO, A., AND PLAPPERT, C. Secure free-floating car sharing for offline cars. In *ACM CODASPY* (2017), pp. 349–360.

[89] DOBBERTIN, H., BOSSELAERS, A., AND PRENEEL, B. RIPEMD-160: A strengthened version of RIPEMD. In *Fast Software Encryption, Third International Workshop, Cambridge, UK, February 21-23, 1996, Proceedings* (1996), pp. 71–82.

[90] DOLEV, D., AND YAO, A. C. On the security of public key protocols. *IEEE Trans. Information Theory 29*, 2 (1983), 198–207.

[91] DOMIN, K., MARIN, E., AND SYMEONIDIS, I. Security analysis of the drone communication protocol: Fuzzing the mavlink protocol. *37th Symposium on Information Theory in the Benelux* (2016), 1–7.

[92] DÖRING, N. Personal home pages on the web: A review of research. *J. Computer-Mediated Communication 7*, 3 (2002), 0.

[93] DRIVENOW. Helping you to improve the quality of life in your city. https://www.slideshare.net/crowdsourcingweek/free-floating-car-sharing. Accessed Oct, 2017.

[94] DUTTON, W. H. *The Oxford Handbook of Internet Studies*. Oxford University Press, 2013.

[95] DWORKIN, M., BARKER, E., NECHVATAL, J., FOTI, J., BASSHAM, L. E., ROBACK, E., AND DRAY, J. F. FIPS PUB 197: Advanced Encryption Standard (AES). *National Institute of Standards and Technology (NIST)* (2001), 1–51.

[96] ELAINE, B., AND JOHN, K. Sp 800-90a rev 1: Recommendation for random number generation using deterministic random bit generators. *National Institute of Standards and Technology (NIST)* (2015), 1–110.

[97] ELECTRONIC FRONTIER FOUNDATION. UPDATED: Uber Should Restore User Control to Location Privacy. https://www.eff.org/deeplinks/2016/12/uber-should-restore-user-control-location-privacy. Accessed Oct, 2017.

[98] ELGAMAL, T. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. Information Theory 31*, 4 (1985), 469–472.

[99] ELLISON, N. B., AND BOYD, D. M. Sociality through social network sites. *The Oxford Handbook of Internet Studies* (2013), 1–21.

[100] ENCK, W., GILBERT, P., CHUN, B., COX, L. P., JUNG, J., MCDANIEL, P. D., AND SHETH, A. TaintDroid: an information flow tracking system for real-time privacy monitoring on smartphones. *Commun. ACM 57*, 3 (2014), 99–106.

[101] ENEV, M., TAKAKUWA, A., KOSCHER, K., AND KOHNO, T. Automobile driver fingerprinting. *PoPETs 2016*, 1 (2016), 34–50.

[102] ERDOS, P., AND RÉNYI, A. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Science 5*, 1 (1960), 17–60.

[103] EUROPEAN COMMISSION. Factsheet on the "right to be forgotten" ruling (c-131/12). https://ec.europa.eu/info/law/law-topic/data-protection_en. Accessed May, 2018.

[104] EUROPEAN COMMISSION. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 (17/EN WP 248). http://ec.europa.eu/newsroom/document.cfm?doc_id=44137. Accessed May, 2018.

[105] EUROPEAN COMMISSION. Opinion 1/2010 on the concepts of "controller' and "processor" (00264/10/EN WP 169). http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf. Accessed Aug, 2017.

[106] EUROPEAN COMMISSION. Opinion 5/2009 on online social networking (01189/09/EN WP 163). http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2009/wp163_en.pdf. Accessed May, 2018.

[107] EUROPEAN COMMISSION. Test phase of the Data Protection Impact Assessment (DPIA) template for smart grid and smart metering systems. https://ec.europa.eu/energy/en/test-phase-data-protection-impact-assessment-dpia-template-smart-grid-and-smart-metering-systems. Accessed May, 2018.

[108] EUROPEAN DATA PROTECTION SUPERVISOR. EDPS opinion on personal information management systems. https://edps.europa.eu/data-protection/our-work/subjects/personal-information-management-system. Accessed Dec, 2017.

[109] EUROPEAN PARLIAMENT AND THE EUROPEAN COUNCIL. Guidelines on automated individual decision-making and Profiling for the purposes of Regulation 2016/679. http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826. Accessed Nov, 2017.

[110] EUROPEAN PARLIAMENT AND THE EUROPEAN COUNCIL. Opinion 15/2011 on the definition of consent (01197/11/EN WP187). http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf. Accessed May, 2018.

[111] EUROPEAN PARLIAMENT AND THE EUROPEAN COUNCIL. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. 1–20.

[112] EUROPEAN PARLIAMENT AND THE EUROPEAN COUNCIL. The ePrivacy Directive. 1–11.

[113] EUROPEAN PARLIAMENT AND THE EUROPEAN COUNCIL. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L119, 4.5.2016. 1–87.

[114] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (FRA). EU Charter of Fundamental Rights. http://fra.europa.eu/en/charterpedia/article/7-respect-private-and-family-life. Accessed Feb, 2018.

[115] EVANS, J. R., AND MATHUR, A. The value of online surveys. *Internet Research 15*, 2 (2005), 195–219.

[116] EVITA. E-safety Vehicle Intrusion Protected Applications (EVITA). http://www.evita-project.org/. Accessed Nov, 2016.

[117] FACEBOOK. https://www.facebook.com. Accessed Feb, 2018.

[118] FACEBOOK. Add Facebook login to your app or website. https://developers.facebook.com/docs/facebook-login/. Accessed Oct, 2017.

[119] FACEBOOK. Candy crush saga. https://www.facebook.com/games/candycrush. Accessed Sep, 2017.

[120] FACEBOOK. Criminal case. https://www.facebook.com/CriminalCaseGame/. Accessed Sep, 2017.

[121] FACEBOOK. Data use policy. https://www.facebook.com/about/privacy/your-info-on-other#friendsapps. Accessed Nov, 2014.

[122] FACEBOOK. Facebook privacy settings and 3rd parties. https://developers.facebook.com/docs/graph-api/reference/v2.10/user. Accessed Aug, 2017.

[123] FACEBOOK. Facebook Application Development (FAQ). https://developers.facebook.com/docs/apps/faq, Accessed April, 2018.

[124] FEDERAL TRADE COMMISSION. Android flashlight app developer settles FTC charges it deceived consumers. https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived. Accessed Aug, 2017.

[125] FEDERAL TRADE COMMISSION. Facebook, Inc. http://www.ftc.gov/enforcement/cases-proceedings/092-3184/facebook-inc. Accessed Aug, 2017.

[126] FEDERAL TRADE COMMISSION. FTC and Facebook agreement for 3rd parties wrt privacy settings. http://www.ftc.gov/sites/default/files/documents/cases/2011/11/111129facebookagree.pdf, Accessed Aug, 2017.

[127] FERRARA, E., AND FIUMARA, G. Topological features of Online Social Networks. *CoRR abs/1202.0331* (2012).

[128] FERRERO, F., PERBOLI, G., VESCO, A., CAIATI, V., AND GOBBATO, L. Car-sharing services–Part A taxonomy and annotated review. Tech. rep., 2015.

[129] FERRERO, F., PERBOLI, G., VESCO, A., MUSSO, S., AND PACIFICI, A. Car-sharing services–Part B business and service models. Tech. rep., 2015.

[130] FEW, S. *Information dashboard design - the effective visual communication of data*. O'Reilly, 2006.

[131] FISCHER-HÜBNER, S., DUQUENOY, P., HANSEN, M., LEENES, R., AND ZHANG, G., Eds. *Privacy and Identity Management for Life - 6th IFIP WG 9.2, 9.6/11.7, 11.4, 11.6/ PrimeLife International Summer School, Helsingborg, Sweden, August 2-6, 2010, Revised Selected Papers* (2011), vol. 352 of *IFIP Advances in Information and Communication Technology*, Springer.

[132] FORBES. Auto supplier continental is preparing for physical car keys to disappear. https://www.forbes.com/sites/dougnewcomb/2017/09/18/auto-supplier-continental-is-preparing-for-physical-car-keys-to-disappear. Accessed Oct, 2017.

[133] FRANK, M., DONG, B., FELT, A. P., AND SONG, D. Mining permission request patterns from android and facebook applications. In *12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012* (2012), pp. 870–875.

[134] FURUKAWA, J., LINDELL, Y., NOF, A., AND WEINSTEIN, O. High-throughput secure three-party computation for malicious adversaries and an honest majority. In *Advances in Cryptology - EUROCRYPT* (2017), pp. 225–255.

[135] GARY, S., CLARK, H., AND FERINGA, A. SP 800-27 Rev A. Engineering principles for IT security (a baseline for achieving security), Revision A. *National Institute of Standards and Technology (NIST)* (2004), 1–35.

[136] GENTRY, C. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009* (2009), pp. 169–178.

[137] GIZMODO. Researchers: Uber's iOS app had secret permissions that allowed it to copy your phone screen. https://gizmodo.com/researchers-uber-s-ios-app-had-secret-permissions-that-1819177235. Accessed Oct, 2017.

[138] GLOBAL MARKET INSIGHTS. Car sharing market size by model (P2P, Station-Based, Free-Floating). https://www.gminsights.com/industry-analysis/carsharing-market. Accessed May, 2018.

[139] GOLBECK, J., AND MAURIELLO, M. L. User perception of Facebook app data access: A comparison of methods and privacy concerns. *Future Internet 8*, 2 (2016), 9.

[140] GOLDREICH, O., GOLDWASSER, S., AND MICALI, S. How to construct random functions. *J. ACM 33*, 4 (1986), 792–807.

[141] GOLDREICH, O., MICALI, S., AND WIGDERSON, A. How to play any mental game or A completeness theorem for protocols with honest majority. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA* (1987), pp. 218–229.

[142] GOLDWASSER, S., MICALI, S., AND RIVEST, R. L. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM J. Comput. 17*, 2 (1988), 281–308.

[143] GOV.UK. Reducing mobile phone theft and improving security. https://goo.gl/o2v99g. Accessed April, 2017.

[144] GRASSI, L., RECHBERGER, C., ROTARU, D., SCHOLL, P., AND SMART, N. P. MPC-friendly symmetric key primitives. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016* (2016), pp. 430–443.

[145] GROSS, R., AND ACQUISTI, A. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, WPES 2005, Alexandria, VA, USA, November 7, 2005* (2005), pp. 71–80.

[146] GUHA, S., TANG, K., AND FRANCIS, P. NOYB: privacy in Online Social Networks. In *Proceedings of the first Workshop on Online Social Networks, WOSN 2008, Seattle, WA, USA, August 17-22, 2008* (2008), pp. 49–54.

[147] GÜRSES, S. *Multilateral privacy requirements analysis in online social networks*. PhD thesis, HMDB, Department of Computer Science, KU Leuven, Belgium, 2010.

[148] GÜRSES, S., AND DIAZ, C. Two tales of privacy in online social networks. *IEEE Security & Privacy 11*, 3 (2013), 29–37.

[149] GÜRSES, S., TRONCOSO, C., AND DIAZ, C. Engineering privacy by design. *COSIC Technical report* (2011), 1–25.

[150] HAFNER, K., AND LYON, M. *Where wizards stay up late: The origins of the Internet*. Simon and Schuster, 1998.

[151] HALL, D. L., AND LLINAS, J. An introduction to multisensor data fusion. *Proceedings of the IEEE 85*, 1 (Jan 1997), 6–23.

[152] HAMARI, J., SJÖKLINT, M., AND UKKONEN, A. The sharing economy: Why people participate in collaborative consumption. *JASIST 67*, 9 (2016), 2047–2059.

[153] HAMILTON, E., KRIENS, M., KARAPANDŽIC, H., YAICI, K., MAIN, M., AND SCHNIFFER, S. Report on trust and reputation models. *ENISA Report* (December, 2011).

[154] HANSEN, M. Marrying transparency tools with user-controlled identity management. In *The Future of Identity in the Information Society - Proceedings of the Third IFIP WG 9.2, 9.6/ 11.6, 11.7/ FIDIS International Summer School on The Future of Identity in the Information Society, Karlstad University, Sweden, August 4-10, 2007* (2007), pp. 199–220.

[155] HARDT, D. The OAuth 2.0 authorization framework. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 6749* (2012), 1–76.

[156] HARKOUS, H., AND ABERER, K. "If you can't beat them, join them": A usability approach to interdependent privacy in cloud apps. In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy, CODASPY 2017, Scottsdale, AZ, USA, March 22-24, 2017* (2017), pp. 127–138.

[157] HAYTHORNTHWAITE, C. Social networks and internet connectivity effects. *Information, Communication & Society 8*, 2 (2005), 125–147.

[158] HEDBOM, H. A survey on transparency tools for enhancing privacy. In *The Future of Identity in the Information Society - 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School, Brno, Czech Republic, September 1-7, 2008, Revised Selected Papers* (2008), pp. 67–82.

[159] HELBERGER, N., AND VAN HOBOKEN, J. Little brother is tagging you – Legal and policy implications of amateur data controllers. In *Computer Law Review International (CRi), 4/2010, 11(4), pp. 101-109* (2010).

[160] HOLLANDS, R. G. Will the real smart city please stand up? *City 12*, 3 (2008), 303–320.

[161] HOLTZ, L., ZWINGELBERG, H., AND HANSEN, M. Privacy policy icons. In *Privacy and Identity Management for Life*. 2011, pp. 279–285.

[162] HOUSING, URBAN AND OTB, MOBILITY STUDIES. Smart cities ranking of European medium-sized cities. http://www.smart-cities.eu/download/smart_cities_final_report.pdf. Accessed May, 2018.

[163] HOWARD, M., AND LIPNER, S. *The Security Development Life-Cycle (SDLC)*. O'Reilly Media, Incorporated, 2009.

[164] HU, H., AHN, G., AND JORGENSEN, J. Multiparty access control for Online Social Networks: Model and mechanisms. *IEEE Trans. Knowl. Data Eng. 25*, 7 (2013), 1614–1627.

[165] HUBER, M., MULAZZANI, M., SCHRITTWIESER, S., AND WEIPPL, E. R. Appinspect: large-scale evaluation of social networking apps. In *Conference on Online Social Networks, COSN'13, Boston, MA, USA, October 7-8, 2013* (2013), pp. 143–154.

[166] HULL, T. E., AND DOBELL, A. R. Random number generators. *SIAM review 4*, 3 (1962), 230–254.

[167] HUMBERT, M., AYDAY, E., HUBAUX, J., AND TELENTI, A. Quantifying interdependent risks in genomic privacy. *ACM Trans. Priv. Secur. 20*, 1 (2017), 3:1–3:31.

[168] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO/IEC 9797-1:2011: Information technology – Security techniques – Message Authentication Codes (MACs) – Part 1: Mechanisms using a block cipher. 1–40.

[169] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO/IEC 27001:2013: Information technology – Security techniques – Information security management systems – Requirements. 1–23.

[170] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 9241-11:2018: Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts. 1–29.

[171] INVERS. Make mobility shareable. https://invers.com/. Accessed April, 2017.

[172] JAHID, S., MITTAL, P., AND BORISOV, N. EASiER: encryption-based access control in social networks with efficient revocation. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, Hong Kong, China, March 22-24, 2011* (2011), pp. 411–415.

[173] JAHID, S., NILIZADEH, S., MITTAL, P., BORISOV, N., AND KAPADIA, A. DECENT: A decentralized architecture for enforcing privacy in online social networks. In *Tenth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2012, March 19-23, 2012, Lugano, Switzerland, Workshop Proceedings* (2012), pp. 326–332.

[174] JANCZEWSKI, L. J., WOLFE, H. B., AND SHENOI, S., Eds. *Security and Privacy Protection in Information Processing Systems, IFIP SEC, Auckland, New Zealand, July 8-10, 2013.* (2013), vol. 405 of *IFIP Advances in Information and Communication Technology*, Springer.

[175] JANIC, M., WIJBENGA, J. P., AND VEUGEN, T. Transparency Enhancing Tools (TETs): An overview. In *Third Workshop on Socio-Technical Aspects in Security and Trust, STAST 2013, New Orleans, LA, USA, June 29, 2013* (2013), pp. 18–25.

[176] JOBBER, D. *Principles and Practice of Marketing*. Principles and Practice of Marketing. McGraw-Hill, 2010.

[177] KALISKI, B., AND STADDON, J. PKCS #1: RSA cryptography specifications version 2.0. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 2437* (1998), 1–39.

[178] KELLER, M., ORSINI, E., AND SCHOLL, P. MASCOT: faster malicious arithmetic secure computation with oblivious transfer. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016* (2016), pp. 830–842.

[179] KELLY, S. G., AND FRANKEL, S. Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 4868* (2007), 1–21.

[180] KERSCHBAUM, F., AND LIM, H. W. Privacy-preserving observation in public spaces. In *Computer Security - ESORICS 2015 - 20th European Symposium on Research in Computer Security, Vienna, Austria, September 21-25, 2015, Proceedings, Part II* (2015), pp. 81–100.

[181] KEYZEE. Keys. Virtual! http://sandbox.keyzee.eu/en/autopartage/. Accessed Oct, 2017.

[182] KHODAEI, M., JIN, H., AND PAPADIMITRATOS, P. Towards deploying a scalable & robust vehicular identity and credential management infrastructure. *CoRR abs/1601.00846* (2016).

[183] KLEINBERG, J. M. The small-world phenomenon: an algorithmic perspective. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA* (2000), pp. 163–170.

[184] KOKOLAKIS, S. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security 64* (2017), 122–134.

[185] KOSINSKI, M., STILLWELL, D., AND GRAEPEL, T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences 110*, 15 (2013), 5802–5805.

[186] KRAWCZYK, H., BELLARE, M., AND CANETTI, R. HMAC: keyed-hashing for message authentication. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 2104* (1997), 1–11.

[187] KRISHNAMURTHY, B. I know what you will do next summer. *Computer Communication Review 40*, 5 (2010), 65–70.

[188] KUHN, R., HU, V., POLK, T., AND CHANG, S. SP 800-32: Introduction to Public Key technology and the federal PKI Infrastructure. *National Institute of Standards and Technology (NIST)* (2001), 1–54.

[189] LEDERER, S., HONG, J. I., DEY, A. K., AND LANDAY, J. A. Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing 8*, 6 (2004), 440–454.

[190] LEONARD, J. Systems engineering fundamentals. *Systems Management College, Department of Defence* (1999), 1–222.

[191] LI, Y., DAI, W., MING, Z., AND QIU, M. Privacy protection for preventing data over-collection in smart city. *IEEE Trans. Computers 65*, 5 (2016), 1339–1350.

[192] LIPMAA, H., AND TOFT, T. Secure equality and greater-than tests with sublinear online complexity. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II* (2013), pp. 645–656.

[193] LIU, K., AND TERZI, E. A framework for computing the privacy scores of users in online social networks. In *ICDM '09* (Washington, DC, USA, 2009), ICDM '09, IEEE Computer Society, pp. 288–297.

[194] LIU, K., AND TERZI, E. A framework for computing the privacy scores of users in Online Social Networks. *TKDD 5*, 1 (2010), 6:1–6:30.

[195] LIU, Y., GUMMADI, P. K., KRISHNAMURTHY, B., AND MISLOVE, A. Analyzing Facebook privacy settings: user expectations vs. reality. In *Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference, IMC '11, Berlin, Germany, November 2-, 2011* (2011), pp. 61–70.

[196] LUCAS, M. M., AND BORISOV, N. FlyByNight: mitigating the privacy risks of social networking. In *Proceedings of the 2008 ACM Workshop on Privacy in the Electronic Society, WPES 2008, Alexandria, VA, USA, October 27, 2008* (2008), pp. 1–8.

[197] LUO, W., XIE, Q., AND HENGARTNER, U. FaceCloak: An architecture for user privacy on social networking sites.

[198] LYFT. Drive towards what matters. https://www.lyft.com/. Accessed Dec, 2017.

[199] LYON, D. Surveillance studies: Understanding visibility, mobility and the phenetic fix. *Surveillance & Society 1*, 1 (2002), 1–7.

[200] MADEJSKI, M., JOHNSON, M. L., AND BELLOVIN, S. M. A study of privacy settings errors in an online social network. In *Tenth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2012, March 19-23, 2012, Lugano, Switzerland, Workshop Proceedings* (2012), pp. 340–345.

[201] MAHAFFEY, K. The new assembly line: 3 best practices for building (secure) connected cars. http://tinyurl.com/omgkvkc. Accessed June, 2016.

[202] MAILONLINE. TripAdvisor links to Facebook to show reviews from your friends... and their friends too. http://www.dailymail.co.uk/travel/article-2128713/TripAdvisor-links-Facebook-reviews-friends.html. Accessed Nov, 2017.

[203] MALHOTRA, N. K., KIM, S. S., AND AGARWAL, J. Internet Users' Information Privacy Concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research 15*, 4 (2004), 336–355.

[204] MARILL, T., AND ROBERTS, L. G. Toward a cooperative network of time-shared computers. In *American Federation of Information Processing Societies: Proceedings of the AFIPS '66 Fall Joint Computer Conference, November 7-10, 1966, San Francisco, California, USA* (1966), pp. 425–431.

[205] MARTÍNEZ-BALLESTÉ, A., PÉREZ-MARTÍNEZ, P. A., AND SOLANAS, A. The pursuit of citizens' privacy: a privacy-aware smart city is possible. *IEEE Communications Magazine 51*, 6 (June 2013), 136–141.

[206] MATELL, M. S., AND JACOBY, J. Is there an optimal number of alternatives for likert scale items? study i: Reliability and validity. *Educational and Psychological Measurement 31*, 3 (1971), 657–674.

[207] MAURER, U. M. Secure multi-party computation made simple. *Discrete Applied Mathematics 154*, 2 (2006), 370–381.

[208] MAXIMILIEN, E. M., GRANDISON, T., LIU, K., SUN, T., RICHARDSON, D., AND GUO, S. Enabling privacy as a fundamental construct for social networks. In *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, Vancouver, BC, Canada, August 29-31, 2009* (2009), pp. 1015–1020.

[209] MCDONNEL, N., TRONCOSO, C., TSORMPATZOUDI, P., COUDERT, F., AND MÉTAYER, L. Deliverable 5.1: State-of-play: Current practices and solutions. FP7 PRIPARE Project. http://prepareproject.eu/research/#wp5-gaps-and-recommendations. Accessed Aug, 2017.

[210] MEDIUM. The Graph API: key points in the Facebook and Cambridge Analytica debacle. https://medium.com/tow-center/the-graph-api-key-points-in-the-facebook-and-cambridge-analytica-debacle-b69fe692d747. Accessed April, 2018.

[211] MENEZES, A., VAN OORSCHOT, P. C., AND VANSTONE, S. A. *Handbook of Applied Cryptography*. CRC Press, 1996.

[212] MICALI, S. Algorand: The efficient and democratic ledger. *arXiv:1607.01341* (2016).

[213] MICROSOFT. Improving web application security: threats and countermeasures. https://msdn.microsoft.com/en-us/library/ff649874.aspx. Accessed May, 2016.

[214] MICROSOFT. Security development lifecycle. https://www.microsoft.com/en-us/SDL. Accessed Oct, 2017.

[215] MILLARD-BALL, A. *Car-sharing: Where and how it succeeds*, vol. 108. Transportation Research Board, 2005.

[216] MILLER, C., AND VALASEK, C. A survey of remote automotive attack surfaces. *Black Hat USA* (2014).

[217] MINKUS, T., AND MEMON, N. On a scale from 1 to 10, how private are you? scoring Facebook privacy settings. In *Proceedings of the Workshop on Usable Security (USEC 2014). Internet Society* (2014).

[218] MISLOVE, A., MARCON, M., GUMMADI, P. K., DRUSCHEL, P., AND BHATTACHARJEE, B. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference, IMC 2007, San Diego, California, USA, October 24-26, 2007* (2007), pp. 29–42.

[219] MORIARTY, K. M., KALISKI, B., JONSSON, J., AND RUSCH, A. PKCS #1: RSA cryptography specifications version 2.2. *Internet Engineering Task Force (IETF), Request for Comments (RFC) 8017* (2016), 1–78.

[220] MORRIS, D. FIPS PUB 202: SHA-3 Standard: Permutation-based hash and extendable-output functions. *National Institute of Standards and Technology (NIST)* (2015), 1–37.

[221] MOTHERBOARD. The data that turned the world upside down. https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win. Accessed April, 2018.

[222] MOTHERBOARD. Stolen Uber customer accounts are for sale on the dark web for 1. https://motherboard.vice.com/en_us/article/z4mk7j/stolen-uber-customer-accounts-are-for-sale-on-the-dark-web-for-1. Accessed Oct, 2017.

[223] MUSTAFA, M. A. *Smart grid security: Protecting users' privacy in smart grid applications*. PhD thesis, The University of Manchester, Manchester, UK, 2015.

[224] MUSTAFA, M. A., ZHANG, N., KALOGRIDIS, G., AND FAN, Z. Smart electric vehicle charging: Security analysis. In *IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2013, Washington, DC, USA, February 24-27, 2013* (2013), pp. 1–6.

[225] MUSTAFA, M. A., ZHANG, N., KALOGRIDIS, G., AND FAN, Z. Roaming electric vehicle charging and billing: An anonymous multi-user protocol. In *2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014, Venice, Italy, November 3-6, 2014* (2014), pp. 939–945.

[226] NAPHADE, M. R., BANAVAR, G., HARRISON, C., PARASZCZAK, J., AND MORRIS, R. Smarter cities and their innovation challenges. *IEEE Computer 44*, 6 (2011), 32–39.

[227] NBC NEWS. Researchers say it was easy to take people's data from Facebook. https://www.nbcnews.com/tech/social-media/researchers-say-it-was-easy-take-people-s-data-facebook-n858701. Accessed April, 2018.

[228] NEBEL, M., BUCHMANN, J., ROSSNAGEL, A., SHIRAZI, F., SIMO, H., AND WAIDNER, M. Personal information dashboard: Putting the individual back in control. *Digital Enlightenment* (2013).

[229] NEPALI, R. K., AND WANG, Y. SONET: A SOcial NETwork model for privacy monitoring and ranking. In *33rd International Conference on Distributed Computing Systems Workshops (ICDCS 2013 Workshops), Philadelphia, PA, USA, 8-11 July, 2013* (2013), pp. 162–166.

[230] NGOC, T. H., ECHIZEN, I., KAMIYAMA, K., AND YOSHIURA, H. New approach to quantification of privacy on social network sites. In *24th IEEE International Conference on Advanced Information Networking and Applications, AINA 2010, Perth, Australia, 20-13 April 2010* (2010), pp. 556–564.

[231] NOORMAN, J., AGTEN, P., DANIELS, W., STRACKX, R., HERREWEGE, A. V., HUYGENS, C., PRENEEL, B., VERBAUWHEDE, I., AND PIESSENS, F. Sancus: Low-cost trustworthy extensible networked devices with a zero-software trusted computing base. In *USENIX, Washington, DC, USA, August 14-16, 2013* (2013), pp. 479–494.

[232] O'NEILL, O. Some limits of informed consent. *Journal of Medical Ethics 29*, 1 (2003), 4–7.

[233] OPENSSL. Cryptography and SSL/TLS Toolkit. https://www.openssl.org/. Accessed April, 2017.

[234] OTAKEYS. Easy, convenient, innovative! http://otakeys.com/. Accessed Oct, 2017.

[235] OWASP. Application threat modeling. https://www.owasp.org/index.php/Application_Threat_Modeling. Accessed May, 2016.

[236] OWASP. Threat dragon. http://docs.threatdragon.org/. Accessed Oct, 2017.

[237] PAAR, C., AND PELZL, J. *Understanding Cryptography - A Textbook for Students and Practitioners*. Springer, 2010.

[238] PALANTIR. Products built for a purpose. https://www.palantir.com/. Accessed Jan, 2018.

[239] PAN, G., QI, G., ZHANG, W., LI, S., WU, Z., AND YANG, L. T. Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine 51*, 6 (2013).

[240] PANG, J., AND ZHANG, Y. A new access control scheme for Facebook-style social networks. *Computers & Security 54* (2015), 44–59.

[241] PAUL, T., STOPCZYNSKI, M., PUSCHER, D., VOLKAMER, M., AND STRUFE, T. C4PS - Helping Facebookers manage their privacy settings. In *Social Informatics - 4th International Conference, SocInfo 2012, Lausanne, Switzerland, December 5-7, 2012. Proceedings* (2012), pp. 188–201.

[242] PEOPLE WHO SHARE, T. The sharing economy experts. http://www.thepeoplewhoshare.com/. Accessed Feb, 2018.

[243] PETIT, J., SCHAUB, F., FEIRI, M., AND KARGL, F. Pseudonym schemes in vehicular networks: A survey. *IEEE Communications Surveys and Tutorials 17*, 1 (2015), 228–255.

[244] PETITCOLAS, F. A. P. Kerckhoffs' principle. In *Encyclopedia of Cryptography and Security, 2nd Ed.* 2011, p. 675.

[245] PFITZMANN, A., AND HANSEN, M. Anonymity, unlinkability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology. 1–98.

[246] POLIT, D. F., AND BECK, C. T. *Essentials of nursing research: methods, appraisal, and utilization*, 6th ed. ed. Lippincott Williams Wilkins, Philadelphia, 2006.

[247] PRENEEL, B. MAC algorithms. In *Encyclopedia of Cryptography and Security, 2nd Ed.* 2011, pp. 742–748.

[248] PRESERVE. Preparing secure Vehicle-to-X communication systems (PRESERVE). https://www.preserve-project.eu/. Accessed Nov, 2016.

[249] PU, Y., AND GROSSKLAGS, J. An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences. In *Decision and Game Theory for Security - 5th International Conference, GameSec 2014, Los Angeles, CA, USA, November 6-7, 2014. Proceedings* (2014), pp. 246–265.

[250] PU, Y., AND GROSSKLAGS, J. Using conjoint analysis to investigate the value of interdependent privacy in social app adoption scenarios. In *Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015, Fort Worth, Texas, USA, December 13-16, 2015* (2015).

[251] PU, Y., AND GROSSKLAGS, J. Towards a model on the factors influencing social app users' valuation of interdependent privacy. *PoPETs 2016*, 2 (2016), 61–81.

[252] PU, Y., AND GROSSKLAGS, J. Valuating friends' privacy: Does anonymity of sharing personal data matter? In *Thirteenth Symposium on Usable Privacy*

and Security, SOUPS 2017, Santa Clara, CA, USA, July 12-14, 2017. (2017), pp. 339–355.

[253] RAMAMURTHY, H., PRABHU, B., GADH, R., AND MADNI, A. M. Wireless industrial monitoring and control using a smart sensor platform. *IEEE Sensors Journal 7*, 5 (2007), 611–618.

[254] RAYA, M., PAPADIMITRATOS, P., AND HUBAUX, J. Securing vehicular communications. *IEEE Wireless Commun 13*, 5 (2006), 8–15.

[255] REDDIT. Identifying Muslim cabbies from trip data and prayer times. https://goo.gl/vLrW1s. Accessed April, 2017.

[256] RIVEST, R. L., SHAMIR, A., AND ADLEMAN, L. M. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM 21*, 2 (1978), 120–126.

[257] ROGAWAY, P. Formalizing human ignorance. In *Progressin Cryptology - VIETCRYPT 2006, First International Conferenceon Cryptology in Vietnam, Hanoi, Vietnam, September 25-28, 2006, Revised Selected Papers* (2006), pp. 211–228.

[258] ROGAWAY, P., AND SHRIMPTON, T. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In *FSE* (2004), vol. 3017 of *LNCS*, Springer, pp. 371–388.

[259] ROTONDO, S. A. Trusted Computing Group. In *Encyclopedia of Cryptography and Security, 2nd Ed.* 2011, p. 1331.

[260] ROVIO. Angry birds. https://www.facebook.com/angrybirds/. Accessed Sep, 2017.

[261] ROYCE, W. W. Managing the development of large software systems: Concepts and techniques. In *Proceedings, 9th International Conference on Software Engineering, Monterey, California, USA, March 30 - April 2, 1987.* (1987), pp. 328–339.

[262] RUST, J., AND GOLOMBOK, S. *Modern psychometrics: The science of psychological assessment*. Routledge, 2014.

[263] SACKMANN, S., STRÜKER, J., AND ACCORSI, R. Personalization in privacy-aware highly dynamic systems. *Commun. ACM 49*, 9 (2006), 32–38.

[264] SAITTA, P., LARCOM, B., AND EDDINGTON, M. Trike v. 1 methodology document. *Technical report* (2005).

[265] SÁNCHEZ, D., AND VIEJO, A. Privacy risk assessment of textual publications in social networks. In *ICAART 2015 - Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 1, Lisbon, Portugal, 10-12 January, 2015.* (2015), pp. 236–241.

[266] SANDHU, R. S., COYNE, E. J., FEINSTEIN, H. L., AND YOUMAN, C. E. Role-based access control models. *IEEE Computer 29*, 2 (1996), 38–47.

[267] SCHMARZO, B. *Big Data: Understanding how data powers big business*. John Wiley & Sons, 2013.

[268] SELENIUM HQ. Browser automation. http://docs.seleniumhq.org/. Accessed Aug, 2017.

[269] SHAHEEN, S., AND COHEN, A. Growth in worldwide carsharing: An international comparison. *Transportation Research Record: Journal of the Transportation Research Board*, 1992 (2007), 81–89.

[270] SHAHEEN, S. A., AND CHAN, N. D. *Electric Vehicle Business Models: Global Perspectives*. Springer International Publishing, 2015, ch. Evolution of E-Mobility in Car Sharing Business Models, pp. 169–178.

[271] SHAHEEN, S. A., MALLERY, M. A., AND KINGSLEY, K. J. Personal vehicle sharing services in North America. *Research in Transportation Business & Management 3* (2012), 71–81.

[272] SLASHDOT. Car manufacturers are tracking millions of cars. https://goo.gl/QvF9WN. Accessed Jan, 2018.

[273] SOFFER, S. N., AND VÁZQUEZ, A. Network clustering coefficient without degree-correlation biases. *Physical Review E 71*, 5 (2005), 057101.

[274] SOLOVE, D. J. A taxonomy of privacy. *GWU University of Pennsylvania Law Review 129* (2006), 1–88.

[275] SPIEGEL. Mobile apps doubleheader: BADASS Angry Birds. https://www.eff.org/files/2015/02/03/20150117-spiegel-exploring_and_exploiting_leaky_mobile_apps_with_badass.pdf. Accessed Feb, 2018.

[276] STATISTA. Number of car sharing users worldwide from 2006 to 2025 (in millions). https://www.statista.com/statistics/415636/car-sharing-number-of-users-worldwide/. Accessed Feb, 2018.

[277] STATISTA. Number of daily active Facebook users worldwide as of 2nd quarter 2017 (in millions). https://www.statista.com/statistics/346167/facebook-global-dau/. Accessed Aug, 2017.

[278] STATISTA. Number of smartphone users in the United States from 2010 to 2021. https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/. Accessed April, 2017.

[279] STINE, K., KISSEL, R., BARKER, W., LEE, A., FAHLSING, J., AND GULICK, J. SP 800-60 Vol. 1 Rev. 1: Guide for mapping types of information and information systems to security categories (2 vols.). *National Institute of Standards and Technology (NIST)* (2008), 1–357.

[280] STINSON, D. R. Some observations on the theory of cryptographic hash functions. *Des. Codes Cryptography 38*, 2 (2006), 259–277.

[281] STOLFI, D. H., AND ALBA, E. Eco-friendly reduction of travel times in european smart cities. In *Genetic and Evolutionary Computation Conference, GECCO '14, Vancouver, BC, Canada, July 12-16, 2014* (2014), pp. 1207–1214.

[282] SUH, G. E., CLARKE, D. E., GASSEND, B., VAN DIJK, M., AND DEVADAS, S. AEGIS: architecture for tamper-evident and tamper-resistant processing. In *Proceedings of the 17th Annual International Conference on Supercomputing, ICS 2003, San Francisco, CA, USA, June 23-26, 2003* (2003), pp. 160–171.

[283] SUNDARARAJAN, A. Local network effects and complex network structure. *The B.E. Journal of Theoretical Economics 7*, 1 (2007).

[284] SWEENEY, L. Simple demographics often identify people uniquely. *Health (San Francisco) 671* (2000), 1–34.

[285] SYMEONIDIS, I., ALY, A., MUSTAFA, M. A., MENNINK, B., DHOOGHE, S., AND PRENEEL, B. SePCAR: A Secure and Privacy-Enhancing Protocol for Car Access Provision. In *Computer Security - ESORICS 2017 - 22nd European Symposium on Research in Computer Security, Oslo, Norway, September 11-15, 2017, Proceedings, Part II* (2017), pp. 475–493.

[286] SYMEONIDIS, I., BEATO, F., TSORMPATZOUDI, P., AND PRENEEL, B. Collateral damage of Facebook Apps: an enhanced privacy scoring model. *IACR Cryptology ePrint Archive 2015* (2015), 456.

[287] SYMEONIDIS, I., BICZÓK, G., SHIRAZI, F., PÉREZ-SOLÀ, C., SCHROERS, J., AND PRENEEL, B. Collateral damage of Facebook third-party applications: a comprehensive study. *Computers & Security, Elsevier 77* (2018), 179 – 208.

[288] SYMEONIDIS, I., MUSTAFA, M. A., AND PRENEEL, B. Keyless car sharing system: A security and privacy analysis. In *IEEE ISC2* (2016), pp. 1–7.

[289] SYMEONIDIS, I., AND PRENEEL, B. SARL: A revocation mechanism for long lived assertions on Shibboleth. *COSIC Technical report* (2014), 1–6.

[290] SYMEONIDIS, I., SHIRAZI, F., BICZÓK, G., PÉREZ-SOLÀ, C., AND PRENEEL, B. Collateral damage of Facebook apps: Friends, providers, and privacy interdependence. In *ICT Systems Security and Privacy Protection - 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30 - June 1, 2016, Proceedings* (2016), pp. 194–208.

[291] SYMEONIDIS, I., TSORMPATZOUDI, P., AND PRENEEL, B. Collateral damage of online social network applications. In *Proceedings of the 2nd International Conference on Information Systems Security and Privacy, ICISSP 2016, Rome, Italy, February 19-21, 2016.* (2016), pp. 536–541.

[292] SYMEONIDIS I. COSIC Blog post: Collateral damage of Facebook apps. https://www.esat.kuleuven.be/cosic/collateral-damage-of-facebook-apps/. Accessed April, 2018.

[293] TALUKDER, N., OUZZANI, M., ELMAGARMID, A. K., ELMELEEGY, H., AND YAKOUT, M. Privometer: Privacy protection in social networks. In *Workshops Proceedings of the 26th International Conference on Data Engineering, ICDE 2010, March 1-6, 2010, Long Beach, California, USA* (2010), pp. 266–269.

[294] TASHAKKORI, A., AND TEDDLIE, C. *Mixed methodology: Combining qualitative and quantitative approaches*, vol. 46. Sage, 1998.

[295] TESLA. Blog. https://www.teslamotors.com/blog/tragic-loss. Accessed June, 2016.

[296] TESLA. Bugcrowd. https://bugcrowd.com/tesla. Accessed June, 2016.

[297] THE GUARDIAN. Facebook announces dating app focused on "meaningful relationships". https://www.theguardian.com/technology/2018/may/01/facebook-dating-app-mark-zuckerberg-f8-conference. Accessed May, 2018.

[298] THE GUARDIAN. Facebook says Cambridge Analytica may have gained 37 million more users' data. https://www.theguardian.com/technology/2018/apr/04/facebook-cambridge-analytica-user-data-latest-more-than-thought. Accessed April, 2018.

[299] THE GUARDIAN. GCHQ taps fibre-optic cables for secret access to world's communications. http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa. Accessed Feb, 2018.

[300] THE GUARDIAN. Hell of a ride: even a PR powerhouse couldn't get Uber on track. https://goo.gl/UcIihE. Accessed April, 2017.

[301] THE GUARDIAN. NSA Prism program taps in to user data of Apple, Google and others. http://www.theguardian.com/world/2013/jun/06/us-tech- giants-nsa-data. Accessed Feb, 2018.

[302] THE GUARDIAN. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. Accessed Feb, 2018.

[303] THE GUARDIAN. The Uber scammers who take users for a (very expensive) ride. https://www.theguardian.com/money/2016/apr/22/uber-scam-hacking-account-phantom-journeys. Accessed Oct, 2017.

[304] THE GUARDIAN. "Tor Stinks" presentation. https://edwardsnowden.com/docs/doc/tor-stinks-presentation.pdf. Accessed Feb, 2018.

[305] THOMAS, K., GRIER, C., AND NICOL, D. M. unFriendly: Multi-party privacy risks in social networks. In *Privacy Enhancing Technologies, 10th International Symposium, PETS 2010, Berlin, Germany, July 21-23, 2010. Proceedings* (2010), pp. 236–252.

[306] TIME. Uber is tracking drivers' phones to watch for speeding. http://time.com/4387031/uber-driver-app-tracking/. Accessed April, 2017.

[307] TOR PROJECT. Metrics. https://metrics.torproject.org/torperf.html. Accessed April, 2017.

[308] TOR PROJECT. Protect your privacy. Defend yourself against network surveillance and traffic analysis. https://www.torproject.org/. Accessed April, 2017.

[309] TRAVERS, J., AND MILGRAM, S. An experimental study of the small world problem. *Sociometry* (1969), 425–443.

[310] TRIPADVISOR. Tripadvisor. https://www.facebook.com/games/tripadvisor. Accessed Jan, 2016.

[311] TRONCOSO, C., DANEZIS, G., KOSTA, E., BALASCH, J., AND PRENEEL, B. PriPAYD: Privacy-Friendly Pay-As-You-Drive Insurance. *IEEE TDSC 8*, 5 (2011), 742–755.

[312] TRUSTED COMPUTING GROUP. TPM 2.0 library profile for automotive-thin. https://trustedcomputinggroup.org/wp-content/uploads/TCG-TPM-2.0-Automotive-Thin-Profile_v1.0.pdf. Accessed May, 2018.

[313] UCEDAVELEZ, T., AND MORANA, M. M. *Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis*. John Wiley & Sons, 2015.

[314] UGANDER, J., KARRER, B., BACKSTROM, L., AND MARLOW, C. The anatomy of the Facebook social graph. *CoRR abs/1111.4503* (2011).

[315] UNITED STATES PATENT AND TRADEMARK OFFICE. APPLICANT: APPLE INC. Accessing a vehicle using portable devices. https://goo.gl/a9pyX7. Accessed Sept, 2017.

[316] USA TODAY. Toyota will test keyless car sharing. https://goo.gl/C9iq34. Accessed Nov, 2016.

[317] VAN TEIJLINGEN, E. R., AND HUNDLEY, V. The importance of pilot studies. *Nursing Standard (through 2013) 16*, 40 (2001), 33–6.

[318] VERDON, D., AND MCGRAW, G. Risk analysis in software design. *IEEE Security & Privacy 2*, 4 (2004), 79–84.

[319] VIEJO, A., AND SÁNCHEZ, D. Enforcing transparent access to private content in social networks by means of automatic sanitization. *Expert Syst. Appl. 62* (2016), 148–160.

[320] VOLVO. Worth a Detour. https://www.sunfleet.com/. Accessed Nov, 2016.

[321] VU, L., ABERER, K., BUCHEGGER, S., AND DATTA, A. Enabling secure secret sharing in distributed online social networks. In *Twenty-Fifth Annual Computer Security Applications Conference, ACSAC 2009, Honolulu, Hawaii, 7-11 December 2009* (2009), pp. 419–428.

[322] WANG, N., GROSSKLAGS, J., AND XU, H. An online experiment of privacy authorization dialogues for social applications. In *Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013* (2013), A. Bruckman, S. Counts, C. Lampe, and L. G. Terveen, Eds., ACM, pp. 261–272.

[323] WANG, N., XU, H., AND GROSSKLAGS, J. Third-party apps on Facebook: Privacy and the illusion of control. In *5th ACM CHIMIT* (NY, USA, 2011), ACM, pp. 4:1–4:10.

[324] WANG, T., SRIVATSA, M., AND LIU, L. Fine-grained access control of personal data. In *17th ACM Symposium on Access Control Models and Technologies, SACMAT '12, Newark, NJ, USA - June 20 - 22, 2012* (2012), pp. 145–156.

[325] WANG, Y., LEON, P. G., SCOTT, K., CHEN, X., ACQUISTI, A., AND CRANOR, L. F. Privacy nudges for social media: an exploratory Facebook study. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume* (2013), pp. 763–770.

[326] WANG, Y., NORCIE, G., KOMANDURI, S., ACQUISTI, A., LEON, P. G., AND CRANOR, L. F. "I regretted the minute I pressed share": A qualitative study of regrets on Facebook. In *Symposium On Usable Privacy and Security, SOUPS'11, Pittsburgh, PA, USA - July 20 - 22, 2011* (2011), p. 10.

[327] WARREN, S. D., AND BRANDEIS, L. D. Das Recht auf Privatheit - The Right to privacy - Originalveröffentlichung in Harvard Law Review, Vol. IV Dec. 15 , 1890 No. 5; übersetzt und mit Zwischenüberschriften versehen von Marit Hansen und Thilo Weichert. *Datenschutz und Datensicherheit 36*, 10 (2012), 755–766.

[328] WÄSTLUND, E., AND FISCHER-HÜBNER, S. End user transparency tools: UI prototypes, 2010.

[329] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of "small-world" networks. *Nature 393*, 6684 (1998), 409–10.

[330] WEIKL, S., AND BOGENBERGER, K. Relocation strategies and algorithms for free-floating car sharing systems. *IEEE Intell. Transport. Syst. Mag. 5*, 4 (2013), 100–111.

[331] WEITZNER, D. J., ABELSON, H., BERNERS-LEE, T., HANSON, C., HENDLER, J. A., KAGAL, L., MCGUINNESS, D. L., SUSSMAN, G. J., AND WATERMAN, K. K. Transparent accountable data mining: New strategies for privacy protection. In *Semantic Web Meets eGovernment, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-06, Stanford, California, USA, March 27-29, 2006* (2006), p. 141.

[332] WESTIN, A. F., AND RUEBHAUSEN, O. M. *Privacy and Freedom*. Ig Publishing, 2015.

[333] WIELINSKI, G., TRÉPANIER, M., AND MORENCY, C. Electric and hybrid car use in a free-floating carsharing system. *International Journal of Sustainable Transportation 11*, 3 (2017), 161–169.

[334] WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K. P. N., AND ZHAO, B. Y. User interactions in social networks and their implications. In *Proceedings of the 2009 EuroSys Conference, Nuremberg, Germany, April 1-3, 2009* (2009), pp. 205–218.

[335] WIRED. The Facebook privacy settings that doesn't do anything at all. https://www.wired.com/story/facebook-privacy-setting-doesnt-do-anything/. Accessed April, 2018.

[336] WOLF, M., AND GENDRULLIS, T. Design, implementation, and evaluation of a vehicular hardware security module. In *Information Security and Cryptology - ICISC 2011 - 14th International Conference, Seoul, Korea, November 30 - December 2, 2011. Revised Selected Papers* (2011), pp. 302–318.

[337] WRIGHT, D., AND DE HERT, P. *Privacy Impact Assessment*, vol. 6. Springer Science & Business Media, 2011.

[338] XU, H., WANG, N., AND GROSSKLAGS, J. Privacy by redesign: Alleviating privacy concerns for third-party apps. In *Proceedings of the International Conference on Information Systems, ICIS 2012, Orlando, Florida, USA, December 16-19, 2012* (2012).

[339] YAO, A. C. How to generate and exchange secrets (extended abstract). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986* (1986), pp. 162–167.

[340] ZHOU, J., AND GOLLMANN, D. Evidence and non-repudiation. *J. Netw. Comput. Appl. 20*, 3 (July 1997), 267–281.

[341] ZHU, X., LI, J., LIU, Z., AND YANG, F. Optimization approach to depot location in car sharing systems with big data. In *2015 IEEE International Congress on Big Data, New York City, NY, USA, June 27 - July 2, 2015* (2015), pp. 335–342.

[342] ZIPCAR. Keyless car sharing. http://www.zipcar.com/. Accessed November, 2016.

List of Publications

International journals

2018 **I. Symeonidis, F. Shirazi, G. Biczok, C. Pérez-Solà, J. Schroers and B. Preneel**, Collateral damage of Facebook third-party applications: a Comprehensive study, Computers & Security (COSY), Elsevier, volume 77, 2018, pp. 179 - 208.

International conferences

2017 **I. Symeonidis, A. Aly, M. A. Mustafa, B. Mennink, S. Dhooghe, and B. Preneel**, SePCAR: A Secure and Privacy-Enhancing Protocol for Car Access Provision, Computer Security - ESORICS 2017 – 22nd European Symposium on Research in Computer Security, Oslo, Norway, September 11-15, 2017, Proceedings, Part II, pp. 475–493.

2016 **I. Symeonidis, M. A. Mustafa, and B. Preneel**, Keyless car sharing system: A security and privacy analysis, IEEE International Smart Cities Conference, ISC2 2016, Trento, Italy, September 12-15, 2016, Proceedings, pp. 1–7.

2016 **I. Symeonidis, F. Shirazi, G. Biczok, C. Pérez-Solà, and B. Preneel**, Collateral damage of Facebook apps: Friends, providers, and privacy interdependence, ICT Systems Security and Privacy Protection - 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30 - June 1, 2016, Proceedings, pp. 194–208.

2016 **I. Symeonidis, P. Tsormpatzoudi, and B. Preneel**, Collateral damage of Online Social Network applications, Proceedings of the 2nd International Conference on Information Systems Security and Privacy, ICISSP 2016, Rome, Italy, February 19-21, 2016, Proceedings, pp. 536–541.

2015 **D. Al-Azizy, D. E. Millard, I. Symeonidis, K. O'Hara and N. Shadbolt**, A Literature survey and classifications on data de-anonymisation, Risks and Security of Internet and Systems - 10th International Conference, CRiSIS 2015, Mytilene, Lesbos Island, Greece, July 20-22, 2015, Proceedings, pp. 36–51.

Miscellaneous

2016 **K. Domin, E. Marin, and I. Symeonidis**, Security analysis of the drone communication protocol: Fuzzing the MAVLink protocol, In Proceedings of the 37th Symposium on Information Theory in the Benelux, Werkgemeenschap voor Informatie- en Communicatietheorie, 2016, pp. 1–7.

2015 **I. Symeonidis, P. Tsormpatzoudi, and B. Preneel**, Collateral damage of Facebook apps: an enhanced privacy scoring model, IACR Cryptology ePrint Archive 2015/456, 2015, COSIC Technical report, pp. 1–7.

2014 **I. Symeonidis**, SARL: A revocation mechanism for long-lived assertions on Shibboleth, 2014, COSIC Technical report, pp. 1–6.

Curriculum Vitae

Look up at the stars and not down at your feet. Try to make sense of what you see, and wonder about what makes the universe exist. Be curious.

STEPHEN HAWKING, *Theoretical physicist, cosmologist, and author*

Iraklis Symeonidis was born on August 22nd in Alexandroupolis, Greece.

He received his engineering degree in Information and Communication Systems Engineering from the University of the Aegean in 2004. He conducted his thesis in collaboration with the National Centre of Scientific Research "Demokritos" in Athens, Greece. In 2005 he served in Hellenic Armed Forces at the Networks and Telecommunication's branch. From 2005 to 2012 he worked as a teacher in programming for the Ministry of Education in Greece and also as a back-end web developer.

In 2013, he received his MS.c degree in Digital Systems Security from the University of Piraeus, Greece. For his MS.c thesis, he joined the COSIC research group at the Department of Electrical Engineering (ESAT) of KU Leuven, Belgium as an exchange student, in February 2013. In November 2013, he started as PhD candidate at COSIC supervised by Prof. Bart Preneel. His current research interests fall in the intersection between information security and privacy with a specific focus on *security by design* and *privacy by design*. Currently, he is working on the analysis and design of solutions for secure and privacy-enhancing information sharing systems considering online social networks and connected cars.

Τὰ πάντα ῥεῖ καὶ οὐδὲν μένει!
HERACLITUS, Greek philosopher

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
IMEC-COSIC
Kasteelpark Arenberg 10 Bus 2452, B-3001 Leuven, Belgium
B-3001 Leuven
iraklis.symeonidis@esat.kuleuven.be
http://cosic.esat.kuleuven.be