Nurul Momen

MEASURING APPS' PRIVACY-FRIENDLINESS

Introducing transparency to apps' data access behavior

Measuring Apps' Privacy-Friendliness

Introducing Transparency to Apps' Data Access Behavior

Nurul Momen

Faculty of Health, Science and Technology

Computer Science

DOCTORAL THESIS | Karlstad University Studies | 2020:24

Measuring Apps' Privacy-Friendliness

Introducing Transparency to Apps' Data Access Behavior

Nurul Momen

DOCTORAL THESIS | Karlstad University Studies | 2020:24

Measuring Apps' Privacy-Friendliness/ Introducing Transparency to Apps' Data Access Behavior

Nurul Momen

DOCTORAL THESIS

Karlstad University Studies | 2020:24

urn:nbn:se:kau:diva-79308

ISSN 1403-8099

ISBN 978-91-7867-132-8 (print)

ISBN 978-91-7867-137-3 (pdf)

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Mathematics and Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2020

WWW.KAU.SE

Measuring Apps' Privacy-Friendliness
Introducing Transparency to Apps' Data Access Behavior

NURUL MOMEN
Department of Mathematics and Computer Science
Karlstad University

Abstract

Mobile apps brought unprecedented convenience to everyday life, and nowadays, hardly any interactive service exists without having an interface through an app. The rich functionalities of apps rely on the pervasive capabilities of the mobile device, such as its cameras and other types of sensors. Consequently, apps generate a diverse and large amount of data, which can often be deemed as privacy-sensitive data. As the mobile device is also equipped with several means to transmit the collected data, such as WiFi and 4G, it brings further concerns about individuals' privacy.

Even though mobile operating systems use access control mechanisms to guard system resources and sensors, apps exercise their granted privileges in an opaque manner. Depending on the type of privilege, apps require explicit approval from the user in order to acquire access to them through permissions. Nonetheless, granting permission does not put constraints on the access frequency. Granted privileges allow the app to access users' personal data for a long period of time, typically until the user explicitly revokes the access. Furthermore, available control tools lack monitoring features, and therefore, the user faces hindrances to comprehend the magnitude of personal data access. Such circumstances can erode intervenability from the interface of the phone, lead to incomprehensible handling of personal data, and thus, create privacy risks for the user.

This thesis covers a long-term investigation of apps' data access behavior and makes an effort to shed light on various privacy implications. It also shows that app behavior analysis yields information that has the potential to increase transparency, to enhance privacy protection, to raise awareness regarding consequences of data disclosure, and to assist the user in informed decision-making while selecting apps or services. We introduce models, methods, and demonstrate the data disclosure risks with experimental results. Finally, we show how to communicate privacy risks through the user interface by taking the results of app behavior analyses into account.

Keywords: Mobile Apps, User data, Transparency, Privacy, Data protection

Acknowledgements

Though this thesis has only my name on top of the cover page, it would not have seen daylight without the help, numerous suggestions and continuous guidance of my supervisor—Lothar Fritsch and co-supervisor—Tobias Pulls. I also would like to express gratitude to my colleagues: Stefan Lindskog and Simone Fischer Hübner, who offered their generous counsel on countless occasions. In fact, I am indeed thankful to all the colleagues at the Department of Mathematics and Computer Science who contributed with their priceless time and inputs.

To my co-authors and colleagues from TU Berlin and Goethe University Frankfurt, I am indeed thankful for your professionalism, knowledge, and immense effort in collaboration. I look forward to have further opportunities to work with you in future.

Thanking my family and friends

To the fellow human beings who make my life worth living for, Mom and Dad, thank you for your unconditional love and affection that kept me going.

To my wife, thank you for being patient, tolerating me, and seeing me when I am invisible. I would not even dare, and perhaps be incapable of loving you in the way that you love me. I will never be able to fathom the depth of your affection, kindness, and support. I would rather prefer to grow old while wondering about it.

To my precious friends whom I greedily collected throughout the journey, you people gave me memories, and I am fond of cherishing them with gratitude. Thanks a ton!

A page from my diary

It has been a couple of years since I had to gather recollection from my childhood to write about research. At that time, I recalled the very first experience with a *telephone connected to wire*. Since that tender moment with my father, nearly three decades went past; I grew older, and that telephone had evolved into today's smartphone. I suppose, I have been longing for another trip down memory lane. I became so desperate that I bought an old-school phone which can be used for making phone calls and sending text messages only. As soon as I got rid of the so-called smartphone, I became unplugged from the captivating world of apps. I can assure you that the experience is just extremely liberating.

Perhaps I have been enjoying and appreciating the time a little too much. All of a sudden, I had the time to call friends from my long-lost childhood—a real phone call. I even found time to write a few words in an old-fashioned diary, which used to be a regular thing for myself. Apparently while writing, I became very sad—I hardly had any memory to write about. I came to work everyday, went back, ate, slept, did household chores, groceries, and repeated. Suddenly, I realized that I had the memory of just one day out of the last few

hundreds of days. Some events could be popped out of the ordinary, but that seldom happened.

It makes me humble by acknowledging my mere existence in the universe. So, I would like to express my gratitude to you, dear Universe. This is the most important artifact that I have produced so far. Hopefully and probably, this piece will remain on your timeline for a bit longer period than my physical self. I do not have any scientific evidence about spiritual existence yet. Hence, I am feeling happy and blessed in this insignificant moment which is certainly less than a dot on the timeline of eternity. Though I can hardly make sense of this oblivious phenomenon, *time* is all that I lost and got to lose. So, why do I keep my eyes glued to this tiny screen and put aside the entire world? However, my gratitude goes to the—*Mobile phone*, due to the fact that its underlying vulnerability kept me employed for the last few years.

Last but not the least, it would be rather unfair if I forget to thank a non-living entity—the coffee machine at the department.

Karlstad University, August 24, 2020 Nurul Momen

List of Included Papers

This thesis is based on the work presented in the following peer-reviewed papers:

I. **Nurul Momen**, Tobias Pulls, Lothar Fritsch, and Stefan Lindskog. "How much Privilege does an App Need? Investigating Resource Usage of Android Apps (short paper)." In *Proceedings of the Fifteenth International Conference on Privacy, Security and Trust (PST)*, pp. 268–273, IEEE, 2017.

II. Lothar Fritsch and **Nurul Momen**. "Derived Partial Identities Generated from App Permissions." In *Proceedings of the Open Identity Summit, Lecture Notes in Informatics (LNI), Volume P-277, (pp. 117-130)*, Gesellschaft für Informatik e.V., 2017.

III. Majid Hatamian, **Nurul Momen**, Lothar Fritsch, and Kai Rannenberg. "A Multilateral Privacy Impact Analysis Method for Android Apps." In *Proceedings of the Annual Privacy Forum, pp. 87–106*, Lecture Notes in Computer Science, vol. 11498, Springer, 2019.

IV. **Nurul Momen**, Majid Hatamian, and Lothar Fritsch. "Did App Privacy Improve after the GDPR?" In *IEEE Security & Privacy*, vol. 17, no. 6, pp. 10-20, Nov.–Dec. 2019.

V. **Nurul Momen** and Lothar Fritsch. "App-generated Digital Identities Extracted through Android Permission-based Data Access—A Survey of App Privacy." In *Proceedings of the Sicherheit—Schutz und Zuverlässigkeit, Volume P-301, (pp. 15-28)*, Gesellschaft für Informatik e.V., 2020.

VI. Sven Bock and **Nurul Momen**. "Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User." To appear in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction (NordiCHI '20)*, ACM, 2020.

VII. **Nurul Momen**, Sven Bock, and Lothar Fritsch. "Accept - Maybe - Decline: Introducing Partial Consent for the Permission-based Access Control Model of Android." *In Proceedings of the 25th ACM Symposium on Access Control Models and Technologies (SACMAT '20)*, 2020.

The included papers have been subjected to minor editorial changes.

Comments on my Participation

- **Paper I** The idea of this paper and the design of the experiment, originated from a discussion with Stefan Lindskog. I carried out the experiment phases, collected data and ran the analysis. I also authored the draft of this paper. Tobias Pulls and Lothar Fritsch helped by offering their insights on writing the background and related work sections. They also provided feedback on the data collection procedure and analysis strategy.

- **Paper II** Lothar Fritsch developed the idea to correlate apps' resource access efforts with the risks associated with partial identity disclosure and I provided the permission classification and empirical insight into it. We discussed and developed a graph-based model for generating partial identities. We both contributed equally in writing the manuscript.

- **Paper III** This paper is the first outcome of our collaboration with the Chair of Mobile Business and Multilateral Security at the Goethe University Frankfurt, Germany. I substantially contributed to this paper with data-gathering, data analysis, and visualization efforts. I conducted the static and dynamic analyses, and compiled the multi-sourced data in order to produce the cumulative ranking of apps.

- **Paper IV** For this paper, we continued the collaboration effort with Goethe University Frankfurt, Germany. As the lead author, I contributed with planning and writing the manuscript, commencing data collection campaigns, analyzing data, and visualizing the results. I conducted both the static and dynamic analyses.

- **Paper V** In this paper, I and Lothar Fritsch, both contributed equally in writing the manuscript, addressing the issues from the previously published identity attribute model, populating the model with data collected from two isolated collection campaigns, and visualizing the changes that were probably caused by regulatory shift.

- **Paper VI** This paper is the outcome of our collaboration with the Chair of Human-Machine Systems at the Technische Universität Berlin, Germany. Both co-authors contributed equally to this paper, and shared responsibilities for design and implementation, experimentation, and preparing the manuscript.

- **Paper VII** The continuation of our research collaboration with the Technische Universität Berlin resulted in this paper. Here, we addressed the research problem from various perspectives, discussed promising research directions, and presented a prototype as the proof of concept. As the lead author, I contributed with planning and writing the manuscript, and developing the prototype.

List of Other Publications

In addition to the previously mentioned publications of my thesis research, the following articles have also been published (or remained under submission):

1. Majid Hatamian, Samuel Wairimu, **Nurul Momen**, and Lothar Fritsch. A privacy and security analysis of COVID-19 contact tracing apps: quality indicators for rapidly deployed apps into a health pandemic context; *Under Submission*.

2. Sven Bock and **Nurul Momen**. A Study on User Preference: Impact of a Privacy Indicator on App Selection Behavior. To appear in *Proceedings of the 22nd International Conference on Human-Computer Interaction*, Communications in Computer and Information Science, vol 1226, Springer, Cham 2020.

3. Sven Bock and **Nurul Momen**. Einfluss einer Datenschutzskala auf das Auswahl-verhalten in einem App-Markt. In Frühjahrskongress 2020-Digitaler Wandel, digitale Arbeit, digitaler Mensch? GfA Press, 2020.

4. **Nurul Momen** and Sven Bock. Neither Do I Want to Accept, nor Decline, Is There an Alternative? In *Proceedings of the 22nd International Conference on Human-Computer Interaction*, Communications in Computer and Information Science, vol 1226, Springer, 2020.

5. **Nurul Momen** and Lothar Fritsch. A Curious Case of "Consent Button"; Synthesis: PET Symposium; 2019 Jul 19; 3:A4.

6. **Nurul Momen** and Sven Bock. User Perception Analysis for Showing Personal Data Access as Privacy Implication Factor; In NordSec 2018 *(Best Poster Award)*, The 23rd Nordic Conference on Secure IT Systems, University of Oslo, 2018.

7. **Nurul Momen**. Turning the Table Around: Monitoring App Behavior; In *Proceedings of the Sicherheit – Schutz und Zuverlässigkeit 2018, Volume P-281, Bonn, Germany*, Gesellschaft für Informatik e.V., (S. 279-284), 2018.

8. **Nurul Momen** and Marta Piekarska. Towards improving privacy awareness regarding apps' permissions. In *Proceedings of the 11th International Conference on Digital Society (ICDS)*, pp. 18-23, 2017.

Acknowledging the Student Projects

My research involved offering student projects on an advanced level, leading to prototypes and exploration of the research project. I would like to thank our students in computer science and engineering for their collaboration. Their results are listed below:

Technical Reports

I. Ludwig Toresson, Sebastian Olars, and Maher Shaker. "Privacy impact self-assessment app"; Technical Project Report (2020), Karlstad University Press [124].

II. Simon Sundberg, Alexander Blomqvist, and Anton Bromander. "KAUdroid-Project Report: Visualizing how Android apps utilize permissions"; Technical Project Report (2019), Karlstad University Press [114].

III. Adrian Carlsson, Christian Pedersen, Fredrik Persson, and Gustaf Söderlund. "KAUdroid: A tool that will spy on applications and how they spy on their users"; Technical Project Report (2018), Karlstad University Press [20].

Master Theses

1. Anton Bromander. "Using Privacy Indicators to Nudge Users into Selecting Privacy Friendly Applications"; Master Thesis (2019), Karlstad University Press [15].

2. Ulf Magnusson. "A tool for visual analysis of permission-based data access on Android phones"; Master Thesis (2019), Karlstad University Press [78].

3. Ashraf Ferdouse Chowdhury; "Introducing Partial Consent to the User Interface: Design, Implementation, and User Study"; Master Thesis (ongoing), Co-supervision with the Department of Computer Science, Stockholm University.

List of Abbreviations

(Used in the Introductory Summary)

- ACM: Association for Computing Machinery.
- AOSP: Android Open Source Project.
- API: Application Programming Interface.
- CS: Computer Science.
- DSRM: Design Science Research Methodology.
- EC: Exclusion Criteria.
- FAIR: Fuzzy Alarming Index Rule.
- GDPR: General Data Protection Regulation.
- HCI: Human Computer Interaction.
- IC: Inclusion Criteria.
- ID: Identity.
- IEEE: Institute of Electrical and Electronics Engineers.
- OS: Operating System.
- PoLP: Principle of Least Privilege.
- PETs: Privacy Enhancing Technologies.
- TETs: Transparency Enhancing Technologies.

Contents

List of Included Papers vii

List of Other Publications ix

List of Student Projects and Theses x

List of Abbreviations xi

Prologue xix

Introductory Summary 1

1 Introduction 3
- 1.1 Research Questions . . . 4
- 1.2 Thesis Structure . . . 5

2 Background and Related Work 6
- 2.1 Technical Background . . . 8
  - 2.1.1 The Android Operating System . . . 8
  - 2.1.2 Permission Types, Manifest and API . . . 8
  - 2.1.3 Problem Definition: Technical Context . . . 9
- 2.2 Related Work . . . 9
  - 2.2.1 App-behavior Analysis . . . 10
  - 2.2.2 Responsibility of Awareness . . . 11
- 2.3 Literature Mapping Study . . . 11
  - 2.3.1 Method of Mapping Study . . . 11
  - 2.3.2 Survey Procedure . . . 13
  - 2.3.3 Findings from the Literature Mapping Study . . . 14

3 Research Methods 16
- 3.1 Method Iteration . . . 17
- 3.2 Methods Used in the Papers . . . 18
- 3.3 Limitations of Our Approach . . . 19

4 Contributions 19

5 Summary of Included Papers 22

6 Concluding Remarks and Outlook 24

Appendices 40

A Appendix: Repositories 40

B Appendix: Tools and Data archives 40

Paper I:
How much Privilege does an App Need? Investigating Resource Usage of Android Apps **41**

1 Introduction **43**

2 Related Work **45**

3 Android and Scenarios **46**
3.1 OS Architecture and Permissions 46
3.1.1 Permission Types 46
3.1.2 The Principle of Least Privilege 46
3.2 Scenario One: The Forgotten Decisions 47
3.3 Scenario Two: The Consistent Inconsistency 47

4 Methodology **48**
4.1 Device and Experimental Platform 48
4.2 Prototype: Monitoring App 48
4.3 Stationary Phase 49
4.3.1 Test Accounts 50
4.3.2 App Selection 50
4.4 User Interaction Phase 50
4.4.1 Pseudonymous Users 50
4.4.2 Constraints 52

5 Results **52**
5.1 Idle-time Usage 52
5.2 Bundled Permission Usage 52
5.3 Comparison Between Phases 53
5.4 Temporal Usage Discrepancy 53

6 Discussions and Limitations **55**

7 Concluding Remarks **55**

Paper II:
Derived Partial Identities Generated from App Permissions **59**

1 Introduction **61**

2 Android and Permissions **62**
2.1 App Permission: Privacy Issues 62
2.2 Access to Identifiable Information Through Permissions . . . 63
2.3 Identification and Partial Identities 63

3 Identity-related Attributes Accessible Through Permissions **64**

4 Model-based Assessment of Permission-accessible Partial Identities 65
4.1 A Model for Permission-based Partial Identity Retrieval . . . 67
4.2 Assessment of an App's Access to Partial Identities 67

5 Stealing My Identity: A Survey of Actual App Behavior 67
5.1 Survey Procedure 68
5.2 Results 70
5.2.1 Idle-time Usage: Phones Never Sleep 70
5.2.2 Partial Identity Extraction 70

6 Limitations and Future Work 72

7 Conclusion 73

Paper III:
A Multilateral Privacy Impact Analysis Method for Android Apps 77

1 Introduction 79

2 Data Acquisition Methodology 81
2.1 Permission Manifest Analysis (A1) 81
2.2 Privacy Policy Analysis (A2) 82
2.3 Permission Usage Analysis (A3) 82
2.4 User Reviews Analysis (A4) 83

3 Multilateral Analysis 83
3.1 Step A1: Permission Manifest Analysis 84
3.1.1 Data Collection 85
3.1.2 Permission Request Analysis 85
3.2 Step A2: Privacy Policy Analysis 85
3.2.1 Data Collection 85
3.2.2 Purpose Specification Analysis 87
3.3 Step A3: Permission Usage Analysis 87
3.3.1 Data Collection 87
3.3.2 Permission Access Analysis 88
3.4 Step A4: User Reviews Analysis 89
3.4.1 Data Collection 89
3.4.2 Privacy Relevant Complaints Analysis 89
3.4.3 The Most Mentioned Permissions 90
3.5 Synthesis of Analysis 90

4 Related Work 92

5 Conclusions and Future Work 95

Paper IV:
Did App Privacy Improve after the GDPR? 101

1 Introduction 103
1.1 Permissions and the GDPR 105
1.2 Expecting changes in apps 107
1.3 Limitations of chosen approach 108

2 Changes in Permission Declaration and Usage Patterns 109
2.1 Permissions: Manifest Changes 111
2.2 Permissions: Changes in use 111

3 Changes in Expressed User Concerns 114

4 Discussion of Observations 116

5 Conclusion 120

Paper V:
App-generated Digital Identities Extracted through Android Permission-based Data Access—A Survey of App Privacy 123

1 Introduction and Research Question 125

2 A Model for Permission-based Partial Identity 126
2.1 Personal Identification in Digital Data 127
2.2 Partial Identities 127

3 Methodology and Data Collection 129

4 Results 130

5 Discussion and conclusion 134

Paper VI:
Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User 145

1 Introduction 147

2 Background 148
2.1 App behavior and Privacy Friendliness 149
2.2 Related Work 149

3 Pre-Study: Indicator Selection **151**
3.1 Finding an Appealing Indicator 151
3.1.1 Designing Indicators 151
3.1.2 Online Survey 152
3.2 Pre-study Results 153

4 Lab-study: User Preference **154**
4.1 Methodology: Empirical Study Procedure 154
4.1.1 Recruitment of participants and Demography 154
4.1.2 Distribution of participants 154
4.1.3 Scenarios and App-categories 156
4.1.4 Workflow of an Experimental-session in Lab 157
4.2 Empirical Results: Users' App Selection behavior 157
4.2.1 Privacy Friendliness of the Selection behavior 157
4.2.2 Ease of App Selection and Apps' Trustworthiness . . 158
4.2.3 Indicator Evaluation by Participants 160
4.2.4 MUIPC—Mobile Users' Information Privacy Concerns 161
4.2.5 Relationship between app selection and privacy concerns (MUIPC) 161
4.2.6 Comparison of Fixation Time from Gaze-behavior Data 162
4.2.7 Familiarity of the selected app 162

5 Discussion **163**
5.1 Assisting the User in App Selection 163
5.2 Influencing App Choice through Nudging 165
5.3 Comparison of Gaze-behavior 166
5.4 Indicator Evaluation by Participants 166
5.5 Impact on Time to Make a Decision 167
5.6 Further Impact on the Selection behavior 167

6 Limitations and Future work **168**

7 Conclusion **169**

Paper VII:
Accept - Maybe - Decline: Introducing Partial Consent for the Permission-based Access Control Model of Android **175**

1 Introduction **177**

2 Technical Context and Scope **179**
2.1 Need and Scope of Partial Consent 179
2.2 Technological building blocks 180

3 Regulatory Perspective **181**

4 Context of partial commitment **184**
4.1 Trust and psychological factors 184
4.2 End-user's perspective 185
4.3 Service Provider's Perspective 186

5 Proof of Concept Implementation **187**
5.1 Design 188
5.2 Implementation 188

6 Discussion **188**
6.1 Limitations of Partial Consent 189
6.2 Potential Research Directions 190
6.3 Future Work 192
6.3.1 Is the 'Maybe' button desired? 192
6.3.2 Is the 'Maybe' button accepted? 192
6.3.3 Is the 'Maybe' button usable? 193

7 Conclusion **193**

Prologue

I personally believe that an overarching perspective of a research project is owed to the audience beyond academia. This section is the popular science summary of the thesis research project for non-specialist audiences.

I could not agree more with the fact that the smartphone apps have introduced numerous and unprecedented conveniences to the user. Over time, we have become habituated with turning to mobile device to find answers, suggestions, or probable solutions to everyday needs. As I try to carry on living in this modern society with a *non-smart* phone, I have to admit that I struggle to cope up with the everyday needs that are entangled with apps, for instance, mobile bank ID for logging in securely to the bank account. What is it that we ought to sacrifice for the sake of the greater good?

At the beginning, a freemium[1] approach was adopted within the mobile app ecosystem that required payment for premium service over free one. Then the nut of behavioral surplus[2] was cracked [11], and the genie of big data began to take over all the other advertising aspects of traditional business entities [26]. Now, our precious time has become the raw material for an economy that is elusive to many [149].

Apps are designed and developed to seek attention from us, recurrently. Can we not just ignore the device? Perhaps the answer is yes, even though it is a 'no' for myself due to having weak control over facing intimidation. However, apps do not ignore the user, not even for a narrow window of time interval. They collect data about users' surroundings that contribute to propel the data-driven economy. This data collection is so extensive that it is very difficult not to be worried about invasion of my private space, if not impossible. But how can I be so sure about apps' excessive data collection?

I guess, you have already figured it out by now that I happened to be an enthusiast about apps. So, I began to look into the underlying mechanism that allows the app to access user data. In other words, I wanted to find out what happens after the user provides consent by pressing *Allow* or *Accept* button to grant access to personal data.

I often faced a hard time to explain the problem to the audience, and to some extent, even to myself. In the early stage of my research, I found myself ill-equipped with respect to the eloquent skill of communicating science. As a naive opportunist, I opted out for metaphors and other nuances to construct an explanation. I would like to explain the problem with one of them:

---

[1]Freemium - Wikipedia: `https://en.wikipedia.org/w/index.php?title=Freemium&oldid=964859444`. [Accessed: 2020-07-02]

[2]Behavioral surplus: predictive information about user behavior that is derived from machine intelligence algorithms, which use cumulative data generated from a ubiquitous environment. Hence, human voices, personalities, and emotions can be used in targeted advertising through intervening in the state of play in order to nudge, coax, tune, and herd behavior toward profitable outcomes [149].

> *Let's say, there are apps to wash hands, hypothetically of course. I open the app store and choose an appropriate one for myself. Upon installation, I run it for the very first use and it shows me an interface telling—'Permission required to access Water.' It seems like a legitimate request to me and I press the button—Allow. Consequently, I use the app to wash my hands and afterwards, I put the phone back into my pocket.*
>
> *Wait a second, did I revoke the permission to access water? Is the water still running?*

Though the user leaves the data tap—permission open to hundreds of services in the current context, the means to monitor and to observe privacy-intrusive data collection remain absent. I kept wondering about addressing the problem from the other end: would it not be convenient for the apps to just respect user privacy? Why do we, the users, need to calculate the trade-off between convenience and privacy? Would it matter if the apps access more data than needed? I think, it is a bigger problem than we can anticipate at this moment, because some anomalies can already be noticed that threaten traditional institutions with individual profiling [63]. Currently, at the time of writing this manuscript, the dilemma is even more prominent regarding the contact tracing apps to monitor the spread of a global pandemic.

I am not an economist, but certainly there is a concern for the "traditional economy" that we used to have in the good old day. Companies claim private human experience to be a source of free raw materials–that is, behavioral data, which they can and may process by advanced computational techniques to create predictions of our behavior, predictions of what we will do now, soon and later [45]. These derived predictions are then sold to other business entities, which often have regular businesses and they use the new information as a competitive advantage [17]. However, they still buy and sell tangible goods or services—things that you can hold onto, or use/experience/feel. They purchase raw materials, process or produce products as well as services, making sure of the transport and marketing of products, pay taxes and salaries to their employees, so on and so forth. At the end of the day, they can make a few bucks that we call revenues.

Here is an observation to think about: data harvesting companies do not pay for user data, it gets generated as the users carry on with their daily life in this pervasive world [45]. So, their cost for raw material is zero. They do not need to produce their product, i.e., information, it is being derived by running algorithms with a bare-bone minimum cost. They employ a significantly small and highly educated workforce that can leverage the power of a massive capital-intensive infrastructure—compared to traditional business entities with similar equity, the ratio between Facebook and General Motors is 1:40 [149]. However, these companies are earning money and the amount is just staggering—behavioral surplus had produced a stunning 3590% increase in revenue in less than four years for Google [108].

The world's richest companies have significantly low number of employees.

Does it matter? Yes, it does. These companies can and are spending their wealth behind tangible goods and services, which prompts inequalities. Especially now, during the time of a global pandemic, all those small, medium, and even some of the large companies will not be able to survive [137]. Their enterprises cannot carry on while being compelled to go through an indefinite hibernation. In contrary, the big-data companies have this luxury—their low maintenance cost is allowing them to thrive even during a global shut-down state. They will also be able to buy out the little ones—as a 'favor' of course. We will be living in a monopolized economy. The question is: Are we ready to accept such inequality within society? Or, are we doing so already without even realizing it?

This thesis makes a mere effort to introduce control over one of the *data-leaking faucets* that can contribute immensely to generate behavioral surplus. It addresses the problem about privacy implication originated from the apps' data access potential. Our research includes empirical studies, app-behavior analyses, visualizing privacy implications, and introducing methods to quantify and to communicate corresponding privacy risks to the user. We hope that this work will contribute to bringing transparency within the ecosystem of apps and thus, encouraging fairness and equality.

Introductory Summary

"All human beings have three lives: public, private, and secret."

— *Gabriel García Márquez,*
Gabriel García Márquez: a Life

1 Introduction

Arguably, a mobile phone is almost the perfect monitoring device that human beings carry along. At one end, we have a powerful device capable of knowing almost everything about a user, and at other end, this device is connected to several hundreds of different entities. So, a user's privacy protection mostly relies on the mobile device's ability to guard the data while keeping all the functionalities uninterrupted. Much of the exertion is devoted for the second purpose while privacy awaits the mercy of the service provider. Though control tools are provided within the settings interface of an operating system, the users face hindrances due to their fuzzy perception of the process and absence of adequate information about apps' privacy implications [2]. This thesis addresses the challenges to quantify, to document, and to communicate apps' privacy-friendliness. So, what do we mean by *apps' privacy-friendliness*?

A proper definition for privacy has been hard to come by, at least the one that is being addressed in this thesis. The Oxford dictionary defines privacy as "a state in which one is not observed or disturbed by other people" [89]. We could also take the "right to be let alone" as the definition of privacy by Warren and Brandeis [131], but the context of this thesis suits well with the definition provided by Westin back in 1967: *"Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others"* [134]. Here, one can correlate the context of this thesis with the significance of the individual right to control visibility of personal information, and associated difficulty to do so successfully, due to variable perceptions and preferences.

Compared to the aforementioned definitions, we would like to take an inverse route to define privacy-*friendliness*. Let us consider that the violation of privacy is the opposite of privacy-friendliness. However, the violation of privacy is rather debatable because of the fact that it is a subjective concept. Besides self-judgment, there are several other reasons behind diverse concepts of privacy violation; such as geography, culture, and law. It compels us to lean onto the legal framework and definitions found in literature. In the context of data protection, Solove described violation as an array of harmful activities and placed them into four different categories: 1) information collection, 2) information processing, 3) information dissemination, and 4) invasion [112]. Though 'invasion' is the most common form to express privacy violation, the rest of the forms are equally, if not more important for information privacy protection. Solove argues that collection, processing, and dissemination of information increase the likelihood of potential invasion. For example, the location of a secret army base was exposed in early 2018 from an unlikely source—soldiers used the fitness tracking app *Strava* which was responsible for chronologically *collecting, processing,* and *disseminating* user data that consequently resulted into *invasion*[3]. Hence, considering the other way around, if

[3]Fitness tracking app Strava gives away location of secret US army bases: `https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases`. [Accessed: 2020-07-05]

an event (originated from an app) is less likely to cause any of the four forms of privacy violation, we can adjudicate the app as privacy-friendly.

The first half of the thesis title—*Measuring Apps* is about measurement. How do we measure if an app is privacy-friendly, or not? Due to the fact that data collection is responsible behind consequential escalation of the other three forms of privacy violation, this thesis focuses on the measurement of apps' data access behavior and its corresponding privacy implications. We consider several aspects within our measurement criteria: *(i)* both static and dynamic methods (by analyzing the app-code and by monitoring the app during runtime, respectively), *(ii)* potential to contrive partial identities, *(iii)* regulatory compliance, and *(iv)* identified threats from user review analysis.

What could be done with the measurement data? The secondary part of the title was added to indicate the purpose of the measured data—*introducing transparency to apps' data access behavior*. Here, our research is focused on bringing transparent app-behavior analysis into the context of personal data privacy. During the final phase of our research, we concentrated on communicating the measurement-yielded privacy risks to the user in both ex-ante and ex-post scenarios [84]. First, our research effort includes addressing the ex-ante scenario by aiding the user with visual cues about privacy risks prior to the selection of an app. Second, we introduce a method and a prototype to enable the user to intervene in ex-post scenario with an indecisive state of consent.

This thesis stands on three complementing pillars: *(a)* measurement, *(b)* conception, and *(c)* communication of privacy risks associated with apps' data access behavior. To achieve comprehensiveness and clarity, these three concepts are not discussed in strict isolation from each other throughout this thesis.

1.1 Research Questions

Variable context, abstraction and complexity, cultural relativity, individual preference, tolerance threshold, and diverse adversaries make privacy undeniably challenging to protect. This thesis makes an effort to provide transparency about apps' data access while trying to ease the users' remorse with privacy. It aims at taking a closer look on the trend to collect more user information than needed by apps [4]. In other words, the goal is to measure privacy-friendly/invasive nature of apps through introducing tools and methods for unearthing some of the risks and consequences. This could allow the user to be aware of app's data harvesting nature which would help them to understand the potential harm and to make informed decisions. The following two research questions are addressed in this thesis:

1. *How can we identify, measure, and conceptualize the privacy risks caused by the apps' data access behavior on Android devices?*

   Mobile app users trade their data for service usage in opaque ways. Accessibility to user data faces the constraints from the access control mechanism—permissions. Nonetheless, approval from the user for sensitive information is required once (during the first use of the correspond-

ing feature), and it provides a *carte-blanche*[4]*–white-card* privilege for an app. Though the user has the option to revoke granted permissions, the absence of monitoring tools and unexpected consequences such as service exclusion, or malfunctions may cause hindrances [4, 38, 126].

In 2015, the permission model in the Android operating system was changed to improve the user experience by introducing API 23 in Marshmallow (Android 6.0)[5]. Instead of asking the user to consent for all the required privileges prior to installation (accept or leave situation), permissions are granted during run-time—once the corresponding feature is used for the first time. So, one could argue that permissions are granted at run-time, are revocable and thus, users' remorse is mitigated. Hence, we take on the challenges to hypothesize, to investigate, to conduct data collection campaigns, and to document apps' privacy-invasive behavior in the present context. Furthermore, we design and develop methods to conceptualize the privacy risks with measurable parameters. *Paper I–V* are appended in this thesis to address this research question, as depicted in Figure 1.

2. *How can we communicate the privilege-induced privacy risks to the user?*

   It has been found as a challenging task in the literature to communicate privacy-risks in an effective manner (see Section 2.2.2). This thesis addresses this question in two phases (*P4* and *P5*), as shown in Figure 1. First, we study the feasibility of an ex-ante privacy cue to aid the user in making informed decisions prior to app installation (*Paper VI*). Second, we investigate the possibility to introduce an intermediate state of decision making that could facilitate intervenability based on apps' evaluation in an ex-post scenario (*Paper VII*).

1.2 Thesis Structure

This introductory summary describes the course of our research project. In the following sections, we discuss the background relevant to the research questions, explain our choice of methods and how they relate to other research, show examples, and brief insights from our contributions. We relate the hypotheses, the research activities, and results to the appended collection of seven peer-reviewed articles in the area of Transparency Enhancing Tools (TETs), Privacy Enhancing Technologies (PETs), and Human Computer Interaction (HCI). Figure 2 presents a graphical overview of this thesis that illustrates the corresponding research activities to address the research questions *(RQ1–RQ2)*, research phases *(P1–P5)*, and included papers *(Paper I–Paper VII)*. Primarily, research phases *(P1 and P3)* were focused on literature study, hypothesis formulation and validation, and building tools and the laboratory setup. Model

[4]Carte-blanche: full discretionary power (Merriam-Webster dictionary, Retrieved: 2020-06-12)

[5]`https://developer.android.com/about/versions/marshmallow/android-6.0-changes.html`; [Accessed: 2020-06-11]

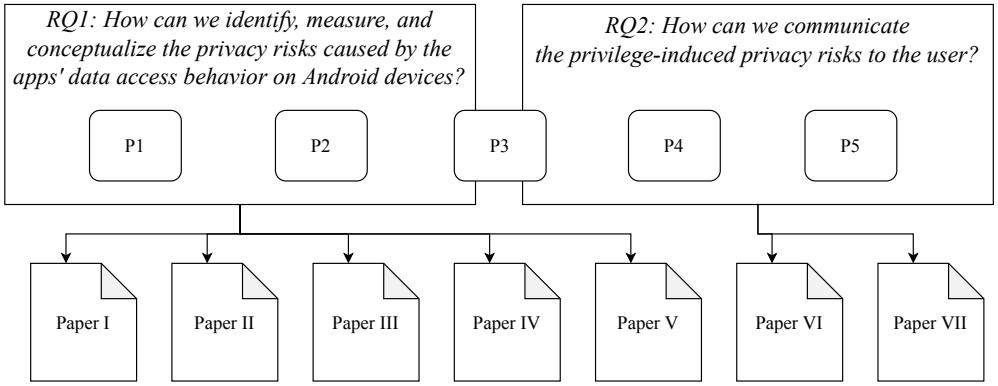

Figure 1: An overview of the thesis, showing correlation among research questions *(RQ1–RQ2)*, research phases *(P1–P5)*, and corresponding publications *(Paper I – Paper VII)*.

development, conducting controlled experiments, and running data collection campaigns were carried out during the secondary research phases *(P2, P3, P4)*. My Licentiate thesis[6] was published concurrently with *P1* and *P2*, as indicated in Figure 2. During the final research phases *(P4 and P5)*, we concentrated on designing and developing means to communicate privacy risks to the user.

The rest of this introductory summary is organized as follows: Section 2 describes the technical background and related work for the research presented in this thesis. The research methods and the limitations of our approach are elaborated in Section 3. The main contributions of this work are discussed in Section 4. Section 5 presents the brief summaries of the included papers. Finally, a discussion containing concluding remarks and future work ends this introductory summary in Section 6.

2 Background and Related Work

In this section, we discuss the technical as well as scientific background of the research area. First, a brief introduction of the technical area is described: the access control model in Android OS—a Linux based mobile OS. This is required to define the problems, to formulate the research questions and to present the arguments for our approach. Second, a discussion of the related literature is presented. Finally, the result from a literature mapping study [25, 28, 144] in the research area of *mobile app privacy* is presented to discuss recent works published during the duration of our research project.

[6]This doctoral thesis presents research that has been extended from my Licentiate thesis published in 2018 [82]. Thus, some of the research descriptions can be found as overlapping material.

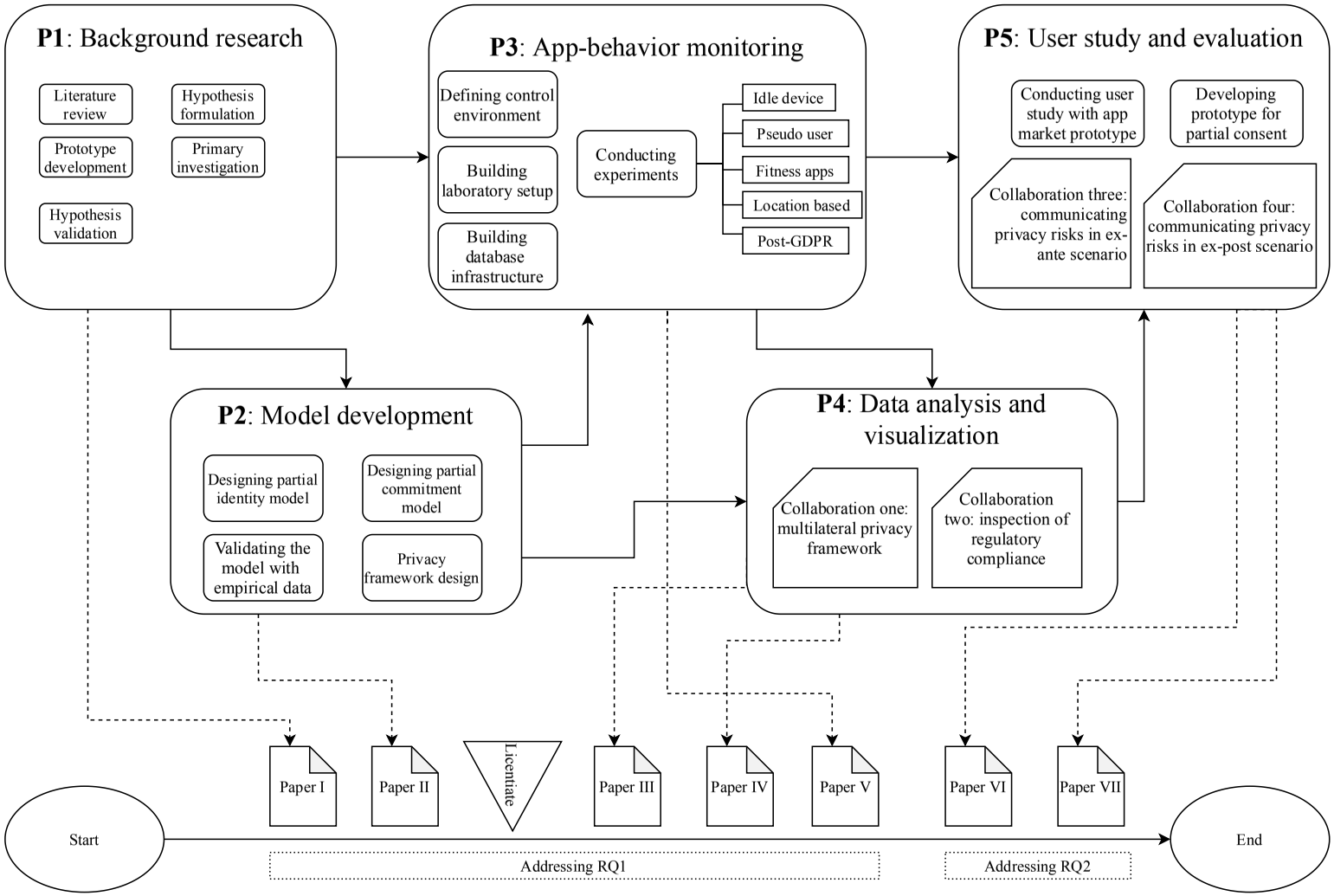

Figure 2: An illustration of the thesis project, showing its research phases (P1–P5), activities to address the research questions (RQ1–RQ2), and publications *(Paper I–Paper VII)*.

2.1 Technical Background

There are several Unix based operating systems available for mobile devices, for example: iOS, Tizen, LineageOS, Android, and Firefox OS. Due to larger user base, project feasibility, and open-source nature of the platform, our research is concentrated on the Android OS. Moreover, it is the most prevalent mobile OS and hence, results would have a large impact. Here, we provide a short description of Android and common terminologies related to its access control model.

2.1.1 The Android Operating System

The Android Open Source Project (AOSP)[7] was launched by Android Inc., which was later acquired by Google in 2005. It is a software stack which supports a wide range of mobile devices. Android runs on a customized Linux kernel which is responsible for running the basic drivers and components (display, audio, binder, etc.). Collectively, these perform the role of a foundation for the Dalvik virtual machine, providing a run-time and native libraries. A tertiary layer, named the application framework, is responsible for accommodating the apps. Each of the installed apps remains virtually isolated in their own sandboxes and they get access to system resources (memory, sensors, etc.) through permission-based access control mechanism.

2.1.2 Permission Types, Manifest and API

In Android, apps can request access to the device's resources through permissions. Depending on the resource types, consent from users is required. Android defines three types of permissions[8]: *normal*, *dangerous*, and *signature*. *Normal* level permissions allow access to resources that are considered low-risk, and they are granted during the installation of any package requesting them, e.g., `ACCESS_WIFI_STATE` allows apps to access information about Wi-Fi networks. The *dangerous* level permissions grant access to resources that are considered to be high-risk, e.g., `ACCESS_FINE_LOCATION` allows an app to access precise location. In this case, the user must provide explicit consent to grant permission. So-called *signature* level permissions are granted at install time, but only when the app that attempts to use a permission is signed by the same certificate as the app that defines the permission, e.g., `REQUEST_INSTALL_PACKAGES` allows an application to request installing packages.

Every app has an `AndroidManifest.xml` file within the app-code that contains information about that particular app (e.g., its name, author, icon, and description) and required permissions that grant access to data such as call logs, contact lists, or location on smartphones. It helps the reviewers as well as users to be informed and to be aware of data access potential of that particular

[7] `https://source.android.com`. [Accessed: 2020-06-12]

[8] `https://developer.android.com/guide/topics/permissions/overview`. [Accessed: 2020-05-28]

app. The following pseudo-code shows the definition of a permission request in a sample manifest file:

```
<!--access request to location by an app-->
<manifest xmlns:android="http://example.website.com/apps/
  android"package="com.website.example">
<uses-permission android:name="android.permission.
  ACCESS_FINE_LOCATION"/>
<application ...>
...
</application>
</manifest>
```

An *Application Programming Interface (API)* can be defined as a collection of features and rules which facilitates the communication between the software modules. In the context of this thesis, it can be defined as the responsible entity for guarding the access and interaction with the device hardware, for example, storage/memory.

2.1.3 Problem Definition: Technical Context

The permission-based access control architecture of Android is placed to guard user data and sensors. Approval from the user for granting a dangerous permission is required during the first use of the app. As the users possess limited knowledge about the magnitude of apps' data access potential, it is difficult for them to perceive the consequence of granting access, and assess the risk [2]. Moreover, information is hardly available about usage of permissions that are allowed access to resources. So, a *carte-blanche or white-card* privilege to the available system resource is given to the app that leaves access decisions about sensitive personal data to arbitrary programs and services. This is a problem for the user (data subject), who suffers from the lack of appropriate monitoring tools in order to support the re-assessment of decisions made earlier [3]. We argue that permission usage data has the potential to reveal apps' behavior and thus, can assist the user in making informed decisions.

2.2 Related Work

Android's access control model has been a popular subject to study because of its vast user base, good coverage over a variety of devices and open source platforms. This section provides the highlights of some key-related works.

Prior to my Licentiate thesis [82], we explored the significant research effort behind malware classifications conducted through static [7, 36, 41, 76, 141, 148], dynamic [6, 10, 13, 16, 33, 66, 109, 115, 140, 147], and hybrid analyses [46, 69, 123, 133]. In general, these works can be viewed as systematic and technically burdensome approaches, because they require a considerable amount of Android middle-ware and kernel modifications. Hence, these approaches face adaptability and usability challenges for both developers and users. However, understanding these approaches to segregate benign and malicious apps was crucial to us for deciding upon a dynamic and user-centric approach. Though

our work can be placed within dynamic analysis category for monitoring apps, it does not address the problem to separate benign apps from the malicious ones. It hypothetically sees user data as currency and tries to evaluate apps based on their consumption or cost for delivering services. Hence, we consider only the benign[9] apps that are available in the app store for the app behavior analysis to evaluate their privacy-friendliness.

2.2.1 App-behavior Analysis

Questions concerning the frequency of resource access by mobile apps are also asked by several researchers in independent studies [4, 38, 126]. Almuhimedi et al. show that data on the frequency of access can encourage users towards privacy-preserving behavior [4]. They introduced a method to warn the user in the form of nudging about the potential implication in order to encourage privacy-preserving behavior. Franzen and Aspinal introduced a policy-based resource usage mechanism for JavaScript apps developed within the PhoneGap framework [38]. The frequency of resource access question is also addressed by Kleek et al. who introduced a prototype with data controller indicators against apps' data sharing practices, in order to help the users to make informed decisions [126]. Hatamian et al. introduced FAIR with a similar objective to ours [53]. They proposed a method based on fuzzy logic to determine a risk score for apps. FAIR can serve a similar purpose which is to provide a post-installation opportunity to judge installed apps based on their resource access efforts.

SecuRank was presented by Taylor et al., which identified 3400 potentially over-privileged apps [119]. It is able to warn the user in advance regarding permission-hungry apps and offer an alternative option which can serve similar needs. Recently (in 2020), Cai and Ryder published a longitudinal study on app's evolutionary structure [18]. However, both of these approaches are concerned with static structure and changes of the app, and they do not cover the resource access efforts made by the apps during run-time. Hence, apps' real-time permission access efforts and their corresponding privacy implications remain unexplored in these research efforts.

Another research direction presents a way to compel the apps to follow a policy. Jeon et al. presented a combination of tools (RefineDroid, Dr. Android and Mr. Hide) to enforce fine-grained permissions [60]. Wang et al. introduced DeepDroid to assist enterprises to implement customized policies [130]. Hammad et al. presented the extensive violation of the principle of least privilege (PoLP) within Android ecosystem [48]. They developed DELDROID and their work introduced a method to detect and enforce the least privilege principle during run-time.

Aligning with these related works, we focused on a long term app behavior analysis using both static and dynamic approaches. Moreover, we took other perspectives (e.g., user review and privacy policy) into consideration and

[9] We cannot guarantee that they are all benign, but they are very unlikely to be malware.

correlated them with the app-behavior analyses. Our app-behavior analyses are presented in *Paper I–Paper V*.

2.2.2 Responsibility of Awareness

Unawareness of app developers was held responsible by some studies for designing and developing of permission hungry apps [79, 91, 132]. Marky et al. introduced design principles of privacy friendly app development [79]. Peng et al. emphasized on the importance of developers' awareness during app development, and recommended the notion of risk score development to diminish exposure of sensitive information [91]. Though today's apps are coming with increasingly rich functionalities, privacy-aware app-development practice is absent. Two studies have shown the insights for the evolution of permission usage, and their investigations indicate the growth of over-privileged apps that ask for more dangerous permissions [121, 132]. In the recent past, a reduction of dangerous permission request was observed in our research, which was influenced by the regulatory shift in Europe [83].

Users' unawareness is well documented in many studies [9, 62, 64, 80, 99]. Reluctance and lack of awareness about granting access to personal information are also found in the user study and user-review analysis conducted by us, which is presented in *Paper VI* and in *Paper IV*, respectively. Paranoid behavior was also observed among the over-concerned users [110]. Diverse preferences make the problem difficult to solve and it was identified earlier by Lin et al. [72] and by Mylonas et al. [85]. Thus, crowd-sourcing user preference and expectations can be considered as a prerequisite and essential parameter to design solutions for raising awareness about apps' privacy-friendliness.

2.3 Literature Mapping Study

This section is intended to offer an overview of recently published related works. Here, we present results from a literature mapping study [25, 28, 144] in the research area of *mobile app privacy*. We also present categorizations of search results illustrating research topics' development, with liberal citations from the literature and a manageable bibliography.

2.3.1 Method of Mapping Study

In the following, we describe the mapping study process that includes the choice of databases, construction of search queries, inclusion-exclusion criteria, and contribution-wise categorization. Figure 3 presents an the overview of systematic literature mapping methodology. Table 3, 4, and 5 present the list of resulted articles along with their contributions' targeted sub-domain.

Database Selection: Due to our overarching goal for the PhD project, we narrowed down the scope of the literature search within the technological focus of the research topic, which also intersects with the regulatory aspects and with the human factors. Thus we decided to limit the scope of literature

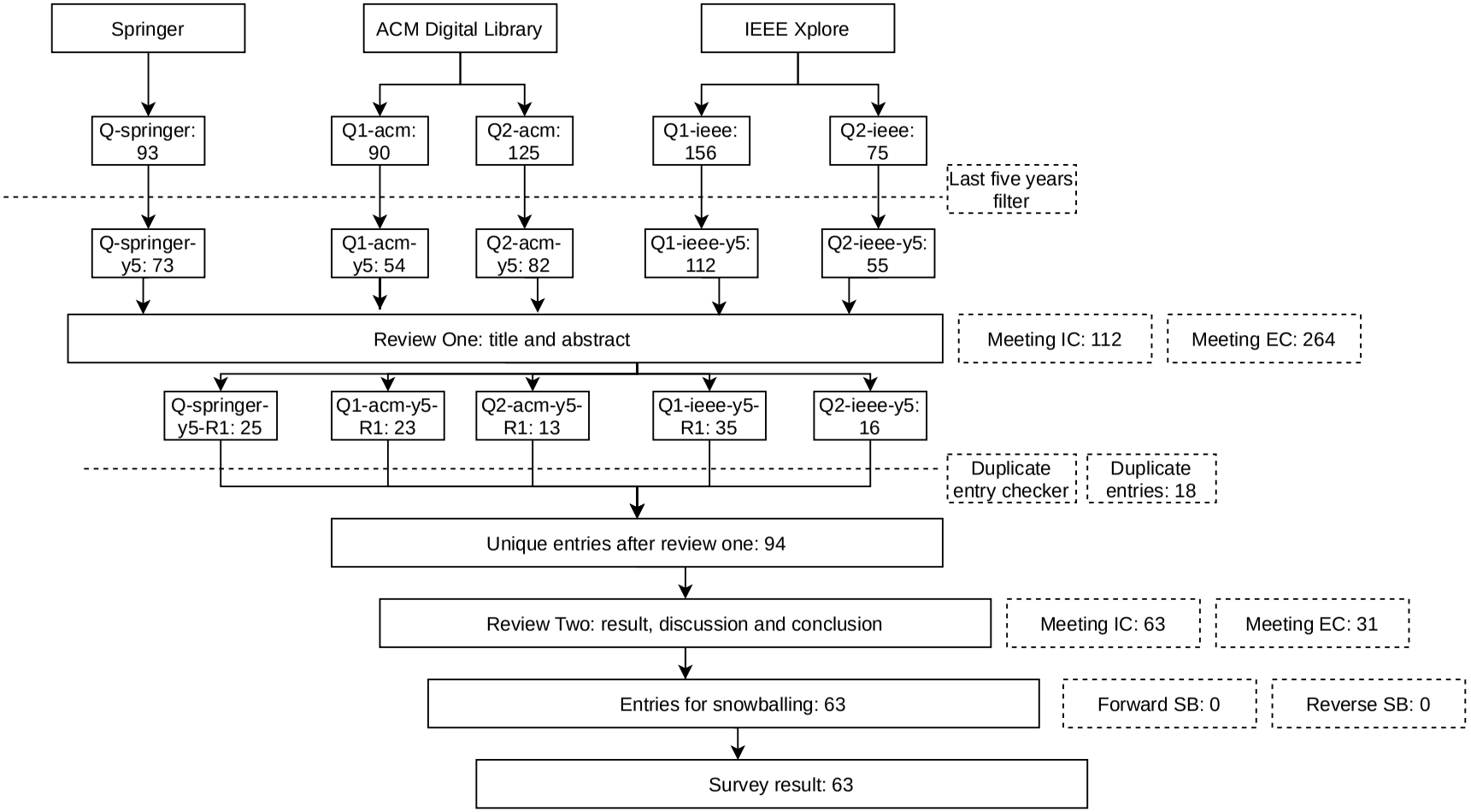

Figure 3: Overall procedure of literature mapping study.

Table 1: Inclusion Criteria (IC) of the literature mapping study.

ICs	Description
IC_1	Technical aspect could include risk identification
IC_2	Technical aspect could include over-privilege issues of Android apps
IC_3	Technical aspect could include solution proposal to address a certain vulnerability
IC_4	Regulatory aspect could reflect upon the corresponding privacy policy document and/or requirement posed by the authority on apps' obligation to address individual's privacy protection
IC_5	User related issues could include analysis of user reviews
IC_6	User related issues could include studies documenting user perceptions about apps' privacy implications
IC_7	User related issues could include usability aspects of privacy concerns originating from apps

search within three technology-focused databases: *(a)* Springer Link, *(b)* IEEE Xplore, and *(c)* ACM Digital Library.

Inclusion Criteria (IC): Records need to be related to the privacy implications of Android's permission-based access control model. Furthermore, they should address the API 23, or higher in order to take the run-time-permission-structure into consideration. In the first review phase (title and abstract inspection), 'privacy implications' could include a form of violation from the four levels as defined by the Solove's taxonomy of privacy [27]. During the second review phase (inspection of result, discussion and conclusion sections), a record needs to meet at least one of the inclusion criteria from Table 1.

Exclusion Criteria (EC): The records should not address issues other than the privacy implications caused by the permission-based access control model in Android. Records should not address the permission-model in API 22 or lower. In the first review phase (title and abstract inspection), records could be excluded due to lack of significant contributions. To be more precise, short paper and poster need not be considered. In the second review phase (inspection of result, discussion and conclusion sections), record could be excluded if it addresses issues beyond privacy risks and/or security issues that do not target benign apps. To be more precise, papers focusing on malware detection and analysis need not be considered.

2.3.2 Survey Procedure

Table 2 demonstrates the gradual progress of the survey procedure. However, some queries suffer from limited exposure to the meta data due to various reasons as described below.

Springer Link: In order to improve the outcome of search query and to achieve a manageable sample size, search query is applied with a limitation—it should include records from two sub-categories provided by Springer Link:

Table 2: Querying processes within the databases.

Search criteria	Springer Link	IEEE Xplore	ACM Digital Library
Date of query execution	22 May 2020	04 May 2020	04 May 2020
$Q_{springer}$	((("All Metadata":android) AND "All Metadata":app) AND "All Metadata": permission) AND "All Metadata":privacy) FROM (Computer Science (Information Systems Applications) AND (Systems and Data Security))	-	-
Q_1	-	((("All Metadata":android) AND "All Metadata":app) AND "All Metadata":permission) AND "All Metadata":privacy)	Abstract:(android) AND Abstract:(permission) AND Abstract:(privacy) AND Abstract:(app) "filter": ACM Content: DL
Q_2	-	(("Abstract":app) AND "Abstract":behavior) AND "Abstract":privacy)	Abstract:(app) AND Abstract:(behavior) AND Abstract:(privacy) "filter": ACM Content: DL
Papers found	$Q_{springer}$ = 93	$Q1_{ieee}$ = 156, $Q2_{ieee}$ = 75	$Q1_{acm}$ = 90, $Q2_{acm}$ = 125
Applying IC=5 years filter	$Q_{springer-y5}$ = 73	$Q1_{ieee-y5}$ = 112, $Q2_{ieee-y5}$ = 55	$Q1_{acm-y5}$ = 54, $Q2_{acm-y5}$ = 82
Applying IC and EC for Review one	$Q_{springer-y5-R1}$ = 25	$Q1_{ieee-y5-R1}$ = 35, $Q2_{ieee-y5-R1}$ = 16	$Q1_{acm-y5-R1}$ = 23, $Q2_{acm-y5-R1}$ = 13

Table 3: Papers found in Springer Link.

Record	Year	Technical issues			Regulatory issues	User related			Comments
		Risk identification	Over-privilege	Solution		Review analysis	Perception	Usability	
Wettlaufer & Simo [135]	2020			×	×				Assessing apps' privacy impact
Liu & Simpson [75]	2020						×	×	User study on privacy preference
Hatamian et al. [52]	2019		×	×	×	×			User study on privacy preference
Alvarez et al. [5]	2019		×	×					Restriction on runtime permissions
Jayakumar et al. [59]	2019						×		User survey on privacy concerns
Hatamian et al. [51]	2018			×			×		User survey on privacy concerns
Hossen & Mannan [55]	2018	×	×						App behavior analysis
Reinfelder et al. [98]	2018						×	×	User survey on runtime permissions
Lu et al. [77]	2018	×	×						Library and app analysis
Shu et al. [111]	2018	×		×					Execution footprint removal
Li et al. [70]	2018	×				×			Forecasting malicious behavior
Hatamian et al. [53]	2017		×	×				×	App behavior analysis

Computer Science (Information Systems Applications), and Systems and Data Security.

IEEE Xplore: The search queries were run on the total meta data and they yielded a manageable data set. Hence, we did not require any special restriction about limiting the exposure of the queries.

ACM Digital Library: In order to improve the outcome of search query and to achieve a manageable sample size, search query is applied to the *Titles* and *Abstracts* only.

2.3.3 Findings from the Literature Mapping Study

From this survey, we found 63 papers that were published during the last five years and addressed the research topic from various perspectives, as illustrated in Figure 3. IEEE Xplore is the highest contributor (33 papers) to the survey

Table 4: Papers found in IEEE Xplore.

Record	Year	Technical issues			Regulatory issues	User related			Comments
		Risk identification	Over-privilege	Solution		Review analysis	Perception	Usability	
Feng et al. [37]	2019		×		×				consistency checker
Yu et al. [142]	2016				×				privacy policy checker
Yu et al. [143]	2018				×				Mapping permissions to privacy policy
Nguyen et al. [87]	2019	×				×			User review analysis
Wang et al. [129]	2019					×	×		User review analysis
Scoccia et al. [107]	2019						×		User review analysis
Peruma et al. [92]	2018						×		User study
Rashidi et al. [95]	2018			×			×		Permission recommendation
Liu et al. [73]	2018			×				×	Permission recommendation
Fratantonio et al. [39]	2017	×							Attack vector
Narain et al. [86]	2017	×							Privilege escalation
Eling et al. [32]	2016						×		User-awareness study
Gao et al. [42]	2020			×					NLP-based permission recommendation
Scoccia et al. [106]	2019	×	×						Analysis of open source apps
Scoccia et al. [104]	2019			×					Framework to offer granular permissions
Scoccia et al. [105]	2019	×		×					Documentation on permission mistakes
Momen et al. [83]	2019		×		×	×			App behavior analysis
Hsu et al. [56]	2019	×		×					Dynamic control to block third party libraries
Gasparis et al. [43]	2019			×					Granular permission control
Onik et al. [50]	2018	×							Risk identification
Calciati et al. [19]	2018	×		×					Framework to identify changes in app-releases
Liu et al. [74]	2018	×						×	Discrepancy in permission rationale
Sadeghi et al. [100]	2018			×					Temporary permission granting
Fu et al. [40]	2017	×		×					Permission control over third party libraries
Tang et al. [116]	2017		×						Detection of over-privileged apps
Chester et al. [21]	2017		×						Identifying permission gaps
Olejnik et al. [88]	2017			×					Option to obfuscate in permission granting
Zhang et al. [146]	2019	×							Attack surface: screen usage
Zhang et al. [145]	2017	×							Contextual ICC checker
Li et al. [68]	2018	×							Contextual protection for images
Kang et al. [61]	2020			×					Location privacy through decoy deployment
Rashidi & Fung [94]	2020	×							Generating risk notifications
Huang et al. [57]	2018	×						×	Privacy dark pattern

Table 5: Papers found in ACM Digital Library.

Record	Year	Technical issues			Regulatory issues	User related			Comments
		Risk identification	Over-privilege	Solution		Review analysis	Perception	Usability	
Baalous & Poet [8]	2018		×		×				privacy policy checker
Gerber et al. [44]	2018						×		Positive reinforcement to privacy-awareness
Jackson et al. [58]	2018						×		Positive reinforcement to privacy-awareness
Wijesekera et al. [136]	2018			×			×		User study on privacy preference
Feichtner et al. [35]	2020					×	×		Mining awareness using deep learning
Raval et al. [96]	2019			×					permission restriction plugins
Wang et al. [127]	2019	×							evolution of Android ecosystem
Diamantaris et al. [30]	2019	×	×						granular run-time permission
Tang et al. [117]	2018	×	×						App behavior monitoring
Srivastava et al. [113]	2017	×							camera permission: visual privacy leaks
Chitkara et al. [22]	2017			×					Context aware location privacy
Taylor et al. [118]	2017			×					Finding privacy-friendly option
Wang et al. [128]	2017	×							Purpose specification of permissions
Taylor et al. [122]	2017	×							Evaluation of apps' update
Taylor et al. [120]	2016	×	×						Safeguarding contextual permissions
Xi et al. [139]	2019							×	Purpose of GUI components
Sadeghi et al. [101]	2019							×	GUI testing tool
Tromer & Schuster [125]	2016	×							Intra-app information flow

Table 6: Cumulative comparison of findings.

Databases	Technical			Regulatory	User aspects		
	Risk identification	Over privilege	Solution		Reviews	Perception	Usability
Springer Link	4	5	6	2	2	4	3
IEEE Xplore	15	5	12	4	3	5	3
ACM Digital Library	8	4	4	1	1	4	2

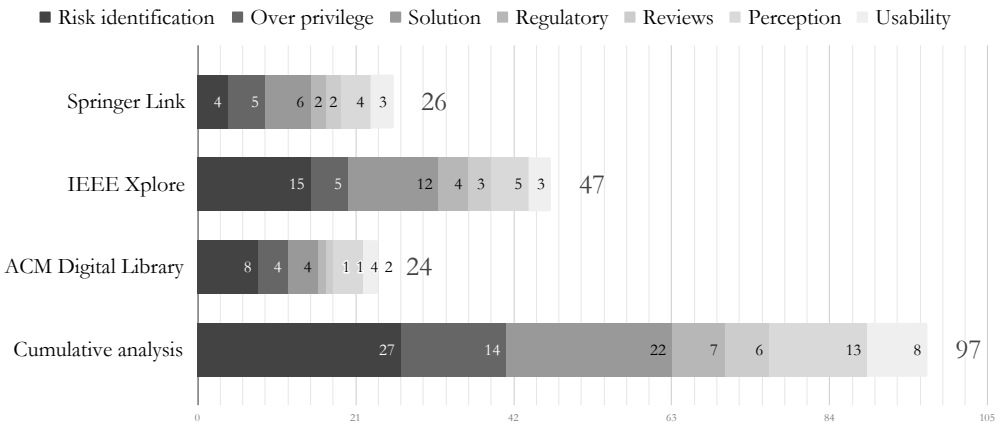

Figure 4: Categorical distribution of papers found in literature mapping study.

result. Our search in ACM digital library and Springer Link yielded 18 and 12 papers, respectively. We also categorized the resulted papers according to their research drive and contributions. Figure 4 shows the overall contribution-wise analysis of papers found from this literature mapping study.

Due to having priority on the technology-focused publishing databases and search terminologies, the majority of the papers found from the literature mapping study are concentrated on the technical aspects; hence contributing to identifying risks, over-privilege issues, and introducing solution proposals to address corresponding privacy vulnerabilities. Due to having a very narrow time-constraint (last five years), neither backward, nor forward snowballing [138] yielded any additional unique publication.

Furthermore, a small volume of research works focusing on regulatory aspects of privacy can be observed in Figure 4 and in Table 6. We acknowledge the limitations that are caused due to finding fewer number of publications focusing on the user-related issues (27 out of total 539 publications) and regulatory aspects (7 out of total 539 publications) of privacy implications.

The search result contained two papers that are included in this thesis: *Paper III* and *Paper IV*. It should be noted that *Paper VI* and *Paper VII* were not available in the database at that time, *Paper I* did not meet the IC for being a short paper, and *Paper II* and *Paper V* belong to a different database—Gesellschaft für Informatik. In comparison with the resulted papers, we observe that our research addresses the topic from various perspectives; namely technical, regulatory, and human factors. Hence, this thesis tries to bring these concepts under the same umbrella and to address the research questions with a holistic overview of the subject matter.

3 Research Methods

This thesis belongs to the realm of Computer Science (CS) which is a vast but comparatively new domain of science [29]. While some authors challenge

the scientific rigor of this discipline, its applications have had an undeniably enormous impact on how the other branches of science are being practiced [29]. The discussion becomes more interesting as CS gets entangled with social science. This area of research is known as Human Computer Interactions (HCI), which is related to studies and experiments concerning how computer and its applications are being interacted with, and affecting humans as well as their surroundings which include both positive and negative aspects [67]. As a result, it entails other research areas—the science of security [54] and privacy engineering [49]. These research fields emerged from the necessity to tackle threats from diverse adversaries, which are of course created by other people. Here, a scientific pursuit for empiricism within this paradigm is rather obvious because it deals with the most vulnerable and the most unpredictable node of the system—human.

Our work remains within the intersection point of these disciplines: security, privacy and human-computer interaction. Hence, this thesis addresses three additional aspects (transparency, intervenability, and unlinkability) compared to the science of security (confidentiality, integrity, and availability) [49]. Aligning with the related literature, the struggle to address a topic associated with philosophically variable privacy-problem is evident in this thesis. It is also challenging to address a subject that is evolving along with the advancement of technology in an exponential pace.

3.1 Method Iteration

Figure 2 presents an overview of the structure and research procedure, which includes five research phases *(P1–P5)*. As these research phases are overlapping on several occasions, their objectives to address research questions got entangled with each other, for instance, *P3* contributes in addressing both *RQ1* and *RQ2*. However, we made an effort to isolate our research approaches for both research questions.

Hence, we leaned onto the Scientific Method [31] and Design Science Research Methodology (DSRM) for Information Systems [90]. Nonetheless, some merged, selected subset and intermediate steps were defined and followed throughout the whole process. Figure 5 shows the iterative research approach to address *RQ1* and *RQ2*. It considers the research phases *(P1–P5)* as nominal processes for method iteration. Upon completion of the iterations, it communicates the contributions *(C1–C5)*. A detail illustration of all contributions is shown in Figure 6.

The iterative process has seven steps: *(a)* motivation, identifying problem, research question formulation, *(b)* literature review, hypothesis formulation, *(c)* model design, subordinate research method selection, *(d)* implementation, prototyping, *(e)* experiment, survey, data collection, *(f)* data processing, data analysis, and *(g)* evaluation, interpretation. Problem definition, specifying research questions, and falsifiable hypothesis formulation had often constructed from brain-storming sessions within the research group.

Once the hypothesis, problem and research questions are formulated, subor-

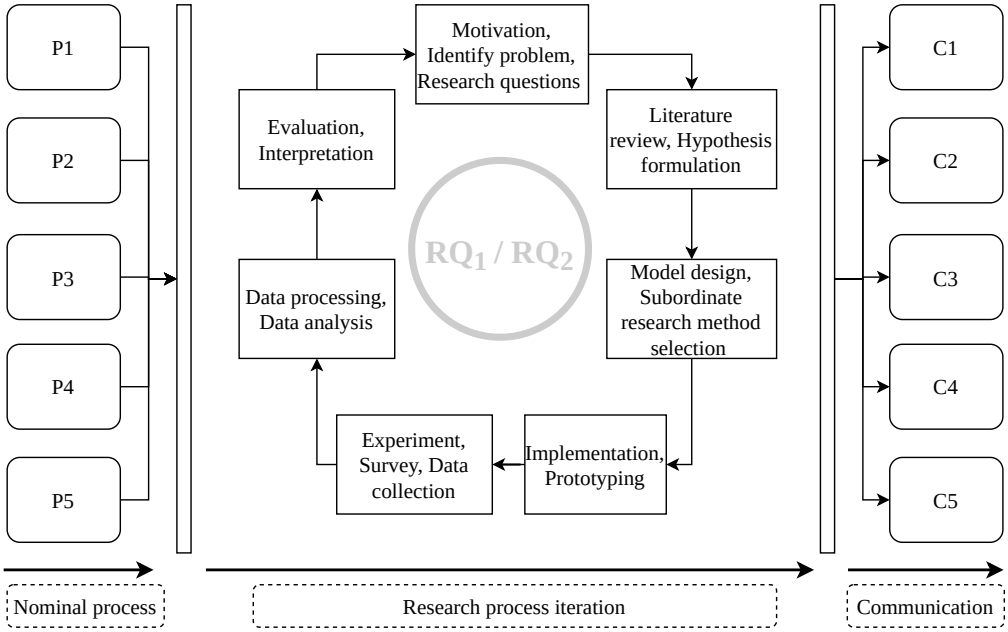

Figure 5: Iterative process of research method that considers research phases *(P1–P5)* as nominal process and communicates contributions *(C1–C5)*.

dinate research method is used to carry out the next phases. DSRM defines four types of research entry points (problem-centered, objective-centered, design and development-centered, and client/context-centered), which put emphasis on the research activity for that particular iteration of research process. Thus, the iteration process can accommodate the research phase according to its feasibility; for instance, design and development-centered approach yielded model-based solutions to the research problems.

3.2 Methods Used in the Papers

Based on the design decision, the implementation phase addresses the necessity of tools in order to facilitate later research activities. Hence in *Paper I, II, VI, VII*, we followed the prototyping approach, which is widely used in the software development process [12, 14]. We also adopted the prototyping approach as an integral part of the experimentation phases. Prototypes were developed as modular tools to provide necessary support for running the experiments. It allowed us to evaluate the performance of an ongoing iterative process and remove, or change modules if necessary. This is an intermediate step for the experiment stage. Outcome of an ongoing experiment can be taken into consideration for primary observation and evaluation. It helped to tune the prototypes and experimental setup if required.

In *Paper VI*, our research also used the survey methodology, on which the social science research relies to a great extent [47]. In this case, a laboratory-bound user study was conducted for collecting and analyzing both quantitative and qualitative data to support and contradict hypotheses. Post-survey and post-experiment phases concentrated on data analyses, interpretation, and

evaluation. We also performed statistical analyses to constitute interpretation in *Paper VI*; namely Kolmogorov-Smirnov test [71], Friedman test [24], Kruskal Wallis test [65], and Mann-Whitney U test [81].

3.3 Limitations of Our Approach

Several limitations arise when investigating privacy aspects of smartphone apps. First, our research is limited to the Android platform only. Hence, we abstain from speculating about apps from other platforms, e.g., Apple's iOS. However, the result from our privacy policy analysis could be considered relevant for the app providers to some extent.

Apart from the user studies, we cannot claim that the presented results are 'reproducible' with respect to any given context, which is one of the fundamental requirements in any branch of science. This is an unavoidable limitation due to the ever-evolving nature of apps as well as of the platform itself. Apps get regularly updated along with their privacy policies, leaving static data snapshots outdated. Thus, it is a very difficult property to achieve because of the challenges associated with retrieving the older versions of apps, privacy policies, and various forks of the Android operating system to create a similar test bed as well as identical data collection campaigns. However, we have archived the corresponding data-sets that were used to produce results and thus, the analyses could be run again. Repositories for the tools and archived data are listed in Appendix A and in Appendix B, respectively.

4 Contributions

In this thesis, our journey includes designing models, developing tools, experimenting with apps in various contexts and longevity to identify privacy risks, taking regulatory aspects into account to conceptualize privacy risks, conducting user study to document empirical privacy preferences, and building prototypes to accommodate conditional consent into an access control mechanism. The contributions of this thesis can be segmented into five areas *(C1–C5)*, as illustrated in Figure 6, and they are discussed below:

- ***C1**—identified privacy risks from the Principle of Least Privilege (PoLP) violation by the apps' data access through permissions.*

  As shown in *Paper I*, granted privileges are accessed in a higher frequency than perceived from interface. By varying the user interaction during the data collection phases, we showed that apps were found to be exercising the privileges that do not meet the requirement and expectation of PoLP [102]. Aligning with the related works, this thesis highlights the risks associated with resource access frequency.

- ***C2**—models to conceptualize the privacy risks associated with Android apps' data access patterns.*

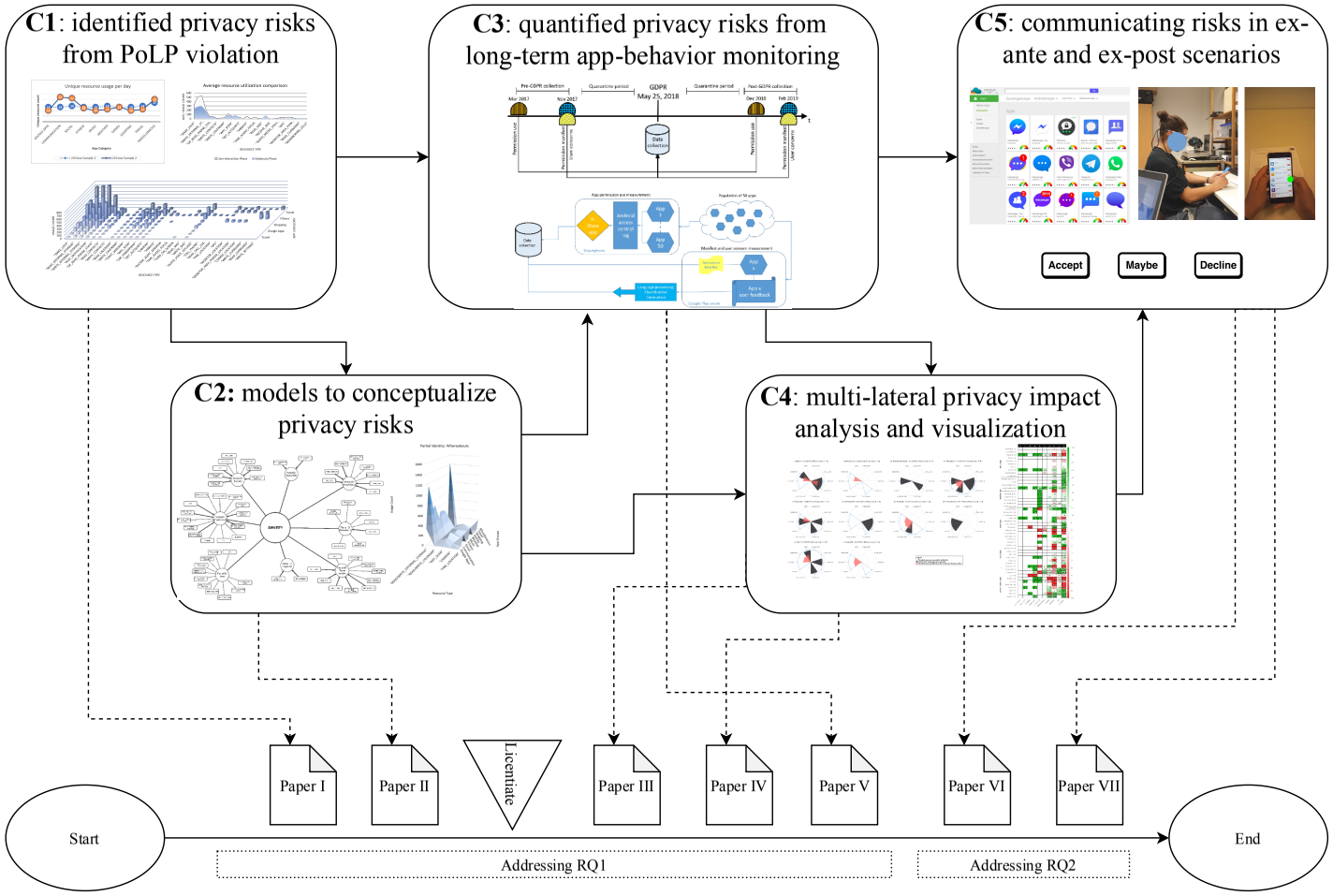

Figure 6: An overview of contributions: *C1*—identifying privacy risks (see *Paper I* for details), *C2*—introducing models to conceptualize privacy implications (see *Paper II & V* for details), *C3*—documenting, measuring and comparing app behavior over a long period of time (see *Paper IV & V* for details), *C4*—introducing methods to conceptualize and to quantify privacy implications from various sources (see *Paper III & IV* for details), and *C5*—communicating privacy risks to the users in ex-ante and in ex-post scenarios (see *Paper VI & VII* for details).

Based on the concept of partial identities [23, 93], we analyze the information gained through app permissions for the partial identity sets that can get retrieved through the permissions. We describe a model for building partial identities from information accessible through permissions on Android devices and an empirical study is presented indicating the likelihood of partial identity extraction. *Paper II* and *Paper V* make an effort to offer a better understanding of underlying risks by correlating them with identity attributes.

- ***C3**—documented apps' permission usage behavior over a long period of time, which offers understandings and insights about privacy implications.*

  We conducted long-term data collection campaigns to record empirical data which supported the assumptions made in hypotheses. We conceptualize the privacy risks by correlating them with the regulatory framework in *Paper III, IV*, and *Paper V*. We also show that the apps' data access patterns can be used in identification, and quantification of regulatory impact assessment for the apps. From our empirical data collection campaigns and analyses, both *Paper IV* and *Paper V* show that General Data Protection Regulation (GDPR) influenced the apps to curb their privacy invasive behavior.

- ***C4**—methods to conceptualize, to quantify, and to visualize privacy risks through multi-sourced data analysis.*

  Besides analyzing apps' corresponding manifest in the code (static analysis) and documenting its run-time access patterns (dynamic analysis), we investigated their privacy policies and user review available in the app market. First, we pay attention to the privacy policy analysis of apps and fulfillment of fundamental legal principles identified in *Paper IV* [83], the extent to which the privacy policy texts of apps are correlated with what developers request (in manifest) and what they do in reality (actual permission usage). Second, crowd-sourced user reviews for apps are an additional reference point for identifying privacy threats. It allows us to take the individual's privacy attitudes into account and map the identified threats to the corresponding cases. We extracted app market user feedback for the app set. We applied this additional information to judge an apps' privacy implications. A cumulative ranking of fitness apps is presented in *Paper III* [52].

- ***C5**—methods for communicating privacy risks to the user in both ex-post and ex-ante scenarios.*

  This thesis addresses the challenge to communicate privacy risks to the user in both ex-ante and ex-post scenarios. First, it introduces a potential solution to the privacy problem concerning users' dilemma about app selection from millions. We also evaluate the feasibility of our proposed solution through a user study. As documented in *Paper VI*, results from this user study show that the mere presence of a privacy indicator

in the interface can trigger statistically significant privacy-preserving app-selection behavior. Second, we address the lack of intervenability within the access control mechanism of Android through introducing a plausible solution. Our proposed method could offer an interim stage for the user to examine apps' potential privacy implications. Thus, it could communicate privacy risk in the ex-post scenario by adding intervenability.

5 Summary of Included Papers

In this section, we present brief summaries of the included papers.

Paper I: How much Privilege does an App Need? Investigating Resource Usage of Android Apps

This paper addresses the first part of *RQ1*—identify privacy risks. In this work, we hypothesized that apps exercise their granted privileges in a manner that does not meet the requirement to ensure PoLP [103]. An experiment was designed to measure apps' resource access efforts and it was carried out in two phases. In principle, access to such privacy sensitive data should be kept to a minimum. In contrary, apps were found to be utilizing their granted privileges frequently, even without interaction from the user. Hence, we validate the hypothesis through inspecting the privilege utilization patterns of apps installed on Android devices.

Paper II: Derived Partial Identities Generated from App Permissions

This paper makes an effort to conceptualize privacy risks (*RQ1*), and presents a model of partial identities derived from app permissions that is based on Pfitzmann and Hansen's terminology for privacy [93]. The article first shows how app permissions can contribute to the accumulation of identity attributes for partial digital identities by building a model for identity attribute retrieval through permissions. Then, it presents an experimental survey of partial identity access for selected app groups. By applying the identity attribute retrieval model on the permission access log from the experiment, we show how apps' permission usage is providing to identity profiling.

Paper III: A Multilateral Privacy Impact Analysis Method for Android Apps

This paper addresses all the three aspects of *RQ1*—identify, measure, and conceptualize privacy risks. Here, we focused on comparing data from different sources and from different locations. So, instead of a single-sourced app behavior analysis, a multi-perspective analysis was performed. First, apps'

static (app-code) and dynamic (run-time permission access patterns) behavior analyses were done at Karlstad University. Second, apps' privacy policy and corresponding user-review analyses were performed at the Goethe University Frankfurt. Finally, a cumulative score was generated to indicate each apps' privacy-friendliness. We applied this method on the top ten fitness apps and ranked them accordingly in order to demonstrate feasibility.

Paper IV: Did App Privacy Improve after the GDPR?

This paper addresses the measurement and conceptualizing issues of *RQ1*, and presents an analysis of app behavior before and after the regulatory change in data protection in Europe. Based on long-term data collection, we present differences in app permission use and expressed user concerns, and discuss their implications. One should expect to find changes in code, program behavior and data collection activities. To investigate this expectation, we analyzed data about Android apps' request and use of permissions to access sensitive group of data on smartphones, and collected user reviews. Our data shows an overall reduction of both permissions used and of expressed user concern. However, in some areas apps have increased access or user complaints while in addition, many apps carry with them several unused access privileges.

Paper V: App-generated Digital Identities Extracted through Android Permission-based Data Access—A Survey of App Privacy

This paper also addresses the measurement and conceptualizing issues of *RQ1*, and presents a renewed version of the model for partial identities derived from app permissions (first introduced in *Paper II*) that is based on Pfitzmann and Hansen's terminology for privacy [93]. This model also refers to identifiability through access to personal data protected by the Android access control mechanism called *permissions*. We populate partial identities with attributes related to permission-protected personal data, and then show how apps accumulate such attributes in a longitudinal study that was carried out over several months. We isolated the data from months before and months after the regulatory (GDPR) change came into effect. Our data visualization shows changes in apps' interest towards identity attributes.

Paper VI: Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User

As described in Section 1.1 and shown in Figure 1, *RQ2* is concerned about communicating privacy risks in two scenarios: ex-ante and ex-post. This paper addresses the ex-ante scenario and presents a lab study on user behavior, decision making, and perception about privacy concern while selecting apps. An app store demo was presented to the user with a minor modification—a privacy indicator for each app. After carrying out several tasks using this modified

mobile interface, participants were interviewed to document reasons behind their decisions, thought process, and perception regarding individual privacy. A total of 82 adults volunteered under the pretext of a usability study. A statistically significant influence of the privacy indicator on their app selection behavior was observed, although this influence decreased in case of familiar apps. By varying the degrees of participatory background information, we show that impact of a privacy indicator on app selection behavior has statistical significance and such privacy-preserving behavior can be invoked by mere presence of the indicator.

Paper VII: Accept - Maybe - Decline: Introducing Partial Consent for the Permission-based Access Control Model of Android

This paper addresses the ex-post scenario to communicate privacy risks ($RQ2$), and inspects the feasibility of partial consent to access permissions in Android. Currently, the user is only able to grant the permission with indefinite validity and has binary options to do so—Accept or Decline. We propose a third option—Maybe, which could potentially enable the user to grant the requested rights with limitations, i.e., for a certain amount of time (hours/days), or number of accesses to system resources. Upon expiration of partially given consent, the user could review the app's performance in terms of privacy preserving attributes and proceed to grant or revoke consent with further extended validity. This paper presents a prototype implementation, and also addresses technical, regulatory, social, individual, and economic perspectives for inclusion of partial consent within an access control mechanism.

6 Concluding Remarks and Outlook

As an evolving technology, Android had been criticized for its flaws and many of them were addressed along with the evolution of the whole platform. One of the most discussed criticism was about mandatory acceptance of all the permissions presented before installing an app. The introduction of run-time permission-granting mechanism made significant improvement for the access control model. It certainly eased the dilemma by eliminating *accept, or leave* situation for the user, but privacy concerns persist due to opaque usage of privileges by the apps. Moreover, communicating privacy risks to the user remains challenging due to the very nature of privacy—it is variable in many possible dimensions.

This thesis addresses problems that are invisible to many mobile phone users and have the potential to cause psychological and/or social harm to them. Though a lot of research effort has already been given to draw the line between benign and malicious behavior of apps, the existence and the perimeter of gray area for good, or bad behavior are rather undefined. This thesis hypothetically considers the user data as currency, assumes that the apps are overpriced in

terms of personal data access, builds tools to run experiments, and shows that the phenomena exist. Then it addresses the problem with models, methods, and tools to conceptualize privacy-risks. Furthermore, it makes an effort to communicate the risks during two crucial use cases: *(i) ex-ante:* prior to choosing an app from the app store which could potentially aid the user in selecting privacy-friendly apps, and *(ii) ex-post:* offering an indecisive stage for consent that could empower the user with transparency and intervenability against privacy implications from apps' covert data access.

We argue that giving away white-card privilege to apps is similar to *leaving a water-tap open*. We also discuss that the users have poor means to assess the amount of personal information already transmitted/consumed by apps and to re-evaluate their earlier decisions. As the modern economy already monetizes user data, personal data/resource consumption of apps needs to be considered within the decision-making process of the end-user. This thesis presents research that could potentially accommodate such consideration within the ecosystem of apps.

For future endeavor, we possess keen interest in continuing research to introduce, to develop, and to evaluate privacy-enhancing technologies. Two immediate steps could be mentioned in this regard: *(i)* construction of a comprehensive method would address the issue on generating a cumulative privacy score that could be used in the app store, and *(ii)* carrying out a user study and a prototype development project to identify the feasibility, necessity, and technical requirements for accommodating partial consent within the access control mechanism of an operating system. The second one is being planned to commence in the near future.

References

[1] A. Acquisti, I. Adjerid, and L. Brandimarte. Gone in 15 seconds: The limits of privacy transparency and control. *IEEE Security & Privacy*, 11(4):72–74, 2013.

[2] E. Alepis and C. Patsakis. Hey Doc, Is This Normal?: Exploring Android Permissions in the Post Marshmallow Era. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*, pages 53–73. Springer, 2017.

[3] E. Alepis and C. Patsakis. Unravelling security issues of runtime permissions in android. *Journal of Hardware and Systems Security*, 3(1):45–63, 2019.

[4] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 787–796. ACM, 2015.

[5] R. Alvarez, J. Levenson, R. Sheatsley, and P. McDaniel. Application transiency: Towards a fair trade of personal information for application services. In S. Chen, K.-K. R. Choo, X. Fu, W. Lou, and A. Mohaisen, editors, *Security and Privacy in Communication Networks*, pages 47–66, Cham, 2019. Springer International Publishing.

[6] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Le Traon, D. Octeau, and P. McDaniel. FlowDroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps. *ACM Sigplan Notices*, 49(6):259–269, 2014.

[7] K. W. Y. Au, Y. F. Zhou, Z. Huang, and D. Lie. PScout: analyzing the Android permission specification. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 217–228. ACM, 2012.

[8] R. Baalous and R. Poet. How dangerous permissions are described in android apps' privacy policies? In *Proceedings of the 11th International Conference on Security of Information and Networks*, SIN '18, New York, NY, USA, 2018. Association for Computing Machinery.

[9] R. Balebako, J. Jung, W. Lu, L. F. Cranor, and C. Nguyen. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 12. ACM, 2013.

[10] A. Bauer, J.-C. Küster, and G. Vegliach. Runtime verification meets Android security. In *NASA Formal Methods Symposium*, pages 174–180. Springer, 2012.

[11] K. Bharat, S. Lawrence, and M. Sahami. Generating user information for use in targeted advertising, Jan. 12 2016. US Patent 9,235,849.

[12] W. R. Bischofberger and G. Pomberger. *Prototyping-oriented software development: Concepts and tools*. Springer Science & Business Media, 2012.

[13] T. Bläsing, L. Batyuk, A.-D. Schmidt, S. A. Camtepe, and S. Albayrak. An Android application sandbox system for suspicious software detection. In *Malicious and unwanted software (MALWARE), 2010 5th international conference on*, pages 55–62. IEEE, 2010.

[14] B. W. Boehm. A spiral model of software development and enhancement. *Computer*, 21(5):61–72, 1988.

[15] A. Bromander. Using privacy indicators to nudge users into selecting privacy friendly applications. Master's thesis, Karlstad University, 2019. Karlstad University Press.

[16] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani. Crowdroid: behavior-based malware detection system for android. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, pages 15–26. ACM, 2011.

[17] R. by Noel Packard. Overlooked history in the age of surveillance capitalism. *Journal of Media Economics*, 0(0):1–7, 2020.

[18] H. Cai and B. G. Ryder. A longitudinal study of application structure and behaviors in android. *IEEE Transactions on Software Engineering*, 2020.

[19] P. Calciati, K. Kuznetsov, X. Bai, and A. Gorla. What did really change with the new release of the app? In *2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)*, pages 142–152, 2018.

[20] A. Carlsson, C. Pedersen, F. Persson, and G. Söderlund. Kaudroid: A tool that will spy on applications and how they spy on their users. Technical report, 2018. Karlstad University Press.

[21] P. Chester, C. Jones, M. Wiem Mkaouer, and D. E. Krutz. M-perm: A lightweight detector for android permission gaps. In *2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pages 217–218, 2017.

[22] S. Chitkara, N. Gothoskar, S. Harish, J. I. Hong, and Y. Agarwal. Does this app really need my location? context-aware privacy management for smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), Sept. 2017.

[23] R. Clarke. A sufficiently rich model of (id) entity, authentication and authorisation. In *The 2nd Multidisciplinary Workshop on Identity in the Information Society, LSE*, volume 5, 2009.

[24] W. J. Conover and R. L. Iman. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3):124–129, 1981.

[25] I. D. Cooper. What is a "mapping study?". *Journal of the Medical Library Association: JMLA*, 104(1):76, 2016.

[26] Y. Cui, N. Shivakumar, A. Carobus, D. Jindal, and S. Lawrence. Content-targeted advertising using collected user behavior data, Jan. 27 2005. US Patent App. 10/649,585.

[27] J. S. Daniel. A taxonomy of privacy. *University of Pennsylvania law review*, 154(3):477–560, 2006.

[28] A. de Lima Fontão, R. P. dos Santos, and A. C. Dias-Neto. Mobile software ecosystem (mseco): a systematic mapping study. In *2015 IEEE 39th Annual Computer Software and Applications Conference*, volume 2, pages 653–658. IEEE, 2015.

[29] P. J. Denning. The science in computer science. *Communications of the ACM*, 56(5):35–38, 2013.

[30] M. Diamantaris, E. P. Papadopoulos, E. P. Markatos, S. Ioannidis, and J. Polakis. Reaper: Real-time app analysis for augmenting the android permission system. In *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy*, CODASPY '19, page 37–48, New York, NY, USA, 2019. Association for Computing Machinery.

[31] G. Dodig-Crnkovic. Scientific methods in computer science. In *Proceedings of the Conference for the Promotion of Research in IT at New Universities and at University Colleges in Sweden, Skövde, Suecia*, pages 126–130, 2002.

[32] N. Eling, S. Rasthofer, M. Kolhagen, E. Bodden, and P. Buxmann. Investigating users' reaction to fine-grained data requests: A market experiment. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 3666–3675, 2016.

[33] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, 32(2):5, 2014.

[34] EU Regulation. 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). *Off J Eur Union*, page L119, 2016.

[35] J. Feichtner and S. Gruber. Understanding privacy awareness in android app descriptions using deep learning. In *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, CODASPY '20, page 203–214, New York, NY, USA, 2020. Association for Computing Machinery.

[36] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner. Android permissions demystified. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 627–638. ACM, 2011.

[37] Y. Feng, L. Chen, A. Zheng, C. Gao, and Z. Zheng. Ac-net: Assessing the consistency of description and permission in android apps. *IEEE Access*, 7:57829–57842, 2019.

[38] D. Franzen and D. Aspinall. PhoneWrap-Injecting the "How Often" into Mobile Apps. In *Proceedings of the 1st International Workshop on Innovations in Mobile Privacy and Security co-located with the International Symposium on Engineering Secure Software and Systems (ESSoS 2016)*, pages 11–19. CEUR-WS.org, 2016.

[39] Y. Fratantonio, C. Qian, S. P. Chung, and W. Lee. Cloak and dagger: From two permissions to complete control of the ui feedback loop. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 1041–1057, 2017.

[40] J. Fu, Y. Zhou, H. Liu, Y. Kang, and X. Wang. Perman: Fine-grained permission management for android applications. In *2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)*, pages 250–259, 2017.

[41] A. P. Fuchs, A. Chaudhuri, and J. S. Foster. SCanDroid: Automated Security Certification of Android. Technical report, 2009.

[42] H. Gao, C. Guo, D. Huang, X. Hou, Y. Wu, J. Xu, Z. He, and G. Bai. Autonomous permission recommendation. *IEEE Access*, 8:76580–76594, 2020.

[43] I. Gasparis, Z. Qian, C. Song, S. V. Krishnamurthy, R. Gupta, and P. Yu. Figment: Fine-grained permission management for mobile apps. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 1405–1413, 2019.

[44] N. Gerber, P. Gerber, H. Drews, E. Kirchner, N. Schlegel, T. Schmidt, and L. Scholz. Foxit: Enhancing mobile users' privacy behavior by increasing knowledge and awareness. In *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust*, STAST '17, page 53–63, New York, NY, USA, 2018. Association for Computing Machinery.

[45] A. Goodman. Age of Surveillance Capitalism: "We thought we were searching Google, but Google was searching us." Democracy Now Interview (Part One) with Shoshana Zuboff., 2019.

[46] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang. Riskranker: scalable and accurate zero-day android malware detection. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 281–294. ACM, 2012.

[47] R. M. Groves, F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. *Survey methodology*, volume 561. John Wiley & Sons, 2011.

[48] M. Hammad, H. Bagheri, and S. Malek. Determination and enforcement of least-privilege architecture in Android. In *Software Architecture (ICSA), 2017 IEEE International Conference on*, pages 59–68. IEEE, 2017.

[49] M. Hansen, M. Jensen, and M. Rost. Protection goals for privacy engineering. In *Security and Privacy Workshops (SPW), 2015 IEEE*, pages 159–166. IEEE, 2015.

[50] M. M. Hassan Onik, N. Al-Zaben, J. Yang, N. Lee, and C. Kim. Risk identification of personally identifiable information from collective mobile app data. In *2018 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, pages 71–76, 2018.

[51] M. Hatamian, A. Kitkowska, J. Korunovska, and S. Kirrane. "it's shocking¡': Analysing the impact and reactions to the a3: Android apps behaviour analyser. In F. Kerschbaum and S. Paraboschi, editors, *Data and Applications Security and Privacy XXXII*, pages 198–215, Cham, 2018. Springer International Publishing.

[52] M. Hatamian, N. Momen, L. Fritsch, and K. Rannenberg. A multilateral privacy impact analysis method for android apps. In M. Naldi, G. F. Italiano, K. Rannenberg, M. Medina, and A. Bourka, editors, *Privacy Technologies and Policy*, pages 87–106, Cham, 2019. Springer International Publishing.

[53] M. Hatamian, J. Serna, K. Rannenberg, and B. Igler. Fair: Fuzzy alarming index rule for privacy analysis in smartphone apps. In *International Conference on Trust and Privacy in Digital Business*, pages 3–18. Springer, 2017.

[54] C. Herley and P. C. van Oorschot. Sok: Science, security and the elusive goal of security as a scientific pursuit. In *Security and Privacy (SP), 2017 IEEE Symposium on*, pages 99–120. IEEE, 2017.

[55] M. Z. Hossen and M. Mannan. On understanding permission usage contextuality in android apps. In F. Kerschbaum and S. Paraboschi, editors, *Data and Applications Security and Privacy XXXII*, pages 232–242, Cham, 2018. Springer International Publishing.

[56] F. Hsu, N. Liu, Y. Hwang, C. Liu, C. S. Wang, and C. Chen. Dpc:a dynamic permission control mechanism for android third-party libraries. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1, 2019.

[57] J. Huang, W. Huang, F. Miao, and Y. Xiong. Detecting stubborn permission requests in android applications. In *2018 4th International Conference on Big Data Computing and Communications (BIGCOM)*, pages 84–89, 2018.

[58] C. B. Jackson and Y. Wang. Addressing the privacy paradox through personalized privacy notifications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2), July 2018.

[59] P. Jayakumar, L. Lawrence, R. L. W. Chean, and S. N. Brohi. A review and survey on smartphones: The closest enemy to privacy. In M. H. Miraz, P. S. Excell, A. Ware, S. Soomro, and M. Ali, editors, *Emerging Technologies in Computing*, pages 106–118, Cham, 2019. Springer International Publishing.

[60] J. Jeon, K. K. Micinski, J. A. Vaughan, A. Fogel, N. Reddy, J. S. Foster, and T. Millstein. Dr. Android and Mr. Hide: fine-grained permissions in android applications. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 3–14. ACM, 2012.

[61] J. Kang, D. Steiert, D. Lin, and Y. Fu. Movewithme: Location privacy preservation for smartphone users. *IEEE Transactions on Information Forensics and Security*, 15:711–724, 2020.

[62] P. G. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402. ACM, 2013.

[63] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15):5802–5805, 2013.

[64] L. Kraus, I. Wechsung, and S. Möller. Using statistical information to communicate android permission risks to users. In *2014 Workshop on Socio-Technical Aspects in Security and Trust*, pages 48–55. IEEE, 2014.

[65] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[66] P. Lantz, A. Desnos, and K. Yang. DroidBox: Android application sandbox, 2012.

[67] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[68] A. Li, D. Darling, and Q. Li. Photosafer: Content-based and context-aware private photo protection for smartphones. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*, pages 10–18, 2018.

[69] S. Li, J. Chen, T. Spyridopoulos, P. Andriotis, R. Ludwiniak, and G. Russell. Real-time monitoring of privacy abuses and intrusion detection in android system. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 379–390. Springer, 2015.

[70] S. Li, S. Kumar, T. Dumitras, and V. S. Subrahmanian. *Breaking Bad: Forecasting Adversarial Android Bad Behavior*, pages 405–431. Springer International Publishing, Cham, 2018.

[71] H. W. Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.

[72] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 501–510. ACM, 2012.

[73] R. Liu, J. Cao, K. Zhang, W. Gao, J. Liang, and L. Yang. When privacy meets usability: Unobtrusive privacy permission recommendation system for mobile apps based on crowdsourcing. *IEEE Transactions on Services Computing*, 11(5):864–878, 2018.

[74] X. Liu, Y. Leng, W. Yang, W. Wang, C. Zhai, and T. Xie. A large-scale empirical study on android runtime-permission rationale messages. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 137–146, 2018.

[75] Y. Liu and A. Simpson. On the trade-off between privacy and utility in mobile services: A qualitative study. In S. Katsikas, F. Cuppens, N. Cuppens, C. Lambrinoudakis, C. Kalloniatis, J. Mylopoulos, A. Antón, S. Gritzalis, F. Pallas, J. Pohle, A. Sasse, W. Meng, S. Furnell, and J. Garcia-Alfaro, editors, *Computer Security*, pages 261–278, Cham, 2020. Springer International Publishing.

[76] L. Lu, Z. Li, Z. Wu, W. Lee, and G. Jiang. Chex: statically vetting android apps for component hijacking vulnerabilities. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 229–240. ACM, 2012.

[77] Y. Lu, Q. Li, P. Su, J. Pan, J. Yan, P. Zhan, and W. Guo. A comprehensive study of permission usage on android. In M. H. Au, S. M. Yiu, J. Li, X. Luo, C. Wang, A. Castiglione, and K. Kluczniak, editors, *Network and System Security*, pages 64–79, Cham, 2018. Springer International Publishing.

[78] U. Magnusson. A tool for visual analysis of permission-based data access on android phones. Master's thesis, Karlstad University, 2019. Karlstad University Press.

[79] K. Marky, A. Gutmann, P. Rack, and M. Volkamer. Privacy Friendly Apps-Making Developers Aware of Privacy Violations. In *IMPS@ESSoS*, pages 46–48, 2016.

[80] A. M. McDonald and L. F. Cranor. The cost of reading privacy policies. *ISJLP*, 4:543, 2008.

[81] P. E. McKnight and J. Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[82] N. Momen. Towards measuring apps' privacy-friendliness. *Licentiate Dissertation, Karlstad University*, 2018.

[83] N. Momen, M. Hatamian, and L. Fritsch. Did App Privacy Improve after the GDPR? *IEEE Security Privacy*, 17(6):10–20, 2019.

[84] P. Murmann and S. Fischer-Hübner. Tools for Achieving Usable Ex Post Transparency: A Survey. *IEEE Access*, 2017.

[85] A. Mylonas, M. Theoharidou, and D. Gritzalis. Assessing privacy risks in android: A user-centric approach. In *International Workshop on Risk Assessment and Risk-driven Testing*, pages 21–37. Springer, 2013.

[86] S. Narain, T. D. Vo-Huu, K. Block, and G. Noubir. The perils of user tracking using zero-permission mobile apps. *IEEE Security Privacy*, 15(2):32–41, 2017.

[87] D. C. Nguyen, E. Derr, M. Backes, and S. Bugiel. Short text, large effect: Measuring the impact of user reviews on android app security privacy. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 555–569, 2019.

[88] K. Olejnik, I. Dacosta, J. S. Machado, K. Huguenin, M. E. Khan, and J. Hubaux. Smarper: Context-aware and automatic runtime-permissions for mobile devices. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 1058–1076, 2017.

[89] Oxford English Dictionary Online, March 2018. *Privacy, noun*. Oxford University Press, 2018.

[90] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[91] H. Peng, C. Gates, B. Sarma, N. Li, Y. Qi, R. Potharaju, C. Nita-Rotaru, and I. Molloy. Using probabilistic generative models for ranking risks of android apps. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 241–252. ACM, 2012.

[92] A. Peruma, J. Palmerino, and D. E. Krutz. Investigating user perception and comprehension of android permission models. In *2018 IEEE/ACM 5th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pages 56–66, 2018.

[93] A. Pfitzmann and M. Hansen. Anonymity, unlinkability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology. In *Designing privacy enhancing technologies*, pages 1–9. Technische Universität Dresden, 10-Aug-2010.

[94] B. Rashidi and C. Fung. Xdroid: An android permission control using hidden markov chain and online learning. In *2016 IEEE Conference on Communications and Network Security (CNS)*, pages 46–54, 2016.

[95] B. Rashidi, C. Fung, A. Nguyen, T. Vu, and E. Bertino. Android user privacy preserving through crowdsourcing. *IEEE Transactions on Information Forensics and Security*, 13(3):773–787, 2018.

[96] N. Raval, A. Razeen, A. Machanavajjhala, L. P. Cox, and A. Warfield. Permissions plugins as android apps. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '19, page 180–192, New York, NY, USA, 2019. Association for Computing Machinery.

[97] J. R. Reidenberg, T. Breaux, L. F. Carnor, and B. French. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkely Technology Law Journal*, 30(1):39–68, 2015.

[98] L. Reinfelder, A. Schankin, S. Russ, and Z. Benenson. An inquiry into perception and usage of smartphone permission models. In S. Furnell, H. Mouratidis, and G. Pernul, editors, *Trust, Privacy and Security in Digital Business*, pages 9–22, Cham, 2018. Springer International Publishing.

[99] S. Rosen, Z. Qian, and Z. M. Mao. Appprofiler: a flexible method of exposing privacy-related behavior in Android applications to end users. In *Proceedings of the third ACM conference on Data and application security and privacy*, pages 221–232. ACM, 2013.

[100] A. Sadeghi, R. Jabbarvand, N. Ghorbani, H. Bagheri, and S. Malek. A temporal permission analysis and enforcement framework for android. In *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*, pages 846–857, 2018.

[101] A. Sadeghi, R. Jabbarvand, and S. Malek. Patdroid: Permission-aware gui testing of android. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*, ESEC/FSE 2017, page 220–232, New York, NY, USA, 2017. Association for Computing Machinery.

[102] J. H. Saltzer. Protection and the control of information sharing in multics. *Communications of the ACM*, 17(7):388–402, 1974.

[103] J. H. Saltzer and M. D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[104] G. L. Scoccia, I. Malavolta, M. Autili, A. Di Salle, and P. Inverardi. Enhancing trustability of android applications via user-centric flexible permissions. *IEEE Transactions on Software Engineering*, pages 1–1, 2019.

[105] G. L. Scoccia, A. Peruma, V. Pujols, B. Christians, and D. Krutz. An empirical history of permission requests and mistakes in open source android apps. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*, pages 597–601, 2019.

[106] G. L. Scoccia, A. Peruma, V. Pujols, I. Malavolta, and D. E. Krutz. Permission issues in open-source android apps: An exploratory study. In *2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pages 238–249, 2019.

[107] G. L. Scoccia, S. Ruberto, I. Malavolta, M. Autili, and P. Inverardi. An investigation into android run-time permissions from the end users' perspective. In *2018 IEEE/ACM 5th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pages 45–55, 2018.

[108] Securities and E. Commission. Amendment no. 9 to form s-1 registration statement under the securities act of 1933 for google inc., 2019.

[109] A. Shabtai, U. Kanonov, Y. Elovici, C. Glezer, and Y. Weiss. "Andromaly": a behavioral malware detection framework for Android devices. *Journal of Intelligent Information Systems*, 38(1):161–190, 2012.

[110] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2347–2356. ACM, 2014.

[111] J. Shu, J. Li, Y. Zhang, and D. Gu. Burn after reading: Expunging execution footprints of android apps. In M. H. Au, S. M. Yiu, J. Li, X. Luo, C. Wang, A. Castiglione, and K. Kluczniak, editors, *Network and System Security*, pages 46–63, Cham, 2018. Springer International Publishing.

[112] D. J. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154:477, 2005.

[113] A. Srivastava, P. Jain, S. Demetriou, L. P. Cox, and K.-H. Kim. Camforensics: Understanding visual privacy leaks in the wild. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, SenSys '17, New York, NY, USA, 2017. Association for Computing Machinery.

[114] S. Sundberg, A. Blomqvist, and A. Bromander. Kaudroid-project report: Visualizing how android apps utilize permissions. Technical report, 2019. Karlstad University Press.

[115] K. Tam, S. J. Khan, A. Fattori, and L. Cavallaro. CopperDroid: Automatic Reconstruction of Android Malware Behaviors. In *NDSS*, 2015.

[116] J. Tang, R. Li, H. Han, H. Zhang, and X. Gu. Detecting permission over-claim of android applications with static and semantic analysis approach. In *2017 IEEE Trustcom/BigDataSE/ICESS*, pages 706–713, 2017.

[117] X. Tang, Y. Lin, D. Wu, and D. Gao. Towards dynamically monitoring android applications on non-rooted devices in the wild. In *Proceedings of the 11th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '18, page 212–223, New York, NY, USA, 2018. Association for Computing Machinery.

[118] V. F. Taylor, A. R. Beresford, and I. Martinovic. There are many apps for that: Quantifying the availability of privacy-preserving apps. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '17, page 247–252, New York, NY, USA, 2017. Association for Computing Machinery.

[119] V. F. Taylor and I. Martinovic. Securank: Starving permission-hungry apps using contextual permission analysis. In *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 43–52. ACM, 2016.

[120] V. F. Taylor and I. Martinovic. Securank: Starving permission-hungry apps using contextual permission analysis. In *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '16, page 43–52, New York, NY, USA, 2016. Association for Computing Machinery.

[121] V. F. Taylor and I. Martinovic. To update or not to update: Insights from a two-year study of android app evolution. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 45–57. ACM, 2017.

[122] V. F. Taylor and I. Martinovic. To update or not to update: Insights from a two-year study of android app evolution. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, ASIA CCS '17, page 45–57, New York, NY, USA, 2017. Association for Computing Machinery.

[123] D. Titze, P. Stephanow, and J. Schütte. App-ray: User-driven and fully automated android app security assessment. *Fraunhofer AISEC TechReport*, 2013.

[124] L. Toresson, S. Olars, and M. Shaker. Privacy impact self-assessment app. Technical report, 2020. Karlstad University Press.

[125] E. Tromer and R. Schuster. Droiddisintegrator: Intra-application information flow control in android apps. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ASIA CCS '16, page 401–412, New York, NY, USA, 2016. Association for Computing Machinery.

[126] M. Van Kleek, I. Liccardi, R. Binns, J. Zhao, D. J. Weitzner, and N. Shadbolt. Better the devil you know: Exposing the data sharing practices of smartphone apps. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5208–5220. ACM, 2017.

[127] H. Wang, H. Li, and Y. Guo. Understanding the evolution of mobile app ecosystems: A longitudinal measurement study of google play. In *The World Wide Web Conference*, WWW '19, page 1988–1999, New York, NY, USA, 2019. Association for Computing Machinery.

[128] H. Wang, Y. Li, Y. Guo, Y. Agarwal, and J. I. Hong. Understanding the purpose of permission use in mobile apps. *ACM Trans. Inf. Syst.*, 35(4), July 2017.

[129] R. Wang, Z. Wang, B. Tang, L. Zhao, and L. Wang. Smartpi: Understanding permission implications of android apps from user reviews. *IEEE Transactions on Mobile Computing*, pages 1–1, 2019.

[130] X. Wang, K. Sun, Y. Wang, and J. Jing. DeepDroid: Dynamically Enforcing Enterprise Policy on Android Devices. In *NDSS*, 2015.

[131] S. D. Warren and L. D. Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.

[132] X. Wei, L. Gomez, I. Neamtiu, and M. Faloutsos. Permission evolution in the Android ecosystem. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pages 31–40. ACM, 2012.

[133] X. Wei, L. Gomez, I. Neamtiu, and M. Faloutsos. Profiledroid: multi-layer profiling of android applications. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 137–148. ACM, 2012.

[134] A. F. Westin. Privacy and freedom. *New York: Atheneum*, 1967.

[135] J. Wettlaufer and H. Simo. *Decision Support for Mobile App Selection via Automated Privacy Assessment*, pages 292–307. Springer International Publishing, Cham, 2020.

[136] P. Wijesekera, J. Reardon, I. Reyes, L. Tsai, J.-W. Chen, N. Good, D. Wagner, K. Beznosov, and S. Egelman. Contextualizing privacy

decisions for better prediction (and protection). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[137] Wikipedia. Companies that filed for chapter 11 bankruptcy in 2020, 2020.

[138] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[139] S. Xi, S. Yang, X. Xiao, Y. Yao, Y. Xiong, F. Xu, H. Wang, P. Gao, Z. Liu, F. Xu, and J. Lu. Deepintent: Deep icon-behavior learning for detecting intention-behavior discrepancy in mobile apps. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, page 2421–2436, New York, NY, USA, 2019. Association for Computing Machinery.

[140] L. Xie, X. Zhang, J.-P. Seifert, and S. Zhu. pBMDS: a behavior-based malware detection system for cellphone devices. In *Proceedings of the third ACM conference on Wireless network security*, pages 37–48. ACM, 2010.

[141] L.-K. Yan and H. Yin. DroidScope: Seamlessly Reconstructing the OS and Dalvik Semantic Views for Dynamic Android Malware Analysis. In *USENIX security symposium*, pages 569–584, 2012.

[142] L. Yu, X. Luo, X. Liu, and T. Zhang. Can we trust the privacy policies of android apps? In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 538–549, 2016.

[143] L. Yu, X. Luo, C. Qian, S. Wang, and H. K. N. Leung. Enhancing the description-to-behavior fidelity in android apps with privacy policy. *IEEE Transactions on Software Engineering*, 44(9):834–854, 2018.

[144] S. Zein, N. Salleh, and J. Grundy. A systematic mapping study of mobile application testing techniques. *Journal of Systems and Software*, 117:334–356, 2016.

[145] D. Zhang, Y. Guo, D. Guo, R. Wang, and G. Yu. Contextual approach for identifying malicious inter-component privacy leaks in android apps. In *2017 IEEE Symposium on Computers and Communications (ISCC)*, pages 228–235, 2017.

[146] H. Zhang, S. Latif, R. Bassily, and A. Rountev. Introducing privacy in screen event frequency analysis for android apps. In *2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pages 268–279, 2019.

[147] M. Zhao, T. Zhang, F. Ge, and Z. Yuan. Robotdroid: A lightweight malware detection framework on smartphones. *JNW*, 7(4):715–722, 2012.

[148] Z. Zhao and F. C. C. Osono. "TrustDroid": Preventing the use of SmartPhones for information leaking in corporate networks through the used of static analysis taint tracking. In *Malicious and Unwanted Software (MALWARE), 2012 7th International Conference on*, pages 135–143. IEEE, 2012.

[149] S. Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 2019.

Appendices

A Appendix: Repositories

1. KAUdroid Server: `https://git.cs.kau.se/nurumome/kaudroid-server`
2. Prototype app: `https://git.cs.kau.se/nurumome/prototypeApp`
3. Log analysis tool: `https://git.cs.kau.se/nurumome/static-log-analysis`

B Appendix: Tools and Data archives

1. Dynamic app-behavior visualizer: `http://193.10.227.39:1337/categories`
2. Visualization tool from Student Project II [114]: `https://git.cs.kau.se/nurumome/kaudroid`
3. Data-set containing apps' permission access logs: `https://git.cs.kau.se/nurumome/kaudroid-databackup-logs`
4. Data-set containing apps' privacy-policies: `https://git.cs.kau.se/nurumome/kaudroid-databackup-policydocs`
5. Data-set containing user-reviews: `https://git.cs.kau.se/nurumome/kau-frankfurt`

In case of difficulties about accessing the tools and repositories, please contact the author, or the Department of Mathematics and Computer Science, Karlstad University.

Measuring Apps' Privacy-Friendliness

Mobile apps brought unprecedented convenience to everyday life, and nowadays, hardly any interactive service exists without having an interface through an app. The rich functionalities of apps rely on the pervasive capabilities of the mobile device. Consequently, apps generate a diverse and large amount of data, which can often be deemed as privacy-sensitive data.

Even though mobile operating systems use access control mechanisms to guard system resources and sensors, apps exercise their granted privileges in an opaque manner. Furthermore, available control tools lack monitoring features, and therefore, the user faces hindrances to comprehend the magnitude of personal data access.

This thesis covers a long-term investigation of apps' data access behavior and makes an effort to shed light on various privacy implications. It also shows that app behavior analysis yields information that has the potential to increase transparency, to enhance privacy protection, to raise awareness regarding consequences of data disclosure, and to assist the user in informed decision-making while selecting apps or services.

ISBN 978-91-7867-132-8 (print)

ISBN 978-91-7867-137-3 (pdf)

ISSN 1403-8099

DOCTORAL THESIS | Karlstad University Studies | 2020:24

Measuring Apps' Privacy-Friendliness

Mobile apps brought unprecedented convenience to everyday life, and nowadays, hardly any interactive service exists without having an interface through an app. The rich functionalities of apps rely on the pervasive capabilities of the mobile device. Consequently, apps generate a diverse and large amount of data, which can often be deemed as privacy-sensitive data.

Even though mobile operating systems use access control mechanisms to guard system resources and sensors, apps exercise their granted privileges in an opaque manner. Furthermore, available control tools lack monitoring features, and therefore, the user faces hindrances to comprehend the magnitude of personal data access.

This thesis covers a long-term investigation of apps' data access behavior and makes an effort to shed light on various privacy implications. It also shows that app behavior analysis yields information that has the potential to increase transparency, to enhance privacy protection, to raise awareness regarding consequences of data disclosure, and to assist the user in informed decision-making while selecting apps or services.

ISBN 978-91-7867-132-8 (print) | ISBN 978-91-7867-137-3 (pdf)

DOCTORAL THESIS | Karlstad University Studies | 2020:24