

Mobile App Privacy in Software Engineering Research: A Systematic Mapping Study

Fahimeh Ebrahimi, Miroslav Tushev, and Anas Mahmoud

Abstract—Mobile applications (apps) have become deeply personal, constantly demanding access to privacy-sensitive information in exchange for more personalized user experiences. Such privacy-invading practices have generated major multidimensional and unconventional privacy concerns among app users. To address these concerns, the research on mobile app privacy has experienced rapid growth over the past decade. In general, this line of research is aimed at systematically exposing the privacy practices of apps and proposing solutions to protect the privacy of mobile app users. In this survey paper, we conduct a systematic mapping study of 54 Software Engineering (SE) primary studies on mobile app privacy. Our objectives are to **a)** explore trends in SE app privacy research, **b)** categorize existing evidence, and **c)** identify potential directions for future research. Our results show that existing literature can be divided into four main categories: privacy policy, requirements, user perspective, and leak detection. Furthermore, our survey reveals an imbalance between these categories—majority of existing research focuses on proposing tools for detecting privacy leaks, with less studies targeting privacy requirements and policy and even less on user perspective. Finally, our survey exposes several gaps in existing research and suggests areas for improvement.

Index Terms—Privacy, mobile application, systematic mapping study.



1 INTRODUCTION

PRIVACY can be hard to define. The word itself is derived from the Latin *Privatus*, which means “withdraw from public life”. However, the majority of modern definitions can be traced back Brandeis and Warren [1] who defined privacy in 1890 as “the right to be let alone”. This definition provided a basis for ensuring the legal protection of privacy as a fundamental human right. Since then, this definition has been revisited numerous times to count for the plethora of political, social, and technological advances in society. For instance, to meet the growing dimensions of the concept, Solove [2] expanded the definition of privacy to include limited access to the self, secrecy, control of personal information, personhood, and intimacy. Another notable definition was introduced by Westin [3], who in 1968, described privacy as “the right to select what personal information about me is known to what people” and later in 2003 as “the claim of an individual to determine what information about himself or herself should be known to others” [4]. Both definitions emphasized privacy as a control over personal information.

The proliferation of mobile devices over the past decade along with their unique operational characteristics have imposed new challenges on end-users privacy. In general, mobile apps are designed with a set of goals in mind. A goal can be described as any desirable user objective that the system under consideration should achieve [5]. However, driven by the fierce market competition, app developers often deviate from their original goals [6]. These deviations typically come in the form of extreme privacy-invading tactics, such as constant location-tracking [7], unsolicited usage data collection [6, 8], or any form of features that are engineered to lure users into sacrificing their privacy in exchange for more personalized services [9, 10].

These intrusive, and sometimes borderline unethical, practices have led to the emergence of new and more significant privacy concerns among mobile app users. Such concerns often revolve around the types of information apps are asking for, who should or should not have access to this information, and how to prevent misuse of this access. In fact, these new challenges have prompted researchers to broaden existing definitions to include, in addition to personal information, device-specific information that can be used as identifiers, including installed apps, connected WIFI, operating system’s build information, and carrier [10].

In response to these challenges, the research on mobile app privacy has witnessed rapid growth over the past decade. In general, studies in this domain cover a broad range of topics, tackling privacy from a user, system, and even legal perspectives. To categorize, summarize, and synthesize this body of research, in this paper, we conduct a survey of existing Software Engineering (SE) literature related to mobile app privacy. Our survey takes the form of a Systematic Mapping Study (SMS). SMSs can be considered a form of Systematic Literature Reviews (SLRs), however, SMSs tend to be more qualitative in nature, mainly focused on the classification and thematic analysis of existing literature in a scientifically rigorous way. SLRs, on the other hand, are concerned with using statistical meta-analysis methods to synthesize the outcomes of empirical studies [11, 12]. SMSs have been long used in research as powerful tools for organizing and categorizing existing evidence and identifying areas for improvement [13]. A thorough discussion of the difference between SLRs and SMS is available in Kitchenham et al. [13] and Petersen et al. [14]. In general, our objectives in this survey paper include:

- Providing descriptive statistics of existing primary studies on mobile app privacy published in SE

venues between 2010 and 2018.

- Systematically categorizing and sub-categorizing these primary studies based on the research problems they address.
- Summarizing existing evidence under each category and identifying research gaps and potential areas for improvement.

The remainder of this paper is organized as follows. Section 2 describes our research method, including our research questions and primary study identification process. Section 3 categorizes and summarizes existing primary studies and identifies directions of future work under each category. Section 4 discusses the main findings and limitations of our work. Finally, Section 5 concludes the paper.

2 METHOD

The survey presented in this paper takes the form of a Systematic Mapping Study. According to Kitchenham et al. [13], a mapping study follows the same principled process as systematic literature reviews [12]. This process consists of three main steps: planning, conducting, and reporting. Under the *planning* phase, the need for the review is justified, the review protocol is established, and the research questions are defined. During the *conducting* phase, the review protocol is put into action, including the identification of primary studies, categorizing, and synthesizing existing evidence. Finally, under the *reporting* phase, the results are reported in a way that is tailored for the intended audience. In this section, we describe the main steps of our survey in detail.

2.1 Research questions

It is essential to identify a set of research questions before taking on a review study. Research questions are necessary to set the *scope* of the primary studies that should be considered in the search process and outline the objectives of the review. In this survey, our research questions are:

- **RQ₁:** *How much evidence is available in Software Engineering research about mobile app privacy?* Under this research question, we seek to determine the number of papers published on mobile app privacy in Software Engineering research. Our objective is to systematically assess the community's effort on this problem. Research venues tackling mobile privacy from other perspectives such Machine Learning or Computer Human Interaction, are excluded.
- **RQ₂:** *What are the main categories of existing research?* Under this question, we seek to build a classification scheme of existing primary studies [14]. Our objective is to narrow down the subjects of mobile app privacy research to few well-defined areas. Such categorization can provide researchers and practitioners with a much needed reference structure, or visual map, that can be easily navigated to get a quick overview of the frequency of publications and their main trends over the past decade.
- **RQ₃:** *What are the limitations of the current research?* Under this question, we seek to identify areas for improvement in existing SE mobile app privacy research. Specifically, through our categorization, we

seek to identify areas where the research has not progressed in comparison to other areas. Furthermore, we analyze the future work sections of exiting papers to outline the authors' visions of future work directions in their areas of research. Our objective is to help junior researchers, or researchers breaking into the field, to identify the areas which are lacking evidence, or the views of experts on research directions that need more attention or worth pursuing.

2.2 Search Process: Identifying primary studies

Our scope includes Software Engineering primary studies (papers that are not surveys, or secondary studies) tackling privacy of mobile apps, published in the period from 2010 to 2018. To identify these papers, we followed a three-step process. At the first step, we searched four scientific databases: Google Scholar, ACM Digital Library, IEEE Xplore, and arXiv. Our search query can be described as follows:

(Mobile AND (app OR apps) AND privacy) OR (mobile AND application AND privacy) OR (mobile AND app AND privacy AND permission) OR (mobile AND app AND privacy AND (policy OR policies)) OR (privacy AND app store)

The number of papers identified at the first step was 524. However, since we are only considering software engineering research, we only considered papers published in SE venues. Limiting the search to SE venues was necessary for two main reasons. First, privacy is a very broad concept, as of June 2019, a trivial search for *mobile privacy* on Google Scholar returns 3,810,000 hits. Therefore, it is important to narrow down the scope of search in order to get a reasonable number of primary studies. Second, our target audience is the SE community, or researchers and practitioners who typically publish in SE venues. To identify our venues, we relied on previous SMS and SLR studies in the literature [15]. In general, any venue with "Software Engineering" in its name, or any venue (workshop) that was held with a major SE venue were included in our analysis. Furthermore, we enforced the time scope of 2010 - 2018 on selected papers. 2010 was the year where initial work on mobile app privacy started emerging as a new line of research in SE, following the launch of the first app store in 2008. A search for primary studies before 2010 did not return any hits. In summary, 85 papers were found in SE venues between 2010 and 2018. These venues, along with the number of papers found in each, are shown in Table. 1, the number of papers identified based on each of our sub-queries are shown in Table. 2.

2.3 Inclusion and exclusion criteria

In SMSs and SRLs, the inclusion and exclusion criteria are used as a basis for selecting primary studies. Such criteria should be determined beforehand during the planning phase. Our inclusion criteria in this paper are:

- Books, papers, technical reports, and grey literature.
- The study explicitly addresses privacy aspects of mobile apps.
- The study is published in English.

TABLE 1: Selected journals and conference proceedings.

Source	Acronym	Count	Included
ACM Transactions on Software Engineering and Methodology	TOSEM	1	1
Communications of the ACM	CACM	1	0
Empirical Software Engineering	EMSE	3	2
IEEE Software	IEEE SW	1	1
IEEE Transactions on Software Engineering	TSE	5	5
Information and Software Technology	IST	2	1
International Conference Fundamental Approaches to Software Engineering	FASE	1	1
Inter. Conference on Automated Software Engineering	ASE	5	6
Inter. Conference on Mobile Software Engineering and Systems	MOBILESoft	15	6
Inter. Conference on Software Engineering	ICSE	14	11
Inter. Conference on Software Engineering Companion	ICSE Comp.	2	0
Inter. Conference Software Analysis, Evolution, and Reengineering	SANER	4	2
Inter. Journal of Secure Software Engineering	IJSSE	1	0
Inter. Requirements Engineering Conference	RE	6	5
Inter. Symposium on Dependable Software Engineering: Theories, Tools, and Applications	SETTA	1	1
Inter. Symposium on Software Reliability Engineering	ISSRE	5	3
Inter. Symposium on Software Testing and Analysis	ISSTA	5	4
Inter. Symposium on Foundations of Software Engineering	FSE	4	1
Inter. Working Conference on Requirements Engineering: Foundation for Software Quality	REFSQ	1	0
Inter. Workshop on Automation of Software Test	AST	1	1
Inter. Workshop on Requirements Engineering and Law	RELAW	1	1
Journal of Systems and Software	JSS	3	0
Software: Practice and Experience	SPE	3	2
Workshop on Requirements Engineering	WRE	0	0
Sum		85	54

TABLE 2: The number of primary studies identified using each search query in the different repositories considered in our survey.

Search Query	Scholar	IEEE	ACM	arXiv
<i>mobile & app & privacy</i>	11	9	4	0
<i>mobile & application & privacy</i>	3	6	2	0
<i>mobile & app & privacy & policy</i>	0	3	1	0
<i>app & privacy & permission</i>	0	4	1	0
<i>app & store & privacy</i>	3	2	0	0
<i>snowballing</i>	5	0	0	0

We used the following *exclusion criteria* to exclude any studies that were irrelevant to our survey goals:

- Short papers (less than 4 pages), editorials, and summaries of keynote, or tutorial papers.
- Secondary studies (i.e., existing SLRs or SMSs).
- Duplicate reports of the same study. In case of duplication, the most recent version is selected.

To include and exclude papers, each paper was examined by each of the three authors individually. Specifically, each author read the title, abstract, and body of each of our 85 papers to determine its relevance to our survey. Each judge flagged each paper as *Include*, *Neutral*, and *Exclude*. The paper was then included or excluded based on the protocol shown in Table 2. In total, 36 papers out of our 85 papers were excluded from the analysis.

To reduce the risk of omitting relevant studies, we also performed a lightweight backward-snowballing on the included papers [16]. We basically inspected the studies cited by each of our included primary studies and the publications that subsequently cited the study, using Google

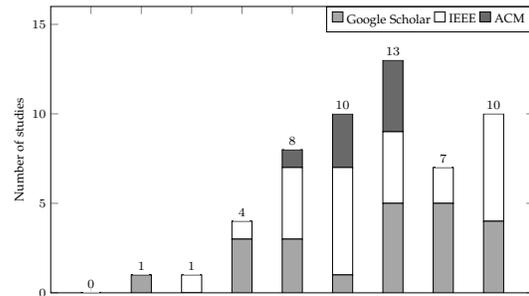


Fig. 1: The number of published papers per year from 2010 to 2018.

Scholar. In total, 6 more papers were identified, raising the number of primary studies to be included in our survey to 54 papers. The histogram in Fig. 1 shows the number of identified studies per year from 2010 to 2018. In this histogram, studies found using snowballing were counted as Google Scholar papers. A summary of our search process is shown in the schematic diagram in Fig. 2 and the distribution of final set of included papers over venue type is shown in Fig. 3. A detailed description of our search process can be found in (<http://seel.cse.lsu.edu/data/TSE2019.xlsx>)

3 PRIMARY STUDY CATEGORIZATION

To develop our classification scheme (RQ_2), we follow the protocol proposed by Petersen et al. [14]. In general, this protocol can be described as follows:

- The reviewers read the abstracts and look for keywords and concepts that reflect the contribution of

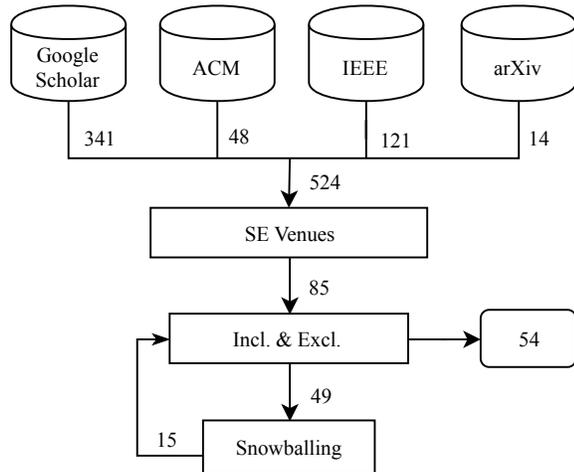


Fig. 2: A schematic diagram of our proposed research plan.

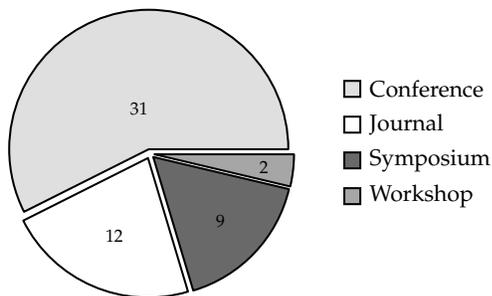


Fig. 3: The distribution of primary study over venue type.

the paper. If the abstract is poorly structured, the body of the paper is examined.

- A set of keywords representative of the context, or contribution of the paper, are extracted.
- The paper is assigned to an existing category based on its keywords. If the category does not exist, a new category is created.

In general, we identified four different categories: privacy policy, privacy requirements, mobile app users' perspective of privacy, and privacy leak analysis. The last category was further classified into two main categories of static and mixed static-dynamic methods. In what follows, we summarize the papers under each category and we highlight any areas for improvement.

3.1 Privacy Policy

Mobile apps' privacy policies act as contracts between app users and providers. In general, policies inform users about how their personal information is collected, used, shared, and protected by the app [8, 17]. Problems in this domain arise from inconsistencies between privacy claims in the policy and the actual behavior of the app. These inconsistencies (also known as violations) often take the form of omission errors, where the policy fails to notify the user about a specific information-privacy practice [8, 18]. In what follows, we summarize the papers classified under this category.

TABLE 2: The decision making process. IN: Include, EX: Exclude, and NU: Neutral.

Researcher ₁	Researcher ₂	Researcher ₃	Final Decision
IN	IN	IN/EX/NU	Include
EX	EX	IN/EX/NU	Exclude
NU	NU	IN/EX/NU	Consensus Meeting
IN	EX	NU	Consensus Meeting

3.1.1 Existing Work

Yu et al. [19, 20] introduced *TAPVerifier*, an automated approach for establishing semantic correlations between apps' privacy policies and their bytecode. The authors' goal was to identify apps' description-to-behavior fidelity, or whether the app behaves as advertised. *TAPVerifier* relied on text analysis to automatically establish and match the semantic meaning of the app's privacy policy, its bytecode, app store description, and permissions. Evaluating *TAPVerifier* on a dataset of 1,200 apps revealed that privacy policies were more likely to describe privacy-related behaviors than descriptions. The results also showed that the proposed approach was able to remove up to 59.4% of false alerts generated by tools such as *Whyper* and *AutoCog*.

Slavin et al. [21] proposed a semi-automated framework to help app developers detect inconsistencies between their privacy policies and code. The proposed framework linked policy phrases to Application Program Interface (API) methods that produced sensitive information. Information flow analysis was then used to detect misalignments by identifying methods that sent data to third-party remote servers. The proposed framework was empirically evaluated on 477 Android apps and discovered 341 potential privacy policy violations.

Aydin et al. [6] introduced *VisiDroid*, a visual configuration interface for end users to understand and configure the way their private information was utilized. The objective was to tailor privacy policies according to the preferences and sensitivities of the user. The authors targeted privacy threats emerging from contextual advertising. Specifically, through a special GUI, end users were able to configure constraints on the context. These configurations were then enforced by the app by anonymizing private fields before being sent to remote servers. The usability of *VisiDroid* was evaluated using a user study of 20 participants. All participants were able to finish all their privacy configuration tasks successfully.

Wang et al. [8] proposed an approach to automatically detect app privacy policy violations based on user input. The proposed approach mapped each Graphical User Interface (GUI) input element of the app to ontology concepts, which in turn were matched with text from the privacy policy. Static information flow analysis was then used to detect inconsistencies between the data collection behavior of the app and the statements of data-collection in the policy (i.e. *privacy leaks*). Evaluating the proposed approach over 120 popular apps, sampled from the finance, health, and dating domains, that it was able to detect 39 strong and weak violations in the studied apps.

Yu et al. [22] presented *PPChecker*, a tool for assessing the trustworthiness of Android apps' privacy policies. *PPChecker* employed natural-language processing (NLP) techniques, such as syntactic analysis, to analyze privacy policies, and adopted static program analysis approaches to inspect if the app's code collected, retained, or disclosed personal information. Based on this analysis, the authors identified five kinds of problems in privacy policies: incompleteness, incorrectness, impreciseness, inconsistency, and not being user-friendly. Applying *PPChecker* to 2,500 popular apps revealed that 1,850 apps (i.e., 74.0%) had at least one kind of the identified privacy problems.

3.1.2 Summary and Future Work

Our survey revealed that privacy policy research revolves around detecting policy violations by mapping privacy claims in the app's privacy policy to information flow in the app's code. Static code analysis tools such as *FlowDroid* [23] are typically used to track information flow in bytecode. Furthermore, majority of the papers under this category propose some sort of a working prototype to implement their findings (e.g., *PPChecker* [22] and *TAPVerifier* [19] and *VisiDroid* [6]) and evaluation is mainly carried out on Android apps as it is possible to get their APK files. Finally, our survey revealed that a large percentage of apps were either non-compliant with their stated privacy claims, did not provide a policy, or provided a policy that was ambiguous or incomprehensible to average users [24].

In general, the line of research under this category is still at its preliminary stages [21]. Specific directions of future work include combining more dynamic and static analysis methods with policy analysis methods [22], developing techniques to further trace the flow of user information to the respective organizations or servers [21], and expanding the scope of privacy threats to areas other than contextual advertising [6, 8]. In terms of tool support, future policy analysis tools should be engineered to meet the demands of their specific userbase (e.g., developers, users, regulators). Such tools should help app developers and policy authors to stay aware of any violations in their apps, enable end-users to determine the trustworthiness of the app [22], and help regulators, such as the U.S. Federal Trade Commission (FTC), or even app stores, to identify questionable or malicious apps [21].

3.2 Privacy in Mobile App User Feedback

App stores provide a mechanism for app users to express their opinions about apps in the form of textual reviews and meta-data (e.g., star ratings). Recent research revealed that around 40% of app store reviews on popular app stores contain actionable maintenance requests [25, 26]. Addressing these requests was found to have a significant positive impact on the rating of the app [27, 28]. Under this category, we summarize papers that identified privacy as a major concern in user feedback.

3.2.1 Existing Work

Khalid et al. [29] conducted an analytical study of user reviews to help developers better anticipate and prioritize possible user complaints. The authors manually examined

and classified thousands of one and two star app reviews, sampled from 20 iOS apps. The analysis uncovered 12 types of common users complaints. Analyzing the impact of each compliant type on the apps' ratings revealed that reviews complaining about privacy invading business practices were often associated with the most negative impact on the app as reflected in the ratings.

McIlroy et al. [30] conducted a qualitative analysis of close to 7,000 user reviews sampled from Google Play and the Apple App Store. The results showed that a substantial amount (30%) of reviews raised more than one technical issue. In terms of privacy, the authors found that 17% of reviews raised some sort of privacy concerns also raised other types of functional or technical concerns. The concentration of these concerns seemed to vary among app categories, where the Shopping category had the highest percentage of privacy complaints. The authors also reported that privacy issues were expressed using more varied language than other technical issues.

Ciurumelea et al. [31] proposed a taxonomy to analyze reviews and codes of mobile apps for a better release planning. The authors defined specific categories of user reviews that were highly relevant for developers during software maintenance, including compatibility, usage, resources, pricing, and privacy. A machine learning based prototype was then introduced to classify reviews and recommend source code files that were likely to be modified to handle issues raised in the reviews. Evaluating the proposed approach over 39 Android apps showed that it was able to detect privacy and protection issues with up to 83% precision and 96% recall.

Scoccia et al. [32] conducted a large-scale empirical study to investigate end users perception of the new Android's run-time permission system. The authors collected over 4.3 million user reviews from 5,572 Google Play apps. A sample of 3,574 permission-related reviews were analyzed using machine learning (Naive Bayes and Support Vector Machines) and NLP techniques. The most common permission-related complaints were then classified into a taxonomy of issues. The results showed that 8% of collected apps had reviews with negative comments about permissions. These permission-related reviews occurred in almost all app categories.

3.2.2 Summary and Future Work

Our survey revealed that end-user privacy concerns, while not as common as other functional issues (reliability and usability), are widespread among almost all app categories and can have a significant negative impact on the ratings of apps [30, 29]. However, our survey exposed a gap in existing research related to mining end-user privacy concerns. Specifically, the majority of existing work proposes generic approaches for mining user feedback for generic maintenance requests (bug reports and feature requests) [33]. These techniques often struggle when it comes to detecting specific user issues [26]. This emphasizes the need for data mining techniques that are specifically tailored for detecting privacy concerns in user feedback, for instance, building taxonomies of user-defined privacy concerns across the different categories of the app store. Extracted information should be

then used to facilitate a privacy-aware app design and development process.

3.3 Mobile App Privacy Requirements

Under this category, we summarize SE primary studies related to mining, specifying, and enforcing privacy requirements of mobile apps.

3.3.1 Existing Work

Young [18] proposed an approach to obtain software requirements from privacy policies. The author used a grounded theory approach to systematically analyze policies and extract commitments, privileges, and rights conveyed within the policies. Extracted information was classified into 12 different procedural (the privacy practice) and legal (the law enforces the practice) categories and then operationalized into software requirements using predefined requirements templates. The analysis was conducted using a case study on 17 health-care privacy policies. The results showed that the proposed approach achieved better coverage in comparison to legal-based approaches since policies often describe procedural practices rather than legal practices.

Tun et al. [34] introduced privacy arguments as a means of analyzing privacy requirements in general and selective disclosure requirements in particular. These arguments enabled users to express their personal privacy preferences through an extended argumentation language. User preferences were then used to represent highly dynamic selective disclosure requirements, and to relate them to the software architecture in order to enable system adaptation to runtime privacy requirements. The practical feasibility of the proposed approach was demonstrated over a mobile app that was developed by the authors.

Omoronyia et al. [35] proposed an adaptive privacy framework to support the selective disclosure of personal information in mobile apps. The framework exploited privacy awareness requirements (PAR) to identify the runtime privacy properties that should be satisfied in order to manage the changing user privacy concerns. The framework was evaluated over multiple cases where the failures and satisfactions of PAR were realized. The results showed that applications that failed to satisfy PAR were unable to regulate information flow based on the utility of disclosure, while applications that satisfied PAR were able to regulate the disclosure of information with changing context.

Thomas et al. [36] proposed a problem analysis framework to extract and refine privacy requirements for mobile apps from data gathered via empirical end-user studies. The framework provided analytical tools such as thematic coding, heuristics, facet questions, and extraction rules to enable a structured analysis of the various privacy problems experienced by users. Furthermore, the framework provided a privacy language to reason about and model privacy requirements. The operation of the framework was demonstrated using data collected through a case study targeting users of the Facebook app. The results showed that the approach was able to provide systematic aid to software engineers in deriving privacy requirements that addressed end-users' privacy concerns and needs.

Mai et al. [37] proposed a use-case driven method to support the specification of security and privacy requirements of multi-device software ecosystems, including mobile and wearable device applications. Specifically, the authors integrated an existing approach of modeling security threats, their mitigation, and their relations to use cases in a *misuse* case diagram. The authors further introduced templates for specifying mitigation schemes and misuse case specifications. A regular expression engine was then used to automatically report inconsistencies among artifacts and between the templates and specifications. The proposed approach was applied over industrial multi-device health-care project and was further evaluated using structured interviews with four software engineers. The proposed approach was able to precisely specify and analyze security threats along with threat scenarios and their mitigations.

Breaux et al. [38, 39] introduced *Eddy*, a methodology to map a well-defined subset of privacy requirements from natural language text to a formal language in description logic. *Eddy* was intended to help developers detect conflicting privacy requirements within a policy and enable the tracing of data flows within these policies. A computer simulation was used to demonstrate the scalability of the proposed language toolset to specifications of reasonable size. In a more recent work, Breaux et al. [40] targeted the privacy risks which arise from the increased information sharing across services (APIs). Specifically, the authors presented a new *Eddy* extension to formally model multi-party data flows requirements. These models were aligned with three critical privacy principles: purpose specification, collection limitation, and use limitation. The proposed approach was evaluated through a case study on the mobile app Waze and three of its service providers: Facebook, Amazon Web Services, and Flurry.com. The results showed that the proposed extensions were able to detect conflicts between Waze and Flurry privacy policies as well as violations of the principle. In addition, the analysis uncovered two privacy specification design patterns whereby specification authors can bypass the limitation principle.

Van Der Sype and Maalej [41] proposed a set of concrete requirements for app developers to protect users' privacy, focusing on the disclosure of users' personal data to third parties. These requirements were derived from the EU Data Protection Directive (Directive 95/46/EC) principles for fair and lawful processing. These principles were categorized into four groups: purpose limitation, data minimization, data security, and transparency. The authors further discussed the tension between law and technology, providing concrete strategies for developers to ease this tension.

Liu and Simpson [42] introduced *PPTMA* (Privacy-Preserving Targeted Mobile Advertising), a solution to achieve a balance between user privacy and utility in the context of mobile ads. The authors established the functional and non-functional requirements, design guidelines, and architecture of *PPTMA* based on the unique privacy challenges of mobile advertising. *PPTMA* was implemented as a background service which can be dynamically configured to enable users to adjust their privacy settings to control ads' access to their private information. The prototype was evaluated on 200 Android apps. The results showed that *PPTMA* was able to operate with low memory and power

consumption, providing effective means for ad-scanning.

3.3.2 Summary and Future Work

Our survey revealed that specifying privacy requirements for mobile apps can be a very challenging process. These requirements are highly dynamic and often context specific, depending on many factors such as time, location, information content, domain, and user preferences [36, 34, 35]. This problem becomes even more challenging in multi-device ecosystems where the app has to communicate information with other devices over the network [37].

In terms of future work, our survey exposed several potential areas of improvement in mobile privacy requirements research. These areas are mainly focused on devising practical techniques for eliciting, modeling, and managing mobile app privacy requirements. Elicitation techniques should take into account the constantly changing nature of privacy requirements under the different usage scenarios of the apps. This emphasizes the need for new ethnographic analysis techniques, such as interaction analyses, to obtain tacit knowledge of users' behavior in different contexts [36]. These techniques should also consider factors such as existing privacy regulations and privacy policies when generating the privacy requirements of the app [41, 18].

Once privacy requirements are specified, modeling techniques should be proposed to represent the relations (synergies and trade-offs) of these requirements to other entities in their specific domain, including user goals and functional features. Existing functional and non-functional requirements modeling techniques, such as Feature Oriented Domain Analysis (FODA) [43] and Softgoal Interdependency Graphs (SIGs) [44, 45] can be used to show the mandatory, alternative, and optional features of the domain along with their commonalities and variabilities [46]. Once these models are generated, they can act as blueprints of privacy-aware requirements models that app developers in different application domains can use as references for their design. Furthermore, such models should be adaptive in order to reflect the constantly changing privacy concerns and the newly emerging threats [35]. Finally, the authors of primary studies emphasize the need for effective automated tool support and extrinsic evaluation strategies to empirically assess the performance of the proposed techniques in realistic settings (e.g., validation through experience) [36, 34, 42, 40].

3.4 Privacy Leak Analysis

Under this category, we summarize papers that target privacy leaks and malicious behavior in mobile apps. In total, 33 primary studies were classified under this category. To better understand the summaries of these studies, we provide the following definitions:

- **Static Analysis:** static analysis methods examine source code or its binary representations without executing the code. Such methods often employ control flow graph analysis [7]. FlowDroid [23] and IccTA [47] are examples of static analysis tools that are commonly used in privacy leak analysis.
- **Dynamic Analysis:** refers to monitoring the code as it executes, thus enabling a precise security analysis based on the run-time behavior of the software. In

general, dynamic methods use data flow methods to track data from untrusted sources as the program executes to detect malicious behaviors [48].

- **Taint Analysis:** is a static/dynamic code analysis method that is commonly used for privacy leak detection. The goal of the process is to observe and track tainted/untainted information flow between sources and sinks. In the context of mobile apps, a tainted source is any system call that access private data and a sink is all the possible ways that make data leave the system (e.g., leak) [49]. TaintDroid [50] is a commonly used tool for taint analysis in Android apps.
- **Privacy Leak:** a privacy leak could be defined as any path that sends user's sensitive information to the outside world (e.g., a third party server) without the user's permission [51].
- **Permission:** In the context of mobile apps, a permission governs the type of user information the app is allowed to access. This could range from the data stored in the mobile device's memory to any peripherals of the device like the camera or the motion sensor. Users must grant permissions to the app before it can access these resources.

Given the large number of papers classified under this category, we combine related papers summaries into two subcategories of purely static, and mixed static/dynamic method analysis.

3.4.1 Static Analysis Methods

Feng et al. [52] presented *Apposcopy*, a semantic-based method to detect apps that leak private user information. *Apposcopy* used a high-level language to describe the semantic signatures of malware families and static analysis to determine whether an app matches a malware signature. Signature matching was enabled through static taint analysis and a novel code representation called Inter-Component Call Graph. These graphs tracked information flow using pointer analysis. *Apposcopy* was evaluated on 1027 malicious and 11,215 benign Android apps. The results showed that it detected malware with high accuracy (90%) and that its signatures were resilient to various program obfuscations.

Huang et al. [53] introduced *AsDroid*, a technique for detecting stealthy malicious behaviors in Android apps. The technique consisted of two parts: a static analysis module for extracting API invocations, and a UI analysis module for extracting the displayed in the interface. Any mismatch between the two components' outputs was considered as a potential stealthy behavior. A dataset of 182 potentially problematic Android apps was used to evaluate *AsDroid*. *AsDroid* was able to report stealthy behaviors in 113 apps, with 28 false positives and 11 false negatives.

Bartel et al. [54] demonstrated that off-the-shelf static analysis was insufficient for extracting mappings between API methods and permissions in Android apps. To overcome these limitations, the authors proposed a set of static analysis techniques. Three of these techniques were based on class hierarchy, including string analysis, service identity inversion, and entry point. Two other techniques: ser-

vice redirection and service initialization, leveraged a field-sensitive module, called `Spark`. Evaluation of the proposed methods over 1,421 Android apps revealed that 9% of the apps were accessed more permissions than necessary.

Shen et al. [55] and Holavanalli et al. [56] proposed Flow Permissions, a technique to enrich the Android permission mechanism with flow analysis. Flow Permissions was implemented through `BlueSeal`, a tool that employed static analysis to extract information flow between permission domains within an app or between apps via IPC (Inter Process Communication) mechanisms. The cross-app permission analysis compared information flows within new apps to those of already-installed apps to derive new Flow Permissions. The effectiveness and utility of Flow Permissions was evaluated using a dataset of 2,992 popular Android apps and 1,047 malicious Android apps and a survey of 540 participants. The results indicated that Flow Permissions was practical to deploy, was able to significantly impact user's decisions to install an app and provided visibility into the holistic behavior of mobile apps.

Corla [57] proposed `CHABADA`, a technique for identifying inconsistencies between advertised behavior of Android apps and their implemented behavior. `CHABADA` performed topic modeling on app descriptions and then clustered apps using k-means. Within each cluster, sensitive APIs which were governed by a user permission were identified. Outliers with respect to API usage were detected using One-Class Support Vectors Machine (SVM). These outliers were considered potentially malicious activity. `CHABADA` was applied on a set of 22,500+ Android apps. The prototype was able to detect several anomalies and flag 56% of novel malware.

Xiao et al. [58] proposed a user-aware privacy control mechanism to track how sensitive data was used inside mobile apps. The authors employed static analysis to extract information flows inside apps. The computed information flows were classified into safe and unsafe flows based on a tamper analysis that tracks whether private data is obscured before escaping through output channels. To evaluate the effectiveness of the proposed method, 50 users opinions were collected, out of which 90% considered the method to be effective to protect their privacy. The results also showed that the proposed approach could be used to automatically provide default privacy settings for apps.

Huang et al. [59] introduced `Dflow`, a context-sensitive information flow type system, and `DroidInfer`, the corresponding type inference analysis for detecting privacy leaks in Android apps. `DroidInfer` improved the scalability of `FlowDroid` by avoiding points-to-based static analysis. The proposed approach was a type-based taint analysis technique which explained information flows using context-free language. Evaluating `DroidInfer` on 144 free Android apps showed that it was able to detect privacy leaks in Android apps with a false-positive rate of 15.7%.

Yang et al. [60] introduced `AppContext`, a mechanism that statically analyzed sensitive behaviors of Android apps. `AppContext` extracted the contexts (events and conditions) that triggered security-sensitive behaviors. The authors applied SVM to classify an app as malware or benign based on the extracted contexts of the app. `AppContext` was evaluated on a dataset of 202 malicious and 633 benign Android

apps. The results showed that `AppContext` was able to correctly identify malicious apps with 87.7% precision and 95% recall.

Li et al. [47] presented `ICCTA`, a static taint analysis tool for detecting inter-component privacy leaks in Android apps (paths from sources to sinks). `ICCTA` relied on propagating context information among components. Specifically, `ICCTA` built a control-flow graph of the whole Android application. This allowed propagating the context between Android components and yielding a highly precise data-flow analysis. `ICCTA` adequately modeled the lifecycle and callback methods to detect ICC (Inter-component communication) based privacy leaks. Evaluating `ICCTA` revealed that it outperformed existing tools (e.g., `FlowDroid`) on two popular benchmark datasets, `DroidBench` and `ICC-Bench`. Furthermore, it was able to detect 534 ICC leaks in 108 apps from the `MalGenome` dataset of malware and 2,395 ICC leaks in 337 apps in a set of 15,000 Google Play apps.

Bagheri et al. [61] presented `COVERT`, a tool for compositional analysis of Android inter-app vulnerabilities. The authors utilized static code analysis to automatically model apps while they were installed, removed, or updated. A formal analyzer was used to match the extracted model with a set of vulnerability patterns to detect leakages occurred due to the interaction of apps comprising a system. `COVERT` was examined against privilege escalation, one of the most prominent inter-app vulnerabilities, using a dataset of 500 Android apps. The results showed that it was able to detect inter-app vulnerabilities in several bundles of popular apps.

Avdienko et al. [62] introduced `MUDFLOW`, a tool for detecting malicious apps based on their treatment of sensitive data. `MUDFLOW` was trained on the data flow of benign apps (the most popular 2,866 Google Play apps). For every sensitive data source, the sensitive APIs to which this data flows was determined. This information was then used to automatically flag apps with suspicious features. Specifically, an app's information flow was extracted using `FlowDroid` [23]. `MUDFLOW` tagged an app as malicious if its data flow from sensitive sources was different from benign apps data flow. Applying `MUDFLOW` on a dataset of 25,577 malicious apps and 2,950 benign apps showed that it was able to correctly identify 86.4% of all malware, and 90.1% of malware leaking sensitive data.

Lee et al. [63] presented `HybridDroid`, a static analysis framework for Android hybrid apps, or apps supporting multiple mobile platforms. The proposed framework analyzed inter-communications between Android Java and JavaScript. The authors relied on extensive testing of source code of apps to understand the semantics of interoperation mechanisms of Java and JavaScript and built call graphs for both Java and JavaScript via an on-the-fly pointer analysis. To demonstrate the types of analysis supported by the framework, the authors implemented a bug detector that identified programmer errors due to hybrid semantics, and a taint analyzer that found information leaks across language boundaries. These tools were empirically evaluated using 88 apps. The results provided an evidence on the practicality of the tools and their ability to detect previously uncovered bugs and privacy leaks in Android hybrid apps.

Li et al. [64, 65] proposed `DroidRA`, an instrumentation-based approach to tackle reflection in Android apps by

propagating string constants. The authors applied code instrumentation to replace reflection calls with standard Java calls. *DroidRA* resolved string values in a context- and flow-sensitive manner. *DroidRA* was evaluated on apps from the *Droidbench* benchmark and a dataset of real-world apps. The results showed that the approach was useful in uncovering dangerous code (e.g., sensitive API calls, sensitive data leaks), outperforming several state-of-the-art analyzers.

Meng et al. [66] proposed a semantic model for Android malware detection and classification. The model represented elements of malicious behaviors in malware variants as Deterministic Symbolic Automata with system APIs as transitions. Based on this model, the authors implemented an automatic malware analysis framework. The proposed framework was evaluated on a malware benchmark and a dataset of 223,170 real-world Android apps. The results showed that the framework outperformed several state-of-the-art malware detection approaches and anti-virus tools.

Rahman et al. [67] used 21 static code metrics, commonly used for defect prediction (e.g., bad coding practice and duplication), to assess the security and privacy risks associated with Android apps. The authors applied four statistical learners to predict the level of privacy risk: Radial-Basis SVM (r-SVM), CART (Classification and Regression Trees), k-Nearest Neighbor (KNN), and Random Forests (RF). Evaluation was carried out over 4,416 Android apps. The results showed that r-SVM was able to predict privacy risks level (e.g., high, medium, and low) with 83% precision.

Narayanan et al. [68] introduced *MKLDroid*, a static analysis framework for malware detection and malicious code localization in Android apps. The framework systematically integrated multiple views (malware detection approaches such as sensitive APIs, system calls, and control-flow) of apps using a graph kernel to capture structural and contextual information from apps' dependency graphs. Multiple Kernel Learning (MKL) was then used to find the weighted combination of views which achieved the best detection accuracy. *MKLDroid* was evaluated through large-scale experiments on multiple datasets. The results showed that it outperformed multiple baselines in terms of accuracy while maintaining comparable efficiency. Furthermore, *MKLDroid* was able to identify malice classes with 94% average recall.

Zhang et al. [69] developed *Ripple*, a static reflection analysis that tackled incomplete information environments of Android apps. Undetermined intents, behavior-unknown libraries, unmodeled services, and unresolved built-in containers were considered as incomplete information environments. *Ripple* applied type inference to resolve reflective calls, even if some data flow were missing. *Ripple* was validated over 17 Android apps with incomplete data-flows. The results showed that it was able to discover reflective targets with a low false positive rate of 21.9%.

Shan et al. [70] proposed an approach to support the characterization and detection of self-hiding behaviors in Android apps. The authors grouped these behaviors into three main categories: hiding an app's presence, blocking or removing traces of remote communications, and subverting the device's reminders. A suit of static analysis methods were employed to detect the defined self-hiding behaviors.

Based on the detected behaviors, apps were labeled as benign or malicious. The proposed approach was evaluated on a data set of 9,452 Android apps. The results showed that each malicious app employed 1.5 self-hiding behaviors, on average. Furthermore, the proposed technique attained 87.19% F-measure in detecting malicious apps.

Dilhara et al. [71] presented *ARPDroid*, a technique to repair incompatible use of permissions in Android apps. *ARPDroid* performed static control flow analysis to extract program points where permission usages must be checked. Based on the results of the analysis, *ARPDroid* detected incompatible permission uses. All incompatibilities were then resolved using bytecode transformation. Evaluation of *ARPDroid* over 23 Android apps showed that it achieved 100% detection accuracy and 90% repair success rate.

Garcia et al. [72] introduced *RevealDroid*, a lightweight machine-learning-based approach for detecting malicious Android apps with privacy leaks and identifying their general families. *RevealDroid* leveraged API-based features (e.g., API method invocations), reflective invocations, and the internal behaviors of apps' native binaries (e.g., external calls made by native binaries). The authors used a linear SVM for malware detection and a CART classifier for family identification. *RevealDroid* was evaluated on a dataset of around 54,000 malicious and benign Android apps. The results showed 98% accuracy in detecting malware and 95% in determination of their families.

Canfora et al. [73] proposed *LEILA*, a tool for Android malware families detection and localizing malicious classes into the infected apps. *LEILA* applied model checking, using Milner's Calculus of Communicating Systems, to analyze and verify the Java Bytecode produced when the app's source code was compiled. This gave the tool an advantage in case the code was not available due to obfuscation. *LEILA* was evaluated on a dataset of malware families released from 2011 to 2016. The results showed accuracy levels between 97% and 100%. The authors also devised a set of rules for specifying malicious behaviors of kinds of malware families.

3.4.2 Mixed Analysis Methods

Xu et al. [74] presented *Permylyzer*, a hybrid framework for automatically analyzing the uses of requested permissions in Android apps. The framework used call stack based analysis to identify where and how a permission is used in the app by analyzing the API calls that triggered the permission check. *Permylyzer* was evaluated using 51 malware/spyware families and over 110,000 Android apps. The results showed that *Permylyzer* was able to provide detailed permission use analysis and discover the characteristics of the permission uses in benign and malicious apps.

Hay et al. [75] proposed *IntentDroid*, a comprehensive testing algorithm for Android Inter-Application Communication (IAC) vulnerabilities. IAC enabled the reuse of functionality across apps via message passing, thus it could be used by malicious apps to infect benign apps. The authors started by describing 8 concrete vulnerability types, along with attack scenarios, that could result from the unsafe handling of IAC messages. To test an app, *IntentDroid* monitored any APIs responsible for security-relevant functionality and access to IAC data in order to

recover IAC-relevant app-level behaviors. All the possible attack scenarios were then implemented to detect malicious activities. *IntentDroid* was evaluated over a dataset of 80 Android apps. The result revealed 150 IAC vulnerabilities with a recall rate of 92%.

Keng [76] proposed a hybrid static/dynamic analysis to detect user-triggered causes and paths of privacy leaks in Android apps. The proposed approach used static analysis to capture activity transitional paths to identify data leakage paths. Technically, the author utilized callback control flow analysis to generate an activity transition graph. This allowed automated testing to directly traverse apps towards privacy related activities for verification. The user-triggered causes of privacy leaks were then displayed to users in the form of messages. The proposed approach was evaluated through a 2-week long field user-study with 47 Android users. The results showed that the generated privacy messages were effective for users in managing their privacy.

In a more recent work, Keng et al. [77] introduced *MAMBA*, a new automated app testing technique that was used to extract test cases from control-flow graphs to simulate privacy-sensitive API calls. *MAMBA* initially searched for user events in control-flow graphs of callbacks generated from static analysis of app bytecode. Based on the extracted paths, the tool built test cases using user events that may trigger access to privacy-sensitive data in the apps. *MAMBA* then simulated the execution of the app using the generated test cases to detect any runtime leaks. Evaluation of *MAMBA* on 24 apps showed that *MAMBA*'s graph-aided directed testing was faster than an exhaustive testing approach.

Lee et al. [78] proposed *SEALANT*, a tool for analyzing Android inter-app vulnerabilities. *SEALANT* applied static code analysis to analyze architectural information of a given set of apps from their APKs. The tool then used data-flow analysis and ICC pattern matching to detect vulnerable ICC where inter-app attacks could be launched. Extracted information was then visualized in integrated and compositional representations. Evaluating *SEALANT* over multiple datasets of Android apps showed that it was able to accurately identify vulnerable ICC paths and successfully aid users' understanding of inter-app vulnerabilities.

Zhang and Feng [51] proposed *AndroidLeaker*, a hybrid analysis tool for detecting privacy leaks in Android apps based on taint analysis. *AndroidLeaker* used static analysis (data flow analysis on call graphs) to detect privacy leaks in individual apps and dynamic analysis to prevent any leaks caused by cooperation of multiple apps. The goal was to reduce runtime overhead false alarms often associated with static and dynamic methods. *AndroidLeaker* was tested on 82 examples from *DroidBench*, a test suite released with *FlowDroid*. The results showed lower precision (82%) but higher recall (69%) than *FlowDroid*.

Xu et al. [79] introduced *SpyAware*, a privacy leakage detection framework which examined the correlation between the data flow of privacy leakage and an app's execution path. *SpyAware* used a profiler module to profile app executions in binder calls and system calls. A feature extractor module was implemented to extract feature vectors of suspicious profile. The authors utilized SVM and Naive Bayes (NB) to train and predict *Spyware* executions based on the feature vectors. The proposed framework was eval-

uated over 100 popular Android apps. Experimental results showed 67.7% accuracy in device ID leakage detection and 78.4% accuracy in location-leakage detection.

Demissie et al. [49] introduced Android Wicked Delegation (AWiDe), a dangerous case of delegation in which benign apps unintentionally exposed their access to sensitive resources to attackers (i.e., permissions are re-delegated to a malicious app). The authors formulated this problem as a problem of inadequate input validation of inter-app communication in Android. The authors compared static and dynamic approaches to automatically detect inadequate message validation. The first approach used static taint analysis to detect if the values used in privileged actions depended on potentially malicious data. The second approach used dynamic analysis with automatically generated input values to detect when data during execution was used in privileged actions by malicious apps. The proposed approach was validated on a dataset of 329 Android apps. The results showed that around 15 popular apps were affected by AWiDe vulnerabilities.

Moussa et al. [80] proposed *ACCUSE*, an approach to help users to assess the risk of installing an Android app. *ACCUSE* assigned an app a risk factor based on the Android classification of permissions (normal, dangerous, and system). The authors rescaled the risk factors with the app's rating and the number of downloads. *ACCUSE* visualized the total risk of installing an app before installation. Evaluation of *ACCUSE* on 12,576 Android apps suggested that it outperformed several existing malware detection tools.

Zhao et al. [81] proposed a system to assess whether an app was vulnerable to location inference attacks. Specifically, the authors employed a series of automatic testing mechanisms including UI match (dynamic testing) and API analysis (static analysis) to extract the location information an app provided. These apps were then classified into two categories based on the type of attacks: with and without distance limitations. Evaluating the proposed system on a dataset of 800 android apps revealed that 34.7% of the passing apps are vulnerable location-leak attacks.

3.4.3 Summary and Future Work

Our survey revealed that majority of work under this category of primary studies is focused on introducing new and more effective methods for privacy leak detection. These methods often take the form of a working tool or a prototype that employs static, dynamic, or mixed models of code analysis to detect components of the apps (e.g., code, API calls, GUI components) that might lead to information leakage.

Our survey also revealed that static code analysis methods are the most common, with fewer number of papers applying dynamic and mixed dynamic-static methods. In fact, even methods applying dynamic analysis often apply some sort of static analysis to complement the approach. This can be attributed to the fact that static methods are typically cheaper to execute and can generate more in depth information about data flow within the app's code. However, there seems to be a consensus that dynamic methods that track the execution traces of apps would give a far more detailed view of apps' behavior [57]. Finally, our survey of

papers under this category revealed three main directions of future work. These directions can be identified as follows:

- **Improving accuracy:** the majority of papers suggest improving the proposed methods' accuracy as a main direction of future work. Accuracy, in general, is related to the ability of the proposed method to capture privacy leakage, or identify malware apps, or malicious behaviors with decent levels of precision and recall [52, 60, 63, 79]. Suggestions for improving accuracy involve incorporate dynamic analysis [58, 62, 61, 68, 69, 71], testing the proposed approach on larger datasets of apps or different benchmarks [60], or validating the effectiveness of the results using human studies [49].
- **Coverage:** another common line of future work aims to broaden the scope of research to investigate other types of attack models [81, 75], including inter-app attacks [78], malware families [73], malicious behaviors [68], and anomalies within specific application domains [62].
- **Supporting other platforms:** evaluation in the majority of existing papers is carried over datasets of Android apps. Therefore, supporting other operating systems, mainly iOS, is identified in several studies as a main direction of future work [75, 81, 58, 54].

4 DISCUSSION

In this section, we provide more quantitative analysis results, describe our study limitations, and describe existing related work.

4.1 Quantitative summary

Our survey revealed that privacy leak analysis, using static and dynamic code analysis methods, is the largest category of primary studies. The categories of privacy requirements and policy are relatively comparable in size. However, they have significantly less primary studies than the privacy leak category. The category of user feedback has the smallest number of primary studies. Table 3 summarizes our categories and lists the primary studies classified under each category, and the Bubble chart in Fig 4 shows the growth of these different categories (number of papers) over time.

A common theme that our survey revealed was that the majority of primary studies proposed some sort of a tool or a working prototype to implement their approach or evaluate their findings. We list these tools along with their descriptions in Table 4. Another observation is that, except for few studies on requirements and policy [30, 29], evaluation in majority of the primary studies was carried out over datasets of Android apps. This emphasizes the need for a community-wide effort to prepare benchmark datasets that support iOS apps.

4.2 Limitations

In terms of limitations, we notice that some of our identified categories were more well-defined than others. For example, several of the primary studies under the user feedback category were not primarily intended to detect privacy

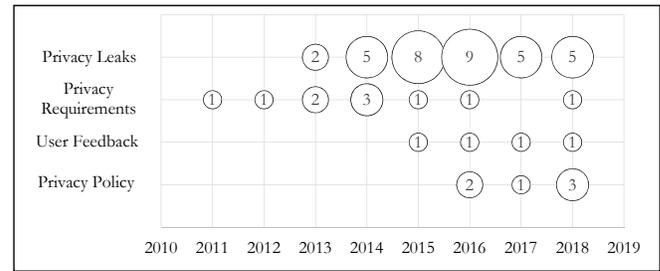


Fig. 4: A Bubble chart of the growth of the different categories over time.

concerns, rather they considered privacy as a user concern that could be extracted from user feedback. Furthermore, there were some studies that could be classified under multiple categories. For example, the primary study in [18] described techniques for extracting privacy requirements from policy. Therefore, it could be classified under both categories. However, since the main focus of the study is deriving privacy requirements, it was summarized under the requirements category.

Other limitations might stem from our inclusion/exclusion protocol. Specifically, we relied on our own assessment of the literature in order to include or exclude papers. To mitigate this threat, we applied a systematic coding protocol, which included individual coding of primary studies and majority voting. We further acknowledge the fact that the quality of summaries might vary across the different categories and sometimes within the same category. We tried to mitigate this threat by imposing a standard structure on our summaries. Finally, our suggested directions of future work, while mainly relied on analyzing the future work sections of our primary studies, were also based on our own assessment of the literature. This might raise some subjectivity concerns as experts in these specific fields might have different views of which specific problems should be perused, or which problems have more value to the research community.

4.3 Related surveys

Since our study takes the form of a systematic survey, other related surveys, or secondary studies, were excluded from our survey. For instance, we identified two mobile app privacy surveys that appeared in SE venues, Sadeghi et al. [82] proposed a taxonomy and qualitative comparison of program analysis techniques for security assessment of Android apps, and Li et al. [7] conducted a systematic literature review of static analysis methods of Android apps.

In general, our mapping study can be distinguished from these existing surveys along two main dimensions. First, we considered the problem from a SE perspective by only considering venues that are commonly known to be SE focused venues, and second, we extended our survey beyond code analysis to include a broader range of SE topics, such as policy, requirements, and user feedback. To that extent, our mapping study bridges an important gap in existing secondary literature on mobile apps privacy.

TABLE 3: Categories and subcategories of primary studies as identified in our survey.

Category	Description	Studies
Privacy Policy	Studies under this category target inconsistencies between apps claimed privacy practices, listed in the privacy policy, and their actual behavior.	[6], [8], [19], [20], [21], [22]
User Feedback	Studies under this category are aimed at mining user privacy concerns from app store reviews.	[29], [30], [31], [32]
Privacy Requirements	Studies under this category are focused on mining, specifying, and enforcing privacy requirements of mobile apps.	[18], [37], [34], [35], [36], [38], [39], [40], [42], [41]
Privacy Leaks	Static Analysis	[47], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73]
	Mixed Methods	[49], [51], [74], [75], [76], [77], [78], [79], [80], [81]

TABLE 4: The list of the tools and working prototype proposed in the primary studies summarized in our survey.

Tool	Category	Brief Description
TAPVerifier [19]	Privacy Policy	An application verification tool that utilizes privacy policies to detect privacy violations.
VisiDroid [6]	Privacy Policy	A visual configuration interface that allows users to configure their privacy preferences.
PPChecker [22]	Privacy Policy	A tool for assessing the trustworthiness of the Android apps' privacy policies.
PPTMA [42]	Privacy Requirements	A service to achieve a balance between user privacy and utility.
Apposcopy [52]	Static Analysis	A semantic-based static taint analysis method to detect malicious apps.
AsDroid [53]	Static Analysis	A privacy leak detector that detects inconsistencies between API invocations and UI.
BlueSeal [55, 56]	Static Analysis	An automatic system to generate Flow Permissions in Android apps.
CHABADA [57]	Static Analysis	A technique to identify inconsistencies between apps' descriptions and their behavior.
Dflow + DroidInfer [59]	Static Analysis	A context-sensitive information flow type system for detecting privacy leaks in Android apps.
AppContext [60]	Static Analysis	A privacy violation detection technique based on the context information.
IccTA [47]	Static Analysis	A static taint analyzer for detecting inter-component privacy leaks.
COVERT [61]	Static Analysis	A compositional analysis tool to detect Android inter-app vulnerabilities.
MUDFLOW [62]	Static Analysis	A static taint analysis tool for distinguishing malicious apps based on their data flow.
HybridDroid [63]	Static Analysis	A static analysis framework for hybrid apps.
DroidRA [64, 65]	Static Analysis	A string inference analysis for resolving reflection.
MKLDroid [68]	Static Analysis	A context-aware, multi-view malware detection tool using graph kernels.
Ripple [69]	Static Analysis	a reflection analysis that tackles incomplete information environments of Android apps.
ARPDroid [71]	Static Analysis	A static control flow technique for detecting incompatible permission uses.
RevealDroid [72]	Static Analysis	A lightweight machine-learning-based approach for detecting malicious apps.
LEILA [73]	Static Analysis	A Java bytecode analyzer exploiting model checking to detect malicious activities.
Premlyzer [74]	Mixed Methods	A hybrid permission analysis tool for Android apps.
IntentDroid [75]	Mixed Methods	A testing approach to detect Android inter-application communication vulnerabilities.
MAMBA [77]	Mixed Methods	A graph-aided directed testing system for the automated checking of privacy behaviors.
SEALANT [78]	Mixed Methods	A tool for automated analysis and visualization of Android inter-app vulnerabilities.
AndroidLeaker [51]	Mixed Methods	A hybrid analysis tool for detecting privacy leaks in Android apps.
SpyAware [79]	Mixed Methods	A privacy leakage detection framework based on data flow and execution path analysis.
ACCUSE [80]	Mixed Methods	An approach to assess apps' installation risks.

5 CONCLUSIONS

We conducted a systematic mapping study of Software Engineering literature on mobile apps privacy. Our objectives were to categorize and summarize the state-of-the-art and enumerate the challenges to be addressed by the Software Engineering research community. Our survey included 54 primary studies published in software engineering venues in the period between 2010 and 2018. These studies were classified, following a systematic coding process, into four main categories: policy, requirements, user feedback, and privacy leak analysis. Our survey revealed that the largest percentage of existing literature tackled problems related to privacy leaks, with less number of papers classified under the other three categories. Primary studies classified under each category were further summarized and their future

work sections were analyzed. The analysis exposed several directions of future work in each of the categories and suggested several areas for improvement.

REFERENCES

- [1] S. Warren and L. Brandeis, "The right to privacy," *Harvard Law Review*, vol. 4, no. 5, pp. 193–220, 1890.
- [2] D. Solove, "Conceptualizing privacy," *California Law Review*, vol. 90, p. 1087, 2002.
- [3] A. Westin and O. Ruebhausen, "Privacy and freedom," *Washington and Lee Law Review*, vol. 25, p. 166, 1968.
- [4] A. Westin, "Social and political dimensions of privacy," *Journal of Social Issues*, vol. 59, no. 2, pp. 431–453, 2003.

- [5] A. van Lamsweerde, "Goal-oriented requirements engineering: A guided tour," in *IEEE Inter. Symposium on Requirements Engineering*, 2001, pp. 249–262.
- [6] A. Aydin, D. Piorkowski, O. Tripp, P. Ferrara, and M. Pistoia, "Visual configuration of mobile privacy policies," in *Inter. Conf. on Fundamental Approaches to Software Engineering*, 2017, pp. 338–355.
- [7] L. Li, T. Bissyandé, M. Papadakis, S. Rasthofer, A. Bartel, D. Oceau, J. Klein, and L. Traon, "Static analysis of Android apps: A systematic literature review," *Information and Software Technology*, vol. 88, pp. 67–95, 2017.
- [8] X. Wang, X. Qin, M. Bokaei, R. Slavin, T. Breaux, and J. Niu, "Guileak: Tracing privacy policy claims on user input data for Android applications," in *Inter. Conf. on Software Engineering*, 2018, pp. 37–47.
- [9] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, Y. Wang, and S. Wilson, "Nudges for privacy and security: Understanding and assisting users' choices online," *ACM Computing Surveys*, vol. 50, no. 3, p. 44, 2017.
- [10] E. Papadopoulos, M. Diamantaris, P. Papadopoulos, T. Petsas, S. Ioannidis, and E. Markatos, "The long-standing privacy debate: Mobile websites vs mobile apps," in *Inter. Conf. on World Wide Web*, 2017, pp. 153–162.
- [11] B. Kitchenham, P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A tertiary study," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009.
- [12] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer, 2012.
- [13] B. Kitchenham, D. Budgen, and P. Brereton, "Using mapping studies as the basis for further research - A participant-observer case study," *Information and Software Technology*, vol. 53, no. 6, pp. 638–651, 2012.
- [14] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Inter. Conf. on Evaluation and Assessment in Software Engineering*, 2008, pp. 68–77.
- [15] B. Kitchenham, P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering a systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009.
- [16] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Inter. Conf. on evaluation and assessment in software engineering*, 2014, p. 38.
- [17] J. Bhatia, T. Breaux, and F. Schaub, "Mining privacy goals from privacy policies using hybridized task re-composition," *ACM Transactions on Software Engineering and Methodology*, vol. 25, no. 3, p. 22, 2016.
- [18] J. Young, "Commitment analysis to operationalize software requirements from privacy policies," *Requirements Engineering*, vol. 16, no. 1, pp. 33–46, 2011.
- [19] L. Yu, X. Luo, C. Qian, S. Wang, and H. Leung, "Enhancing the description-to-behavior fidelity in Android apps with privacy policy," *IEEE Transactions on Software Engineering*, vol. 44, no. 9, pp. 834–854, 2018.
- [20] L. Yu, X. Luo, C. Qian, and S. Wang, "Revisiting the description-to-behavior fidelity in Android applications," in *Inter. Conf. on Software Analysis, Evolution, and Reengineering*, 2016, pp. 415–426.
- [21] R. Slavin, X. Wang, M. Bokaei, J. Hester, R. Krishnan, J. Bhatia, T. Breaux, and J. Niu, "Toward a framework for detecting privacy policy violations in Android application code," in *Inter. Conf. on Software Engineering*, 2016, pp. 25–36.
- [22] L. Yu, X. Luo, J. Chen, H. Zhou, T. Zhang, H. Chang, and H. Leung, "PPChecker: Towards accessing the trustworthiness of Android apps' privacy policies," *IEEE Transactions on Software Engineering*, 2018.
- [23] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Traon, D. Oceau, and P. McDaniel, "Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps," *Acm Sigplan Notices*, vol. 49, no. 6, pp. 259–269, 2014.
- [24] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. Bellovin, and J. Reidenberg, "Automated analysis of privacy requirements for mobile apps," in *AAAI Fall Symposium Series*, 2016, pp. 286–296.
- [25] C. Iacob and R. Harrison, "Retrieving and analyzing mobile apps feature requests from online reviews," in *Mining Software Repositories*, 2013, pp. 41–44.
- [26] N. Jha and A. Mahmoud, "Mining non-functional requirements from app store reviews," *Empirical Software Engineering*, 2019.
- [27] E. Noei, F. Zhang, S. Wang, and Y. Zou, "Towards prioritizing user-related issue reports of mobile applications," *Empirical Software Engineering*, 2019.
- [28] F. Palomba, M. Linares-Vasquez, G. Bavota, R. Oliveto, M. Penta, D. Poshyvanyk, and A. Lucia, "User reviews matter! Tracking crowdsourced reviews to support evolution of successful apps," in *Inter. Conf. on Software Maintenance and Evolution*, 2015, pp. 291–300.
- [29] H. Khalid, E. Shihab, M. Nagappan, and A. Hassan, "What do mobile app users complain about?" *IEEE Software*, vol. 32, no. 3, pp. 70–77, 2015.
- [30] S. McIlroy, N. Ali, H. Khalid, and A. Hassan, "Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews," *Empirical Software Engineering*, vol. 21, no. 3, pp. 1067–1106, 2016.
- [31] A. Ciurumelea, A. Schaufelbühl, S. Panichella, and H. Gall, "Analyzing reviews and code of mobile apps for better release planning," in *Inter. Conf. on Software Analysis, Evolution and Reengineering*, 2017, pp. 91–102.
- [32] G. Scoccia, S. Ruberto, I. Malavolta, M. Autili, and P. Inverardi, "An investigation into Android run-time permissions from the end users' perspective," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2018, pp. 45–55.
- [33] W. Martin, F. Sarro, Y. Jia, Y. Zhang, and M. Harman, "A survey of app store analysis for software engineering," *IEEE Transactions on Software Engineering*, vol. 43, no. 9, pp. 817–847, 2017.
- [34] T. Tun, A. Bandara, B. Price, Y. Yu, C. Haley, I. Omoronyia, and B. Nuseibeh, "Privacy arguments: Analysing selective disclosure requirements for mobile

- applications," in *Inter. Requirements Engineering Conf.*, 2012, pp. 131–140.
- [35] I. Omoronyia, L. Cavallaro, M. Salehie, L. Pasquale, and B. Nuseibeh, "Engineering adaptive privacy: On the role of privacy awareness requirements," in *Inter. Conf. on Software Engineering*, 2013, pp. 632–641.
- [36] K. Thomas, A. Bandara, B. Price, and B. Nuseibeh, "Distilling privacy requirements for mobile applications," in *Inter. Conf. on Software Engineering*, 2014, pp. 871–882.
- [37] P. Mai, A. Goknil, L. Shar, F. Pastore, L. Briand, and S. Shaame, "Modeling security and privacy requirements: A use case-driven approach," *Information and Software Technology*, vol. 100, pp. 165–182, 2018.
- [38] T. Breaux, H. Hibshi, and A. Rao, "Eddy, a formal language for specifying and analyzing data flow specifications for conflicting privacy requirements," *Requirements Engineering*, vol. 19, no. 3, pp. 281–307, 2014.
- [39] T. Breaux and A. Rao, "Formal analysis of privacy requirements specifications for multi-tier applications," in *Inter. Requirements Engineering Conf.*, 2013, pp. 14–23.
- [40] T. Breaux, D. Smullen, and H. Hibshi, "Detecting repurposing and over-collection in multi-party privacy requirements specifications," in *Inter. Requirements Engineering Conf.*, 2015, pp. 166–175.
- [41] Y. Van Der Sype and W. Maalej, "On lawful disclosure of personal user data: What should app developers do?" in *Inter. Workshop on Requirements Engineering and Law*, 2014, pp. 25–34.
- [42] Y. Liu and A. Simpson, "Privacy-preserving targeted mobile advertising: Requirements, design and a prototype implementation," *Software: Practice and Experience*, vol. 46, no. 12, pp. 1657–1684, 2016.
- [43] K. Kang, S. Cohen, J. Hess, W. Novak, and S. Peterson, "Feature-oriented domain analysis (FODA) feasibility study," Software Engineering Institute, Carnegie Mellon University, Tech. Rep. CMU/SEI-90-TR-21, 1990.
- [44] B. Baixauli, J. Leite, and J. Mylopoulos, "Visual variability analysis for goal models," in *Inter. Requirements Engineering Conf.*, 2004, pp. 198–207.
- [45] E. Yu, "Social modeling and i," *Conceptual Modeling: Foundations and Applications*, pp. 99–121, 2009.
- [46] K. Pohl, G. Böckle, and F. Linden, *Software Product Line Engineering: Foundations, Principles, and Techniques*. Springer, 2005.
- [47] L. Li, A. Bartel, T. Bissyandé, J. Klein, Y. Traon, S. Arzt, S. Rasthofer, E. Bodden, D. Ocateau, and P. McDaniel, "IccTA: Detecting inter-component privacy leaks in Android apps," in *Inter. Conf. on Software Engineering*, 2015, pp. 280–291.
- [48] J. Newsome and D. Song, "Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software," in *Network and Distributed System Security Symposium*, 2005.
- [49] B. Demissie, D. Ghio, M. Ceccato, and A. Avancini, "Identifying Android inter app communication vulnerabilities using static and dynamic analysis," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2016, pp. 255–266.
- [50] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B. Chun, L. Cox, J. Jung, P. McDaniel, and A. Sheth, "Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Transactions on Computer Systems*, vol. 32, no. 2, p. 5, 2014.
- [51] Z. Zhang and X. Feng, "AndroidLeaker: A hybrid checker for collusive leak in Android applications," in *Symposium on Dependable Software Engineering: Theories, Tools, and Applications*, 2017, pp. 164–180.
- [52] Y. Feng, S. Anand, I. Dillig, and A. Aiken, "Apposcopy: Semantics-based detection of Android malware through static analysis," in *Inter. Symposium on Foundations of Software Engineering*, 2014, pp. 576–587.
- [53] J. Huang, X. Zhang, L. Tan, P. Wang, and B. Liang, "Asdroid: Detecting stealthy behaviors in Android applications by user interface and program behavior contradiction," in *Inter. Conf. on Software Engineering*, 2014, pp. 1036–1046.
- [54] A. Bartel, J. Klein, M. Monperrus, and Y. Traon, "Static analysis for extracting permission checks of a large scale framework: The challenges and solutions for analyzing Android," *IEEE Transactions on Software Engineering*, vol. 40, no. 6, pp. 617–632, 2014.
- [55] F. Shen, N. Vishnubhotla, C. Todarka, M. Arora, B. Dhandapani, E. Lehner, S. Ko, and L. Ziarek, "Information flows as a permission mechanism," in *Inter. Conf. on Automated Software Engineering*, 2014, pp. 515–526.
- [56] S. Holavanalli, D. Manuel, V. Nanjundaswamy, B. Rosenberg, F. Shen, S. Ko, and L. Ziarek, "Flow permissions for Android," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2013, pp. 652–657.
- [57] A. Gorla, I. Tavecchia, F. Gross, and A. Zeller, "Checking app behavior against app descriptions," in *Inter. Conf. on Software Engineering*, 2014, pp. 1025–1035.
- [58] X. Xiao, N. Tillmann, M. Fahndrich, J. Halleux, M. Moskal, and T. Xie, "User-aware privacy control via extended static-information-flow analysis," *Automated Software Engineering*, vol. 22, no. 3, pp. 333–366, 2015.
- [59] W. Huang, Y. Dong, A. Milanova, and J. Dolby, "Scalable and precise taint analysis for Android," in *Inter. Symposium on Software Testing and Analysis*, 2015, pp. 106–117.
- [60] W. Yang, X. Xiao, B. Andow, S. Li, T. Xie, and W. Enck, "Appcontext: Differentiating malicious and benign mobile app behaviors using context," in *Inter. Conf. on Software Engineering*, 2015, pp. 303–313.
- [61] H. Bagheri, A. Sadeghi, J. Garcia, and S. Malek, "Covert: Compositional analysis of Android inter-app permission leakage," *IEEE Transactions on Software Engineering*, vol. 41, no. 9, pp. 866–886, 2015.
- [62] V. Avdiienko, K. Kuznetsov, A. Gorla, A. Zeller, S. Arzt, S. Rasthofer, and E. Bodden, "Mining apps for abnormal usage of sensitive data," in *Inter. Conf. on Software Engineering*, 2015, pp. 426–436.
- [63] S. Lee, J. Dolby, and S. Ryu, "Hybridroid: Static analysis framework for Android hybrid applications," in *Inter. Conf. on Automated Software Engineering*, 2016, pp. 250–261.
- [64] L. Li, T. Bissyandé, D. Ocateau, and J. Klein, "Reflection-aware static analysis of Android apps," in *Inter. Conf. on Automated Software Engineering*, 2016, pp. 756–761.
- [65] L. Li, T. Bissyandé, D. Ocateau, and J. Klein, "Droidra: Taming reflection to support whole-program analysis

- of Android apps," in *Inter. Symposium on Software Testing and Analysis*, 2016, pp. 318–329.
- [66] G. Meng, Y. Xue, Z. Xu, Y. Liu, J. Zhang, and A. Narayanan, "Semantic modelling of Android malware for effective malware comprehension, detection, and classification," in *Inter. Symposium on Software Testing and Analysis*, 2016, pp. 306–317.
- [67] A. Rahman, P. Pradhan, A. Partho, and L. Williams, "Predicting Android application security and privacy risk with static code metrics," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2017, pp. 149–153.
- [68] A. Narayanan, M. Chandramohan, L. Chen, and Y. Liu, "A multi-view context-aware approach to Android malware detection and malicious code localization," *Empirical Software Engineering*, vol. 23, no. 3, pp. 1–53, 2018.
- [69] Y. Zhang, Y. Li, T. Tan, and J. Xue, "Ripple: Reflection analysis for Android apps in incomplete information environments," *Software: Practice and Experience*, vol. 48, no. 8, pp. 1419–1437, 2018.
- [70] Z. Shan, I. Neamtiu, and R. Samuel, "Self-hiding behavior in Android apps: Detection and characterization," in *Inter. Conf. on Software Engineering*, 2018, pp. 728–739.
- [71] M. Dilara, H. Cai, and J. Jenkins, "Automated detection and repair of incompatible uses of runtime permissions in Android apps," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2018, pp. 67–71.
- [72] J. Garcia, M. Hammad, and S. Malek, "Lightweight, obfuscation-resilient detection and family identification of Android malware," *ACM Transactions on Software Engineering and Methodology*, vol. 26, no. 3, p. 11, 2018.
- [73] G. Canfora, F. Martinelli, F. Mercaldo, V. Nardone, A. Santone, and C. Visaggio, "Leila: Formal tool for identifying mobile malicious behaviour," *IEEE Transactions on Software Engineering*, 2018.
- [74] W. Xu, F. Zhang, and S. Zhu, "Permylzer: Analyzing permission usage in Android applications," in *Inter. Symposium on Software Reliability Engineering*, 2013, pp. 400–410.
- [75] R. Hay, O. Tripp, and M. Pistoia, "Dynamic detection of inter-application communication vulnerabilities in Android," in *Inter. Symposium on Software Testing and Analysis*, 2015, pp. 118–128.
- [76] J. Keng, "Automated testing and notification of mobile app privacy leak-cause behaviours," in *Inter. Conf. on Automated Software Engineering*, 2016, pp. 880–883.
- [77] J. Keng, L. Jiang, T. Wee, and R. Balan, "Graph-aided directed testing of Android applications for checking runtime privacy behaviours," in *Inter. Workshop on Automation of Software Test*, 2016, pp. 57–63.
- [78] Y. Lee, J. Bang, G. Safi, A. Shahbazian, Y. Zhao, and N. Medvidovic, "A SEALANT for inter-app security holes in Android," in *Inter. Conf. on Software Engineering*, 2017, pp. 312–323.
- [79] H. Xu, Y. Zhou, C. Gao, Y. Kang, and M. Lyu, "Spyaware: Investigating the privacy leakage signatures in app execution traces," in *Inter. Symposium on Software Reliability Engineering*, 2015, pp. 348–358.
- [80] M. Moussa, M. Penta, G. Antoniol, and G. Beltrame, "Accuse: Helping users to minimize Android app privacy concerns," in *Inter. Conf. on Mobile Software Engineering and Systems*, 2017, pp. 144–148.
- [81] F. Zhao, L. Gao, Y. Zhang, Z. Wang, B. Wang, and S. Guo, "You are where you app: An assessment on location privacy of social applications," in *Inter. Symposium on Software Reliability Engineering*, 2018, pp. 236–247.
- [82] A. Sadeghi, H. Bagheri, J. Garcia, and S. Malek, "A taxonomy and qualitative comparison of program analysis techniques for security assessment of Android software," *IEEE Transactions on Software Engineering*, vol. 43, no. 6, pp. 492–530, 2017.



Fahimeh Ebrahimi received the B.S. and M.S. degrees in Information Technology Engineering in 2013 and 2015 from Amirkabir University of Technology (Tehran Polytechnic). She is currently a PhD student of Computer Science and Engineering at Louisiana State University. Her main research interests include requirements engineering, app store analysis, natural language processing, and data mining.



Miroslav Tushev received the B.S. in management and organization in 2013 from the Russian Academy of National Economy and Public Administration. He is currently a PhD student of Computer Science and Engineering at Louisiana State University. His main research interests include open source systems, program comprehension, code navigation, and natural language analysis of software.



Anas Mahmoud received the M.S. and PhD degrees in Computer Science and Engineering in 2009 and 2014 from Mississippi State University. He is currently an assistant professor of Computer Science and Engineering at Louisiana State University. His main research interests include requirements engineering, software testing, program comprehension, code navigation, natural language analysis of software, program refactoring and information foraging.