Survey of Apps, Privacy Leaking and Regulations

Philipp Reindl-Spanner*
philipp.spanner@tum.de
Technical University Munich
Munich, Germany

ABSTRACT

Mobile applications have seen an immense growth over the past decade. During this time the tracking of users through third-party services has become a common part of the mobile ecosystem. Developers often use third-party services to either ease their app development or implement those to maximize their apps' monetization functions. However often the apps' users are unaware of the companies accessing and processing their private data. This results in potential privacy leaking without the users' consent.

In this survey I compare different contemporary research papers to give an overview of the mobile tracking ecosystem, insight in established and new privacy leak detection methods and also a look at the legal issues that arise with privacy leaks and insufficient privacy policies. The symbiosis of these papers demonstrates that with the right choice of tools it is possible to uncover many privacy leaks that have not been discovered so far. Furthermore, combining different research results enabled me to look at the mobile (tracking) ecosystem from a wider angle. Ultimately, I try to find concepts for regulations to improve the current situation for app users.

The results of this paper show that third-party trackers have a big influence on the mobile ecosystem and are a factor in enabling privacy leaking. Another key finding is that by combining different privacy leak detection methods their accuracy can be improved. Unfortunately when looking at inconsistencies in privacy policies my results show that user protection is still insufficient, especially when considering vulnerable audiences.

CCS CONCEPTS

• **Security and privacy** → **Economics of security and privacy**; **Privacy protections**; *Mobile and wireless security*; • **Networks** → *Mobile networks*; • **Applied computing** → Law.

KEYWORDS

apps, trackers, privacy leaks, privacy policies

ACM Reference Format:
Philipp Reindl-Spanner. 2018. Survey of Apps, Privacy Leaking and Regulations. In *Connected Mobility Seminar '19: Apps and Privacy Leaking, Summer '19, Munich, GER.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145

CMS '19, Summer '19, Munich, GER
© 2018 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$0
https://doi.org/10.1145

1 INTRODUCTION

"We work with third-party partners who help us provide and improve our Products or who use Facebook Business Tools to grow their businesses, which makes it possible to operate our companies and provide free services to people around the world [...]"[1]. This statement from Facebook's privacy policy reinforces what seems to be the current trend in (mobile) applications. Developers try to improve user experience, ease app maintenance and maximize app revenue by using third-party services. As the usage of mobile applications has increased drastically over the last decade (Figure 1), new markets for user data have been created. Those new markets include ways to monetize apps and especially make use of targeted advertisement. To increase their revenue developers utilize different ways to collect the app users' private data. This collection of private data becomes problematic when it is not fully disclosed in the apps' privacy policy. This for the users unknown collection of their private data is called privacy leaking.

Herewith, two equally problematic types of privacy leaking can be defined: privacy leaking that is within the knowledge of the developers and privacy leaking that occurs through bad code or other problematic sources unknown to the developers. As both leaking types have the same negative effects for the app user, this work will consider both when talking about privacy leaks.

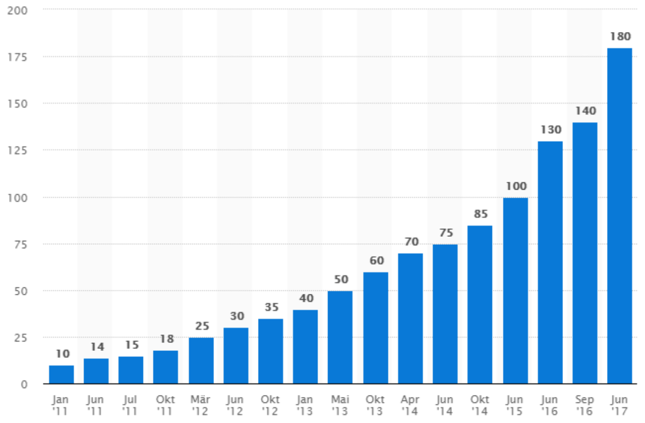

Figure 1: Accumulated number of global downloads from the apple app-store from January 2011 till June 2017 (in billions) [1]

In this paper I will argue for the importance and motivate research in the field of privacy protection in mobile applications by addressing the following three questions:

[1] https://www.facebook.com/about/privacy

(1) How do third-party services influence the mobile tracking ecosystem and in what manner do these connect to privacy leaking?
(2) What are the different possibilities to detect privacy leaks?
(3) What are the main legal issues with privacy leaks and how can user experience be improved?

I will attempt to answer these questions based on four papers as presented in the related works section. In addition I will support or argue the authors' results taking into account various different resources. Regarding the postulated research questions the focus for this paper is to highlight the connections between the different papers.

2 RELATED WORKS

For this survey I have chosen four different research papers that address different problems within the context of apps and privacy leaking. These are not the only sources this paper is based on, but they are the central resources I used for my research.

With more and more third-party services being integrated into mobile applications, many aspects for the developers are simplified. With this development a new ecosystem for third-party tracking has been created. However user experience may suffer through this trend [14]. By using different techniques, especially the Lumen Privacy Monitor[2] the authors of [14] uncover privacy leaks in apps. With their results the authors introduce techniques to identify Advertising and Tracking Services (ATS), which they then use to uncover the parent companies. The data proves the existence of cross-device tracking and other new mechanisms used by third-party services. The collected data set is essential for understanding the global flow of data within the tracking ecosystem and the implications of current and proposed regulations.

To better understand privacy leaking the authors of [3] contribute to the field of privacy leak detection. In this work the authors introduce black-box differential analysis for privacy leak detection. This approach solves the problem of eliminating non-determinism from the network. For this paper the authors developed a tool called AGRIGENTO, which performs root cause analysis of non-determinism in the network behaviour of Android apps. Using this tool the authors show that non-determinism in many cases can be explained and eliminated. This approach gives new insights into how modern apps use custom encoding and obfuscation techniques to leak private user information [3].

With people using different devices simultaneously in different context (mobile, stationary, etc.), online tracking is evolving from tracking browser- and device-tracking to people-tracking. This calls for new methods in the field of privacy leak detection that involve multiple devices to detect the leaks. The authors of [16] introduce new techniques to prove the existence of device tracking and demonstrate how tracking companies use cross-device tracking to improve predictions on app users.

Mobile apps are often non-compliant with what the developers state in the apps' privacy requirements [17]. With their paper the authors of [17] introduce an automated way to check mobile apps' privacy policies. For different sized data sets the authors first checked whether the apps even included a privacy policy. For apps with a privacy policy they then applied their machine-learning algorithm to analyze those. After this step the authors checked whether the apps behaved as stated in the privacy policy [17].

3 THE MOBILE TRACKING ECOSYSTEM

The following subsections will describe numerous important aspects of and influences on the mobile tracking ecosystem. First of all, when looking at mobile applications it is essential to compel multiple domain-specific differentiations. Very important for a wider understanding of the ecosystem is, which purpose third-party services are used for. Considering the global tracking ecosystem, many services not only serve the purpose of advertising but also comprise further interesting use-cases. Moreover I will give insight into the evolution of the tracking ecosystem on the basis of cross-device tracking. The last section will focus on the influence of pricing model and app-store rating in regard to privacy leaking.

3.1 First- and Third-Party Domains

Domains in mobile applications can be divided into first- and third-party domains. According to [14] the main reason for this differentiation is that first-party domains enable essential functionalities to the apps themselves and with that tend to be trusted by users when installing the application on their smartphone. Adding to this conjecture the paper [17] states: "Both[3] have to be analyzed independently as one may be allowed while the other may not." On the contrary, third-party domains may have completely different functionalities from the apps' main functionalities. Still, these domains can collect and transfer private user data. For this reason my work is focused on third-party domains.

A further differentiation has already been indicated by the authors of [14], who introduce two categories of Apps and Tracking Services (ATS)[4] in third-party domains:

- ATS domains
- ATS-capable domains (ATS-C)

This categorization is based on the third-party service business model and observed behaviour. While ATS domain's primary services are advertising and tracking services, an ATS-capable domain's primary service might have another purpose. For instance [17] discovered that 17% of their analyzed apps could be sharing data with third-parties without disclosing so in their policies.

All these differentiations may help understanding the mobile tracking ecosystem and clarify how companies collect and share private user data through third-party services.

3.2 Use-Cases for Third-Party Services

The most prominent use-case for third-party services to users might be advertising. The reason for this is the users' awareness of the directed advertising resulting from the trackers. Advertising also is the only use-case where the user gets a visual clue inside the apps, resulting from the developers' dependency on those services for monetization [14]. Furthermore with rising competition on the mobile application market, developers are always eager to surpass their competitors. Development can easily be accelerated by using

[2] https://www.haystack.mobi/

[3] first- and third-party domains

[4] For an in-depth description please see [14] page 3

third-party services that automatically collect and analyze user behaviour data, generate crash reports or provides social network integration.

Another use-case, user preference prediction through third-party services, can even be improved through usage of methods like cross-device tracking [16]. Cross-device tracking is a technique that enables the tracker to link multiple devices to one unique user. This is not new to the tracking industry, but its importance has risen over the last few years as more and more users use multiple devices on a daily basis. With these multi-device-users (an average user uses at least one mobile- and one web-based device on a daily basis) the objective shifts from device to user tracking.

3.3 Spread of Cross-Device Tracking

In the scope of cross-device tracking it is assumed by the authors of [16] that most cross-device tracking might be applied through Google and Facebook as they have the most profound access to the various devices of their users. Still, they do not name a specific percentage to which a typical Internet user might be tracked unknowingly. Opposing to this the authors of [14] discovered that 39% of all cross-device tracking capable ATSes are present as third-parties in at least one of the Alexa Top 1,000 websites. This concludes that cross-device tracking is already widespread.

For the mobile ecosystem cross-device tracking will become an even bigger factor in the future. The authors of [16] state that cross-device tacking in its current form will just mark the beginning of a larger trend. With the Internet of Things evolving just as rapidly as it does right now, cross-device tracking adds a lot to the inter-connectivity of devices. Unfortunately this also challenges all current user privacy laws and regulations.

3.4 Paid and Free Apps

One component of the mobile ecosystem is the pricing model for apps. As stated above, many developers include third-party services in their "free" apps to monetize them. The motivation for this chapter is that it is common sense that paying for an app increased your privacy protection. This stems from the believe that when choosing to only use the free version of an app, you "pay" with your private user data and the app developer generates money by using advertising directed at your interests.

The results of [13] show that over the sample of free and paid pairs, there is no clear evidence that paying for an app will guarantee protection from extensive data collection. The authors of [14] differentiated the categorization of free and paid apps more precisely. They also introduced the category of free apps with in-app purchases (also called "freemium"). Contrary to [13] the results showed the presence of ATS and ATS-Cs was the highest in free apps with the possibility of in-app purchases. This category is followed by free apps, whereas paid apps appeared to have the least trackers. In addition the authors of [14] state that even when opting out of advertisement through acquiring the paid version of an application many paid apps still included third-party services for other purposes without clear indication. A reason for this might be the re-use of code in the paid and free version of the same application. Because of those results [13] call the benefits of paying for an app (privacy concerning) at best tenuous and likely to mislead consumers.

In conclusion, the common misbelief that the paid or "freemium" version of an app provides increased privacy protection as opposed to a free counterpart offered by the same developer could not be confirmed.

3.5 Rating of Apps

When looking at the differences between app pricing models, another interesting differentiation would be the app store rating of apps. A common perception of users is that apps with high download numbers or better app rating have lower tendencies to leak and/or make use of user data. Interestingly, the authors of [17] discovered that apps with high overall Google Play store scores in fact do not have fewer odds for potential inconsistencies. While the rating itself has no impact on the probability of privacy policy inconsistencies, the results also showed that apps with a high amount of ratings showed a lower likelihood for inconsistencies.

3.6 Influence of Third-Party Services

The described use-cases create the market for third-party services. Overall, the analysis and further use of user data by the developers seems very natural. For example, developers need crash reports from the real world to improve their app. However through the usage of third-party tracking, one problem the authors of [14] came across was that many big tracking and advertising companies do not have strict data-sharing policies. A common issue is that parent companies often claim not to share data with third-party companies but tolerate and even allow data sharing between their subsidiaries. In some cases user data acquired by this method even ended up for sale on data exchange websites without the users' knowledge. This can be considered a data breach [14].

With the tracking industry on the rise, new industries for privacy control tools and anti-tracking services emerge. Nowadays many web users already use different ad-blockers like "AdBlock"[5] while developers recently also created ad-blockers for smartphones. One example for a smartphone application would be the open-source project "Blokada"[6].

Overall it is obvious that privacy leaking is closely connected to the usage of third-party services in mobile applications. In all resources for this paper the authors were able to link privacy leaking to third-party services. To detect such privacy leaks sophisticated methods are necessary.

4 TECHNIQUES TO DETECT PRIVACY LEAKS

Nowadays many mobile applications are closed source or obfuscated, often making static analysis impossible. Even if the source code of an application is available, run-time events and different app configurations often dictate information use [7]. Therefore advanced privacy leak detection techniques are vital.

4.1 Static Analysis

This classic analysis approach - which is inspired by research like [5] and [8] - contains the analysis of the app code —if available— or

[5] 562.985 Users [11]

[6] https://blokada.org/

the program's binaries. While static analysis is highly efficient and provides good scale in many studies, this analysis strategy does not reflect the behaviour of the app during its execution [14]. Studies in the field of mobile privacy leaks that are based on static analysis are presented in [2] or [17]. An example for a static analysis approach would be [13]. In their static analysis phase, they identified third-party libraries by eliminating package names that share the same first two levels as the app package. With this method they were able to reveal which third-party libraries are shared between the compared free and paid apps.

4.2 Dynamic Analysis

Dynamic analysis of app binaries requires running an app in a carefully monitored environment [14]. For this approach the app is closely monitored during its execution. The results indicate how the app behaved during the test. In contrast to static analysis this technique makes it possible to analyze applications which are closed source or where the code is not available. However this analysis type lacks the ability to inspect the network traffic, which means dynamic analysis cannot observe which data is really leaked to third-party trackers.

4.3 Network Traffic Analysis

Network Traffic Analysis extends dynamic analysis. During runtime those methods intercept all network traffic sent over Wi-Fi or the cellular network by the mobile device. A good example for this kind of analysis would be the Lumen Privacy Monitor which was used by [14] for their study. This app works by capturing and analyzing app traffic in user-space. However this app cannot decrypt encrypted traffic and with that cannot reliably identify all privacy leaks.

4.4 Black-Box Differential Analysis

One of the key perquisites for performing a differential analysis is to eliminate any sources of non-determinism, such as random values from random number generators, timing values, system values, encrypted values or network values between different executions [3]. With this, even encrypted traffic can be decrypted and analyzed. However the authors of [3] created a tool named AGRIGENTO which performs black-box differential analysis. This tool operates in two phases:

(1) Network Behaviour Summary Extraction: The app is executed and observed multiple times in order to collect network traces and contextual information.
(2) In this stage the app is run again, this time with changed input. The results are then compared with the results of the first phase. This happens in two steps: Differential analysis and risk analysis.

When running on an app, AGRIGENTO can eliminate most kinds of non-determinism in apps and very precisely uncover privacy leaks. Still, this method comprises several shortcomings. One major deficit of this method might be the amount of time it consumes. This method always requires at least two runs on the application. Compared to the other methods (especially static analysis for large scale studies), this is a major disadvantage. In addition, the authors of [3] had issues with false negatives because of AGRIGENTO's limited code coverage. This means during the analysis the app does not behave as it would in real-world circumstances and potential privacy leaks might not occur in a testing environment. Another problem the creators of AGRIGENTO ran into was the detection over covert channel attacks.

4.5 Cross-Device Tacking

As mentioned above, tracking evolves from user tracking to linking multiple devices to a single user. That is why cross-device tracking usually consists of two tasks. The first task is to uniquely identify a user's device and the second task is to connect the identified devices with the same user. For this reason methods to uncover cross-device tracking involve multiple devices. As this adds another layer of complexity, discovering the presence of cross-device trackers requires a more sophisticated approach. In their paper the authors of [16] used pairs of freshly installed web and mobile-based devices connected to the same router. The web-based devices were used to visit different websites where the presented ads were observed. After two months the mobile devices have been used to search Google for consumer products. Following this step, the authors could observe ads on the desktop PCs that could be linked to products searched on the mobile devices.

4.6 Comparison of Privacy Detection Techniques

Comparing the different presented techniques, static and dynamic analysis either seem to not be precise enough or too complex — for example when the code is not open-source— to get any viable results. Moreover the results of those types of analysis often lack preciseness. However, many authors decide to use static analysis because of its scalability properties. To compensate the lack of preciseness while still being able to do a large scale analysis, the authors of [17] chose a conservative approach for their static code analysis. This means whenever they were unsure if privacy leaking occurred they classified the app in a way as if a privacy leak was present.

As network analysis extends dynamic analysis by the possibility to see all (encrypted or unenecrypted) network traffic, this method depicts a solid base in analyzing mobile app behaviour. An example for this are the authors of [14], as they were able to achieve reliable and consistent results through network traffic analysis by using Lumen.

In conclusion, [3] states: It is still relatively easy for app developers to hide privacy leaks from state-of-the-art tools. When evaluating AGRIGENTO the authors of [3] compared their approach to several state-of-the-art privacy leak detection techniques. One example is ReCon, a multi-platform system for detecting privacy leaks [15]. In comparison to ReCon, when running AGRIGENTO on the same test set of applications AGRIGENTO identified the same apps leaking private data as ReCon. In addition AGRIGENTO detected 49 apps that ReCon indicated as non-leaking applications. When auditing the additional flags the developers of AGRIGENTO could confirm that about 71% of the marked apps did indeed leak private data [3]. In additional comparison to other tools AGRIGENTO showed similar results. Those results show that despite having some shortcomings, obfuscation-resilient privacy leak detection not only works well but also outperforms current state-of-the-art privacy leak detection methods.

The authors of [13] used a combination of static and dynamic analysis for different purposes:

(1) static analysis: to determine the requested permissions and third-party SDKs.
(2) dynamic analysis: to detect sensitive data collected by remote services at the network traffic level.

This approach of combining different techniques seems to be the advisable choice when looking at big data sets. From my point of view an "increasing" complexity approach would be sensible. This means the first group of apps should be analyzed using dynamic and static methods to get a coarse overview over the test. For more detailed work network traffic analysis should be used. Finally for the apps whose traffic, due to the non-determinism in encryption methods, cannot be analyzed by tools like Lumen, the technique developed by [3] can be used.

Another approach several already presented authors chose, is to use one analysis method for the first run on the test set and to use another analysis method to confirm the results. An example for this would be the possibility to use dynamic analysis after static analysis already has been used on the test corpus. With the dynamic analysis privacy leaks can be detected which would have been false negatives during the static code analysis.

In order to discover cross-device tracking different approaches are necessary, since cross-device tracking involves multiple devices that correlate to one unique user. When uncovering cross-device tracking there should be many possibilities to combine single device privacy leak detection techniques with cross-device tracking detection methods to create more sophisticated and reliable methods. One use case would be controlling supposed cross-device tracking traffic by decrypting the traffic and analyzing where ends up (analysis of the datasinks). With this uncovering the companies behind the tracking had higher chances as two devices connected to the same tracking company could be analyzed.

5 PRIVACY POLICIES, LEGAL ISSUES AND REGULATIONS

Up until now it has been a very complex and lengthy process to analyze apps' privacy policies on a greater scale. However, with the introduced technique by [17] this problem seems to be solved. Still, all the possibilities of the new technique have yet to be fully explored. The prevalently occurring legal issue throughout all the underlying work was that most third-party tracking services did not comply with the privacy policies of the apps.

5.1 Self-Regulation and Legal-Regulation

One way to increase regulation of third-party tracking is self-regulation through the platforms (app stores) [2]. Most users use one app store on their device. This strong position for the platforms could be used to actively reduce dishonest use of third-party trackers in mobile applications. Moreover the authors of [2] postulate that further action against third-party trackers might have been held back as Google and Apple both have had a stake in the digital advertising industry. Another problem for self-regulation is that it often leaves room for interpretation. In conclusion, the directives often do not cover all possibilities that companies could use and with that, most of the times companies will use those to their benefit.

For cross-device tracking the authors of [16] discovered that many companies that use cross-device tracking and where self-regulation would be applicable are not transparent about their practices. In their group of controlled cross-device tracking companies over 40% omitted their cross-device tracking activities, despite being members of the Digital Advertising Alliance (DAA). This concerning lack of transparency leads to the believe, that self-regulation does not work in this area or the guidance is not enforced enough.

Furthermore a way to regulate this mobile tracking ecosystem would be legal regulation through laws. At this point in time several directives apply to sensitive user data that is sent over the internet. On example for this is the ePrivacy Directive[7] [6]. This directive forces anyone who collects user data to inform them of the involved third-parties (which includes the receiver of the data) and to request the user's consent. This directive is going to be replaced with the new ePrivacy regulation (ePR)[8]. This regulation has the goal to keep up with the fast pace at which IT-based services are developing and evolving [9]. The proposal of this regulation has several very important key points which have been listed by the EU [9][9]:

(1) Privacy rules will in the future also apply to new players providing electronic communications services;
(2) Stronger rules in such a form that all people and businesses in the EU will enjoy the same level of protection;
(3) Privacy is guaranteed for communications content and meta data;
(4) Once consent is given for communications data to be processed, traditional telecoms operators will have more opportunities to provide additional services and to develop their businesses;
(5) The cookie provision, which has resulted in an overload of consent requests for internet users, will be streamlined;
(6) Bans unsolicited electronic communications by emails, SMS and automated calling machines;
(7) The enforcement of the confidentiality rules in the Regulation will be the responsibility of data protection authorities;

Those are important innovations regarding the web and mobile regulations. Moreover this regulation will introduce penalties of up to € 20 million for noncompliance.

Another big step in the right direction has been the introduction of the GDPR in Europe [10]. Furthermore the authors of [14] add that a major shortcoming of the regulations being proposed is the absence of strong limitations on how harvested data may be shared with subsidiaries and third-party organizations.

Here the problem arises that it is still a grey area which country is responsible for data that origins in one country and ends up on servers located in another country. This special case is called cross-jurisdictional data flow.

[7] Directive 2002/58/EC
[8] the implementation of this regulation is expected to be in 2019 (https://www.eprivacy.eu/en/about-us/news-press/news-detail/article/what-does-the-eprivacy-regulation-mean-for-the-online-industry/)
[9] full proposal: https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications

5.2 Cross-Jurisdictional Data Flow

Cross-jurisdictional data flow is a problem that has yet to be solved. The GDPR is the first step into the right direction to reduce the transfer of private user data from the EU to non-EU countries. However, there will be no difference for companies located in the EU [2].

For identifying and understanding the flow of user-data acquired by third-party services the authors of [14] looked at the locations of Lumen users and the sinks where the data ended up. Looking at Figure 2 a frightening amount of data ended up in the US even though only about a quarter of all the users had been located in the United States. Note that this figure had been produced prior to the GDPR becoming effective. This Figure illustrates how data from countries with strict privacy laws (like Germany and other EU countries) tend to have a big chunk of user data flowing away from their servers. If you look closer at EU-countries 89.27% of data ended up on servers in the United States [14].

A big problem with this behaviour is that countries like the United States and China do have privacy laws that allow companies to easily trade user data they acquire through third-party services. In addition in the case of data loss punishments for the companies are almost non-existent. With the GDPR it is still to prove, that those flows of user data away from the EU have been reduced.

5.3 Privacy Policies

Regarding Privacy Policies a huge milestone has been achieved by [17] in creating a new method to automate privacy policy analysis. With their machine learning approach, they managed to check a set of 17,991 Android apps of which only 9,295 had a privacy policy. Comparing their approach to state-of-the-art methods it became obvious that the speedup through an automated system could be enormous. For example, 26 data protection agencies needed one week to analyze the privacy policies of 1,211 apps [12].

Concerning users' privacy the results for this subsection tend to be rather unsatisfying. As stated above only about 51% [17] of the analyzed apps had a link to their privacy policies in the app store. They also identified that many apps (46%) lack a notification for privacy policy updates.

The results as presented in [13] are very similar. The authors' results showed, that only 45% of the paid and free app pairs in their test corpus provided a privacy policy. Another concerning discovery was that less than 1% of the pairs had policies that differed between the free and paid apps.

With their machine learning approach the authors of [17] were able to uncover potential app privacy policy inconsistencies on a large scale. The most prominent examples here are the collection of device IDs (50%), leaking locations (41%) and sharing device Ids (63%). These results show that privacy policy inconsistencies in apps are a serious problem. It seems like many of the privacy policies fail to fulfill privacy requirements [17]. This behaviour can seriously harm the company-customer trust relationship when users become aware.

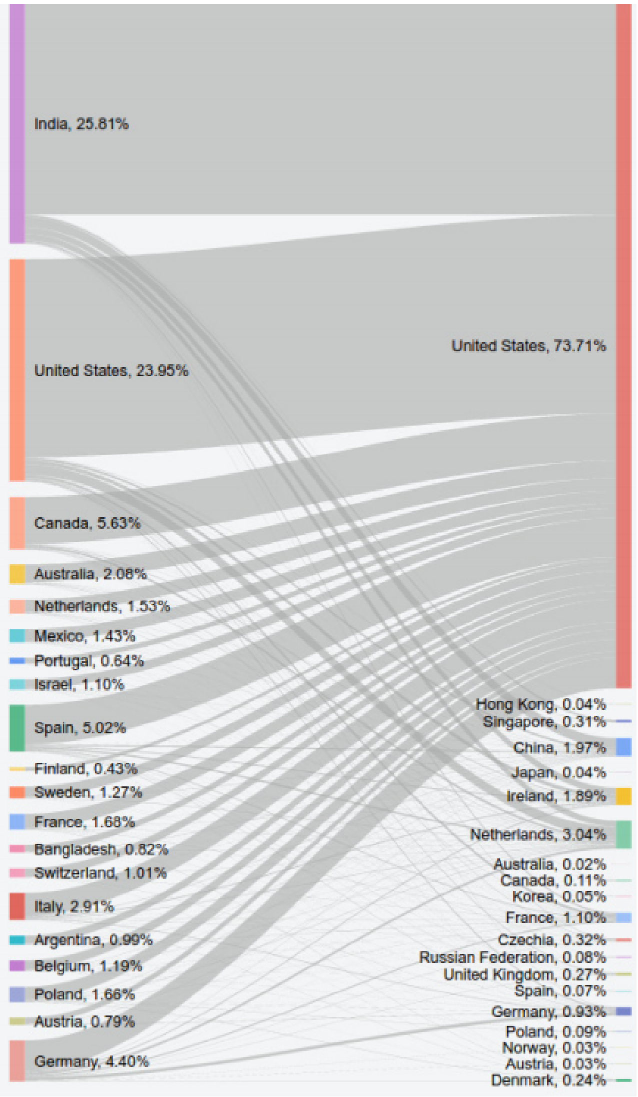

Figure 2: Interactions observed between the 20 most common locations of Lumen users and ATS server locations. Percentages indicate the fraction of flows originating (or, terminating)at the corresponding country. [14]

5.4 Vulnerable Audiences

For the protection of vulnerable audiences, such as underage children[10] regulations have been formed in the "Children's Online Privacy Protection Act" of 1998 (COPPA) [4]. This Act was designed to improve children's safety by making it a requirement for operators of websites or online services to obtain consent of a guardian. Adding to the COPPA profiling and marketing children was also targeted by the GDPR. Here it states that companies should "refrain from profiling them for marketing purposes" [10].

Not less concerning are the findings of [2]. After identifying distinct trackers in apps, they grouped the app genres into different

[10] COPPA defines a Child as an individual under the age of 13

"super-genres". This enabled them to give a high level analysis of distinct trackers by genre. The results showed that "News" and "Family" apps comprise the highest median number of tracker companies associated with them. In numbers an average of 7 distinct third-party trackers is used in "Family" genre apps ("News" genre apps comprise even more trackers). Unfortunately the authors of [2] do not provide information regarding the kind of trackers used in those apps. Nonetheless all the apps in the "Family" genre should be COPPA compliant which means users (and their guardians) should be aware of the trackers.

In summary, the overall consent on how trackers and third-party services are used in connection with children seems more than concerning.

5.5 How to Improve User-Experience

One first step for companies to gain their users' trust is to reduce the overall opacity of their third-party tracking infrastructure. In addition companies should adapt their privacy policies in such way that they comply with regulatory guidelines like the COPPA or GDPR. Another big factor in the third-party tracking ecosystem is, as stated above, self-regulation. Companies should have an increased interest in being as transparent as possible with their tracking behaviour. Also stricter regulations through the app stores are conceivable. An example here would be that every offered app —independent from pricing model or audience— has to have an up-to-date privacy policy. Also sanctions for developers that violate the rules are possible. These could be monetary sanctions or a system where the offered app is banned from the app store for a set amount of time and can only be re-listed after inspection by the app store.

6 DISCUSSION

This paper is oriented around the three questions I phrased in the introduction. The first question of this survey is to address the influence of third-party services on the mobile tracking ecosystem and how privacy leaking is connected to those services. The results show that the most prominent use case for third-party services and trackers is the creation of revenue for the app developer. This can be achieved through a pricing model that includes the user paying for the application he wants to use, or involving third-party services that apply advertisement directed towards the users' preferences in the app itself. This inclusion of third-party services enables the possibility of privacy leaking. The common misbelief that paid apps do not make use of trackers as much as free apps could be disproved. In addition the usage of third-party services influences the mobile ecosystem in such way, that new markets for privacy protection and tracking countermeasures are created.

The second very important discovery of this study is that combining different privacy leak detection methods would be a possibility to improve the overall effectiveness of uncovering privacy leaks. In general this is the result of [3]. With their new obfuscation resilient approach a big improvement has been made in the area of privacy leak detection. Despite the shortcomings this highly precise approach seems to add a much needed component to the variety of privacy detection methods. In addition contributions like [16] show us the importance of looking very closely at the evolution of the mobile tracking ecosystem. Using the example of cross-device tracking it is possible to see the shift from tracking single devices or users to linking several devices to the same unique person.

The last question I tried to answer is, which legal issues exist and how the user experience regarding privacy policy inconsistencies can be improved. Still, a very big concern is the current lack of child protection. Despite many good approaches and new regulations the results show that vulnerable audiences are still not protected in an acceptable manner. Even thought in the last few years many innovations have been made the results show that the advertising and tracking infrastructure nowadays is still a big issue for app users' privacy. Based on the outcome of the present study it is safe to say that the current way users' privacy is treated is insufficient.

7 CONCLUSION

For this paper an analysis of the data sets, which for most papers have been published by the authors, would have gone beyond the scope of this project. Therefore this survey presents a study of different aspects within the mobile tracking ecosystem. The main goal was to create a symbiosis of recent research within this field. The results clearly show that the current regulations are insufficient and there is still room for improvement. The GDPR in 2018 seems like a good step in the right direction. Also the new ePrivacy regulation is expected to improve the users' overall situation. Furthermore I was able to link third-party services to privacy leaking and show ways to improve state-of-the-art privacy leak detection tools through joint application.

The indicated problems and deficits motivate further research in different fields of privacy protection. Especially current research in the field of privacy leak detection seems promising and can be used to further investigate the currently still opaque third-party tracking ecosystem. Unfortunately with more sophisticated techniques to uncover privacy leaks, companies will likely develop new methods to hide their user exploitation.

In general it can be concluded through the results of this paper that the advertising and tracking ecosystem will keep growing within the next years.

8 FUTURE WORK

My survey accentuates the need for further research in the uncovering of privacy leaking. Most authors as cited in this paper exclusively analyzed apps from the Google App Store running on Android phones. Very interesting to see in the future would be whether new methods as presented in [3] are also applicable to iOS-based apps. Right now it still seems very complex to decompile and analyze iOS apps for static analysis [17].

Adding to the detection of privacy leaks, making obfuscation resilient black box analysis viable for large scale analysis would be a major improvement for the method. This could be achieved through parallel execution on multiple devices [3].

Very interesting for future directions would be looking into the changes of the cross-jurisdictional data flow after the GDPR became effective, especially the flow of user data originating in the EU. As already stated above, figure 2 was created before the GDPR became effective.

Concerning the privacy of users, extending [17] looks very promising. Future research should aim at improving their machine learning algorithm and applying different kinds of privacy leak discovery methods in their app analysis.

REFERENCES

[1] Apple. June 5th 2017. Kumulierte Anzahl der weltweit heruntergeladenen Apps aus dem Apple App Store von Januar 2011 bis Juni 2017 (in Milliarden) [Chart]. https://de-statista-com.eaccess.ub.tum.de/statistik/daten/studie/20149/umfrage/anzahl-der-getaetigten-downloads-aus-dem-apple-app-store

[2] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third party tracking in the mobile ecosystem. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 23–31.

[3] Andrea Continella, Yanick Fratantonio, Martina Lindorfer, Alessandro Puccetti, Ali Zand, Christopher Kruegel, and Giovanni Vigna. 2017. Obfuscation-Resilient Privacy Leak Detection for Mobile Apps Through Differential Analysis.. In *NDSS*.

[4] COPPA. 1998. COPPA. https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

[5] Patrick Cousot and Radhia Cousot. 1977. Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *Proceedings of the 4th ACM SIGACT-SIGPLAN symposium on Principles of programming languages*. ACM, 238–252.

[6] E-Privacy Directive. 2002. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). *Official Journal L* 201, 31 (2002), 07. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML

[7] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. 2014. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* 32, 2 (2014), 5.

[8] Michael D Ernst. 2003. Static and dynamic analysis: Synergy and duality. In *WODA 2003: ICSE Workshop on Dynamic Analysis*. New Mexico State University Portland, OR, 24–27.

[9] EU. 2019. Proposal for an ePrivacy Regulation. https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation

[10] GDPR. 2019. GDPR Portal: Site Overview. https://eugdpr.org/

[11] Google. 2019. Chrome Web Store AdBlocker. https://chrome.google.com/webstore/detail/adblock/dgpfeomibahlpbobpnjpcobpechebadh?hl=de

[12] GPEN. Mar 2015. 2014 annual report. https://www.privacyenforcement.net/node/513

[13] Catherine Han, Irwin Reyes, Amit Elazari Bar On, Joel Reardon, Álvaro Feal, Serge Egelman, and Narseo Vallina-Rodriguez. 2019. Do You Get What You Pay For? Comparing The Privacy Behaviors of Free vs. Paid Apps. (2019).

[14] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. 2018. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. (2018).

[15] Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David Choffnes. 2016. Recon: Revealing and controlling pii leaks in mobile network traffic. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 361–374.

[16] Sebastian Zimmeck, Jie S Li, Hyungtae Kim, Steven M Bellovin, and Tony Jebara. 2017. A privacy analysis of cross-device tracking. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 1391–1408.

[17] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven Bellovin, and Joel Reidenberg. 2016. Automated analysis of privacy requirements for mobile apps. In *2016 AAAI Fall Symposium Series*.