Did app privacy improve after GDPR?

Nurul Momen, Majid Hatamian (IEEE Member), Lothar Fritsch

Abstract—What are the effects of the GDPR on consumer apps? This article presents an analysis of app behavior before and after the regulatory change in data protection in Europe. Based on long-term data collection, we present differences in app permission use and expressed user concerns and discuss their implications. In May 2018, the General Data Protection Regulation (GDPR) changed the data protection obligations of the information industry with the European Union users substantially. One should expect to find changes in code, program behavior and data collection activities. To investigate this expectation, we analyzed data about Android apps request and use of permissions to access sensitive group of data on smartphones, and collected user reviews. Our data shows an overall reduction of both permissions used and of expressed user concern. However, in some areas apps have increased access or user complaints while in addition, many apps carry with them several unused access privileges.

Index Terms—Apps, data protection, GDPR, information privacy, survey.

1 Introduction

In May 2018, stronger regulation of the processing of personal data became law in the European Union, known as the General Data Protection Regulation (GDPR) [1]. The expected effect of the regulation was better protection of personal data, increased transparency of collection and processing, and stronger intervention rights of data subjects, with some authors claiming that GDPR will change the world, or at least that of data protection regulation [2]. The GDPR had a two-year (2016–2018) implementation period that followed four years of preparation. At the time of this writing, in November 2019, one and a half year have passed since the implementation of GDPR.

Has GDPR had an effect on consumer software, then? Has the world of code changed, too? Did the GDPR have a measurable effect on mobile apps behavior? How should such a change in behavior be measured?

In our study, we decided to use two indicators for measurement: Android *dangerous permission* [3] privileges and user feedback from the Google Play app market. We collected data from smartphones with an installed app set for months before GDPR implementation on May 25, 2018, and months after that date. Both Fig. 1 and Fig. 2 show how the data collection was organized. The set of 50 apps in our observation is listed in Fig. 3.

We ran a long-term data collection about Android apps use of permission privileges through the Android operating system permissions mechanism. We focused on the so-called *dangerous permissions* [3], a group of access privileges defined as sensitive by Android developers, as they may have an effect on the users sensitive data that regulate access to location, contacts, phone log, sensors, and other data sources. We monitored app permission access request data in March 2017. To compare, we installed a subset of the post-GDPR version of the respective apps in December 2018 and ran a one-week data collection campaign as highlighted in both Fig. 1 and Fig. 2. The data collection was done with the *A-ware* data capture tool described in [4], [5]. The data is stored in an online collection database [6].

Figure 2 shows how we collected three different types of data from the Google Play server, from the app manifest and from observing apps at run time.

Permissions are Androids' access control mechanisms that regulate an apps' access to various system resources. To retrieve data protected by dangerous permissions, the app code contains a declaration of the permissions requested by the app programmers in its *manifesto*. Upon the apps' first presentation to the operating system, the user of the device is prompted to confirm the apps' permission request. If the user consents, the app stores this granted permission and then can use it without user interaction to access resources. Apps do not face any restriction in terms of the period, frequency or amount of data extracted, except for the latest version (Android 9.0—Pie) that has introduced some granular conditions for permission access recently. The permission declaration in the app is, therefore, an indication of the apps' likely data access (however there may be effects that obfuscate this interpretation, see Section 1.3). To evaluate the extent of access, we logged and archived app permission use at the operating system level with the *A-ware* logging tool. The logs were collected for further analysis. From these logs, we obtained data about the apps' actual usage of its permissions.

User feedback is the second indicator of app change. Such feedback sometimes comprises valuable information regarding different aspects of apps ranging from user-friendliness to privacy aspects. Thus, based on our previous study [7], we collected 121,991 and 130,065 user reviews of the 50 most downloaded apps within five app sets on the Google Play app market for the pre- and post-GDPR period, respectively. We then applied natural language processing (NLP) and

- *N. Momen and L. Fritsch work at the Department of Mathematics and Computer Science, Karlstad University, Sweden. E-mail: {nurul.momen,lothar.fritsch}@kau.se*
- *M. Hatamian works at the Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Germany. E-mail: majid.hatamian.h@ieee.org*

This paper has been accepted in IEEE Security & Privacy. Published version here: https://ieeexplore.ieee.org/document/8845749, DOI: 10.1109/MSEC.2019.2938445

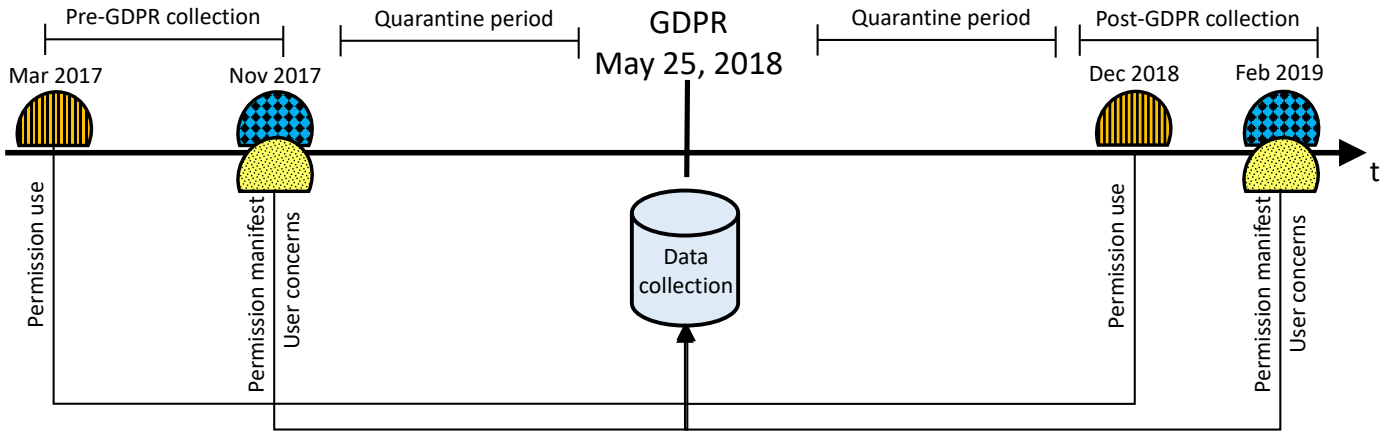


Fig. 1. Data was collected with an approximate 6-months quarantine before and after GDPR implementation. App manifest data (list of required permissions) and user concern data (retrieved from Google Plays' server) were collected after app permission use profiling. Quarantining ensured that app producers had time to adapt apps to GDPR between the two data collection periods.

machine learning (ML) techniques to collected user reviews and our approach was focused on vocabulary referring to privacy threats. The resulting set of threats for the two data collection periods is an indicator of end users' concerns and experiences before and after the adoption of the GDPR.

Before we focus on the data analysis, we discuss the expected impact of the GDPR on apps and their permission use.

1.1 Permissions and the GDPR

One would expect that many of the aspects of GDPR are now implemented and this should show in the app code. Such expectation should be reflected in the software because of the enforcement of expensive violation sanctions. In particular, mobile apps have been known to extract large amounts of personal data [8], [9]. In theory, the observed data collection and processing behavior of apps should have adapted to the GDPR either (or both) by improved privacy statements and consent collection interfaces or through software updates that changed functionality. We target functionality when we measure permission usage.

The GDPR places various requirements on personal data collection and processing. Some of these requirements are discussed below.

The *principle of purpose specification* requires any processing of personal data to be bound to a declared purpose. Article 5-1(b) of the GDPR states that personal data shall be *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ("purpose limitation") [1].* The choice of requested permissions should, therefore, correlate to the app functionality and the privacy policy. Moreover, the use of permission should be bound to its propose—for example, the `MICROPHONE` permission can be used for purpose *A* only, not for purpose *B*.

The *principle of data minimization* requires personal data processing to be reduced to the minimum amount necessary to fulfill the app purpose. Article 5-1(c) of the GDPR states that personal data shall be *adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ("data minimization") [1]*. Thus, the use of permissions should be restricted to the minimum needed to deliver a transaction with the app.

The *principle of transparency* requires all personal data processing to be clearly transparent to the data subject. Article 5-1(a) of the GDPR states that personal data shall be *processed lawfully, fairly and in a transparent manner in relation to the data subject ("lawfulness, fairness, and transparency") [1]*. For permissions, transparency means that information about the kind of data accessed, the frequency and the amount of data extracted should be available. Currently, only one-stop consent for any amount of permission-based access is provided.

Data protection by design is a principle that requires apps to respect privacy from the start, with safe configurations and minimum necessary data processing. Article 25-2 of the GDPR highlights the requirement: *the controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individuals' intervention to an indefinite number of natural persons [1].* The implications for permission use are that the minimum number of permissions should be requested and a lower number of actual permissions should be used in such ways that they reduce the loss of personal data by default.

Data Protection Impact Assessment, which is mandatory for high-risk or high-magnitude data processing or handling of sensitive personal data may uncover risks for data subject privacy. Article 35-7(a–d) of the GDPR elaborates on the criteria: *the assessment shall contain at least: (a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller; (b) an assessment of the necessity and proportionality of the processing operations in*

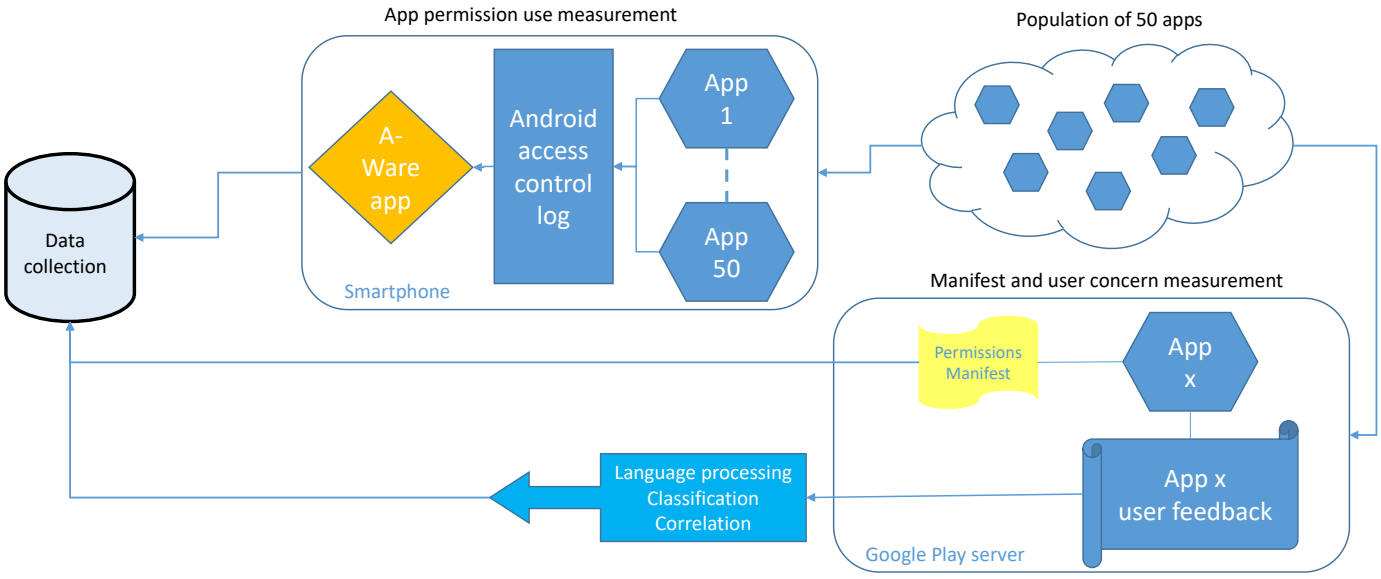


Fig. 2. An overview of the data collection methods used. App manifest data and user concern data were retrieved from Google Play's server, respectively extracted from app manifests and user feedback forum. App use data were observed with the *A-Ware* tool installed on smartphones.

relation to the purposes; (c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1 (Art. 35-1); and (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned [1]. GDPR compliance activities may uncover such risks materializing as a consequence of excessively broad or deep permission use in apps. Consequently, risk reduction will decrease the number of permissions consumed and shown.

Freely given and unambiguous data subject consent is a precondition for lawful data processing. Article 7-2 of the GDPR emphasizes this aspect: *If the data subjects' consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding [1].* Freely given consent would require permission to access data to be bound to a declared purpose, confirmed by the data subject through consent. Therefore, one can expect post-GDPR apps to change behavior so that permissions will not be confirmed in bulk from the start, but in more selective and interactive ways.

The *right to withdraw consent* is also an important right for implementing individual privacy preferences. Historically, an app would just show permissions with a binary (accept/decline) consent collection form to the data subject which was addressed in newer Android versions (6.0–9.0). Now the right to withdraw consent is preserved in these versions and therefore they comply with Article 7-3: *The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent [1].* However, the spirit of freely given and revocable consent is yet to be fully captured, because there is no permission usage monitoring mechanism in the user interface. Such unavailability could influence individual decisions on consents given earlier. Thus, permission usage monitoring efforts can potentially support informed decision-making.

Software changes in reaction to the new regulation are hard to predict, though. While it has recently been argued that the inertia of regulation adaption that is caused by national exceptions and local adaption of EU data protection traditions predating the GDPR [2], we believe that the threat of fines and public exposure actually has created a momentum to improve software, at least when provided by professional software firms or service providers. However, another argument is presented in [10]: the reaction of software vendors to regulation must not always be in support of regulation. As software can be updated very quickly, and standard practices can be established quicker than new regulations can be enforced, there is a risk that software may intentionally use loopholes, camouflage its compliance with regulation, and pursue its own agenda.

1.2 Expecting changes in apps

Based on our discussion of potential GDPR effects on software behavior, we formulate these hypotheses:

- H1 Reduced permission declaration: Code should have been cleaned up and data collection minimized, which should be visible in a reduced number of dangerous permission declared.
- H2 Reduced permission use: A lower permission use frequency should occur. As the user provides consent in a granular fashion, permissions should be used more selectively, based on apps' real functionality.

H3 Reduced user concern: User feedback should consequently show reduced worries about privacy threats.

Before we proceed to the analysis of the collected data, we discuss the constraints and limitations of our data collection project.

1.3 Limitations of chosen approach

Several difficulties arise when interpreting the intention and the relevance of permission use in apps. While our captured data measures static permission declaration and actual use during run-time, it will be difficult to estimate the reasons for and intentions behind the particular permission use. Therefore, not all permission uses may relate to an actual privacy risk.

Simple programming issues, such as re-used code asking for too many permissions or code that has not been cleaned up from testing could be one issue that affects our measurements. Program libraries have been observed as a possible source for excessive use of permissions. In [11], a growing consumption of permissions is identified that is imported through advertising libraries. The intentions of the app programmers and the library programmers may diverge in some cases. For the interpretation of permission use before and after the GDPR, we can only speculate about the possibility that permission-hungry libraries had actually been removed from the app code, since there is not an updated survey for [11] available. In addition, our data was collected on idle smartphones that were not actively used, thus we measured only the app activity that took place without user interaction. In later experiments with app interaction, we got the impression that some apps will collect approval for permission use after a certain time period or if certain conditions occur. Possibly, permissions have not yet been asked for by idling apps.

We point out that while we measured apps showing their permissions credentials to the operating system, we did not observe whether they access or collect data. Neither did we consider personal information extraction, transfer to connected online servers or processing done at such servers.

There are practical limitations to data captured with respect to the control of context variables. Many apps are highly interactive. Social media and news-related apps, as well as advertising, are controlled by external activities. Apps may be updated during data collection, thus automatic updating has to be turned off during the measurement. However, Google Play enforces updates after a while, which limits measurement campaigns to shorter period. Our app sample was for the sake of stable conditions installed to the devices, activated and if necessary personalized with an artificial account. The apps were then left as they were and measured with no user interaction. Through this, we aimed to exclude interaction bias and the influence of social network activities, which both may change app behavior.

For user review analysis, we implemented and followed a machine learning-based scheme to mine user reviews for finding privacy relevant complaints. It is important to note that, the overall usefulness of such a technique is dependent on both (1) how well it understands the user reviews and (2) how truthful those user reviews are. Our approach mainly covers the former, and still, there is little discussion of the accuracy, veracity, and clarity of the reviews because this is generally out of our control.

We would like to highlight that performing a comprehensive app privacy analysis for both the pre- and post-GDPR periods is not an easy task. This is mainly because of the lack of historic app data (manifest, permission usage and user reviews) which are not fully available for the pre-GDPR period. However, we were able to partially correlate our app data sets obtained from separate studies conducted by two research groups (in Sweden and Germany). Nevertheless, we would like to emphasize that there is a mismatch between the studied app categories in user reviews analysis and app permission usage analysis (only *Health & Fitness* and *Music & Audio* app categories overlapping).

Reproducibility is difficult in this setting. Apps get updated, often several times per week. Most apps receive messages from their background services that impact their behavior. Thus it will be challenging to re-create the exact same test setting.

The app sets composition will have an impact on the result. The current app set is mainly taken from the globally most downloaded apps on the Google Play app market. Those apps reflect a global, English-speaking consumer community with an expected over representation in English-speaking countries. Therefore, the data may contain an implicit bias towards non-EU privacy regulation and attitudes. A focused data set representing apps from European vendors, available in EU languages may show different results. We also acknowledge that the chosen app set for permission analysis is a very small subset of the millions of apps available in various app markets. However, it is plausible to assume that choosing globally popular apps allows correlation of privacy expectations of a vast user base, and with industry-standard practice.

2 Changes in Permission Declaration and Usage Patterns

App developers are required to define permissions needed for their apps to function in the app permission manifest. Based on the permission request level [12], users have to consent to data access by granting or rejecting permission requests. Therefore, our study looks for the changes in the developers permission declarations before and after the adoption of the GDPR. Previous studies showed that app developers usually request more permissions than needed for the proper functionality of their apps [4], [8], [9], [13]. The number of declared permissions has historically risen over time [11]. We can examine hypothesis H1 by comparing the extent to which the GDPR had a measurable effect on the permissions requested in the app manifest.

2.1 Permissions: Manifest Changes

This subsection investigates hypothesis H1. We collected the permission manifest for the 50 most-used apps across five app categories on the Google Play app market in November 2017. Our analysis mainly focuses on those permissions defined by Android as dangerous permission requests. Figure 3a visualizes the results for both the pre- and post-GDPR

(a) Case I: Manifest analysis

(b) Case II: Permission usage analysis

Fig. 3. A visualization of pre-post-GDPR differences in permission requests and permission access pattern during run-time. Case I (left): Analysis of permissions requested in manifest (Green = reduced declaration, Red = additional declaration, White = no change from unused, Grey = no change from declared.) and Case II (right): Comparison of run-time permission access pattern (shades of green: percentage of reduced access & shades of red: percentage of increased access). + and − symbols (left) and percent values (right) were added to support viewers with achromatopsia. Reading example: App `Accuweather` (row 19). Case I: `CALENDAR` permission removed (green, −). `MICROPHONE` permission added (red, +), `PHONE` permission removed (green, −). Case II: Number of `LOCATION` permissions showed: 100 percent less (green, −100), `SENSORS` permission increased by 10% (light red, +10), `SMS` permission showing reduced by 6% (light green, −6).

periods. The quantification of app permission requests is based on a three-level indicator.

A reduction of dangerous permissions declarations can be observed. In total, we observe fewer dangerous permission requests in 44 incidents. For an overview, Fig. 3a shows how individual apps have increased and decreased their permission declaration. Moreover, this was also confirmed in total changes shown in Fig. 4a where the reduction in permission declaration is shown as blue columns. An app category that has the highest reduction bar has a lower data access potential, and thus be considered more privacy-friendly. All app groups show a reduction of permission declaration after the adoption of the GDPR.

With respect to category, visualization of change in permission declaration is shown in Fig. 3a which depicts that weather forecasting-related apps represent a significant improvement in terms of requesting less dangerous permissions, namely `PHONE` (6 incidents), `CONTACTS` (5 incidents) and `CALENDAR` (4 incidents). Interestingly, this was also confirmed in permission usage changes shown in Fig. 3b. After *Weather* apps, both Fig. 4a and Fig. 4b confirm that *Music* apps dominate the second best privacy performance. Although the privacy performance of *Fitness* apps is better than *Social* and *Communication* apps regarding the permission manifest changes, all the three categories were assigned similar delta values for permission usage changes.

Our analysis shows that apps are less permission-hungry when it comes to requesting permissions `CALENDAR`, `SMS`, `PHONE`, `CONTACTS` and `LOCATION`. However, some of the permissions were requested more excessively in post-GDPR period: `CAMERA` (4 incidents), `MICROPHONE` (3 incidents) and `SENSORS` (1 incident).

2.2 Permissions: Changes in use

In order to investigate hypothesis H2, we monitored apps' use of *dangerous permissions* at run time. The first phase of the data collection campaign was in March 2017, which is referred to as the pre-GDPR period. We carried out the second phase in December 2018 and it represents the post-GDPR period. In this subsection, we discuss the changes and other notable observations between the two collected data sets.

In the pre-GDPR phase, apps showed dangerous permissions more frequently while idling than would be expected [5], [14]. We monitored several sets of popular apps from different categories and their permission access patterns while keeping the devices idle (without user interaction). The apps were installed on several devices that had a pre-configured prototype app *(A-ware)* installed to record API access from the operating system. The collected logs were then analyzed to determine apps permission access patterns. Recently, we repeated the same experiment with a post-GDPR app set installation.

The post-GDPR phase faced several obstacles; including keeping the catalyst parameters for data collection intact and finding an app-set that matches with the corresponding data sets from independently conducted studies in two different geographic locations (Sweden and Germany). In order to reproduce the earlier collection context and to keep the other influencing parameters constant, none of the apps were actively used after installation. We collected logs of their permission access patterns. Figure 3 shows a correlation visualization of the permission data. Figure 3b shows the comparison of the pre-GDPR and post-GDPR permission usage data. The following changes in patterns can be observed:

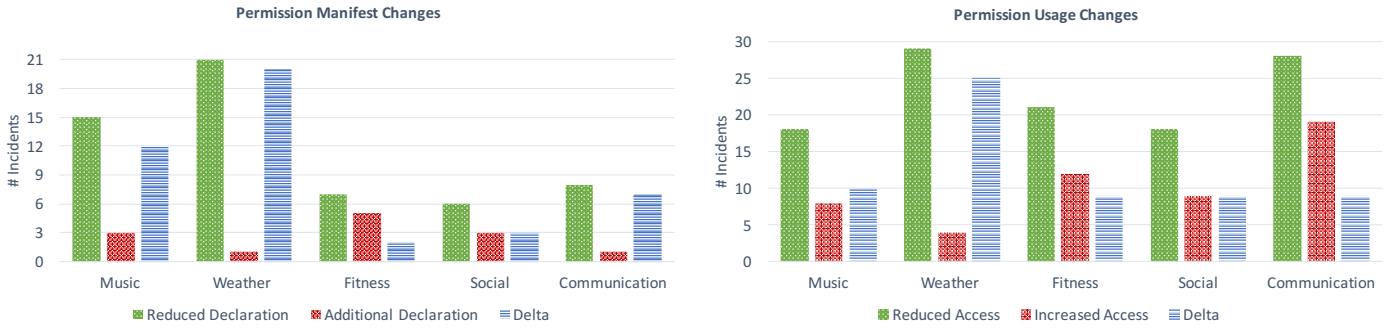
- In general, the frequent presence of green fields in Fig. 3b indicates that many apps reduced their permission access counts compared to the pre-GDPR phase.
- Considering the area between K11 & T20, *Weather* apps significantly reduced permission access and `YahooWeather(14)` shows the most significant improvement. The area between O14–O20 also shows reductions, as well as rows 8, 14, and 21.
- *Music* apps also reduced permission access during idle time with one notable exception: `GooglePlayMusic(3)` (see K3–T3: increasing `SENSOR & STORAGE` access while reducing access to the rest of permission groups.)
- Columns R & T indicate the overall increment in accessing `SENSOR & STORAGE` permissions. *Fitness*, *Communication* and *Social* apps are mainly responsible for such increase.
- Column O highlights the overall reduction of access to `LOCATION` permission group with a couple of exceptions: *Fitness* and *Social* apps have increased access frequency.
- `Line(43)`, `WeChat(44)`, `Imo(45)`, `Runtastic(28)` and `Pedometer(29)` are the five apps that increased idle time permission access frequency compared to the pre-GDPR period.

Some noteworthy observations can be made by comparing Fig. 3a to Fig. 3b:

- App 14 has both reduced its declared permissions in the manifest and shows reduced use of the remaining permissions. Apps 8 and 21 have reduced their use, but only removed few of the dangerous permissions from the manifest.
- The groups of social apps and communication apps (rows 31–50) show a reduced idle permission use pattern in columns K–P, however, they kept nearly all dangerous permissions in their manifesto ready to be used, as seen in columns C–J.
- While `CAMERA` and `CONTACTS` permissions are the most-added post-GDPR permissions in the manifest (column C–D), their idle use has strongly decreased (column M–N).
- Both the permission requests for and the use of `PHONE` (column G and Q) and `SMS` (column I and S) have reduced in the post-GDPR data set.

3 Changes in Expressed User Concerns

We collected a large number of user reviews from the Google Play app market for both the pre- and post-GDPR periods [7]. Applying text classification techniques, the goal was to investigate how many privacy-related concerns can be extracted from publicly available user reviews that would



(a) Manifest analysis

(b) Permission usage analysis

Fig. 4. A comparison of app permission declaration reductions and permission usage reductions from pre- to post GDPR periods. The blue columns illustrate the magnitude of reduction of (a) permission declarations in manifests and (b) actual permission usage.

ultimately enable us to compare users attitudes towards apps for both time frames. Thus, this section examines hypothesis H3.

Based on the mobile app threat categorization in [7], we processed, analyzed, and classified the collected user reviews (before the adoption of the GDPR, in Nov 2017) associated with the top 50 most downloaded apps within five app categories on the Google Play app market. As depicted in Fig. 5, we detected 799 privacy complaints concerning our app sets. When it comes to the most reported threats, *Targeted Ads*, *Spam* and *General* have the biggest portion with a share of 389, 136 and 105 complaints, respectively. It is worth mentioning that the *General* category mostly contains complaints regarding the over-privileged (permission hungry) apps that are requesting irrelevant permissions to their proper functionality. By contrast, *Tracking & Spyware*, *Phishing*, and *Unintended Data Disclosure* comprise the lowest number of complaints (16, 21, and 37 user reviews, respectively). In terms of the most and the least reported app categories, we can refer to *Lifestyle* (255 complaints) and *Music & Audio* (120 complaints) respectively. Both categories have *Targeted Ads* as the most reported threat. However, for *Lifestyle* category, the second most reported threat is *Spam* (62 complaints) and for *Music & Audio* it is *Unauthorized Charges* (30 complaints).

In February 2019, we collected a new data set of user reviews corresponding to the same app set to check the privacy perception of mobile users after the GDPR came into effect. Overall, a decent reduction can be observed in the total number of privacy relevant complaints (from 799 to 704). As for the most reported threat, similar to the pre-GDPR period, it involves complaints about *Targeted Ads* (347). As opposed to the pre-GDPR period, *General* category takes the second place with 130 complaints. In addition, the third most reported threat is *Spam* (97 complaints). The same scenario repeated for the least reported threats for the post-GDPR period with some permutations: *Phishing* (13 complaints), *Unintended Data Disclosure* (27 complaints), *Tracking & Spyware* (27 complaints). In terms of the app category with the maximum number of privacy complaints, *Lifestyle* still dominates the biggest share (165 complaints). In contrast to the pre-GDPR period where *Music & Audio* had the lowest number of privacy relevant complaints, this time it has the third biggest portion (150 complaints). The minimum number of privacy relevant complaints are found in *Health & Fitness* this time (113 complaints).

4 Discussion of Observations

We see an overall reduction in all three data sets that were collected during the post-GDPR period:

- The declared intent to use permissions in the manifest data shows a general reduction of the number of permissions declared in the manifesto.
- There are substantially fewer permissions shown in the post-GDPR data set that measured idling apps, although some permissions are being used more often.
- User concern expressed in the Google Play forum has somewhat decreased for all app categories, except for worries about targeted advertising and general security concerns.

It is difficult to correlate the observed phenomena directly with the GDPR. In this section, we try to interpret the results to find potential causes and explanations.

As shown in Fig. 3 and in Fig. 4, there is a significant reduction in the number of permissions used by apps. At first glance, we observe that the *data minimization principle* may be more strongly followed by app developers in the post-GDPR period. Such improvement in data minimization is also directly connected to the *purpose specification principle* as it requires developers to clarify the need for requesting relevant permissions. As an important observation, we found that apps are greedier in requesting sensor-related permissions (`CAMERA`, `MICROPHONE`, and `BODY_SENSOR`) in the post-GDPR period. Our explanation is that access to other permissions (`CONTACTS`, `LOCATION`, and `MICROPHONE`) may have become conditional and is triggered by motion sensors using `SENSORS`. Should there be changes in `SENSORS` data, other permissions can be invoked. We plan to investigate this phenomenon in future experiments. A possible explanation for this may be the fact that advertising content is shown more effectively when users actually look at the screen. Consequently, apps may use the sensors for motion, acceleration and bearing to determine when to show advertisements. Increasing interest

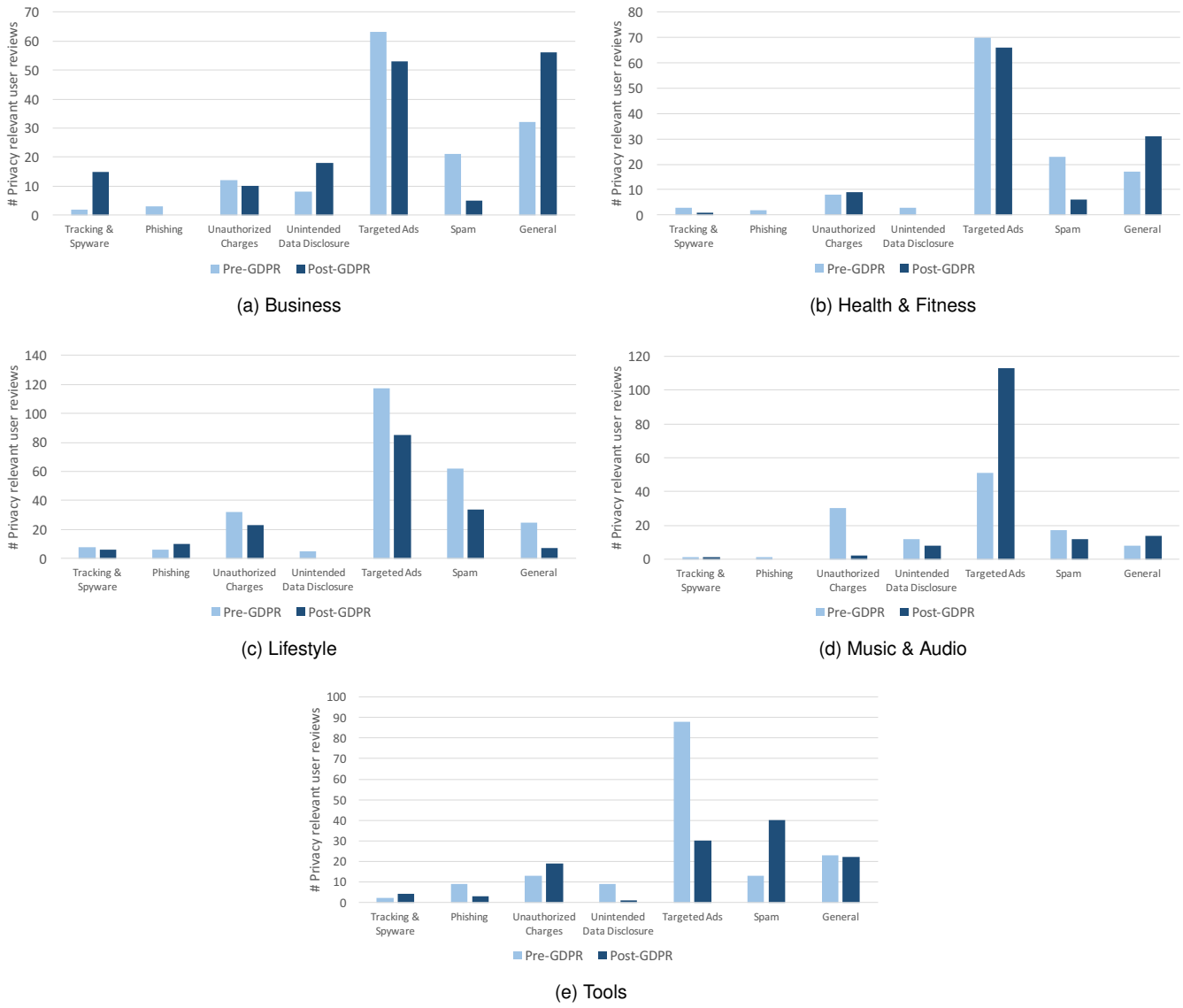


Fig. 5. A set of graphs comparing numbers of privacy-related user reviews associated with different app categories in pre- and post-GDPR scenarios: (a) business, (b) health & fitness, (c) lifestyle, (d) music & audio, and (e) tools.

in user location may express an increased trend towards location-aware advertising.

We see a large number of declared permissions remaining unused, particularly, in the communications and social apps categories. What does this imply? This might point to sleeping functionality that is not used in idling apps, and could indicate that some apps deploy continuous and invasive tactics to harvest more data (i.e. progressively collecting consent for dangerous permissions after a period of interaction). Such tactics have been observed and described in the literature [15], [16]. However, it is hard to judge whether this is the result of data minimization work, or of malicious intent. To clarify this, a further study will have to document apps efforts to collect partial consent in similar ways as those documented in a report about nudging users towards less privacy [17].

The significant reduction of permission declaration and use in the weather app group is best explained with a large amount of press and media attention spent on weather apps that extract personal data. From 2017 to 2019, there were frequent press stories in, for example The Times, the Wall Street Journal and PC World, that exposed weather apps greedy data collection and sharing practice. Such massive negative publicity may have led to the re-engineering of the weather apps for the simple reason of preventing further negative public exposure.

The reduction of apps' interest in the `PHONE` and `SMS` permissions (see Sec. 2.2) may provide slightly more privacy for those who use those communication channels, however, the decrease may also point to a general decline in these channels' popularity and relevance. A decline in telecom operators' text messaging profits caused by chat and messenger software is a well-established observation, with SMS revenue dropping for the first time in 2014.

The user feedback shows a descending trend in the total number of concerns voiced from the pre- to post-GDPR periods. This is a somewhat puzzling fact, since most of the global user population must have received

a multitude of GDPR compliance messages, seen media and press coverage about 'spy apps' and read news about major data breaches. All these should increase the attention, awareness, and wariness of consumers. However, it seems that the expressed user concern actually declined. This is evident for the *Lifestyle* category as the total number of privacy relevant complaints decreased by 90. However, in two cases we observed a gentle increase in the total number of users' complaints, namely *Business* and *Music & Audio* categories. Both categories target users' general interests and are supposed to give certain relevant functionalities. A closer look at the *Business* category shows a significant increase in the reporting of *General* threats in the post-GDPR period (almost twice as many). The same also happened to *Music Audio* where *Targeted Ads* were reported more than twice as often in the post-GDPR period.

A probable cause remains unexplored in this work, which is the possibility of establishing formal channels by companies to handle individual privacy-related complaints from the user. Therefore, the reduction of user complaints in public forums may not be able to adjudicate proper assessment of the real situation. Furthermore, implementation of GDPR has compelled many companies to provide several privacy notices and transparent services to the data subject (i.e. ability to see, download and modify data stored by the corresponding service provider), which may cause the user to be annoyed and reluctant in this regard respectively.

One possible interpretation of this result is the capacity and effect of software to establish de-facto laws and principles ignoring or bypassing regulation, see page 466 in [10]. The practices of the information industry may have constituted a precedence widely accepted by consumers, in spite of regulation. Another interpretation is that consumers may not really notice the subtle changes in post-GDPR app behavior. Previous research in human-computer interaction has shown that the permission-based access control models' consequences and implications are mostly incomprehensible to end-users [13], who may seldom have first-hand-experience with negative privacy impact.

5 Conclusion

In our data, we have seen changes in app behavior and in user feedback that point towards the positive impact of the GDPR on apps. The number of permissions demanded in app manifestos has somewhat declined (H1 confirmed), with the strongest decline in the weather app group. Idling apps seem to use fewer of the permissions than they are actually prepared to use, with observed reduction in permission use (H2 confirmed). In user feedback, a moderate decline in concerns related to privacy can be seen, though awareness and worries about targeted advertising seem to have increased. This is an overall confirmation of H3.

Apps seem to have become more interested in sensor data and location, as well as memory access. However, they still have the capacity to use other dangerous permissions and may just not yet use them while the app lacks the expected degree of interaction.

We speculated about reasons for our observations in section 4. Some apps may have undergone re-engineering for better privacy. Other apps may have moved their use of certain permissions where there is interaction with users, and where apps are actively being used, measurable by sensor input. Our findings are inconclusive concerning our expectation that regulatory compliance would show in app behavior and experience. The user feedback results are showing that consumers seem the most worried about targeted advertising, both before and after the GDPR came into effect.

Possibly the compliance process has not led to a reduction in the number of permissions used, but it may have changed privacy policies and the ways in which permission consent is obtained from the app users.

Overall, we conclude that app privacy has moderately improved after the GDPR was implemented.

Acknowledgments

The work presented in this article is partially funded by the ALerT project, Research Council of Norway, IKTPLUSS 2017-2021 and by the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 675730 Privacy&Us.

References

[1] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)," *Off J Eur Union*, p. L119, 2016.

[2] J. P. Albrecht, "How the GDPR will change the world," *Eur. Data Prot. L. Rev.*, vol. 2, p. 287, 2016.

[3] *Dangerous permissions*, Accessed on 27-Jan-2019. [Online]. Available: https://developer.android.com/guide/topics/permissions/overview#dangerous_permissions

[4] N. Momen, "Towards measuring apps' privacy-friendliness (licentiate dissertation)," Karlstad University, Department of Mathematics and Computer Science, Tech. Rep. 2018:31, 2018.

[5] N. Momen, T. Pulls, L. Fritsch, and S. Lindskog, "How much privilege does an app need? investigating resource usage of android apps (short paper)," in *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, Aug 2017, pp. 268–2685. [Online]. Available: https://ieeexplore.ieee.org/document/8476943

[6] A. Carlsson, C. Pedersen, F. Persson, and G. Söderlund, "Kaudroid: A tool that will spy on applications and how they spy on their users," Karlstad University, Department of Mathematics and Computer Science, Tech. Rep., 2018. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-66090

[7] M. Hatamian, J. Serna, and K. Rannenberg, "Revealing the unrevealed: Mining smartphone users privacy perception on app markets," *Computers & Security*, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404818313051

[8] D. Barrera, H. G. Kayacik, P. C. van Oorschot, and A. Somayaji, "A methodology for empirical analysis of permission-based security models and its application to android," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, Conference Proceedings, pp. 73–84.

[9] M. Hatamian, J. Serna, K. Rannenberg, and B. Igler, "Fair: Fuzzy alarming index rule for privacy analysis in smartphone apps," in *Trust, Privacy and Security in Digital Business*, J. Lopez, S. Fischer-Hübner, and C. Lambrinoudakis, Eds. Cham: Springer International Publishing, 2017, pp. 3–18.

[10] R. P. Wagner, "On software regulation," *Southern California Law Review*, vol. 78, pp. 457–520, 2004.

[11] T. Book, A. Pridgen, and D. S. Wallach, "Longitudinal analysis of android ad library permissions," Rice University, Report arXiv preprint arXiv:1303.0857, 18-Apr-2013 2013. [Online]. Available: https://arxiv.org/pdf/1303.0857.pdf

[12] *Permissions overview*, Accessed on 27-Jan-2019. [Online]. Available: https://developer.android.com/guide/topics/permissions/overview

[13] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall, "A conundrum of permissions: Installing applications on an android smartphone," in *Financial Cryptography and Data Security*, J. Blyth, S. Dietrich, and L. J. Camp, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 68–79.
[14] L. Fritsch and N. Momen, "Derived partial identities generated from app permissions," in *Proceedings of the Open Identity Summit (OID) 2017, Lecture Notes in Informatics LNI, 277*. Gesellschaft für Informatik, 2017.
[15] L. Fritsch, "Partial commitment–try before you buy and buyers remorse for personal data in big data & machine learning," in *Trust Management XI: 11th IFIP WG 11.11 International Conference, IFIPTM 2017*. IFIP AICT 505, 14-Jun-2017 2017, pp. 3–11. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-55017
[16] L. Fritsch, "Privacy dark patterns in identity management," in *Open Identity Summit (OID), 5-6 october 2017, Karlstad, Sweden.* Gesellschaft für Informatik, 2017, pp. 93–104.
[17] "Deceived by design - how tech companies use dark patterns to discourage us from exercising our rights to privacy," Consumer Council of Norway, Report, 27-Jun-2018 2018. [Online]. Available: https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf

BIOGRAPHIES

Nurul Momen is a Ph.D. candidate in the Department of Computer Science and Mathematics at Karlstad University, Sweden. His research interests focus on privacy-enhancing technologies, transparency, usability, mobile communications, and data protection, particularly the security and privacy aspects of access-control models for mobile operating systems. Momen received an M.S. in security and an M.S. in privacy from the double-degree program at the Technical University of Berlin, Germany, and the University of Trento, Italy. Contact him at nurul.momen@kau.se.

Majid Hatamian is a Ph.D. candidate in computer science at the Goethe University of Frankfurt, Germany, where he also serves as a research and teaching assistant to the chair of mobile business and multilateral security. His research interests are in wireless communications, security and privacy in peer-to-peer networks, nano-communications, and machine-learning techniques, and his current focus is on privacy-risk analysis in smartphone ecosystems. He is a Member of the IEEE. Contact him at majid.hatamian.h@ieee.org.

Lothar Fritsch is an associate professor in information security in the Department of Computer Science and Mathematics at Karlstad University, Sweden. His research interests include cyber-security, privacy-enhancing technologies, privacy-preserving identity management, data protection, cryptography, public key infrastructure, mobile signatures, mobile computing, mobile communications, m-commerce, e-commerce, location-based services, and platform accountability. He received PhD from Goethe University of Frankfurt, Germany. Previously he worked as senior research scientist at the Norwegian Computing Center. Dr. Fritsch is a member of International Federation for Information Processing (IFIP) and Gesellschaft für Informatik (GI). Contact him at lothar.fritsch@kau.se.