Where does your data go? Comparing network traffic and privacy policies of public sector mobile applications

Robin Carlsson[1], Timi Heino[1], Lauri Koivunen[1], Sampsa Rauti[1], and Ville Leppänen[1]

University of Turku, Turku, Finland
{crcarl,tdhein,lamkoi,sjprau,ville.leppanen}@utu.fi

Abstract. As services increasingly move online and mobile devices become ubiquitous, mobile applications are widely used by ordinary people with little technical knowledge. Consequently, user privacy has become an essential matter to consider when developing mobile applications. In this paper, we study the privacy of 32 mobile applications provided by Finnish public sector bodies. First, we investigate with network traffic analysis what kind of personal data these application send out to third party analytics services. We then analyze the privacy policy documents of these applications and assess their clarity and transparency. Our findings show that there are several inconsistencies between the actual traffic of the studied applications and what is said about processing personal data in privacy policies. This underlines the need for software developers and organizations to be better aware of privacy regulations and data their applications send out. There is also lots of work to be done in making the privacy policies less vague and more informative, for example when it comes to explaining what technical data items are sent to third parties and how this can potentially affect the user privacy.

Keywords: Mobile application privacy · network traffic analysis · privacy policies · personal data

1 Introduction

In digitalizing society, smart mobile devices are increasingly being used to accomplish everyday tasks. At the same time, many mobile applications collect potentially sensitive personal data and deliver it to third parties such as *analytics services*. Authorities are attempting to regulate the usage of personal data with GDPR (General Data Protection Regulation) with the aim of improving individuals' control and rights over their personal data. According to GDPR, any data collection must be disclosed to users and they must be clearly informed about the extent of data collection.

For the large part, however, users remain oblivious to the amount and nature of the potentially sensitive data mobile applications send out. Privacy policy documents meant to inform users about the nature of collected data are often

overly long and use vague expressions. In this study, we are interested in what kind of personal data mobile applications send out to analytics services, and how transparently and accurately the privacy policy documents describe this data and the related data handling procedures of the analyzed applications. Application vendors typically use third party analytics services for better understanding how users use their application. The downside of using analytics services is that third parties collect data on the users.

The contributions of this paper are as follows. We present a study of network traffic of mobile applications provided by Finnish public sector bodies. We analyze the network traffic in order to identify personal and sensitive information sent out by the applications. Moreover, we also compare these findings to the privacy policies to find out whether there are inconsistencies between their statements about personal data and the real network traffic. We chose to study mobile applications provided by organizations and companies in the public sector because they often handle sensitive data. One can also argue that these bodies should take special care in building their applications, have clear and accurate privacy policy documents, and generally act as an example to other actors in the digital world.

The rest of the paper is structured as follows. Section 2 explores related work. Section 3 covers the setting of the study, including the technical setup we used to capture network traffic of the analyzed mobile applications, and the approach we used to analyze the privacy policy documents. Section 4 presents the results of traffic analysis, and assesses the transparency and clarity of the related privacy policies. Section 5 discusses the implications of our findings and also covers the limitations of the study. Finally, Section 6 concludes the paper.

2 Related work

Brandtzaeg et al. [1] conducted a mixed-methods study that included a user survey, an analysis of personal data flows in mobile applications, and a content analysis of privacy policies of 21 popular mobile applications. The user survey showed that more than half of the respondents had avoided downloading mobile application because of privacy concerns. Their findings also show that 19 of the 21 applications sent out personal data to a total of approximately 600 different domains. It was found that, in three cases, the application shared data in violation with its terms of use. As a solution to the data leakage problem, the authors suggest visualizations to improve transparency of personal data flows in mobile applications.

Papageorgiou et al. [8] present an in-depth security and privacy analysis of popular free mobile health applications. The study is carried out by analyzing the applications by both static and dynamic means and by testing each application's functionalities. Applications are also audited for general data protection regulation compliance. The findings of the study reveal that most of the studied applications do not conform to any well-known privacy guidelines or practices. Even data protection regulations are often ignored, jeopardizing

the privacy of users. On a similar note, O'Loughlin et al. [7] investigated privacy policies of 116 mobile phone applications intended for depressed people to gauge the transparency of these documents. Of the analyzed policies, 4% got a score of acceptable, 28% questionable, and 68% unacceptable. Only 49% of the studied applications had a privacy policy document. However, applications collecting identifiable personal information were more likely to have a privacy policy compared to applications which did not collect such data.

Jia et al. [2] proposed an approach for personal data leakage detection based on association mining. With this method, the authors then searched for hidden privacy leakages in traffic recorded from mobile applications. They tested 509 applications with the developed algorithm. The findings revealed that 76.23% of the applications collected and transmitted personal data insecurely and 34.06% of them sent personal data to third parties.

Liu [4] et al. studied the privacy of the mobile operating systems by investigating six different variants of the Android OS. They found that all Android variants transmitted significant amounts of data to the OS developer and also to third parties (such as Google, Microsoft and Facebook) that have pre-installed system apps. The authors found that regular data transmissions go well beyond occasional communication with OS servers, raising several privacy concerns. The fact that the user cannot opt out from this data collection makes the situation worse.

The findings of these studies show that despite regulatory measures such as GDPR, the privacy of mobile applications and systems in general is still not adequately addressed. Privacy policies of mobile applications often are not sufficiently transparent and do not reflect the actual data flowing out of the user's device. Our paper contributes to this body of existing research by studying mobile applications provided by the Finnish public sector. We also introduce an experimental setting used for recording traffic of mobile applications, and provide a detailed analysis of personal data items leaked by applications and the destinations this data is sent to. Lastly, the clarity and transparency of privacy policies is analyzed from an ordinary user's point of view.

3 The setting of the study

In the current study, we analyzed 32 mobile applications of Finnish public sector bodies. For the purposes of this study, public sector applications are mobile applications provided by the Finnish government or cities, publicly controlled or funded entities or enterprises, and other bodies that offer public services or goods. These include, among others, several applications related to public transport, mobile services of the Finnish public broadcasting company Yleisradio, and guide applications for Finnish cities.

The applications were searched from Google Play's regional list of most popular applications[1], and also by extending the search to cover other applications

[1] https://play.google.com/store/apps/top

created by the same developers. Several well-known services provided by Finnish public sector bodies were also directly searched by name in Google Play. Only applications with over 1000 users were included in our analysis.

In our experiment, the mobile phone was connected to the internet only through a Linux computer which acted as a WiFi access point. The access point has been configured to record the phone's network traffic. Because the operating system has pre-installed applications which produce network traffic and because our experimental environment does not allow us to detect which exact application has produced the traffic, the system's normal background noise was recorded and this noise was omitted from application-specific recordings.

Network traffic was captured with mitmproxy and tcpdump tools. Mitmproxy is an open-source proxy tool that can be used for intercepting, inspecting, modifying, and replaying web traffic such as HTTP, SSL/TLS, HTTPS and WebSocket. The application acts as a man-in-the-middle between a web client and a web server. Tcpdump is a packet analyzer program which can read and display contents of packets transmitted through a network. The traffic logs of each mobile application were analyzed and any personal or sensitive data that can be used to identify or profile a user was recorded. The analytics providers (such as Google and Facebook) the applications connected to were also listed. Figure 1 shows a sample excerpt of the recorded mobile application traffic. When recording the network traffic, each application was actively used for a few minutes[2], with the aim of invoking the most important functionality of the application.

The privacy policies of analyzed mobile applications were then read carefully to assess whether the data they send out – especially to third parties – was clearly mentioned in the documents. The privacy policies were searched for mentions of third parties and the analytics services detected during traffic analysis. The observations made were confirmed by another researcher, and any disagreements were discussed until a consensus was reached. Moreover, we also gauged the clarity of the documents, as privacy policies with vague expressions confuse users and often do not answer the questions users may have about how their personal data is handled. Overly generic privacy policies trying to cover every service of the same service provider, for example, have a similar problem as well [10].

Finally, briefly clarifying the concept of *personal data* is in order here. We borrow the definition of the Finnish office of the data protection ombudsman, and say that personal data is "all data related to an identified or identifiable person"[3]. Therefore, an IP address, accurate location data, or an ID that an analytics service uses to track the user is personal information. Moreover, there is also the concept of so called special category personal data, which we simply call *sensitive personal data* in this paper. This refers to data concerning the user's health, political opinions, ethnicity etc. that requires a higher level of protection.

[2] We have written detailed descriptions of how each application was used. Due to space constraints, these details are not included in this paper. The detailed descriptions of our experiments as well as a list of researched mobile applications are available upon request.

[3] https://tietosuoja.fi/en/what-is-personal-data

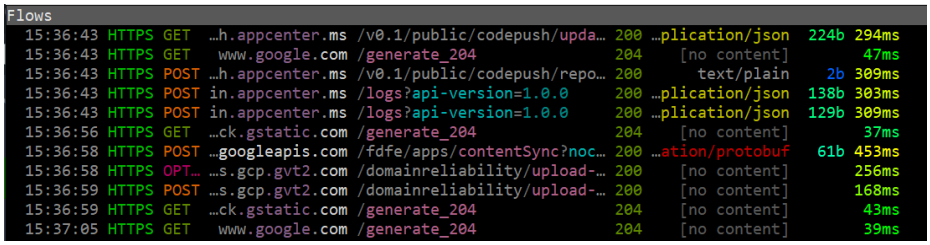

Fig. 1. An excerpt from the traffic log of an analyzed application. The log generated by *mitmproxy* lists requests sent by the inspected application. The contents of each network packet can also be inspected with the tool.

4 Results

4.1 Network traffic analysis

The network traffic recorded from the studied mobile applications was inspected for any personal data or data that can be used when profiling a specific user. Table 1 lays out the personal data the studied applications were found to send out to any third party analytics services. Note that also analytics services or libraries implemented by a third party but located on a Finnish service provider's or application developer's servers are included in the table. These locally hosted services may still circulate the received data through a third party.

Unsurprisingly, the most frequently sent piece of data is the device's IP address. Note that not all studied application connected to the internet at all, and the table only includes the data applications sent to analytics services. In light of this, the percentage of applications that transmit an IP address, over 85%, can be considered quite large. While the fact that the device IP address is sent to all analytics services the applications communicate with, many ordinary users may not be aware that a unique address, which can usually be used to identify them, is being delivered to third parties.

Other frequently sent data includes several types of technical details such as the phone brand and the operating system. Screen and window sizes as well as processor reveal additional details that can be used to identify the user. Note that some of these details are being transmitted quite frequently, for example over half of analyzed applications send out the phone brand and model. Interestingly, in two cases, even somewhat sensitive information on whether the phone is rooted is sent out – it is likely that this is functionality of the development platform and not intended by the application developers. Other pieces of contextual information such as the country, language and internet service provider were also often delivered to third parties and analytics services.

The aforementioned pieces of technical and contextual data are not personal alone but can easily be used to build a profile for a specific user when different pieces of data – possibly also from different applications and websites in the case of large analytics companies – are combined together.

Table 1. Data items sent to analytics services by the analyzed mobile applications.

Sent data	Number of applications	Percentage
IP address	29	85.3
Phone brand and model	21	61.8
Phone OS	21	61.8
Phone OS version	20	58.8
Processor	2	5.9
Screen size	21	61.8
Window size (viewport)	6	17.6
Is the phone rooted?	2	5.9
Internet service provider	14	41.2
User identifier	3	8.8
Device identifier	3	8.8
Other unknown identifier	19	55.9
Timezone	7	20.6
Location (country)	11	32.4
Location (city, area)	1	2.9
Timestamp	20	58.8
Language	12	35.3
List of purchased products/services	1	2.9
Performed actions	12	35.3

On the other hand, user and device identifiers are very personal. We also found there was a high number of unknown identifiers – usually long strings that likely act as identifiers for either the users, devices or sessions. Almost half of the applications sent out these kind of identifiers. Finally, over third of the studied applications also reported actions the user performed back to the analytics services. One of the applications even leaked a list of items the user had chosen in an online store.

It is also worth taking a look at the destinations of the data sent to third parties. Figure 2 shows an alluvial diagram of analytics services the analyzed applications were found to send data to. We can see that Microsoft's App Center and Google's different analytics services (which are bundled together here for simplicity) are the most popular destinations for user data. App Center is used by 9 applications and Google's services are used by 8. Adobe, Facebook, Firebase and Hotjar are also used by several applications.

4.2 Analysis of privacy policies

The scopes of the studied privacy policies greatly varied. Only 5 applications (15.6%) had privacy policies which focused on a single mobile application, while 18 (56.3%) applications had documents that covered several services (for example, the mobile application and the organization's website) at once. In the case of 9 applications (28.1%), the privacy policy was even more generic, encompassing several applications and all digital services of the organization. Lastly, two

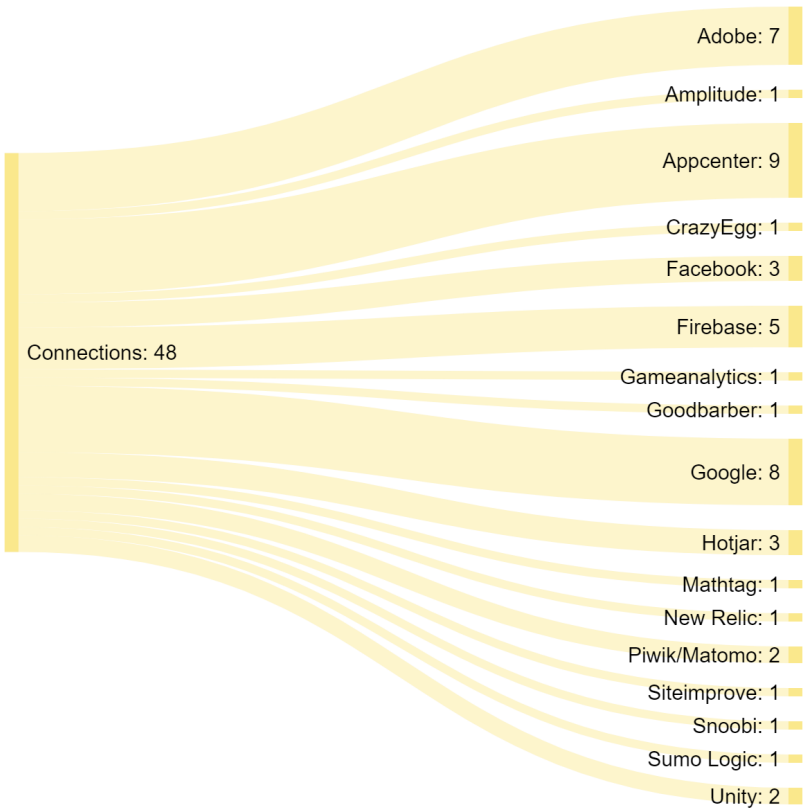

Fig. 2. An alluvial diagram of analytics providers used by public sector mobile applications tested with mitmproxy. One connection per provider per application is shown in the diagram. For example, three separate applications used Hotjar.

(6,3%) of the mobile applications did not appear to have privacy policies even after a careful search.

In many cases, the generic privacy policies that discussed privacy of the services provided by the whole organization were quite vague and difficult to follow from the point of view of a mobile application user. For example, the only privacy policy document that was found to be associated with the "City of X" (name of the city withheld) application was the generic privacy policy of the city, which did not separately discuss privacy of the application. With these kinds of generic privacy documents, it is often not clear to the reader what data is handled and collected by the mobile application, especially if the application is not discussed in a separate section in the privacy policy.

Even when the mobile application was mentioned in the more generic privacy policies, it was often not sufficiently separated from other digital services such as the organization's website. While the boundary between mobile applications

and websites may be blurring in many cases (for example in the cases of progressive web applications or browsers embedded in native mobile applications), in most cases different applications and services could be better separated from each other to correctly and accurately inform the user about the usage of their personal data.

Generic documents often alternated between generic and application-specific privacy issues and often got very confusing about what service exactly is being discussed (e.g. mobile application, website, all digital services of an organization). In some cases, the user has to go through several privacy policy documents to get a good understanding of mobile application's privacy. One privacy policy even prompted the reader to read Firebase's English privacy policy in order to understand how the data is moved and handled outside Europe.

When a privacy policy indicates the data is given to a third party, it is not always clear who handles the data. For example, the Finnish broadcasting company Yleisradio states that it uses "subcontractors and service providers" to help with the analysis of user data. However, there is no way for users to know who these other involved parties are and where their personal data goes exactly, unless they analyze the traffic of mobile applications, which of course is an unreasonable assumption for ordinary users. Moreover, in many cases there was no mention whatsoever about how long the collected data is going to be stored.

4.3 Comparison between observed data collection, privacy policies and permissions

When comparing privacy policies with the captured traffic, several inconsistencies between privacy policies and the actual functionality and network traffic of the studied applications were found.

One clear difference between the analyzed privacy policies and the delivered data was the fact that lots of technical data was sent out to third parties but not always reported in the privacy policies. This data included details on the device and network connection, such as the network operator, the device brand and model, processor, the operating system, window size and the screen size. While these pieces of data are not sensitive and personal individually, they can be combined to build a profile for a specific user. Many analytics services are very good at this, and a digital fingerprint is often created without the user's consent. Moreover, many ordinary users may not even understand the meanings of these pieces of technical data [5], much less understand the repercussions of sending such data to third parties, which arguably gives even more reason to cover them in a privacy policy.

Another piece of technical data frequently ignored in privacy policies is the device's IP address. What makes an IP address different from the previously discussed pieces of technical information is the fact that, in most cases, it can be directly used to identify a specific user and constitutes personal data[4]. Like other

[4] The Finnish data protection board has ruled that IP addresses should be treated as personal data.

technical pieces of data transmitted from mobile applications, an IP address is also something all users may not completely understand or even know about, which makes it an important detail to mention in privacy policies.

In some cases, the information provided in the privacy policy document was outright incorrect. For example, the privacy policy of one application promised that data is not moved outside European Union or European Economic Area. Still, our traffic analysis revealed that the data on the user's approximate location and the device's IP address was sent to the USA.

There was also one case which involved the mobile application sending clearly sensitive personal data to analytics services. More specifically, an application of a Finnish government enterprise has sent the data about the items the customer has chosen in an online store to third parties (Facebook and Google). Based on the line of business in question, this data can be considered highly sensitive. At the same time, in the related privacy policy, the company claims it does not handle any sensitive personal data, not to mention sending the data to analytics services that also operate outside Europe.

In some cases, it is quite obvious that the effort put in privacy matters has been minimal. A guide application for a large Finnish city, for which we were unable to find a privacy policy document altogether, still sent identifying information on the user to the USA. There were also cases in which the link to the privacy policy was broken in the application store and we had to use a search engine to find it.

Some of the applications had an embedded web browser so that the user's device could easily display a related website as a part of the application. The analytics services used on such websites were usually not included in the privacy policy documents. For now, this seems to remain a grey area and most application developers do not seem to deem it necessary to take the embedded website into consideration in mobile applications' privacy policy documents.

Finally, even if the collected data is clearly mentioned in the privacy policy document, the questions still remains whether collecting information is necessary or ethical. For example, Pikku Kakkosen Eskari, which is a game application for children, records the actions taken in the game and also sends this information to the server of the development platform used to build the application (Unity). It should also go without saying that special care should be taken when handling any data on children.

Altogether, we found that over third of the analyzed applications handled the data in a way inconsistent with the privacy policy, or at least the privacy policy was too unclear for the user to get a good understanding what kind of personal data is sent and where. To summarize, these inconsistencies and unclear issues can be categorized as follows:

- Technical details about the device are collected but this is not stated in the privacy policy.
- Tracking mechanisms are used in the application, but this is not clearly stated in the privacy policy (in generic privacy policies, cookies and analytics

on the organization's website are often mentioned but the mobile application is ignored).
- Personal data is sent outside EU/EEA, while the privacy policy says this does not happen.
- As an important part of the mobile application's functionality, a browser embedded in the application takes the user to a website that collects lots of analytics, but this data collection is not mentioned in the privacy policy.
- The privacy policy says sensitive information is not handled by the application, but information about the sensitive products/services the user has purchased is sent to analytics services (Facebook and Google) along with the IP address.

5 Discussion

In the current study, we have seen that privacy policies of mobile applications do not often fully reflect the personal data that is sent to third parties. For example, IP addresses and several device related details are often leaked to third parties without mentioning them in the privacy policy. It seems that developers may not consider these details personal data and may not fully understand possible consequences of such data leaking out.

Collecting technical information about the device and application usage is a part of user profiling but it may also be necessary to improve user experience. The latter is also the reasoning often used by developers (and analytics services) to justify the collection of the data. This information collection easily becomes a problem, however, when lots of details about the user accumulate over time, or the user simply happens to use a phone model that is rare in a certain area, for example. Profiling users on the internet based on the technical details their browsers sends out has been honed to a fine art over the years [3] and the same is quickly happening with mobile applications.

Mobile application developers should be better aware of what constitutes personal data. It is especially important to be careful with bits of information that can be pieced together. The European commission clearly states that "Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data"[5]. It is also essential for the developers to know where the used third party analytics tools deliver the data. Application network traffic analysis similar to our analysis in the current study should be a part of mobile application testing, and developers should take care to read the privacy policies of the third party services and libraries they use. For critical applications in the public sector, developers should also look into the possibility of implementing the necessary analytics themselves or using a solution that is guaranteed to keep the data on safe local servers. In general, public sector agencies should work to bring privacy of their applications to a high standard, encouraging private companies to follow the example [11].

[5] https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

Another obvious challenge is writing clear and concise privacy policy documents [6]. The wide scopes and vague expressions were the main observed problems of privacy policy documents analyzed. GDPR states that privacy policy documents have to be written "in a concise, transparent, intelligible, and easily accessible form", encouraging the use of "clear and plain language". A possible way to improve the readability of these documents would be introducing a few different privacy policy templates or structures that privacy policy documents have to follow. Standardized formats would make the documents both easier to create and easier to understand [9].

Analyzing application network traffic as a part of testing will also help the developers to review what kind of data their application really sends out (especially when third party libraries are used) and makes it easier to write accurate and realistic privacy policy documents. Privacy policies could also be improved when it comes to informing users about possible consequences of information being transferred to a third party. This does not mean frightening users with unrealistic threats, but laying out possible scenarios about, for example, how the technical details of the mobile device can be used to profile the users.

The current study has a few limitations. First, our traffic analysis only reveals the clearly identifiable personal information sent out from the mobile applications. There may be other transmitted data, for example some details that have purposefully been made difficult to detect. Second, although we can say with certainty that specific personal details have been sent to a specific destination, we cannot say whether this data (for example, an IP address) is stored and made use of, or whether it is immediately discarded. Our analysis is only limited to the client side. The analysis also does not cover some functionality available in the studied applications, such as purchasing products or filling in and sending forms.

Third, with the widespread use of cloud services and conflicting information of their server locations, it is in many cases difficult to say whether data has been sent outside Europe. This is why we have not focused that much on geographical location. Finally, we have mostly analyzed the traffic and privacy policies from the perspective of an ordinary user on the one hand and from the technical perspective on the other. Any kind of legal analysis is outside the scope of this study, however.

6 Conclusions

In this paper, we have studied the personal data the mobile applications provided by Finnish public sector bodies send out, especially to third parties, and assessed how clearly and transparently the privacy policy documents of the studied applications report the processed personal data items and the data handling process. We found several inconsistencies between actual network traffic of the mobile applications and their privacy policies. While most of the found inconsistencies are unlikely to have immediate serious consequences, there is definitely a need to stop practices such as delivering device details and user identifiers to

third parties without mentioning this in the privacy policy, or claiming to not send anything outside EU/EAA but still doing so.

Overall, our findings underline the need for software developers to be better aware of privacy regulations and data their applications send out. Developers and organizations also need to make privacy policies clearer by avoiding vague expressions and by taking care to explain what kinds of technical data is sent to third parties and how it can potentially be used to profile users.

Acknowledgements

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

References

1. Brandtzaeg, P.B., Pultier, A., Moen, G.M.: Losing control to data-hungry apps: A mixed-methods approach to mobile app privacy. Social Science Computer Review **37**(4), 466–488 (2019)
2. Jia, Q., Zhou, L., Li, H., Yang, R., Du, S., Zhu, H.: Who leaks my privacy: Towards automatic and association detection with gdpr compliance. In: International Conference on Wireless Algorithms, Systems, and Applications. pp. 137–148. Springer (2019)
3. Kaur, N., Azam, S., Kannoorpatti, K., Yeo, K.C., Shanmugam, B.: Browser fingerprinting as user tracking technology. In: 2017 11th International Conference on Intelligent Systems and Control (ISCO). pp. 103–111. IEEE (2017)
4. Liu, H., Patras, P., Leith, D.J.: Android Mobile OS Snooping By Samsung, Xiaomi, Huawei and Realme Handsets (2021)
5. Liu, Y., Song, H.H., Bermudez, I., Mislove, A., Baldi, M., Tongaonkar, A.: Identifying personal information in internet traffic. In: Proceedings of the 2015 ACM on Conference on Online Social Networks. pp. 59–70 (2015)
6. Mulder, T.: Health apps, their privacy policies and the gdpr. European Journal of Law and Technology (2019)
7. O'Loughlin, K., Neary, M., Adkins, E.C., Schueller, S.M.: Reviewing the data security and privacy policies of mobile apps for depression. Internet interventions **15**, 110–115 (2019)
8. Papageorgiou, A., Strigkos, M., Politou, E., Alepis, E., Solanas, A., Patsakis, C.: Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access **6**, 9390–9403 (2018)
9. Rowan, M., Dehlinger, J.: A privacy policy comparison of health and fitness related mobile applications. Procedia Computer Science **37**, 348–355 (2014)
10. Sunyaev, A., Dehling, T., Taylor, P.L., Mandl, K.D.: Availability and quality of mobile health app privacy policies. Journal of the American Medical Informatics Association **22**(e1), e28–e33 (2015)
11. Thompson, N., Ravindran, R., Nicosia, S.: Government data does not mean data governance: Lessons learned from a public sector application audit. Government information quarterly **32**(3), 316–322 (2015)