Privacy in Popular Children's Mobile Applications: A Network Traffic Analysis

Robin Carlsson*, Sampsa Rauti*, Samuli Laato†, Timi Heino*, Ville Leppänen*
* University of Turku, Department of Computing, Turku, Finland
† Tampere University, Gamification Group, Tampere, Finland
crcarl,sjprau,sadala,tdhein,ville.leppanen@utu.fi

***Abstract*—Children increasingly download and use mobile applications from marketplaces such as Apple's App Store or the Google Play Store. One would expect that applications intended for children are free of third-party analytics, or at least make sure parents give their consent for collecting personal data from children. In this study, we performed an in-depth technical analysis of a representative snapshot of 15 applications from Google Play Store aimed at children (age group classifications 0–5, 6–8 and 9–12). We recorded the network traffic of these applications and compared it to the privacy policies of the applications. Across the applications, a significant number (13/15) were delivering more information about the users to various third parties than what was admitted in the respective privacy policies. We elaborate on details regarding the observed network traffic, and discuss implications of these findings on strategies for preserving user privacy, particularly for sensitive audiences such as children.**

***Keywords*—privacy, web services, data leaks, vulnerable groups**

I. Introduction

In recent years, the use of mobile devices such as smartphones and tablets has become widespread. In addition to adult audiences, there are tens of thousands of games and educational applications aimed at children on app stores such as Google Play Store and Apple App Store [1]. Due to phenomena such as the abstraction of development tools, online connectivity and the availability of online services, increasingly many applications, including applications aimed at children, now operate online - with multiple features such as advertisements, updates and multiplayer elements operating through online services [2]. These developments have led to a situation where there is a huge amount of network traffic being generated by mobile devices, and accurately understanding and classifying this traffic has become a challenge [2].

Several analytics companies such as Alphabet/Google, capture and record users' digital fingerprints [3], [4]. Personal data items such as device and user specific identifiers and gaming behavior are used to better understand the user base, more optimally target ads [5] or to improve the usability and performance of applications. However, in addition to this data being collected for developers, it is also leaked to third-party companies [6], [7]. Unfortunately, all of this tracking is usually not visible for users, and children in particular can be considered a group vulnerable to privacy violations [8].

This situation is concerning from the perspective of end user privacy. The GDPR recognizes the need for extra protection for children's personal data and sets specific requirements for its processing. The regulation requires that parents or guardians provide their consent for the processing of children's data, and that application developers implement appropriate measures to ensure that children's rights to privacy are respected. However, past work has showcased that parents and children alike may not understand digital privacy very well [9], and that application developers and app stores do not help by following privacy regulations poorly [4]. Therefore, it is critical to examine the network traffic of children's applications on app marketplaces for understanding the magnitude of the problem of privacy violations in children's mobile applications.

Former studies have explored the privacy of children's applications at a large scale, often by automatic methods. Binns et al. [10] used static analysis to study apps from Google Play stores. They discovered that apps aimed at children had a high number of third-party trackers. Reyes et al. [11] conducted a study of network traffic from the most popular free Android children's applications and discovered that most of the applications had potential violations of Children's Online Privacy Protection Act. The aim of this study is to take a closer look at the privacy of children's applications. Therefore, we carried out an in-depth investigation of a representative snapshot of mobile applications aimed for children. This included analyzing the applications' network traffic and studying their privacy policies and Google Play Store's data protection section to see how transparently the personal data leaked to third parties is reported. We also tested whether the applications request consent from parents before a child can use them.

In doing so, this study contributes important practical information on the status quo of end user privacy in available children's applications in popular app marketplaces. Our findings also have contributions to academic literature on privacy, as we discuss measures for improving user privacy from a software engineering point of view based on our findings.

The rest of the paper is structured as follows. Section II outlines our research approach. Section III presents the

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

results of the network traffic analysis, and compares the actual traffic to Google Play's data protection sections and privacy policy documents. Section IV summarizes our main findings, discusses implications for software development and addresses limitations of the current study. Finally, Section V concludes the paper.

II. Research approach

Methodologically our work followed a three-step approach. First, we searched for a representative snapshot of children's applications on the popular app marketplace of Google Play Store. Second, using an intermediary Linux computer between the mobile phone and the internet we recorded all network traffic passing between the surveyed applications and any third-party servers. Third, we analyzed the data from two perspectives: a descriptive look of what data is sent and where; and a comparison between the applications' Google Play data protection sections, privacy policies and observed data traffic. Next we elaborate on these three steps in more detail.

A. The application search process

We selected Google Play Store as the marketplace for searching for applications primarily for two reasons. First, it is a hugely popular marketplace with over 100 billion annual app downloads across the globe [12]. Second, it provides a specific section of applications aimed at children, and search tools for obtaining these applications. Previous research looking for applications in Google Play Store have used specific keywords and obtained the first results for analysis, with the justification that these applications are the ones average users are also most likely to find [1], and those studies looking at very specific applications have cut off the search when a new page of results did not yield any new matches [13]. In the case of this study, we had potential applications in the tens of thousands, meaning we also had to cut off the analysis at some point, and here we referred to the criteria of saturation: when the latest search results provided no additional insights, we could cut off the search. Hence, we commenced searching for children's applications on Play Store in summer 2022 from the "Kids" section. All apps that are accepted here have to go through a specific review process, and Google advertises that these are accepted by teachers [1].

In order to ensure we obtained a representative snapshot, we started reviewing applications from three age groups: 0–5, 6–8 and 9–12. Due to the authors' location, we focused on apps aimed at Finnish children, thus setting the preferred language to Finnish and English. Subsequently, without any further sorting, we downloaded the first applications from each of the three categories. We continued this process until the researchers were confident that we had reached a sufficient saturation in terms of the number of applications. Saturation here was measured so that once we had gone through two rounds of applications without any further significant findings from the network traffic analysis, we determined that observing any further applications would provide diminishingly small added value. Following this criteria, in the end we went through 15 applications, 5 for each age group category.

B. Recording network traffic

In order to record all mobile traffic, we set up an Android phone that was connected to the internet through a computer, which acted as a wireless access point. This Linux computer was used to record the phone's network traffic. Because there are some pre-installed applications related to the phone's operating system which generate network traffic, the system's background noise produced by these applications was recorded and removed from application-specific recordings.

On the Linux computer, mitmproxy and tcpdump tools were used to record the traffic. Mitmproxy is a free proxy tool that allows intercepting and inspecting network traffic by acting as a man-in-the-middle between the client and the server. Tcpdump was used to display and analyze contents of network packets transmitted through the access point. A traffic log file was generated for each studied mobile application, and the files were analyzed to find any personal data that can be used to identify a user. In addition to recording these data items, the third parties receiving personal data from applications were also listed. While recording the traffic, the main functionality of each children's application was actively tested for a few minutes[2]. In Figure 1, a sample snippet of captured mobile application traffic is shown.

C. Data analysis

Once we had the network traffic recorded we proceeded with the analysis. As we were interested in whether any personal data was being shared by these applications, it is important to clarify what it means. Both GDPR and Finnish office of the data protection ombudsman define personal data as "all data related to an identified or identifiable person"[3]. Therefore, data items such as a user's IP address, or accurate data revealing the user's location are chiefly considered personal data. However, it is also important to note that a combination of several pieces of data can often be used to identify a person and can constitute personal data. For instance, although a device's screen size as a technical data item does not directly identify a specific device or user, it can definitely be a very useful piece of information in profiling a user when combined with other technical data items.

First, we looked into whether any of such personal data was being shared, and where. We recorded all instances

[1]For more information, see Google's description of the Kids section here: https://play.google.com/store/apps/category/FAMILY?hl=en&gl=US &pli=1, visited February 1, 2023

[2]While the tested functionalities could not be included in this paper due to space constraints, detailed descriptions of how each application was tested are available upon request.

[3]See https://gdpr.eu/eu-gdpr-personal-data/ and https://tietosuoja.fi/en/what-is-personal-data, visited February 1, 2023

Fig. 1: A sample view from the mitmproxy tool. Network traffic with HTTP requests is shown.

of data being shared, and did some investigation on the domains to whom it was being sent, trying to identify the company behind the servers, as well as the probable cause for the traffic (e.g. Google Analytics, an elaborate data collection scheme or something else). Second, we read the privacy policies of all 15 applications and compared the observed network traffic to the data collection disclosed in these texts. Third, we then also compared these two to the Google Play Data Protection section, where developers must disclose whether any data in their application is being sent to third parties.

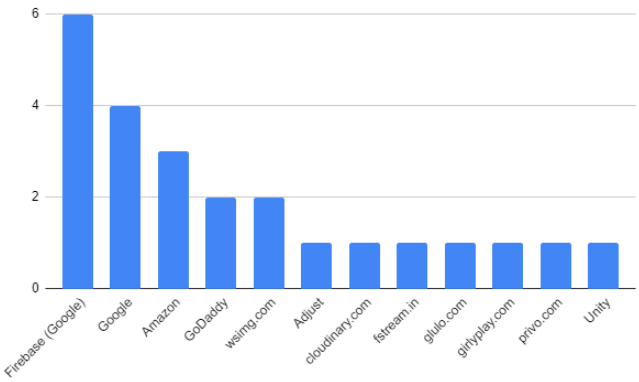

Fig. 2: The third-party analytics services found in the studied applications and number of occurrences (counted once per application).

III. Results

A. Findings from the network traffic analysis

The third parties receiving traffic from the analyzed applications are shown in Figure 2. Firebase, an app development platform provided by Google, is the most frequent receiver of personal data, followed by Google and Amazon.

The personal data items the studied applications sent to third parties are shown in Table I. The most frequently leaked pieces of personal data were the device's IP address and information on the phone's brand and OS. While the IP address is always sent out with every network connection an application forms, two of the analyzed applications did not connect to any third parties. Therefore, the IP address, as well as information about the phone and OS were leaked in 13 out of 15 cases.

TABLE I: The personal data items sent to third-party services by the studied mobile applications.

Sent data	Number of applications	Percentage
IP address	13	86.7
Phone brand and model	13	86.7
Phone OS	13	86.7
Phone OS version	13	86.7
Timestamp	5	33.3
Screen size	2	13.3
Processor	2	13.3
Is the phone rooted?	1	6.7
App store	1	6.7
Internet service provider	1	6.7
User identifier	1	6.7
Device identifier	2	13.3
Installation identifier	6	40.0
Instance identifier	4	26.7
Session identifier	1	6.7
Other unknown identifier	6	40.0
Timezone	2	13.3
Country	4	26.7
Language	4	26.7
Child's first name	1	6.7
Email address	1	6.7
Use of camera	1	6.7
Use of microphone	1	6.7

While the device IP address is sent to all third parties a mobile application communicates with, many users may not know that a unique address, which can be used to identify them, is leaked to third parties and analytics services. Although IP addresses of home users are usually dynamic, it is common that the same IP address remains in the possession of the same device and user for time periods exceeding one month [14]. Moreover, large global analytics companies can be expected to be able to effectively connect even dynamic IP addresses to specific users.

Other frequently sent data items include numerous technical details such as screen size and processor, revealing additional information which can be used to profile users. In one case, even a sensitive piece of information on whether the phone has been rooted was leaked. We suspect that this is part of the functionality of the development platform and not a feature developers have intended to add. The findings also included leaked contextual information such as the app store the application was downloaded

from, user's country, language, timezone and internet service provider.

The technical and contextual data items mentioned before cannot be used to identify a person by themselves (except the IP address), but when combined together, they can be very useful when generating a profile for a specific user. This is especially the case when large analytics companies combine pieces of data from different mobile applications and websites.

Then again, several device or user specific identifiers shown in the list are usually very personal. Our findings also include many unknown identifiers. These are long strings that are probably used to identify users, devices or sessions. Unknown identifiers were sent out by 40% of the analyzed applications. One application also sent out the child's first name and email address, although this information was explicitly requested from the user on registration. Finally, one of the tested applications was also found to transmit data about the user's actions, revealing the use of camera and microphone.

B. Findings from the analysis of privacy policies and data protection sections

Table II shows whether the applications ask for parental consent and how they announce their target groups. We can see that none of the studied applications ask for a consent for collecting data from a parent or guardian. Still, 13 applications collected personal data, and almost all of the applications indicated (in the app name, on their product page, in the application itself or on a separate linked resource such as privacy policy) that the application is targeted for children. Only one of the applications, Indian Royal Wedding Game, did not clearly indicate that the application was targeted for children. In all cases, however, the target audience is quite clear from the Google Play age group as well as the contents of the application.

Only two of the studied applications, "Drawing Games: Draw and Color for Kids" and "Kids Carwash Service Auto Workshop" did not collect any data. It is quite clear that other applications, by collecting data on children while not requiring a consent, are violating GDPR.

Table III shows how the studied privacy policies and the data protection section in Google Play correspond to actual network traffic containing personal data. We can see that in several (5) cases, the data protection section in Google Play Store does not correspond to the actual network traffic. There are 4 applications that claim they do not collect personal data or deliver it to third parties despite doing so. Interestingly, there is also one application that reports collecting personal data but did not do so in our experiments.

On the other hand, we can see privacy policy documents usually mentioned that data was being collected and sent to third parties – although they did not always list the transmitted personal data items or receiving parties in sufficient detail. Also, several applications (4) did not have proper privacy policies at all, showing quite obvious disregard for privacy regulations. Privacy policies of two applications – "Tonka: Trucks Around Town" and "Color by Numbers: Cars" – claimed not to collect any user information at all, but at least IP addresses could still be collected by the developer and the associated third-party server.

Interestingly, developers of two applications, "Kids Art & Drawing Game" and "Indian Royal Wedding Game", seemingly try to wriggle out of their responsibilities by including statements such as "our services are not intended for use by children under the age of sixteen (16)" and "these services do not address anyone under the age of 13" in their privacy policies. It is obvious that this strongly contradicts the age group and the contents of the applications.

IV. Discussion

A. Key findings

The key findings of our study can be summarized as follows:

- Of the 15 studied applications, 13 sent out personal data to third parties, identifying the child using the application. This information can also be used to profile the user and their behavior, especially when other applications send information to the same third party, enabling them to build a more comprehensive picture of what applications the user uses and how. Because the user is a child, implications can be even more serious.
- None of the 13 applications sending personal data to third parties asked for parental consent. This violates the GDPR.
- When it comes to personal data delivered to third parties, there were significant discrepancies between 1) the studied application's data safety sections in Google Play, 2) the associated privacy policies, and 3) the recorded network traffic.
- Some application developers may try to escape from their responsibility by claiming in their privacy policy documents that their application is not intended for children. This blatantly contradicts the age range given for the applications in Play Store.

B. Implications for software development

It is apparent that the third-party libraries and platforms developers have chosen to use when building their application are a significant factor in what kind of personal data and technical details are sent out to third parties. Oftentimes, it seems developers and data protection officers do not pay sufficient attention to this side effect of using third-party libraries and analytics services, and the task of finding out how the used external services handle personal data is left to the user of the application.

Software companies and developers are, however, responsible for reporting the personal data items and the parties they are delivered to in the application's privacy

TABLE II: Consent and target groups of the applications.

			Does the developer indicate that the app is for children?			
Age group	Application	Parental consent	In app name	On product page	In the app	Linked file/website
0–5	Kids Art & Drawing Game	No	X	X		
	Coloring book - games for kids	No	X	X	X	X
	Drawing Games: Draw and Color for Kids	No	X	X		X
	Indian Royal Wedding Game	No				
	Paint for Kids	No	X	X		
6–8	Daddys Hair Salon	No				
	Jigsaw Puzzles: Games for Kids	No	X	X	X	X
	Kids Carwash Service Auto Workshop: Fun Game	No	X	X		
	Miga Town: My TV Shows	No				X
	Toca Hair Salon 4	No			X	X
9–12	ChatterPix Kids by Duck Duck Moose	No No	X		X	
	Color by Numbers: Cars	No		X	X	X
	Kids Coloring Book	No	X	X	X	X
	Montessori Nature	No		X	X	X
	Tonka: Trucks Around Town	No		X	X	X

TABLE III: Transparency of privacy policies and data protection section compared to actual network traffic.

Application	Sends personal data to 3rd parties	According to the Google Play Data Protection section, is data sent to 3rd parties	According to the privacy policy, is data sent to 3rd parties
Kids Art & Drawing Game	Yes	No	Yes
Coloring book - games for kids	Yes	No	N/A
Drawing games: Draw and Color for kids	No	No	No
Indian Royal Wedding Game	Yes	Yes	Yes
Paint for kids	Yes	No	Yes
Daddys hair salon	Yes	No	N/A
Jigsaw Puzzles: Games for Kids	Yes	Yes	Yes
Kids Carwash Service Auto Workshop: Fun Game	No	Yes	N/A
Miga Town: My TV Shows	Yes	Yes	Yes
Toca Hair Salon 4	Yes	Yes	Yes
Chatterpix Kids by Duck Duck Moose	Yes	Yes	Yes
Color by Numbers: Cars	(developer uses a 3rd party server)	N/A	No
Kids Coloring Book	Yes	Yes	Yes
Montessori Nature	Yes	N/A	N/A
Tonka: Trucks Around Town	(developer uses a 3rd party server)	N/A	No

policy document. That is why analyzing applications' network traffic should always be an integral part of the software development and testing process. This will help the developers to monitor what kind of personal data their mobile applications leak out to third parties. This knowledge is necessary to compile realistic and accurate data safety sections and privacy policy documents.

Privacy policy documents could also do much better when informing users about potential consequences of transferring information to a third party. This does not mean scaring users with unrealistic doomsday scenarios. It would be beneficial, however, to lay out some examples of how technical data items sent to analytics services can be used to profile children, for instance. In general, a potential method to improve privacy policy documents could be introducing a small number of different templates or structures that privacy policy documents would have to follow. Following a standardized format helps to include the necessary information (such as personal data items and their destinations) sufficient detail makes privacy policies easier to write and understand [15].

The discrepancies between network traffic, data safety sections and privacy policies may also speak of a gap in understanding between software developers and data officers. It is not only important to understand what kind of data the application sends out but also have a common appreciation of what constitutes personal data. With this renewed understanding, it is also worth reading the privacy policies of the used third-party libraries and reconsidering whether it is necessary to include analytics services in a mobile application aimed for children.

Google Play Store should also take steps to better ensure applications targeted for children are GDPR compliant. Applications that send personal data to third parties without consent should be disallowed. Also, when developers of applications clearly intended for kids declare their applications are designed for adults, Google Play Store seems to simply look the other way. Because the tech giants behind the app stores have the real power over the app economy, they should enforce stricter privacy.

C. Limitations

There are few limitations in the current study. First, the network traffic analysis we performed only reveals clearly detectable personal data mobile applications transmit. The analyzed traffic may contain some data items that have been obscured and purposefully made difficult to uncover to avoid detection or hide implementation details.

Second, while we can be certain that specific personal data items have been delivered to a specific destination, we cannot say what happens to this data (e.g. an IP address) once it reaches a third party server. The data can either be stored and possibly made use of, or immediately deleted. The network traffic analysis carried out in this study only covers the client side. Our analysis also is not exhaustive in terms of functionality available in the selected applications, despite our best efforts to cover the most essential functions.

Third, we only looked at applications in the Google Play Store, and it remains unclear whether similar findings would emerge in other Android app marketplaces such as Samsung Galaxy Store, and further in app stores for alternative operating systems, primarily the iOS App Store. Thus, so far our findings merely speak about the status quo of network traffic in children's apps in the Kids section of Google Play Store, and should be understood as such. However, the findings do raise broader and more critical questions regarding modern ecosystems and user privacy.

V. Conclusion

In this study, we demonstrated via network traffic analysis that several applications targeted for children, available in the world's most popular application marketplace, send data to third parties without parents' content, which violates the GDPR. We also found that there are discrepancies in what application developers disclose as part of the Google Play Data Protection section and the applications' privacy policies. Our network traffic analysis further showed that this information also does not align with what data is actually being sent and where, primarily in that more data was being shared than what was officially disclosed.

These findings have significant implications for both practice and research. On the practical side, users, and particularly the parents and guardians of children, should be aware that the applications they use and trust with their personal information, and the information of their offspring, may not be as privacy-friendly as they believe. Our work suggests that even if the parents and guardians read the data protection sections and privacy policies carefully, they may not get an accurate picture of what the applications really do.

On the research side, this study opens the door for further investigation into the extent of data leakage in applications and the impact it has on user privacy. It also highlights the need for further regulations and guidelines to ensure that application developers are transparent about their data collection practices and that user data is protected. Furthermore, as developer tools operate at higher and higher abstraction levels, it may be increasingly difficult for developers to keep track of whether there are some unintended functionalities embedded in some of the blocks of code they use, in our case, unintended network traffic. Overall, this study sheds light on the urgent need for greater awareness and action on the issue of data privacy in the age of digital applications, particularly for audiences in vulnerable positions, such as children.

References

[1] S. Laato, R. Lindberg, T. H. Laine, P. Bui, B. Brezovszky, L. Koivunen, O. De Troyer, and E. Lehtinen, "Evaluation of the pedagogical quality of mobile math games in app marketplaces," in *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, 2020, pp. 1–8.

[2] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Multi-classification approaches for classifying mobile app traffic," *Journal of Network and Computer Applications*, vol. 103, pp. 131–145, 2018.

[3] I. Ullah, R. Boreli, and S. S. Kanhere, "Privacy in targeted advertising on mobile devices: a survey," *International Journal of Information Security*, pp. 1–32, 2022.

[4] F. Zhao, S. Egelman, H. M. Weeks, N. Kaciroti, A. L. Miller, and J. S. Radesky, "Data collection practices of mobile applications played by preschool-aged children," *JAMA pediatrics*, vol. 174, no. 12, pp. e203 345–e203 345, 2020.

[5] M. Shumanov, H. Cooper, and M. Ewing, "Using ai predicted personality to enhance advertising effectiveness," *European Journal of Marketing*, vol. 56, no. 6, pp. 1590–1609, 2022.

[6] T. Chen, I. Ullah, M. A. Kaafar, and R. Boreli, "Information leakage through mobile analytics services," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, 2014, pp. 1–6.

[7] S. Rauti and S. Laato, "Location-based games as interfaces for collecting user data," in *Trends and Innovations in Information Systems and Technologies: Volume 2 8*. Springer, 2020, pp. 631–642.

[8] M. Liu, H. Wang, Y. Guo, and J. Hong, "Identifying and analyzing the privacy of apps for kids," in *Proceedings of the 17th international workshop on mobile computing systems and applications*, 2016, pp. 105–110.

[9] S. Livingstone, M. Stoilova, and R. Nandagiri, "Children's data and privacy online: growing up in a digital age: an evidence review," 2019.

[10] R. Binns, U. Lyngs, M. Van Kleek, J. Zhao, T. Libert, and N. Shadbolt, "Third party tracking in the mobile ecosystem," in *Proceedings of the 10th ACM Conference on Web Science*, 2018, pp. 23–31.

[11] I. Reyes, P. Wijesekera, J. Reardon, A. Elazari Bar On, A. Razaghpanah, N. Vallina-Rodriguez, S. Egelman *et al.*, ""won't somebody think of the children?" examining coppa compliance at scale," in *The 18th Privacy Enhancing Technologies Symposium (PETS 2018)*, 2018.

[12] Statista, "Annual number of app downloads from the google play store worldwide from 2016 to 2021," *ONLINE, available at: https://www.statista.com/statistics/734332/google-play-app-installs-per-year/*, vol. 1, 2021.

[13] D. Fernández Galeote and J. Hamari, "Game-based climate change engagement: analyzing the potential of entertainment and serious games," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CHI PLAY, pp. 1–21, 2021.

[14] V. Mishra, P. Laperdrix, A. Vastel, W. Rudametkin, R. Rouvoy, and M. Lopatka, "Don't count me out: On the relevance of ip address in the tracking ecosystem," in *Proceedings of The Web Conference 2020*, 2020, pp. 808–815.

[15] M. Rowan and J. Dehlinger, "A privacy policy comparison of health and fitness related mobile applications," *Procedia Computer Science*, vol. 37, pp. 348–355, 2014.