Technology and Regulation

# How Decisions by Apple and Google obstruct App Privacy

Konrad Kollnig*, Nigel Shadbolt**

app stores, Digital Services Act, privacy, data protection, online platforms, Apple, Google, iOS, Android, apps

konrad.kollnig@maastrichtuniversity.nl
nigel.shadbolt@cs.ox.ac.uk

**Ample past research highlighted that privacy problems are widespread in mobile apps and can have disproportionate impacts on individuals. However, doing such research, especially through automated methods, remains hard and has become an arms race with those who engage in invasive data practices. This paper analyses how decisions by Apple and Google, the makers of the two primary app ecosystems (iOS and Android), currently hold back (automated) app privacy research and thereby create systemic risks that have previously not been systematically documented. Such an analysis is timely and pertinent since the newly enacted EU Digital Services Act (DSA) obliges Very Large Online Platforms to enable 'vetted researchers' to study systemic risks (Article 40) and to put in place reasonable, proportionate and effective mitigation measures against systemic risks (Article 35).**

## 1. Introduction

Ample previous research has found that privacy problems in mobile apps are and remain widespread.[1] There have been some legislative counter initiatives, such as the EU General Data Protection Regulation (GDPR)[2] and the California Consumer Privacy Act of 2018 (CCPA),[3] but many of the known problems persist. For example, a recent study from researchers at the University of Oxford" with "our research group recently showed that the extent to which apps can communicate with Google, Facebook and other tracking companies has largely remained unchanged since the GDPR came into force in 2018.[4]

Outside of China, the two main app ecosystems are those by Apple and Google: iOS and Android. In governing their ecosystems, both companies follow somewhat different strategies. This is arguably rooted in the business model of each company.

Google generates most of its revenues from ads. Mobile apps are central to this business model, both as a platform for advertising and as a source of personal and behavioural data. On this basis, Google built an extremely lucrative digital advertising company:[5] in 2020 alone, the parent company of Google, Alphabet, generated an estimated $147bn (80%) of its revenue from advertising,[6] most coming from mobile devices. Google's reliance on advertising has arguably had the direct result that Google has tended to be more lenient about apps' practices on Android, especially when it comes to data (since Google's business relies on access to vast amounts of data).[7] This,

1 Joel Reardon and others, '50 Ways to Leak Your Data: An Exploration of Apps' Circumvention of the Android Permissions System', *28th USENIX Security Symposium (USENIX Security 19)* (USENIX Association 2019) <https://www.usenix.org/conference/usenixsecurity19/presentation/reardon>; Reuben Binns and others, 'Third Party Tracking in the Mobile Ecosystem', *Proceedings of the 10th ACM Conference on Web Science - WebSci '18* (ACM Press 2018) <http://dl.acm.org/citation.cfm?doid=3201064.3201089> accessed 14 February 2020; Paul Vines, Franziska Roesner and Tadayoshi Kohno, 'Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob', *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society* (ACM Press 2017) <http://dl.acm.org/citation.cfm?doid=3139550.3139567> accessed 14 February 2020.

2 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, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016 OJ L 119/1.

3 California Consumer Privacy Act of 2018.

* Konrad Kollnig is Assistant Professor in the Faculty of Law of Maastricht University.
** Nigel Shadbolt is Professor of Computing Science in the Department of Computer Science of the University of Oxford.

Received 25 Oct 2022, Accepted 28 Jun 2023, Published 15 Sep 2023

4 Konrad Kollnig and others, 'Before and after GDPR: Tracking in Mobile Apps' (2021) 10 *Internet Policy Review* <https://policyreview.info/articles/analysis/and-after-gdpr-tracking-mobile-apps> accessed 21 December 2021.

5 Competition and Markets Authority, 'Online Platforms and Digital Advertising' (2020) <https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf> accessed 28 October 2022.

6 Alphabet, 'Form 10-K' (2020) <https://abc.xyz/investor/static/pdf/20210203_alphabet_10K.pdf> accessed 24 October 2022.

7 Daniel Greene and Katie Shilton, 'Platform Privacies: Governance, Collaboration, and the Different Meanings of "Privacy" in iOS and Android

in turn, contributes to the fact that privacy-invasive mobile apps on Android remain widespread.

Meanwhile, Apple has increasingly been building a unique selling point for its iOS ecosystem around privacy.[8] Apple can do so because it mainly relies on device sales, and not data-driven digital ads. Apple's reliance on device sales has encouraged the company to pursue a strategy of vertical integration, in which it controls most aspects of the value chain. To maintain its control over this ecosystem, Apple widely uses closed-source and proprietary technologies on iOS. This has, in the past, severely restricted independent analysis of Apple's privacy claims. For example, the last large-scale study into app privacy on iOS was conducted in 2013,[9] until the recent release of our paper on this subject in 2022.[10] Since iOS research remains difficult,[11] there currently exists much more research on Android. In the near future, it seems that Apple will increasingly move into the ad space and become a data company because it faces pressure (e.g. by the right to repair movement) to further increase revenues, move away from its reliance on device sales and diversify its revenue stream,[12] thereby potentially heightening privacy concerns in app ecosystems.

While Android – in principle – adopts an open-source strategy, Google has also been restricting researchers' capabilities to study Android app privacy over recent years; we will discuss this later in this paper. Researchers have previously struggled to assess important questions, too, in the Google Play ecosystem. For example, a pre-print from May 2022 argued that the introduction of the GDPR caused the disappearance of a third of apps on the Google Play Store.[13] Our research group cast doubt over the validity of these claims.[14] We highlighted that this pre-print had actually not considered the content moderation by Google on the Play Store and that the Apple App Store had not seen a similar decline; only with additional data provided by Google on its removals of apps could the impact of the GDPR on apps genuinely be assessed. This debate further underlines the current challenges of studying important research questions (like the material impacts of the GDPR) in app ecosystems.

It is widely accepted that transparency is a key facet of privacy. Despite this, designing and conducting studies into app privacy remains hard and only done on occasion. Without such studies and privacy audits, end-users continue to have limited (and often outdated or inaccurate) information at hand to make decisions on what apps and what smartphones to use. It also makes the work of data protection authorities more difficult because they will not be able to keep up with the scale of the app ecosystem without robust app analysis tools. This paper analyses what currently makes research on app privacy harder than necessary. Specifically, we focus on how conduct and technical design decisions (whether intentional or not) by Google in Android and by Apple in iOS currently hold back privacy research, and what should change.

One might argue that the focus on app stores in this work is unwarranted and that one should instead focus on the obligations of app developers. Yet, it is well known from previous research that app developers often struggle with their legal obligations, especially when it comes to data protection and privacy,[15] and that – at least for now – it is insufficient to rely on the information provided by app developers through their privacy notices and other means.[16] Indeed, previous research highlighted that the providers of app stores play an important role in regulating app privacy. Greene and Shilton found from engaging with app developers in 2018 that the providers of app stores might be able to move quicker than the relevant authorities as regards privacy problems in apps.[17] Considering this, they argued that there needs to be greater transparency around platform governance and enforcement of privacy rules. Similarly, Van Hoboken and Ó Fathaigh argued in 2021 that Google and Apple increasingly act as important regulators of data protection and privacy, but with limited regulation, oversight, and accountability.[18] To increase transparency, these authors argued for mandatory disclosures about the privacy-related activities of smartphone platforms – as a minimally invasive but realistic intervention. In short, the providers of app stores are uniquely positioned to help data protection and other laws scale across millions of mobile apps, due to their central role in the app ecosystem and their leading expertise within their own operating system.

Our authority to speak about these topics emerges from studying privacy, compliance and challenges to fundamental rights in app ecosystems for many years. Through our research, we have previously made important contributions to the development of methodology and technology in this field, and regularly engage with relevant regulators, civil society, and the interested public. We also create a range of educational materials, so as to raise awareness of privacy challenges in digital ecosystems.

Development' (2018) 20 *New Media & Society* 1640.

8 Kelly D Martin and Patrick E Murphy, 'The Role of Data Privacy in Marketing' (2017) 45 *Journal of the Academy of Marketing Science* 135.

9 Yuvraj Agarwal and Malcolm Hall, 'ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing', *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (ACM Press 2013) <http://dl.acm.org/citation.cfm?doid=2462456.2464460> accessed 14 February 2020.

10 Konrad Kollnig and others, 'Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps' (2022) *Proceedings on Privacy Enhancing Technologies Symposium* <https://petsymposium.org/popets/2022/popets-2022-0033.php> accessed 24 October 2022.

11 Sebastian Zimmeck and others, 'MAPS: Scaling Privacy Compliance Analysis to a Million Apps' (2019) *Proceedings on Privacy Enhancing Technologies Symposium* 66; Binns and others (n 1).

12 Benjamin Mayo, 'Hiring Trends Indicate Apple Plans to Significantly Expand Its Ads Business' *9to5Mac* (3 August 2022) <https://9to5mac.com/2022/08/03/apple-ads-expansion/> accessed 24 October 2022.

13 Rebecca Janßen and others, 'GDPR and the Lost Generation of Innovative Apps' (National Bureau of Economic Research 2022) <http://www.nber.org/papers/w30028.pdf> accessed 23 September 2022.

14 Konrad Kollnig and Reuben Binns, 'The Cost of the GDPR for Apps? Nearly Impossible to Study without Platform Data' <http://arxiv.org/abs/2206.09734> accessed 11 July 2022.

15 Anirudh Ekambaranathan, Jun Zhao and Max Van Kleek, '"Money Makes the World Go around": Identifying Barriers to Better Privacy in Children's Apps From Developers' Perspectives', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (ACM Press 2021) <https://doi.org/10.1145/3411764.3445599> accessed 27 October 2022; Sean Sirur, Jason RC Nurse and Helena Webb, 'Are We There Yet?: Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR)', *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security* (ACM Press 2018) <http://dl.acm.org/citation.cfm?doid=3267357.3267368> accessed 14 February 2020; Abraham H Mhaidli, Yixin Zou and Florian Schaub, '"We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks' (2019) *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*.

16 Zimmeck and others (n 11).

17 Greene and Shilton (n 7).

18 Joris van Hoboken and R w Ó Fathaigh, 'Smartphone Platforms as Privacy Regulators' (2021) 41 *Computer Law & Security Review* 105557.

Lastly, it is important to highlight that our observations do not only relate to privacy-related research, but also other domains. For the sake of argument however, we confine the discussion mostly to data protection and privacy issues. Indeed, less technical disciplines might be particularly affected by platform measures that make research on apps more difficult. For example, the authors of one economics paper on the Google Play Store criticised the currently "extremely significant data collection effort, which was very intense as the collection process could only be performed manually for most of the variables of interests and the controls".[19]

The rest of this paper is organised as follows. We briefly review other academic work on app ecosystems and their regulation in Section 2. We then introduce past academic literature on automated app analysis in Section 3. Next, we present our findings in Section 4. Finally, we draw conclusions in Section 5.

## 2. Current App Store Regulation

The centrality of app stores makes them a target for effective regulation. Yet, such regulation has so far been limited.[20] The Federal Trade Commission (FTC) established some baseline rules for app stores in 2013. They strongly encouraged app stores to require just-in-time consent for sensitive data access, to seek privacy policies from app developers, and to implement system-wide opt-out mechanism from data collection.[21] Despite not being law, Google and Apple followed many of the recommendations, and have not seen further public recommendations from the FTC since.

In the EU and UK, there existed limited targeted regulation of app stores until recently. The Regulation on platform-to-business relations (P2BR)[22] contains general provisions for online intermediaries, including app stores, but does little to enact better privacy protections.[23] Data protection laws, such as the GDPR and the ePrivacy Directive,[24] arguably place the primary responsibility for data protection with the app developers, not usually with app store providers – although this is subject to ongoing debate (specifically, the concept of 'joint controllership' over software development processes).

App stores also face increasing scrutiny by courts and regulators. In the case *Epic Games v Apple Inc*[25] running since 2020, a US District Court judge largely found no monopolistic behaviour of Apple, but did identify some anticompetitive conduct in Apple's business practices. The judge ordered Apple to allow app developers to inform app users of alternative payment methods. Both Apple and Epic Games have appealed the ruling. In the EU, following a complaint of Spotify against Apple from 2019, the European Commission opened formal proceedings against Apple and identified multiple anticompetitive aspects about the company's app ecosystem – the case is, however, still ongoing.[26] In January 2022, the Dutch competition authority demanded changes from Apple to its App Store policies.[27]

The challenges in keeping up with regulation of online platforms have spurred a recent countermovement by lawmakers. In South Korea, parliament amended the *Telecommunication Business Act* to force app stores to allow alternative payment methods and reduce commissions. In response, Apple lowered the share it takes from App Store revenues of small developers (making less than $1 million per year) from 30% to 15%. In the US, Congress is debating a new *Open App Markets Act* that aims to address common competition concerns around app stores and passed the Senate Judiciary Committee with a strong 20—2 bipartisan vote in February 2022. Yet, there has been no further publicly documented progress since, in part because of heavy lobbying by the tech industry.[28]

In the EU, in late 2022, lawmakers adopted two new pieces of legislation that aim to improve the regulation of digital markets, the Digital Markets Act (DMA)[29] and the Digital Services Act (DSA).[30] With the DSA, the EU revised the rules of the e-Commerce Directive from 2000, motivated by drastic changes to the online environment over the past 20 years. Online platforms and search engines now have increasing influence over our day-to-day lives, and so the DSA amended the e-Commerce Directive such that online platforms and search engines face more explicit rules. Micro- or small enterprises are exempt from many of the obligations.[31] Meanwhile, very large online platforms (VLOPs) face a set of additional rules.[32] These VLOPs are those online platforms that have a number of average monthly active recipients of the service in the Union equal to or higher than 45 million.[33] These obligations for VLOPs include giving 'vetted researchers' to study 'systemic risks',[34] among others. The European Commission recently clarified that both app stores, those by Google and Apple, classify as VLOPs and thus fall into the category with the strictest measures under the DSA.[35]

19 Paolo Roma and Daniele Ragaglia, 'Revenue Models, in-App Purchase, and the App Performance: Evidence from Apple's App Store and Google Play' (2016) 17 *Electronic Commerce Research and Applications* 173.
20 van Hoboken and Ó Fathaigh (n 18); R Ó Fathaigh and J van Hoboken, 'European Regulation of Smartphone Ecosystems' (2019) 5 *European Data Protection Law Review* 476.
21 Federal Trade Commission, 'Mobile Privacy Disclosures: Building Trust Through Transparency' (2013) <https://www.ftc.gov/sites/default/files/documents/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission-staff-report/130201mobileprivacyreport.pdf>.
22 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services 2019 OJ L 186/57.
23 Ó Fathaigh and van Hoboken (n 20).
24 Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws 2009 OJ L 337/11.
25 *Epic Games v Apple Inc* 493 F Supp 3d 817 (ND Cal 2020).
26 *Apple* (Case AT.40437-Apple-App Store Practices (music streaming)) *Commission Proceedings*.
27 *Autoriteit Consument en Markt v Apple* [2021] ROT 21/4781 and ROT 21/4782.
28 Taylor Giorno, 'Big Tech Lobbying Push Helped Block Bipartisan Bills That Aimed to Curb Alleged Anti-Competitive Behavior' (*OpenSecrets News*, 20 December 2022) <https://www.opensecrets.org/news/2022/12/big-tech-lobbying-push-helped-block-bipartisan-bills-that-aimed-to-curb-alleged-anti-competitive-behavior/> accessed 27 January 2023.
29 Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) 2022 OJ L 265/1.
30 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) 2022 OJ L 277/1.
31 DSA, arts 15(2), 19 and 29.
32 DSA, ch III sec 5.
33 DSA, art 33.
34 DSA, art 40.
35 European Commission, 'DSA: Very Large Online Platforms and Search Engines' (*European Commission - European Commission*, 25 May 2023) <https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413>

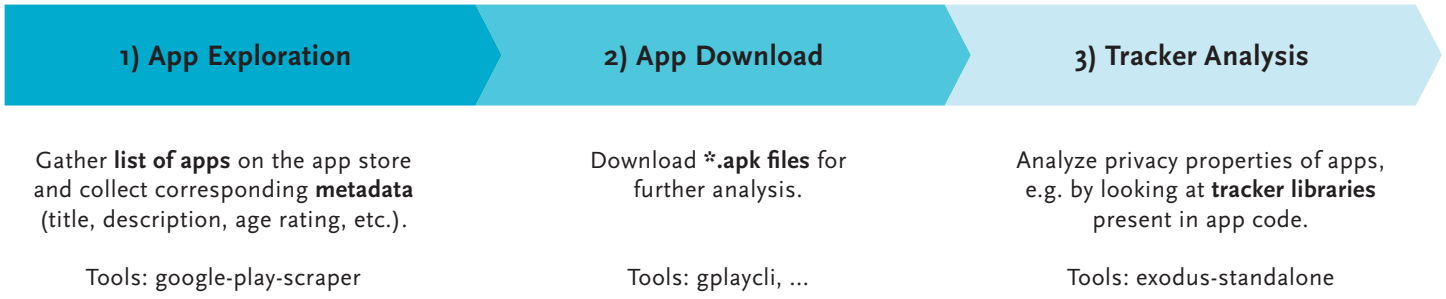


**Figure 1:** Typical steps for static tracking library analysis of an Android app. This is usually supported through a range of open-source tools. On iOS, much fewer tools exist for each step.

Beyond the DSA, the DMA brings further obligations for the providers of app ecosystems, including the ability for end-users to install apps from alternative app stores (which is currently difficult on iOS).

## 3. Automated Methods for the Study of App Privacy

Privacy problems in apps are a well-known risk. This is why we first review the relevant literature that has studied such risks. This literature review enables us, in the rest of this paper, to describe common challenges faced in mobile privacy research, and how design decisions and conduct by app platforms make this research more difficult to pursue.

We restrict this literature review to *automated methods* for the study of app privacy since only these approaches can scale with the vastness of the app stores. Automated methods are thus essential to the study of risks in app stores. For example, this excludes interview- or survey-based studies about app ecosystems, which are difficult to scale across millions of apps and developers.

Overall, there has been a wealth of past research that analysed privacy in mobile apps. There emerged two main methods for doing so: dynamic and static analysis.

The usual steps for app download and analysis are visualised in Figure 1. A usual analysis comprises 1) an exploration and selection of apps for further analysis (including the scraping of app metadata, like title and app description), 2) the download of selected apps (either as an *.apk on Android or a *.ipa file on iOS), and 3) the analysis of the downloaded app packages.

### 3.1 Dynamic Analysis

Dynamic analysis investigates the run-time behaviour of apps by executing them on a real smartphone operating system and observing their data practices. Early research focused on *OS instrumentation*, i.e. modifying Android[36] or iOS.[37] Enck et al. modified Android so that sensitive data flows through and off the smartphone could be monitored easily.[38] Agarwal and Hall modified iOS so that users were asked for consent to the usage of sensitive information by apps,[39] before the introduction of run-time permissions by Apple in iOS 6 in 2012.

With growing complexity of mobile operating systems, recent work has shifted to *network traffic analysis*. For example, Ren et al. instrumented the VPN functionality of Android, iOS, and Windows Phone to expose leaks of personal data over the Internet.[40] Conducting a manual traffic analysis of 100 Google Play and 100 iOS apps, they found regular sharing of personal data in plain text, including device identifiers (47 iOS, 52 Google Play apps), user location (26 iOS, 14 Google Play apps), and user credentials (8 iOS, 7 Google Play apps). Van Kleek et al. used dynamic analysis to expose unexpected data flows to users and design better privacy indicators for smartphones.[41]

There has been an increasing focus on regulatory issues over recent years. Reyes et al. used dynamic analysis to assess the compliance of children's apps with COPPA,[42] a US privacy law to protect children. Having found that 73% of studied children's apps transmit personal data over the Internet, they argued that none of these apps had obtained the required 'verifiable parental consent' because their automated testing tool could trigger these network calls, and a child could likely do so as well. Okoyomon et al. found widespread data transmissions in apps that were not disclosed in apps' privacy policies, and raised doubts about the efficacy of the commonly used notice & choice regime in privacy.[43] Kollnig et al. observed that most apps on the Google Play Store use third-party tracking, but few retrieve the legally required user consent (less than 3.5%).[44] Despite the increas-

accessed 9 June 2023.

36 William Enck and others, 'TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones', *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation* (2010).

37 Agarwal and Hall (n 9).

38 Enck and others (n 36).

39 Agarwal and Hall (n 9).

40 Jingjing Ren and others, 'ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic', *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (ACM Press 2016) <http://dl.acm.org/citation.cfm?doid=2906388.2906392> accessed 14 February 2020.

41 Max Van Kleek and others, 'Better the Devil You Know: Exposing the Data Sharing Practices of Smartphone Apps', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (ACM Press 2017) <http://dl.acm.org/citation.cfm?doid=3025453.3025556> accessed 14 February 2020.

42 Irwin Reyes and others, '"Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale' (2018) *Proceedings on Privacy Enhancing Technologies Symposium*.

43 Ehimare Okoyomon and others, 'On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies', *Workshop on Technology and Consumer Protection (ConPro '19)* (2019).

44 Konrad Kollnig and others, 'A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps' (2021) *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*.

ing focus on policy, no explicit analysis of hurdles to (automated) app privacy research exists so far.

Dynamic analysis is largely device-independent, but it can easily give incomplete results, if not all privacy-relevant aspects of an app are observed during the analysis. It also does not scale well across a large number of apps, because every app must be executed individually. Recent versions of Android made such analysis even more difficult, as we will discuss in Section 4.2.2.

### 3.2 Static Analysis

Static analysis dissects the behaviour of apps without executing them. Usually, apps are decompiled (i.e. the low-level CPU instructions within a downloaded app are turned into a more easy-to-understand program code), and the obtained program code is analysed.[45] The key benefit of static analysis is that it can analyse apps quickly, allowing it to scale to millions of apps.[46]

Egele et al. developed an iOS decompiler and analysed 1,407 iOS apps in 2011. They found that 55% of those apps included third-party tracking libraries.[47] 38.2% of apps could share data with Google Ads. Viennot et al. analysed more than 1 million apps from the Google Play Store in 2014, and monitored the changing characteristics of apps over time.[48] They found a widespread presence of third-party tracking libraries in apps (including Google Ads in 35.73% of apps, the Facebook SDK in 12.29%, and Google Analytics in 10.28%). Similarly, Binns et al. found in analysing nearly one million Google Play apps in 2018 (using a different method than Viennot et al.) that about 90% may share data with Google, and 40% with Facebook.[49] Kollnig et al. analysed 12,000 apps from the Google Play and Apple App Store each and found common compliance problems and data sharing in apps from both platforms in 2022.[50] There has been some recent research on the new privacy nutrition and data labels[51] on the app stores,[52] but there are also concerns around the accuracy of these labels.[53]

Static analysis can involve substantial computational effort and – unlike dynamic analysis – does not allow the observation of real data flows because apps are never actually run. Compared to dynamic analysis, static analysis enables the analysis of apps at much larger scale (often millions instead of thousands of apps), but may suffer from both false positives (e.g. if certain parts of the app are not run in practice, but detected as potentially privacy-invasive by the analysis) and false negatives (e.g. if apps load and execute additional invasive programme code from an external source at runtime). Despite recent advances, the static analysis of iOS apps remains much harder than on Android; we discuss reasons for this later in Section 4, e.g. the encryption of all iOS apps with Apple FairPlay DRM.

### 3.3 Systemic Risks and Data Access under the DSA

Previous scholarship underlined that there is a widespread absence of compliance with fundamental provisions of the GDPR, which aims to protect private life and personal data, among other fundamental rights. This potential GDPR infringement include widespread invasive tracking of app activities,[54] sending of personal data to the US,[55] and lack of consent implementations.[56] This points to systemic risks to exercising data protection rights and privacy in apps. These protections are especially lacking when it comes to the protection of personal data relating to children,[57] with potentially negative effect for their development.[58] This points to conflicts with the rights of the child. Furthermore, the dominance of Apple and Google in the app ecosystem might have negative effects on consumer protection, due to the imbalance of power between consumers and platforms.[59]

In light of these concerns, 'vetted researchers' will likely be able to gain data access to study 'systemic risks' from the app store platforms under Article 40 DSA. Indeed, Article 34(1) DSA explicitly mentions the right to respect for private and family life, to the protection of personal data, of the child, and to a high-level of consumer protection as needing to be protected under the DSA. The definition of systemic risks is broad, since it covers 'any' negative effects for the exercise of fundamental rights, no matter if 'actual or foreseeable'. As a result, many potential research studies into privacy, compliance, and gatekeeper power in the app ecosystem will likely classify as 'systemic risks' as defined by the DSA and will thus have to be enabled by the providers of VLOPs. However, as we will discuss in the following, the current design of app ecosystems poses significant challenges to researchers doing such research.

## 4. Impediments to App Privacy Research

We now give an overview of current impediments to app research, due to decisions by the providers of app stores. This overview emerged from a review of the relevant literature.

Based upon our long-standing expertise in the field, we created a list of some of the most important and impactful works in app privacy

45 Manuel Egele and others, 'PiOS: Detecting Privacy Leaks in iOS Applications', *Proceedings of NDSS 2011* (2011); Jin Han and others, 'Comparing Mobile Privacy Protection through Cross-Platform Applications', *Proceedings of the 2013 Network and Distributed System Security Symposium* (Internet Society 2013).

46 Nicolas Viennot, Edward Garcia and Jason Nieh, 'A Measurement Study of Google Play', *Proceedings of the 2014 ACM International Conference on Measurement and Modeling of Computer Systems* (ACM Press 2014) <http://dl.acm.org/citation.cfm?doid=2591971.2592003> accessed 14 February 2020; Binns and others (n 1).

47 Egele and others (n 45).

48 Viennot, Garcia and Nieh (n 46).

49 Binns and others (n 1).

50 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).

51 Patrick Gage Kelley and others, 'A "Nutrition Label" for Privacy', *Proceedings of the Fifth Symposium on Usable Privacy and Security* (ACM Press 2009) <http://portal.acm.org/citation.cfm?doid=1572532.1572538> accessed 10 December 2021.

52 Tianshi Li and others, 'Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (ACM Press 2022) <https://dl.acm.org/doi/10.1145/3491102.3502012> accessed 24 October 2022; Li and others.

53 Konrad Kollnig and others, 'Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels', *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM Press 2022) <https://doi.org/10.1145/3531146.3533116>.

54 Kollnig and others, 'Before and after GDPR' (n 4).

55 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).

56 Kollnig and others, 'A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps' (n 44).

57 Reyes and others (n 42); Binns and others (n 1); Ekambaranathan, Zhao and Van Kleek (n 15).

58 Jun Zhao and others, ''I Make up a Silly Name': Understanding Children's Perception of Privacy Risks Online', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (ACM Press 2019) <http://dl.acm.org/citation.cfm?doid=3290605.3300336> accessed 23 April 2020.

59 'Mobile Ecosystems Market Study Final Report' (*GOV.UK*) <https://www.gov.uk/government/publications/mobile-ecosystems-market-study-final-report> accessed 2 December 2022; Kollnig and others, 'Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels' (n 53); Kollnig and others, 'A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps' (n 44).

research. We then reviewed each of these papers for challenges faced in conducting app research, due to decisions by Apple and Google. In total, we reviewed 101 papers.

We complemented these findings by drawing on our own experience and identified further challenges. We enrich this analysis through a comparative analysis of research capabilities and challenges on iOS and Android.

We organise our findings into three categories: data collection, data analysis, and platform conduct.

Our review is likely not exhaustive and is not meant to be either. Indeed, it would be impossible to give a complete overview of those decisions because many are not known and may not even be taken intentionally. Instead, this section aims to give an overview over some of the most pressing issues and highlight that there is an issue that has not previously been covered in detail.

Ultimately, we aim to underline how providers of app stores make deliberate efforts to undermine transparency and accountability around apps' practices, and thereby create systemic risks to the individuals' exercise of fundamental rights.

## 4.1 Data Collection and Public APIs

**Lack of Public APIs.** Public APIs (i.e. programmatic access points for data and other resources) for researchers to interact with app stores could address the current problems with data collection and the reliance on open-source volunteers mentioned in the previous paragraph. Unfortunately, few such APIs exist. There do exist public APIs provided by Apple to collect metadata about the App Store, but no app downloading tools for iOS. There exist no public APIs provided by Google to collect app metadata or to download Android apps.[60] This is despite the fact that Apple and Google necessarily need to maintain APIs to provide smartphones with access to the app stores. It would be rather straightforward to open up these APIs to academic researchers. Interestingly, Apple actually provides affiliate partners with rather direct and immediate access to App Store data through its Enterprise Partner Feed. This programme, however, is not open for academic researchers. We tried to sign up for this programme as academic researchers on 21 June 2022, but have not heard back from Apple at the time of writing. Furthermore, downloading always requires an account on the app stores, which are tied to specific countries and do not get access to all apps, due to geo-blocking.[61]

As a result of the lack of public APIs, the collection of information about apps is currently not straightforward. Researchers do not even know with certainty how many and what apps exist on the app stores.[62] On Android and iOS, researchers would usually use a tool like *google-play-scraper* or *app-store-scraper* to explore what apps are on the respective app stores and collect metadata (i.e. title, description, age rating, genre, etc.) about them. In the next step, researchers would use another tool to download any identified apps. These tools are usually maintained by open-source volunteers. They tend to be fragile and rely on the continued dedication of the project maintainers. For example, the tool *app-store-scraper*, even though the best tool available for the App Store, has not seen any updates by the project maintainer over the past 2.5 years. The tool is currently partly broken because it relies on unofficial APIs. Apple and Google can change the APIs that these tools rely on at any time, and thereby break them; this happens regularly. The informality and instability of the data collection process currently acts as a deterrent to conducting app research and costs unnecessary research time.

**Restrictions on scraping.** Apple and Google currently implement countermeasures against the data collection about apps by researchers and other individuals.[63] These countermeasures include the use of CAPTCHAs (when downloading apps from Google Play), the throttling of connections based on IP and user account (on both app stores),[64] and the use of TLS fingerprinting (also when downloading from Google Play). While Google and Apple have a legitimate interest in protecting their services against abuse (e.g. DoS attacks), these measures also make legitimate research much harder. Viennot et al. paid individuals through Amazon MTurk to create legitimate Google accounts and rented servers in different locations to work around the limitations when interacting with the Play Store.[65]

To mitigate the need to obtain information directly from the app stores, third-party providers like *42matters* and *data.ai* have emerged to ease the process of obtaining information on apps. These organisations tend to charge up to tens of thousands of dollars for their services.[66] This is not usually feasible for academia. One reason for this cost being so high is that these data providers mainly cater to commercial organisations that would like to monitor their competitors in the app space. The high price also reflects the difficulty of the data collection process.

### *4.1.1 Provision of App Metadata*

**Limited insights into app ranks and installs.** A common subject of app analysis study is the evolution of ranks and install counts; this has been especially studied in literature from economics on the app ecosystem.[67] Unfortunately, Apple and Google currently provide limited insights into the app ranks and installs. Google only provides an approximation of the install count (e.g. '10,000–50,000 installs'), whereas Apple does not provide any such information. Moreover, insights into app ranks are restricted to a couple of hundreds. This is despite the fact that Apple (and potentially Google, too) maintains app ranks for larger numbers of companies and makes them available through its Enterprise Partner Feed (see Section 4.1).

**Unverified, and potentially misleading privacy labels.** Apple and Google have recently been rolling out new privacy nutrition labels, following loosely the design put forward by Kelley et al. in 2009.[68]

60 A good overview of the challenges faced when trying to download Android apps from Google Play exists in Viennot, Garcia and Nieh (n 46).
61 Kevin Allix and others, 'AndroZoo: Collecting Millions of Android Apps for the Research Community', *Proceedings of the 13th International Conference on Mining Software Repositories* (ACM Press 2016) <https://dl.acm.org/doi/10.1145/2901739.2903508> accessed 26 January 2023.
62 Pierre Laperdrix and others, 'The Price to Play: A Privacy Analysis of Free and Paid Games in the Android Ecosystem', *Proceedings of the 2022 ACM Web Conference* (ACM Press 2022) <https://doi.org/10.1145/3485447.3512279>; Viennot, Garcia and Nieh (n 46).
63 Laperdrix and others (n 62); Viennot, Garcia and Nieh (n 46).
64 Laperdrix and others (n 62).
65 Viennot, Garcia and Nieh (n 46).
66 42matters AG, 'Pricing Plans and Products' <https://42matters.com/pricing> accessed 31 January 2023.
67 Reuben Binns and others, 'Measuring Third-Party Tracker Power across Web and Mobile' (2018) 18 *ACM Transactions on Internet Technology*; Roma and Ragaglia (n 19); Wen Wen and Feng Zhu, 'Threat of Platform-owner Entry and Complementor Responses: Evidence from the Mobile App Market' (2019) 40 *Strategic Management Journal* 1336.
68 Kelley and others (n 51).

While a potentially positive development for transparency around apps' data practices, these labels are self-reported by app developers and not usually verified by Apple or Google. In fact, previous research has revealed notable discrepancies between reported and actual data practices.[69] Despite this, a range of new and emerging research studies have embarked on analysing these labels and on deriving claims about app privacy from their findings.[70] This underlines the need to disclose more clearly when and to what extent some privacy labels might be verified by Apple and Google, and when they are not.

**Lack of insights into iOS app permissions.** On both iOS and Android, permissions form a cornerstone of the security model. Certain pieces of data can only be accessed once users have given apps the permission to do so (called 'dangerous permissions' on Android and 'protected resources' on iOS); some further non-opt-in permissions exist on both Android and iOS (called 'non-dangerous permissions' on Android and 'entitlements' on iOS). All permissions are currently disclosed publicly on the Google Play Store. Meanwhile, Apple does not provide any information about permissions or entitlements on the App Store. To obtain this information, researchers instead must download each app of interest from the App Store in a laborious process.[71] Since permissions have been an important subject for privacy research on Android and gave insights into the permission (over) use of apps,[72] Apple should allow more insights into permissions and entitlements on iOS.

**No reporting of third-party libraries in individual apps, despite possessing this information.** Much previous research has focused on the use of invasive third-party tracking libraries in apps.[73] These libraries are widely used by app developers to ease the integration of ads and analytics in their apps. The most prominent example is Google Analytics, which was declared in violation of the GDPR by various data protection authorities,[74] but is still widely used in apps (by about 17%[75]). This is why the provision of information on the use of such libraries by apps on the app stores could help research and transparency efforts immensely. Unfortunately, such information is currently provided neither by the Google nor Apple app store. The community-driven projects like Exodus Privacy, ClassyShark3xodus, TrackerControl and App Warden have set out to close this important gap, but still face challenges, such as a de-facto ban on self-signed certificates (see Section 4.2.2) and code obfuscation (see Section 4.2.1). Such community tools also only exist for Android, and not on the App Store. Importantly, Google already maintains information on the use of third-party libraries in apps, and provides descriptive statistics about their use as part of the Google Play SDK Index; Google just does not currently make this information available about single apps. For Apple, as demonstrated in previous research,[76] it would be rather straightforward to obtain this information as part of its already existing app vetting process.

### 4.1.2 Access to App Packages

As highlighted above in Section 4.1, the download of apps from app stores is currently a tedious and laborious process because researchers tend to rely on unofficial, fragile community tools for app downloading. In fact, up until 2021, there did not exist any public app download tool for iOS until the release of ipatool. Partly as a result of the difficulty in obtaining app packages, the last large-scale study into app privacy on iOS was conducted in 2013,[77] until the recent release of our paper on this subject in 2022.[78] Meanwhile, Google and Apple implement heavy restrictions on app downloading at scale, as also discussed above in Section 4.1.

**Encryption of *all* iOS apps and paid Android apps.** Both Google and Apple implement measures to protect their app ecosystems against piracy. On Google Play, such protection measures are applied to paid apps only, which seems reasonable.[79] Meanwhile, Apple applies its FairPlay DRM encryption to *all* apps, even free ones. Decryption of iOS apps is possible, but relies on access to a physical device and takes time.[80] Depending on the jurisdiction, there might also exist legal challenges related to the decryption of iOS apps, since this might circumvent effective copyright protections.[81] In other words, the application of encryption to free apps – which are the most common subject of app privacy studies – drives researchers into legal grey areas.

## 4.2 Data Analysis

**Use of closed-source and proprietary technologies.** Both Apple and Google heavily rely on closed-source technologies as part of their respective app ecosystems. On Android, this involves the Google Play Store and Play Services technologies. Due to the centrality of Google in the app ecosystem, these technologies are a foundational part of the Android app ecosystem. On iOS, Apple goes a step further and applies closed-source and proprietary technologies to almost every part of their app ecosystem; limited documentation of internals

69 Kollnig and others, 'Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels' (n 53).
70 Li and others (n 52); Yue Xiao and others, 'Lalaine: Measuring and Characterizing Non-Compliance of Apple Privacy Labels at Scale' (2023).
71 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).
72 Adrienne Porter Felt and others, 'Android Permissions Demystified', *Proceedings of the 18th ACM Conference on Computer and Communications Security* (ACM Press 2011) <http://dl.acm.org/citation.cfm?doid=2046707.2046779> accessed 16 April 2020; Reardon and others (n 1); Jinseong Jeon and others, 'Dr. Android and Mr. Hide: Fine-Grained Permissions in Android Applications', *Proceedings of the second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '12* (ACM Press 2012) <http://dl.acm.org/citation.cfm?doid=2381934.2381938> accessed 19 December 2020.
73 Binns and others (n 1); Abbas Razaghpanah and others, 'Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem', *Proceedings 2018 Network and Distributed System Security Symposium* (Internet Society 2018) <https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_05B-3_Razaghpanah_paper.pdf> accessed 14 February 2020; Anastasia Shuba, Athina Markopoulou and Zubair Shafiq, 'NoMoAds: Effective and Efficient Cross-App Mobile Ad-Blocking' (2018) *Proceedings on Privacy Enhancing Technologies Symposium* 125.
74 CNIL, 'Use of Google Analytics and Data Transfers to the United States: The CNIL Orders a Website Manager/Operator to Comply' <https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply> accessed 31 January 2023; noyb, 'Austrian DSB: EU-US Data Transfers to Google Analytics Illegal' <https://noyb.eu/en/austrian-dsb-eu-us-data-transfers-google-analytics-illegal> accessed 31 January 2023.
75 Exodus, 'Google Analytics' <https://reports.exodus-privacy.eu.org/en/trackers/48/> accessed 31 January 2023.
76 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).
77 Agarwal and Hall (n 9).
78 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).
79 Han and others (n 45).
80 Damilola Orikogbo, Matthias Büchler and Manuel Egele, 'CRiOS: Toward Large-Scale iOS Application Analysis', *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices* (ACM Press 2016) <https://dl.acm.org/doi/10.1145/2994459.2994473> accessed 1 March 2021; Kai Chen and others, 'Following Devil's Footprints: Cross-Platform Analysis of Potentially Harmful Libraries on Android and iOS', *2016 IEEE Symposium on Security and Privacy* (IEEE 2016) <http://ieeexplore.ieee.org/document/7546512/> accessed 15 April 2020; Egele and others (n 45).
81 Kollnig and others, 'Are iPhones Really Better for Privacy?' (n 10).

exists, which complicates iOS research.[82] As a result, there still does not exist a universal decompiler for iOS apps; previous research only managed to decompile a subset of iOS apps.[83]

**A shift towards server-side code and less platform accountability.** Due to increasing end-user demands and regulatory scrutiny, pressure has increased on Apple and Google over recent years to implement better privacy solutions. These companies have tended to opt for centralised, proprietary solutions. One example of this is Apple's SKAdNetwork, which allows the tracking of users after the introduction of Apple's new privacy measures since iOS 14.[84] As part of this, Apple collects data about users' interactions with apps and the ads of other advertisers. This data, in turn, allows Apple to compute a privacy threshold and make sure that no personal data is leaked. The system also puts more power and data into the hands of Apple. Since Apple provided no specific information on SKAdNetwork in its privacy policy, we asked the company for more information with reference to our information rights under Article 13 GDPR. Apple took more than 8 months to provide us with more information on the SKAdNetwork system and reply to our GDPR requests relating to this system adequately; under the GDPR, companies usually have one month to provide a detailed response. This underlines that transparency in such platform-centric, proprietary solutions remains an important problem. Recently, in late 2022, the French data protection authority imposed an 8 million euro fine on Apple over these practices.[85]

Perhaps more worryingly, Google has increasingly been championing the use of server-side tagging in its Google Tag Manager technology. The Google Tag Manager allows companies to manage their tracking technologies. If server-side tagging is enabled, then part of the tracking of end-users is moved to Google's servers. This means that app privacy researchers get even less insights into apps and websites' data practices than currently because it is not clear anymore with whom data gets shared and how. Server-side tagging has been introduced in response to the increasing adoption of privacy-preserving methods in web browsers and in mobile apps – particularly in Mozilla's Firefox, the Brave browser, and Apple's devices (such as Safari's Intelligent Tracking Prevention). The widespread use of server-side tagging would make both dynamic and static analysis much more difficult, if not render it completely impossible. It is undoubtedly one of the most concerning developments in the app privacy field so far.

**Risks of the Google Privacy Sandbox for Android.** As part of the latest Android 13, Google is introducing a 'Privacy Sandbox'.[86] This Privacy Sandbox separates *certain* third-party libraries and apps. This ends the current problematic practice of permission sharing between apps and their integrated third-party libraries: a user granting a maps app the permission to read the current GPS location no longer also grants the same permission to all third-party libraries (e.g. advertising libraries) used by that maps app. This separation also allows the independent update of third-party libraries without developer intervention, leading to fewer vulnerabilities in apps due to outdated libraries. These are positive developments for app privacy.

There is, however, a risk that the introduction of the Privacy Sandbox will make independent privacy analysis even harder when the activities of third-party libraries cannot be attributed to specific apps anymore. This already happens on Android with the Google Play Services app, which facilitates user tracking for Google Ads and Analytics for other apps.[87] This makes it harder to attribute network traffic to apps, when part of the network traffic and tracking is conducted by other parts of the system, often in aggregate. This is not currently a problem because processes for third-party libraries are not shared between apps, but this might happen in the future. The Privacy Sandbox could then replicate these existing transparency issues around the Google Play Services. The Privacy Sandbox also introduces a separate storage of apps and their third-party libraries, potentially making the analysis of the data stored by those libraries more difficult. Understanding what data those libraries store is important to assess compliance with the EU ePrivacy Directive, which mandates consent for reading and storing of data for most tracking libraries.[88]

```
10        minSdk 21
11        targetSdk 32
12        versionCode 1
13        versionName "1.0"
14
15        testInstrumentationRunner "androidx.test.runner.AndroidJUnitRunner"
16    }
17
18    buildTypes {
19        release {
20            minifyEnabled false
21            proguardFiles getDefaultProguardFile('proguard-android-optimize.t
22        }
23    }
24    compileOptions {
25        sourceCompatibility JavaVersion.VERSION_1_8
26        targetCompatibility JavaVersion.VERSION_1_8
27    }
```

**Figure 2:** Code obfuscation can be enabled with the change of a single line of code in the Android Studio IDE (highlighted).

Importantly, if the Privacy Sandbox gets implemented as currently planned, there would be a strong incentive for the developers of third-party libraries to sign up for the programme. This is because Google is phasing out the existing Android Advertising Identifier and will replace it with new APIs (FLEDGE and Topics) that can only be accessed by apps running inside the Privacy Sandbox. Those third-party libraries that want to be the most competitive and lucrative will thus need to sign up.

The Privacy Sandbox also introduces a range of other functionality. A full analysis is beyond this current paper.

### *4.2.1 Static Analysis*

We now characterise problems in pursuing static analysis, that is, the analysis of apps without running them on a real device – often at scale.

82 Xiao and others (n 70).
83 Egele and others (n 45); Zimmeck and others (n 11).
84 Kollnig and others, 'Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels' (n 53).
85 CNIL, 'Identifiant Publicitaire : Sanction de 8 Millions d'euros à l'encontre de APPLE DISTRIBUTION INTERNATIONAL' <https://www.cnil.fr/fr/identifiant-publicitaire-sanction-de-8-millions-deuros-lencontre-de-apple-distribution-international> accessed 27 January 2023.
86 Google, 'SDK Runtime' (*Android Developers*) <https://developer.android.com/design-for-safety/privacy-sandbox/sdk-runtime> accessed 24 October 2022.
87 microg, 'Implementation Status' (*GitHub*) <https://github.com/microg/GmsCore> accessed 24 October 2022.
88 Kollnig and others, 'A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps' (n 44).

**Ease of obfuscation of apps.** Many apps nowadays use code obfuscation techniques. These techniques primarily aim to reduce the size of app packages and reduce bandwidth of app downloads and updates. Code obfuscation, however, can also be used to hide potentially problematic data practices and make independent analyses significantly harder. Sophisticated code obfuscation techniques are often used by leading developers of apps and third-party libraries, but are increasingly used by less popular apps, too, particularly on Android. One reason for the increased usage of obfuscation is that Google has been simplifying the process. Inside the Android Studio IDE, obfuscation can be enabled with the change of a single line of code in a central location, see Figure 2. Additionally, when app developers upload apps to the Google Play Store, Google sometimes urges them to reduce the size of the app – and obfuscation is the primary approach to accomplish this goal.

From a legal point of review, it could be argued that apps actually need to allow a certain level of transparency of their data practices, and that enabling obfuscation might clash with transparency requirements under the GDPR[89] and other legal regimes. However, as of yet, no legal precedent exists in the app space to our knowledge.

On iOS, Apple actively discourages the use of code obfuscation.[90] Any app using obfuscation would be rejected from being published on the App Store, unless developers explain why they need obfuscation. Apple's reasoning for this policy is that the use of obfuscation hinders its own app review process. This underlines that the use of obfuscation not only makes app review more difficult, but also legitimate app privacy research. At the same time, the code in iOS remains difficult to grasp for third-parties, since iOS apps rely on low-level machine code (but not Android apps).[91]

### *4.2.2 Dynamic Analysis*

We now characterise challenges in conducting dynamic analysis, that is, the analysis of apps by running them on a real device and watching apps' behaviour (e.g. network traffic).

**De-facto ban of self-signed certificates on Android and challenges with certificate pinning.** Traditionally, Android used to be the significantly more open app ecosystem compared to iOS. This has been changing over recent years. A good example of these increasing restrictions is the fact that, as of version 7 from 2016, Android apps do not trust self-signed certificates anymore (unless app developers manually disable this behaviour for their apps). The reason behind this change is to better protect end-users against human-in-the-middle attacks (i.e. the snooping on users' HTTPS-encrypted network traffic by adversaries), which previously were also used for legitimate research. This small change thus represents a significant blow to app privacy research using network traffic analysis (see Section 3.1). Due to these new limitations, even relatively recent studies still use the outdated Android 6, which may lead to non-representative results, since most users use more current Android versions.[92] The research community was developing promising network traffic analysis tools,[93] which are not compatible with the latest versions of Android (version 7 or higher) – unless one makes modifications to either the Android system or specific apps. iOS does not implement similar restrictions on self-signed certificates. Instead, Apple makes it rather difficult to install and trust self-signed certificates on iOS in the first place. At the same time, both Google and Apple increasingly urge developers to implement certificate pinning. It would be a positive addition for researchers to have a method to trust self-signed certificates on Android (as they can already on iOS) and ban certificate pinning from apps (except for rare exceptions). Indeed, Google and Apple discourage third-party app developers from using certificate pinning, but also use certificate pinning themselves for many system-level communications (with no option to disable this), which makes the analysis of such difficult for researchers.

**Restrictions on automated instrumentation on iOS.** The automated instrumentation of Android apps is rather straightforward through the accessibility API and the ADB tools. Google even provides its own automated app testing tool, the so-called *monkeyrunner*. As a result, the automation of Android apps has been well-studied in the academic literature.[94] Meanwhile, there exists almost no study on the automated instrumentation of iOS apps. One reason for this is that Apple does not provide a similar solution as the Android *monkeyrunner*. Another reason is that iOS instrumentation tools are heavily restricted in runtime devices. They can only send one user interaction per second to the device. This limitation is artificial and does not exist in the iOS Simulator. The use of such emulated devices is, however, not suitable for research purposes because apps usually behave differently when run in such an environment.[95]

**Restrictions on system modification, jailbreaks, root, and hook framework.** Much previous app privacy research relied on the ability to make modifications to the operating system (see Section 3). For example, this is currently necessary to gain visibility into encrypted network traffic,[96] to install and trust self-signed certificates, and to study the tracking of users by Google and Apple.[97] Although such approaches represent the state-of-the-art, ideally, app privacy research should ideally be possible without system modification. Unfortunately for researchers, such modifications have become a lot harder in recent versions of operating systems. At the time of writing, no jailbreak exists for iOS 15 and higher. Thus, the study of iOS apps' privacy on the latest iOS versions is rather restricted. These restrictions extend beyond privacy research: Apple is being taken to court by the developer of Cydia, one of the most important alternatives iOS App Stores, over being overly restrictive as to custom technologies.[98]

89 GDPR, art 5 para 1(a).

90 Jingyi Guo and others, 'iLibScope: Reliable Third-Party Library Detection for iOS Mobile Apps' <http://arxiv.org/abs/2207.01837> accessed 9 July 2022.

91 Egele and others (n 45).

92 Anastasia Shuba and Athina Markopoulou, 'NoMoATS: Towards Automatic Detection of Mobile Tracking' (2020) *Proceedings on Privacy Enhancing Technologies Symposium*; Reyes and others (n 42); Shuba, Markopoulou and Shafiq (n 73); Reardon and others (n 1); Catherine Han and others, 'Do You Get What You Pay For? Comparing The Privacy Behaviors of Free vs. Paid Apps.', *Workshop on Technology and Consumer Protection (ConPro '19)* (2019).

93 Shuba, Markopoulou and Shafiq (n 73); Shuba and Markopoulou (n 92); Ren and others (n 40); Yihang Song and Urs Hengartner, 'PrivacyGuard: A VPN-Based Platform to Detect Information Leakage on Android Devices', *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices* (ACM Press 2015) <http://dl.acm.org/citation.cfm?doid=2808117.2808120> accessed 14 February 2020.

94 Yuanchun Li and others, 'DroidBot: A Lightweight UI-Guided Test Input Generator for Android', *2017 IEEE/ACM 39th International Conference on Software Engineering Companion* (IEEE 2017) <http://ieeexplore.ieee.org/document/7965248/> accessed 4 July 2022.

95 Song and Hengartner (n 93).

96 Shuba and Markopoulou (n 92).

97 Douglas J Leith, 'Mobile Handset Privacy: Measuring The Data iOS and Android Send to Apple And Google' (2021) *SecureComm*.

98 Sarah Perez, 'Cydia's Antitrust Case against Apple Is Allowed to Proceed'

A particularly concerning development for app research is the increasing rollout of the Google SafetyNet and hardware attestation. As part of this SafetyNet, Google tries to restrict attempts to modify Android by preventing certain apps from running on such devices. This is meant to protect sensitive apps (e.g. banking apps) from running on unsafe devices, but SafetyNet is also used by other popular apps such as Pokemon GO and Snapchat. Some Internet outlets have declared the 'end for Android rooting, [and] custom ROMs'.[99] The roll-out of the SafetyNet implies that many approaches involving system modification for research face additional challenges, and that there, ideally, should exist ways to study app privacy without such modifications.

**Locking down bootloaders.** The iOS ecosystem is highly locked down. On most devices, it is impossible to install a custom boot chain, e.g. to develop a research version of iOS. In the Android ecosystem, Google usually delivers its own Android devices (i.e. the Pixel series) with a bootloader that can be unlocked. However, third-party manufacturers are still given the option to implement bootloader restrictions. This further holds back app privacy research.

## 4.3 Platform Conduct

**No programmes for academic researchers.** We reached out to multiple points of contact at both Apple and Google to find out how they might support academic researchers. These requests were met with confusion and denial of responsibility. Instead, we were redirected to third-party consultations (Apple facilitates this through consultants.apple.com) or the app developers themselves.

**Bans of privacy software on app stores.** Both Google and Apple review apps before and after they get released on their respective app stores. Partly due to a conflict of interests of these companies in protecting their business interests and user privacy, they have held back app privacy research in the past. For example, Apple banned the ProtectMyPrivacy Lite app from the App Store back in 2013.[100] This app was developed by researchers and would have given users detailed insights into apps' data practices. Apple did not like the fact that this data was obtained from jailbroken iOS devices (even though the proposed ProtectMyPrivacy Lite app did not generate such information itself and only downloaded information from a central server instead) and decided to prevent the publication of this app.

Apple also banned the app 'Sift' from the App Store in 2018.[101] This app would have allowed users to inspect to which domains other apps send data. However, as Apple argued, the use of so-called network filters is not usually permitted for non-Apple apps. As a result, the app is unavailable to the wider public, and only on GitHub for those who are able to compile and install custom iOS apps. Fortunately, Apple has now accepted the need to allow users to inspect apps' network traffic and has integrated such functionality into iOS with version 15 – but to its own, non-customisable design.

Similarly, Google previously banned the Disconnect.me app in 2014.[102] This app would have allowed end-users to inspect and block other apps' network connections to tracking companies, but also conflicted with Google's business model around data-driven mobile ads. Google took down the app five days after its first publication citing that the interference with other apps' functionality is not permitted by apps on the Google Play Store.[103]

All these incidents arguably serve as deterrents to developing future privacy tools for iOS and Android.

**Lack of compliance guidance.** Both Apple and Google are central to the app ecosystem and are uniquely positioned to improve apps' privacy practices.[104] Despite this, neither company provides much guidance on how to comply with legal obligations under key data protection and privacy laws. While this is somewhat understandable for a company like Apple (which engages less in third-party tracking), Google also develops a range of invasive tracking technologies and tightly integrates these technologies into its Play Store ecosystem. Previous research has underlined that existing compliance guidance for app developers tends to be difficult to find, hard to read, and poorly maintained.[105]

**Contractual obligations on researchers.** The use of both the Apple and Google app stores comes with certain contractual obligations, some of which can conflict with the work of app researchers. A set of examples is shown in Table 1. In these contractual obligations, research is not usually expressly permitted, only 'personal, non-commercial use'. Related to this, Apple reserves the unconditional right to restrict the amount of content that can be downloaded by a user. Lastly, these obligations ban unofficial ways of interacting with their services and the circumvention of security features; these strategies, however, lie currently at the heart of most app research.

These current contractual obligations can drive researchers to violate terms of service and hope that they will not be prosecuted. This might lead to service bans for researchers, and potentially have significant legal consequences in some jurisdictions. For example, the UK Computer Misuse Act 1990 makes 'unauthorised' acts in relation to a computer illegal (Section 3). In the most severe cases, doing so may result in imprisonment up to 14 years (Section 3ZA).

Lastly, Apple's app rejection notices (which include the reasons for rejection) are usually subject to non-disclosure requirements. This currently inhibits transparency around why Apple prevents the publication of certain kinds of app designs on the App Store.

## 4.4 Systemic Risks due to Gatekeeper Decisions and Mitigation

In section 3, we elaborated that there are currently widespread known shortcomings regarding app privacy and that these can cause systemic risks. As argued in this Section, these arise, in part, from the

(*TechCrunch*, 31 May 2022) <https://techcrunch.com/2022/05/31/cydias-antitrust-case-against-apple-is-allowed-to-proceed-judge-rules/> accessed 24 October 2022.

99 JC Torres, 'Google SafetyNet Update Might Be The End For Android Rooting, Custom ROMs' (*SlashGear*, 30 June 2020) <https://www.slashgear.com/google-safetynet-update-might-be-the-end-for-android-rooting-custom-roms-01627121/> accessed 24 October 2022.

100 Agarwal and Hall (n 9).

101 Alex Grinman, 'Sift App' <https://github.com/agrinman/sift-ios> accessed 24 October 2022.

102 Reed Albergotti, Alistair Barr and Elizabeth Dwoskin, 'Why Some Privacy Apps Get Blocked From the Android Play Store' (*Wall Street Journal*, 28 August 2014) <https://www.wsj.com/articles/BL-DGB-37413> accessed 24 October 2022.

103 Google, 'Device and Network Abuse' <https://support.google.com/googleplay/android-developer/answer/9888379> accessed 24 October 2022.

104 van Hoboken and Ó Fathaigh (n 18); Greene and Shilton (n 7).

105 Kollnig and others, 'A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps' (n 44).

| Contractual Obligation | App Store Terms of Service | Google Play Terms of Service | Problem |
|---|---|---|---|
| Personal use only | 'You may use the Services and Content only for personal, noncommercial purposes (except as set forth in the App Store Content section below).' | 'for your personal, non-commercial use only' | Research is not usually personal. |
| Potential restrictions of data collection | 'You may be limited in the amount of Content you may download, and some downloaded Content may expire after a given amount of time after downloaded or first played.' | (no equivalent found) | Research often relies on downloading vast amounts of data from the app stores. |
| No circumvention of security features | 'You may not tamper with or circumvent any security technology included with the Services.' | 'You may not ... attempt to, or assist, authorize or encourage others to circumvent, disable or defeat any of the security features or components that protect, obfuscate or otherwise restrict access to any Content or Google Play.' | Some state-of-the-art app research relies on disabling security features, e.g. to circumvent certificate pinning. |

**Table 1:** Contractual obligations of app stores can conflict with app privacy research.

decisions of the providers of the two leading app stores to undermine transparency and accountability around their app ecosystems, and to make app research more difficult than necessary.

These decisions likely constitute, by themselves and in the meaning of Article 34(1) DSA, a systemic risk to the exercise of fundamental rights, including to data protection and respect for private life. Those risks are systemic since they affect nearly every user of app stores since they arise from the normal and intended use of app stores, and with consumers putting much trust in the security measures taken by Apple and Google.[106] This is also true given that some of these decisions have arguably been deliberate and systemic; that the imbalance of power between app store providers and its users is profound; and that there are significant impacts on individuals' exercise of fundamental rights in the absence of app research, transparency, and accountability.

This, in turn, implies that the providers of app stores would need to implement reasonable, proportionate and effective mitigation measures (Article 35). Importantly, these mitigation measures might not just encompass changes to app stores, but also to the research impediments in the underlying mobile app operating systems – that are also developed by the same companies. This is because the operating systems are linked to every transaction on the app stores and because such changes might well classify as 'reasonable, proportionate and effective'.

## 5. Recommendations & Conclusions

Apple and Google wield enormous power over their respective app ecosystems. Given their centrality in the app economy and the relative lack of current power of regulators in digital ecosystems, previous research even termed these platforms 'privacy regulators'.[107]

Despite their central role, our analysis reveals that these platforms currently use numerous technical and non-technical measures that make legitimate (automated) app privacy research more difficult than necessary. These include the widespread use of code obfuscation (potentially in violation of the EU GDPR's transparency requirements) and de-facto ban of self-signed certificates on Android, the encryption of all downloaded iOS apps by Apple (FairPlay DRM), the lack of public APIs and programmes to support independent researchers (e.g. for academic researchers, journalists, and NGOs), and the lack of research provisions in the applicable terms and conditions. Current contractual obligations and the encryption of all iOS apps are particularly problematic decisions because they drive researchers into legal grey areas. Over recent years, this situation has worsened on Android, while iOS had never been particularly permissive.

Overall, these decisions hold back app research and impede users' understanding and choice regarding privacy. They also hinder regulators in their work, restrict organisations in building new privacy and app analysis tools, and foster apps' non-compliance with data protection and privacy laws, which is known to be common. Since some of these decisions were taken *deliberately* by the platform providers with profound impacts on individuals' exercise of their fundamental rights, we conclude that these impediments to app research pose a systemic risk and need to be mitigated, e.g. under Article 35 DSA. For example, as in China, app developers may need to be required to deposit parts of their app source code in a central repository, so as to fulfil the transparency stringent requirements under the GDPR.

Even with the DSA and DMA, explicit regulation of app ecosystems remains rather limited (see Section 2) and privacy problems in apps continue to be widespread. Thus, there might be a need for more explicit transparency and accountability obligations for the providers of app ecosystems, as argued for in previous research[108]. Addressing these current limitations and impediments to (automated) app privacy research is ever more important in a world that keeps increasing its reliance on data, and in which fundamental human rights get chal-

106 Patrick Gage Kelley, Lorrie Faith Cranor and Norman Sadeh, 'Privacy as Part of the App Decision-Making Process', *Proceedings of the 2013 Conference on Human Factors in Computing Systems* (ACM Press 2013) <http://dl.acm.org/citation.cfm?doid=2470654.2466466> accessed 14 February 2020.

107 van Hoboken and Ó Fathaigh (n 18); Greene and Shilton (n 7).

108 van Hoboken and Ó Fathaigh (n 18); Greene and Shilton (n 7).

lenged through the mass adoption of digital systems, and app-based services in particular.

## 6. Acknowledgments

Konrad Kollnig was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/R513295/1. Konrad Kollnig and Nigel Shadbolt have been supported by the Oxford Martin School EWADA Programme.