

Mirror Mirror on the Wall, which APK Mirror Site is the Largest of them All?

Neal Keating

tpp121@gwu.edu

The George Washington University
Washington, DC, USA

Wellington Esposito Barbosa

wellington@gwu.edu

The George Washington University
Washington, DC, USA

Leo Phan

phan@gwmail.gwu.edu

The George Washington University
Washington, DC, USA

Viraj Prakash

vcp@gwmail.gwu.edu

The George Washington University
Washington, DC, USA

Gianluca Stringhini

gian@bu.edu

Boston University
Boston, MA, USA

Adam J. Aviv

aaviv@email.gwu.edu

The George Washington University
Washington, DC, USA

Abstract

Android apps or Android Application Packages (APKs) are commonly distributed through official app stores, but a significant parallel ecosystem of APK mirror sites has emerged, providing users with alternative access to APK packages. These mirror sites host APKs for direct download and sideloading, bypassing security checks typical of official stores and offering access to packages otherwise unavailable to some users. Despite their growing prominence, academic researchers have underexplored the APK mirror ecosystem. In this paper, we analyzed metadata from over 34M versions of approximately 27M unique Android packages collected from seven prominent APK mirror sites, alongside data from the Google Play Store and Amazon Appstore for comparison. Our findings reveal substantial variation in catalog size and package versioning across mirror sites. The smallest, APK Mirror, has only 17K packages while the largest, APK Combo, hosts over 12M packages, compared to Google Play Store's 3.1M packages, at the time of measurement. Mirror sites differ markedly from official stores in both breadth - hosting more unique packages - and depth - retaining multiple package versions, often serving as semi-historical archives. This offers a potentially rich record for researchers to access.

CCS Concepts

• General and reference → Measurement; • Information systems → Digital libraries and archives; Mobile information processing systems.

Keywords

Mirror Sites; APK Repositories; Android; Metadata

ACM Reference Format:

Neal Keating, Wellington Esposito Barbosa, Leo Phan, Viraj Prakash, Gianluca Stringhini, and Adam J. Aviv. 2025. Mirror Mirror on the Wall, which APK Mirror Site is the Largest of them All?. In *Proceedings of the 2025 ACM*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IMC '25, Madison, WI, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1860-1/2025/10
<https://doi.org/10.1145/3730567.3764499>

Internet Measurement Conference (IMC '25), October 28–31, 2025, Madison, WI, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3730567.3764499>

1 Introduction

Traditional app stores [1, 14, 20], e.g., Google Play Store, are the primary mechanism by which users install packages on their Android devices. Alongside these, a substantial secondary ecosystem of APK mirror sites has emerged — platforms like APK Mirror [28], APK Combo [5], APK-DL [3], APK Pure [12], Aptoide [15], APK Fab [6], and APK Monk [8] that allow users to directly download Android Application Packages (APKs) to their devices, bypassing the distribution and security models enforced by official stores.

Despite their prevalence, the scope, scale, and characteristics of these mirror sites remain largely unmeasured, unlike traditional marketplaces. To address this gap, we collected and analyzed metadata from seven prominent mirror sites and compared them to two official marketplaces: Google Play Store and the Amazon Appstore. We observed significant variation in catalog size, ranging from 17K packages on APK Mirror [28] to over 12M packages on APK Combo [5], compared to 3.1M on Google Play Store [20]. Unlike official stores, many mirror sites retain extensive version histories of packages, with one averaging over 25 versions per package.

These discrepancies reflect differing curation strategies. Some mirrors intentionally serve as semi-historical archives of package versions over time [7], and store 25+ average versions of each package, while others operate as strictly commercial ventures, keeping only a few of the recent versions to prioritize monetization through search engine optimization [17] (SEO) and advertising.

Importantly, while mirror sites may act as unofficial app distributors, they do not simply replicate the content of official platforms. Their catalogs diverge both in breadth, often hosting more unique packages, and in depth, retaining longer version histories. The stores also differ both in their app collection practices and in their security measures. While some allow user uploads and claim to verify developer signatures, provide security reports, assign trusted-app badges, and carefully curate their catalogs, others do not disclose if and how they perform similar security checks. These variations offer important opportunities for researchers to view historical and varied records of APKs, for example, to support research on malware analysis, app development, and software engineering.

2 Methods

Identifying APK Mirror Sites. We broadly define official app stores as those associated with traditional providers in the mobile community, such as Google Play Store and Amazon Appstore. At the same time, we consider mirror sites to be third-party, non-official app distributors that, at some level, attempt to (imperfectly) reflect the content found in the official stores. To identify APK mirror services, we performed initial targeted searches on Google for "android mirror sites" and "APK mirror downloads," and while it may be the case that other sites are active, these constitute the largest, easily available sets that a typical user would find. Through this process, we identified seven prominent Android mirror sites: APK Mirror [28], APK Combo [5], APK-DL [3], APK Pure [12], Aptoide [15], APK Fab [6], and APK Monk [8] to scrape metadata from, as well as two traditional app markets: Amazon Appstore [1] and Google Play Store [20]. We performed all measurements from 06/21/2022 to 10/14/2023 using industry best practices for web scraping [24] by restricting the rate and distributing requests as best as possible. We intentionally excluded two platforms often mentioned in prior work from our analysis: AndroZoo[2] and F-Droid [18]. AndroZoo is a researcher-only site that requires formal access agreements and is not a public distribution channel. F-Droid is a repository of free and open-source software (FOSS) applications. Every app on F-Droid must be open source, is built from source by the maintainers, and passes through a structured submission and review pipeline, excluding many proprietary apps. Therefore, these sites do not serve general Android users in the same way as the analyzed publicly accessible mirror sites.

Data Organization. The metadata available varied slightly (details available in Table 1), but all sites included several standard fields, including package name, version number, version code, name of package publisher, and the package category. We also enriched each package-version combination with additional fields, namely the name of the mirror site where the package-version combination, the time date group when the package-version combination was scraped, and the full URL on the mirror site where the package-version combination's metadata was scraped from. Additionally, when available, we also captured other relevant metadata fields for each package-version combination, such as SHA1 hashes, SHA256 hashes, certificates, URL of APK download location, and number of downloads from the mirror site.

Further details on the data organization and the datasets used in our analysis can be found in the artifact available here.

Limitations. While we believe our analysis captures some of the most relevant mirror sites within the Android distribution ecosystem, the vast number of platforms distributing Android apps means we may have missed some important or emerging sites, particularly those serving primarily non-English speakers.

Additionally, some technical limitations affected data collection. Several sites employed protections like Cloudflare, which constrained automated scraping. Moreover, our custom parsers occasionally failed to extract full metadata for certain package-version combinations. These failures impacted approximately 0.4% of the dataset, and while this small percentage could still introduce minor inaccuracies in our analysis, particularly in precise calculations of

averages or distributions, we believe the overall trends and conclusions remain. Therefore, we do not believe they materially affect our overall analysis.

3 Results

Number of Packages on Mirror Sites. We find that there are stark differences in the number of packages hosted on each site and metadata provided. The number of package-version combinations, category distribution, and available versions per package vary dramatically, potentially indicating differing strategic approaches and business models. The primary measurement results are presented in Table 1.

APK Combo hosts the most extensive catalog, with over 12M packages, followed by APK Pure at nearly 11M. APK Mirror, while the smallest with 17K packages, describes itself as "a highly curated community" [7], which likely explains its reduced volume. The heatmap in Figure 1 illustrates the bidirectional overlap of apps across the analyzed app stores. For example, while 87.7% of the Google Play Store's catalog is covered by APK Combo, only 23.1% of APK Combo's catalog is covered by Google Play, highlighting significant differences in catalog size and coverage. This reveals that not all non-official app stores effectively replicate the content of the leading distributor of Android packages and raises the question of what type of packages are available in these third-party stores. This suggests that the mirror sites are likely serving a different purpose than the official app store and are instead offering packages that are either antiquated, deprecated, removed, modified, or not distributed on official channels. We also observe a significant overlap between the two largest providers: APK Combo and APK Pure. Notably, almost 88% of APK Pure's catalog is replicated on APK Combo. Incidentally, nearly 79% of APK Combo's packages are on APK Pure, suggesting that the sites may have used similar app acquisition methods to build their catalog. Meanwhile, the Amazon Appstore has low overlap with nearly all of the mirror sites, particularly the largest ones, indicating that it might not be a meaningful reference point for catalog completeness. Although no individual mirror site covers the entire Google Play Store catalog, we find that the combined coverage of the analyzed app stores includes 100% of the 3M unique apps available on Google Play. Furthermore, we observe that over 10M unique packages are available for download exclusively outside the Google Play Store.

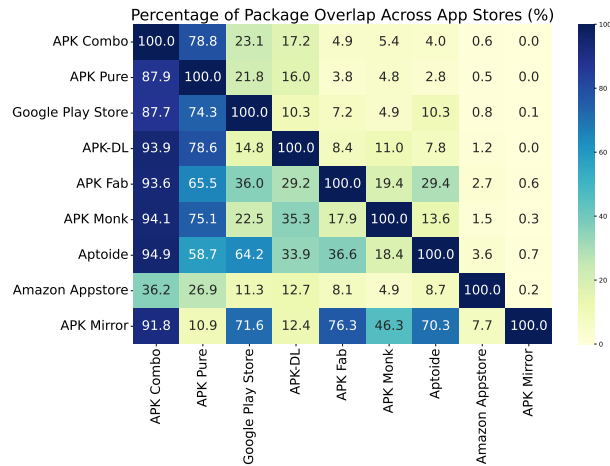
Package Release and Upload Dates. Figure 2 provides a cumulative distribution of (a) packages' release dates and (b) packages' upload dates across several APK mirror sites – after the release of Android's first commercial version on September 23, 2008 – reflecting the evolution of app offerings over time. The steepness of the curves indicate the rate new packages are added to each platform. All sites have shown significant growth in their app collections, particularly in recent years, signaling a robust expansion of mirror sites.

The Google Play Store can be utilized as a baseline for comparative analysis due to its status as the most established and globally recognized app store. Most mirror sites, except APK Fab, exhibit curves that are positioned to the left of the Google Play Store's curve. This indicates that a more significant percentage of their packages were released during earlier periods than those released in

Table 1: Comparison of Various Android App Markets/Mirror Sites

Market	APK Mirror	APK Pure	APK Combo	APK-DL	Aptoide	APK Fab	APK Monk	Google	Amazon
# Packages	17,260	10,925,324	12,078,901	2,207,919	508,618	632,806	687,838	3,178,986	211,187
Avg Num. of Versions/Package	25.4	1.4	1.2	1.0*	4.2	6.6	1.7	1.0*	1.0*
# Developers	6,384	2,502,978	2,786,039	791,154	221,386	239,156	364,849	1,073,332	64,345
Founded Date	Jul 2014	Apr 2014	May 2018	Jan 2014	Oct 2010	Jun 2016	Oct 2016	Sep 2008	Mar 2011
Earliest Release Date	7/6/10	9/22/08	1/5/09	10/22/08	2/19/10	11/6/08	10/7/08	2/9/09	10/11/08
Earliest Upload Date	10/2/14	4/29/14	10/26/18	3/16/14	7/8/11	12/2/16	10/27/16	3/25/09	3/15/11
Ads Displayed	✓	✓	✓	✓	✓	✓	✓		
Ad Distributor	Google	Google	Google	Google	Google	Google	Google	N/A	N/A
Permission Display					✓		✓	✓	✓
User Ratings		✓			✓	✓		✓	✓
User Reviews	✓	✓		✓	✓			✓	✓
User Direct Upload	✓	✓	✓			✓	✓		
DMCA Request Support	✓	✓	✓	✓	✓	✓	✓	✓	✓
Claim App Vetting	✓	✓		✓	✓	✓	✓	✓	✓
APK Certificate	✓			✓					
MD5 Hash	✓								
SHA-1 Hash	✓	✓		✓	✓	✓			
SHA-256 Hash	✓								
Site Map	✓	✓	✓	✓	✓	✓			
User Accounts	✓	✓						✓	✓
Membership	✓	✓							
About or FAQ	✓	✓	✓		✓	✓		✓	✓

* Only provide a single version per app

**Figure 1: Site Overlap using Package Name**

the Google Play Store. In another way, the packages on the mirror sites tend to be older than those on the Google Play Store. Most likely, the mirror sites are not removing packages, and thus, older packages that have since been removed (or updated) on the Google Play Store persist on the mirror sites.

This could also reflect a strategic choice to host packages released in a specific era, possibly to cater to users seeking legacy applications or to fill niches not covered by the official store. For instance, Figure 2(b) shows that over half of the packages on APK Combo, APK Pure, APK Monk, and APK-DL were uploaded before 2020, indicating that they may serve as valuable repositories for

older packages no longer available on the Google Play Store. This can be particularly beneficial for users with older devices or those searching for packages that have been removed from the official market for various reasons, such as policy changes or updates. However, these legacy packages might lead to possible security risks due to not getting timely updates and possibly malware that was removed from the official market yet remained in the third-party mirror site [21].

Package Categories. We developed a unified category system to allow for app store comparison by manually aligning semantically similar package categories from Google Play Store (and mirror sites) and Amazon Appstore. Table 2 (in the appendix) shows how we grouped similar categories (e.g., "Business," "Productivity," "Tools," and "Utilities") under a unified label (e.g., "Productivity"). The heat density in Figure 3 shows the results and demonstrates a vibrant array of package categories, reflecting Android users' diverse interests and the platform's strategic content.

Notably, the category "Games" stands out with a consistently high percentage across most sites, which indicates that gaming packages are a primary focus for APK mirror sites, aligning with the global popularity of mobile gaming. The heat map also shows the relatively low presence of Finance packages on APK mirror sites compared to the Google Play Store, where the category is more represented. This could be attributed to Finance packages' dependence on security features, which users prefer to obtain from official sources.

Non-Reflection of Mirror Sites. Although many third-party app stores are known as mirror sites of the Google Play Store, their app catalogs often diverge significantly from one another and from the Play Store itself. Figure 1 shows that APK Pure and APK Combo

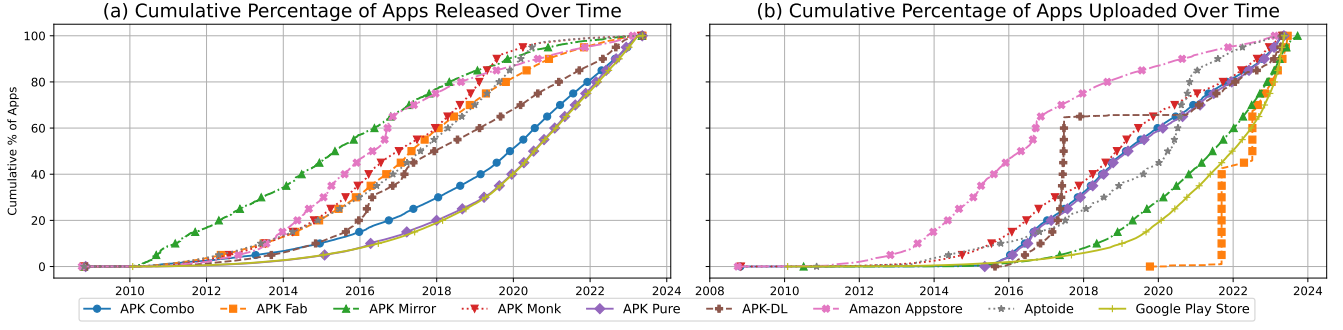


Figure 2: (a) App Release and (b) Upload Date as Cumulative Distribution Functions

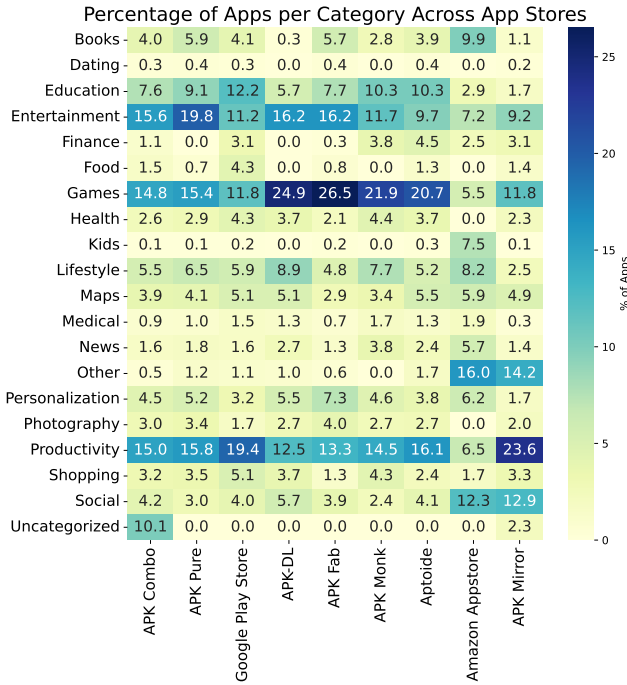


Figure 3: Category Distribution

exhibit a typical "mirror site behavior," closely tracking app availability from Google Play Store. However, most mirror sites allow users to submit uploads and claim to comply with DMCA take down requests (see Table 1), leading to catalogs that differ substantially in content over time. For instance, Figure 2 shows that APK-DL - despite explicitly claiming to be a mirror site [4] - illustrates an irregular upload pattern, having accumulated over 60% of its catalog in 2017, spending 4 years without significant uploads, and resuming app acquisition more consistently after 2021. Similarly, APK Fab added over 40% of its catalog within a short window in 2021 and increased its activity again in late 2022.

There are several possible explanations for these discrepancies. First, some sites may aim to replicate the Play Store comprehensively but lack the technical resources or infrastructure to do so

constantly and at scale. Second, specific platforms intentionally position themselves not as passive mirrors but as curated third-party app stores. APK Mirror, despite its name, acts as a highly curated platform that supports user uploads, enforces signature verification to verify if the app comes from a reliable developer, and conducts manual reviews before publishing APK files [7], indicating a deliberate shift away from pure mirroring. Moreover, platforms like Aptoide and APK Pure increasingly distinguish themselves by offering exclusive features to both users and developers, including APK signature checks [10], developer dashboards [9], monetization tools [13], APK website builder [11] and simultaneous distribution to multiple alternative app stores [16]. Finally, future research could investigate if mirror sites use selection criteria, such as app popularity, when deciding which packages to collect from official stores to reduce resources on uploading less relevant packages. These strategies indicate a broader trend in which non-official app stores evolve to differentiate in the app distribution ecosystem.

4 Related Work

Researchers have made significant efforts in compiling Android packages' metadata and datasets [25, 26], as well as in measuring vulnerability progression, API field changes, and security issue evolution [19, 23, 27]. The closest related works to our research is Wang et al. [29], which investigated and analyzed more than 6 million Android packages downloaded from 16 Chinese app markets and Google Play, and Ishii et al. [22], which investigated 4.7M Android packages covering 27 app markets, mainly obtained from AndroZoo [25]. Here, we explore third-party marketplaces that attempt to mirror official stores, categorize the content on these sites, and observe how curation strategies differ from those of the official marketplaces. Furthermore, previous research has focused on packages and metadata collection per individual site, instead of a comparative study on third-party Android markets. To our knowledge, our work is the most exhaustive comparative analysis made between different mirror sites.

5 Discussion and Conclusion

We presented a large-scale measurement of metadata collected from seven popular mirror sites for Android app distribution platforms, in order to shed light on this relatively underexplored segment of the Android app ecosystem. Our findings reveal that the scale,

content, and roles of these services vary significantly, highlighting a complex and diverse ecosystem beyond the official app stores.

Scale and Diversity. We identified that, during our collection period, the targeted websites hosted a total of over 34M unique Android package-version combinations. We measured characteristics such as the package-version combinations' upload dates, categories, versions, and overlap between mirror sites. We find that not all mirror sites aim or can replicate the content of the Google Play Store in practice. This divergence can be observed in both the breadth (more unique package names than the official store) and depth (more package-version combinations of same package name) of the catalogs they offer. Some mirror sites host a broader range of packages, including several unique packages (e.g., APK Pure and APK Combo) or multiple versions of the same package (e.g., APK Mirror). In contrast, others include only a limited subset of what is available in official app stores (e.g., Aptoide and APK Mirror).

Additionally, some mirror sites effectively allow the upload of modified versions of packages, which further distinguishes their catalogs from the official stores or strictly curated repositories. This practice poses significant risks for users, especially when these app stores do not securely vet modified packages or when they keep outdated versions of packages with known security vulnerabilities to increase the site's catalog and number of visits.

Role of Mirror Sites. Our findings indicate that APK mirror sites may serve a valuable role in providing access to applications for users in regions where official app distribution platforms are unavailable or restricted. Additionally, they give an audience seeking modified packages that cannot be distributed on Google Play due to rights infringement, aka 'mod' or 'modded' packages, that offer new themes, unlocked features, or unique versions of games and messaging packages that also cannot be found on official stores.

At the same time, the business model of APK mirror sites may also introduce notable risks. These platforms heavily rely on Search Engine Optimization (SEO) strategies to attract traffic, which in turn generates substantial advertising revenue. The maintenance of extensive app repositories enhances SEO performance but simultaneously increases the likelihood of hosting unvetted, potentially vulnerable, or malicious applications, given the lack of curation inherent in many of these services.

Implications and Future Research. The differences between non-official app stores show that they target different user needs and have different goals. Understanding this variation on the Android app distribution system is relevant to support future research in various areas, such as malware analysis, app development, and software engineering, by helping researchers select repositories that match their goals and clarifying the limitations involved in each app store analyzed. Our study also lays the groundwork for future research on why users utilize these app stores, what security practices and vulnerabilities are present in these platforms, and to what extent these app stores contribute to malware dissemination in the Android ecosystem.

Acknowledgments

We thank the reviewers and shepherd for their thoughtful feedback. We also thank Tanim Khan and Juehao Lin for their helpful collaboration to this project. This material is based upon work supported in part by a Google ASPIRE Fund Award and the National Science Foundation under Grant Number 1845300.

Artifact Availability

The datasets containing the metadata of packages from the analyzed app stores are available for download at Open Science Framework (OSF): <https://doi.org/10.17605/OSF.IO/DGWNB>

References

- [1] AMAZON. Amazon Appstore App For Android. <https://www.amazon.com/gp/mas/get/amazonapp>. Last Accessed: May 5, 2025.
- [2] ANDROZOO. Androzoo home. <https://androzoo.uni.lu/>. Last Accessed: May 6, 2025.
- [3] APK-DL. APK-DL : Download Android APK Files. <https://apk-dl.com/>. Last Accessed: September 27, 2024.
- [4] APK-DL. DMCA Removal Request. <https://apkdl.net/dmca>. Last Accessed: May 5, 2025.
- [5] APKCOMBO. APKCombo - #1 APK Downloader. <https://apkcombo.com/>. Last Accessed: May 6, 2025.
- [6] APKFAB. Free Android Games and Apps Online Download - FREE APK Downloader. <https://apkfab.com/>. Last Accessed: May 5, 2025.
- [7] APKMIRROR. FAQ - APKMirror. https://www.apkmirror.com/faq/#What_is_the_purpose_of_APKMirrorcom_What_APks_are_accepted_I_just_uploaded_an_APK_but_it8217s_not_going_live. Last Accessed: May 6, 2025.
- [8] APKMONK. Download Android App Apks Free. <https://www.apkmonk.com/>. Last Accessed: May 5, 2025.
- [9] APKPURE. APKPure Developer. <https://developer.apkpure.com/?hl=en>. Last Accessed: May 5, 2025.
- [10] APKPURE. APKPure Signature Verification Online, Official Version APK Detection Service. <https://apkpure.com/apk-signature-verification>. Last Accessed: May 5, 2025.
- [11] APKPURE. Best Website Builder Tailored for App Developers. https://apk.com/?utm_source=APKPure&utm_medium=com_footer_link. Last Accessed: May 5, 2025.
- [12] APKPURE. Download APK on Android with Free Online APK Downloader - APKPure. <https://apkpure.com/>. Last Accessed: May 5, 2025.
- [13] APKPURE. Monetize with us | APKPure for Business. <https://apkpure.com/traffic-monetization>. Last Accessed: May 5, 2025.
- [14] APPLE. App Store - Apple. <https://www.apple.com/app-store/>. Last Accessed: September 27, 2024.
- [15] APTOIDE. Aptoide - The alternative Android app store. <https://en.aptoide.com/>. Last Accessed: May 5, 2025.
- [16] APTOIDE. Aptoide Connect - The future of app distribution. <https://connect.aptoide.com/>. Last Accessed: May 5, 2025.
- [17] BERMAN, R., AND KATONA, Z. The role of search engine optimization in search marketing. *Marketing Science* 32, 4 (2013), 644–651.
- [18] F-DROID. F-Droid - Free and Open Source Android App Repository. <https://f-droid.org/>. Last Accessed: May 6, 2025.
- [19] GAO, J., LI, L., KONG, P., BISSYANDÉ, T. F., AND KLEIN, J. On vulnerability evolution in android apps. In *Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings* (2018), pp. 276–277.
- [20] GOOGLE. Android Apps on Google Play. <https://play.google.com/>. Last Accessed: May 5, 2025.
- [21] HASSAN, S., SHANG, W., AND HASSAN, A. E. An empirical study of emergency updates for top android mobile apps. *Empirical Software Engineering* 22 (2017), 505–546.
- [22] ISHII, Y., WATANABE, T., KANEI, F., TAKATA, Y., SHIOJI, E., AKIYAMA, M., YAGI, T., SUN, B., AND MORI, T. Understanding the security management of global third-party android marketplaces. In *Proceedings of the 2nd ACM SIGSOFT International Workshop on App Market Analytics* (2017), pp. 12–18.
- [23] KALYSCH, A., SCHILLING, J., AND MÜLLER, T. On the evolution of security issues in android app versions. In *Applied Cryptography and Network Security Workshops: ACNS 2020 Satellite Workshops, AIBlock, AIHWS, AIoTS, Cloud S&P, SCI, SecMT, and SiMLA, Rome, Italy, October 19–22, 2020, Proceedings 18* (2020), Springer, pp. 523–541.
- [24] KROTOV, V., AND JOHNSON, L. Big web data: Challenges related to data, technology, legality, and ethics. *Business Horizons* 66, 4 (2023), 481–491.

- [25] LI, L., GAO, J., HURIER, M., KONG, P., BISSYANDÉ, T. F., BARTEL, A., KLEIN, J., AND TRAON, Y. L. Androzoo++: Collecting millions of android apps and their metadata for the research community, 2017.
- [26] LINDORFER, M., VOLANIS, S., SISTO, A., NEUGSCHWANDTNER, M., ATHANASOPOULOS, E., MAGGI, F., PLATZER, C., ZANERO, S., AND IOANNIDIS, S. Andradar: fast discovery of android applications in alternative markets. In *Detection of Intrusions and Malware, and Vulnerability Assessment: 11th International Conference, DIMVA 2014, Egham, UK, July 10–11, 2014. Proceedings 11* (2014), Springer, pp. 51–71.
- [27] MAHMUD, T., CHE, M., AND YANG, G. Android api field evolution and its induced compatibility issues. In *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (2022), pp. 34–44.
- [28] MIRROR, A. APKMirror - Free APK Downloads - Free and safe Android APK downloads. <https://www.apkmirror.com/>. Last Viewed May 6, 2025.
- [29] WANG, H., LIU, Z., LIANG, J., VALLINA-RODRIGUEZ, N., GUO, Y., LI, L., TAPIADOR, J., CAO, J., AND XU, G. Beyond google play: A large-scale comparative study of chinese android app markets. In *Proceedings of the Internet Measurement Conference 2018* (New York, NY, USA, 2018), IMC '18, Association for Computing Machinery, p. 293–307.

A Appendix

Ethical Considerations

In our research, ethical considerations played a pivotal role, particularly given the methodologies used for data collection. One

significant concern was the rate at which our web scrapers operated. Rapid scraping can impose financial burdens on site operators due to increased bandwidth and server load, potentially leading to substantial costs depending on their hosting agreements. To address this, we carefully balanced efficient data collection with minimizing site impact. We chose an average rate of 10 requests per second per site, which we deemed a fair compromise, ensuring timely acquisition while preserving fair access for other users. Additionally, we deliberately avoided large-scale downloads of APK files, which could have caused considerable egress traffic and associated costs. By focusing exclusively on metadata collection, we significantly reduced the potential financial and operational burden on the mirror sites. We believe our methodology balances the need for large-scale empirical insight with the imperative to respect infrastructure limits and site integrity.

Table 2: Mapping the relationship between categories from Google Play Store and Amazon Appstore.

Google Play Store*	Unified Category	Amazon Appstore
Books & Reference	Books	Books & Comics
Comics		Magazines
Business	Productivity	Business
Productivity		Productivity
Tools		Utilities
Communication	Social	Communication
Social		Social
Dating	Dating	-
Education	Education	Education
Entertainment	Entertainment	Movies & TV
Video Players & Editors		Music & Audio
Music & Audio		Novelty
Events	Lifestyle	Lifestyle
Lifestyle		Local
House & Home		
Finance	Finance	Finance
		Personal Finance
Food & Drink	Food	Food & Drink
Games	Games**	Games
Health & Fitness	Health	Health & Fitness
Parenting	Kids	Kids
Maps & Navigation	Maps	Transportation
Medical	Medical	Medical
News & Magazines	News	News
Personalization	Personalization	Customization
Photography	Photography	Photo & Video
Art & Design		
Shopping	Shopping	Shopping

*All mirror sites followed a similar categorization structure to the Play Store, which allowed us to treat their category labels as equivalent. For APK Combo, an app store that did not provide category information, we estimated each package's category by cross-referencing it with listings on the other stores. For example, if we found the package com.facebook.katana without a category associated, we would search for how it was categorized by Google Play Store, Amazon Appstore, and the other mirror sites – in that order – to later categorize it as a 'Social' app.

**The 'Games' category includes the following subcategories: 'action', 'adventure', 'arcade', 'board', 'card', 'casino', 'casual', 'educational', 'music', 'puzzle', 'racing', 'role playing', 'simulation', 'sports', 'strategy', 'trivia', and 'word'.