Majed Almansoori*, Andrea Gallardo*, Julio Poveda, Adil Ahmed, and Rahul Chatterjee

A Global Survey of Android Dual-Use Applications Used in Intimate Partner Surveillance

Abstract: Intimate partner violence (IPV) is a pervasive societal problem that affects millions of people around the world. IPV perpetrators increasingly weaponize digital technologies like mobile applications ("apps") to spy on, monitor, and harass victims. Surveillance-capable apps can have legitimate use cases, for example, locating children, and are therefore easily available on various mobile app stores like the Google Play Store. Nevertheless, these applications are easily repurposed by abusers to track their victims. The problem of such *dual-use* apps in IPV is global. However, current understanding of the ecosystem of such apps is limited to English-language apps, potentially limiting its relevance to non-English speaking IPV survivors across the world. In this paper, we study the prevalence of dual-use applications found in 15 languages and 27 countries. We collected 51,868 unique apps in 2020 from the Google Play Store, using queries such as "track wife's location." Through a semi-manual analysis of a subset of these apps, we discovered 854 unique dual-use apps, and estimate that among the apps collected from Google Play, 3,988 are dual-use apps. We found notable differences in app search results, suggested queries, and marketed capabilities of dual-use apps across different languages. For instance, we identified that 18% of dual-use apps do not have an English description, and 28% could not be found using English queries. Google Play (cursorily) blocks certain queries referring explicitly to intimate partner surveillance (IPS) to discourage potential abusers, but the blocking efficacy varies across languages. For example, we found that 80% of explicit IPS queries for English are blocked, but none for Bengali, Chinese, Hindi, Malay, Thai, and Vietnamese. Thus, abusers fluent in those languages can evade such blocking with no effort.

Keywords: intimate partner violence, dual-use android apps, technology-facilitated abuse

DOI 10.56553/popets-2022-0102
Received 2022-02-28; revised 2022-06-15; accepted 2022-06-16.

***Corresponding Author: Majed Almansoori:** University of Wisconsin-Madison, malmansoori2@wisc.edu
***Corresponding Author: Andrea Gallardo:** Carnegie Mellon University, agallar2@andrew.cmu.edu
Julio Poveda: University of Maryland, jpoveda@umd.edu
Adil Ahmed: University of Wisconsin-Madison, oahmed4@wisc.edu
Rahul Chatterjee: University of Wisconsin-Madison, rahul.chatterjee@wisc.edu

1 Introduction

Intimate partner violence (IPV) and technology-facilitated abuse are severe societal problems that affect a large number of people in the US and around the world [49, 54, 60, 70]. A recent study reports that globally one in four ever-partnered women between the age of 15-49 years experience IPV [54]. Digital technologies are increasingly being used to spy on, stalk, and harass intimate partners [19, 24, 25, 62, 64, 65]. Among other tools, abusers regularly use mobile applications ("apps") to monitor or control the victim's phone and thereby track their whereabouts. A number of apps available on mobile application stores for benign use cases, such as locating a phone if it is lost or recording calls for business purposes, can be used to surreptitiously surveil an intimate partner. Chatterjee et al. [19] use the term *dual-use* apps to refer to apps that have legitimate use cases but can easily be used for intimate partner surveillance (IPS). These dual-use apps not only harm victims emotionally [71], but can also lead to physical violence [25] or even murder [61].

Dual-use apps are used globally to surveil victims [39, 47, 67], and their use has soared following COVID-19 pandemic shelter-in-place restrictions [16, 18, 42, 55, 66]. Yet, prior work on dual-use apps has only focused on apps available in English, potentially limiting its applicability in non-English contexts.

In this work, we conduct the first multilingual measurement study of Android apps available on Google Play that can be used to conduct IPS. We attempt to answer the following three research questions:

(1) How many and what types of dual-use apps are

available in non-English languages on Google Play?

(2) What is the overlap of these apps with English-language dual-use apps?

(3) Are there differences in the ways dual-use apps are described in different languages in the context of IPS?

Knowing answers to these questions is essential to build effective defense tools for victims of IPS around the globe and the support services that help them.

The main challenge in conducting such a study consists in language constraints. Dual-use apps are hard to identify through static or dynamic analysis, as their behavior is identical to other legitimate apps, even when they are being used in malicious use cases (as we show in Section 3.4). Thus, the only known way to detect dual-use apps at scale is by looking at app descriptions on app marketplaces and identifying app capabilities using natural language processing tools. (App developers report app capabilities to advertise to potential users — legitimate or abusive.) Therefore, detecting dual-use apps in different languages requires language-dependent approaches.

Chatterjee et al. [19] developed a semi-manual pipeline to identify dual-use apps with English descriptions. We extended their search pipeline to look for dual-use apps in 15 different languages (with more than 5 billion native speakers) and 27 different countries. First, for each language, we created a seed set of queries by translating the seed queries used in prior work [19] (queries such as "track wife's location," which are likely to be used by abusers seeking dual-use apps on popular mobile app stores) and adding new queries as appropriate, based on grammar and usage, with the help of native speakers of each language. We expanded the query set to include semantically related queries in each language, using Google Play's query suggestion API. We crawled the Google Play Store with these queries for each language for 47 days in 2020, downloading query suggestions, search results, and app metadata.

In total, 51,868 unique apps were returned for our queries across 15 languages. We filtered the apps using a machine learning (ML) classifier created by Chatterjee et al. [19], after translating the non-English descriptions to English using Google Translate. We then manually coded the capabilities of 350 randomly sampled apps from each language based on their descriptions and images posted on Google Play, and flagged them as dual-use or not. We labeled 854 dual-use apps, and we estimate that there are around 3,988 apps that can be used for IPS in our dataset of 51,868 apps. These apps provide a broad range of capabilities, such as tracking location, recording audio/video, and remotely controlling the phone.

Of the labeled dual-use apps, 563 are found in English searches using the US country code (English-US). This is significantly fewer than the 2,473 dual-use apps found in 2018 using only English-US search queries [10, 19, 36], of which 2,062 (about 83%) have been taken down from Google Play. As our search pipeline mirrored that of Chatterjee et al. [19] and was conducted for a longer period of time, we believe this reduction in dual-use apps may also be due to Google's August 2020 addition of a policy against stalkerware [29, 33]. However, we also found some new egregious dual-use apps in our new crawl of Google Play Store (Section 4.3).

Of all 51,868 apps we collected, more than half (52%) have an English description available on Google Play, but only 13% were found using only English search queries. Of the dual-use apps we manually flagged, 72% were found via searches using English queries. This shows that searching with multiple languages and countries provides better visibility into the ecosystem of dual-use apps available on Google Play. We estimate that in our dataset more than 500 dual-use apps (per language) can be found with Chinese, English, and Spanish queries, and that fewer than 200 such apps can be found with Bengali and Japanese searches, as shown in Fig. 4.

We also audited Google Play's query suggestions based on our queries for finding dual-use apps. We found that in some languages, Google Play blocks queries referring to spying or tracking an intimate partner, for example, "app to track girlfriend". However, such blocking is not consistent across languages, and Google Play often fails to block translations of blocked queries (e.g., the same query in Arabic *"tatbiq litatabue sadiqatih"*) or misspelled versions (e.g., "app to track girl friend"). In general, query suggestions led to better coverage of dual-use apps, but the quantity and quality of suggestions vary widely among languages. For example, we did not receive any suggestions in Japanese and the majority of the suggestions in Bengali were not relevant. Nevertheless, we also found several queries suggested by Google Play that directly show IPS intention, such as "*rastrear a tu esposo*" (track your husband). Since Google Play is blocked in China, we could not get visibility into one of the largest Android user groups [59] via our analysis of Google Play. Therefore, we also crawled four popular Android app stores available in China and discovered 110 dual-use apps. We observe that, similar to Google Play, the policies of these stores prevent overt spyware

from being distributed, but dual-use apps are still available for download.

Contribution. Our main contributions are:

- We did the first comprehensive study of dual-use apps found in 15 different languages and 27 countries. We collected 51,868 apps from Google Play, which is the largest dataset of dual-use apps analyzed in multiple languages. We hand-labeled 350 apps found in each language. We found 854 unique dual-use apps and estimate that there are 3,988 potentially dual-use apps among the apps we collected from Google Play across all 15 languages.
- We compare the distribution of dual-use apps and their capabilities found in different languages, and also compare our findings with prior work.
- We also analyzed the query suggestions by Google Play and show that (a) query suggestions lead to better coverage of dual-use apps, and (b) query suggestions vary widely between languages (e.g., we did not receive any suggestions in Japanese, and the majority of the suggestions in Bengali were not relevant).

Dual-use apps can be dangerous when used for stalking or IPS, and one of the ways to combat them is by creating awareness. We have therefore added our list of dual-use apps found in English and non-English languages to the database used by ISDi [10] to scan survivors' devices. We reported our findings to Google, specifically the issue of IPS query suggestions in English and non-English languages, easy evasion of query blocking, and the dual-use apps we found. We also reported our results to the Chinese app stores we crawled, namely Xiaomi, Baidu, Tencent, and Huawei. Our data is publicly available for use by researchers. [1]

2 Background and Related Work

Technology Abuse and Intimate Partner Violence. Prior work has shown that abusers exploit technology, including spyware and dual-use apps, to harass, impersonate, threaten, monitor, intimidate, stalk, and harm their victims [19, 24–26, 44, 62, 71]. There is a vast marketplace for spyware apps and hacking services in English [35, 48]. In 2018, Chatterjee et al. [19] compiled a list of dual-use apps that an abuser might be able to find by simply searching Google Play and the Apple App Store. They also showed that existing tools for programmatically detecting such apps are ineffective, as they do not flag IPS-capable dual-use apps [19]. The list of dual-use apps found in this study was incorporated in a novel IPS-oriented dual-use detection tool called ISDi [10], which can scan IPS survivors' devices for dual-use apps [36]. The tool was tested in a technology clinic that supports IPS survivors in New York City (NYC), and it proved to be a valuable tool for both IPV professionals and survivors during their consultations [36]. The efficacy of such tools depends on knowledge about dual-use apps available on app stores.

ISDi cannot comprehensively help victims whose abuser might use non-English languages to search for dual-use apps, a limitation noted by Havron et al. [36]. While studies have pointed out that dual-use apps are used in countries around the world [39], prior work has not analyzed the dual-use app ecosystem in other languages. Thus, there is a need to expand current knowledge about dual-use apps across countries and languages. In this study, we focus on Android apps found on Google Play because Android is the most used smartphone operating system in the world [27], and Google Play is the official store for Android apps, hosting more than 3.5 million apps that are available in 77 languages and in more than 150 countries [12, 22].

App Usage Across the Globe. User behavior and app installations ("installs") can vary based on region. Prior work has looked into how mobile users acquire and use apps, predicting user traits such as language based on apps present on the device [40, 50, 56]. Guo et al. [34] provide a comprehensive survey of studies that analyze user behavior, including user reviews and app installs across different languages and regions. Peltonen et al. [50] studied how cultural background affects app usage and installs, analyzing app usage of 25,323 Android users from 44 countries, and show that app usage differs significantly by country. However, this prior work has not considered the factor of users' languages in app search results.

Auditing Algorithms. Our work could be viewed as auditing the Google Play search mechanism when searched with queries related to spying, tracking, or monitoring intimate partners. Prior work has audited algorithms, such as those used by YouTube [37] and Amazon [38], by using queries and analyzing search results. Despite not having access to proprietary algorithms, such algorithmic auditing is one way to gain

1 https://github.com/majed-almansoori/IPS-dual-use-android-apps

insight into how recommendation or ranking systems work [46, 52].

Prior work has looked into web search engine behavior in different languages or locations. For example, Arif et al. [15] audited Google search results for anti-vaccine web pages in five languages and found that the information quality algorithm used by Google varied by language. Ballatore et al. [17] investigated the localness of Google's HTML search results pages for 144 countries and 99 languages when searching for information about the capital city of the given country, and found that "wealthy and well-connected countries tend to have much more locally produced content that is visible about them than poor and poorly connected countries." In our study, we also use location and language parameters to analyze aspects of search, here for Google Play's query suggestions, blocking, and search results.

3 Crawling Apps in 15 Languages

We adapted the pipeline created by Chatterjee et al. [19] to analyze apps in multiple languages. We first describe how we picked the languages and countries for the study, then explain our modified crawling pipeline for finding dual-use apps on Google Play in those languages. We also discuss our methods for crawling several popular app stores in China.

3.1 Choosing Languages and Countries

We picked 15 languages and 27 countries for this study based on three factors: number of speakers, number of Android users, and availability of Google Play in that language and country. We first considered languages with over 100 million speakers worldwide [13, 69]. From these languages we removed Urdu, as Google does not support localization in Urdu [12]. Additionally, we added Turkish (tr), Vietnamese (vi), and Thai (th), as they are the languages spoken by most residents in Turkey, Vietnam, and Thailand—countries that have one of the highest Android app download rates [7, 21]. Next, we decided which countries to crawl by considering the ones with the highest number of speakers for each chosen language. The list of the languages and countries we crawled is shown in Fig. 1. Since Google Play is blocked in China, for Chinese-Mandarin (zh) we considered countries with the largest Chinese diaspora populations: Indonesia (id), Thailand (th), and the United States (us) [51]. We also conducted a brief investigation into dual-use apps available in other popular Android app stores in China (see Section 4.4).

#	Language	Countries we crawled
5	Arabic (ar)	Egypt (eg), Algeria (dz), Saudi (sa)
6	Bengali (bn)	Bangladesh (bd)
2	Chinese (zh)	Thailand (th), US (us), Indonesia (id)
1	English (en)	US (us), UK (gb), Bangladesh (bd), Australia (au), Nigeria (ng), Pakistan (pk), India (in), South Africa (za)
7	French (fr)	France (fr), Algeria (dz), Morocco (ma), Canada (ca)
12	German (de)	Germany (de)
3	Hindi (hi)	India (in)
13	Japanese (ja)	Japan (jp)
11	Indonesian / Malay (ms)	Indonesia (id)
9	Portuguese (pt)	Brazil (br), Portugal (pt)
8	Russian (ru)	Russia (ru), Ukraine (ua)
4	Spanish (es)	Mexico (ms), Colombia (co), Spain (es), US (us)
30	Thai (th)	Thailand (th)
16	Turkish (tr)	Turkey (tr)
21	Vietnamese (vi)	Vietnam (vn)

Fig. 1. Combinations of 15 languages and 27 countries crawled. The left-most column shows the rank of the language based on the number of native speakers according to Ethnologue [13]. The language and country codes are shown in parentheses.

3.2 Defining Dual-Use Apps

We aim to find applications that can be used by an abuser to remotely spy on, track, or stalk their victims, also called *IPS-relevant* applications [19, 53]. Not all IPS-relevant apps, however, are designed for spying or tracking intimate partners; many apps designed for non-IPS-relevant purposes, such as tracking a child's location, have features that can be used for IPS-relevant purposes, making them "*dual-use*" apps, as noted by Chatterjee et al. [19].[2] Such an app enables a non-tech-savvy abuser to remotely monitor a victim's whereabouts after installing it on the victim's device. We expand their definition of dual-use apps to assert that: (a) the app must provide capabilities that *can be used* for spying, track-

2 Chatterjee et al. [19] consider an app to be dual-use if: (1) its primary purpose is giving another person the ability to collect data, track location, and/or remotely control a device; (2) it functions, after initial installation and configuration, without interaction with the current user of the device; and (3) the victim most likely does not want it on their device.

ing, and/or stalking, even though they are not meant for such (ab)use, (b) such capability is automated; does not require (ab)user interaction with the victim's device after installation and configuration (if the app requires access to the victim's device), (c) such capability is described in app metadata (text or screenshots) (We acknowledge we might miss apps advertising full capabilities outside of the Google Play Store), and (d) the app might not be installed on the victim's device and might use external hardware such as GPS trackers. Apps in the last category leave no trace on the victim's device; thus, they are difficult to detect and are dangerous from the perspective of an IPS threat model [53]. As we included new apps in the dual-use category, when applicable, we report the number of dual-use apps according to the old definition [19] as well.

3.3 Crawling Pipeline

There are six steps in our measurement pipeline (shown in Fig. 2).

(1) Creating Seed Queries. Our pipeline begins with the manual translation of the English seed queries used in prior work [19] into 14 languages. Specifically, we used Google Translate [2] to create the initial translations and received the assistance of native speakers of those languages who are also proficient in English to verify and correct the translations. Additionally, we added new seed queries recommended by the native speakers, who paid special attention to sociolinguistic differences. The number of seed queries varies across languages (see Fig. 3) due to such differences.

(2) Query Snowballing. As in the study by Chatterjee et al. [19], we used Google Play's query completion API [30] to expand the set of seed queries for each language using a *"query snowballing"* approach. We started with the seed set of queries, and then for each searched query, we collected search suggestions from Google Play until no new queries were found or the total number of collected search terms for a language reached 10,000. Due to query suggestions' variation between languages and countries, the final set of queries we obtained also differed across languages.

(3) Collecting Apps. We used a modified version of an unofficial scraper for Google Play called google-play-scraper [11] to search for apps, adding language (*hl*) and geolocation (*gl*) parameters to conduct a global search without requiring physical servers in multiple countries (Google Play search results may vary slightly based on the IP address of the client). For each language-country pair, we searched Google Play with the final set of seeds and snowballed queries with language (*hl*) and location (*gl*) set accordingly. For each query, we collected metadata, such as app title, descriptions, appId, genres, and permissions, for the top 50 apps shown on the Google Play search results page. Overall, we crawled Google Play for 47 days (from April 28, 2020 to June 13, 2020) per language-country pair listed in Fig. 1. We also distributed crawling across multiple Amazon EC2 instances and grouped all the downloaded data onto a single server for further analysis.

(4) Translating App Metadata. We used spaCy [8] to detect the language of the apps' description, as not all apps' metadata language matches with the language used in the search query or the language parameter *hl*. This situation happens because some developers do not localize their apps (i.e., they do not provide translated versions of their apps for certain search query languages) [31]. Thus, if an app is not translated for the queried language, it is returned in a default language. After detecting the languages of the apps' descriptions, we used the Google Translate API [2] to translate the titles, descriptions, and summaries of the apps into English (if they were not already available in English), in order to utilize our pipeline's rich English-language training data to build machine learning (ML) classifiers for identifying dual-use apps. Hence, we avoided building separate classifiers and preparing the required training data for each language. Further, since many apps are found in multiple languages, we translated each unique app only once if not already found in English.

(5) Classifying Apps. To identify dual-use apps from the set of all apps we found, we used an adapted supervised linear classifier from prior work [19]. The classifier was trained on English apps only. Due to the difficulty of preparing training data and creating separate classifiers for each language, we applied the classifier to translated metadata instead. We recorded the classifier confidence score for each app, picking a threshold of 0.4, such that apps scoring below 0.4 are classified as *not* dual-use. After examining the classification scores, we chose a lower-than-normal threshold (which is usually 0.5) to avoid failing to flag dual-use apps due to low classification scores. Chatterjee et al. [19]) used a threshold of 0.3, but we found that in this translated multilingual app descriptions the threshold of 0.4 provides as few false negatives as threshold of 0.3, and the number of false positives increases significantly from 0.4 to 0.3; there-

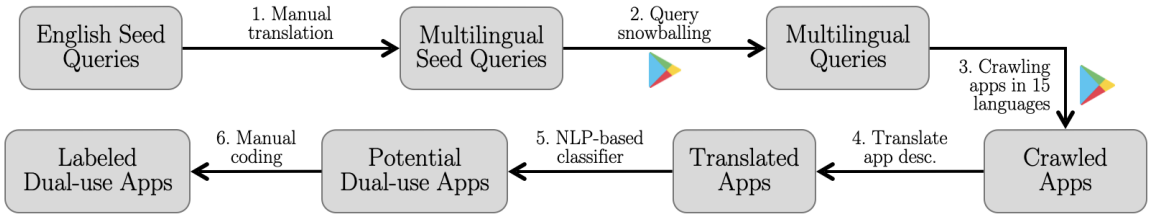

Fig. 2. Measurement pipeline expanded from [19] to find dual-use apps on Google Play in fifteen different languages. This semi-automated pipeline has six stages.

fore, we used 0.4 as our threshold for considering an app as dual-use.

(6) Human Labeling. Finally, we randomly sampled 250 apps that scored more than 0.4 and 100 apps that scored less than 0.4 for each queried language (5,250 apps in total) and manually flagged them. One researcher reviewed every app from the samples and labeled it as dual-use or not based on the app's metadata (mainly title, summary, and description, with occasional reference to the app page on Google Play). Although each app had a single reviewer, the research team discussed the reviewed apps and their labels as a group. We also assigned each app to a category (e.g., vehicle tracker, auto call recorder, etc.) and coded its spying capabilities. For many apps, the metadata is not conclusive as to whether the app has spying capabilities or not, usually due to short descriptions. In these cases, the researcher evaluated screenshots, reviews and permissions obtained by searching for the app identifiers of these apps on Google Play or other Android apps download websites (e.g., APKPure), if the app was no longer available in Google Play. For example, we found several call recorder apps advertising screenshots showing an "automatic recording" option not mentioned in the description. Furthermore, some apps, such as "Phone Number Locator" apps, received high scores from the classifier (considered dual-use), despite having no dual-use capabilities. Thus, using screenshots, reviews, and permissions in such cases helped correctly determine the dual-use capabilities of ambiguous apps. When we could not conclude whether apps were dual-use due to the lack of descriptive metadata and screenshots, we decided to label them as dual-use, to err on the side of caution. Based on our labeling, the estimated percentage of such apps in our dataset is 1.22%. In total, we found 854 unique dual-use apps through manual coding.

3.4 Limited Efficacy of Static Analysis

We explored the use of static analysis of Android package (APK) files to further prune the results returned by the NLP classifier, developing a static analysis classifier. Apps that received an NLP classifier score ≥ 0.4 were downloaded using a utility called gplaycli [43]. We selected a sample of 1,655 apps (about 50% of them had been manually flagged as dual-use apps). Of these, 9% could not be downloaded by gplaycli, often due to the app being a paid app. We downloaded 1516 apps in total. We then used the droidlysis [20] framework to extract static features from the apps. Droidlysis can extract several properties, e.g., imported libraries, use of certain permissions, etc. but does not work well for APIs. We used androguard [23] for extracting API calls from the APKs. The features we included in the analysis are permissions, receivers, content providers, services, API calls, and additional properties, such as whether a set of specific methods are used that could be a sign of exfiltrating sensitive data or malicious behavior (e.g., code for hiding the app icon from the app drawer, which was often used by some APKs from the 2018 study). Some apps could not be parsed correctly using droidlysis. In the final sample, 1333 were used for analysis. We treat each app as a "bag of features" and then apply one-hot encoding. We tried several ML algorithms, including decision trees, logistic regression, and random forest. We used $k = 4$ fold cross-validation to compute the accuracy of the classifier. For the 1,333 apps we analyzed, the best precision we could achieve was 60% when using Random Forest. However, this also has a low recall value of 70%, indicating a high number of false negatives i.e. dual-use apps missed by the classifier. To reduce the false negatives, we used a lower than normal threshold of 0.29, which resulted in a recall of 96%, but the precision decreased significantly to 40%. The high false discovery rate means that a large number of apps flagged by the static analysis classifier are not dual-use apps. We estimate that a static analysis classifier could

prune approx. 30% of apps that are assigned ≥ 0.4 score by the NLP classifier.

We may have achieved low accuracy because the classifier was used only on apps that had a high NLP classifier score (≥ 0.4) and also because these apps were likely to have similar functionalities with minor differences that are hard to differentiate. For instance, the app "Smart GPS Tracker" was correctly identified as dual-use while the app "Find lost phone: Anti-theft protection" is incorrectly flagged as dual-use while having more dangerous admin permissions and similar location tracking functionalities. Yet, the latter app can only help find lost phones by responding to claps and whistles, and dangerous device admin permissions are requested to prevent unwanted removal of the app by the thief. Therefore, this app cannot be used for stalking. Although static features such as permissions and API calls have been used with some success to detect malware [14, 72, 73], we hypothesize that the dual-use apps returned by the NLP classifier do not differ significantly on a static level. Thus, static analysis cannot replace manual analysis. We decided that the overhead — downloading APK files and extracting static analysis features — incurred in integrating static analysis into our app analysis pipeline outweighs the potential benefit of saving some manual labeling and, therefore, did not pursue this approach further.

3.5 Ethical Considerations

Minimal-impact automated scraping of publicly available information not containing personal identifiers is generally considered acceptable research behavior. We made sure to not make more than ten queries per second, and we believe such a load would have a negligible impact on the Google Play service's normal operation. Similarly, we ran rate-limited search queries on Chinese app stores. Our research does not introduce new harms or exploit existing vulnerabilities in novel ways. Instead, we audit readily available search results and analyze them to shed light on what types of dual-use apps are available in app stores.

4 Results

From our 47-day multilingual search on Google Play using our measurement pipeline, we collected thousands of query suggestions and applications. We analyzed the types of queries suggested (Section 4.1), the prevalence of dual-use apps in different languages (Section 4.2), and their capabilities to surveil an intimate partner (Section 4.3).

4.1 Google Play Query Suggestions

Google Play provides search query recommendations to help users search for apps more effectively. Chatterjee et al. [19] used this feature (in English) to gather relevant search terms that an abuser might use to search for apps to conduct IPS. They refer to this process as *query snowballing*. We extended the approach to gather and analyze search queries in 15 different languages.

Query Snowball Sizes. We find that query suggestions differ significantly by language and also, within the same language, by country. For each language, the number of seed queries used and the cumulative number of new queries gathered over the whole measurement period is shown in Fig. 3 (columns: "Seed", "New", and "Total"). As can be seen, there is a huge variance in the number of unique queries obtained and used for searching in each language. For example, we searched in eight countries with English and thus obtained 2,576 new unique query suggestions in total. Yet, for some languages, such as Bengali, Thai, Turkish, and Vietnamese, we received only a handful (≤ 100) of query suggestions. For Japanese, we received no query suggestions.

IPS Query Suggestions. We define an IPS query as a query explicitly referring to tracking or monitoring an intimate partner. We reviewed all 7,244 search queries to identify IPS queries, by translating queries into English using Google Translate and then manually flagging whether they were IPS queries. If a translation was not clear, possibly due to an error in translation, we used Google search to determine if the query refers to IPS. We found 1,012 IPS queries across all languages.

Fig. 3 shows the number of IPS queries we used for searching. The majority of seed queries are IPS queries, as intended. However, Google Play also suggested several IPS queries, especially in non-English languages. For example, in Indonesian/Malay (ms), the store suggested 97 additional IPS queries, like "*aplikasi melacak no hp pacar*" ("application to track girlfriend/boyfriend's cellphone number") and "*apk pelacak lokasi pacar*" ("boyfriend/girlfriend location tracker apk"). Similarly, several IPS queries were suggested in Spanish (es), such as "*rastreador de mi esposo*" ("my husband tracker") and "*aplicación para rastrear a mi*

Lang	Queries		IPS		Blocked	
	Seed	New	Seed	New	Total	IPS
Arabic	**190**	454	77	6	13	9
Bengali	117	58	72	**0**	**0**	**0**
Chinese	73	136	37	**0**	**0**	**0**
English	79	**2,576**	41	12	**23**	**22**
French	128	231	66	18	4	2
German	119	320	**79**	7	6	**0**
Hindi	68	239	**20**	**0**	**0**	**0**
Japanese	76	**0**	42	**0**	1	**0**
Malay	133	674	77	**97**	**0**	**0**
Portuguese	113	322	56	34	9	9
Russian	79	269	43	2	4	2
Spanish	101	606	48	56	8	6
Thai	**48**	73	29	**0**	**0**	**0**
Turkish	69	48	28	1	1	**0**
Vietnamese	86	90	51	**0**	**0**	**0**

Fig. 3. Numbers of unique queries used by our scraper: original "Seed" queries, "New" queries suggested by query snowballing, queries showing explicit "IPS" intent, and "Blocked" queries that return no search results. The highest and least values in each column are highlighted in bold.

mujer" ("application to track my wife").

Minimal Blocking of IPS Queries. We suspect that Google Play is blocking some queries to prevent users from searching for apps with malicious intent, such as "tracking my wife". For these blocked queries, Google Play returns no apps and no query suggestions. Though the majority of such blocked terms are IPS queries, not all IPS queries are blocked (as shown in Fig. 3, "Blocked" columns). Also, query blocking varies widely between languages. For example, among the 41 English seed queries with IPS intent, 23 (56%) are blocked. However, 16% of IPS seed queries are blocked in Portuguese, Arabic, and Spanish and 5% in French and Russian. None of the IPS seed queries are blocked in 9 of the 15 languages. This also means it is easy to bypass the query blocking by simply translating a query (see Section 5). Google Play does not block queries that it suggests.

4.2 App Search Results

By querying with all the (not blocked) seed terms and their suggestions for 47 days, we obtained 51,868 unique apps across 15 languages. On average, we downloaded 5,628 apps per day across all languages (some apps were downloaded multiple times in different languages). The total number of apps obtained per language (shown in Fig. 4, "Apps Found" column) differs significantly. Due to technical issues, our scraper failed to collect any apps on certain days for some stores, and our scraper for Vietnamese failed to collect any data after May 1, 2020. Since we still obtained thousands of apps for all languages, we kept them for further analysis. We believe the number of apps obtained for a language is directly related to the size of the query snowballs and the number of countries searched for that language. For example, the five lowest counts of apps found correspond to the five languages with the lowest number of queries searched, which were also each searched for one country. While some of our queries were blocked by Google Play, the blocking did not significantly reduce the number of results.

App Localization. We found that app descriptions are not always in the language used in the search query or specified by the language parameter *hl*. Also, the metadata of an app does not specify the languages the app is localized for — the languages into which the developer translated the app metadata. Therefore, we consider an app *localized* for a language if the app description can be found in that language. We used spaCy [8] to detect the language of each app. The detected language results are shown in Fig. 4 ("App Language" columns). In total, nearly 52% of the apps obtained across all queried languages had English descriptions, although only 13% of the apps were found via English queries. This shows that some apps were not localized for the given search language. However, we also found that at least 46% of apps in each language are available in the queried language, except for Hindi (33%). We also found that a small number of apps per language had descriptions in languages other than English and the queried language. Understanding these distributions requires further investigation into how Google Play works, which is left for future research.

To find the number of apps that are localized for English, we downloaded the description of each of the 51,868 apps after setting $hl = en$ and $gl = us$, and then detecting the language of the downloaded apps using spaCy. We found that 27,065 (52%) of these apps have English descriptions on the Google Play Store, and thus that 48% of them are not available in English. Thus, multilingual searches resulted in better app coverage. We evaluated the accuracy of language detection by randomly sampling 20 apps for each detected language with more than 100 apps. We found that all languages were detected with an accuracy of more than 95% except for Korean, which was detected for a set of mostly Chinese

Lang	Apps Found	App Language (%) Eng.	Query	Other	Clf. Acc. (%) Recall	Precision	# dual-use Lab.	Estim.	Capabilities (%) L	C	A	S	R
Arabic	7,969	31	68	2	**100**	33	83	581	**46**	24	12	63	24
Bengali	2,861	48	48	4	73	**30**	**78**	315	**77**	24	12	51	29
Chinese	6,649	46	48	**7**	84	**53**	**135**	846	62	21	16	40	8
English	**15,967**	**99**	-	**1**	86	39	98	**1,120**	47	**13**	6	43	36
French	6,742	36	60	5	83	40	103	783	50	21	6	54	27
German	7,660	50	48	3	74	42	107	596	68	17	13	40	**2**
Hindi	3,735	65	**33**	3	**100**	46	116	**245**	54	19	**5**	53	**37**
Japanese	2,756	**21**	**74**	5	91	42	107	293	74	**28**	8	51	8
Malay	5,432	51	48	2	**100**	33	83	321	55	27	13	**65**	14
Portuguese	7,997	42	52	7	**69**	38	98	760	55	15	9	52	21
Russian	5,869	31	67	2	75	40	103	691	49	17	15	45	26
Spanish	9,915	28	71	**1**	89	47	118	886	64	14	7	**39**	15
Thai	2,821	51	46	3	79	52	133	391	66	25	14	53	14
Turkish	3,241	24	**74**	2	**100**	39	97	303	58	23	14	48	29
Vietnamese	**2,354**	39	60	2	84	50	128	292	48	22	**20**	51	25

Fig. 4. This table shows: 1) the total number of unique apps we collected by searching in each language; 2) the detected language of app description — "Eng." if in English, "Query" if in the same language as the queried language, and "Other" for any other language (Note that an app might exist in multiple languages); 3) the accuracy of the ML classifier ("Clf. Acc.") in terms of "Recall" and "Precision"; 4) the number of dual-use apps found by manually labeling a sample of 350 apps in each language ("Lab.") followed by the estimated number of dual-use apps in this language ("Estim."); and 5) the distribution of capabilities of found dual-use apps to Locate (L), Control (C), Access (A), Share (S), or Record (R) a victim's information (a single app can have multiple capabilities). The highest and least values in each column are highlighted in bold.

apps; thus, we counted these apps as Chinese.

Dual-Use Apps Found in Multiple Languages. Of the 1,587 apps that we labeled as dual-use across all languages, only 854 are unique. This suggests that many dual-use apps are found in more than one queried language. To understand how these apps are distributed across languages, we plot the cumulative distribution of apps found in multiple query languages; specifically, the fraction of apps that are found in at least x languages. We first plot the curve for all 51,868 apps we found through our crawling (solid "All apps" line in Fig. 5). Then, we also plot the curve for 854 dual-use apps (solid "Dual-use apps" line in Fig. 5), showing that they are typically found in multiple languages. While only 5% of apps, in general, are found in six or more languages, 45% of dual-use apps are found in six or more languages. Indeed, 77.4% of them are found in at least two languages. Thus, dual-use apps are typically localized for many languages, making it easier for abusers to find them.

We also found that 192 (22.5%) of labeled dual-use apps are unique to only one queried language, indicating that some dual-use apps have limited availability and may not be found using other languages but only via a language-specific app search (as in our pipeline).

In Fig. 5, we also plotted the percentages of apps with metadata found in multiple languages as detected by spaCy. About 53% of dual-use apps had metadata in only one language, suggesting that half of dual-use apps are not localized — even though they are found using those languages; even fewer have metadata in more than two languages. This suggests that although many apps are found using multiple query languages, fewer apps are localized.

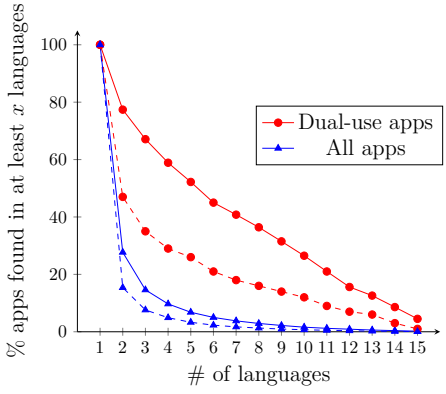

Fig. 5. The percentages of dual-use apps and all apps found using at least x queried language(s) (solid lines) and the metadata found in x detected language(s) (dashed lines).

Estimating the Number of Available Dual-Use Apps on Google Play. Although our search procedure found 51,868 apps, not all qualify as dual-use. We

pruned the list using a semi-manual filtering process. We used a machine learning classifier to remove the apps that are obviously not relevant, scoring below 0.4 (see our measurement pipeline (5) in Section 3).

For each language, we manually labeled a sample of 100 apps scoring < 0.4 and 250 apps scoring ≥ 0.4 to compute the precision and the recall of the classifier. We compute the precision as $\frac{\mathsf{TP}}{250}$, where TP is the number of apps (scoring ≥ 0.4) that are manually flagged as dual-use. For recall, we need to estimate the total number of dual-use apps for each language. We estimate the total number of dual-use apps in a language as $\hat{\mathsf{P}} = \frac{\mathsf{TP}}{250} \times n_+ + \frac{\mathsf{FN}}{100} \times n_-$, where n_+ and n_- are the total numbers of apps scoring ≥ 0.4 and < 0.4, respectively. Here, FN is the number of apps flagged as dual-use in the sample of 100 apps scoring < 0.4. Then, recall of the classifier in a language is calculated as $\mathsf{TP}/\hat{\mathsf{P}}$. The precision and recall are shown in Fig. 4 ("Clf. Acc." columns).

In Fig. 4, we also show the estimated number of apps ($\hat{\mathsf{P}}$) that were found by searching each language individually. We estimate that in our sample, searching in Chinese and English resulted in more than 600 apps, and in French and Spanish, we estimate having found more than 500 dual-use apps in Google Play.

Not all apps across languages are distinct. Therefore, we cannot add them to get the total number of estimated dual-use apps. Instead, we consider the total number of unique apps with a score ≥ 0.4 — which is 8,226 — and multiply it with the average precision of our classifier across languages, which turned out to be 41.7%. Thus we obtain the total estimated number of dual-use apps as $3,988$. Note, this is a lower bound, as our classifier's recall is not always 100%, and this estimate is missing those dual-use apps that score below 0.4.

ML Classifier is Fooled by Irrelevant Apps. We obtained lower precision for classifying dual-use apps than prior work [19], despite raising the classifier score threshold to 0.4. Many apps were unrelated to spying or tracking but used keywords such as "GPS," while some apps mention tracking capabilities in their description but do not actually possess any of them. Some apps repeatedly mention a single term such as "GPS" or "SMS" in their descriptions, which likely made the classifier assign a very high score, although the app is not dual-use.

Many of these apps are obtained through seed terms such as "SMS tracker," which suggests that the store search engine also gets confused, or "poisoned," by such terms. Search engine poisoning ("SEP") [41] is a well-known strategy for boosting websites' search engine rankings. App developers may be using similar methods to boost the Google Play search rankings of their apps, even if some queries are related to IPS or stalking. Across all languages, we noticed that app descriptions are normally poisoned with "family locator", "GPS locator", and "location finder", hinting that developers might be trying to exploit the interests of some users for such apps. We did not classify some apps as dual-use because they lack automated spying capabilities. For example, many call recorders are manual, making them benign since the abuser cannot automate recording. Similarly, some location-tracking apps do not allow continuous location tracking, and they require explicit user interaction each time the location is shared. Such semantic details are hard to parse with the ML classifier we used, and thus these apps received high scores from the classifier, leading to false positives. Future work could explore ways to improve the classifier so that it considers such nuanced details in app descriptions. For now, we rely on manual inspection.

4.3 Characteristics of Dual-Use Apps

Among the 5,250 apps we manually labeled, 832 are considered dual-use, and only three dual-use apps explicitly stated that their app can be used for IPS. We assessed the stalking capabilities of the dual-use apps, including differences across languages.

Distribution of Stalking Capabilities. We analyzed the labeled dual-use apps and categorized them based on five stalking capabilities, described below and shown in Fig. 6. Detailed descriptions of the capabilities are given in Appendix B. We modified the categories created by Chatterjee et al. [19] slightly to include additional types of dual-use apps found by our crawlers.

The five categories we used are:

- Locate (L): Apps that can find the precise location of a device and share it remotely.
- Record (R): Apps that can automatically record calls, videos, the screen, or voice audio.
- Share (S): Apps that can record and share data other than location data or data recorded by the app. This includes call logs, SMS, keystroke logging, etc.
- Access (A): Apps that can remotely access the camera, microphone, or screen of another device.
- Control (C): Apps that can provide any type of control over another device's settings (including and beyond camera and microphone).

An app can have multiple capabilities, or even all of

Category	App Type	Description
Locate (L)	Personal Tracking	Track your own device's location remotely
	Mutual Tracking	Mutual location sharing with friends, family and partners
	Subordinate Tracking	Track the location of children or employees
	GPS Trackers Apps	Track GPS trackers installed on cars, attached to pets, etc.
Record (R)	Voice Call Recorder	Record incoming and outgoing voice phone calls automatically
	Video Call Recorder	Record incoming and outgoing video calls with audio automatically
	Screen Recorder	Record the screen in the background at given intervals
	Voice Recorder	Schedule automatic voice recording in the background
	Video Recorder	Use the camera to record videos in the background at predefined times
Share (S)	Auto Backup/Sync	Automatically backup data locally or sync to the cloud
	SMS/Logs Forwarder	Forward SMS or call logs to another phone or send them via email
	SIM Change Detector	Get notified when a phone's SIM is removed or changed
	Last Seen Tracker	Get notified when people appear online on WhatsApp and other apps
	WhatsApp Cloner	Use the same WhatsApp account on two or more devices
	Keylogger	Record which keys were selected by the user; logs keystrokes
	Device Reports	Provide reports about battery, app usage, websites visited etc.
Access (A)	Camera Access	Control IP cams remotely or turn your phone into an IP cam
	Microphone Access	Access the microphone of a device remotely
	Screen Mirroring	Mirror the screen of another device onto your device
Control (C)	Parental Control	Limit screen time, delete apps, block websites, and more
	Phone/Tablet/PC Control	Gain remote access and control of the device and its data

Fig. 6. Categories of dual-use apps based on their capabilities. The "Locate" category includes apps that can track the location of another device or share the location of the current device automatically. The "Record" category contains apps that can record the screen, calls, or videos automatically or in timed intervals. The "Share" category includes apps that collect or send any information to other devices automatically. The "Access" category is for apps that grant the user real-time access to the device's camera, microphone, or screen. The "Control" category is for apps allowing full or partial control over the device.

them. One researcher manually assigned relevant capability flags to each app. The research team randomly picked a sample of apps to discuss and came to agreements regarding definitions and assignments of the labels. The percentage of dual-use apps that possess each capability is shown in Fig. 4.

At least two-thirds of dual-use apps in Bengali, German, and Japanese can locate (L) a device remotely through a phone connection, smart device, or a dedicated GPS device. Locate (L) is indeed the most prevalent capability that dual-use apps have in all languages, except in Arabic, French, and Malay (where most apps have share capability). These apps are advertised with the use-case of tracking family members, friends, and children, finding stolen or lost devices, storing and sharing location history, and keeping track of vehicles.

The second most prevalent capability was Share (S). Dual-use apps had various types of sharing capabilities, such as a remote (pre-configured) alert about SIM card changes or wrong unlock patterns. Some apps share reports about battery, app usage, browser logs. We also include social media "last seen trackers" as having the share capability.

Access (A) was the least prevalent capability in our dataset, with less than 10% of apps in six languages and less than 21% in the remaining nine languages having the capability of remotely accessing the camera, microphone, or screen. A higher number of apps had *Parental Control* capabilities that prevent the user from accessing certain apps or websites at certain times. Although apps with this type of control (C) capability might not enable spying on someone, they can be used to attempt to control a victim by limiting their access to the device.

Most dual-use apps in our samples had one or two dual-use capabilities, and less than 20% had at least three capabilities. We found 6 ($< 1\%$) dual-use apps that had all five capabilities. Typically, these are anti-theft applications, such as the app titled "Lost Android". This dual-use app allows location tracking, remote recording, camera control, device control, and data sharing. Such apps seem to be limited, even in our dataset. Overall, the results suggest that dual-use apps found using our search terms across all 15 languages would likely be capable of tracking location, sharing information, or controlling access, which can be dangerous in IPV contexts.

Popularity of Dual-Use Apps Based on Installs. To learn more about the use and popularity of dual-use apps, we looked at the number of installs for each dual-use app. Most apps were downloaded at least a thousand times, while few were downloaded more than a hundred million times. We found that 80 English and 73 non-English dual-use apps were downloaded more than a million times. The distribution of download counts was similar for English and non-English apps in general.

4.4 Prevalence of Dual-Use Apps in Chinese Android App Stores

China has nearly a billion Android users [59], but Google Play is blocked in China. Therefore, we crawled four popular app stores in China, namely Xiaomi [6], Baidu [1], Tencent [9], and Huawei [4], and measured the prevalence of dual-use apps in these stores.

Crawling Chinese App Stores for Two Days. We searched for dual-use apps using seed queries translated into Mandarin and the set of relevant search queries in Chinese obtained from Google Play (step 2 in our pipeline described in Section 3.3). (Some app stores suggest related queries, as Google Play does, but we did not look into those suggested queries.) We observed that the results of search queries for some stores varied based on the user's region. Hence, we ran the crawler for all stores from a server based in Hong Kong. We then repeated the remaining stages of the pipeline for each app store. We crawled the stores twice, on September 24, 2021, and October 1, 2021. We collected consolidated data from both runs for analysis. A total of 2,484 apps were collected across stores (Xiaomi: 445, Baidu: 244, Tencent: 562, Huawei: 1,233), of which 2,325 were unique. Of the 2,325 unique apps, 654 were also available on Google Play, based on their app IDs. We manually coded all 2,325 apps after translating their metadata into English and found 110 dual-use apps, only 17 of which are available on Google Play. Of the stalkerware apps repeated across the Chinese app stores, 13 apps appear in at least two stores, two apps appear in at least three stores, and one app appears in all four stores. To verify that app IDs across different stores correspond to the same app, we randomly sampled 25 apps and found all apps sharing an app ID to be the same. The majority of the 110 dual-use apps (65%) can locate devices and some also possess capabilities to Record (20%), Share (20%), Access (27%), and Control (25%).

5 Takeaways and Discussion

Dual-use apps present a significant threat to victims of IPV, and identifying them is challenging not only for victims but also for developers building tools to help victims and advocates automatically detect dual-use apps. As our results suggest, thousands of these apps are marketed across languages for different use cases, many of them being dual-use apps that are not easily classified as dual-use. Comparing searches across four years (2018, 2020, and 2022), we found that there is significant turnover in the Google Play Store market and that even the wording of app descriptions has changed over time. Yet, our findings suggest that a multilingual approach can detect more dual-use apps than a monolingual approach, encouraging a broader scope in IPS research.

Comparison With Prior Work. Prior work [19] found 2,473 dual-use apps in 2018. Of these apps, only 411 (17%) were still present on Google Play in 2020, 387 were found via our crawling, and only 320 of these were found using English queries. This is surprising, as the English seed terms we used for finding dual-use apps were exactly the same as the ones used by Chatterjee et al. [19]. We believe that differences in suggested terms and query blocking may have contributed to this difference, though we do not have access to all search queries found in this study [19] and therefore cannot verify the reason for this disparity concretely.

The majority (83%) of dual-use apps flagged in the 2018 study are no longer available from Google Play. We suspect this is due to the 2020 revisions to Google's policies regarding spyware and stalkerware, in which apps that explicitly promote stalkerware, as well as the ads that market apps for IPS, were banned [3, 32, 33]. Among the 411 apps that are still present in Google Play, there are apps for tracking family members' or children's locations, such as Life360 or MMGuardian, apps for finding stolen or lost phones, such as Mobile GPS Location Tracker, and apps for automatic call recording. All couple tracking apps reported in the 2018 study were discontinued. However, we note these apps are still available on unofficial Android application stores and can be found via a simple Google Search.

Among the 854 unique apps we manually flagged, we only found 605 (71%) apps that meet the definition of dual-use from 2018 (ignoring the vehicle tracking apps and apps for tracking social media activity). Thus, the precision of the NLP-based classifier is significantly lower than what was seen in 2018. We tried retraining the classifier, but that did not improve the accuracy.

We suspect that the set of dual-use apps is smaller due to a few reasons. First, we believe the overall number of dual-use apps has decreased significantly on Google Play since 2018, also due to the aforementioned changes in Google's policies. Second, there could have been a significant concept drift [58] among how dual-use apps describe themselves since 2018, for example, by suppressing mentions of their (spying or IPS) capabilities. Thus, our NLP classifier may fail to identify potential dual-use apps.

Multilingual Search Helps Find More Dual-Use Apps. Searching Google Play using non-English search queries helped find many dual-use apps that we would not have found otherwise. We found 1,018 dual-use apps in our 2020 app dataset (including apps manually flagged in 2018, in addition to the 854 we manually flagged in 2020). We analyzed the language data for all 1,018 dual-use apps' descriptions and found that a total of 849 (83%) apps had English descriptions on Google Play, but only 739 (73%) were found via English searches. Thus, multilingual searches help to find more apps, even in English. In our dataset of apps, the likelihood that a dual-use app is available in English is higher (83%) than the likelihood that an app, in general, is available in English (52%). However, there are 279 (27%) dual-use apps that were never found in English searches. Then, to obtain better coverage of dual-use apps, we can search in multiple languages.

As shown in Fig. 5, dual-use apps are mostly available in only one to three languages, suggesting that localization in more languages (among our 15 languages) is not that common across dual-use apps, or that our crawler did not manage to find apps localized in many languages within the top 50 apps returned per query.

Surveillance Capabilities of Dual-Use Apps. The majority of dual-use apps across all languages had the capability to locate (L) (see Fig. 4). While most of the dual-use apps we found were advertised as family safety tools to track family members or anti-theft apps to find stolen devices, we also found several GPS device trackers that work with specific proprietary GPS devices. Other surveillance capabilities of dual-use apps vary across languages. For example, in German, we found only one app (out of 105) with automatic recording (R) capability, and very few in Chinese and Japanese; whereas in Hindi, we found 37% of apps had recording (R) capability. Few apps, on average, had the capability to remotely access (A) information in a device. However, we found that 20% of dual-use apps in Vietnamese had this capability. Differences in capabilities could stem from privacy regulations across countries we crawled, but investigating such possibilities was not within the scope of this study.

	2018	2020	2022	Total
# apps found	14,461	51,868	27,584	78,544
New apps found	–	46,370	18,177	–
# apps ≥ 0.4 score	3,884	8,226	4,979	15,139
Estm. # dual-use	2,473	3,988	2,025	–

Fig. 7. Survey of dual-use apps on Google Play across four years. Note, in 2018, Chatterjee et al. [19] only looked for apps in English (`en`) and in the USA (`us`).

Searching Other App Stores. In some countries, other stores are more popular than Google Play (especially if it is blocked in the country) and may have different dual-use apps. Of the 110 dual-use apps found on Chinese app stores, 93 (84%) were not found in the Google Play Store. The number of apps found varies greatly across the analyzed stores, indicating that the ecosystem of dual-use apps in these stores is varied and complex. Prior work has shown that dual-use apps are available for iOS, the second most popular mobile operating system [19, 48]. Future work should seek to understand the ecosystem of dual-use apps across more app stores in different languages and regions.

Survey of Dual-Use Apps on Google Play Over Multiple Years. To understand the changes in the dual-use app ecosystem on Google Play since our crawling in 2020, we crawled Google Play again in all fifteen languages for one day in February 2022 with the queries we obtained and used in our 2020 crawling. We found a total of 27,584 unique apps, out of which 18,177 (68%) were only found in 2022. This shows a high churn rate of Google Play apps and search results. We report the apps found for each language in Fig. 9. We found the distribution of apps in different languages was similar to what we found in 2020. For each language, we also flagged 50 random apps classified as potentially dual-use by the NLP classifier with confidence ≥ 0.4, and 50 random apps with confidence < 0.4. Based on this sample, we estimate that nearly 2,025 dual-use apps are in the 2022 dataset. The apps retrieved via crawling and the estimated number of dual-use apps we found in the crawl are noted in Fig. 7 for different years. We also include a Total column, because while many apps from 2018 were removed from Google Play, they can be found on unofficial Android application stores. As we can see, there is a drop in the number of dual-use apps after the 2018 study.

Evading Query Blocking. Google Play blocks some IPS-specific queries, such as "track your husband", probably to prevent potential abusers from finding dual-use apps (Section 4.1). However, we found that this query blocking can be easily bypassed. We tested a small set of variations and obtained relevant results that contain several dual-use apps. These are some variations we tested to avoid query blocking: (a) translating the query to a different language, e.g., "track my wife" (en, blocked) to "menjejaki isteri saya" (ms, not blocked), (b) inserting a simple typo, such as "track my wifes" and "track my wifè", (c) adding irrelevant words, as with "spy wife may", (d) replacing words with synonyms (e.g, "track my couple" instead of "track my spouse", and (e) switching word order (e.g., "wife my track"). Of course, this is not an exhaustive list of changes and there are potentially other variations to evade query blocking. While this blocking might have affected our survey of Google Play (as we did not try query evasion during our automated crawling), how current IPS query blocking is handled by perpetrators needs to be further analyzed. A persistent abuser — a real threat, as observed in prior work [36] — can easily bypass Google Play query-blocking to obtain dual-use apps. Such easy evasion tactics could be prevented by blocking semantically similar queries, as evidenced by Varelas et al. [68]. Implementing these types of strategies to prevent successful blocking evasion can leverage the fact that Google Play search utilizes semantic search (we empirically evidenced this by obtaining relevant results even after modifying the query). We recommend future research on the efficacy of blocking, as well as interventions, such as warning boxes like those shown for searches on self-harm, that discourage people from engaging with the search result content.

Dual-Use Apps Outside Official Stores. We only focused on popular application stores such as (the official Android app store) Google Play, and application stores popular in China, such as Xiaomi, Baidu, Tencent, and Huawei app stores. Installing apps from these stores is much easier than installing apps from outside of these stores; for example, users do not have to explicitly enable the flag for installing from "unknown sources" in Android settings. Nevertheless, it is possible that many dual-use apps are distributed outside these stores, and our current study will miss those apps. Future work should look into the prevalence of dual-use apps outside of official stores in different languages and countries.

Disclosure of Dual-Use Apps. Prior work has influenced Google Play's policies and helped reduce the number of apps that promote IPS explicitly. All manually-flagged dual-use apps, including three dual-use apps that explicitly stated that their app can be used for IPS, as well as recording apps (in violation of a recently implemented Google policy [28]), were reported to Google. We also reported the suggested IPS-relevant queries and different effective query-blocking evasion techniques we found. For the Chinese app stores, we reported the dual-use apps and any policy-violating apps we found.

Limitations of Our Pipeline. Although we tried to capture a comprehensive view of dual-use apps, our approach is not without limitations. First, we did not download, install, and execute each app to measure its dual-use capabilities. Instead, we used Google Play app descriptions and other content (such as images, reviews, and permissions) to manually determine if an app can be used as dual-use. This approach was also used by Chatterjee et al. [19]. This can lead to both false positives and false negatives: apps might promote capabilities not actually offered, and apps might hide capabilities from the description and disclose them after installation. We argue the former false positive is harmless [19, 36], though it might inflate our estimates of dual-use apps available on the Google Play Store. The latter issue of false negatives is concerning, as we might not account for a dual-use app that poses a real threat, since we assume that apps marketed to users on Google Play tend to disclose and highlight all of their capabilities. Future research could use static and dynamic analysis [57] to improve the accuracy of detecting dual-use apps. However, such analysis will require access to the app executable file, and obtaining tens of thousands of app files could be challenging. We do not know all the dual-use apps in our dataset, as this would require manually labelling at least 8,250 apps scored higher than 0.4 by our ML classifier. We instead took fixed-size stratified samples from the apps found in each language. This might not give us an unbiased sample from all dual-use apps in our dataset but provides an understanding of dual-use apps available in different languages.

Our study was limited in its global reach. First, Google, an American company, runs the Google Play Store, which may influence its availability or appeal in some countries. Since Google Play is blocked in multiple countries, including China, we did a small study on app stores available in China (Section 4.4). For a more comprehensive understanding of the dual-use ecosystem in multiple languages, future work should look at other popular app stores. Second, we often cannot tell where an app originated. While Google Play provides

the names of companies and developers, it does not always indicate their addresses. Language and country data for apps do not necessarily indicate their origin.

Our seed queries are not a comprehensive set of variations on our original English queries that abusers might use to search for IPS apps. Our team is multilingual and can understand seven of the fifteen languages, and we verified each search term's translation with a native speaker and made recommended changes. Yet, some languages or dialects have great varieties of expression. (e.g., Malay and Indonesian, or various dialects of Spanish or Arabic). Thus, we might miss out on some specific synonymous queries for some languages. We believe Google Play query suggestions made up for some of the IPS queries missed in our seed sets, in some languages more than others, as we show in Section 4.1.

App search results vary significantly between languages and countries, and when analyzing our results, we observed patterns in app results, such as a significant number of certain types of apps for some languages, e.g., "love SMS" apps that provide romantic quotations. We hypothesize that this may be based on the popularity of searches or search results, but our lack of insight into Google Play analytics prevents us from confirming this. App results were also influenced by query snowballing. We do not treat patterns identified on the basis of query suggestions or app search results as representative of a culture [45]. The locations (IP addresses) from which the searches were done may affect the results returned by Google Play. The EC2 servers we used for our crawling in 2020 and 2022 were all located in the US. We believe the geolocation (gl) and language (hl) are the two primary parameters affecting the search results and did a small study to test this hypothesis: we used two machines, one located in the UAE and another in the US, and searched with 10 Arabic search queries from our seed set with $gl = ar$ and $hl = sa$. We did not find any significant difference in the search results (<3% apps). To avoid this small discrepancy, future work could try to use an IP address in the same country as the gl.

Finally, we relied on machine translation to translate app metadata, since it would have been impractical to rely on human translation for the large number of apps found. We admit that machine translation has limitations and inaccuracies; however, recent work showed that Google Translate has high accuracy for many languages [63]. Moreover, we observed that the translated text was still relevant and understandable; thus, we believe that machine translation was sufficient in our case.

6 Conclusion

This measurement study expands the current understanding of dual-use apps available in various languages and countries in the largest mobile application store, Google Play. We searched for apps related to spying on, tracking, monitoring, or controlling intimate partners in 15 different languages in 27 countries for 47 days. We collected 51,868 unique apps, out of which 24,803 apps are not available in English. We found that 17% of dual-use apps do not have an English description and 26% of them were not found using English queries, underscoring the utility of using multilingual search techniques to better understand the availability of dual-use apps on Google Play. We highlight several key findings in the way Google Play suggests and blocks queries. Furthermore, search results found in multiple languages with IPS-specific queries yielded a high rate of dual-use apps. We also identified problems in Google Play's current blocking mechanisms to prevent IPS-specific queries: an abuser could use simple blocking-avoidance techniques, such as translating the query, to find apps that can be used for IPS. Compared to prior work, we found fewer dual-use apps, but our results indicate that the dual-use app market is still thriving globally. More work is needed to mitigate the threats posed by apps with surveillance capabilities.

Acknowledgement

We thank the anonymous reviewers for their insightful feedback and Urs Hengartner for shepherding this paper. This work was partially supported by the University of Wisconsin—Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

References

[1] Baidu App Store. https://shouji.baidu.com.
[2] Google Cloud Translation API. https://cloud.google.com/translate.
[3] Google Play Terms of Service. https://play.google.com/about/play-terms/index.html.
[4] Huawei AppGallery. https://appgallery.huawei.com.
[5] ips_queries.py. https://go.wisc.edu/4e0q68.
[6] MIUI App Store. https://app.mi.com.
[7] Product: Store intelligence. https://sensortower.com/solutions/store-intelligence.

[8] SpaCy-Langdetect. https://spacy.io/universe/project/spacy-langdetect.
[9] Tencent My App Store. https://android.myapp.com.
[10] ISDi: IPV Spyware Discovery Tool. https://github.com/stopipv/isdi, 2018.
[11] Google-play-scraper. https://github.com/facundoolano/, 2021.
[12] Play Console Help: Translate and localize your app. https://support.google.com/googleplay/android-developer/answer/9844778?#zippy=%2Cview-list-of-available-languages, 2021.
[13] What are the top 200 most spoken languages? https://www.ethnologue.com/guides/ethnologue200, 2021.
[14] Y. Aafer, W. Du, and H. Yin. Droidapiminer: Mining api-level features for robust malware detection in android. In *International conference on security and privacy in communication systems*, pages 86–103. Springer, 2013.
[15] N. Arif, M. Al-Jefri, I. H. Bizzi, G. B. Perano, M. Goldman, I. Haq, K. L. Chua, M. Mengozzi, M. Neunez, H. Smith, et al. Fake news or weak science? visibility and characterization of antivaccine webpages returned by google in different languages and countries. *Frontiers in immunology*, 9:1215, 2018.
[16] Avast. 51% increase in the use of online spying and stalking apps during lockdown. 2020.
[17] A. Ballatore, M. Graham, and S. Sen. Digital hegemonies: the localness of search engine results. *Annals of the American Association of Geographers*, 107(5):1194–1215, 2017.
[18] L.-P. Beland, A. Brodeur, J. Haddad, and D. Mikola. Covid-19, family stress and domestic violence: Remote work, isolation and bargaining power. 2020.
[19] R. Chatterjee, P. Doerfler, H. Orgad, S. Havron, J. Palmer, D. Freed, K. Levy, N. Dell, D. McCoy, and T. Ristenpart. The spyware used in intimate partner violence. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 441–458. IEEE, 2018.
[20] Cryptax. Droidlysis property extractor for android apps. https://github.com/cryptax/droidlysis, 2022.
[21] David Curry. Android statistics (2022), January 2022.
[22] S. R. Department. Number of apps available in leading app stores as of 1st quarter 2021. https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/, September 2021. Last accessed: Oct, 2021.
[23] A. Desnos. Androguard. https://github.com/androguard/androguard, 2020.
[24] J. P. Dimond, C. Fiesler, and A. S. Bruckman. Domestic violence and information communication technologies. *Interacting with Computers*, 23(5):413–421, 2011.
[25] D. Freed, J. Palmer, D. Minchala, K. Levy, T. Ristenpart, and N. Dell. Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders. *PACM: Human-Computer Interaction: Computer-Supported Cooperative Work and Social Computing (CSCW)*, Vol. 1(No. 2):Article 46, 2017.
[26] D. Freed, J. Palmer, D. Minchala, K. Levy, T. Ristenpart, and N. Dell. "A Stalker's Paradise": How intimate partner abusers exploit technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.
[27] S. GlobalStats. Mobile operating system market share worldwide.
[28] Google. Developer program policy (effective may 11, 2022, unless otherwise stated). urlhttps://support.google.com/googleplay/android-developer/answer/11987217.
[29] Google. Developer Program Policy: September 16, 2020 announcement - Play Console Help. https://support.google.com/googleplay/android-developer/answer/10065487.
[30] Google. Method: query.suggest - cloud search api. urlhttps://developers.google.com/cloud-search/docs/reference/rest/v1/query/suggest.
[31] Google. Translate and localize your app - Play Console Help. https://support.google.com/googleplay/android-developer/answer/9844778.
[32] Google. Update to enabling dishonest behavior policy (august 2020). https://support.google.com/adspolicy/answer/9726908, addendum = "(accessed: 12.16.2020)",.
[33] Google. Updates to Google Play Policies - Play Console Help. https://support.google.com/googleplay/android-developer/answer/9934569.
[34] B. Guo, Y. Ouyang, T. Guo, L. Cao, and Z. Yu. Enhancing mobile app user understanding and marketing with heterogeneous crowdsourced data: A review. *IEEE Access*, 2019.
[35] D. Harkin, A. Molnar, and E. Vowles. The commodification of mobile phone surveillance: An analysis of the consumer spyware industry. *Crime, Media, Culture*, 16:33–60, 2020.
[36] S. Havron, D. Freed, R. Chatterjee, D. McCoy, N. Dell, and T. Ristenpart. Clinical computer security for victims of intimate partner violence. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 105–122, 2019.
[37] E. Hussein, P. Juneja, and T. Mitra. Measuring misinformation in video search platforms: An audit study on youtube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):048:1–048:27, May 2020.
[38] P. Juneja and T. Mitra. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, page 1–27. Association for Computing Machinery, May 2021.
[39] Kaspersky. The state of stalkerware in 2021.
[40] H. Li, X. Lu, X. Liu, T. Xie, K. Bian, F. X. Lin, Q. Mei, and F. Feng. Characterizing smartphone usage patterns from millions of android users. In *Proceedings of the 2015 Internet Measurement Conference*, pages 459–472, 2015.
[41] L. Lu, R. Perdisci, and W. Lee. Surf: detecting and measuring search poisoning. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 467–476, 2011.
[42] Malwarebytes. State of Malware. 2021.
[43] Matlink. Google play downloader via command line. https://github.com/matlink/gplaycli, 2020.
[44] T. Matthews, K. O'Leary, A. Turner, M. Sleeper, J. P. Woelfer, M. Shelton, C. Manthorne, E. F. Churchill, and S. Consolvo. Stories from survivors: Privacy & security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2189–2201. ACM, 2017.
[45] B. McSweeney. Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. *Human relations*, 55(1):89–118, 2002.

[46] D. Metaxa, J. S. Park, J. A. Landay, and J. Hancock. Search media and elections: A longitudinal investigation of political search results. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):129:1–129:17, Nov 2019.

[47] A. Mirian, J. DeBlasio, S. Savage, G. M. Voelker, and K. Thomas. Hack for hire: Exploring the emerging market for account hijacking. In *The World Wide Web Conference*, pages 1279–1289. ACM, 2019.

[48] A. Molnar and D. Harkin. The consumer spyware industry: An australian-based analysis of the threats of consumer spyware. Aug 2019.

[49] W. H. Organization. Violence Against Women. https://www.who.int/news-room/fact-sheets/detail/violence-against-women.

[50] E. Peltonen, E. Lagerspetz, J. Hamberg, A. Mehrotra, M. Musolesi, P. Nurmi, and S. Tarkoma. The hidden image of mobile apps: geographic, demographic, and cultural factors in mobile usage. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 10. ACM, 2018.

[51] D. L. Poston and J. H. Wong. The chinese diaspora: The current distribution of the overseas chinese population. *Chinese Journal of Sociology*, 2(3):348–373, Jul 2016.

[52] I. D. Raji and J. Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, page 429–435. Association for Computing Machinery, Jan 2019.

[53] K. A. Roundy, P. B. Mendelberg, N. Dell, D. McCoy, D. Nissani, T. Ristenpart, and A. Tamersoy. The many kinds of creepware used for interpersonal attacks. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 626–643. IEEE, 2020.

[54] L. Sardinha, M. Maheu-Giroux, H. Stöckl, S. R. Meyer, and C. García-Moreno. Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018. *The Lancet*, 2022.

[55] A. Scroxton. Use of spyware apps linked to domestic abuse soars in lockdown. https://www.computerweekly.com/news/252485842/Use-of-spyware-apps-linked-to-domestic-abuse-soars-in-lockdown, 2021.

[56] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti. Predicting user traits from a snapshot of apps installed on a smartphone. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):1–8, 2014.

[57] P. Shijo and A. Salim. Integrated static and dynamic analysis for malware detection. *Procedia Computer Science*, 46:804–811, 2015.

[58] A. Singh, A. Walenstein, and A. Lakhotia. Tracking concept drift in malware families. In *Proceedings of the 5th ACM workshop on Security and artificial intelligence*, pages 81–92, 2012.

[59] D. Slotta. Market share of mobile operating systems in china from january 2013 to march 2021. https://www.statista.com/statistics/262176/market-share-held-by-mobile-operating-systems-in-china/, April 2021. Last accessed: Oct, 2021.

[60] S. G. Smith, K. C. Basile, and M.-j. Kresnow. *The National Intimate Partner and Sexual Violence Survey: 2016/2017 Report on Stalking*. Jan 2022.

[61] C. Southworth, S. Dawson, C. Fraser, and S. Tucker. A high-tech twist on abuse: Technology, intimate partner stalking, and advocacy. *Violence Against Women*, 2005.

[62] C. Southworth, J. Finn, S. Dawson, C. Fraser, and S. Tucker. Intimate partner violence, technology, and stalking. *Violence against women*, 13(8):842–856, 2007.

[63] B. R. Taira, V. Kreger, A. Orue, and L. C. Diamond. A pragmatic assessment of google translate for emergency department instructions. *Journal of General Internal Medicine*, 36(11):3361–3365, 2021.

[64] T. N. Y. Times. Thermostats, locks, and lights: Digital tools of domestic abuse. https://www.nytimes.com/2018/06/23/technology/smart-home-devices-domestic-abuse.html.

[65] N. N. to End Domestic Violence. Inside the 'stalkerware' surveillance market, where ordinary people tap each other's phones. https://www.vice.com/en/article/53vm7n/inside-stalkerware-surveillance-market-flexispy-retina-x.

[66] N. N. to End Domestic Violence. Tech abuse in the pandemic & beyond. https://static1.squarespace.com/static/51dc541ce4b03ebab8c5c88c/t/61674c082419497a370af990/1634159630368/2021_T2E+Needs+Assessment+Report.pdf.

[67] M. Valites. The commoditization of mobile espionage software. https://blog.talosintelligence.com/2019/10/the-commoditization-of-mobile-espionage.html.

[68] G. Varelas, E. Voutsakis, P. Raftopoulou, E. G. Petrakis, and E. E. Milios. Semantic similarity methods in wordnet and their application to information retrieval on the web. In *Proceedings of the 7th annual ACM international workshop on Web information and data management*, pages 10–16, 2005.

[69] Wikipedia. List of languages by total number of speakers — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=List%20of%20languages%20by%20total%20number%20of%20speakers&oldid=1027421582, 2021. [Online; accessed 09-June-2021].

[70] U. Women. Facts and figures: Ending violence against women.

[71] D. Woodlock. The abuse of technology in domestic violence and stalking. *Violence against women*, 23(5):584–602, 2017.

[72] D.-J. Wu, C.-H. Mao, T.-E. Wei, H.-M. Lee, and K.-P. Wu. Droidmat: Android malware detection through manifest and api calls tracing. In *2012 Seventh Asia Joint Conference on Information Security*, pages 62–69, 2012.

[73] W. Z. Zarni Aung. Permission-based android malware detection. *International Journal of Scientific & Technology Research*, 2(3):228–234, 2013.

A Multilingual Seed Terms

We translated seed terms used by Chatterjee et al. [19] with the help from native speakers who are also proficient in English. The full list of seed terms is available in GitHub [5]. We provide examples in Fig. 8.

B Capabilities of Dual-Use Apps

Dual-use apps can have different capabilities of collecting information and sharing it with the abuser. They also provide different levels of control over the victim's device. We categorize these capabilities based on what dual-use apps we found in our dataset. The five categories are shown in Fig. 6.

Locate. Apps in this category allow the abuser to track the location of the victim. Four types of location tracking dual-use apps belong to this category:

(1) *Personal Tracking*: These apps are designed to allow users to track their own devices. Usually, different devices are linked to the same account, and users can track the location of their devices by accessing their account on the web or the app itself, which should be installed on a second device. No consent is required to track devices since the app assumes all devices belong to the same user. Disabling tracking, which can be done remotely or on the device itself, usually requires inputting a user password. An example is "Google Find My Device."

(2) *Mutual Tracking*: Apps belonging to this group are different from personal tracking apps in that you cannot track the second device without getting consent from that device. Moreover, the second device can track the location of the first device as well since it is mutual tracking. Each party has the option to disable tracking and usually does not require inputting any passwords. Moreover, disabling tracking can be done only by accessing the device itself. Widespread apps include family and friend tracker apps.

(3) *Subordinate Tracking*: These apps provide one-way tracking of remote devices. They usually come with two types of users: administrators (admins) and members. Admins can track all members linked to their accounts, but members cannot track the admin. In some cases, members are allowed to track each other. The primary purpose of these apps is to allow a person or group of people to track another group. Such apps imply non-consensual but socially approved tracking, such as tracking children or employees.

(4) *GPS Tracker Apps*: These tracking apps rely on GPS trackers that can be installed on vehicles, bikes, etc. Users must purchase a GPS tracker to be able to use such apps. Popular examples are vehicle and pet trackers.

Record. These apps allow the user to record audio and video automatically. We identified five different types of recording apps:

(1) *Voice Call Recorder*: Records voice calls of the device automatically, including incoming and outgoing calls. Such apps only record audio, so they may not work with video calls or they might just record the audio of the call, depending on the app.

(2) *Video Call Recorder*: Similar to voice call recorders, these apps record video calls automatically. These recorders usually work on third-party video chat apps, such as Skype, and not just built-in video calls.

(3) *Screen Recorder*: Allows the user to schedule times to record the screen, with audio, in the background. Recording can be scheduled to be daily, weekly, and so on.

(4) *Voice Recorder*: Uses the device's microphone to record the surrounding audio in the background. The app starts recording automatically at times specified by the user.

(5) *Video Recorder*: These recorders use at least one of the cameras (front, back or both) to record video automatically in the background. Recording times are specified by the user.

Share. Apps that share data, other than location, with a remote device. Seven types of apps were identified and observed:

(1) *Auto Backup/Sync*: Automatically backs up some or all device data and syncs it with another device or a cloud account. These data could be gallery data only (photos and videos), recorded calls, contacts, or even a full backup of the device. Data is usually uploaded to Google Drive, Dropbox, or similar cloud services.

(2) *SMS/Logs Forwarder*: Automatically forwards SMS and call logs from the device to a given phone number, an email or even to a server.

(3) *SIM Change Detector*: Sends a text message to a pre-specified phone number whenever the SIM card of the target device, on which the app is installed, is removed.

(4) *Last Seen Tracker*: Tracks the time spent online by users on WhatsApp, Telegram, and sometimes other texting apps using only the user's phone number linked to these apps, without the need to install an app on the target's device.

Language	Agent terms	Language	Agent terms
Arabic (ar)	hbib , habibati , zujati , zuji , shrik , shrikati	Indonesian / Malay (ms)	pacar, kekasih, istri, suami, pasangan, cewek, cowok, pelakor
Bengali (bn)	premik, premika, bondhu, bandhabi, stri, sbami, pati, patni, bou	Portuguese (pt)	namorado, namorada, esposa, marido, cônjuge, esposo, parceiro, parceira, mulher
Chinese (zh)	nan peNjou, ny peNjou, lao puo, laokoN, peiou	Russian (ru)	paren, podruga, zhena, muzh, suprug, muzh
English (en)	boyfriend, girlfriend, wife, husband, spouse	Spanish (es)	novio, novia, esposa, esposo, hombre, mujer
French (fr)	petit ami, petite amie, copain, copine, mari, femme, épouse, époux, conjoint	Thai (th)	faen, phrrya, sami, khusmrs
German (de)	freund, freundin, ehefrau, mann, ehepartnerin, ehepartner	Turkish (tr)	erkek arkadaş, kız arkadaş, eş, koca, eş, koca
Hindi (hi)	premi, premika, pətni, pəti, pəti, pəti	Vietnamese (vi)	ban trai, ban gái, vo, chong, nguoi đàn ong, nguoi phu nu
Japanese (ja)	kareshi, kanojo, tsuma, otto, haigūsha, otto		

Language	Example queries	Language	Example queries
Arabic (ar)	altajasus ealaa al{agent}, taeaqub al{agent}, tatabae hatif al{agent}, altanasut ealaa almukalamat, taeaqub almawqie	Indonesian / Malay (ms)	melacak {agent}, memata-matai {agent}, membaca email pribadi {agent}, mencari tahu perselingkungan {agent}, tidak setia {agent}
Bengali (bn)	{agent}-er protarona dhorar app, Thog {agent}-er abosthan track kora, oporer call sonar app, amar {agent}-er opor nojordari kora, amar agent er opor spy kroa	Portuguese (pt)	descobrir {agent} traindo, espionar meu {agent}, infiel {agent}, rastrear {agent}, rastrear o telefone do {agent}
Chinese (zh)	bù zhōngshí de {agent}, zhuā zhù bù zhōng de {agent}, gēnzōng {agent}, gēnzōng {agent} de diànhuà, jiānshì wǒ de {agent}	Russian (ru)	nevernyy {agent}, poymatńevernykh {agent}, poymat́{agent} obman, otslezhivat́{agent}, treker {agent}
English (en)	catch {agent} cheating, cheating {agent}, spy my {agent}, track {agent}, apps for spying on my {agent}	Spanish (es)	apps para espiar mi {agent}, atrapar {agent} infiel, descubrir a {agent} infiel, leer el correo electrónico de mi {agent}, saber si {agent} te está engañando
French (fr)	attraper {agent} infidèle, des applications pour espionner mon {agent}, espionner mon {agent}, suivre téléphone de {agent}	Thai (th)	nxkcı {agent}, pêuua dtìt dtaam {agent}, pêuua dtìt dtaam toh-rá-sàp kŏng {agent}, pêuua àan ee men jàak toh-rá-sàp kŏng {agent}
German (de)	apps zum ausspionieren meines {agent}, {agent} betrügen, fang betrügerischen {agent}, track {agent}	Turkish (tr)	{agent} betrügen, {agent} casusluk için başvurular, {agent} hile yakalamak, sadakatsiz {agent}, {agent} üzerinde casusluk yapmak
Hindi (hi)	sunne ke liye app, {agent} ko pakadne ke liye app, paribar tracker, bebafa {agent} ko pakadne ke liye app	Vietnamese (vi)	khon chung thwi {agent}, de băt xoŋ cuŋ then {agent}, de băt {agent} zan lan, de thew zoj {agent}, de thew zoj dien thwaj kwa {agent}
Japanese (ja)	fuseijitsuna {agent}, fuseijitsuna {agent} o tsukamaeru, {agent} o tsuiseki suru, {agent} torakkā, {agent} no denwa kara mēru o yomu		

Fig. 8. Example of (transliterated) seed terms used in 15 languages. Top table shows the different `{agent}` terms that are replaced in the {agent} in the bottom table. The full list of terms was posted in GitHub [5].

(5) *WhatsApp Cloner*: Although WhatsApp does not allow two devices to be logged into the same account, these apps "clone" a WhatsApp account on another device, allowing two devices to access the same WhatsApp account.

(6) *Keylogger*: Apps that record or log all keystrokes on the device on which they are installed.

(7) *Device Reports*: Sends reports about the target device. These reports could be as simple as battery level, or as comprehensive as a report including what apps were used and for how long, newly installed apps, deleted apps, surfed web pages, and more. The reports are usually accessible from an admin account in the same app (or a companion app by the same developer) or from a web page.

Lang	Apps Found	App Language (%) EN	Query	Other	Clf. Acc. (%) Recall	Precision	# dual-use Lab.	Estim.	Capabilities (%) L	C	A	S	R
Arabic	2,665	65	29	6	**100**	**8**	**4**	23	-	-	**100**	-	-
Bengali	1,589	93	**3**	4	**69**	28	15	88	33	20	27	60	-
Chinese	2,602	74	18	5	85	40	21	280	**81**	14	-	24	14
English	**8,146**	**99**	-	**1**	**100**	**58**	**29**	**911**	41	**3**	7	28	**31**
French	3,643	47	50	3	90	48	25	569	36	30	16	44	20
German	4,525	62	36	2	85	48	25	511	72	12	4	48	**4**
Hindi	1,356	80	13	7	**100**	10	8	**12**	-	-	88	-	13
Japanese	1,648	77	17	7	**100**	10	8	24	**25**	-	63	25	-
Malay	3,052	60	37	3	**100**	50	25	344	40	24	16	**72**	-
Portuguese	3,531	46	48	6	88	52	27	489	48	**33**	11	44	26
Russian	2,306	55	37	**8**	**100**	42	21	177	67	5	24	**14**	-
Spanish	5,301	**32**	**68**	**1**	**100**	**58**	**29**	575	55	17	**3**	41	17
Thai	**1,336**	80	14	7	74	48	26	170	46	12	27	42	8
Turkish	2,258	53	41	6	79	44	24	334	58	8	13	38	13
Vietnamese	3,131	80	14	6	**100**	48	24	225	42	-	46	17	**4**

Fig. 9. Summary of the data collected in Feb 2022. The figure is structured similarly to Fig. 4.

Access. Some applications allow users to gain live access to other devices. However, this access is usually partial and does not grant control over the device. Three types of apps belong to this category:

(1) *Camera Access*: Apps that use and access the camera of a remote device to stream a live video. These accessed devices could be IP cameras, baby cameras, or even a smart device's camera.

(2) *Microphone Access*: Allows access to the microphone of a remote device and broadcasts its surrounding audio.

(3) *Screen Mirroring*: Projects and shares the screen of a remote device with the user's phone, without necessarily granting control over that remote device.

Control. These apps are powerful, as they allow the user to remotely control other phones and smart devices. Only two types of apps belong to this category:

(1) *Parental Control*: These apps target parents as primary users and provide them with multiple features to control their children's devices remotely. Some features include but are not limited to: limit total screen time, limit the usage of apps, block and delete apps, block websites, restrict app installation, and remotely turn off the device. These apps usually come in two versions, one to be installed on the child's device, and the second to be installed on the parent's device. In many cases, such apps hide themselves from the child's device.

(2) *Phone/Tablet/PC Control*: Unlike parental control apps, device control apps allow users to fully control other devices. Users can access another device remotely and access it as if the device were in their hands. A few apps in this category do not provide complete control over the other device, but only a few features, such as turning the device on and off or adjusting volume.

We coded a total of 21 types of apps belonging to five categories, which were identified based on app capabilities. We note that apps can belong to multiple categories. For example, many parental control apps can find the location of the child and collect reports from the child's device, such as app usage and browsing history. Such an app belongs to Locate, Share, and Control categories. Moreover, an app can belong to several types within the same category. For example, some apps can access the camera and microphone at the same time, or record both voice and video calls automatically.

C Current Status of Google Play

We present the results of our February 2022 crawl in Fig. 9. While qualitatively the results are very similar to our crawl in 2020 (Fig. 4) for all languages except for Arabic, Hindi, and Japanese, we found significantly fewer dual-use apps in these languages. We are not sure what caused this change.