



STATE OF DATA PRIVACY OF MOBILE APPS & WEBSITES FROM INDIA

THE ARRKA STUDY 2018





INDEX

Foreword	04
Executive Summary	05
Setting the Context	07
Methodology and Approach	08
What Personal Data is being accessed?	10
Who is your Personal Data being shared with?	14
Is your Personal Data secure?	17
Are Organizations being transparent with you?	18
Are Children's Apps safe?	20
Conclusion	21
Authors	22



Foreword

Today, India has the fastest growing digital population and is the second largest 'digital nation' in the world. A majority of India's digital population accesses the net via their mobile phones. The quantum of data they generate coupled with the size of the Indian economy and the demographic profiles involved makes this one of the most 'prized' digital populations in the world.

While this is certainly an impressive statistic and offers tremendous opportunities, it comes with its associated risks. One of these key risks is that of Data Protection & Privacy. The absence of a dedicated Data Protection & Privacy law exacerbates this risk. The combination of the large digital population with no real curbs on who uses this population's data and how makes digital Indians particularly vulnerable.

As part of the work we do at Arrka in the Data Privacy domain, we set up the Arrka Privacy Lab. The lab enables organizations to test their various digital properties for Data Privacy. When we set up this lab in mid-2017, we realised that there were no real data points available to help Indian (and international) organizations benchmark their digital properties for privacy. So we decided to make a start here ourselves by conducting a privacy study of Indian Android Apps. The results we published in Dec 2017 via the 'State of Privacy of Indian Android Apps' report.

The world has meanwhile galloped ahead on the Data Privacy front in this last one year since this report was published. So this year we decided to test not only Android Apps but their IOS versions and associated websites as well. Further, to benchmark this against global properties, we studied a small sample of Global Android Apps too. The result is this report - the Second Edition of the 'State of Privacy of Indian Mobile Apps and Website - 2018'.

We look forward to this report equipping various stakeholders with some valuable data points about the reality on the ground of the State of Privacy in India today.

Coming to Websites, basically the concept of permissions is very different from that of Mobile Apps and they are not to be equated. Permissions on Websites being just four in number with very limited data being accessible, tracking mechanisms are the preferred mode of capturing personal data. Of these, our study is restricted to testing for only two types of tracking mechanisms (that are possible via external-only tests). The study found that over 95% websites deploy at least 1 tracking mechanism while permissions were limited to only 20% organizations using at least 1 permission.

To identify which 3rd parties Personal Data was being shared with, we conducted an analysis of the Network traffic flowing from the App/Website. Almost all Apps & Websites were found sharing data with external third parties. The average number of third parties an App or Website was sending out data to was 5.5. More than 90% of the 3rd parties with whom personal data was being shared fell into the following four key categories: Advertising, Analytics, Development-Supporting tools and Authentication Entities. Advertising was the top category that Websites sent data to with 30% of the overall traffic headed there. Analytics took center-stage for iOS Apps with 47% of the traffic going out to this category. Development-Supporting tools was the primary category for Android Apps(48%). The highest percentage of traffic was found to be headed to a Google property (30-58% instances across channels) with Facebook coming a distant second (9-14%). Personal data crossed geographical boundaries of the country (cross border) in almost all Apps/Websites. USA is where the primary chunk of the traffic was found headed to – which is likely owing to the fact that most Advertising and Analytics companies' data centres are located there. Data center destinations like Ireland and Singapore were a distant 2nd and 3rd.

The results of testing for security parameters was heartening with the organizations studied showing significant maturity in this area.

Transparency and Disclosure of what really an App or Website does about one's personal data has often been a bone of contention. Hence, we analysed the privacy notices of all the properties we studied. We studied the Notices along three lines: How easily was the Notice available to the user, how complete

were the contents of the Notice and how easy to read and understand was the content. On availability, while the Notice is readily available at the first user touchpoint – like the App playstore or Website landing page, it is not easily available post that during actual usage. In terms of completeness of content, we observed that organizations revealed less than half the information they should ideally be disclosing. Interestingly, critical categories like banks fared poorly in this regard. Not surprisingly, most Privacy Notices studied qualified as "Difficult to Read" with some being "Very Confusing" as per the Fleisch Scale, which is an industry standard used for readability of content.

The study took a special look at Android Apps targeting children, given how they are a vulnerable group. The findings were rather alarming. 71% children's Apps were observed accessing Location, Phone Details and storage. More than half the permissions accessed were not required for the App to function. Majority of the Apps did not take consent and, wherever consent was being taken, no verification was done to ensure that the individual was an adult. In-App purchase options were available and In-App ads were present in majority of the Apps.

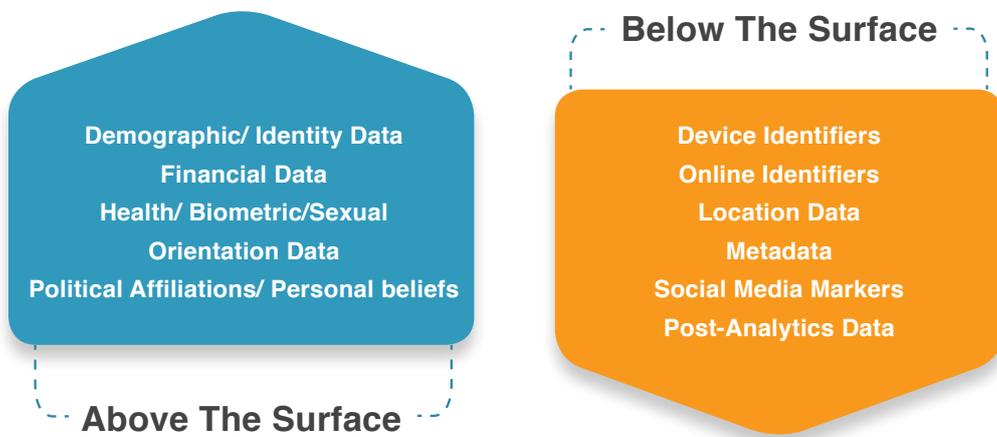
Finally, benchmarking the above findings against their global counterparts (only for Android Apps), Indian Android Apps explicitly ask for 45% more permissions than Global Android Apps. This difference was even more striking in categories like Travel Booking, Shopping & Mobile Wallets where Indian Apps took 60-80% more permissions. Access to sms, microphone, phone and contacts were the permissions that were accessed by significantly higher number of Indian Apps.

Setting the Context

The Arrka Study is based on 2 key concepts of Privacy: A. Personal Data and B. Privacy Principles

A. Personal Data: Personal Data can be categorized into ‘Above-the-Surface’ data (visible to an individual) and ‘Below-the-Surface’ data (not visible to an individual). ‘Below-the-Surface’ data that gets collected by Mobile Apps, Websites and Devices is usually not noticed by people. Our study focuses largely on “Below the Surface” data

Why is ‘Below-the-Surface’ data important? Various elements of this data are run through sophisticated algorithms, combined with data from ‘Above-the-Surface’ and 3rd parties to build individuals’ in-depth profiles, understand preferences and predict behaviour. This info is then used to not only dish out ads, provide ‘recommendations’ on what individuals should read, view, shop for, etc but also influence & shape their views and opinions.



B. Privacy Principles: Privacy principles represent the core of privacy protection and they provide a holistic lens to analyse Personal Data Privacy. They also form the underlying components around which data protection or privacy protection laws across the world are based. In our study we have selected 4 principles: Collection Limitation, 3rd Party Disclosure, Notice and Security which can be tested externally without involving the target organization.

Principles	User Concerns
Notice	Are you clearly and unambiguously telling me all that you do with my Personal Data?
Consent	Is my consent being taken on all that is being done with my Personal Data?
Access and Correction	Can I access my Personal Data & make corrections as and when required
Collection Limitation	Is the Personal Data being collected more than what is required?
Usage and Purpose	Why are you collecting my Personal Data? What are you going to use it for? Are you going to use it to track me & build my profile?
Disclosure/3rd Party Transfer	Are you sharing my Personal Data with a 3rd party? Is my Personal Data being sent outside geographical boundaries?
Security	Is my Personal Data adequately protected and safeguarded?

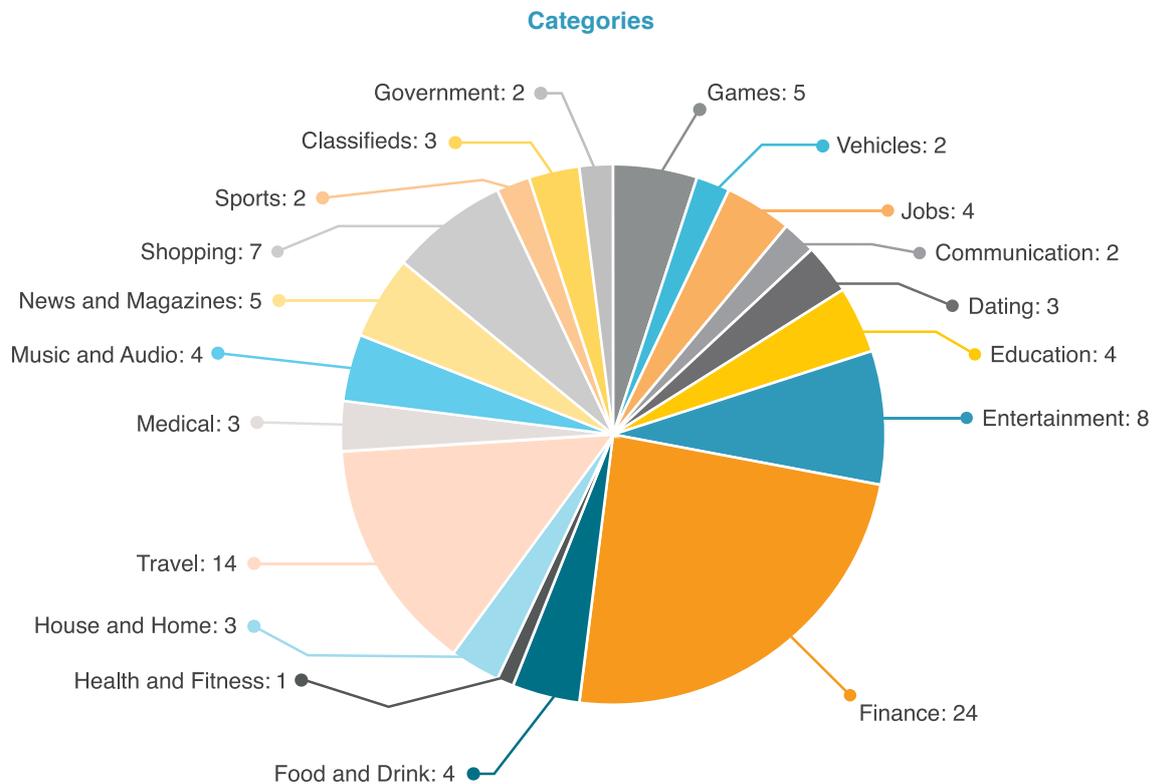
Methodology and Approach

In this section, we detail the “What” and the “How” of the study. A. What - Apps we covered in the study and the Sampling methodology. B. How - Privacy Principles we selected and how we tested each principle.

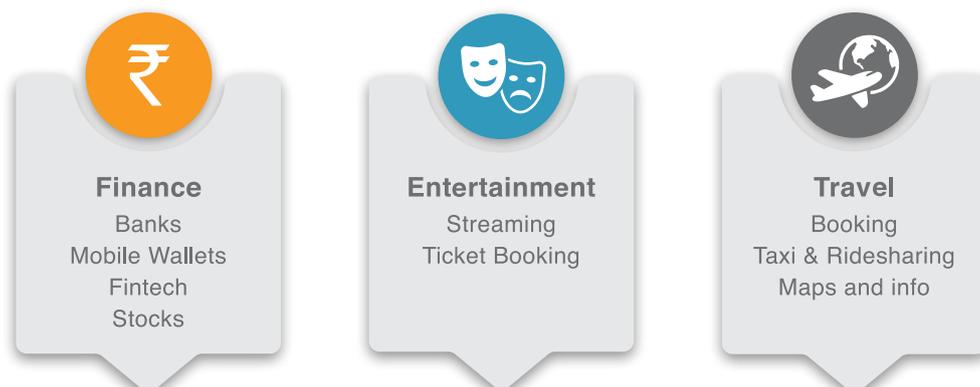
A. Sampling Methodology

The study covered 100 organizations from India and 3 digital properties of each organization: An Android App, its iOS counterpart and the associated Website. The organizational categories have been defined based on Google Playstore definitions and industry nomenclature. We chose organizations across categories to ensure fair representation.

Within each category, organizations were chosen based on popularity. For eg., more that 90% of the Android Apps chosen have had greater than one million downloads on Android and 58% of the Websites chosen were among the Top 500 Indian websites based on Alexa ratings.



Sub-Categories





What Personal Data is being accessed?

Mobile Apps and Websites obtain Personal Data of users from their devices via Permissions and Tracking Mechanisms. The types of Permissions/Tracking Mechanisms used are different for Mobile Apps and Websites. Further there is a difference between the Android and iOS Permission groupings. The study does an analysis of all of these.

Section A: Android Apps

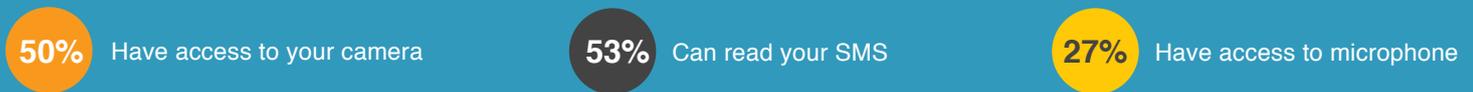
A.1: Top Dangerous Permissions Accessed

The following table highlights the most deployed dangerous permissions by Apps. Read and Write External Storage tops the most accessed permissions list. The possible reason could be because these permissions are needed by Apps to store intermediate results

Top 4 Dangerous Permissions



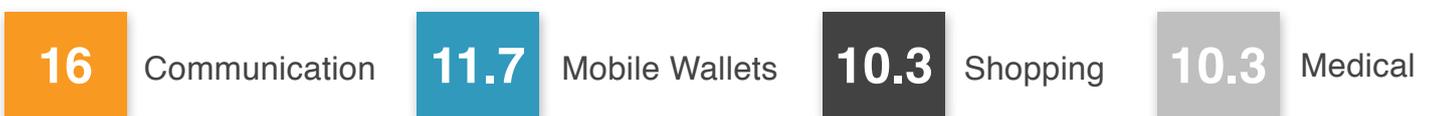
Other Key Dangerous Permissions



A.2: Category-wise Dangerous Permissions

Avg. no. of Dangerous Permissions per App - 8

Categories using the most no. of Dangerous Permissions per App



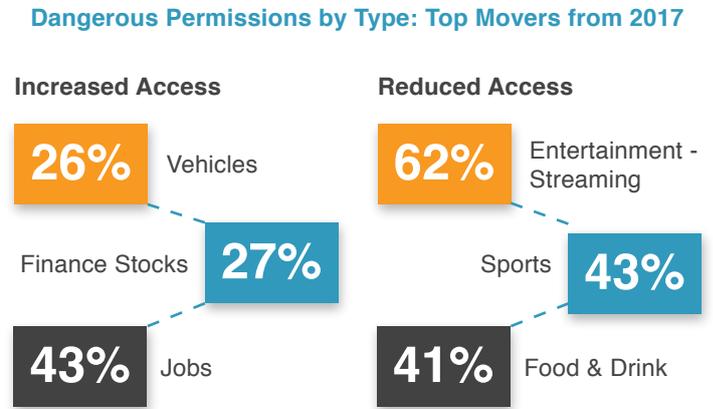
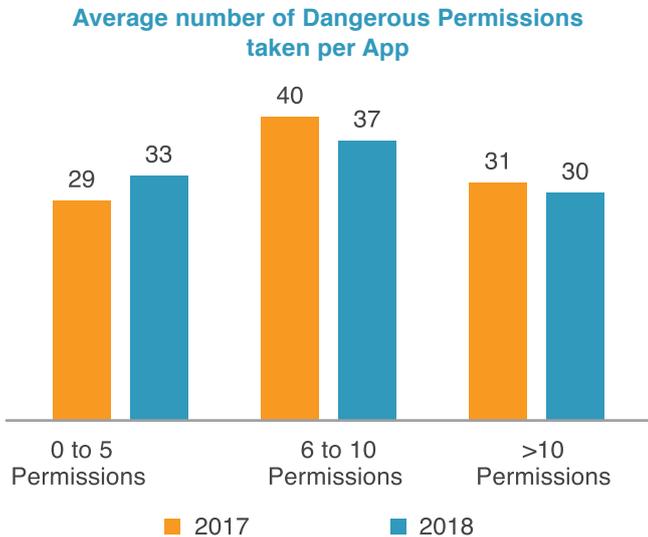
Categories using the least no. of Dangerous Permissions per App



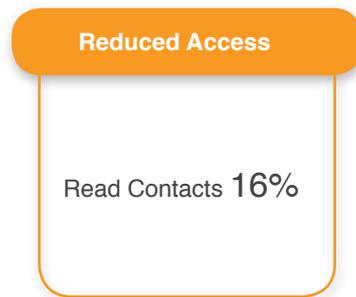
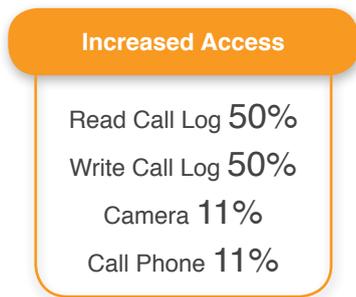
The study observed a significant variation (5X) between the category taking the highest number of Dangerous Permissions (Communication - 16) and the category taking the lowest number (Games-3). This could be possibly attributed to the features required by different categories.

A.3: Comparison with the 2017 Study

There has been no significant change in the average number of Dangerous Permissions taken per App from 2017. It stands at ~ 8. About 25% of the Apps showed more than a 10% variance (increase or decrease) in the number of permissions accessed. This indicates that a majority of the Apps don't make drastic changes to Permissions.



----- Dangerous Permissions by type: Top Movers from 2017 -----

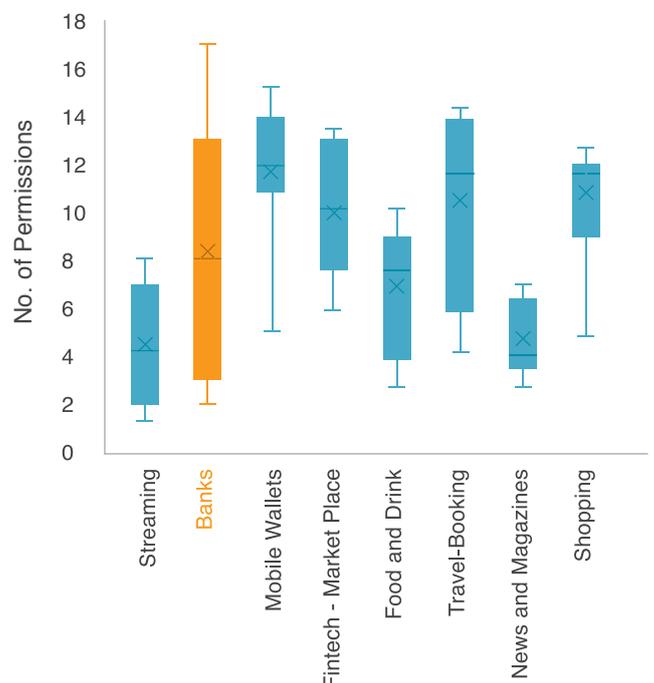


A.4: Intra Category Variation (Permissions)

- Though Inter-Category variation in Permissions taken by Apps could be justified as the features provided by different categories could vary widely, however intra-category variation becomes harder to explain.
- When we studied Dangerous Permissions in Android Apps and conducted an intra category variation analysis, we observed an average difference of 3X between Apps from the same category taking the maximum and minimum number of permissions.
- Categories with highest Intra Category variations were: Banks, Streaming, Food & Drink.

When we studied Dangerous Permissions in Android Apps and conducted an intra category variation analysis, we observed

Intra-Category Permission Analysis



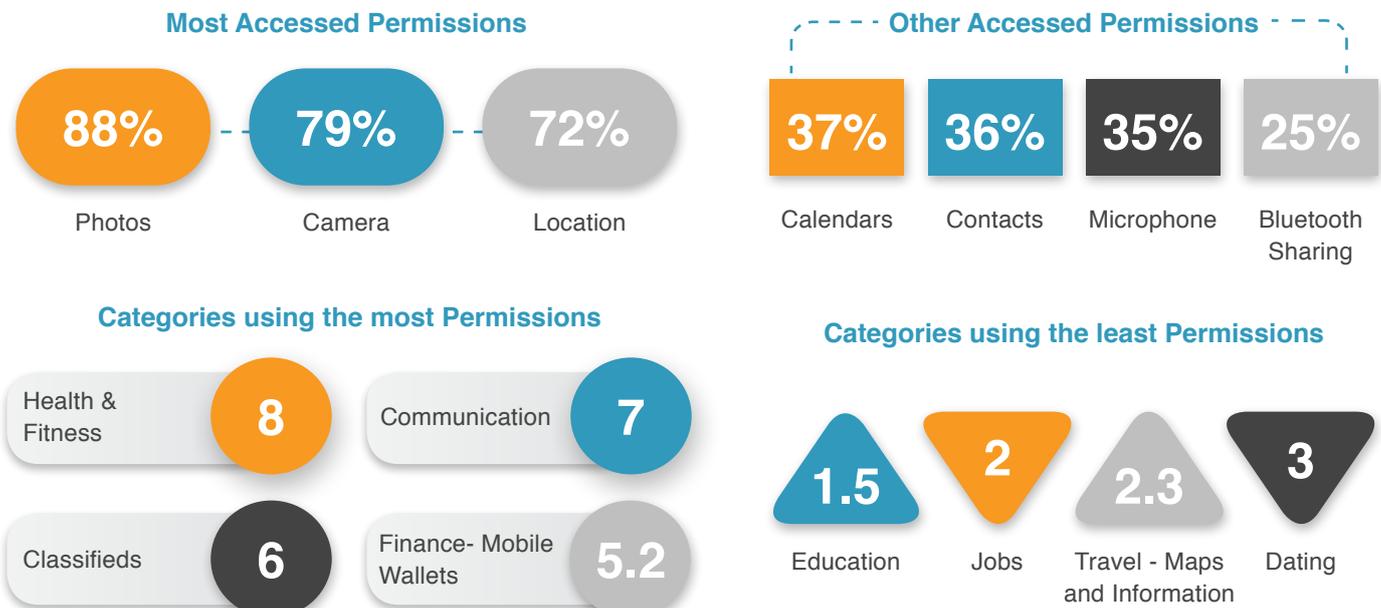
B. iOS - Permissions

The Permission structure of iOS is a little different from that of Android. There are 16 Permissions in all, some of which are common with Android (eg: Contacts, Camera) while some are different (Apple Music, TV Account). Moreover, certain permissions (i.e. Microphone and Camera) can be configured such that they can be accessed in one of two modes – ‘While Using the App’ or ‘Always’

On an average an iOS App accesses 3.9 (out of 16) Permissions

79% iOS Apps accessing Camera only track it when the App is in Use

29% iOS Apps “Always” track user location regardless of whether the App is in use or not.

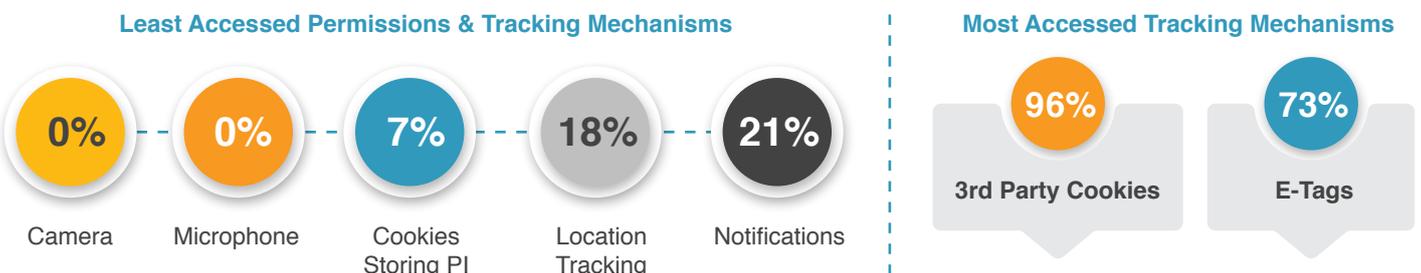


C. Websites – Permissions and Tracking Mechanisms

The Study analyzed 7 Permissions & Tracking Mechanisms deployed by Websites. Permissions have a far lesser impact in the world of Websites (as compared to Mobile Apps) with the number being very small (just 4 possible Permissions) and the data gathered via that being very limited. Websites rely far more on sophisticated Tracking Mechanisms than on Permissions to access a user’s data. This was reflected in the study – with more than 96% Websites seen using Tracking Mechanisms vs only 21% using Permissions. Among the Permissions used, Location Tracking was used only by 18% websites (as compared to 70% of Mobile Apps). However, it is important to note that any Website can track a user via her IP address anyways.

Of Tracking Mechanisms used, we tested only for Cookies and E-tags – the only ones possible to test externally.

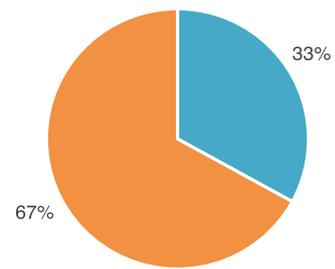
Cookies included 3rd Party Cookies too



D.1 Excess Permission Analysis: Are Apps taking excess Permissions?

- The Arrka Permission-To-Function (P2F) Correlation Index is an indicator of excessive Permissions accessed by an App without providing any corresponding functionality. i.e. A P2F score of 25% indicates that only 25% of the Permissions accessed by the App could be justified, the remaining are excessive Permissions.
- Our analysis concluded that 1/3rd of Permissions taken were excessive

Permissions in Android

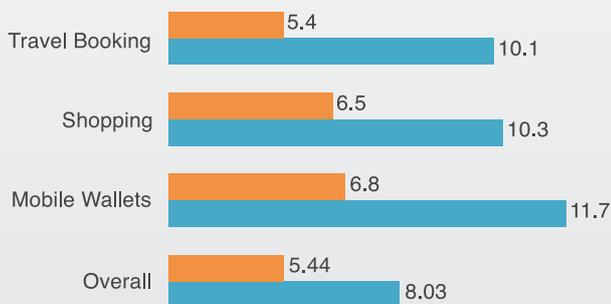


■ Excessive Permissions ■ Justified Permissions

D.2 Comparison with Global Apps

In the absence of relevant and available global data points, we tested 50 Global Android Apps and compared the permissions accessed with those by Indian Apps. No real surprises here though - Indian Apps explicitly ask for 45% more permissions than Global Apps. In categories like Travel Booking, Shopping and Mobile Wallets, the difference is particularly striking with Indian Apps taking 1.6X to 1.8X times more number permissions.

Avg. Permissions Per App



■ Global Apps ■ Indian Apps

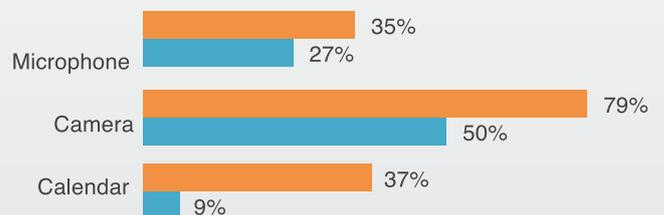
* This test was conducted on a sample of 50 popular Global Android Apps picked Apps across categories .

D.3 Cross-Device Comparison

Some interesting results emerged when we compared the results of Android and IOS Apps. While almost 60% of the Permissions were similarly accessed by Android and IOS versions of an App, specific Permissions displayed significant variance. Permissions like Microphone, Camera and Calendar were accessed 1.3-1.4 X more by iOS Apps as compared to their Android versions.

This indicates that either the same organization may be providing different features on Android and iOS versions or that excess Permissions were being taken on iOS. Another possible reason could be different development teams/vendors may be approaching Permissions differently, in the absence of an internal standard.

Cross Device Comparison of Permission Access



■ iOS ■ Android

% indicates % of Apps accessing permission



Who is your Personal Data being shared with?

The study analyzed the traffic flowing out of each App & Website to understand where data was headed out to.

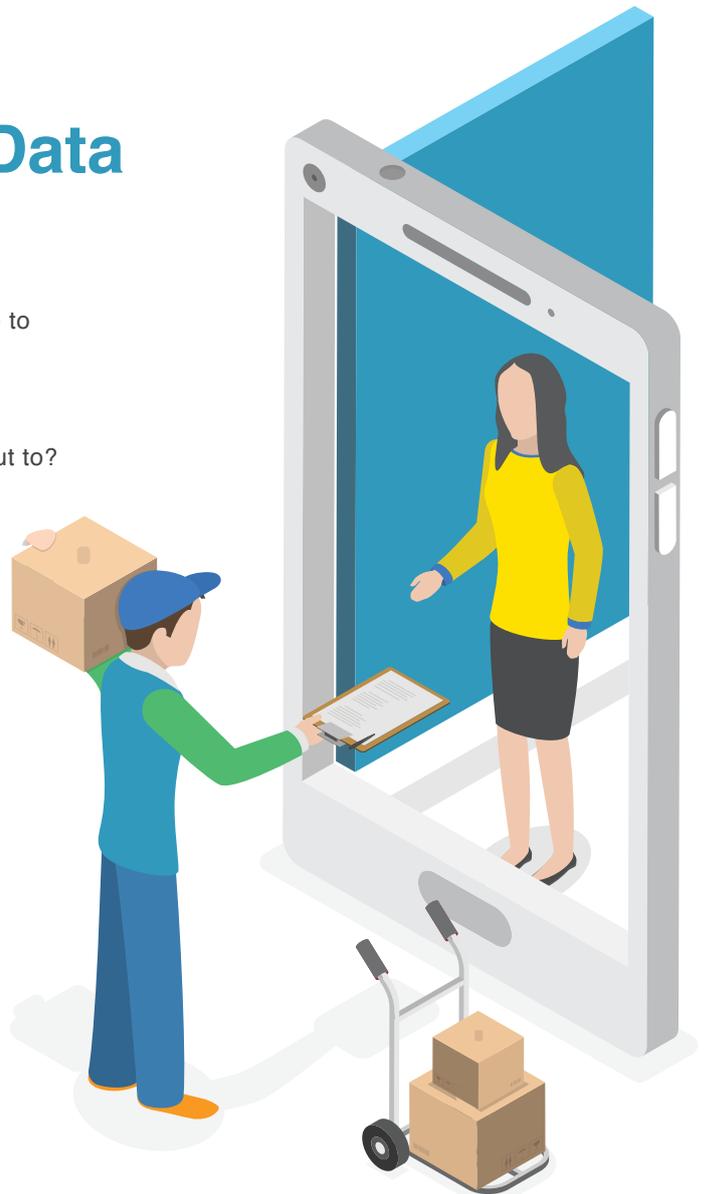
We looked for answers to the following questions:

- A. How many 3rd Parties was each App/Website sending data out to?
- B. Where were these 3rd Parties geographically located?
- C. Which functional categories did these 3rd Parties belong to?
- D. Which parent organizations did these 3rd Parties belong to?

A. 3rd Parties Data was being sent out to

On an average, organizations were found to share data with 5.5 3rd Parties per digital channel (Mobile Apps/Websites) with minor variations across channels

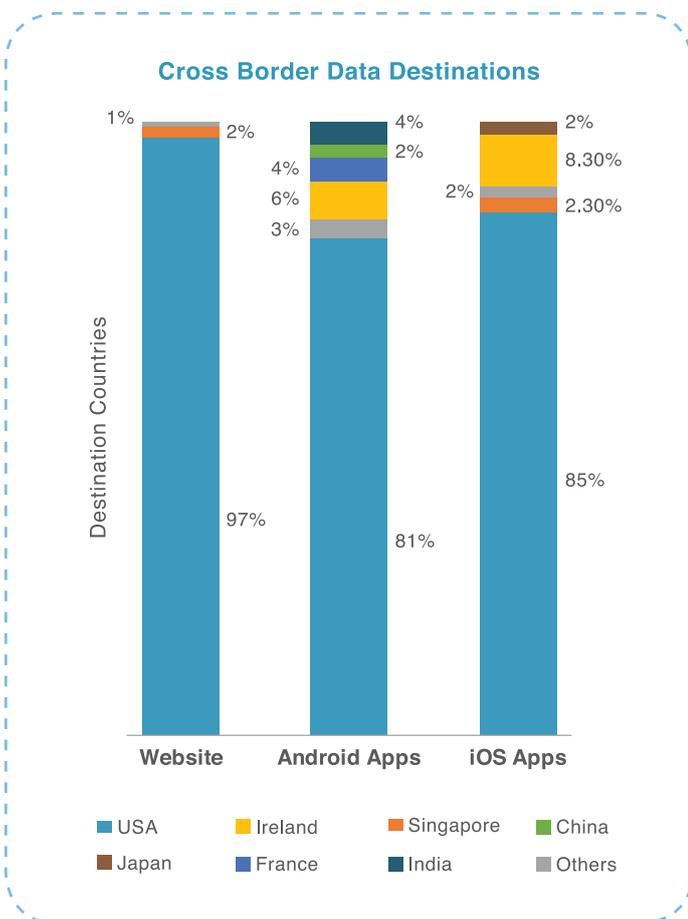
Across channels, Banks, Education and Communication shared data with the least number of 3rd parties. Across channels, Vehicles, Classifieds and Sports were the worst performers, sharing data with 7.5 3rd parties on an average.



			
% of Apps/Websites sending data to 3rd Parties	99%	94%	94%
Average No. of 3rd Parties Data is being shared with	5.6	6.2	4.5
Categories sending data to maximum number of 3rd parties'	Entertainment – Streaming - 9.6 Sports - 7.5 Games - 7.4	Classifieds - 11.3 Sports - 10 News and Magazines – 10	Vehicles - 12.5 Jobs - 12 Food and Drink - 5.8
Categories sending data to least number of 3rd Parties	Finance – Banks 3 Government 3 Finance - Stocks 3	Finance – Banks 1.6 Education 1.8 Communication 2	Communication 1 Education 2.5 Finance - Mobile Wallets 3

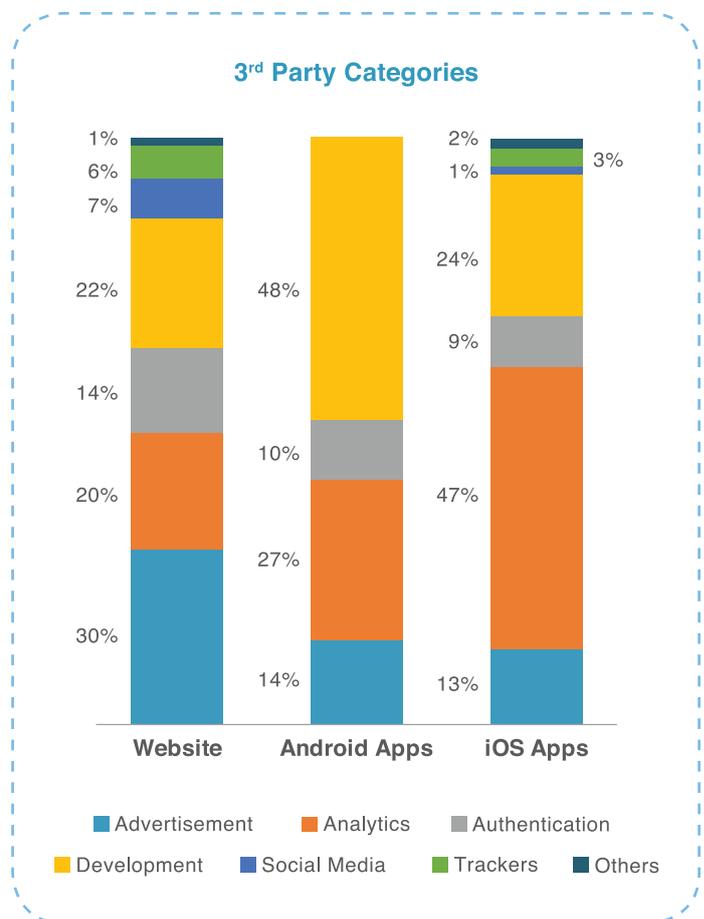
B. Cross Border Data Destinations

The study identified the first destination country data was heading out to 99% of organizations studied sent data across the borders. Not surprisingly, the US is the primary destination of all the data being transferred outside India with more than 81-97% of the traffic being directed there. This is probably owing to the fact that most of the 3rd Party Advertisers and Analytics companies are based out of the US. At a distant second came Ireland, Singapore and France. We also found two outlier countries which featured only in a single instance of data transfer: Russia and Tanzania.



C. 3rd Party Categories

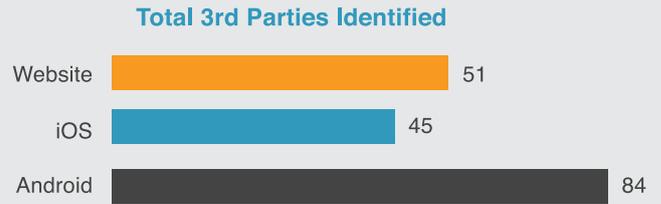
The top categories of 3rd Parties with whom data was being shared with were Advertising, Analytics and Development (used to add functionality to Apps), Authentication (where platforms like Google are used to authenticate users), Social Media and Trackers (usage statistics). However, there was high variation in 3rd Party Categories composition across Mobile Apps and Websites. While advertising was the highest category that data was being shared with by Websites (30%), Analytics was the highest Category in iOS (47%) and Development was the highest in Android (48%).



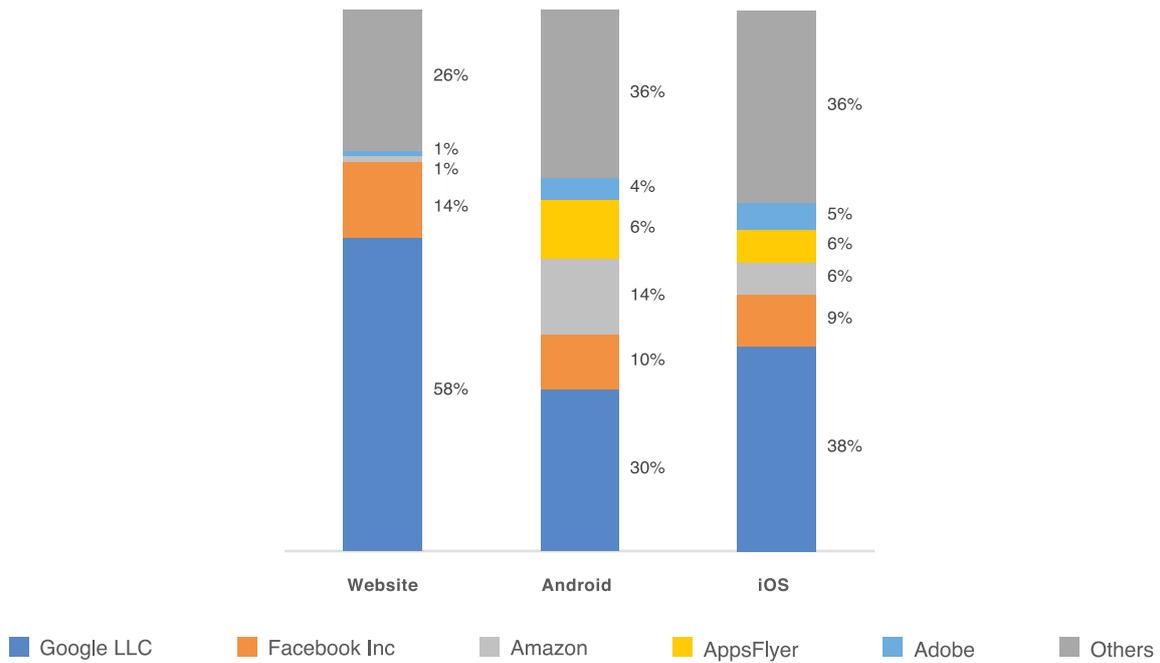
D. 3rd Party Organizations

We looked at which specific 3rd Parties data was being shared with and categorized them based on their parent organizations. Google as an entity (aggregated across all their properties) was found to be where the highest percentage of traffic was headed out to. Interestingly, the share of Google-bound traffic from Websites (58%) was found to be significantly higher than from their Android and iOS (30-38%) counterparts. Facebook came a distant 2nd with a presence across all 3 channels studied (9-14%). We also observed that some 3rd Parties were channel-specific ie. Microsoft was primarily seen on websites whereas Amazon was primarily seen in Mobile Apps

Number of 3rd Parties identified in Android Apps is significantly higher (1.6X) as compared to Websites and iOS Apps. This can be possibly attributed to the Open Source coding and exchanging of codes within the Web developer community.



3rd Party Organizations



Only the top 5 organizations have been featured and the remaining organizations have been moved into the "Others" Category

Top 5 3rd Parties



Amazon	Crashlytics - Google	DoubleClick Ads - Google
Crashlytics - Google	Google	Google
Facebook	Facebook	Google Analytics
Google	Google Analytics	Facebook
DoubleClick Ads - Google	DoubleClick Ads - Google	Google Ad Services/ Syndication



Is your Personal Data secure?

The study reviewed Security from the perspective of the entire Information Lifecycle of a user's Personal Data with an App or a Website. We looked at security from the points of view of how it is Stored, Transmitted, Used in the Code, and Destroyed or Deleted. Overall Security was found to be the most mature area among the ones covered as part of this Study. This is a heartening statistic as it shows that the maturity of organizations when it comes to security is significantly higher than that of privacy.





Websites

Only **4%** websites were observed transmitting data in an unencrypted manner; **80%** of the "Games" websites failed in this regard. Interestingly, although websites were found to use HTTPS, **41%** websites did not encode/encrypt username and password at the client side prior to transmission. We also observed that only **3%** session cookies are stored in a secure manner. This may be because these session cookies were generated prior to login into the website. Generally, session cookies are generated once the user logs into the website.



iOS

Although none of the Apps took Root level access of the device, we observed that **22%** of the Apps had certificates/encryption keys visible in the source code. Only **29%** Apps were transmitting data in a secure manner using HTTPS prior to login. We are confident that they probably switch to HTTPS once a user logs into the application.



Android

Only **4%** of Apps did not clear the data post uninstallation of the App. Communication Apps comprised half the Apps that failed in Secure disposal

Are Organizations being Transparent with you?

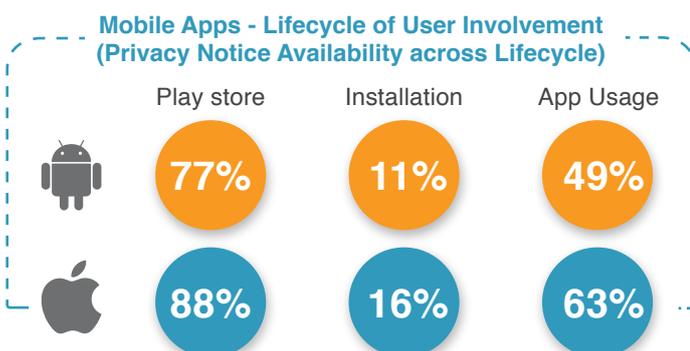
To test how transparent organizations were being with users via their Privacy Notices and how easy were they making it for users to understand their practices, we tested Privacy Notices via three lenses:

- A. How easily was the Notice available to a user at all stages of her engagement lifecycle with the organization’s App or Website?
- B. How complete were the contents of the Notice?
- C. How easy was it to read?

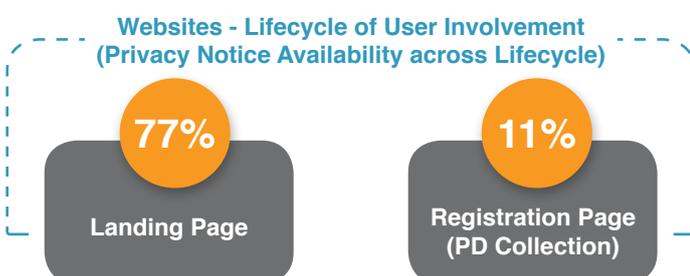
A. Notice Availability

Key Findings:

For an App, the lifecycle of user engagement extends from the time of procurement of the product from the App Store, to installation, registration, usage and deletion. We found that the Privacy Notice is not easily available consistently at all points of the User lifecycle. It is seen to be high (>75%) at the first touchpoint (playstore or landing page), which drops at the subsequent touchpoints like installation, registration and usage. This pattern is observed across Apps and Websites. This may probably be because playstore policies require Notices to be made available - subsequently, there is no such push from anyone.



Above percentages indicate the % of organizations where Privacy Notice was available at the particular stage of the lifecycle



To understand more about permissions and tracking, look up www.arrka.com



B. Notice Analysis - Content

We at Arrka have developed a Notice Completeness Framework that checks for how comprehensive the contents of a Notice are. We used this framework to test for completeness of the Notice.

Key Findings:

Notice content was found to be consistent across channels. In most cases, the same Notice was used across channels. Overall Notice completeness scores across Websites & Apps were around 43-47%. Critical categories like Banks were found to be amongst the worst performers with completeness scores of 12%. High Inter-Category variation was observed which indicates that the industry has not yet reached any level of standardization or consistency in Notice content. We expect this variation to reduce in the coming years with the coming of the India Data Protection Bill. Only 32% organizations have provided the details of a Grievance officer or any person to whom issues and complaints pertaining to Privacy could be raised.

Notice Content Completeness – Overall Score: 45%

Notice Content Completeness - Categories	
Best Performers	Worst Performers
65-75% Completion Sports, Vehicles, Medical Music and Audio	<25% Completion Ticket Booking, Banks, Travel - Maps & Info Education

Notice Content Completeness - Topics in Notice



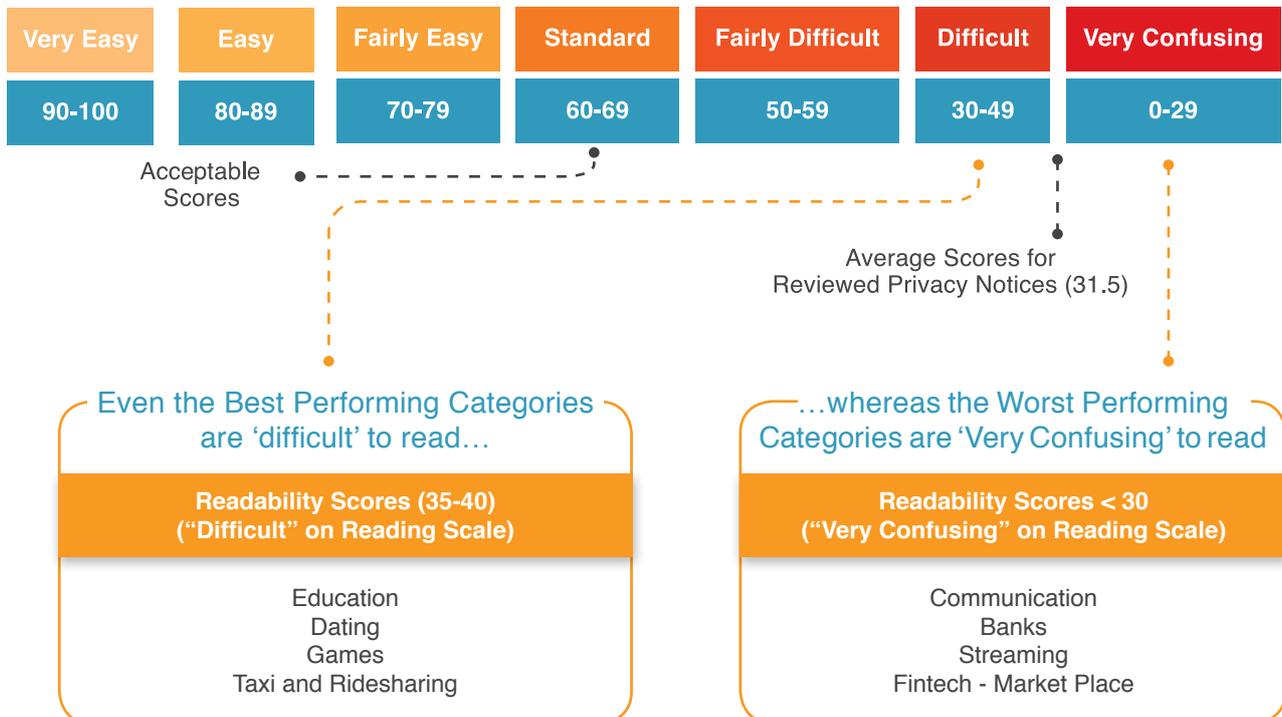
C. Notice Analysis - Readability

To analyze Notice Readability, we used the Industry Standard “Fleisch Reading Ease Scale”. The Fleisch Reading Ease scores are being used as a standard readability formula by many US Government Agencies. Standard Acceptable scores on the Fleisch Reading Ease Scale are 60-70 (on a scale of 0-100).

Key Findings:

The average readability score of Privacy Notices studied is 31, which puts it in the ‘Difficult to Read’ bucket and bordering on the ‘Very Confusing’. Even the best performing Notices still fall in the ‘Difficult to Read Category’ whereas the worst ones fall in the “Very Confusing” category, which is the lowest on the Fleisch Scale. Banks and Fintech market places which are critical categories were amongst the worst performers.

Fleisch Reading Ease Scale



Are Children's Apps safe?

Children are a particularly vulnerable category. Hence, one area we specifically studied over and above the 100 base Apps were Android Apps from India targeting children. The Google Play store categorizes children's Apps based on age group. We selected Apps from each category there and studied them from a privacy perspective.



Key Findings:

First of all, it was heartening to see that **29%** Apps took NO Permissions at all. **29%** Apps had access to Location & phone details. **71%** Apps had access to Storage. Our analysis also found that **56%** of the permissions accessed were not required.

On the other hand, **100%** Apps have links to other Apps.

Additionally, **71%** Apps were found to contain In-App ads. And the Ads were not found to be child safe. Ads shown were for shopping, part-time studies, women entrepreneurs, real estate, physiotherapy and healthcare. Ads were also found redirecting the user to other websites without any consent.

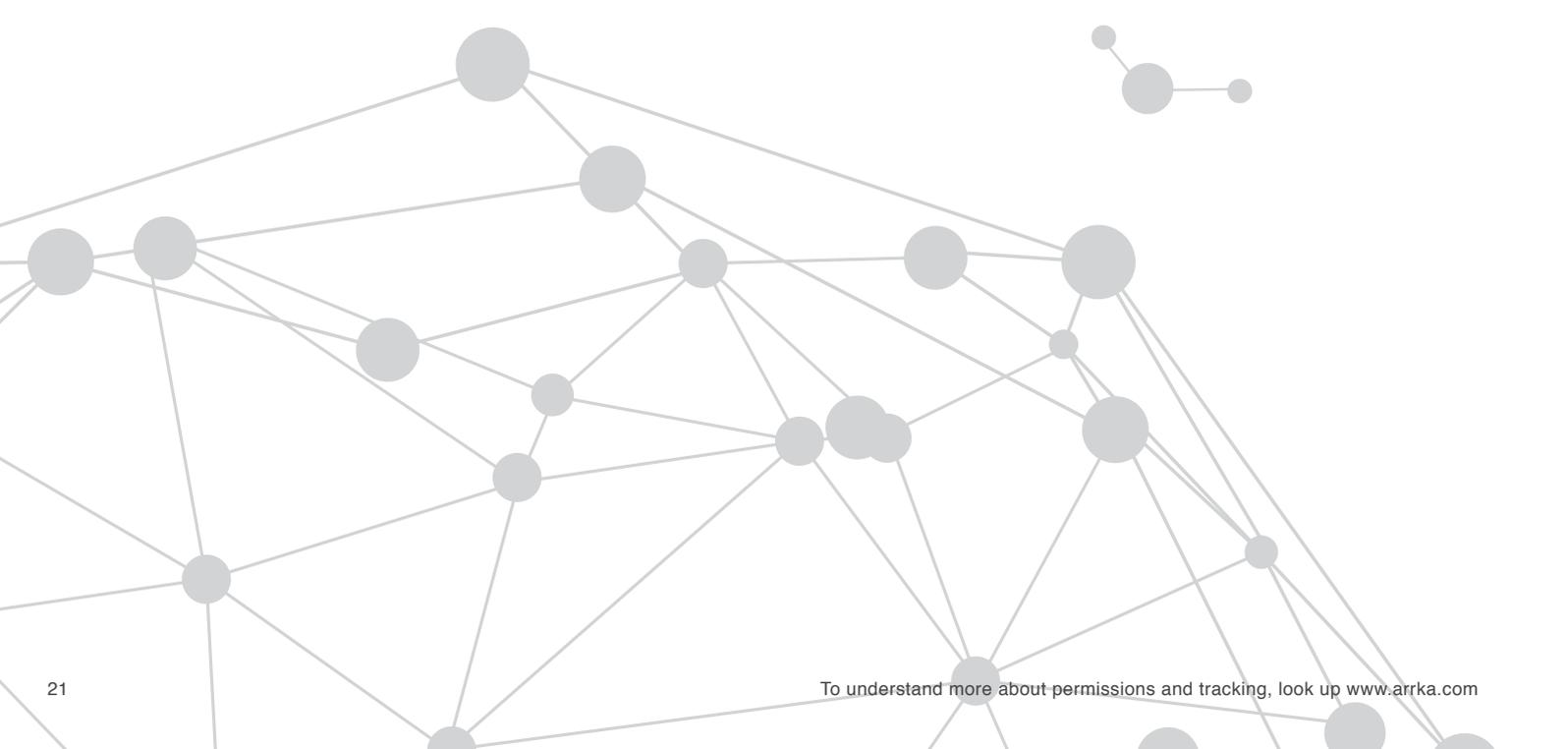
43% Apps offered In-app purchase options. No consent or verification of an adult was found required to make the purchases.

29% of Apps did not have a Notice addressing children under age 13. In **86%** Apps, consent was not being taken. Even when consent was taken there was no verification to check if the person was an adult.

Conclusion

Data Privacy is still at a nascent stage in India. While there is plenty of dialogue going on and progress has been made on the policy and regulatory front, the average Indian organization seems to be quite far from taking any concrete steps to translate privacy related concerns into specific actions on the ground. Many basic steps towards ensuring privacy are yet to be taken. The only exceptions seem to be organizations whose digital properties have exposure to privacy - mature markets like the EU (with the GDPR), Singapore, Canada, US, etc.

However, one specific aspect of privacy - that of the security of Personal Data - seems pretty mature. This heartening statistic shows that the maturity of organizations when it comes to security is significantly higher than that of privacy as a whole. This leads us to conclude that as awareness increases and is combined with regulatory and legal pushes, adoption and maturity of privacy would go up in India, that too rapidly and significantly.



Authors



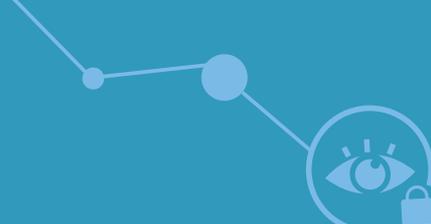
Shivangi Nadkarni
Co-Founder & CEO - Arrka



Sandeep Rao
Principal Consultant - Arrka



All the testing for this study was carried out at the Arrka Privacy Testing Lab. A one-of-its-kind lab in India, it is dedicated exclusively to privacy testing of mobile apps, websites and other digital properties & technology infrastructure.



Arrka: Empowering organizations to manage their Data Privacy and Information Security Programs

Arrka enables organizations to Assess, Design, Implement, Maintain and Manage their Data Privacy and Information Security Programs. Using a combination of the Arrka Intelligent Platform, Custom Toolkits, One-of-Its-kind Testing Lab, Learning Academy and Domain experts, we offer separate solutions for Large Enterprises and Small & Mid-Sized Businesses (SMBs) to address their unique challenges in these domains.



Pune Office:

44/1, First Floor, Namrata Apts.
Amar Housing Society, Gulmohar Path
Off Law College Road, Pune 411004

Mumbai Office:

201, 2nd Floor, Marathon Future IT Park Building
Mafatlal Chamber 'A', Mafatlal Mills Compound
N.M. Joshi Marg, Lower Parel, Mumbai 400 013

www.arrka.com

✉ privacy@arrka.com

🐦 [@arrka2](https://twitter.com/arrka2)

🌐 www.linkedin.com/company/Arrka

📘 <https://www.facebook.com/arrkaconsulting/>

All brand names, logos and digital properties referred to in this report are the property of the respective organisations. This material and the information contained herein has been prepared by Arrka Infosec Private Limited ("Arrka"). It is intended to provide general information on the subjects under consideration and is not an exhaustive treatment of the said subjects. The information is not intended to be relied upon as the sole basis for any decision which may affect you as an individual or your business. Arrka shall not be responsible for any loss whatsoever sustained by any person who relies on this material.

©2018 Arrka Infosec Private Limited

