An International View of Privacy Risks for Mobile Apps

Jingjing Ren, Daniel J. Dubois, David Choffnes
Northeastern University, Boston, MA, USA
ren.ji@northeastern.edu,
choffnes@ccs.neu.edu,
d.dubois@northeastern.edu

ABSTRACT

Current understanding of mobile apps privacy is mostly based on studies conducted in markets with the highest penetration of Internet-connected mobile devices among the general population, such as the US. However, such studies may not represent well the mobile apps privacy situation in the rest of the world. For example, China and India have a lower penetration, but a higher number of mobile Internet users than the US, making them notable markets. To give an idea of how mobile privacy may be differently affected in other countries, we analyze the 100 most popular iOS and Android apps in India and China, and compare them with the US. Key findings from our analysis show that China and India have a worse overall privacy situation than the US, and that China has a worse overall privacy situation than India.

1. INTRODUCTION

The use of mobile apps on Internet-connected mobile devices has always raised the question on how they affect the privacy of their users. Studying the privacy of mobile apps in general is not easy since the application ecosystem depends on: (*i*) the operating system (OS), (*ii*) the region in which the app if offered. Although the OS is typically considered, most existing studies mainly focus on western countries. One reason is that they have typically the highest percentage of mobile Internet users: the penetration of mobile Internet among the general population in the US in 2017 was 73% [31]. However, the situation in the rest of the world is not the same. For example, in China such penetration was 47% [29], and in India it was 24% [30].

In this work, we advance the state of the art by investigating what privacy differences there are between one of the most studied mobile app markets (i.e., the US one) with some of the fastest growing, highly populated, and differently regulated international markets. We found that China and India fulfill these requirements. Both of these countries have different customs, cultural roots and privacy laws when compared to the US; moreover, their markets evolved in different ways, resulting in different sets of apps and apps popularity. China has also notable particularities: due to local laws and strict Internet control, it has not only different apps when compared to the US, but also the widespread use of unofficial app stores, due to the unavailability of the Google Play store in the Android platform.

Our main goal is to analyze the privacy differences between the top US mobile apps and the top apps of China and India. Understanding such differences can be of critical importance by different types of stakeholders: (*i*) *persons* or *organizations* living in such markets or using apps from such markets; (*ii*) local governments and policy-makers to define new regulations; (*iii*) *researchers* as motivation to spend more effort in considering additional countries in follow-up research.

As personal information, by definition, is disseminated to the Internet from mobile apps, we focus our measurements on the network traffic. In our experiments, we gather network traffic (both cleartext and encrypted) from manual interactions with the 100 most popular apps available in the US, in China, and in India. We then use ReCon [26], which leverages machine learning algorithms, to detect any PII (Personally Identifiable Information) dissemination in the network traffic. In the privacy analysis, we consider the amount and type of disseminated PII, the destination of such information (first party vs third party), and whether such information was sent encrypted or not. Our results show that popular apps from each country expose user privacy in different ways, different amounts, and to different parties, thus confirming our hypothesis that results from privacy studies in a mobile app market may be difficult to generalize.

Overall, we have found that, according to the privacy metrics we defined, China and India have both a worse privacy situation than the US, with China having a significantly wider gap, and India sharing several similarities with the US. We correlate these findings with symptoms of possible technological lag of China's apps compared to other markets. For example, in China we have measured a lower adoption of Google guidelines [13] and lower adoption of encrypted protocols over secure ones, which in a previous study [25] has been correlated with outdated apps. We have also correlated our findings

with local regulations. For example, China tends to use local advertising and analytics (A&A) services as opposed to international ones, probably as a result of its stricter Internet controls, whereas India, which suffers from unenforced privacy laws [17], shows both a wide range of local and international A&A services.

2. RELATED WORK

Mobile App Privacy. Early studies showed that popular mobile apps expose location, usernames, passwords, and phone numbers [32], demonstrating the need for further investigation. Then, follow-up studies observed similar behavior at scale [33, 20, 26]. Several efforts systematically identify situations in which PII are diffused over the Internet, and develop defenses against them [26, 36, 15, 5, 10, 34, 21, 12, 16, 35, 37, 4, 14, 8]. One problem of these studies is that, although they are successful in demonstrating that privacy in mobile apps is a problem, their measurements lack an international perspective, which is what this work addresses.

Mobile Privacy Dynamics. Understanding what affects mobile privacy has been widely studied across several dimensions. Some studies focused on the web app vs native app dimension [19, 23], thus comparing the privacy variances between mobile applications and their web counterparts. Other studies focused on the OS dimensions [26], thus analyzing the differences in privacy between the same apps deployed for different OSes. Another dimension that has been analyzed is *time*, i.e., how the privacy of apps changes overtime [25]. In this work, we provide insights for a new international dimension, starting with the case of China and India.

Mobile Experimentation Methods. There are two main techniques for analyzing mobile app privacy. The first is based on static [6, 28, 27] and/or dynamic [11] analysis, meaning that the app code (static) and/or its runtime execution trace (dynamic) are inspected to detect access and dissemination of PII stored or sensed by the devices. This technique is hard to employ since it needs modification of mobile OSes or access to app code. The second is based on network traffic analysis, where app functionalities are tested, and the resulting network traffic is analyzed. The idea behind network traffic analysis is that PII exposure almost always occurs over the Internet. Testing app functionalities for network traffic analysis can either be done automatically or manually. Automated approaches are popular for scalability reasons [22, 14, 7]. However, they do not work well in every situation, for example they cannot automatically explore apps that require signing in [9] and they are more prone to miss exposing PII [26]. Since in this work we prioritize accuracy and also consider apps requiring sign-in, we use manual tests.

Network Traffic Analysis Approaches. Network traffic analysis has the benefit of being agnostic with respect to platforms and OSes, but it requires the ability to reliably identify PII (which may be encrypted and/or obfuscated) in network traffic. Several approaches support TLS interception to access plaintext traffic to search for PII, but differ in what they search for: most approaches need to know in advance the list of PII they are looking for [24, 28, 18], whereas ReCon [26] does not, since it uses machine-learning to infer a broader range of PII. The explanation is that, not relying on a predefined list of PII allows the approach to detect PII in situations where they are encoded differently (e.g., GPS location), unpredictable (e.g., user input), or encrypted. In order to maximize the number of PII we can detect, in this work we will use ReCon [26].

3. METHODOLOGY

Our approach can be summarized as follows. We first select the 100 most popular apps for iOS and Android from each country (US, China and India). Then, we manually interact with each app and collect all the traffic it has sent over the Internet. Finally, we analyze such traffic to determine the PII contained, its destinations, and whether it has been sent in plaintext or encrypted.

3.1 Selecting Mobile Apps

The first step to our experiments is to select mobile applications. Since our goal is to gain insights on the privacy practices of mobile applications that are prevalent in China and India with respect to the ones prevalent in the US, we select popular apps that are used by the majority of the population. In the case of iOS, we use, for all three countries, the list of the top 100 apps ranked as most popular by Apple in its official App Store. In the case of Android, we use the same criteria with the Google Play Store for the US and India. For China, since the Google Play Store is not available, we use the top apps available in the most popular unofficial stores, which provide a ranking based on the number of downloads: Tencent MyApps [3] and 360 Mobile Assistant [1]. When selecting apps from multiple stores, such in the case of China, we consider the first 100 highest ranked apps that appear on either or both stores.

3.2 Experiments

Testbed Setup. Our experiments involved one Nexus 5 with Android 6 and one iPhone 5 with iOS 10 for apps in India and the US, one Nexus 5 with Android 6 and one iPhone 5s with iOS 10 for apps in China. Each device is pre-configured with *Mitmproxy* [2] to intercept both HTTP and the plaintext content of HTTPS traffic. Before each interaction, we uninstall all non-stock apps and clear caches from previous experiments.

Manual Interaction. Each experiment consists of manually interacting with a given app for five minutes and test all the main features. For each app requiring a login, we created a new account using a previously

unused email address. The traffic generated to create new accounts is not considered in our analysis.

3.3 Identifying Privacy Metrics

As established in previous work [25], we first separate network traffic into network flows, each one defined as a single outgoing HTTP or HTTPS request, then – for each flow – we identify the following metrics:

- *PII*, which measures the number of PII and PII types included in the flow;
- *communication protocol*, which can be either HTTP or HTTPS;
- *destination domain party*, which can be either first-party (when such domain is related to the app provider) or third party (when such domain is not related to the app provider, for example advertisement and analytics services).

PII Identification. To identify PII we use ReCon [26], which uses a machine learning algorithm for effective PII identification without knowing the PII values in advance. Not needing PII values in advance allows us, for example, to detect GPS locations with different precision levels, passwords hashed in different ways, devices IDs encoded in non-standard ways, user input, etc. Most of these variations would be missed if we simply search the traffic for PII using a string-matching approach against known PII values. Once we have identified the PII, we classify each PII according to a *type*. We use types to group different PII containing the same type of information. The full list of PII types we consider is reported in the first column of Table 3.

Protocol Identification. We distinguish if a flow is HTTP or HTTPS using the output provided by *Mitmproxy* [2]. In our setup, we instrumented Mitmproxy to collect HTTP(s) flows that are transmitted over standard ports (i.e., 80, 443, and reasonable variations).

Destination Domain. To classify the destination domain of a network flow, we match the second level part and the WHOIS information for such domain against the app name and the app organization name that generated the flow. If a match is found, such domain is considered a first-party, otherwise it is a third-party.

4. RESULTS

In this section, we describe our dataset and compare the measured privacy metrics by *market*, where we define market as a combination of country (US, China, India) and mobile OS (iOS, Android).

4.1 Dataset Description and Summary

We consider the 100 most popular apps in each market on August 2017, with experiments running between August and November 2017. Table 1 shows which apps overlap among different markets. As we can see, the set of popular apps from China is disjoint from the US and has only one app in common with India in the iOS market. In the case of India, we can see 7 (iOS) and 8 (Android) common apps with the US, meaning that also in this case, the set of apps is mostly disjoint. These differences in app popularity explain some of the privacy differences we will find in our analysis.

OS	C1	C2	#	Shared Apps
Andriod	US	CN	0	-
	US	IN	8	Amazon Kindle, Amazon Shopping, ESPN, Fitbit, OkCupid, Peel Smart Remote, Pinterest, SURE Universal Smart Remote.
	CN	IN	0	-
iOS	US	CN	0	-
	US	IN	7	Google Translate, Medscape, Nike+ Run Club, Pinterest, Sarahah, Tinder, Waze.
	CN	IN	1	UC Browser.

Table 1: **Common apps for each region, by OS**. We can see that regions have very few apps in common.

Market	Flows	PII flows	PII instances	PII types	3rd-party domains	Cleartext flows
CN Android	57.7K	19.5%	19,773	12	362	79.1%
CN iOS	61.1K	15.0%	12,209	6	363	52.2%
IN Android	7.5K	29.9%	3,290	14	234	30.8%
IN iOS	19.8K	16.3%	3,949	9	335	21.1%
US Android	12.4K	14.1%	2,369	12	262	25.7%
US iOS	20.5K	7.8%	1,847	8	286	23.4%

Table 2: **Summary by market.** For each market, we first show, from left to right, the following: (*i*) the number of flows generated by the top 100 apps; (*ii*) the percentage of flows containing at least one PII; (*iii*) the number of PII instances found; (*iv*) the number of PII types found; (*v*) the number of second-level third-party domains that have been contacted; (*vi*) the percentage of cleartext flows. Higher numbers (as in the case of China) represent higher privacy exposure.

A summary of the data we measured is reported in Table 2. For each market, we can see the total number of flows we measured, the percentage of them containing at least one PII instance, the number of PII instances, the list of types of such PII instances, the number of third-party domains contacted, and the percentage of cleartext flows. This summary clearly shows that China has in general the highest numbers, followed by India, and then the US. This means that, under the assumption that such metrics are correlated with privacy risk [25], China may be riskier than India, and India may be riskier than the US, from a privacy perspective. The remainder of this section will do a more detailed analysis of our dataset to explain possible causes and privacy implications of these patterns.

4.2 PII Dissemination Analysis

Amount of PII disseminated. By looking at the percentage of the flows containing at least one PII and

	Android			iOS		
PII Type	**US**	**CN**	**IN**	**US**	**CN**	**IN**
Google Ad ID	84.7%	22.3%	**87.3%**	0	0	0
GSF ID	**74.1%**	0	58.2%	0	0	0
Android ID	62.4%	67.0%	**83.5%**	0	0	0
Location GPS	40.0%	**47.9%**	44.3%	39.7%	51.2%	**51.8%**
Email	36.5%	62.8%	**63.3%**	13.2%	0	**42.4%**
Zip Code	**10.6%**	0	1.3%	4.4%	0	**4.7%**
Gender	5.9%	**22.3%**	12.7%	11.8%	13.8%	**27.1%**
IMEI	4.7%	**79.8%**	10.1%	0	0	0
MAC Address	1.2%	**31.9%**	2.5%	0	0	0
Serial Number	1.2%	**23.4%**	10.1%	0	0	0
First Name	1.2%	1.1%	**19.0%**	4.4%	7.5%	**24.7%**
Phone Number	0	2.1%	**11.4%**	0	0	**28.2%**
Last Name	0	0	**13.9%**	2.9%	0	**16.5%**
iOS Ad ID	0	0	0	**82.4%**	88.8%	**95.3%**

Table 3: **PII type transmitted by apps in each market.** Bold values represent the most disseminated PII types. Notable is the high dissemination of hardware identifiers in the Chinese Android market.

the total number of PII instances transmitted in Table 2, we can see that both China and India have a higher amount of PII dissemination for both Android and iOS when compared to the US, making them riskier. In the comparison between China and India we can see that India has a higher percentage of flows with PII, while China has a higher number of PII instances. This means that a flow from India has a higher probability to contain a PII, while an app in China is likely to share more PII in absolute terms.

Type of PII disseminated. The "PII types" column of Table 2 shows that Android shares more PII types than iOS across all markets, moreover we can see slight differences between countries. To better understand such differences, we show in Table 3 how the actual set of PII types differs among markets. In the case of Android markets, we notice that location is disseminated in a comparable way among the three countries, while the other types follow many different patterns. For example, apps in India have a higher tendency of disseminating personal information such as names and phone numbers when compared to the other countries. We also notice that in the Chinese Android market the Google Ad ID has a very low dissemination rate (22.3%), while in the other countries it is over 80%; on the contrary, China has a very high IMEI dissemination rate (79.8%), while the other countries have 10% or less. The same pattern can be seen with other hardware identifiers such as MAC Address and Serial Number. Sharing hardware identifiers is a serious privacy problem since they cannot be easily changed by users. Previous work [25] reported that new Android apps and new versions of existing Android apps are moving from IMEI and other hardware identifiers to Google Ad id, following Google guidelines [13]. Therefore, having a high amount of apps transmitting hardware identifiers may be a symptom that the development of top Android apps in China is lagging with respect to adopting Google guidelines.

Takeaways. A takeaway from this analysis is that, with respect to the quantity of PII disseminated, China is worse than India and India is worse than the US, where by *worse* we mean more chances of privacy violations regarding PII. Regarding the type of PII disseminated, we see in general too much diversity to support any meaningful generalization. We can, however, say that apps targeted at different global audiences share some types of PII collection, but differ in others. To support this, we show in bold in Table 3 the measurements for the markets with the highest PII dissemination, and therefore privacy risk, for a specific PII type. Another takeaway is the delayed adoption of the Google guidelines on Android identifiers, which is a symptom of technological lag concerning mobile apps.

4.3 Third Parties Comparative Analysis

Number of contacted third parties. To compare the contacted parties we have grouped all actual domains according to their second-level domain name, and then determined whether each one is a first party (i.e., the app provider) or a third-party (i.e., anything that is not the app provider) using the approach described in §3.2. By looking at the "3rd-party domains" column of Table 2, we can see that China's apps contact more third parties domains than the ones of the US and India for both OSes, while India's apps contact less third parties than both China and the US. Third parties, by our definition, are distinct entities with respect to the app provider (with whom the user has a direct relationship), therefore we consider a privacy risk to give third parties any data. Based on this consideration, we consider China riskier than the US and the US riskier than India, with respect to the number of third parties.

Third-party domains by app. Table 4 shows the number of apps that contacted each particular third-party domain. By analyzing the table we can notice that Chinese markets contact different third-party domains when compared to the US and Indian markets. For example, Google, Crashlytics, Appsflyer, Cloudfront, and Amazon AWS dominate the US and Indian markets, while their appearance is marginal in the Chinese market, where domains like Umeng, QQ, Baidu, Taobai, etc. are much more popular. We also spot significant differences between the US and Indian markets. For example, India has a higher number of Google-related third-parties with respect to the US. These results show that the "third-party ecosystem" is heavily regional and that third-party domain analyses that are specific for a region cannot be necessarily generalized.

PII types similarities by domain. In Table 5 we can see that there is a significant overlap of commonly contacted domains among markets. This raises the question of *how similar* the PII types sent across these common domains are. Figure 1 answers this by showing the Jaccard similarity index for PII types shared to the

	Android			iOS		
Domain	**US**	**CN**	**IN**	**US**	**CN**	**IN**
google.com	78	46	**79**	3	1	**11**
crashlytics.com	49	10	**58**	2	1	**4**
appsflyer.com	**25**	4	16	19	5	**24**
doubleclick.net	**20**	6	17	**39**	4	**39**
googleadservices.com	18	14	**34**	20	10	**31**
googleapis.com	16	**71**	0	19	4	**25**
cloudfront.net	14	0	**31**	17	0	**21**
amazonaws.com	13	0	**20**	**21**	0	18
moatads.com	**12**	1	4	**14**	0	4
branch.io	**11**	0	10	1	0	0
scorecardresearch.com	**10**	1	11	**22**	0	10
facebook.com	**9**	7	0	11	1	**24**
umeng.com	5	**38**	2	0	**37**	4
gstatic.com	**5**	3	0	23	3	**25**
google-analytics.com	4	**12**	0	25	5	**46**
googlesyndication.com	**3**	2	0	**31**	3	27
mopub.com	**3**	0	2	**22**	0	14
flurry.com	**2**	**2**	0	13	2	**22**
qq.com	1	**48**	1	3	**60**	2
baidu.com	0	**41**	0	0	**28**	0
taobao.com	0	**29**	1	0	**19**	2
amap.com	0	**26**	0	0	**5**	0
qlogo.cn	0	**24**	0	0	**30**	0
irs01.com	0	**17**	0	0	**4**	0
igexin.com	0	**17**	0	0	**1**	0
gtimg.cn	0	**16**	0	0	**13**	0
alipay.com	0	**15**	1	0	**13**	2
weibo.com	0	**11**	0	3	**55**	3
p3-group.com	0	0	**22**	0	0	0
apple.com	0	0	0	**97**	74	82
icloud.com	0	0	0	4	15	**22**

Table 4: **Third party domains by number of apps across markets.*** Values corresponding to the markets in which a domain is contacted by the most apps are reported in bold. US and India share some similarity, while China has very different third party domains.
*We only show domains among the top 10 most contacted domains by number of apps in at least a market.

same third party domains across every pair of countries. From the figure we can see the following trends: US/India and China/India have the highest similarities for Android and iOS, while China/India (Android) and US/China (iOS) have the lowest similarities. Since the results vary heavily by OS and country, we cannot consistently say whether the same third parties collect the same/different PII types across markets. However, we can see enough PII type similarities to claim that common domains, due to their presence in more than one market, may be able to track a user that moves from a country to another.

Takeaways. The main takeaway from this analysis is that, from a quantitative point of view, China's top apps are riskier than the US ones since they share data with more third parties, while India's ones are less risky for the same reason. We have also noticed a much more different set of third parties in the case of China. A possible explanation is that China has much tighter Internet controls than India and the US, resulting in a proliferation of locally regulated third parties (i.e., advertising and analytics service providers) with respect to international ones.

OS	C1	C2	#Domains	#1st Party	#3rd Party
Andriod	US	CN	33	0	33
	US	IN	98	7	91
	CN	IN	33	2	31
iOS	US	CN	44	0	44
	US	IN	137	11	126
	CN	IN	89	1	88
ALL			10	0	10

Table 5: **Common domains among markets.** We can see that Indian domains have much more similarities to the US ones than to the Chinese ones.

4.4 HTTPS Adoption Analysis

Encrypted Flows Analysis. The "Cleartext flows" column of Table 2 shows that the Chinese market has the largest fraction of cleartext (HTTP) flows, followed by India, and then by the US. Having a high amount of cleartext HTTP is risky from a privacy perspective since such traffic can be intercepted.

Destinations Analysis. Figure 2 analyzes how HTTP and HTTPS capabilities are distributed by destination party, i.e., the percentage of domains that are contacted using HTTP, HTTPS, or both. This analysis is important to understand if cleartext traffic is mostly due to domains related to the apps providers (first-party) or to external third-party services. In the case of China, we have the highest percentage of contacted domains that only support HTTP among all countries, for both OSes, without significant differences regarding destination parties. In the case of India, its third-party domains that only support HTTP are significantly less than the US in the Android platform, while they are more than the US in the iOS platform, meaning that the risk caused by third parties using HTTP depends on the OS for these countries. Regarding first parties, the domains in India and the US behave in a comparable way with respect to HTTP-only destinations.

Takeaways. Previous work [25] showed correlation between old apps or older versions of existing apps and lower HTTPS adoption, meaning that a lower trend of HTTPS adoption in a certain market, such as we observed in China, can be a symptom of apps not being properly updated or properly actively developed. Our results also show that a considerable percentage of domains across all markets are contacted with both HTTP and HTTPS (see Figure 2). This is consistent with the fact that HTTPS is being increasingly adopted by new apps, but HTTP is still supported by the same domains for backward compatibility with old apps.

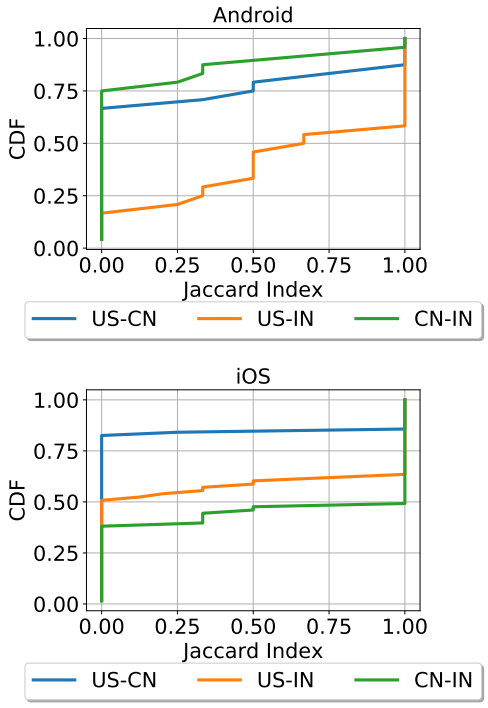

Figure 1: **CDF of the generalized Jaccard similarity index for PII types shared to the same third party in each country pair.** OSes show a different trend, with more similar shared PII types between US and India to common Android third parties, and more similar shared PII types between India and China to common iOS third parties.

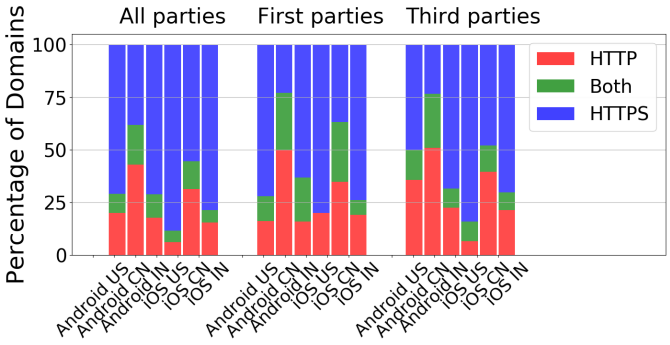

Figure 2: **Protocol distribution of domains contacted across market.** The y-axis shows the percentage of domains that are contacted by HTTP only, HTTPS only or both. China has the highest HTTP adoption across all markets.

5. CONCLUDING DISCUSSION

In this paper, we wanted to answer the question: "*is mobile privacy affected differently in different parts of the world?*". We have found out that the answer is in most cases *yes* when comparing the top 100 apps of China and India to the ones of the US, meaning that the region plays a key role in privacy. To better understand these differences, we have analyzed the trends in PII dissemination, third-parties contacted, and percentage of HTTPS adoption across all the three markets.

We have observed that the privacy situation in China is consistently worse than the US and India according to all three trends: more PII is disseminated, more third-parties contacted, with a lower level of HTTPS adoption. A possible explanation for this phenomenon is that China, due to its tight Internet control and different privacy regulations, has a mobile apps ecosystem that is evolving differently from the US one. Not only the apps and the third party destinations tend to be different from the other regions we considered, but we also observed many signals of slower adoption of newer technologies. For example, Chinese apps are responding more slowly to the Google guidelines regarding not using hardware identifiers with respect to the other markets. Another trend is the low HTTPS adoption, which has been correlated with old app versions in other markets [25]. The implication of this technological lag may go well beyond what we can measure from our dataset. For example, apps not being updated with current guidelines and technologies may expose the users to other privacy and security risks we have not analyzed, such as the presence of exploitable bugs due to the use of outdated libraries.

For what concerns India, the privacy situation is more similar to the US, but still worse overall. Similarities can be explained by some overlap in popular apps and third party destinations, which often tend to be US companies (e.g., Google). Also, the additional PII dissemination and third-party destinations we have observed in India may be partially explained by the fact that India has a less effective (and more difficult to enforce) legislation on data protection [17]. Regarding HTTPS adoption, differences between India and the US are minimal: US is slightly worse than India on Android and India is slightly worse than the US on iOS. We interpret this result as a signal that, as opposed to China, there is no obvious technological lag between US and Indian apps, and that the relatively small variations are mostly due to differences in popular apps.

In conclusion, this comparative analysis has shown significant differences across the three countries. As we have seen, such differences can be correlated to differences in local customs (e.g., different set of popular apps), local legislation (e.g., Internet control and privacy laws), and technological advancement. This motivates the development of future measurement studies taking into consideration also the regional dimension. As a future work, we plan to extend this privacy comparative analysis to other countries and to other apps selected in a different way.

6. REFERENCES

[1] 360 mobile assistant. http://zhushou.360.cn. (Accessed on 05/15/2018).

[2] mitmproxy. https://mitmproxy.org/.

[3] Tecent myapps. http://android.myapp.com/. (Accessed on 05/15/2018).

[4] Agarwal, Y., and Hall, M. ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing. In *Proc. of MobiSys* (2013).

[5] Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Le Traon, Y., Octeau, D., and McDaniel, P. FlowDroid: Precise Context, Flow, Field, Object-sensitive and Lifecycle-aware Taint Analysis for Android Apps. In *Proc. of PLDI* (2014).

[6] Au, K. W. Y., Zhou, Y. F., Huang, Z., and Lie, D. PScout: Analyzing Android Permission Specification. In *Proc. of ACM CCS* (2012).

[7] Azim, T., and Neamtiu, I. Targeted and Depth-first Exploration for Systematic Testing of Android Apps. In *Proc. of OOPSLA* (2013).

[8] Chen, X., and Zhu, S. DroidJust: Automated Functionality-aware Privacy Leakage Analysis for Android Applications. In *Proc. of WiSec* (2015).

[9] Choudhary, S. R., Gorla, A., and Orso, A. Automated Test Input Generation for Android: Are We There Yet? In *Proc. of the IEEE/ACM International Conference on Automated Software Engineering (ASE)* (2015).

[10] Egele, M., Kruegel, C., Kirda, E., and Vigna, G. PiOS: Detecting Privacy Leaks in iOS Applications. In *Proc. of NDSS* (2011).

[11] Enck, W., Gilbert, P., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A. N. TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones. In *Proc. of USENIX OSDI* (2010).

[12] Gibler, C., Crussell, J., Erickson, J., and Chen, H. AndroidLeaks: Automatically Detecting Potential Privacy Leaks in Android Applications on a Large Scale. In *Proc. of TRUST* (2012).

[13] Google. Best practices for unique identifiers. https://developer.android.com/training/articles/user-data-ids. (Accessed on 05/16/2018).

[14] Hao, S., Liu, B., Nath, S., Halfond, W. G., and Govindan, R. PUMA: Programmable UI-Automation for Large-Scale Dynamic Analysis of Mobile Apps. In *Proc. of MobiSys* (2014).

[15] Jeon, J., Micinski, K. K., and Foster, J. S. SymDroid: Symbolic Execution for Dalvik Bytecode. Tech. Rep. CS-TR-5022, University of Maryland, College Park, 2012.

[16] Kim, J., Yoon, Y., Yi, K., and Shin, J. SCANDAL: Static Analyzer for Detecting Privacy Leaks in Android Applications. In *Proc. of MoST* (2012).

[17] Kshetri, N., and Kshetri, N. Cybersecurity in india: Regulations, governance, institutional capacity and market mechanisms. *Asian Research Policy 8*, 1 (2017), 64–76.

[18] Le, A., Varmarken, J., Langhoff, S., Shuba, A., Gjoka, M., and Markopoulou, A. AntMonitor: A System for Monitoring from Mobile Devices. In *Proc. of Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data* (2015).

[19] Leung, C., Ren, J., Choffnes, D., and Wilson, C. Should you use the app for that?: Comparing the privacy implications of app-and web-based online services. In *Proc. of IMC* (2016).

[20] Liu, Y., Song, H. H., Bermudez, I., Mislove, A., Baldi, M., and Tongaonkar, A. Identifying personal information in internet traffic. In *Proceedings of the 3rd ACM Conference on Online Social Networks (COSN'15)* (Palo Alto, CA, November 2015).

[21] Lu, L., Li, Z., Wu, Z., Lee, W., and Jiang, G. CHEX: Statically Vetting Android Apps for Component Hijacking Vulnerabilities. In *Proc. of ACM CCS* (2012).

[22] Machiry, A., Tahiliani, R., and Naik, M. Dynodroid: An Input Generation System for Android Apps. In *Proc. of the Joint Meeting on Foundations of Software Engineering (ESEC/FSE)* (2013).

[23] Papadopoulos, E. P., Diamantaris, M., Papadopoulos, P., Petsas, T., Ioannidis, S., and Markatos, E. P. The long-standing privacy debate: Mobile websites vs mobile apps. In *Proceedings of the 26th International Conference on World Wide Web* (2017), International World Wide Web Conferences Steering Committee, pp. 153–162.

[24] Razaghpanah, A., Vallina-Rodriguez, N., Sundaresan, S., Kreibich, C., Gill, P., Allman, M., and Paxson, V. Haystack: In Situ Mobile Traffic Analysis in User Space. *arXiv preprint arXiv:1510.01419* (2015).

[25] Ren, J., Lindorfer, M., Dubois, D. J., Rao, A., Choffnes, D. R., and Vallina-Rodriguez, N. Bug Fixes, Improvements, ... and Privacy Leaks – A Longitudinal Study of PII Leaks Across Android App Versions. In *Proc. of NDSS* (2018).

[26] Ren, J., Rao, A., Lindorfer, M., Legout, A., and Choffnes, D. R. ReCon: Revealing

and Controlling Privacy Leaks in Mobile Network Traffic. In *Proc. of MobiSys* (2016).

[27] Seneviratne, S., Kolamunna, H., and Seneviratne, A. A Measurement Study of Tracking in Paid Mobile Applications. In *Proc. of ACM WiSec* (2015).

[28] Song, Y., and Hengartner, U. PrivacyGuard: A VPN-based Platform to Detect Information Leakage on Android Devices. In *Proc. of ACM SPSM* (2015).

[29] Statista. Mobile phone internet user penetration in china from 2015 to 2022. https://www.statista.com/statistics/309015/china-mobile-phone-internet-user-penetration/. (Accessed on 05/16/2018).

[30] Statista. Mobile phone internet user penetration in india from 2015 to 2022. https://www.statista.com/statistics/309019/india-mobile-phone-internet-user-penetration/. (Accessed on 05/16/2018).

[31] Statista. Mobile phone internet user penetration in the united states from 2015 to 2022. https://www.statista.com/statistics/275587/mobile-phone-internet-user-penetration-us/. (Accessed on 05/16/2018).

[32] The Wall Street Journal. What They Know - Mobile. http://blogs.wsj.com/wtk-mobile/, December 2010.

[33] Vallina-Rodriguez, N., Shah, J., Finamore, A., Grunenberger, Y., Papagiannaki, K., Haddadi, H., and Crowcroft, J. Breaking for commercials: Characterizing mobile advertising. In *Proc. of IMC* (2012).

[34] Xia, M., Gong, L., Lyu, Y., Qi, Z., and Liu, X. Effective Real-time Android Application Auditing. In *IEEE Symposium on Security and Privacy* (2015).

[35] Yan, L. K., and Yin, H. DroidScope: Seamlessly Reconstructing the OS and Dalvik Semantic Views for Dynamic Android Malware Analysis. In *Proc. of USENIX Security* (2012).

[36] Yang, Z., Yang, M., Zhang, Y., Gu, G., Ning, P., and Wang, X. AppIntent: Analyzing Sensitive Data Transmission in Android for Privacy Leakage Detection. In *Proc. of ACM CCS* (2013).

[37] Zhang, Y., Yang, M., Xu, B., Yang, Z., Gu, G., Ning, P., Wang, X. S., and Zang, B. Vetting undesirable behaviors in Android apps with permission use analysis. In *Proc. of ACM CCS* (2013).