

IMC 2018, Boston, MA, USA

Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets

Haoyu Wang, Zhe Liu, Jingyue Liang, Narseo Vallina-Rodriguez, Yao Guo, Li Li, Juan Tapiador, Jingcun Cao, Guoai Xu



uc3m | Universidad Carlos III de Madrid

institute
imdea

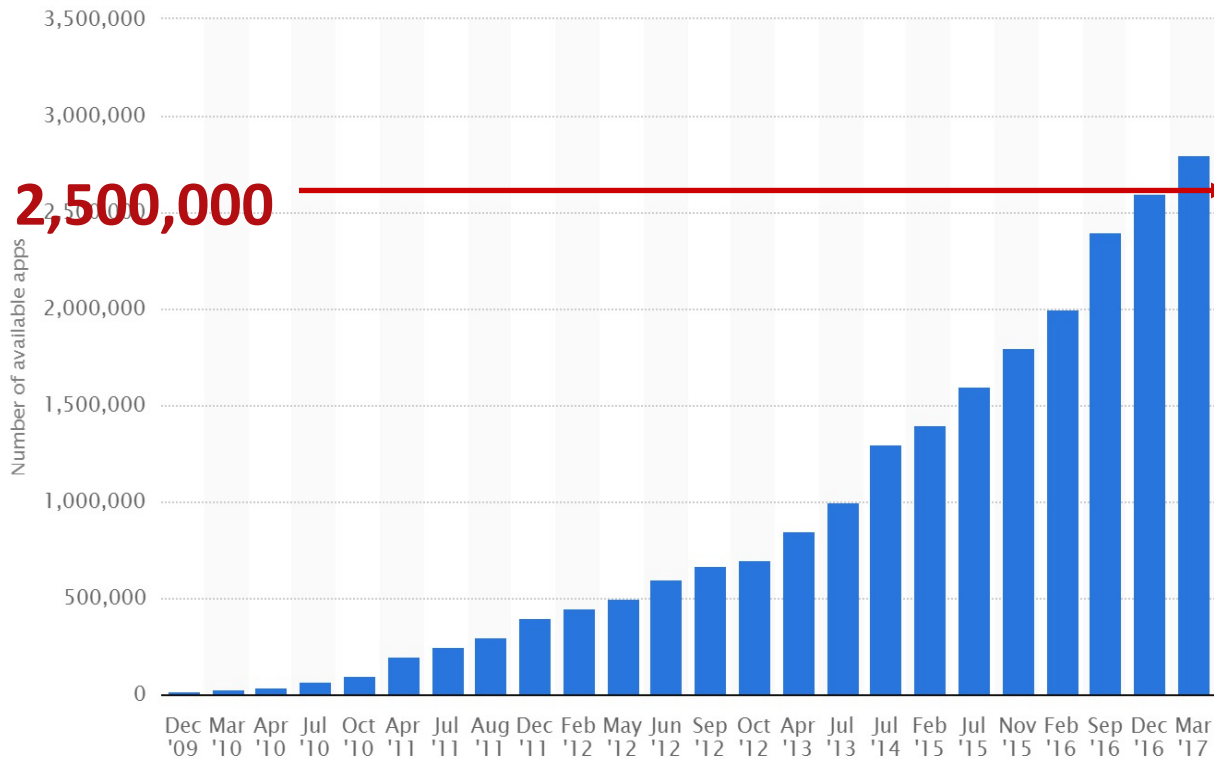


MONASH
University



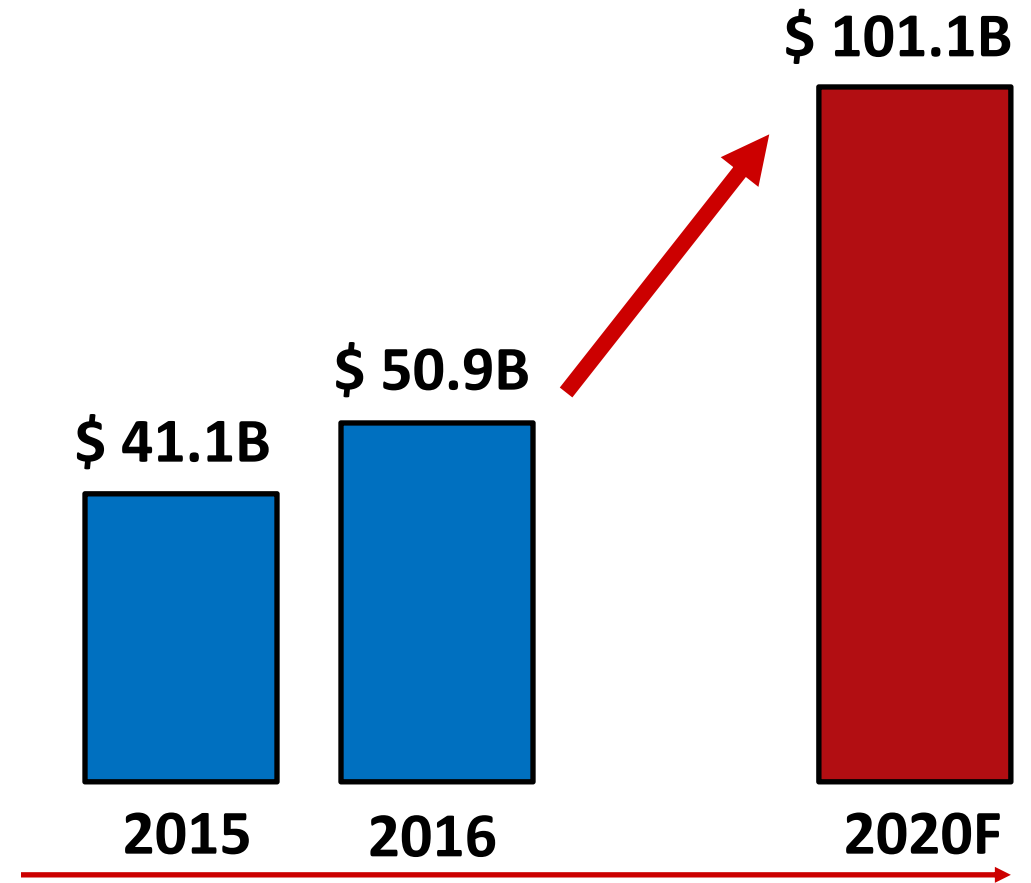
INDIANA UNIVERSITY





**Number of Apps in Google Play:
2.51 million by the end of Oct, 2018**

Source: AppBrain



**App Economy Forecasts:
\$100 Billion Revenue by 2020**

Source: App Annie

Google Play is banned in China

- Great business opportunity (**700+ million users**) for Chinese Internet companies and smartphone manufacturers



TECH

CYBERSECURITY | ENTERPRISE | INTERNET | MEDIA | MOBILE | SOCIAL MEDIA

Google is missing out on billions of dollars by not having an app store in China, new data shows






















- Chinese smartphone users spent nearly \$35 billion on apps in 2017, according to new data from App Annie, which means that Google is missing out on an ever-growing revenue opportunity as it fails to relaunch its Play Store there.
- The company has tried to bring Play back to China for years, to no avail.
- App Annie predicts that consumer app spend will continue to swell because many areas are still in the process of adopting mobile.

Beyond Google Play ...

App Distribution in the West



Making Sense of Chinese App Markets

 Tencent MyApp	 PP Assistant	 Wandoujia
 360 Mobile Assistant	 Meizu Flyme	 HiMarket
 Baidu Mobile Assistant	 Lenovo Le Store	 历趣 用安卓·上历趣
 Huawei App Market	 2345 Mobile Assistant	 机锋市场 apk.gfan.com
 Oppo Software Store	 Anzhi Market	 AppChina.com 应用汇
 MIUI App Store	 Gionee Software Store	 优亿市场 E08MARKET.COM
 VIVO App Store	 Coolmart	 China Telecom 189 Store
 Sogou Mobile Assistant	 China Mobile MM Store	

Our work: roadmap and research questions

- ① Characterization of Chinese markets
 - Market types
 - Copyright checks
 - App auditing (inc. security) process
 - Incentives to attract developers
 - Transparency efforts

Our work: roadmap and research questions

- ② Market contents
 - Catalog diversity
 - User downloads
 - User ratings
 - Targeted users
 - App updates
 - Third-party libraries

Our work: roadmap and research questions

- 3 Publishing dynamics, i.e., relationships among
 - Developers
 - Apps
 - Markets

Our work: roadmap and research questions

- 4 Misbehaviors, i.e., presence of
 - Fake apps
 - Cloned apps
 - Malware

The dataset: Android app markets

■ Official



■ Vendor-specific app markets

Xiaomi Oppo Huawei Lenovo Meizu



■ Web companies

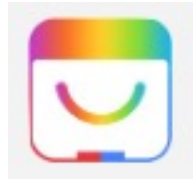
Tencent
Myapp



360



Baidu



■ Specialized markets

25PP Wandoujia HiAPK AnZhi LIQU PC Online Sogou AppChina



Features of app markets — copyright & security checks

■ Official



■ Vendor-specific app markets

Xiaomi Oppo Huawei Lenovo Meizu



■ Web companies

Tencent
Myapp



360



Baidu



■ Specialized markets

25PP Wandoujia HiAPK AnZhi LIQU PC Online Sogou AppChina



PConline
太平洋电脑网



Features of app markets — privacy policy

■ Official



■ Vendor-specific app markets

Xiaomi Oppo Huawei Lenovo Meizu



■ Web companies

Tencent
Myapp



360



Baidu



■ Specialized markets

25PP Wandoujia HiAPK AnZhi LIQU PC Online Sogou AppChina



Dataset: 6.2 million apps crawled

■ Official



2.03M

■ Vendor-specific app markets

Xiaomi Oppo Huawei Lenovo Meizu



91K 426K 51K 37K 80K

■ Web companies

Tencent
Myapp



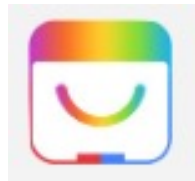
636K

360



163K

Baidu



227K

■ Specialized markets

25PP Wandoujia HiAPK AnZhi LIQU PC Online Sogou AppChina



1.01M



554K



246K



223K



179K

PConline
太平洋电脑网

134K



128K



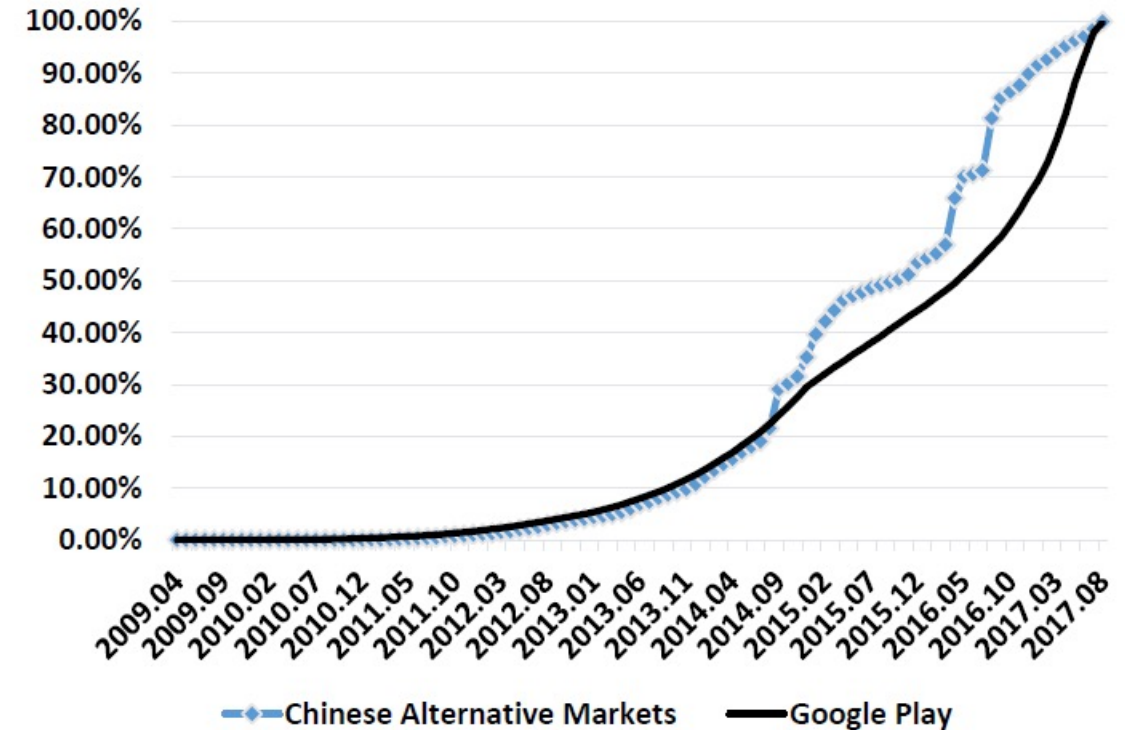
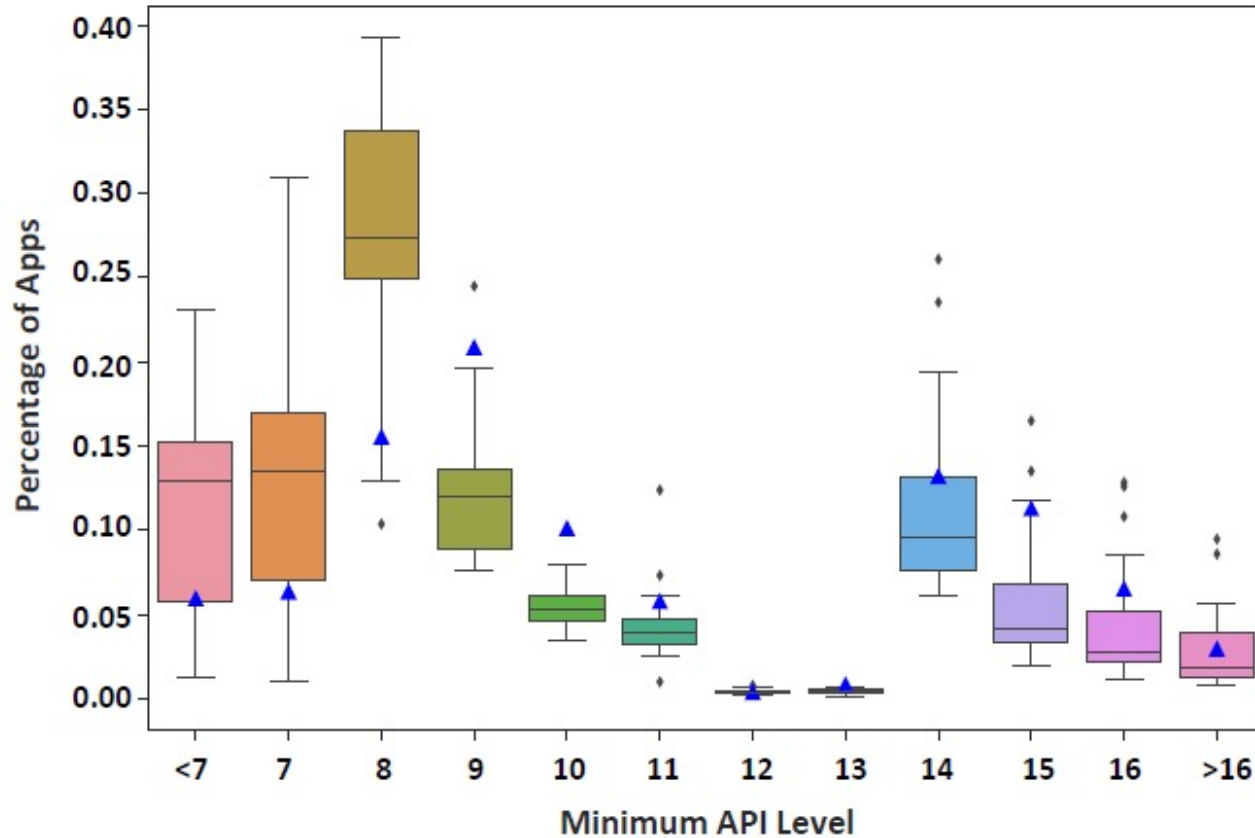
42K

General Overview – App Downloads

	0-10	10-100	100-1K	1K-10K	10K-100K	100K-1M	>1M
Google Play	4.05%	17.90%	30.52%	25.38%	15.15%	5.62%	1.21%
Tencent Myapp	55.87%	12.37%	15.50%	10.38%	4.21%	1.21%	0.35%
Baidu Market	0.00%	34.98%	25.91%	23.21%	7.65%	5.40%	2.26%
360 Market	16.54%	16.08%	19.25%	25.79%	12.78%	7.24%	1.97%
Huawei Market	0.10%	0.00%	38.05%	27.33%	17.64%	11.73%	4.16%
Xiaomi Market	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Wandoujia	1.96%	4.74%	43.66%	35.24%	12.17%	1.77%	0.38%
HiApk	0.00%	0.00%	78.24%	13.15%	5.93%	2.05%	0.53%
Anzhi Market	0.10%	1.35%	49.72%	42.83%	4.86%	0.84%	0.23%
OPPO Market	0.00%	0.00%	84.31%	10.47%	3.16%	1.55%	0.43%
25PP	0.27%	4.63%	68.02%	20.34%	4.82%	1.49%	0.37%
Sougou	0.77%	17.83%	55.13%	22.27%	2.51%	1.15%	0.31%
MeiZu Market	7.63%	13.50%	45.37%	19.54%	7.97%	4.28%	1.42%
LIQU	0.01%	0.03%	0.01%	71.83%	22.32%	5.14%	0.61%
App China	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lenovo MM	0.04%	14.70%	0.00%	53.54%	16.78%	11.02%	3.19%
PC Online	13.07%	74.19%	8.62%	2.98%	0.91%	0.21%	0.02%

- **long-tail distribution:** top 1% of apps usually accounting for over 80% of total downloads across the 17 markets
- Significant differences in the popularity and quality of the apps published in certain app stores

General Overview — Minimum API Level



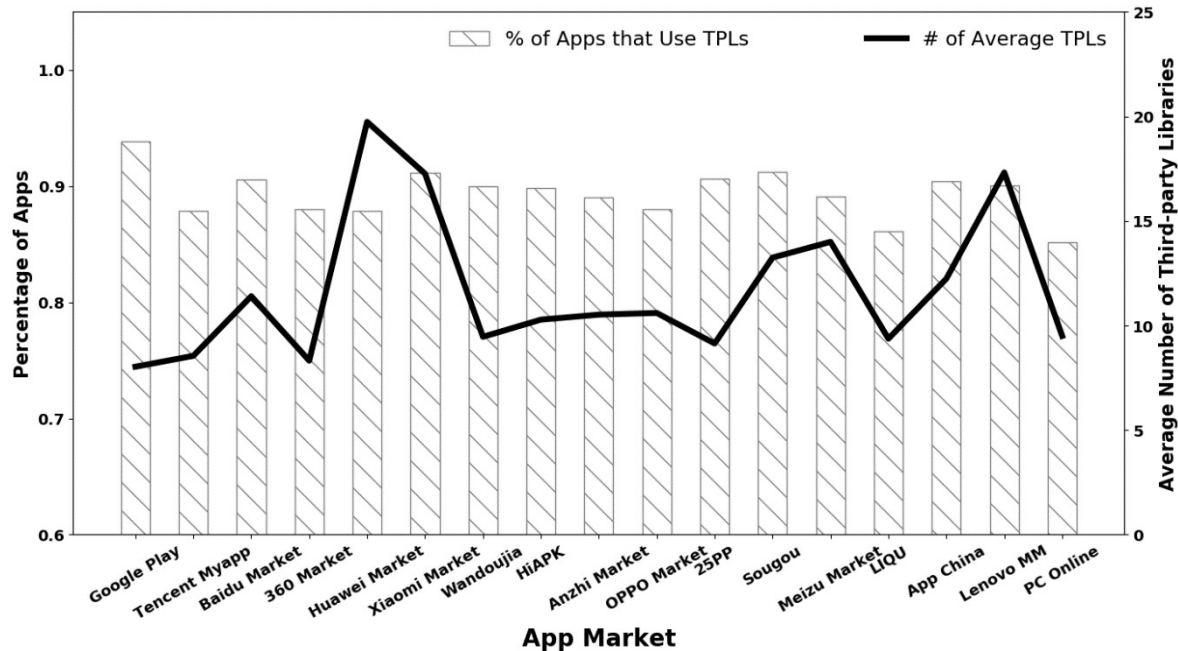
Distribution of App Release/Update Dates

- API levels 7-9 (i.e., Android 2.1.x to 2.3.2) are the most widely supported minimum API levels.
- Percentage of apps in Chinese markets supporting low API levels is 3x higher than that of Google Play

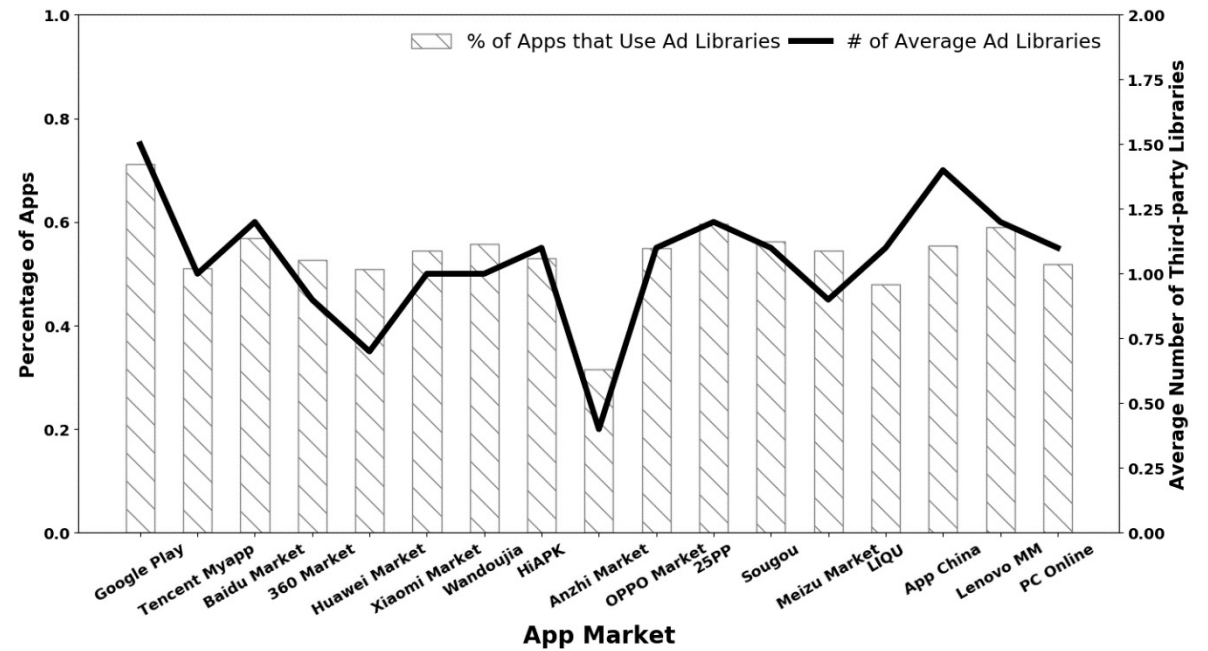
- 5% (Chinese) vs 23% (GPlay) apps updated in the 6 months prior to our crawling

General Overview — Third-party Libraries

- **Approach:** clustering-based approach [LibRadar, 2016] to the 6 million apps collected
 - build a new and complete feature dataset of TPLs
 - 5,102 libraries with 672,052 different versions



(1) Third-party Libraries



(2) Advertisement Libraries

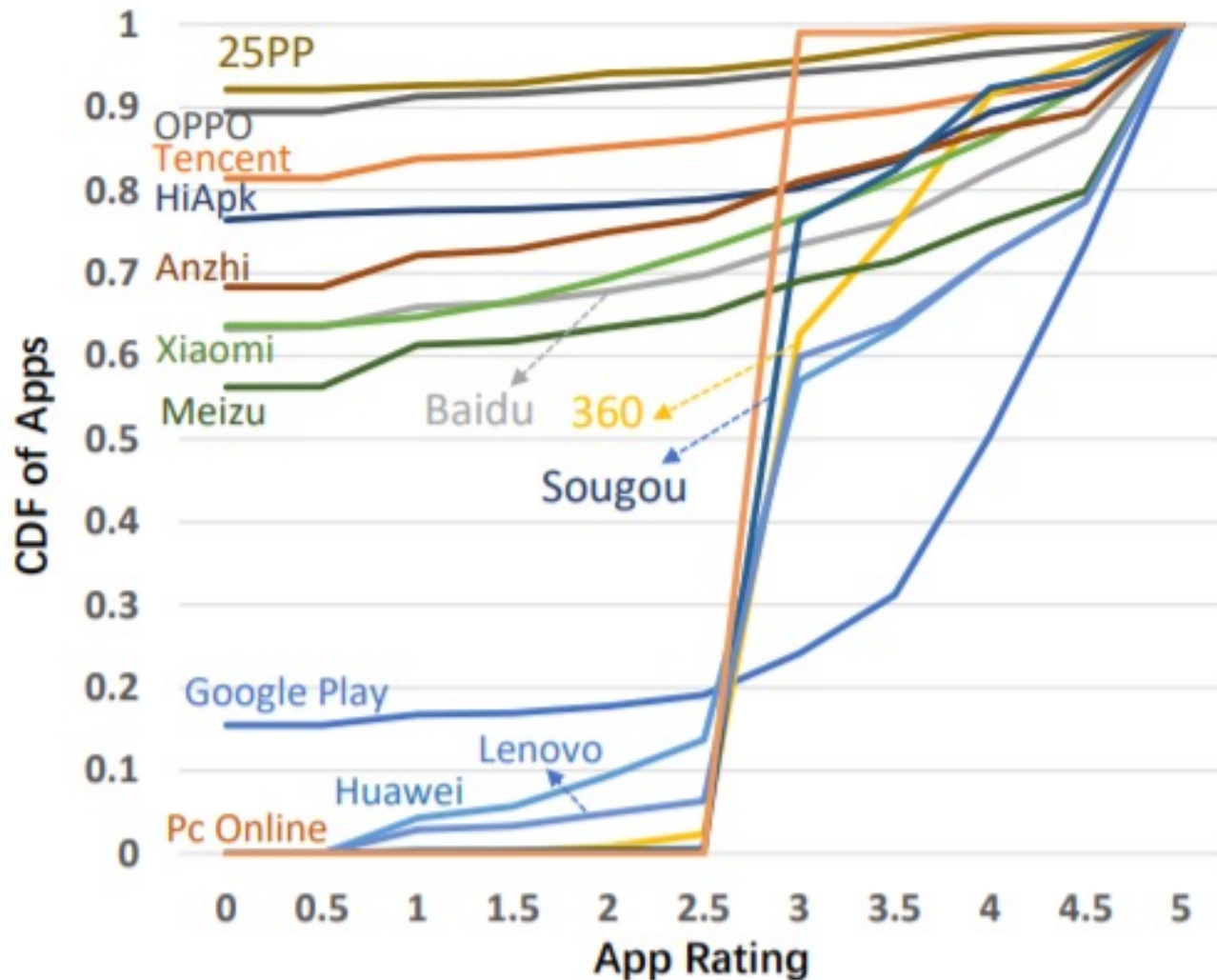
General Overview — Third-party Libraries

Top 10

Google Play		
Package Name	Type	Usage (%)
com.google.android.gms	Development	66.1
com.google.ads	Advertisement	62.1
com.facebook	Social Networking	21.5
org.apache	Development	20.5
com.squareup	Payment	13.8
com.google.gson	Development	12.9
com.android.vending	Payment	12.5
com.unity3d	Game Engine	11.8
org.fmod	Game Engine	9.6
com.google.firebase	Development	9.0
Chinese Markets		
com.google.ads	Advertisement	25.7
org.apache	Development	24.1
com.google.android.gms	Development	20.5
com.tencent.mm	Social Networking	17.3
com.baidu	Development, Map	16.9
com.umeng	Analytics, Advertisement	16.5
com.google.gson	Development	16.3
com.alipay	Payment	11.0
com.facebook	Social Networking	10.7
com.nostra13	Development	10.6

- Presence of Google-related ad libs because
 - some markets crawl apps from GPlay
 - Developers target GPlay and Chinese markets with the same app
- 282 specific ad libraries
 - More fragmented/competing market than GPlay

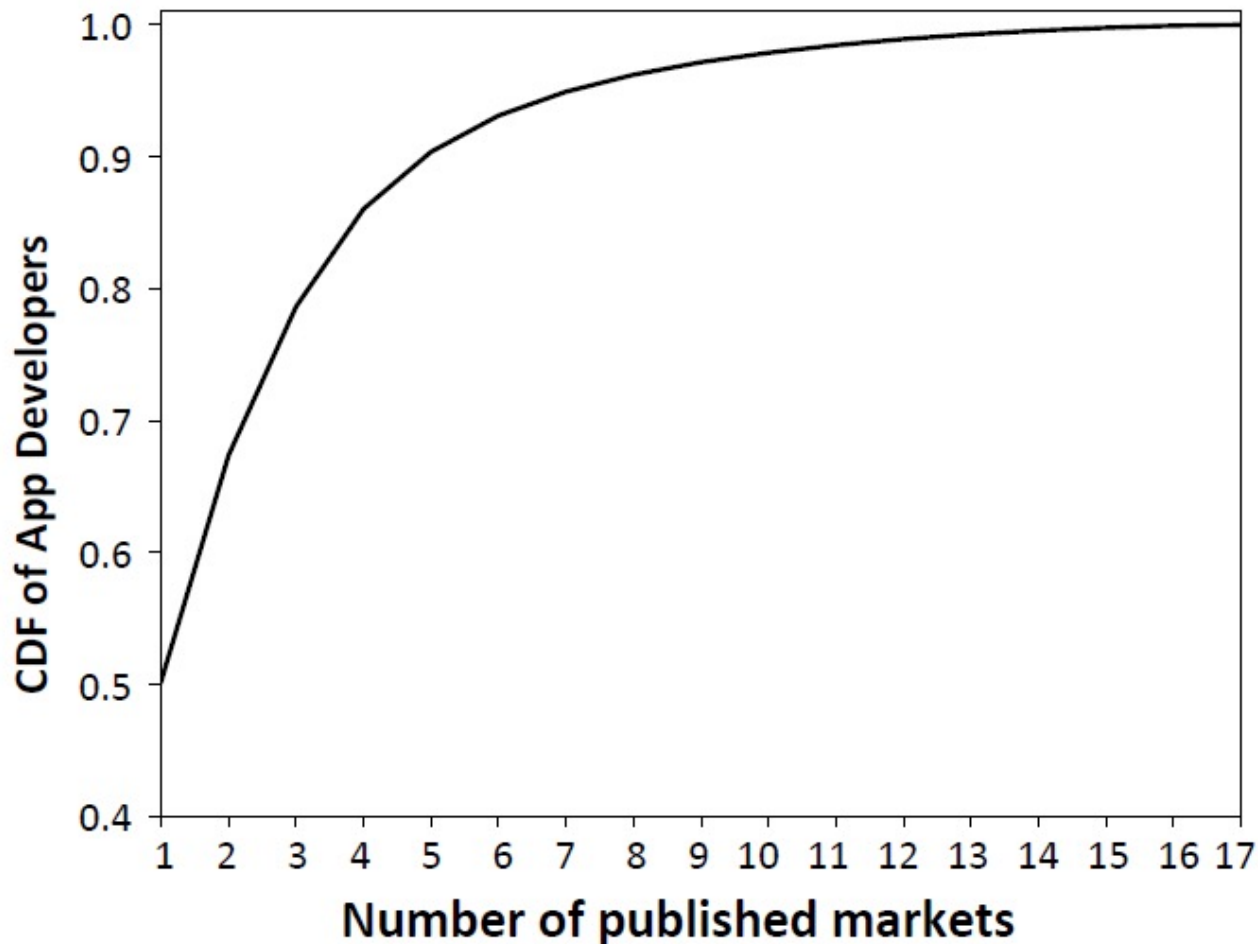
General Overview — App Ratings



- **Pattern #1:** More than 80% of apps in Chinese third-party app markets do not receive user ratings at all
- **Pattern #2:** Several markets use a default rating of 3, instead of a default rating 0
- Google Play presents a pattern completely different to that of any Chinese app market.

Publishing Dynamics — App Developers

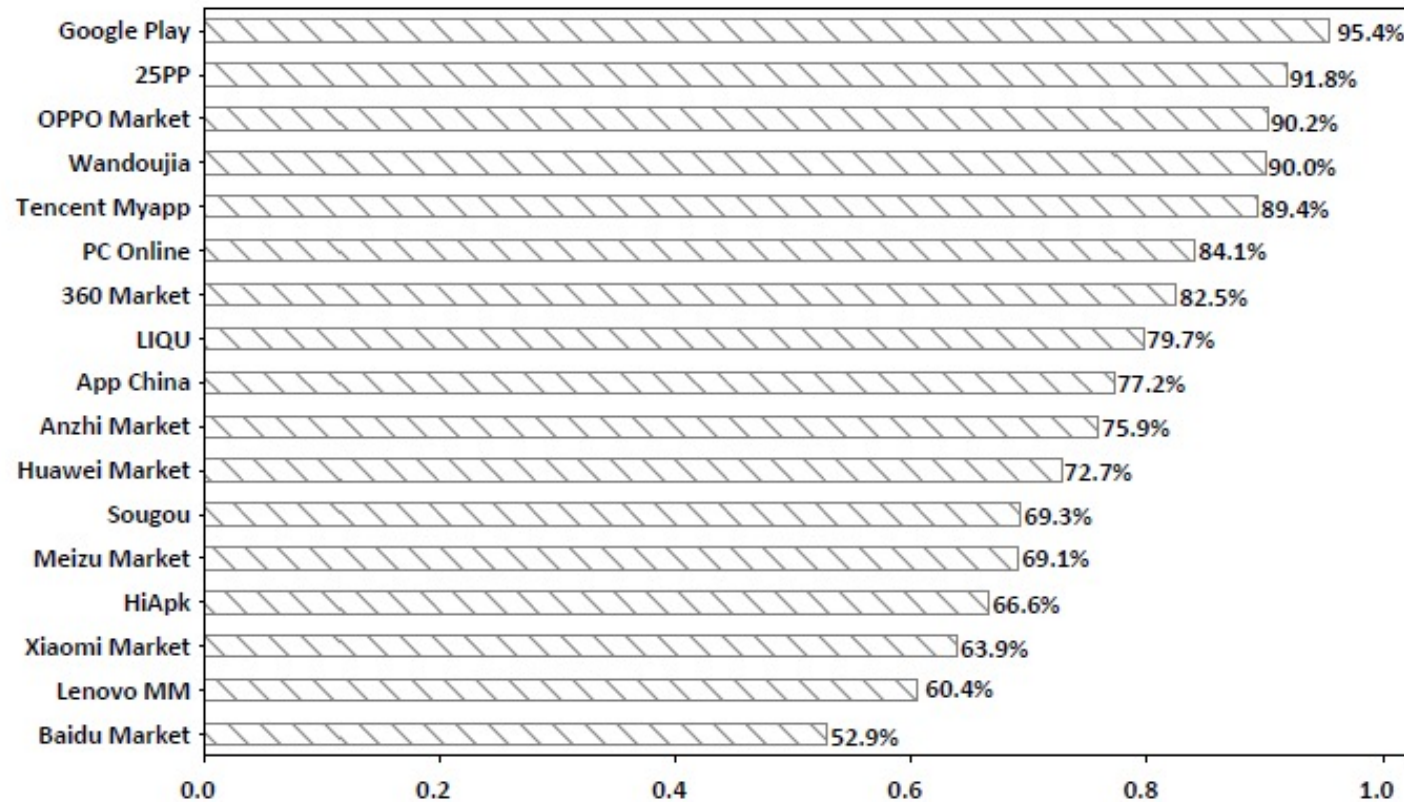
■ Over 1 million app developers



- More than half of the developers release their apps in Google Play
- Around 48% of developers focus solely on Chinese markets
- Around 20% of the app developers publish apps in more than 3 app stores
- Over 10% of the developers target exclusively one single Chinese store

Publishing Dynamics — App Updates

- Multistore released apps are unsynchronized roll-outs of new app versions across stores



A comparison of app updates across markets

Developer Misbehaviors – Fake and Cloned Apps (1/2)

- **Clustering-based method to efficiently identify fake apps**
 - build a cluster enforcing a strict matching of app names
 - heuristic rule to remove legitimate clusters
- **App Clone Detection**
 - *Signature-based clones*: Same package name, different signature
 - *Code-based Clones*
 - Implemented based on WuKong [ISSTA 2015]
 - Two-phase detection combining coarse-grained and fine-grained detection techniques

Developer Misbehaviors – Fake and Cloned Apps (2/2)

Market	Fake (%)	Clones	
		SB (%)	CB (%)
Google Play	0.03	4.01	17.82
Tencent Myapp	0.53	8.24	22.73
Baidu Market	0.48	10.98	17.38
360 Market	0.50	5.43	23.26
Huawei Market	0.33	11.54	18.76
Xiaomi Market	0.0	8.00	20.11
Wandoujia	0.39	5.98	21.23
HiApk	0.64	7.51	20.08
AnZhi Market	0.57	4.92	20.71
OPPO Market	0.38	5.85	20.94
25PP	0.35	7.16	24.08
Sougou	1.83	4.86	18.28
MeiZu Market	1.14	6.65	18.42
LIQU	0.40	5.32	16.68
App China	0.0	10.17	13.23
Lenovo MM	0.67	7.81	16.37
PC Online	1.89	8.60	23.34
Average	0.60	7.24	19.61

- Fake apps are present in nearly all app stores, including Google Play
- Code-based clones (~20%) are generally more common than signature-based clones (~10%)
- Source market of app clones
 - Google Play is the premier source for cloning apps
 - Market 25PP has the largest number of cloned apps

Developer Misbehaviors – Malware Prevalence (1/2)

Market	AV-rank (% apps)		
	>= 1	>= 10	>= 20
Google Play	17.03	2.09	0.32
Tencent Myapp	34.15	11.16	3.45
Baidu Market	42.77	12.24	3.30
360 Market	41.40	12.35	3.10
OPPO Market	42.97	16.43	6.00
Xiaomi Market	55.11	9.12	1.82
MeiZu Market	51.40	10.70	3.14
Huawei Market	57.48	4.71	0.57
Lenovo MM	54.20	7.53	1.52
25PP	32.36	8.26	2.06
Wandoujia	31.99	7.98	2.19
HiApk	41.89	11.12	2.72
AnZhi Market	55.32	11.37	2.41
LIQU	45.91	13.00	4.27
PC Online	55.93	24.01	8.37
Sougou	52.41	16.53	4.59
App China	48.55	14.13	4.27
Average	36.49	12.30	3.69

- Upload all the apps to VirusTotal
- AV-Rank: how many engines flag a sample as malware
- ~2% of the apps in Google Play are labeled as malware
- Percentage of malware in Chinese markets is much higher
 - 11 out of the 16 Chinese markets, the percentage of malware exceeds 10%

Developer Misbehaviors – Malware Prevalence (2/2)

- Only 38.3% of these malware samples are repackaged apps
 - app repackaging is no longer the main way for malware spreading
- Malware Families
 - used AVClass to obtain the family name of each identified malware
 - Different family distribution in Chinese markets and Google Play
 - most popular malware family in Chinese markets is kuguo (12.69%), while it only corresponds to 0.6% of malware in Google Play.
 - ~45% of the malware present in Google Play belong to the family airpush (29.04%) and revmob (15.09%)

Post Analysis

- Performed a second crawl for each market 8 months later

Market	%Malware Removed	#Overlapped with GPRM	%Removed
Google Play	84%	-	-
Tencent Myapp	8.75%	7,157	3.1%
Baidu Market	23.99%	1,422	34.53%
360	43%	1,198	34.22%
Xiaomi	32.50%	636	31.13%
Meizu	29.18%	668	26.20%
Huawei	26.92%	169	23.08%
Lenovo MM	22.75%	263	16.35%
25PP	19.63%	7,804	17.31%
Wandoujia	34.51%	5,289	44.74%
AnZhi	27.61%	632	25.78%
LIQU	14.08%	1,878	11.18%
PC Online	0.01%	1,117	0.00%
Sougou	24.24%	1,082	22.00%
App China	20.51%	546	30.24%

- ~84% of the malware found in GPlay have been removed
- the percentages of malware removal in Chinese markets vary from 0.01% to 34.51%
- apps with an AV-rank ≥ 10 removed from Google Play
 - over 70% of them are still hosted by at least one Chinese market

Conclusion and Take-away Points

- A few similarities and lots of differences between GPlay and Chinese app markets
- Heterogeneous behaviors across markets
- Higher prevalence of various types of malicious behaviors
 - Fake apps, app clones, over-privileged, malware
 - Google Play clearly outperforms Chinese markets
- Some markets claim to perform manual inspection, but our empirical observations seem to contradict this

Dataset and Tools

- **LibRadar++**
 - a clustering-based third-party library detection tool
- **5,000 labeled Android malware samples**
 - package name, MD5, app markets, AV-rank, and its malware family



<http://market.orangeapk.com/>

Thank you!