

Understanding Worldwide Private Information Collection on Android

Content

Overview	1
Data	2
Findings	3
Discussion	4

Authors

Yun Shen	NortonLifeLock Research Group
Pierre-Antoine Vervier	NortonLifeLock Research Group
Gianluca Stringhini	Boston University

Presented by William Que

Overview



Overview

As mobile devices become more critical in the computing experience of users, the potential threats of private information collection (PIC) increase.



● Privacy leaked ● Privacy not leaked



Pervasiveness

How many apps collect private information?



Data

What kind of data is collected?



Actor

Who collects the data?



Geolocation

Where does the data end up?

Data



Workflow

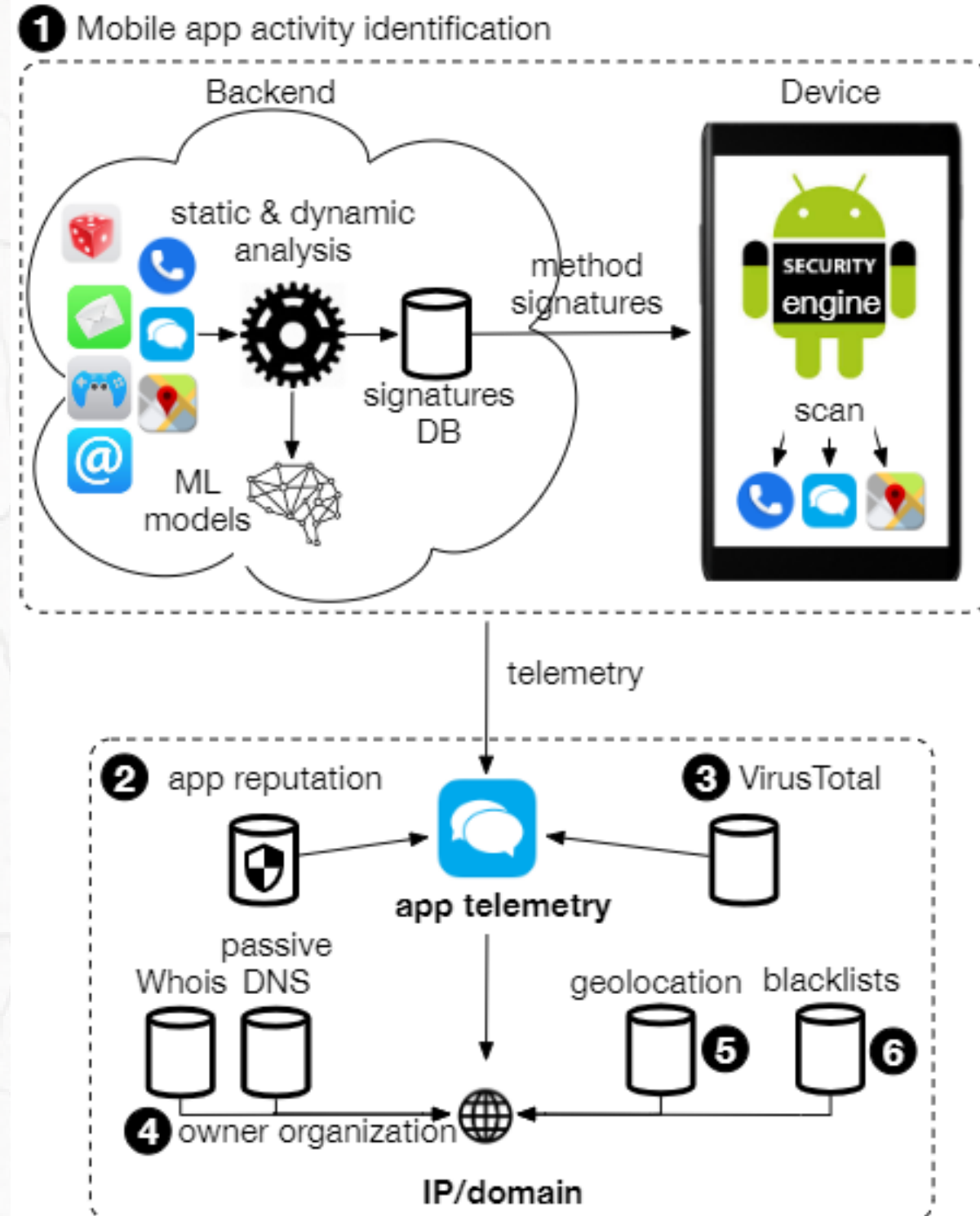


Fig. 1: Workflow of our measurement study.

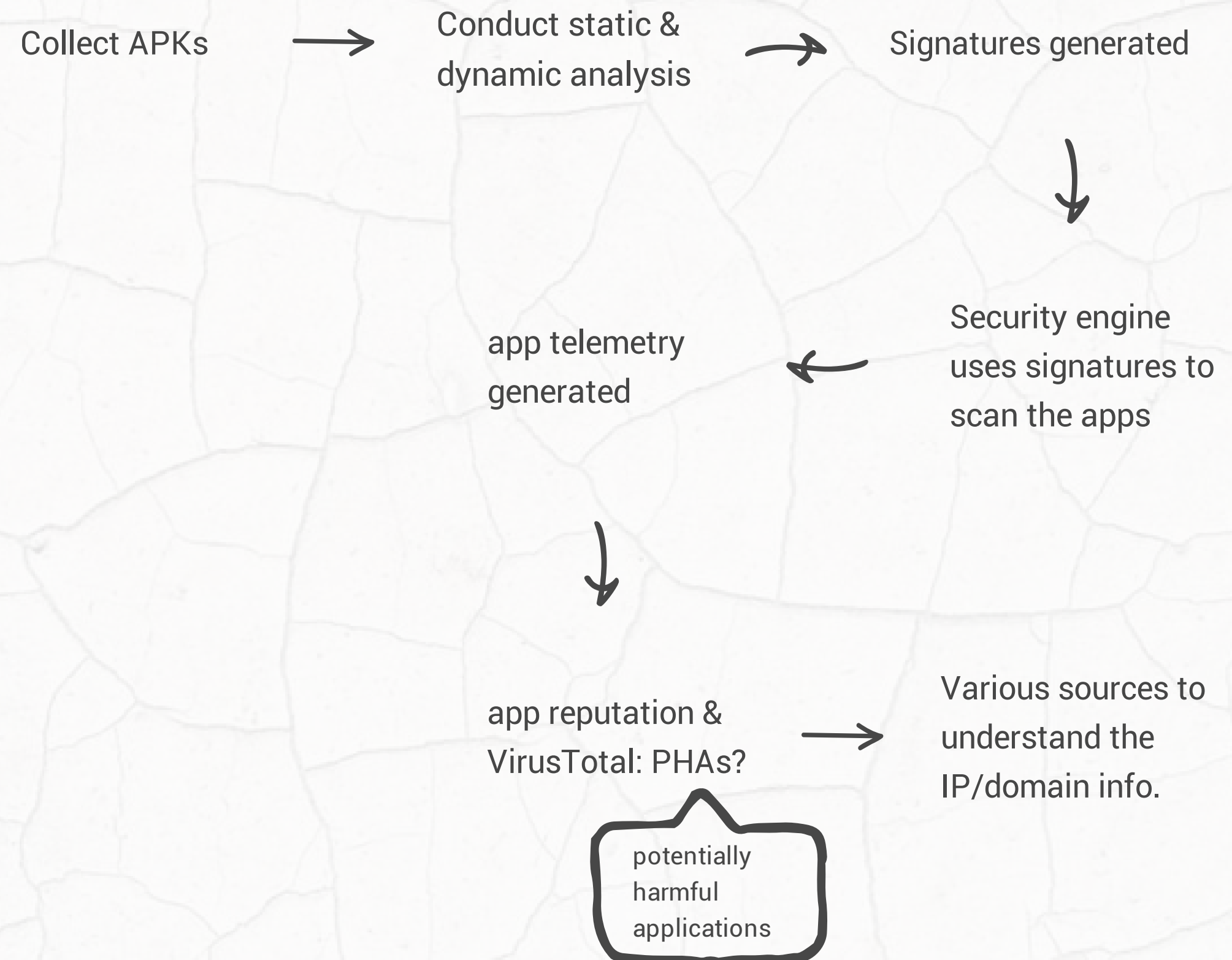


TABLE II: The 22 types of private information monitored by the security app.

Group	Category	Description	Previously studied or novel
Tracking	Phone number	Phone number	[36], [58]
	Device info	IMEI, OS/kernel version, phone producer, phone model	[36], [58], [61]
	SIM card info	Information about SIM serial number, IMSI, voicemail number	[58]
	Location info	GPS or cell tower coordinates	[36], [61]
	Operator info	Information about the network operator	✓
	Setting info	Information about the device configurations	✓
Activity and social profiling	Account info	Details about the configured accounts can be exported (including user names of entries under Settings/Accounts)	[36], [61] (partially)
	Email info	Details about the email address such as Gmail address can be exported	[36]
	Contact info	Contact list can be exported	[61]
	Social network account	Details about the social network accounts such as Facebook account can be exported	✓
	Voice mail account	Details about the voice mail accounts can be exported	✓
	Call log	Call log can be exported	✓
	SMS info	App can send the content or sender/recipient details from SMS/MMS messages	✓
	Calendar info	Calendar can be exported	✓
Usage preference	Installed app info	Details about apps installed on the phone are/can be exported (full or partial list of installed package names, or app titles)	✓
	Running app info	Details about apps running at a certain time are/can be exported	✓
	Browser history info	Browser history can be exported	✓
	Browser bookmark info	Browser bookmarks can be exported	✓
Audio/Video	Audio info	Recorded audio clips can be exported (e.g., recorded by the app, or picked from saved)	[52]
	Photo info	Photo can be exported	✓
	Video info	Video can be exported	[52]
	Camera info	App can take pictures or picks them from gallery and exports them	✓

Dataset

TABLE I: Summary of datasets used.

Dataset	Data	Count
Mobile app activity log (01/2018 - 09/2019)	Total records	6B
	Days	634
	Countries and regions	201
	Devices	17.3M
	Distinct app names	2.13M
	Distinct app SHA2s	6.5M
	Distinct PIC FQDNs	76,451
	Distinct PIC domains	40,851
Mobile app reputation log	Low reputation SHA2s	3.4M
VT	Total reports	6.5M
	PHA SHA2s (detections \geq 6)	3.5M
	Benign SHA2s (no detection)	2.3M
	Not found SHA2s	401K
Domain to owner org. (01/2018 - 09/2019)	Domains	10,736
	Organizations	9,593
Blacklists (01/2018 - 09/2019)	Domains/IPs	7,670
Geolocation (01/2018 - 09/2019)	Domains/IPs	40,851

Actor

☒

Geolocation

☒

Findings



PIC domains: App Presence

PIC organizations generally benefit from collecting data about more users.

To achieve that, PIC organizations tend to increase their presence in mobile apps to reach out to more users.

These domains are attributed to three functions: Metrics/Analytics (M), Advertising (A), and Development (D).

The majority of these PIC domains (15 out of 20) offer advertising services.

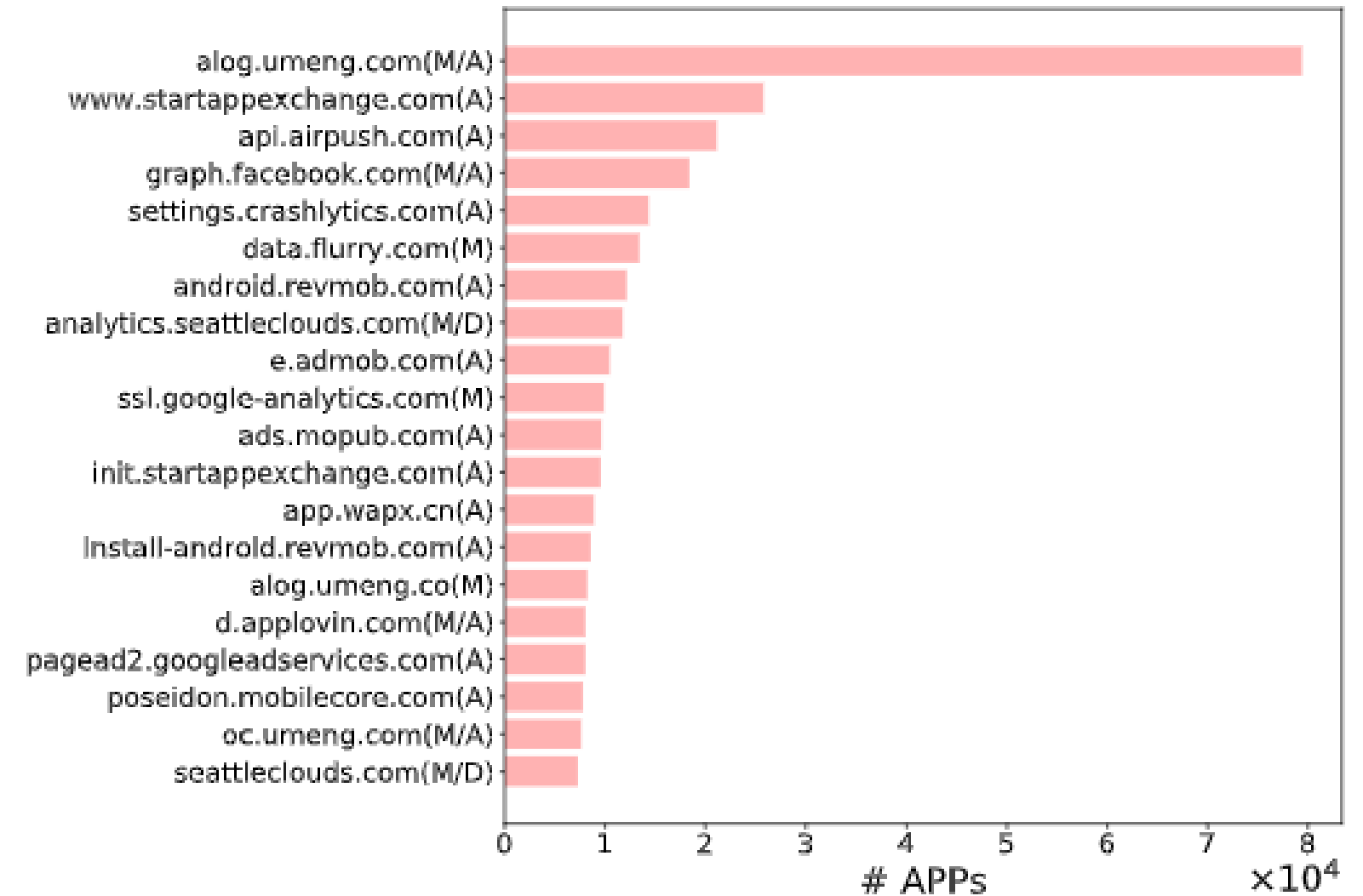


Fig. 3: Global top 20 PIC domains ranked by app presence. Domain's primary function- **M**: Metrics/Analytics, **A**: Advertising, and **D**: Development.

PIC domains: Device Penetration

- Higher app presence rate does not necessarily lead to higher device penetration.
- It is linked to the amount of actual information an app is able to access from the device.
- Looking at device penetration provides different results than looking at apps only.
- Designing measurement studies focused on executing apps could lead to conclusions that are biased.

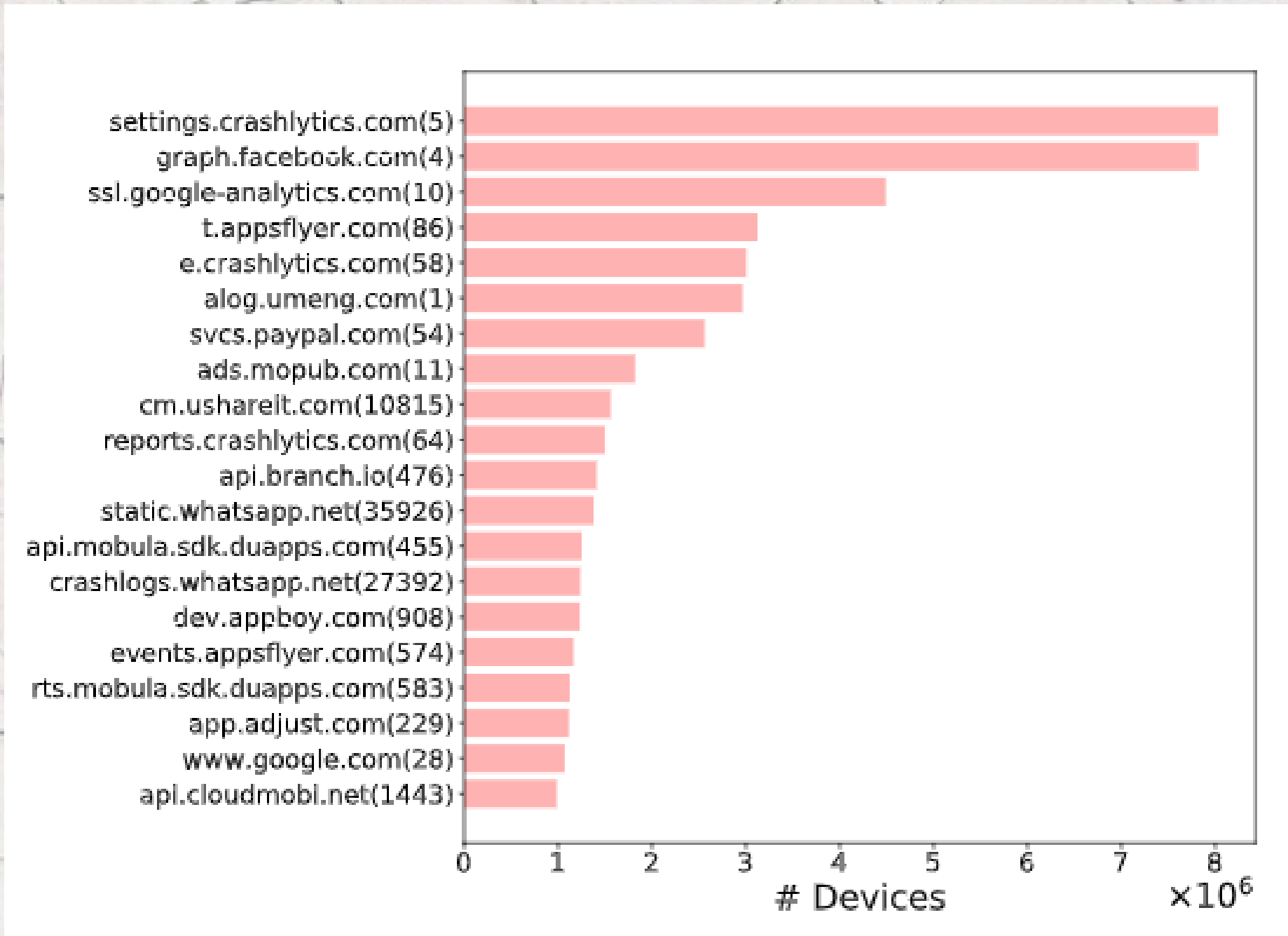
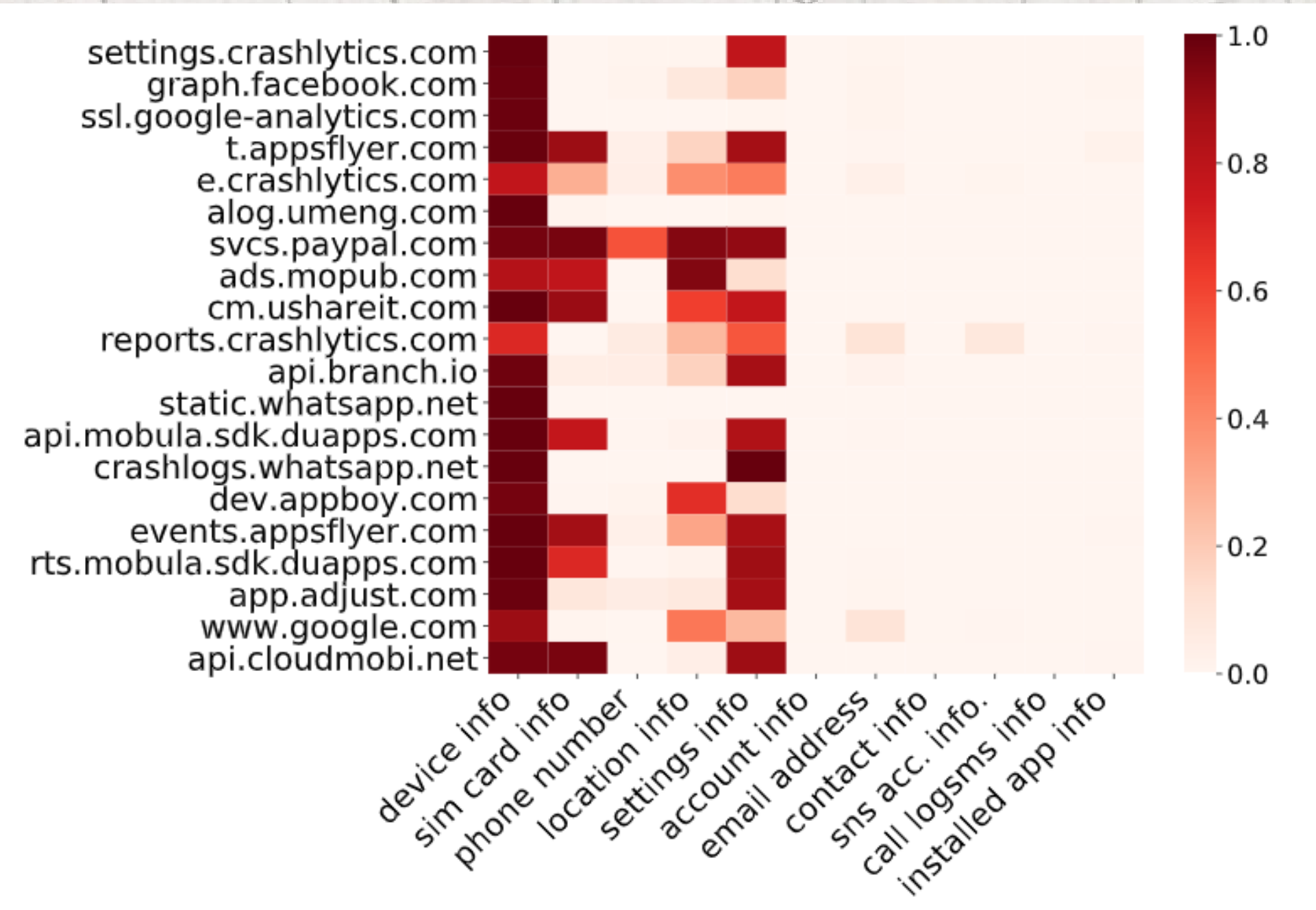


Fig. 5: Top 20 PIC domains ranked by device penetration rate. The number next to a PIC domain represents its ranking by app presence.

Types of information collected

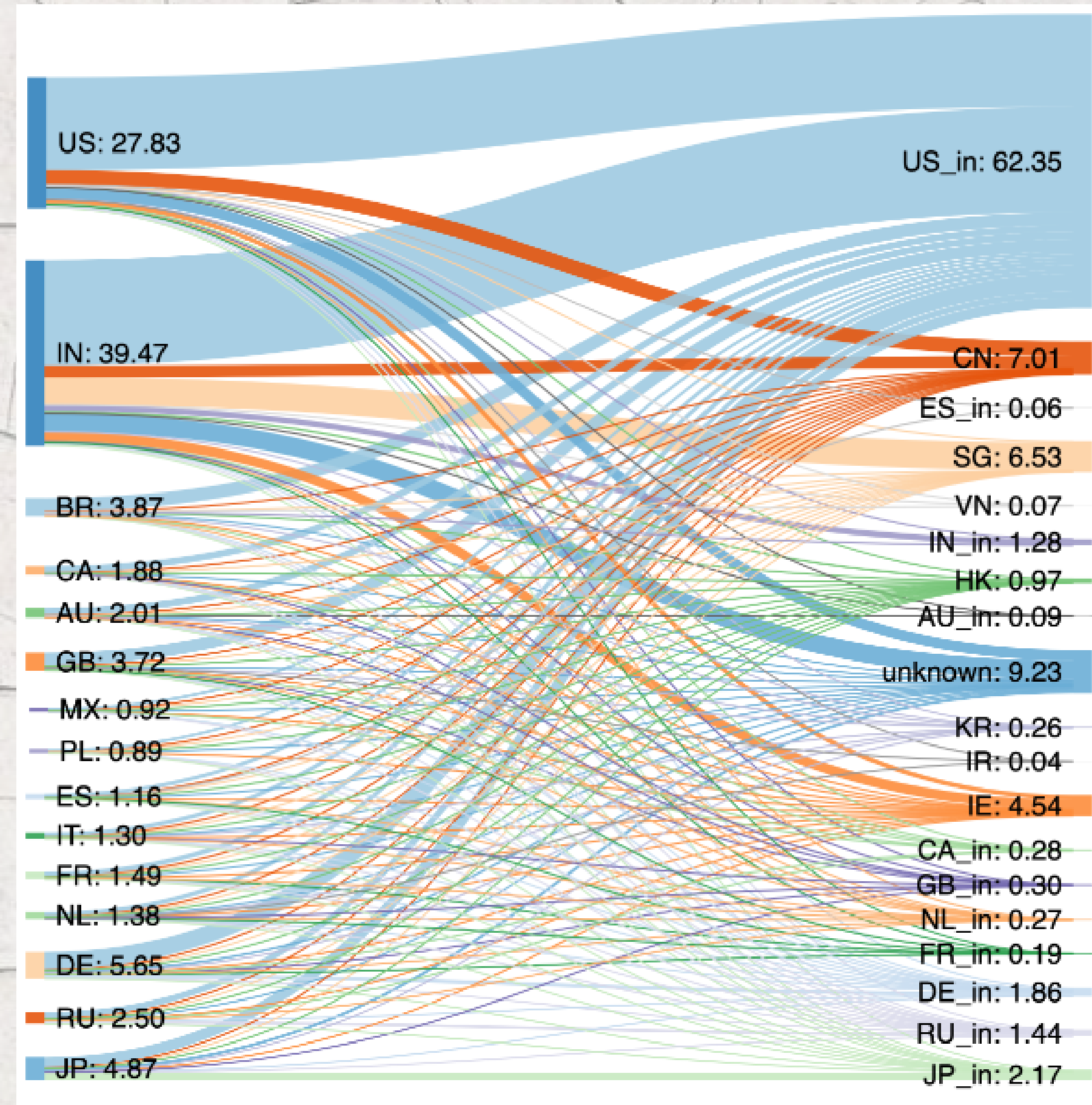
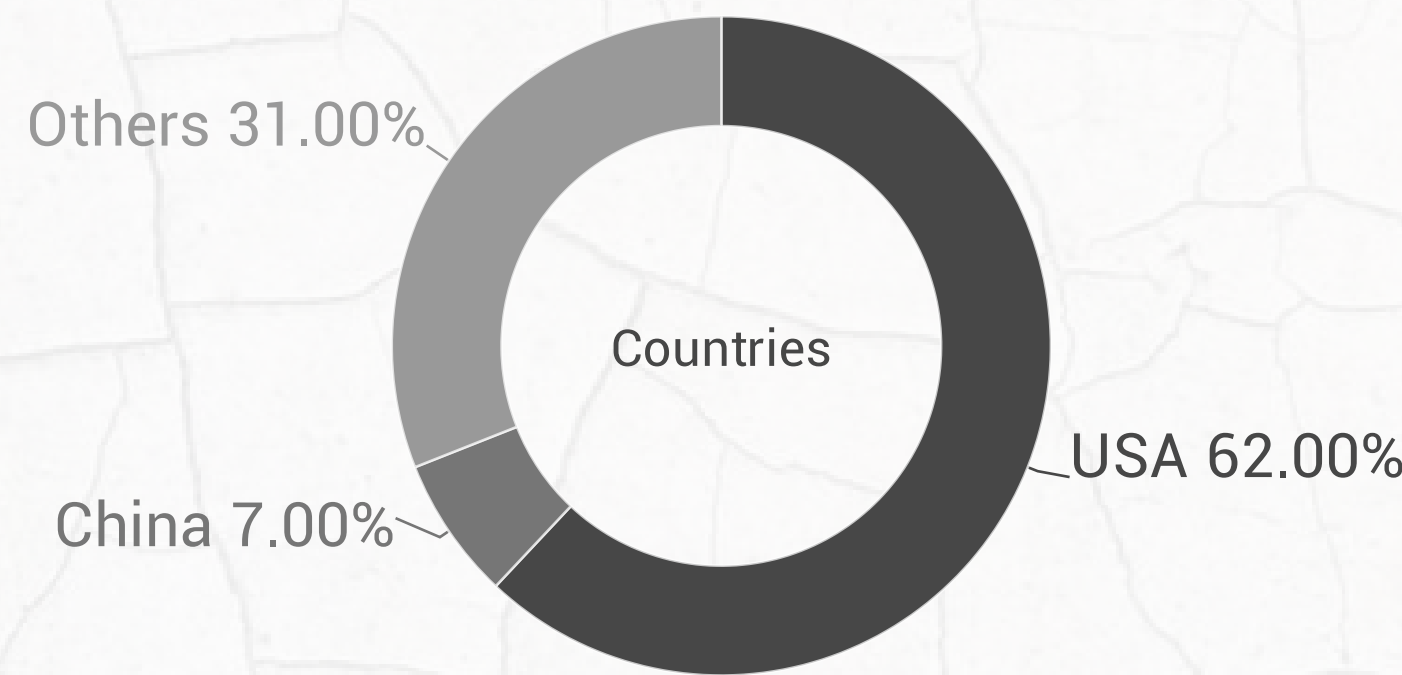
1. **Device info**
2. **SIM card info**
3. **Location info**
4. **Setting info**

It potentially enabling them to track the end users more systematically



Global private information flow

- PIC domains hosted in the United States collect 62% (of which 42.3% coming from out of the country) of global private information
- PIC domains hosted in China collect 7% of private information flows from 4.59M devices globally.



Discussion

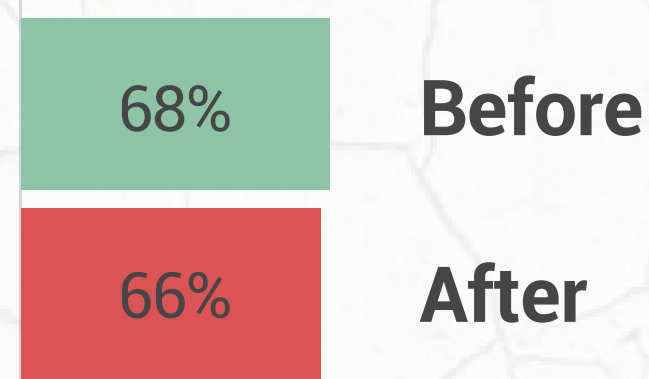


Implications for Policymakers

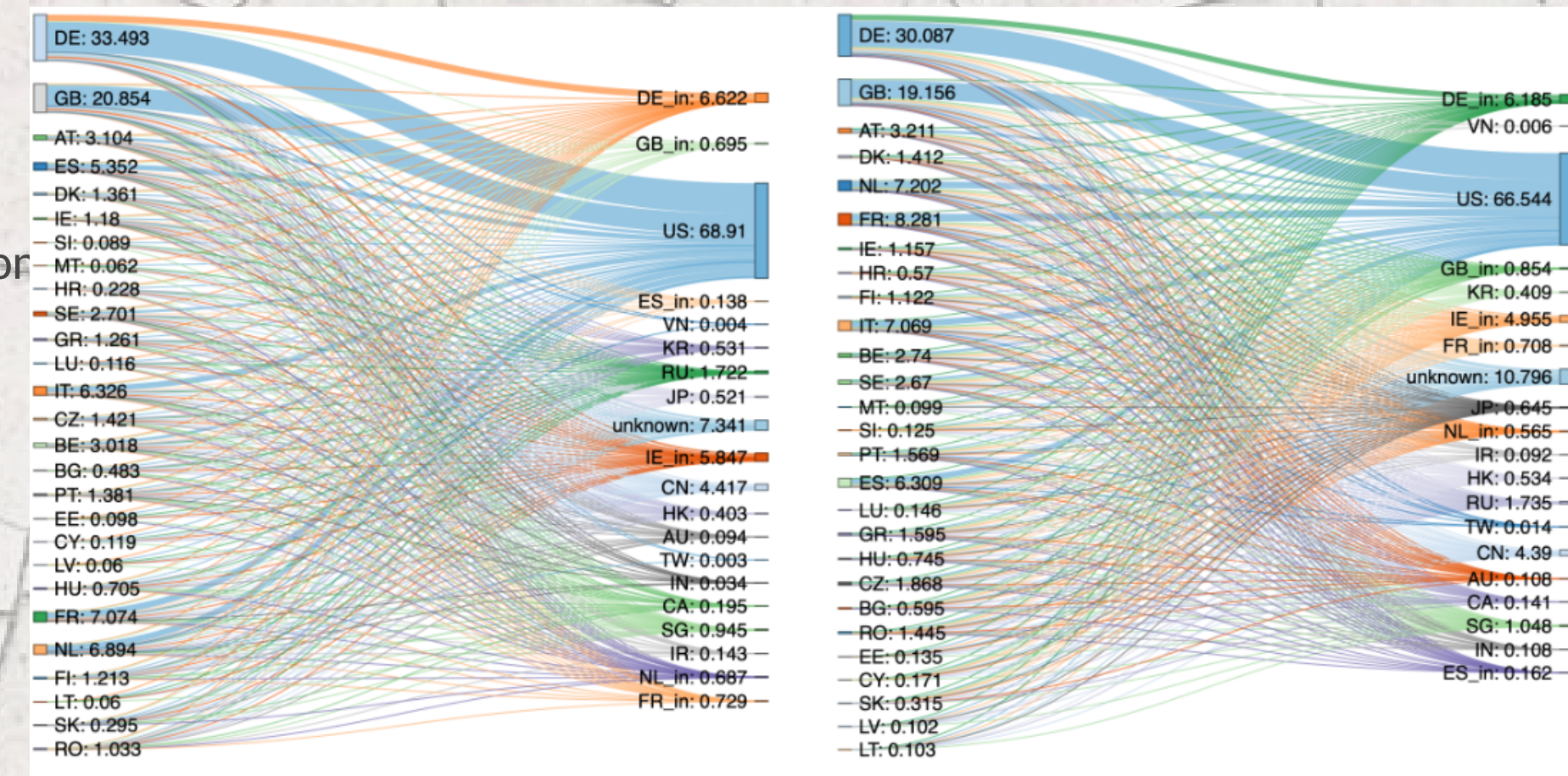
The European Union's (EU) General Data Protection Regulation (GDPR) entered into effect on May 25th, 2018.

It imposes obligations onto organizations in any country as long as they target or collect data related to people in EU countries

PIC domains hosted in the United States dominate the private information collection in the EU.



% of the US hosted PIC domains in EU before, and after GDPR



Questions?

