CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android Apps

Trung Tin Nguyen, Michael Backes, Ninja Marnau, and Ben Stock
CISPA Helmholtz Center for Information Security

WHAT IS GDPR?

General Data Protection Regulation (GDPR) governs all processing of personal data related to individuals situated in the EU and EEA

May 2018

Name

Phone Number

Email

IP Address or Location, Unique Identifier

User Personal Data

DETECTING VIOLATIONS OF GDPR

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

In mobile apps, researchers have analyzed the app privacy policies to identify legislation violations, i.e., determining whether an app's behavior is consistent with the privacy policy

App Behavior

```
@Override
public boolean onOptionsItemSelected(MenuItem item) {
    switch (item.getItemId()) {
        case R.id.menu_location:
            // ...
            break;
        case R.id.menu_fare:
            TelephonyManager manager = (TelephonyManager)t
            String phoneNumber = manager.getLine1Number();
            // ...
            break;
        case R.id.menu_refresh:
            // ...
            break;
    }

    LocationManager manager = (LocationManager) this.getSy
    manager.requestLocationUpdates(LocationManager.NETWORK
            minTime: 0,  minDistance: 0, locationListener);
```

analyzing the app code

≠

Privacy Policy

What Personal Data Do We Receive?

Personal data is any information from or about an identified or identifiable person, including information that Zoom can associate with an indivi
person. We may collect, or process on behalf of our customers, the following categories of personal data when you use or interact with Zoom P

- **Account Information:** Information associated with an account that licenses Zoom Products, which may include administrator name, conta

How Do We Use Personal Data?

Zoom employees do not access meeting, webinar, or messaging content (specifically, audio, video, files, and messages) unles
owner, or as required for legal, safety, or security reasons, as discussed below. Zoom uses personal data to conduct the follo

- **Provide Zoom Products and Services:** To provide Products, features, and services to account owners, their users, and
meetings and webinars hosted on their accounts, including to customize Product features and recommendations for
Zoom also uses personal data, including contact information, to route invitations and messages to recipients when pe
messages using
and messages,
account owner
- **Product Resea**
background filt
- **Marketing and**
based on your
and when you
features, and s
- **Authentication**
experiences, ac

How Do We Share Personal Data?

Zoom provides personal data to third parties only with consent or in one of the following circumstances (s
under applicable law):

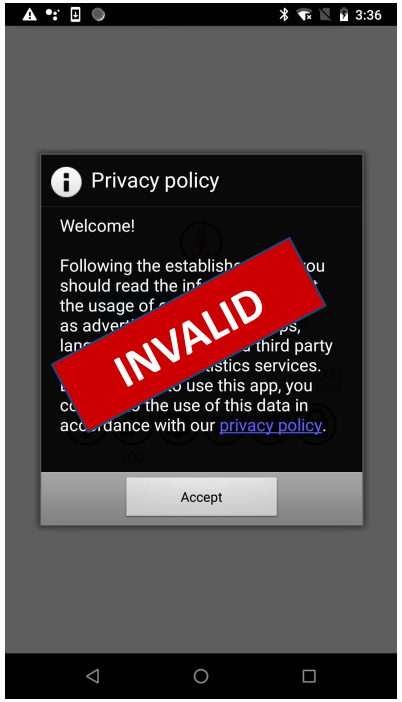
- **Resellers:** If an account owner licensed Zoom from a third-party reseller of Zoom Products, the rese
content for users, including meetings, webinars, and messages hosted by the account owner.
- **Vendors:** Zoom works with third-party service providers to provide, support, and improve Zoom Pro
also work with third-party service providers to provide advertisements and business analytics regar
access personal data subject to contractual and technical requirements for protecting personal data
data for any purpose other than to provide services to Zoom or as required by law.
- **For Legal Reasons:** Zoom may share personal data as needed to: (1) comply with applicable law or r
legal process and proceedings, including from law enforcement or government agencies; (2) enforce
of Service or policies; (3) detect, prevent, or investigate potential fraud, abuse, or safety and securit
meet our corporate and social responsibility commitments; (5) protect our and our customers' right

The GDPR knows several legal justifications for processing of personal data.

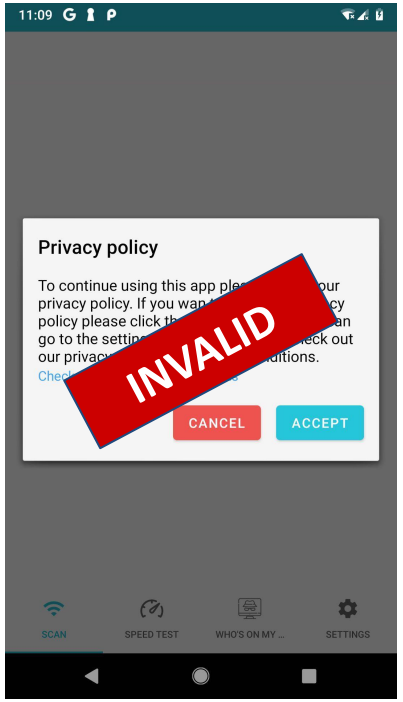
In the case of transferring data to third party data controller for advertising purposes explicit consent is the viable option of these justifications

GDPR CONSENT REQUIREMENTS

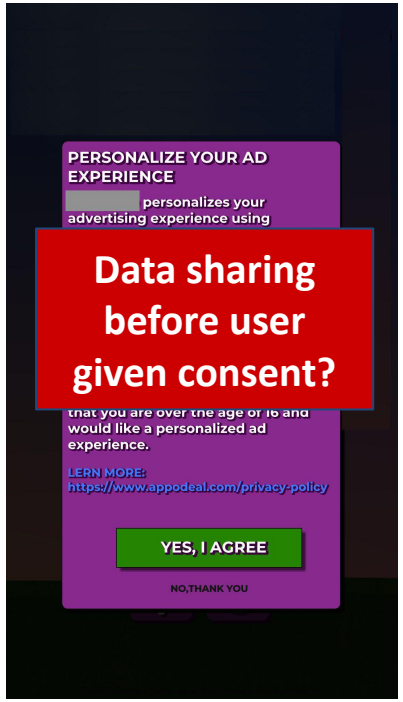
The GDPR requires the consent to be ***freely given***, ***specific***, ***informed***, and ***unambiguous***.



(a)



(b)



(c)

Personal data transfer must only occur after the user has actively agreed (e.g., by clicking accept), i.e., “consent” packaged in terms and conditions or privacy policies is not compliant

MOTIVATION AND RESEARCH QUESTIONS

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

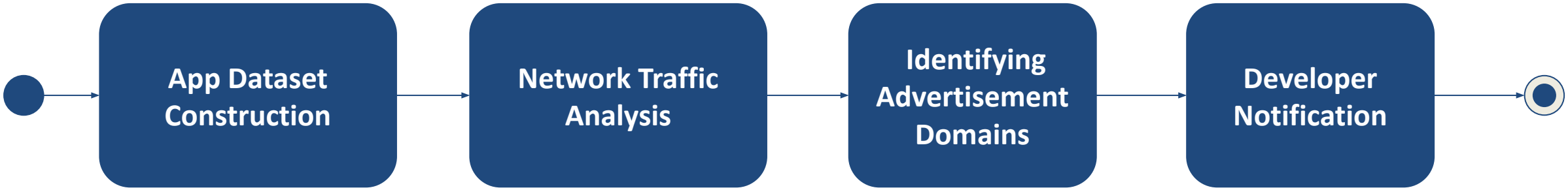
The community lacks insight into such GDPR violations in the mobile ecosystem.

Our research aims at answering the following research questions:

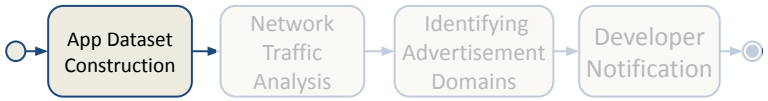
- **RQ1**: How many apps send out personal data without any prior consent?
- **RQ2**: Of the apps which send out any data, how many send it towards parties that act as data controllers under the GDPR?
- **RQ3**: Are developers aware of the requirements of GDPR and the issues that might arise from not following the outlined laws?

METHODOLOGY

Overview of the methodology to identify violations of GDPR's explicit consent in Android apps



APP DATASET CONSTRUCTION



Aiming to assess the state of GDPR violations in both high-profile and long-tail apps on the Play Store, and to understand if the violations are specific to either of them

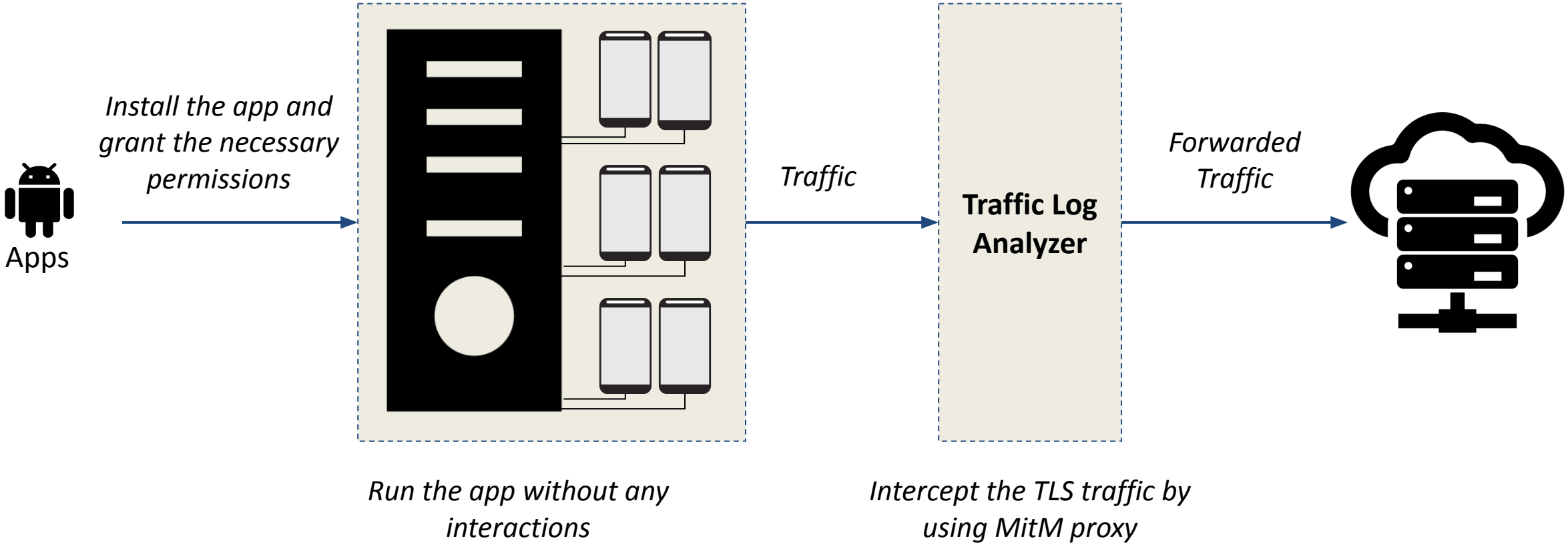
High-profile app dataset

16,163 top free high-profile apps from 33 app categories (i.e., AppBrain statistic).

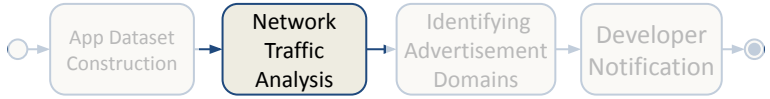
Long-tail app dataset

70,000 distinct apps with at least 10,000 downloads and excluded those in the high-profile set.

NETWORK TRAFFIC ANALYSIS



PERSONAL DATA TIED TO A PHONE

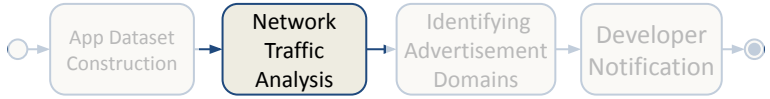


CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Data Type	Description
AAID	Android Advertising ID
BSSID	Router MAC addresses of nearby hotspots
Email	Email address of phone owner
GPS	User location
IMEI	Mobile phone equipment ID
IMSI	SIM card ID

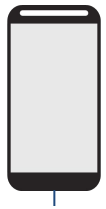
Data Type	Description
MAC	MAC address of WiFi interface
PHONE	Mobile phone's number
SIM_SERIAL	SIM card ID
SERIAL	Phone hardware ID (serial number)
SSID	Router SSIDs of nearby hotspots
GSF ID	Google Services Framework ID

STRING-MATCHING DEVICE-BOUND DATA



CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Using simple string-matching to identify personal data that is ***known***, common transformations such as upper/lower case, hashing (e.g., MD5), encoding (e.g., base64) are considered



Extracting

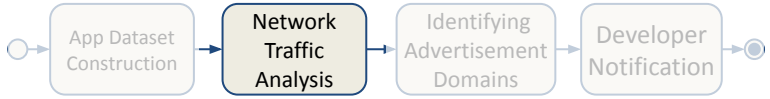
Advertising ID:
70831fd5-c2df-4b75-94bd-915a2046fe14

Searching

POST https://api.uca.cloud.unity3d.com HTTP/1.1

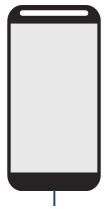
{"header": {"appid": "323c504d-fae5-449d-acd1-a89f2cf06b09", "userid": "8190a000-0b24-4f36-a981-c535f57ff164", "sessionid": 11219588307516230, "platform": "Android", "sdk_ver": "u5.3.3f1"}, "events": [{"type": "deviceInfo", "ts": 1607086918599, "make": "Android", "model": "Google Pixel 3a", "processor_type": "ARMv7 VFPv3 NEON", "system_memory_size": "3593", "engine_ver": "5.3.3f1", "app_name": "com.xxxx", "app_install_mode": "dev_release", "debug_build": false, "license_type": "personal", "os_ver": "Android OS 9 / API-28 (PQ3B.190801.002/5674421)", "deviceid": "cf9f2bb31b46f4871094b3217b8349a9", "app_ver": "1.0.21", "changed": ["app_ver", "os_ver", "sdk_ver"]}, {....}, {"type": "deviceInfo", "ts": 1607086919086, "adsid": **"70831fd5-c2df-4b75-94bd-915a2046fe14"**, "ads_tracking": false, "changed": ["adsid"]}]}

STRING-MATCHING DEVICE-BOUND DATA



CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Using simple string-matching to identify personal data that is ***known***, common transformations such as upper/lower case, hashing (e.g., MD5), encoding (e.g., base64) are considered



Extracting

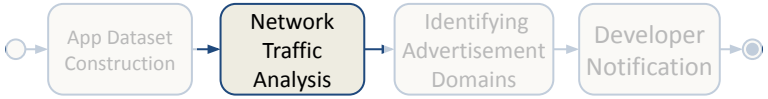
IMEI:
354787113965960
MD5 Hashing:
cf9f2bb31b46f4871094b3217b8349a9

Searching

POST https://api.uca.cloud.unity3d.com HTTP/1.1

{"header": {"appid": "323c504d-fae5-449d-acd1-a89f2cf06b09", "userid": "8190a000-0b24-4f36-a981-c535f57ff164", "sessionid": 11219588307516230, "platform": "Android", "sdk_ver": "u5.3.3f1"}, "events": [{"type": "deviceInfo", "ts": 1607086918599, "make": "Android", "model": "Google Pixel 3a", "processor_type": "ARMv7 VFPv3 NEON", "system_memory_size": "3593", "engine_ver": "5.3.3f1", "app_name": "com.xxxx", "app_install_mode": "dev_release", "debug_build": false, "license_type": "personal", "os_ver": "Android OS 9 / API-28 (PQ3B.190801.002/5674421)", "deviceid": "**cf9f2bb31b46f4871094b3217b8349a9**", "app_ver": "1.0.21", "changed": ["app_ver", "os_ver", "sdk_ver"]}, {....}, {"type": "deviceInfo", "ts": 1607086919086, "adsid": "70831fd5-c2df-4b75-94bd-915a2046fe14", "ads_tracking": false, "changed": ["adsid"]}]}

POTENTIALLY UNIQUE TRACKING IDENTIFIERS DETECTOR

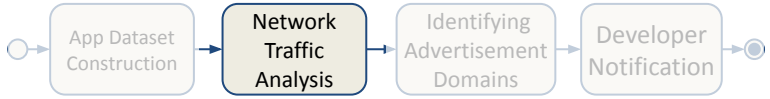


CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Performing multiple runs with a different set of devices to identify parameters that could be used to track and profile an individual, but do not obviously string-match

Domains	**Parameter**
appsflyer.com	deviceFingerPrintId=<UUID>
branch.io	hardware_id=6fd9a2e0f2721498
tapjoy.com	managed_device_id=tjid.36cec2b4196...
unity3d.com	common.deviceid=d3d55baf21d8f31839...

NETWORK TRAFFIC ANALYSIS RESULT

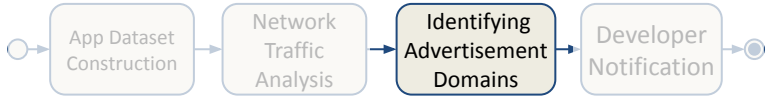


CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Third-party domains representing only 12.0% of domains which received personal data, are responsible for 94,7% of cases of receiving personal data without prior consent

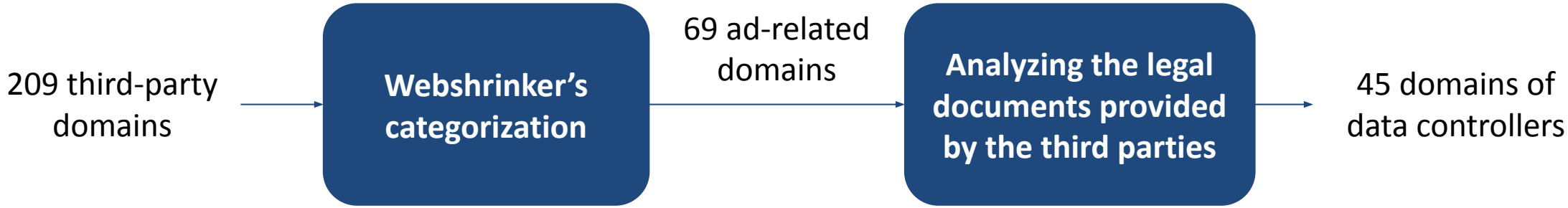
86,163 apps → **72,274 (83.9%) apps were successfully analyzed** → **28,665 (39.66%) sent personal data to 1,744 domain names (209 third-party domains)**

IDENTIFYING ADVERTISEMENT DOMAINS

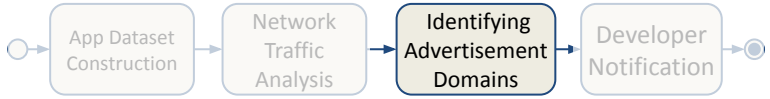


CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

An app which relies on external data controllers for targeted advertising needs to explicitly ask for the user's consent to share her personal data with the third party



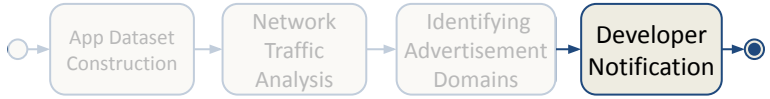
IN-DEPTH ANALYSIS OF VIOLATIONS



CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

- 24,838 (88.5% 28,065) apps sent personal data to ad-related domains, thereby violating GDPR's mandated consent.
  - More than half of the apps which sent data without consent sent data to (at least) Facebook, then Unity, and Flurry.
  - Library providers make it very cumbersome for developers to be compliant with GDPR.
- 3,840 apps that combined the AAID with some other type of personal information (e.g, IMEI).
  - Due to developers' opt-in or the usage of outdated libraries that do not support GDPR.
- The phenomenon of sending out personal data without prior explicit consent happens as frequently and with as many parties in both dataset.

DEVELOPER NOTIFICATIONS



CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

- 11,914 developers were notified (responsible for 17,795 apps).
- Until February 1, 2021: there are 2,083 apps accessed the notification reports.
  - 448 distinct developers that answered the survey.

GDPR issues are **widespread**, often **misunderstood**, and require effort from advertisement providers, app stores, and developers alike to mitigate the problems.

CALL TO ACTIONS

- Third Parties Should Take Responsibility
  - Limiting the data collection.
  - Providing the automatically consent mechanism.
  - Making their documentation transparent and easy to access.
- App Stores Should Take Actions
  - Employing such techniques as our to identify the potential violations of GDPR explicit consent, or the usage of outdated SDK.
- Support for Developers
  - Strongly call on third-party vendors for better documentation and transparency in legal documents.

CONCLUSION

MOTIVATION AND RESEARCH QUESTIONS

The community lacks insight into such GDPR violations in the mobile ecosystem.

Our research aims at answering the following research questions:

- **RQ1**: How many apps send out personal data without any prior consent?
- **RQ2**: Of the apps which send out any data, how many send it towards parties that act as data controllers under the GDPR?
- **RQ3**: Are developers aware of the requirements of GDPR and the issues that might arise from not following the outlined laws?

5

METHODOLOGY

Overview of the methodology to identify violations of GDPR's explicit consent in Android apps

6

IN-DEPTH ANALYSIS OF VIOLATIONS

- 24,838 (88.5% 28,065) apps sent personal data to ad-related domains, thereby violating GDPR's mandated consent.
  - Library providers make it very cumbersome for developers to be compliant with GDPR.
- 3,840 apps that combined the AAID with some other type of personal information.
  - Due to developers' opt-in or the usage of outdated libraries that do not support GDPR.
- The phenomenon of sending out personal data without prior explicit consent happens as frequently and with as many parties in both dataset.

16

DEVELOPER NOTIFICATIONS

- 11,914 developers were notified (responsible for 17,795 apps).
- Until February 1, 2021: there are 2,083 apps accessed the notification reports.
  - 448 distinct developers that answered the survey.

GDPR issues are **widespread**, often **misunderstood**, and require effort from advertisement providers, app stores, and developers alike to mitigate the problems.

15

CALL TO ACTIONS

- Third Parties Should Take Responsibility
  - Limiting the data collection.
  - Providing the automatically consent mechanism.
  - Making their documentation transparent and easy to access.
- App Stores Should Take Actions
  - Employing such techniques as our to identify the potential violations of GDPR explicit consent, or the usage of outdated SDK.
- Support for Developers
  - Strongly call on third-party vendors for better documentation and transparency in legal documents.

23


Thank you!
tin.nguyen@cispa.de