

# Reliability Analysis of Privacy Policies Based on Android Static Analysis

---

Yoonkyo Jung

Assistant Professor

Dept. of Computer Science, Republic of Korea Air Force Academy

# Outline

---

- Introduction
- System Design
- Result
- Conclusion

# Introduction

---

# Privacy Risk in Mobile

---

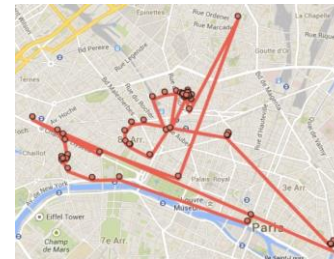
- With the increase of mobile users, privacy issues in the mobile environment have also grown
  - Many privacy threats have received attention from mobile users and regulators
- Mobile apps collect more sensitive data than the data web collects
  - Mobile device has more personal information that identifies user contexts



IMEI



Mobile Carrier



Location History

# Privacy Policy

---

- On the google play store, Google requires android developers to disclose how their apps collect, use, and share user data in privacy policies
  - App users can easily know how the apps collect data by checking the privacy policy

## User Data

You must be transparent in how you handle user data (e.g., information collected from or about a user, including device information). That means disclosing your app's access, collection, use, and sharing of the data, and limiting the use of the data to the purposes disclosed. In addition, if your app handles personal or sensitive user data, please also refer to the additional requirements in the **"Personal and Sensitive Information"** section below. These Google Play requirements are in addition to any requirements prescribed by applicable privacy and data protection laws.

## WhatsApp Privacy Policy

*If you live in the [European Region](#), WhatsApp Ireland Limited provides the services to you under this [Terms of Service](#) and [Privacy Policy](#).*

## WhatsApp Legal Info

If you live outside the [European Region](#), WhatsApp LLC ("WhatsApp," "our," "we," or "us") provides our Services to you under this [Terms of Service](#) and [Privacy Policy](#).

## Developer contact ^

### Website

<http://www.whatsapp.com/>

### Email

[android@support.whatsapp.com](mailto:android@support.whatsapp.com)

### Address

1601 Willow Road Menlo Park, CA 94025

### Privacy policy

<http://www.whatsapp.com/legal/#Privacy>

# Privacy Regulation

---

- Many strict regulations have been adopted for data protection
  - App developers modify the privacy policy to comply with the legal requirements



CCPA



GDPR



COPPA

## CalOPPA

California Online Privacy Protection Act

CalOPPA

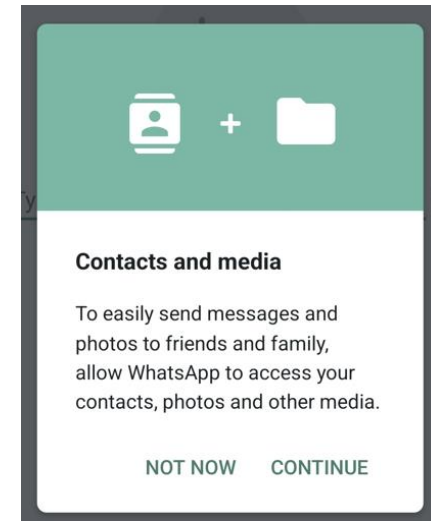
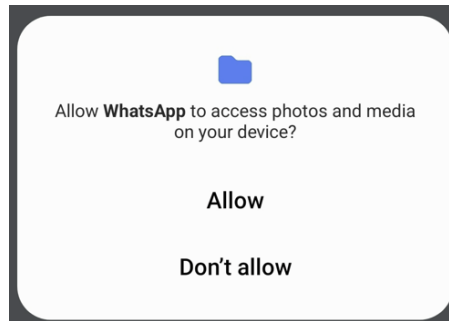
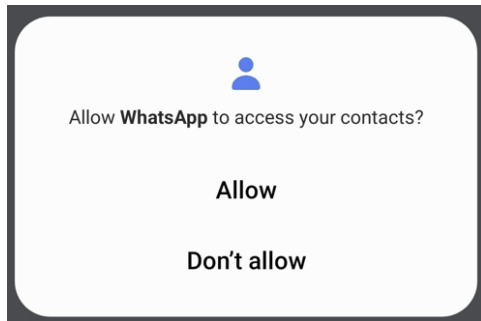


PIPEDA

# App Permission

---

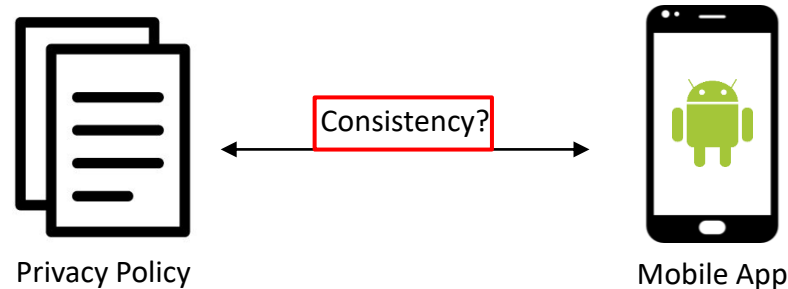
- Previous works focus on preventing excessive requests for permissions
  - Recently developed apps require only permission of essential data
- The permissions are requested at runtime
  - Users can allow or deny permissions in person



# Is the Privacy Policy Reliable?

---

- Privacy policy is the only way to know how the app collect user data
  - Users have to rely on the privacy policy to identify the app activities
- There is no technical solution to make sure that app developers specify all of the app practices in their privacy policy
  - It may be an inconsistency in privacy policy and app practices

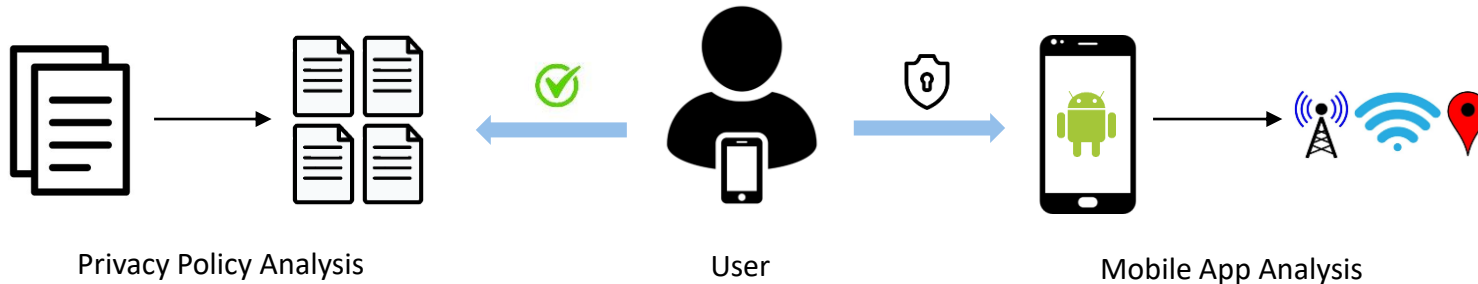




# Policy and App Analysis

---

- To ensure privacy transparency while using the app, we
  - Perform an mobile app analysis to check which data the app has access to
  - Perform a policy analysis to check the app practices described by the developers
  - Compare the results to check the reliability

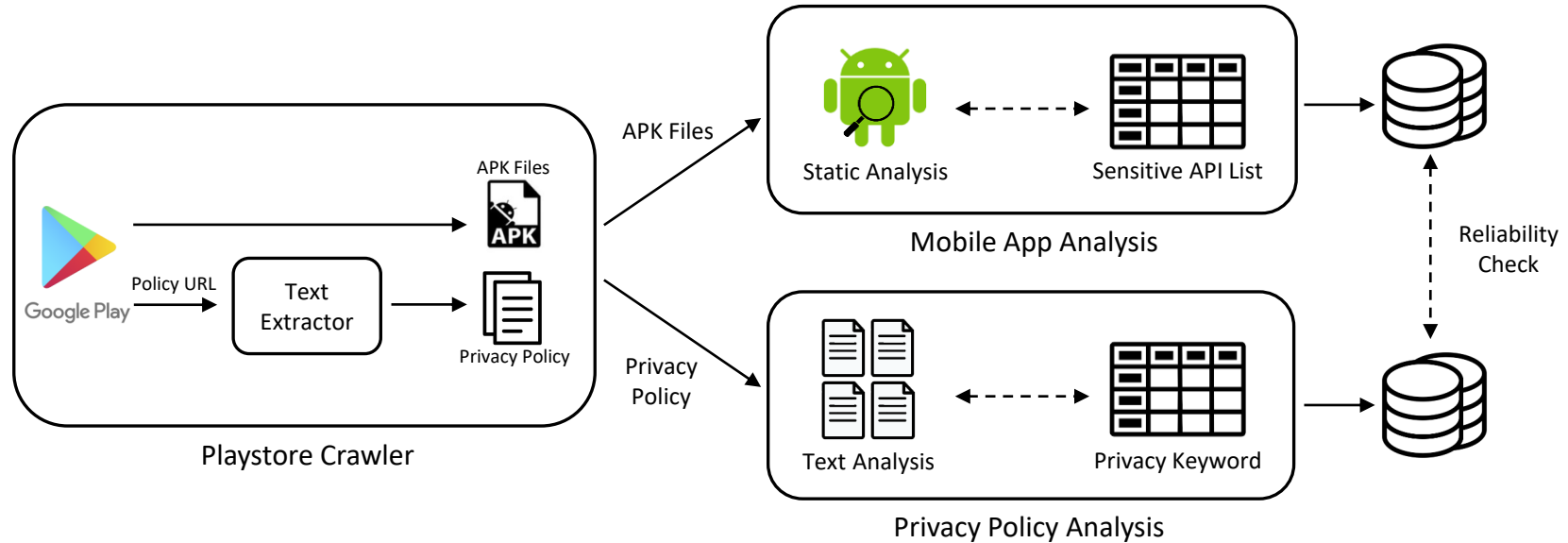


# System Design

---

# System Design

- Overview



# Privacy Category

---

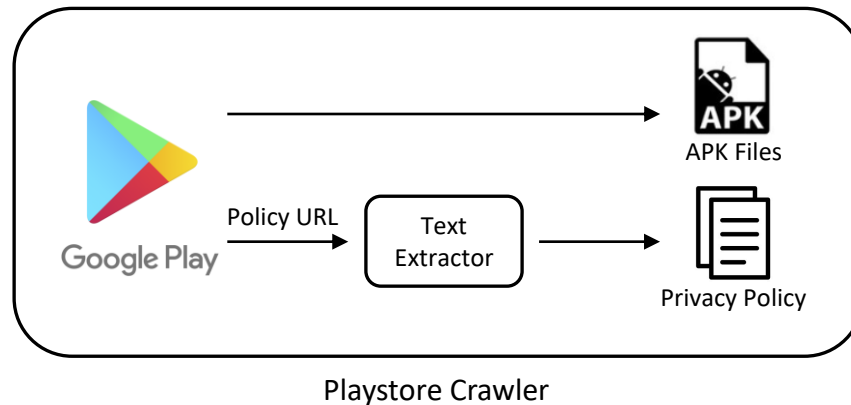
- We categorize personal data for analysis
  - The categories are based on Google User Data Policy
- For app analysis, we check the API lists that collect each personal data
  - The API lists are based on Android Documentation
- For policy analysis, we make a keyword list corresponding to each category

Category	Sub Category
Contact	Email
	Phone Number
Unique Identifier (ID)	Cookie
	Android ID
	IMEI
	IMSI
	MAC
	Mobile Carrier
	SIM Serial
	SSID and BSSID
	Cell Tower
Location	GPS and WiFi

# Playstore Crawler

---

- The playstore crawler downloads app's APK files and metadata from app store
  - We use open source crawler for collecting initial dataset
- The metadata contains app title, package name, policy URL, etc
  - Crawler stores this metadata on MongoDB in Json format



# Mobile App Analysis

---

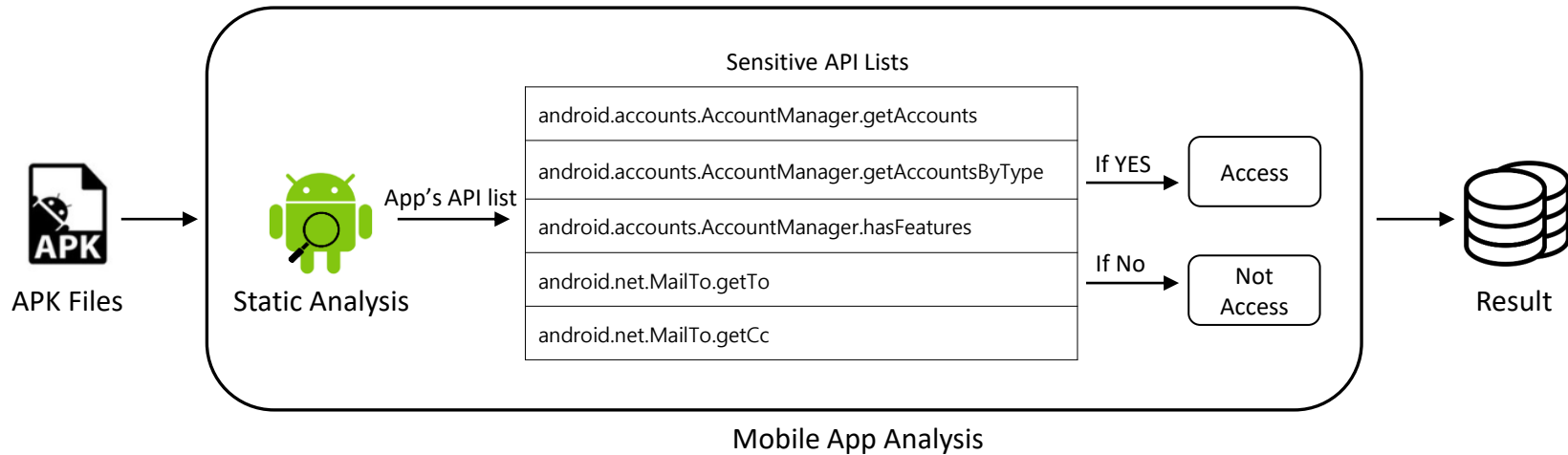
- We check the Sensitive API lists that collect each personal data
  - The API lists are based on Android Documentation

< Table : Example of Sensitive API Lists >

Category	Sensitive API
IMEI	android.telephony.TelephonyManager.getDeviceId
	android.telephony.TelephonyManager.getImei
GPS and WiFi	FusedLocationProviderClient.getLastLocation
	android.location.LocationManager.requestLocationUpdates
	android.location.LocationManager.requestSingleUpdate
	android.location.LocationManager.getLastKnownLocation

# Mobile App Analysis

- The system analyzes whether the apps have APIs that access personal data
  - It conduct the analysis by comparing the app's API and Sensitive API Lists
  - We use Androguard for static analysis



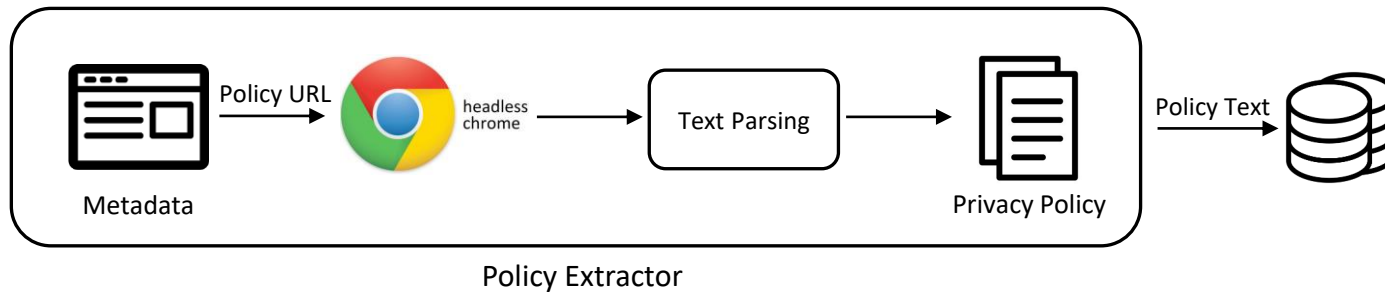
# Privacy Policy Analysis

- We extract privacy policy text from policy URL

Developer Contact ^

privacy policy  
<http://www.whatsapp.com/legal/#Privacy>

- We create a text extractor to handle errors while extracting privacy policy
  - First, the extractor renders the policy URL by chrome headless browser
  - Next, it uses Readability.js to parse the main text of the page
  - If Readability.js is not working, make an exception (E.g. different policy URL format)





# Privacy Policy Analysis

- We select 50 apps to effectiveness of privacy keyword
  - We check correctness of policy analysis results in person

		Prediction (Policy Analysis Results)	
		True	False
Actual Result	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

Precision	Recall	F1 score
0.814	0.778	0.796

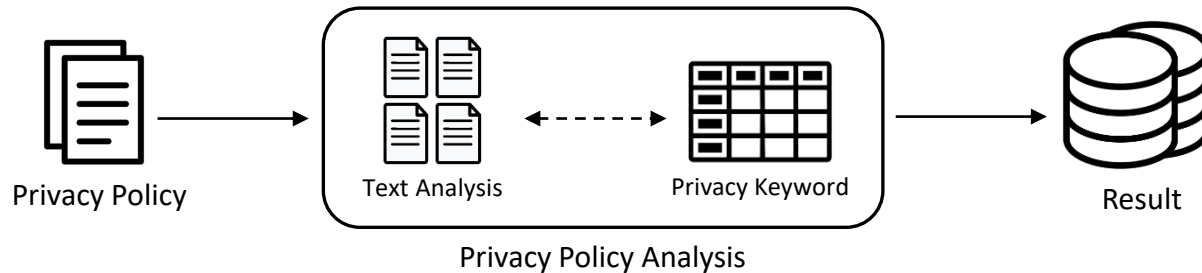
$$precision = TP / (TP + FP)$$

$$recall = TP / (TP + FN)$$

# Privacy Policy Analysis

---

- The system classifies the privacy policy to identify the specific policy text that describe a certain practice
  - The system analyzes whether the policy text describes the use of personal data
- The system conducts policy analysis to identify if the policy text has contents about the personal data



# Privacy Policy Analysis: Result

---

- The analysis results are recorded for each personal data

App Name	Personal Data Category
Youtube	Contact, Email, Phone Number, Identifier, Cookie, Device ID, Location, GPS and WiFi
Whatsapp	Contact, Email, Phone Number, Identifier, Cookie, Device ID, Location
Nexflix	Contact, Phone Number, Identifier, Cookie, Device ID, IMEI
Twitter	Contact, Email, Phone Number, Identifier, Cookie, Device ID, Location
Tiktok	Contact, Email, Phone Number, Identifier, Cookie, Device ID, IMEI, MAC, Location, GPS and WiFi

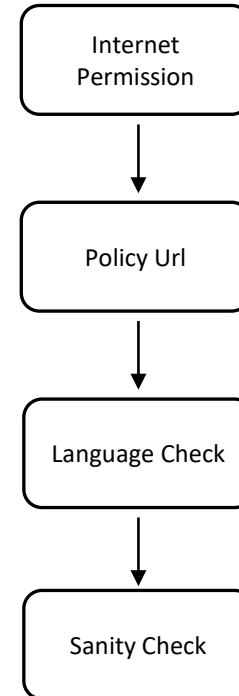
# Result

---

# Preprocessing Dataset

---

- We preprocess the dataset to select the app for analysis
- The dataset is preprocessed under 4 conditions
  - Condition 1: Internet permission
  - Condition 2: Policy URL
  - Condition 3: Language check (English)
  - Condition 4: Sanity check
- We conduct policy and app analysis on selected apps

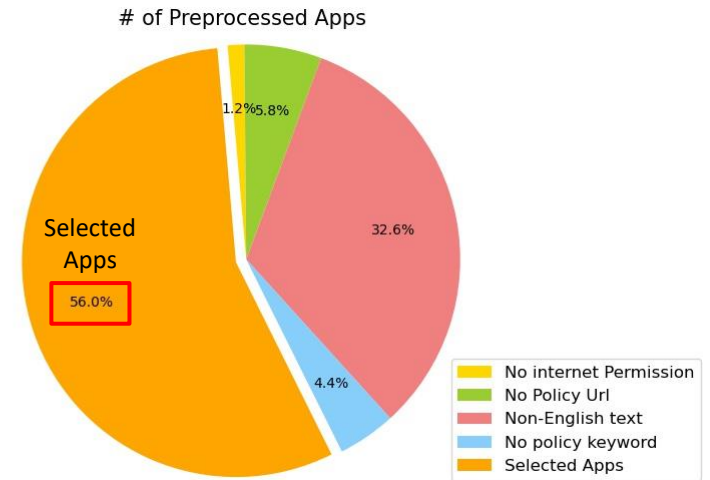


# Dataset

- The playstore crawler collected over 13K android apps' APK files and metadata

- The system preprocessed the collected apps based on their metadata

- Internet permission (13,223 → 13,058)
- Policy Url (13,058 → 12,286)
- English text (12,286 → 7,981)
- Policy keyword (7,981 → 7,401)

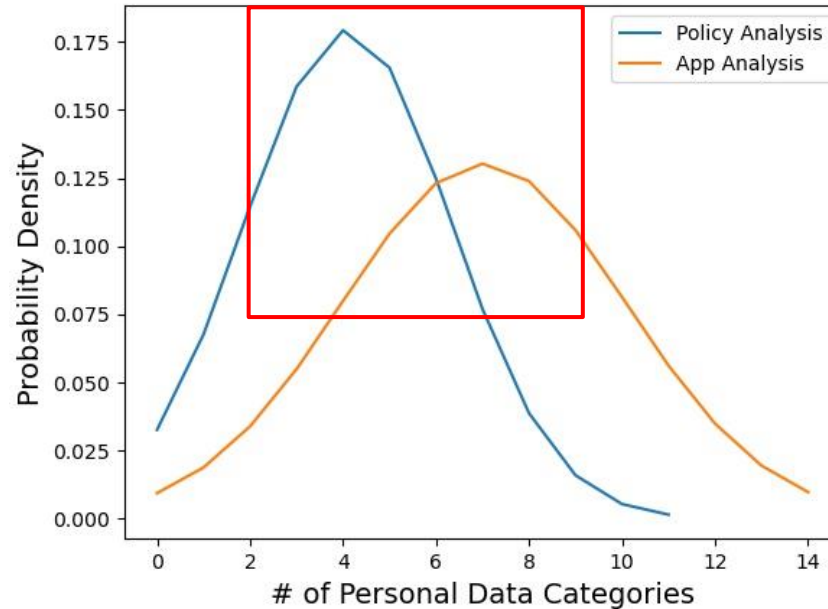


- The system conducted policy and app analysis on 7.4K apps

# Number of Personal Data Categories

---

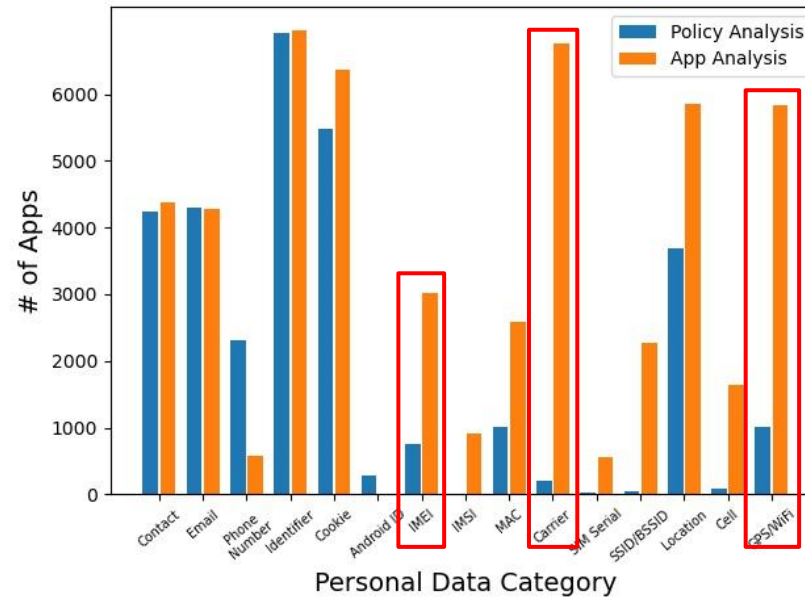
- App analysis results include more personal data categories



Number of Personal Data Categories by Policy and App Analysis

# Number of Apps for Each Category

- Some personal data categories make a big difference



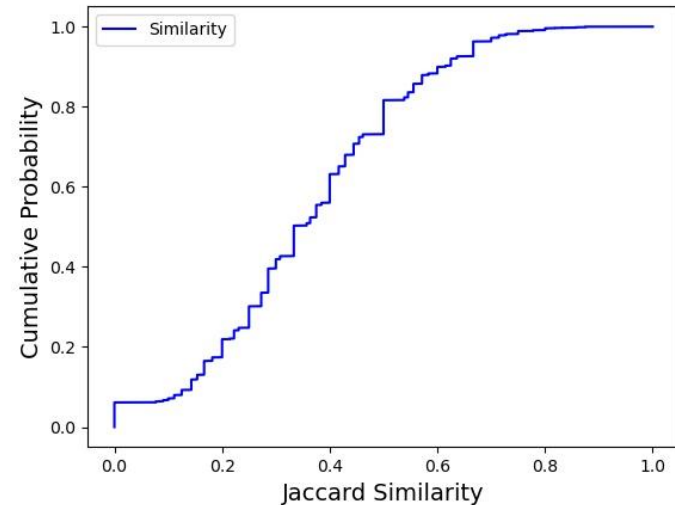
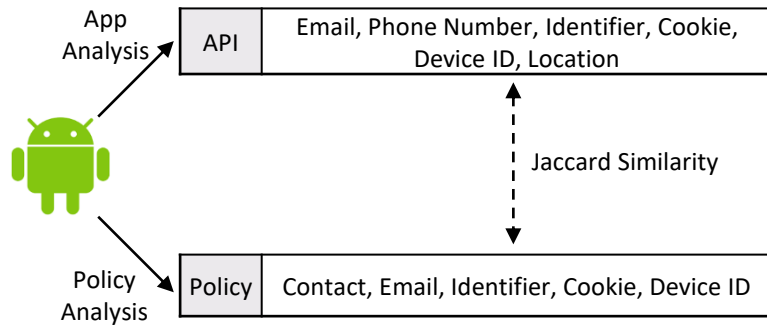
Number of Apps for Each Category



# Jaccard Similarity

- Jaccard similarity computes the similarity of policy and app practice

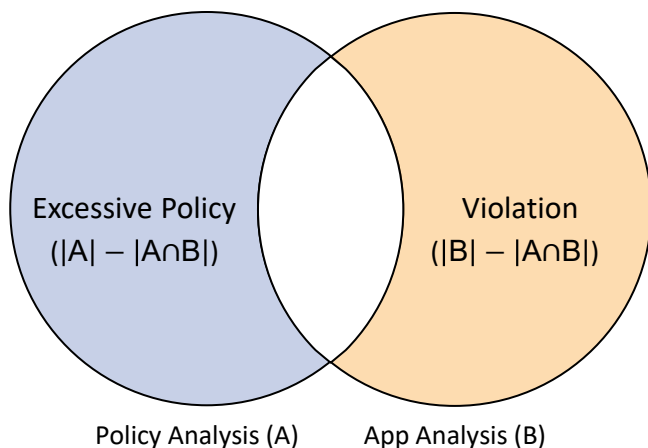
$$\text{Jaccard Similarity: } J(A, B) = \frac{A \cap B}{A \cup B} = \frac{\text{App Analysis} \cap \text{Policy Analysis}}{\text{App Analysis} \cup \text{Policy Analysis}}$$



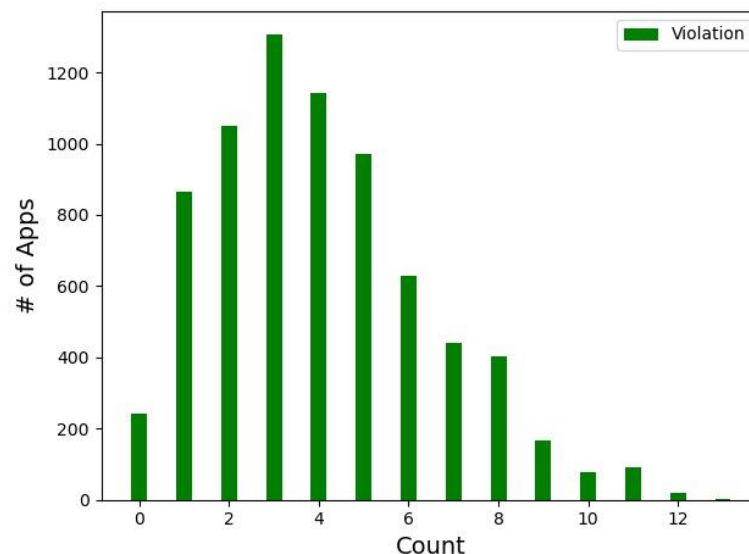
Similarity between Policy and App Practice

# Excessive Policy and Violation

- We defined Excessive Policy and Violation for checking reliability



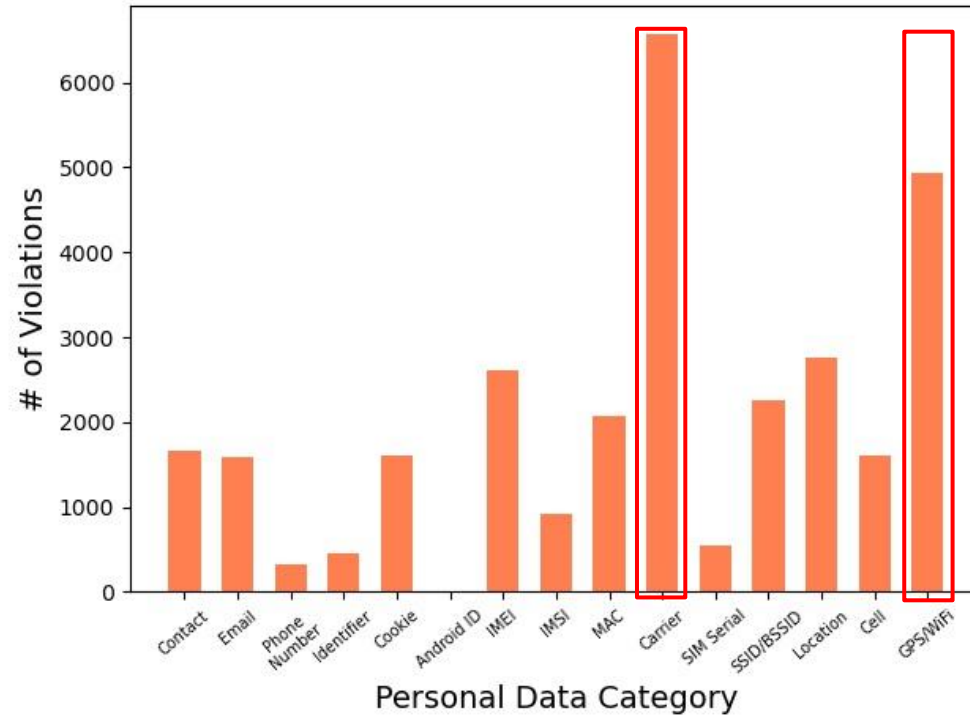
Definition of Excessive Policy and Violation



Number of Apps by Violation Count

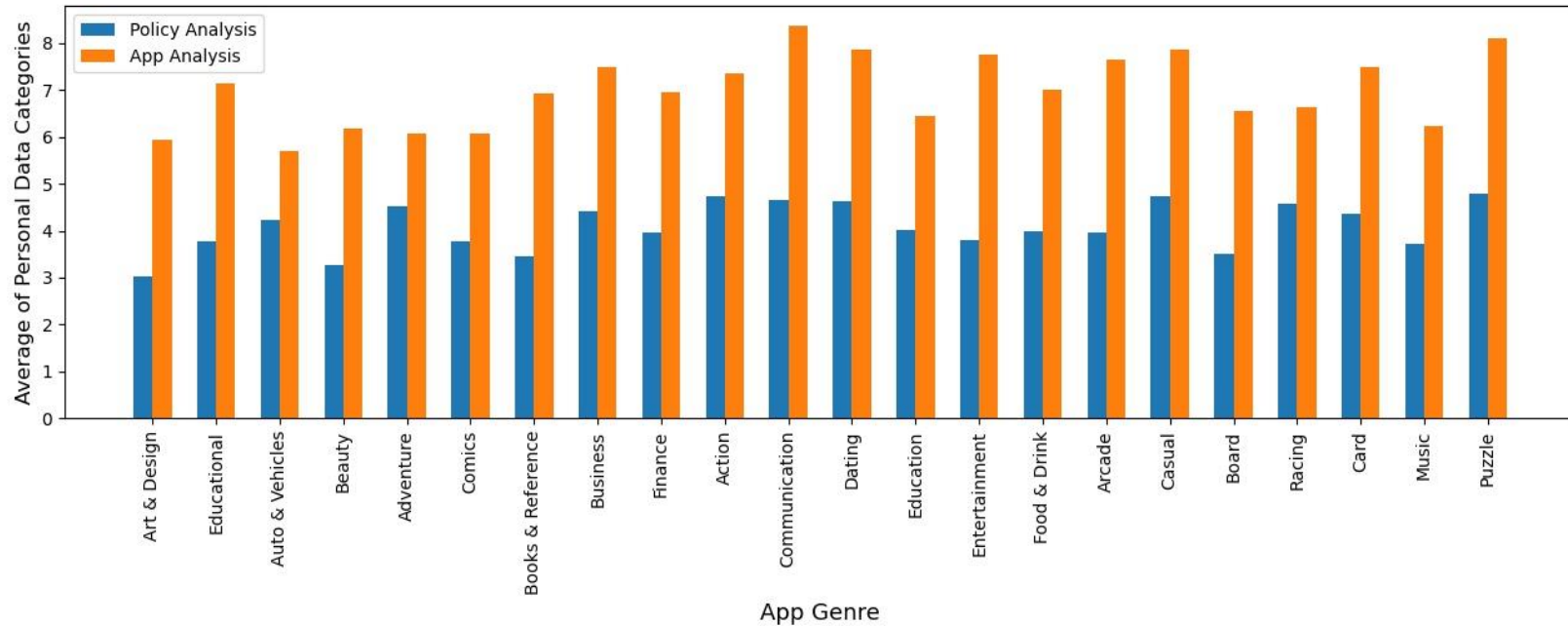
# Violation by Category

---



Number of Violations by Personal Data Category

# Average of Personal Data Categories by Genre



Average of Personal Data Categories by Genre

# Conclusion

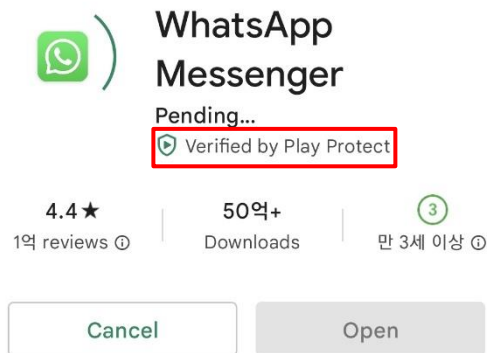
---

- Privacy policy is the only way to know how the app collect user data
  - It may be an inconsistency in app practices and privacy policy
- Google use Play Protect to check apps and devices for harmful behavior
  - It is hard to check all the app's practices

## Use Google Play Protect to help keep your apps safe and your data private

Google Play Protect checks your apps and devices for harmful behavior.

- It runs a safety check on apps from the Google Play Store before you download them.
- It checks your device for potentially harmful apps from other sources. These harmful apps are sometimes called malware.
- It warns you about potentially harmful apps.
- It may deactivate or remove harmful apps from your device.
- It warns you about detected apps that violate our [Unwanted Software Policy](#) by hiding or misrepresenting important information.
- It sends you privacy alerts about apps that can get user permissions to access your personal information, violating our [Developer Policy](#).
- It may reset app permissions to protect your privacy on certain Android versions.



# Conclusion

---

- We analyzed the app practices and privacy policy to check the reliability
  - Mobile app analysis shows whether the apps have APIs that access personal data
  - Privacy policy analysis identify the specific policy text that describe a certain practice
  - We compared these result to check if privacy policies are reliable
- Results show that apps can access more personal data than expected
  - App has more violation than excessive policy
  - We need more solutions to enable users to check the reliability

Thank you

---

Q&A

---



# Appendix

---

# Privacy Table: Policy Analysis

Category	Keyword	Sub Category	Keyword
Contact	<ul style="list-style-type: none"> <li>· contact information</li> <li>· account</li> </ul>	Email	<ul style="list-style-type: none"> <li>· email address</li> <li>· e-mail address</li> <li>· email id</li> </ul>
		Phone Number	<ul style="list-style-type: none"> <li>· phone number</li> <li>· telephone number</li> </ul>
Unique Identifier (ID)	<ul style="list-style-type: none"> <li>· personal information</li> <li>· device id</li> <li>· unique id</li> <li>· phone status</li> </ul>	Cookie	<ul style="list-style-type: none"> <li>· cookie</li> </ul>
		Android Id	<ul style="list-style-type: none"> <li>· android id</li> </ul>
		IMEI	<ul style="list-style-type: none"> <li>· international mobile device identification code</li> <li>· IMEI</li> </ul>
		IMSI	<ul style="list-style-type: none"> <li>· IMSI</li> <li>· subscriber ID</li> </ul>
		MAC	<ul style="list-style-type: none"> <li>· MAC</li> <li>· network card address</li> <li>· Media Access Control</li> </ul>
		Mobile Carrier	<ul style="list-style-type: none"> <li>· mobile carrier</li> <li>· mobile network operator</li> <li>· MNO</li> <li>· sim operator</li> <li>· wireless service provider</li> </ul>
		SIM Serial	<ul style="list-style-type: none"> <li>· sim serial</li> <li>· sim card</li> </ul>
		SSID BSSID	<ul style="list-style-type: none"> <li>· wifi router</li> <li>· wi-fi router</li> <li>· SSID</li> <li>· BSSID</li> </ul>
Location	<ul style="list-style-type: none"> <li>· location</li> </ul>	Cell Tower	<ul style="list-style-type: none"> <li>· cell tower</li> <li>· cellular tower</li> <li>· cell information</li> </ul>
		GPS and WiFi	<ul style="list-style-type: none"> <li>· network state</li> <li>· GPS</li> <li>· wifi</li> <li>· wi-fi</li> <li>· geolocation</li> </ul>

# Privacy Table: Example

Category	Keyword	Sub Category	Keyword	API
Unique Identifier (ID)	<ul style="list-style-type: none"><li>· personal information</li><li>· device id</li><li>· unique id</li><li>· phone status</li></ul>	IMEI	· international mobile device identification code	android.telephony.TelephonyManager.getDeviceId
			· IMEI	android.telephony.TelephonyManager.getImei
		IMSI	· IMSI	android.telephony.TelephonyManager.getSubscriberId
		SIM Serial	· sim serial	android.telephony.TelephonyManager.getSimSerialNumber
			· sim card	
		SSID BSSID	· wifi router	android.net.wifi.WifiManager.getConfiguredNetworks
			· wi-fi router	android.net.wifi.WifiInfo.getBSSID
			· SSID	
			· BSSID	android.net.wifi.WifiInfo.getSSID

# Privacy Table: App Analysis

Category	Sub Category	API
Contact	Email	android.accounts.AccountManager.getAccounts
		android.accounts.AccountManager.getAccountsByType
		android.accounts.AccountManager.getAccountsByTypeAndFeatures
		android.accounts.AccountManager.getAccountsByTypeForPackage
	Phone Number	android.telephony.TelephonyManager.getLine1Number
Unique Identifier (ID)	Cookie	android.webkit.CookieManager.getInstance
	Android Id	android.provider.Settings.Secure.getString
	IMEI	android.telephony.TelephonyManager.getDeviceId
		android.telephony.TelephonyManager.getImei
	IMSI	android.telephony.TelephonyManager.getSubscriberId
	MAC	android.net.wifi.WifiInfo.getMacAddress
	Mobile Carrier	android.telephony.TelephonyManager.getNetworkOperator
		android.telephony.TelephonyManager.getNetworkOperatorName
		android.telephony.TelephonyManager.getSimOperator
		android.telephony.TelephonyManager.getSimOperatorName
	SIM Serial	android.telephony.TelephonyManager.getSimSerialNumber
Location	Cell Tower	android.net.wifi.WifiManager.getConfiguredNetworks
		android.net.wifi.WifiInfo.getBSSID
		android.net.wifi.WifiInfo.getSSID
		FusedLocationProviderClient.getLastLocation
		android.location.LocationManager.requestLocationUpdates
		android.location.LocationManager.requestSingleUpdate
		android.location.LocationManager.getLastKnownLocation
		android.telephony.gsm.GsmCellLocation.getCid
		android.telephony.gsm.GsmCellLocation.getLac
		android.telephony.TelephonyManager.getCellLocation
		android.telephony.TelephonyManager.getAllCellInfo
	GPS and WiFi	FusedLocationProviderClient.getLastLocation
		android.location.LocationManager.requestLocationUpdates
		android.location.LocationManager.requestSingleUpdate
		android.location.LocationManager.getLastKnownLocation

# Related Work: Privacy Policy Analysis

---

- Privacy policy corpus creation

- Wilson et al. (2016) create a privacy policy corpus (OPP-115) of 115 web privacy policies with crowdsourcing
- Lebanoff et al. (2018) create a corpus for privacy policies for vague words prediction
- Zimmeck et al. (2019) create an app privacy policy corpus (APP-350) for App executable and privacy policy compliance check

- Automatic privacy policy analysis

- Ramanath et al. (2014), Liu et al. (2018) focus on the structure of privacy by identifying policy sections relating to different topics
- Sathyendra et al. (2017) provide labels on opt-out choices similar consumer choice options on websites
- Libert (2018) finds that the names of third parties are usually not explicitly disclosed in website privacy policies

# Related Work: Mobile App Analysis

---

- Static Analysis

- Mutchler et al (2015) analyze Android web apps and find that 28% of those have at least one vulnerability
- Story et al. (2018) study the metadata of apps on the Play Store and find that many apps lack privacy policies, even when developers describe the data collection
- Reyes et al. (2018) reveal that many Android apps collect persistent device identifiers to track users, which is not allowed for advertising purposes

- Dynamic Analysis

- Abbas et al. (2015) develop Haystack (Lumen app) to monitor traffic from apps
- Ren et al.(2018) study privacy leak using Lumen and UI monkey generator

# Work

---

- We built an automated system that analyzes the privacy policy and app practices to check the consistency
  - We created privacy table to analyze the privacy policies and android methods regarding sensitive data
- The system crawled play store to collect APK files and policy text
- We analyzed privacy policy text to identify which data is informed to be collected
- We analyzed app source code to identify whether the app has sensitive API

# Readability.js

- Node.js library that provides Firefox Reader Mode

Cookies Policy  
Open Source  
Virtual Items  
Intellectual Property Policy  
Law Enforcement  
Privacy Policy  
Terms of Service

## Privacy Policy

(If you are a user having your usual residence in the US)

*Last update: January 1, 2020.*

Welcome to TikTok (the "Platform"). The Platform is provided and controlled by TikTok Inc. ("TikTok", "we" or "us"). We are committed to protecting and respecting your privacy. This Privacy Policy covers the experience we provide for users age 13 and over on our Platform. For information about our under-13 experience ("Children's Platform") and our practices in the United States regarding children's privacy, please refer to our [Privacy Policy for Younger Users](#).

Capitalized terms that are not defined in this policy have the meaning given to them in the [Terms of Service](#).

### What information do we collect?

We collect information when you create an account and use the Platform. We also collect information you share with us from third-party social network providers, and technical and behavioral information about your use of the Platform. We also collect information contained in the messages you send through our Platform and information from your phone book, if you grant us access to your phone book on your mobile device. More information about the categories and sources of information is provided below.

### Information you choose to provide

For certain activities, such as when you register, upload content to the Platform, or contact us directly, you may provide some or all of the following information:

✕  
As  
13  
18

### What information do we collect?

We collect information when you create an account and use the Platform. We also collect information you share with us from third-party social network providers, and technical and behavioral information about your use of the Platform. We also collect information contained in the messages you send through our Platform and information from your phone book, if you grant us access to your phone book on your mobile device. More information about the categories and sources of information is provided below.

### Information you choose to provide

For certain activities, such as when you register, upload content to the Platform, or contact us directly, you may provide some or all of the following information:

## Firefox Reader mode

### What information do we collect?

We collect information when you create an account and use the Platform. We also collect information you share with us from third-party social network providers, and technical and behavioral information about your use of the Platform. We also collect information contained in the messages you send through our Platform and information from your phone book, if you grant us access to your phone book on your mobile device. More information about the categories and sources of information is provided below.

### Information you choose to provide

For certain activities, such as when you register, upload content to the Platform, or contact us directly, you may provide some or all of the following information:

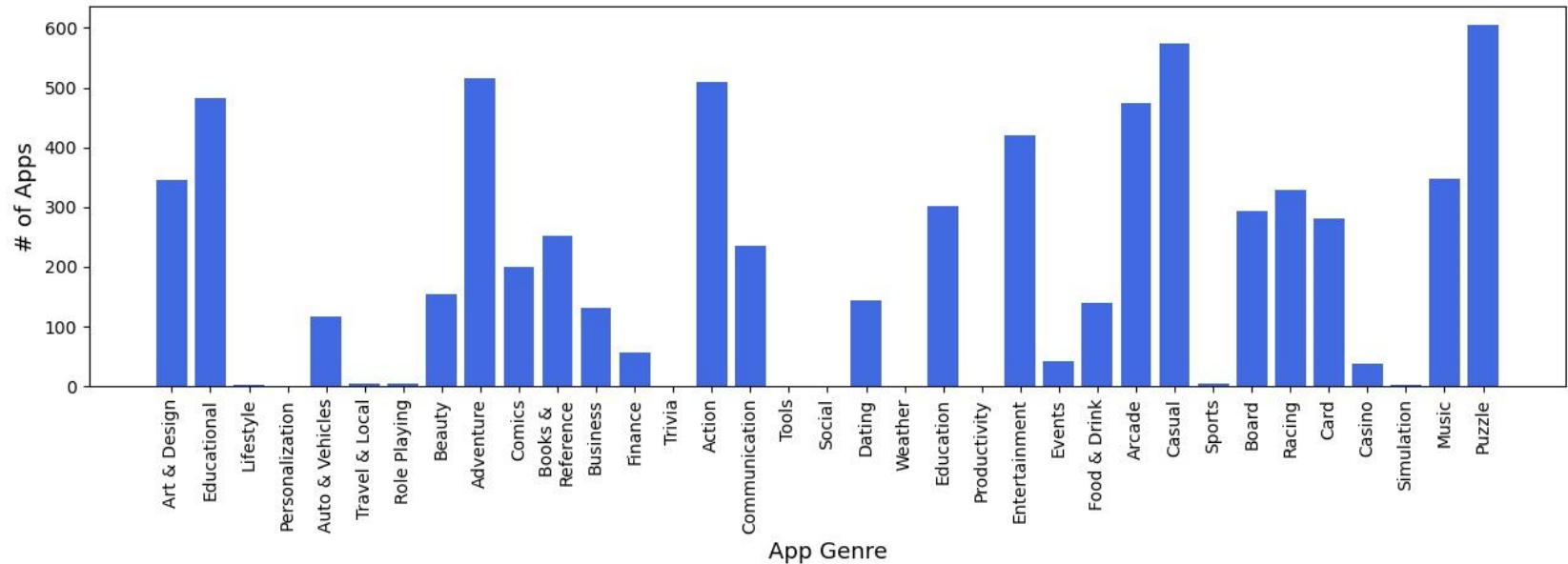
- Registration information, such as age, username and password, language, and email or phone number
- Profile information, such as name, social media account information, and profile image
- User-generated content, including comments, photographs, videos, and virtual item videos that you choose to upload or broadcast on the Platform ("User Content")
- Payment information, such as PayPal or other third-party payment information (where required for the purpose of payment)
- Your phone and social network contacts, with your permission. If you choose to find other users through your phone contacts, we will access and collect the names and phone numbers and match that information against existing users of the Platform. If you choose to find other users through your social network contacts, we will collect your public profile information as well as names and profiles of your social contacts

Readability



# Number of Apps by genre

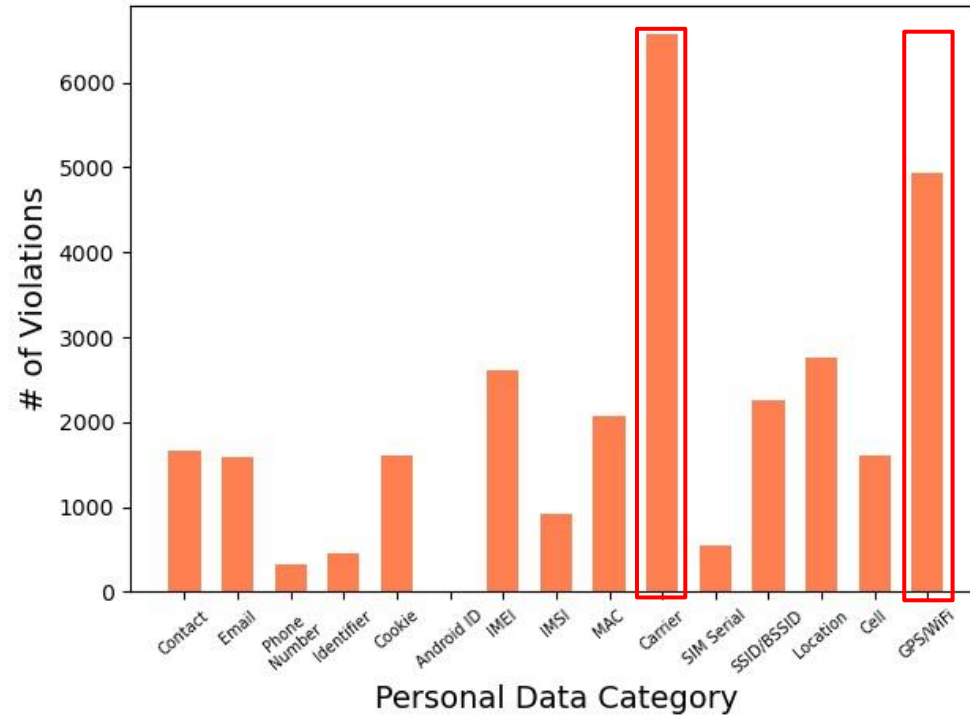
- Fewer than 100 apps were collected in some genres
  - We excluded the genres from the analysis to obtain a more accurate result



Number of Apps by genre

# Violation by Category

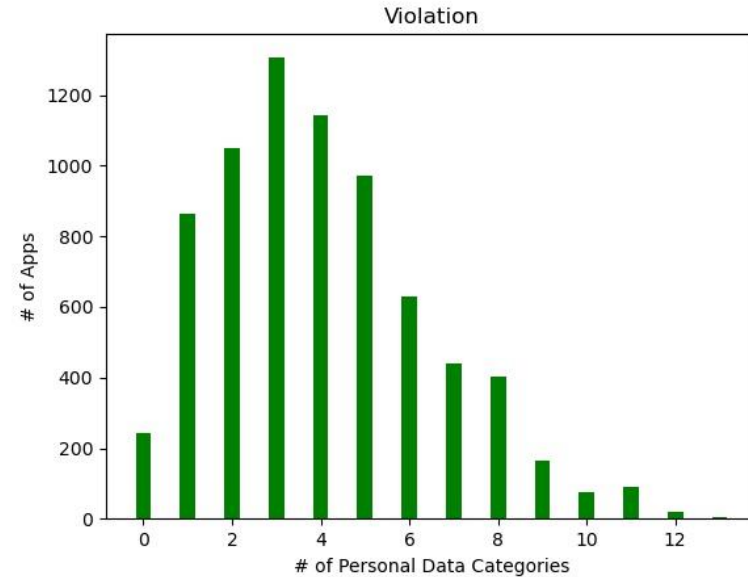
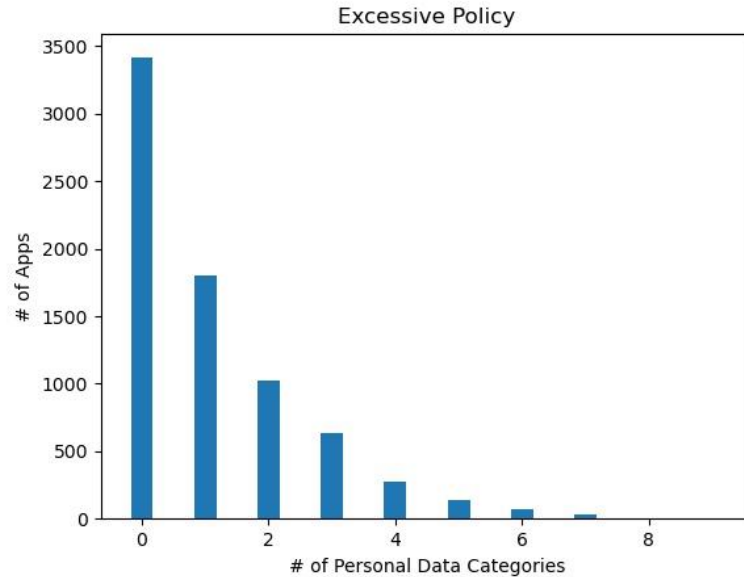
---



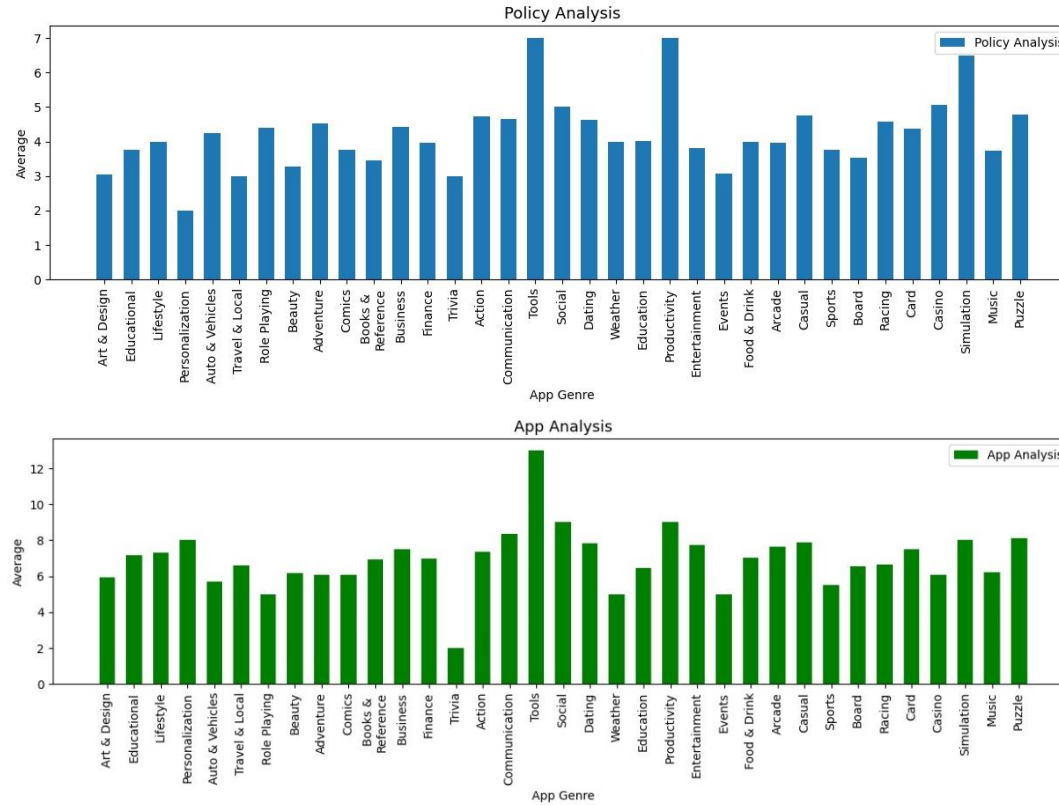
Number of Violations by Personal Data Category

# Excessive Policy and Violation Count

- Number of Apps by Excessive Policy and Violation Count



# Average of Personal Data Categories by Genre



# App Analyzer

---

- The system analyzes whether the apps have APIs that access personal data
  - It conduct the analysis by comparing the app's API and privacy table
  - We use Androguard for static analysis
- After the analysis, the system store the API list to the database

