Bug Fixes, Improvements,… and Privacy Leaks:

A Longitudinal Study of PII Leaks Across Android App Versions

Jingjing Ren*, Martina Lindorfer†, Daniel J. Dubois*,
Ashwin Rao‡, David Choffnes* And Narseo Vallina-Rodriguez§

*Northeastern University †UC Santa Barbara

‡University Of Helsinki §IMDEA Networks Institute And ICSI

Sponsored by:

DATA TRANSPARENCY LAB

Outline

Motivation

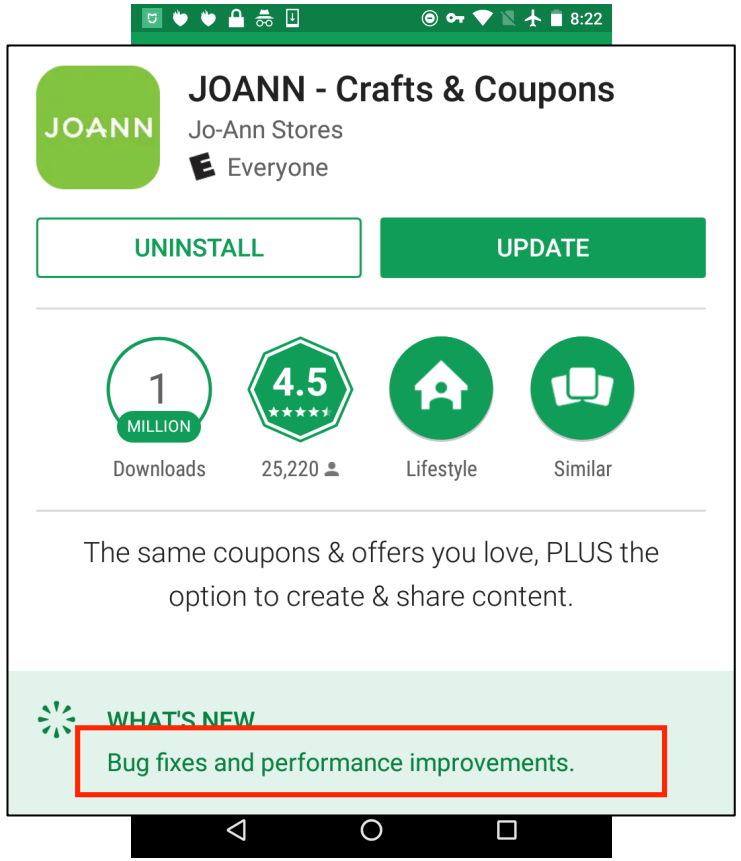
Threat Model

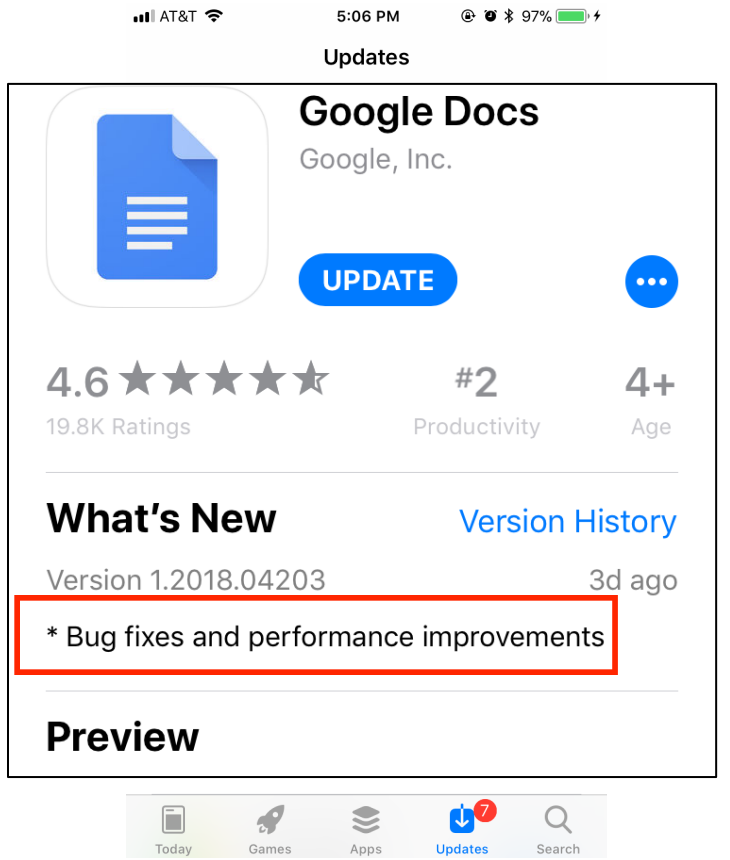
Methodology

Macroscopic Trends in Privacy

Conclusion

Motivation



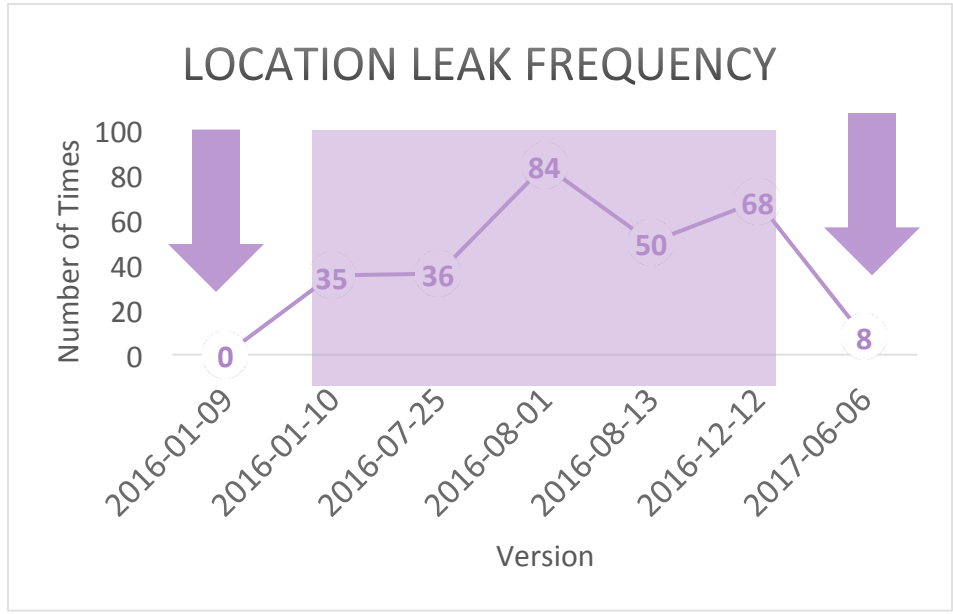


Are there any changes in privacy when I update the app?

The Evolution of Privacy in Mobile Devices

3,545 times in a week

JOANN



How does mobile privacy evolve over time? (☺ or ☹)

Outline

Motivation

Threat Model

- Privacy definition
- Leak definition

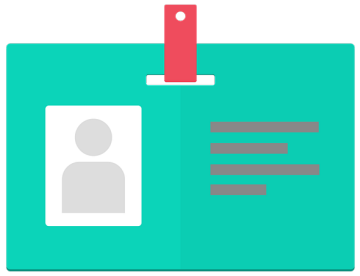
Methodology

Macroscopic Trends in Privacy

Conclusion

What Do I Mean by “Privacy” in This Work?

What information is shared?

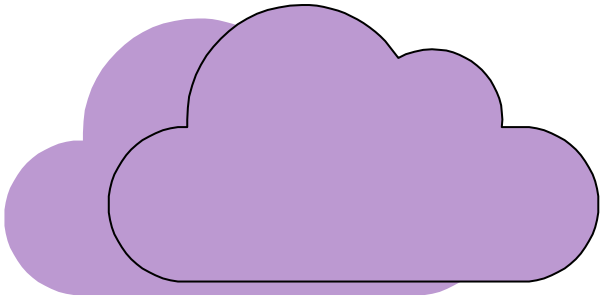


Personally Identifiable Information (PII)
Tracking ID
User information
Location
Contact
…

How is it being shared?

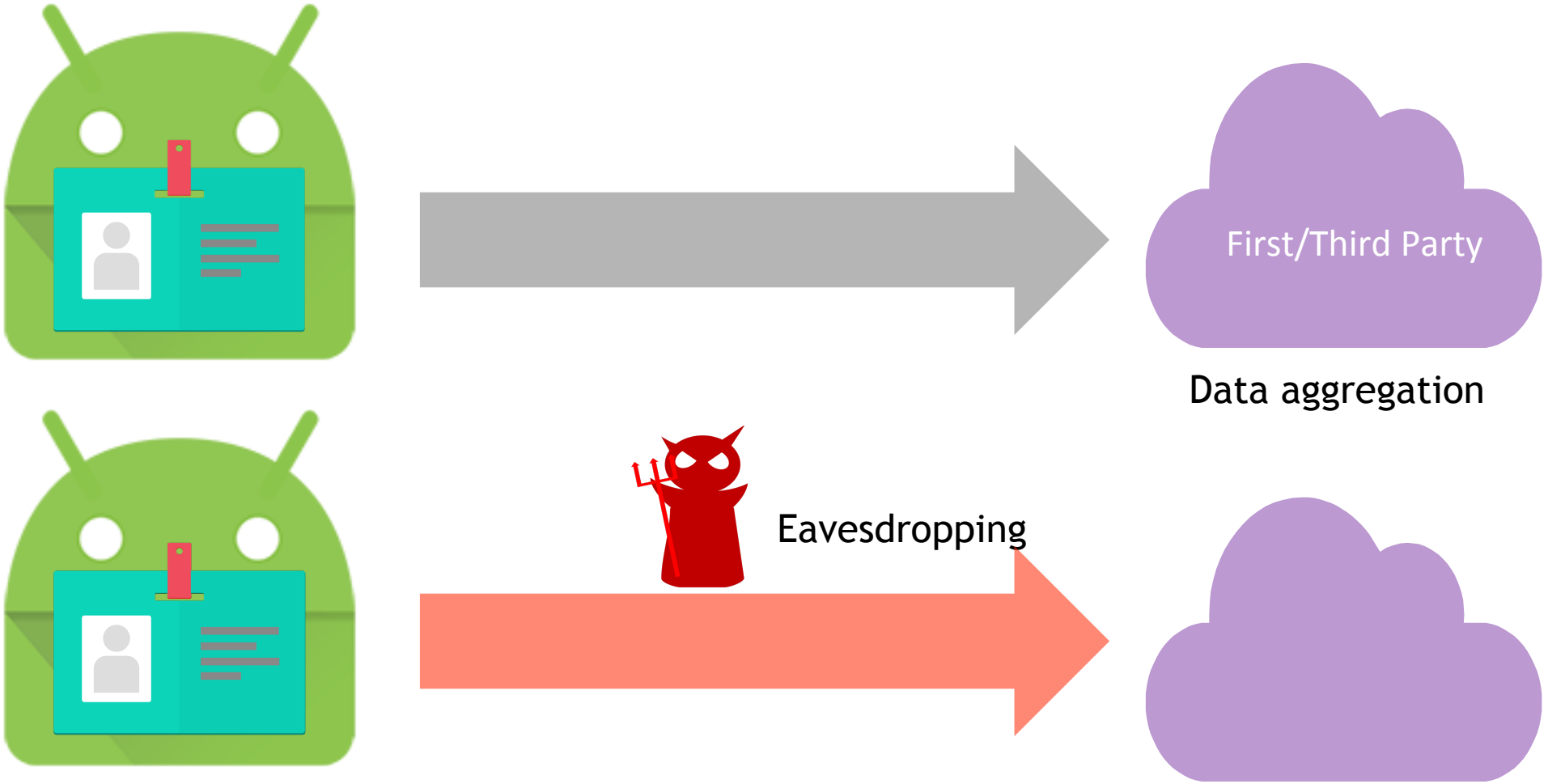
Transport security
Encrypted (HTTPS)
or
Plaintext (HTTP)

Where is it going?



Destination:
First party (app owner)
or
Third party (advertising & analytics)

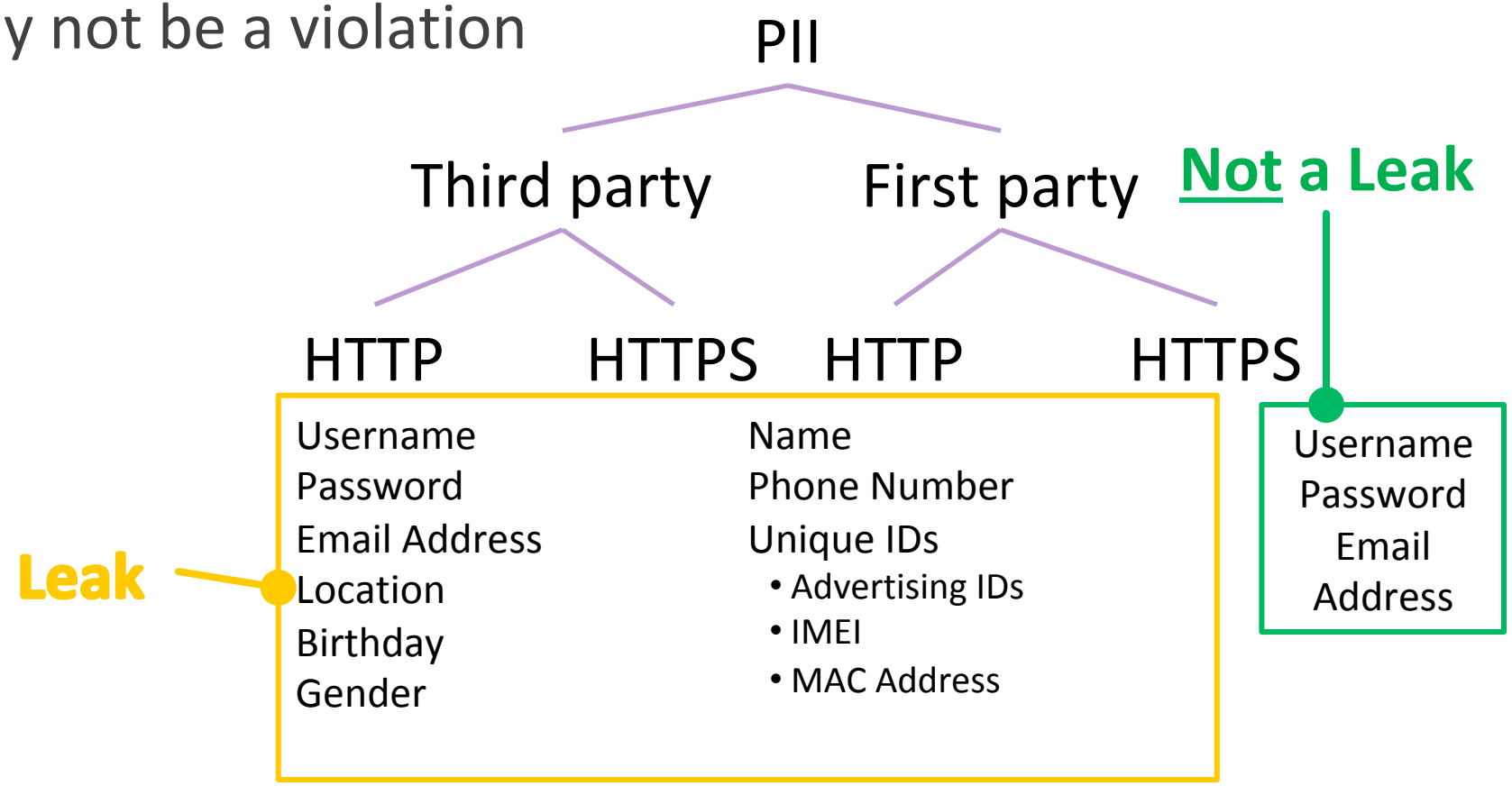
Threat Model



What is a “Leak” in This Work?

Goal: Understand user information shared with other parties

- A leak may or may not be a violation



Outline

Motivation

Threat Model

Methodology

- Controlled Experiments
- Detecting PII Leaks
- Privacy Attributes

Macroscopic Trends in Privacy

Conclusion

Controlled Experiments

App selection criteria
- Multiple versions [1]
- Popularity
- Amenable to traffic analysis
  - MITM TLS connections

Privacy measurements
1. Interact with apps
2. Detect privacy leaks
3. Validate manually

512 Android apps
7,665 versions (APKs)
8 years

[1] M. Backes, S. Bugiel, and E. Derr, “Reliable Third-Party Library Detectionin Android and its Security Applications,” In *Proc. of CCS*, 2016.

Interaction

Inducing privacy leaks requires interaction
- Real, controlled user interactions are good, … but not scalable ☹

Automated and scripted interaction
- Monkey: randomly generated events with good coverage
- login and replay across the versions
- ~10 minutes per experiment

Test environment
- 5 Android phones
- MITM proxy to intercept both plain-text and encrypted network traffic

Detecting PII Leaks

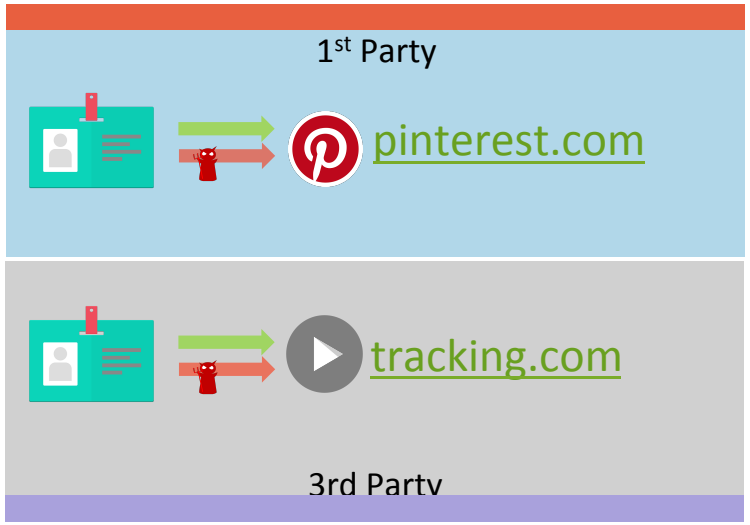
Network traffic analysis

- PII leaks, by definition, leak over Internet
- ReCon: using ML to detect without prior knowledge of PII values [Mobisys'16]
- Manual validation

Privacy Attributes

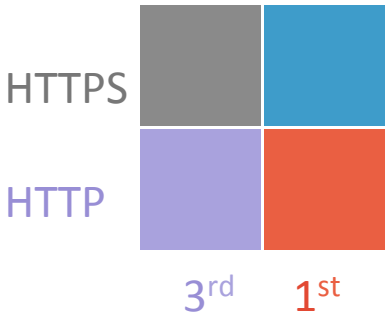
PII Types

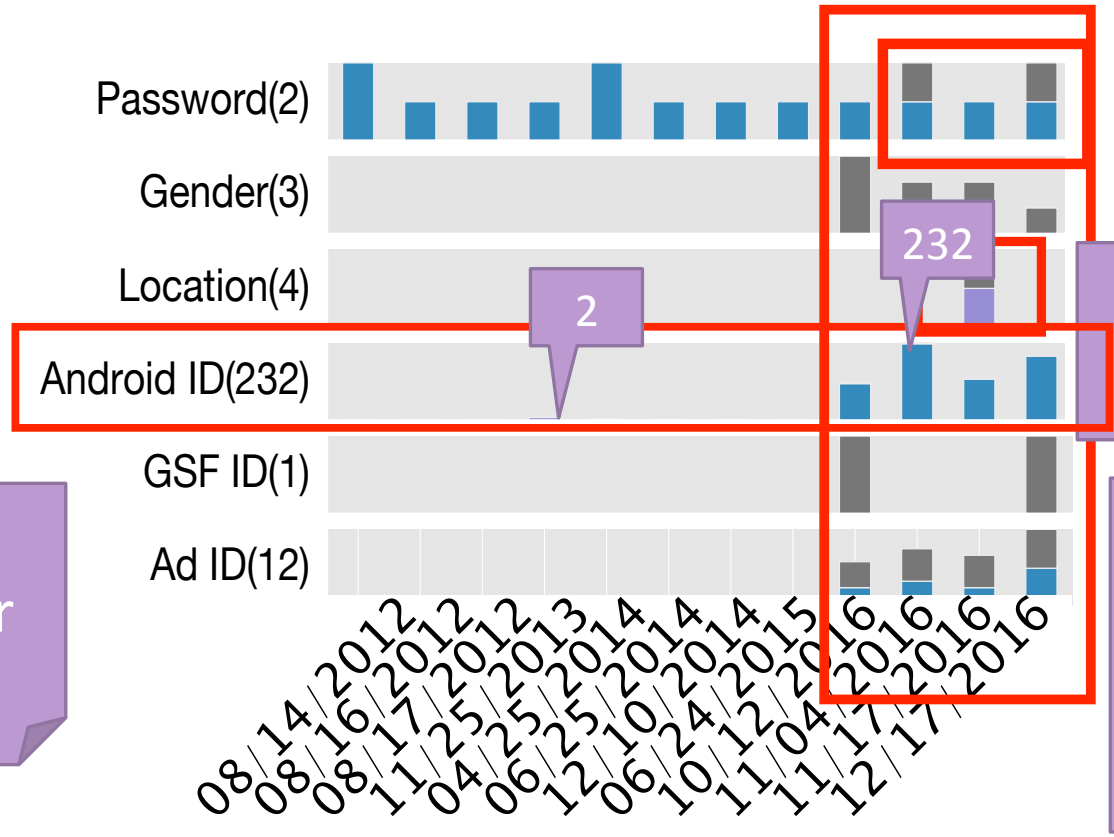
- Password
- Gender
- Location
- Android ID
- GSF ID
- Ad ID



08\14\2012
08\16\2012
08\17\2012
11\25\2013
04\25\2014
06\25\2014
12\10\2014
06\24\2015
10\12\2016
11\04\2016
11\17\2016
12\17\2016

Case study: Pinterest





Sending password to a third party in 2 out of 12 versions

HTTPS is used on

Jan. 5, 2017: disclosed
Feb. 7, 2017: fixed

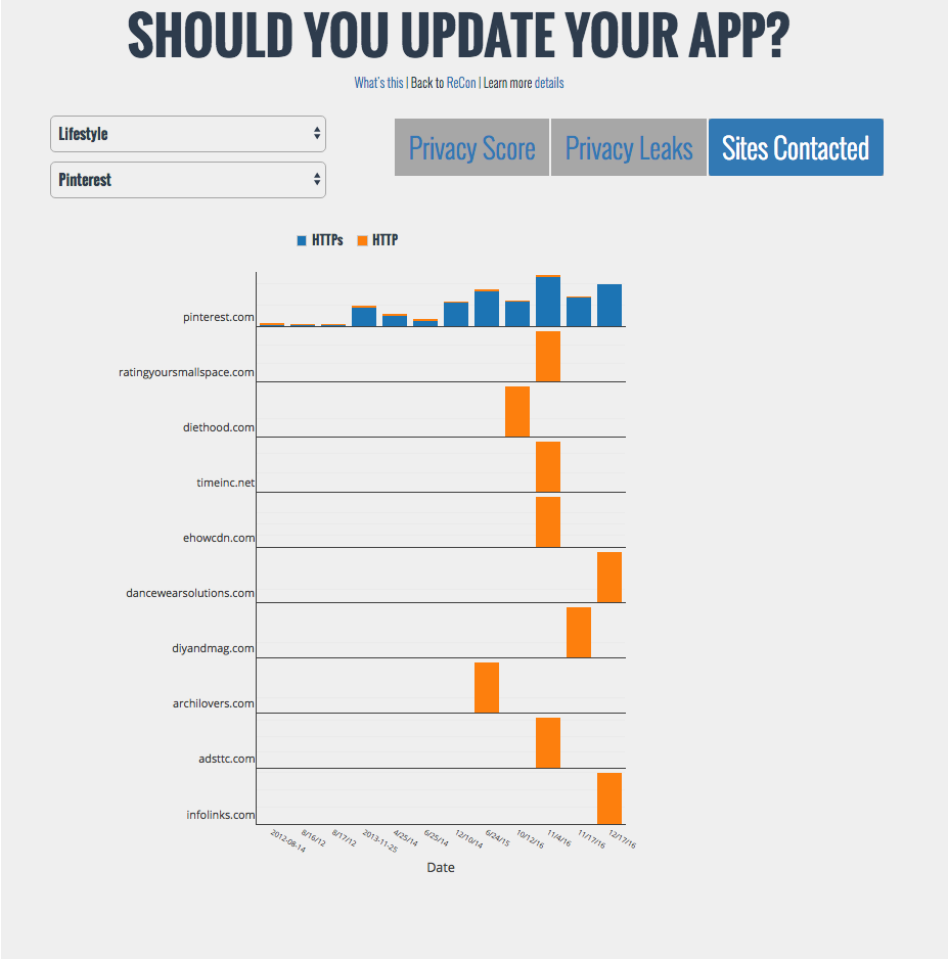
Increased frequency for Android ID

More types (gender, location, android ID etc.) are leaked

Takeaway: High variance in privacy risks across versions

- password leaks, PII types, frequency, encryption

Privacy Leaks For Individual Apps



https://recon.meddle.mobi/appversions/

Outline

Motivation

Threat Model

Methodology

Macroscopic Trends in Privacy

Conclusion

Macroscopic Trends in Privacy

Summary of Results

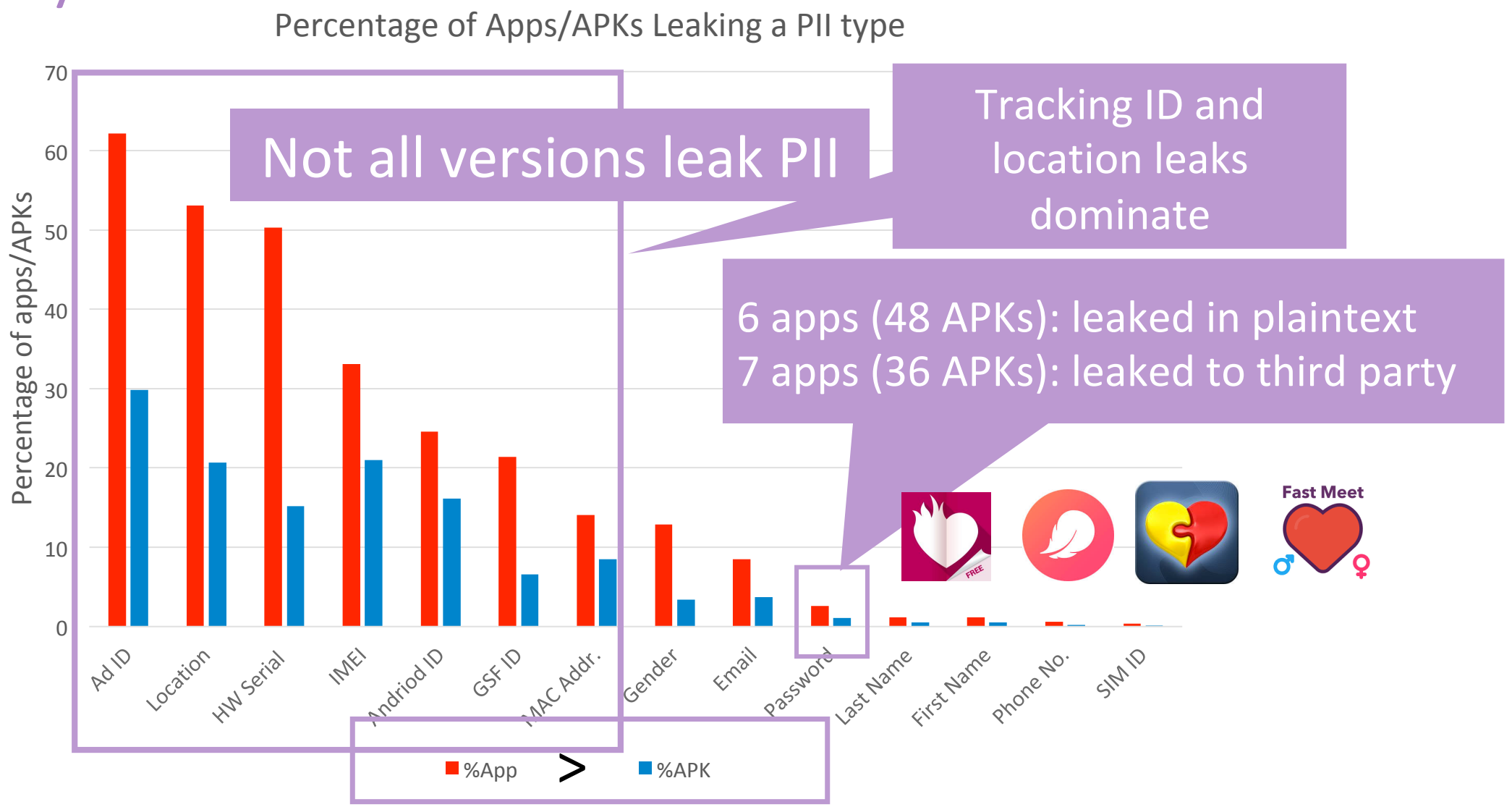
Variations in PII Leaks

HTTPS Adoption Trends

Third-Party Characterization

Multidimensional analysis

Summary of Results



Macroscopic Trends in Privacy

Summary of Results

Variations in PII Leaks

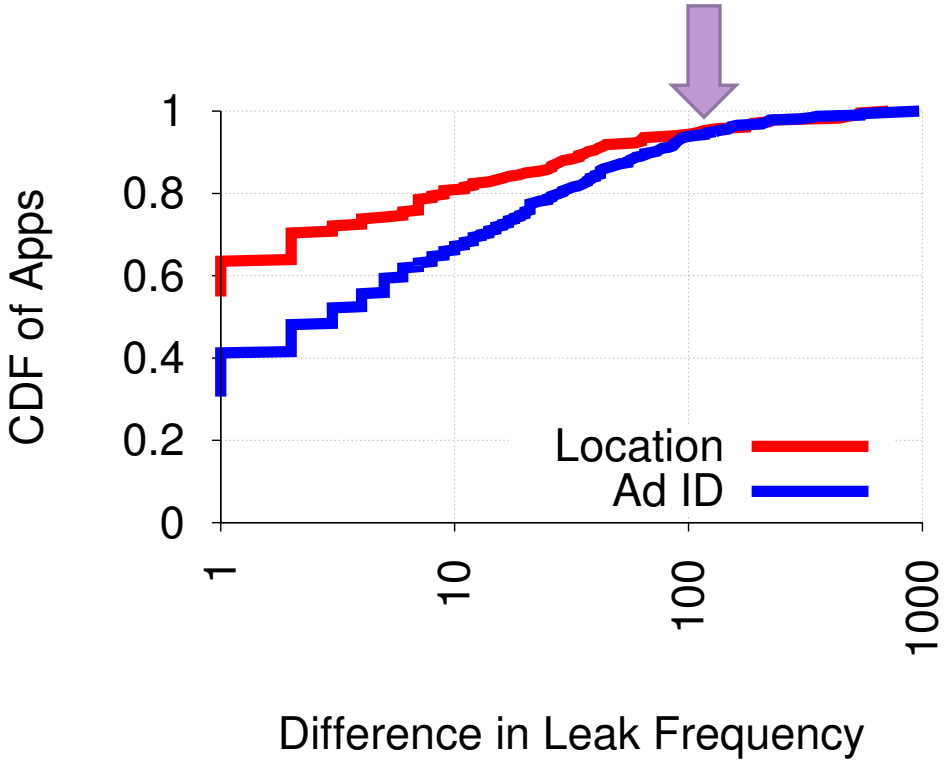
- How different are leaking frequencies

HTTPS Adoption Trends

Third-Party Characterization

Multidimensional analysis

Frequency of PII Leaks



5.6% apps see a several orders of magnitude difference

- fine-grained location tracking
- increased opportunities for network eavesdroppers to invade user privacy

Macroscopic Trends in Privacy

Summary of Results

Variations in PII Leaks

HTTPS Adoption Trends

- Extremely slow, for half of the domains:
  - 10% apps, 2 years
  - 50% apps, 5 years

Third-Party Characterization

Multidimensional analysis

Macroscopic Trends in Privacy

Summary of Results

Variations in PII Leaks

HTTPS Adoption Trends

Third-Party Characterization

Multidimensional analysis

High-risk Tracking Across Apps

Tracking ID: IMEI, Android ID, advertising ID, MAC address etc.

Other PII: location, gender, name, email

Third Parties	(Tracking ID +) Other PII	#Apps	#APKs
google[*]	Location, Gender, First/Last Name, Email	124	387
kochava.com	Email, Gender	8	36
vungle.com	Location, Gender	7	34
mopub.com	Location	6	13
…			
56txs4.com	Gender	3	11
…			

Might *permanently* link individuals/personal information to a tracking ID

>100 domains

Macroscopic Trends in Privacy

Summary of Results

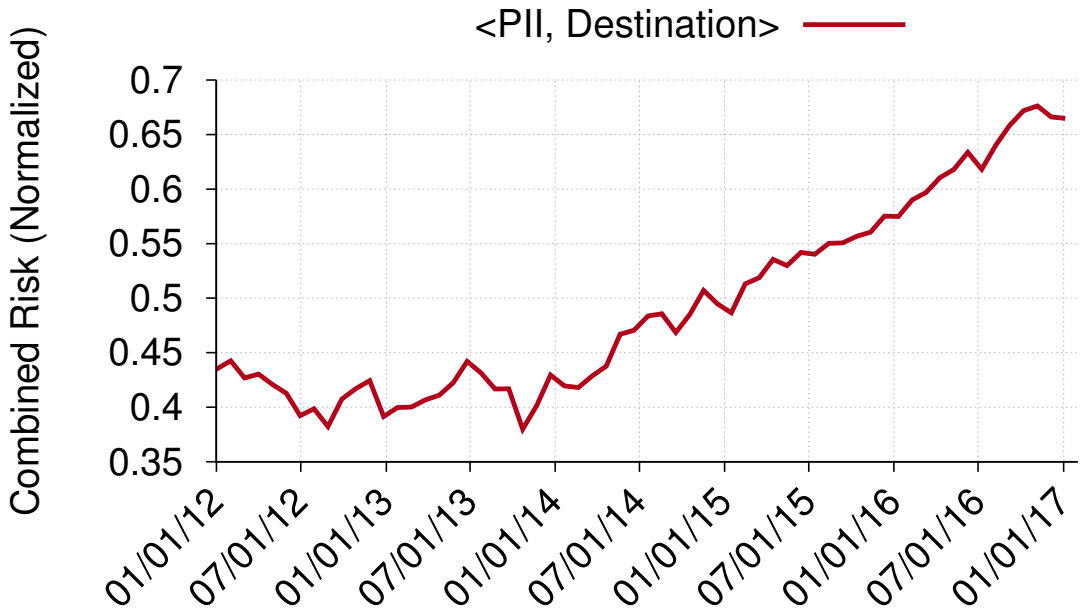
Variations in PII Leaks

HTTPS Adoption Trends

Third-Party Characterization

Multidimensional analysis

Is Privacy Getting Better or Worse?



Combined privacy worsens over time

- mainly due to more PII types and more domains

Outline

Motivation

Threat Model

Methodology

Macroscopic Trends in Privacy

Conclusion

Conclusion

- Privacy has *worsened* over time
  - PII can *change* substantially across versions
  - HTTPS Adoption is *slow*
  - Third-party tracking is *pervasive* ***and*** *broad*

- Need for continuous monitoring
  - Using systems: ReCon, Lumen, AntMonitor etc.

Disclaimer: we recommend updating apps for security reasons

https://recon.meddle.mobi/appversions/