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*, Narseo Vallina-Rodriguez§

*MNortheastern University †UC Santa Barbara 
‡University Of Helsinki §IMDEA Networks Institute And ICSI

Sponsored by:
Motivation
Evolution of Privacy in Mobile Devices

How does mobile privacy evolve over time? 😊 or 😞

3,545 times in a week

LOCATION LEAK FREQUENCY

Version

Number of Times
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
  - Plaintext
  - or
  - Encrypted

Where is it going?

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

Why do we care?

- Data Aggregation
- Eavesdropping Attack
Experiment Methodology

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

Interaction with apps
   - manual is good, … but not scalable 😞
   - Automated and scripted interaction: Monkey!
     • randomly generated events with good coverage
     • login and replay across the versions

MITM proxy to intercept HTTP(S) traffic

PII detection: ReCon + manual validation
Privacy Attributes

PII Types:
- Password
- Gender
- Location
- Android ID
- GSF ID
- Ad ID

HTTP/HTTPS

1st Party
- instagram.com

3rd Party
- tracking.com

08/14/2012
08/16/2012
11/25/2013
04/25/2014
12/24/2015
10/24/2016
11/17/2016
12/17/2016
Case study: Pinterest

High variance in privacy risks across versions

- Sends password to a third party in 2 out of 12 versions
- HTTPS is used on Jan. 5, 2017: disclosed
- Feb. 7, 2017: fixed
- More types (gender, location, Android ID etc.) are leaked

Increased frequency for Android ID

HTTPS vs. HTTP: 3rd party vs. 1st party
Aggregate Trends in Privacy

- PII leaks can *change* substantially across versions
- HTTPS Adoption is *slow*
  Takes apps years to adopt HTTPS for supported domains
- Third-party tracking is *pervasive* and *broad*
  Evolving tracking systems: hard-coded ID => resettable Ad ID
  Permanent linkage between tracking ID and PII: >100 domains
Is Privacy Getting Better or Worse?

Combined privacy worsens over time
• mainly due to more PII types and more domains
Conclusion

• Privacy has *worsened* over time
  • PII leaks can *change* substantially across versions
  • HTTPS Adoption is *slow*
  • Third-party tracking is *pervasive* and *broad*
• Need for continuous monitoring
  • ReCon, Lumen, AntMonitor etc.
SHOULD YOU UPDATE YOUR APP?

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