Free for All! Assessing User Data Exposure to Advertising Libraries on Android

Soteris Demetriou, Whitney Merrill, Wei Yang, Aston Zhang, Carl A. Gunter
University of Illinois at Urbana-Champaign

Problem
- What’s the data exposure risk of embedding an ad library to an mHealth app?
- How can we determine such a risk?
- Can it be done automatically?

Evaluation
- In-app user data discovery performance
  - Random Forest: 88.6%, Marital Status: 93.8%, Sex: 92.9%
  - SVM: 44.8%, 66.9%, 80.9%
  - KNN: 85.7%, 93.8%, 89.9%

Utility of Pluto
- Risk Score: (higher values indicate higher privacy risk)
  - App score $z_a = \frac{x_{a} - \min(X)}{\sum_{i=1}^{n} x_i - \min(X)}$

Conclusion
- We have modeled an opportunistic ad library that aggressively collects targeted data from Android devices.
- We demonstrate that the access channels considered are realistic.
- We have designed a reliable and extensible framework that can be leveraged to assess user data exposure by an app to a library.

Pluto: in-app exposure discovery
- DB Miner
- XML Miner
- Generic Miner
- Manifest Miner
- API

Pluto: out-app exposure discovery
- Device Co-Installation Pattern Miner
- App Bundles
- Random Forest
- SVM
- KNN
- AGGR