

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

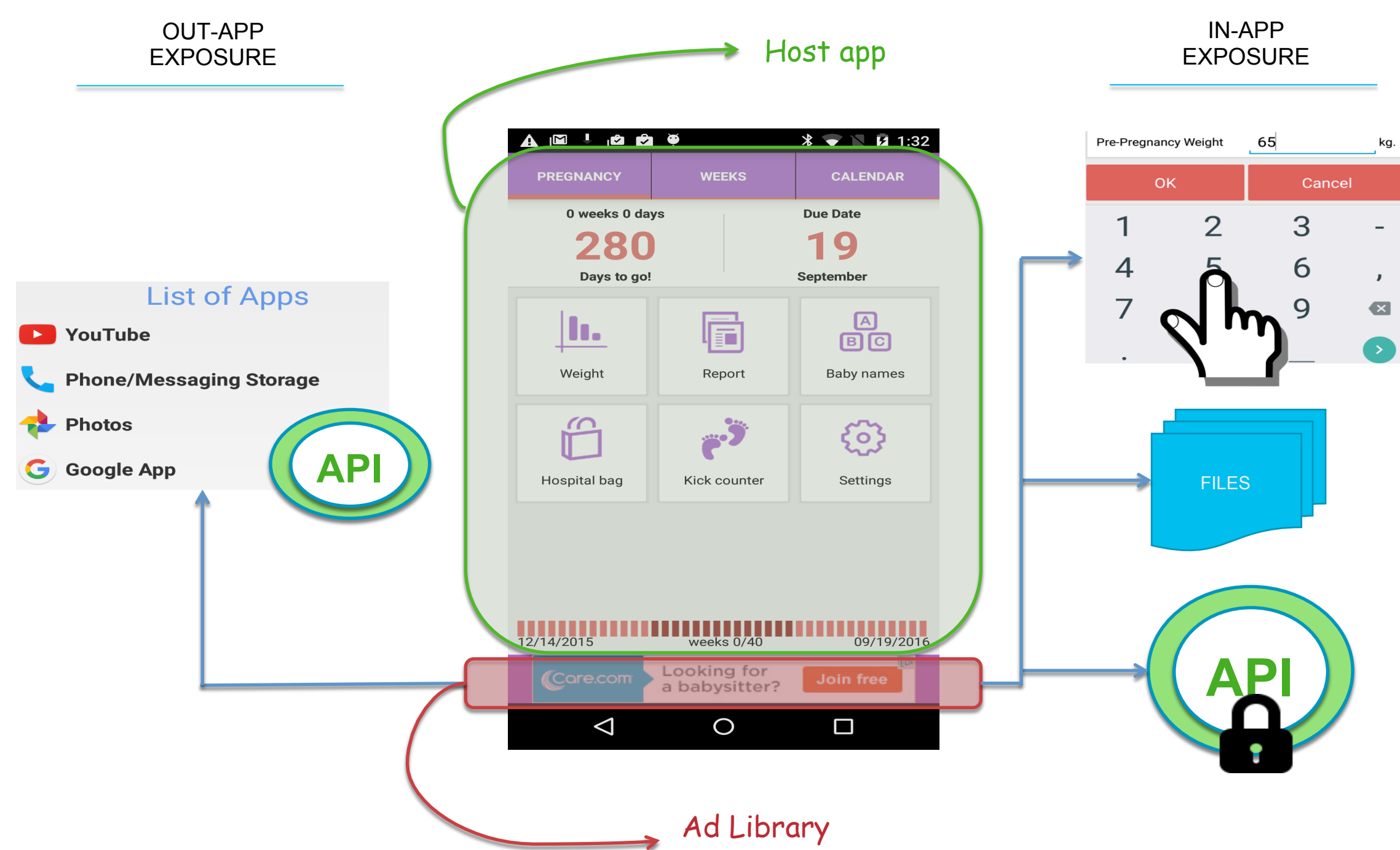


Soteris Demetriou, Whitney Merrill, Wei Yang, Aston Zhang, Carl A. Gunter
University of Illinois at Urbana-Champaign

Problem

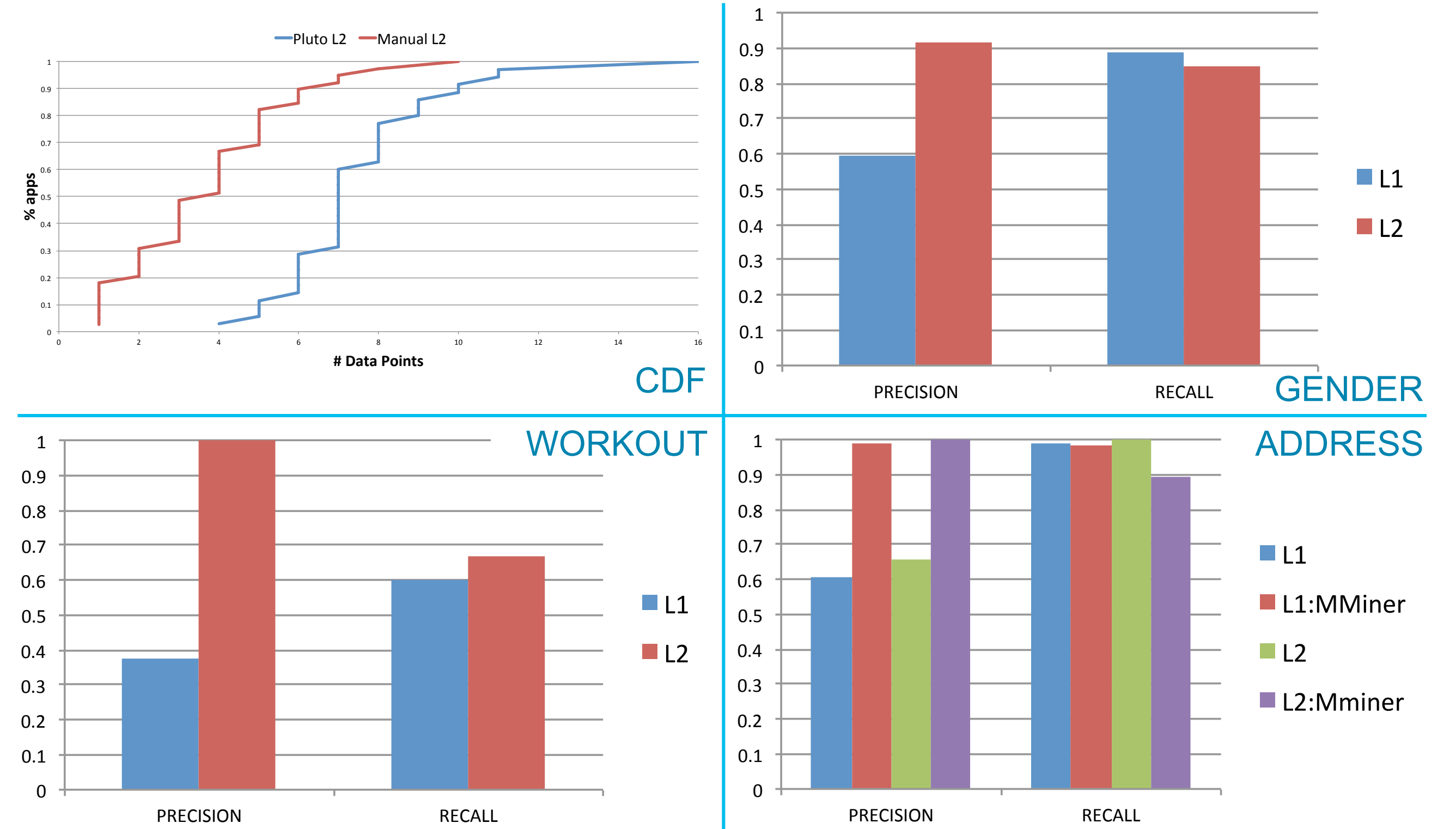
- What's the data exposure risk of embedding an ad library to an mHealth app?
- How can we determine such a risk?
- Can it be done automatically?

Approach



Evaluation

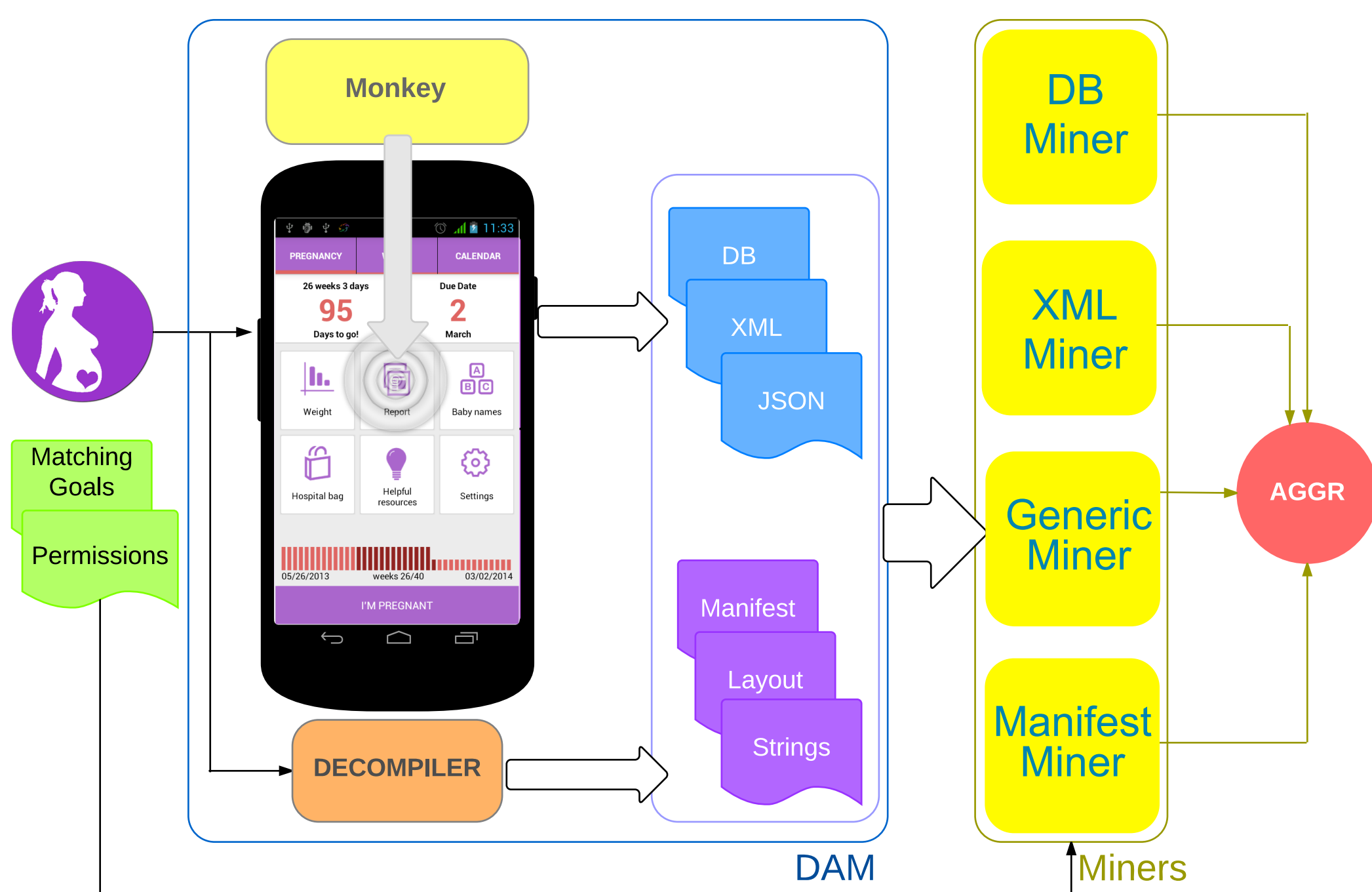
- In-app user data discovery performance



- Out-app user data discovery performance

Classifier	Age		Marital Status		Sex	
	P(%)	R(%)	P(%)	R(%)	P(%)	R(%)
Random Forest	88.6	88.6	95.0	93.8	93.8	92.9
SVM	44.8	35.4	66.9	50.5	80.9	70.1
KNN	85.7	83.6	92.5	91.2	91.6	89.9

Pluto: in-app exposure discovery



Utility of Pluto

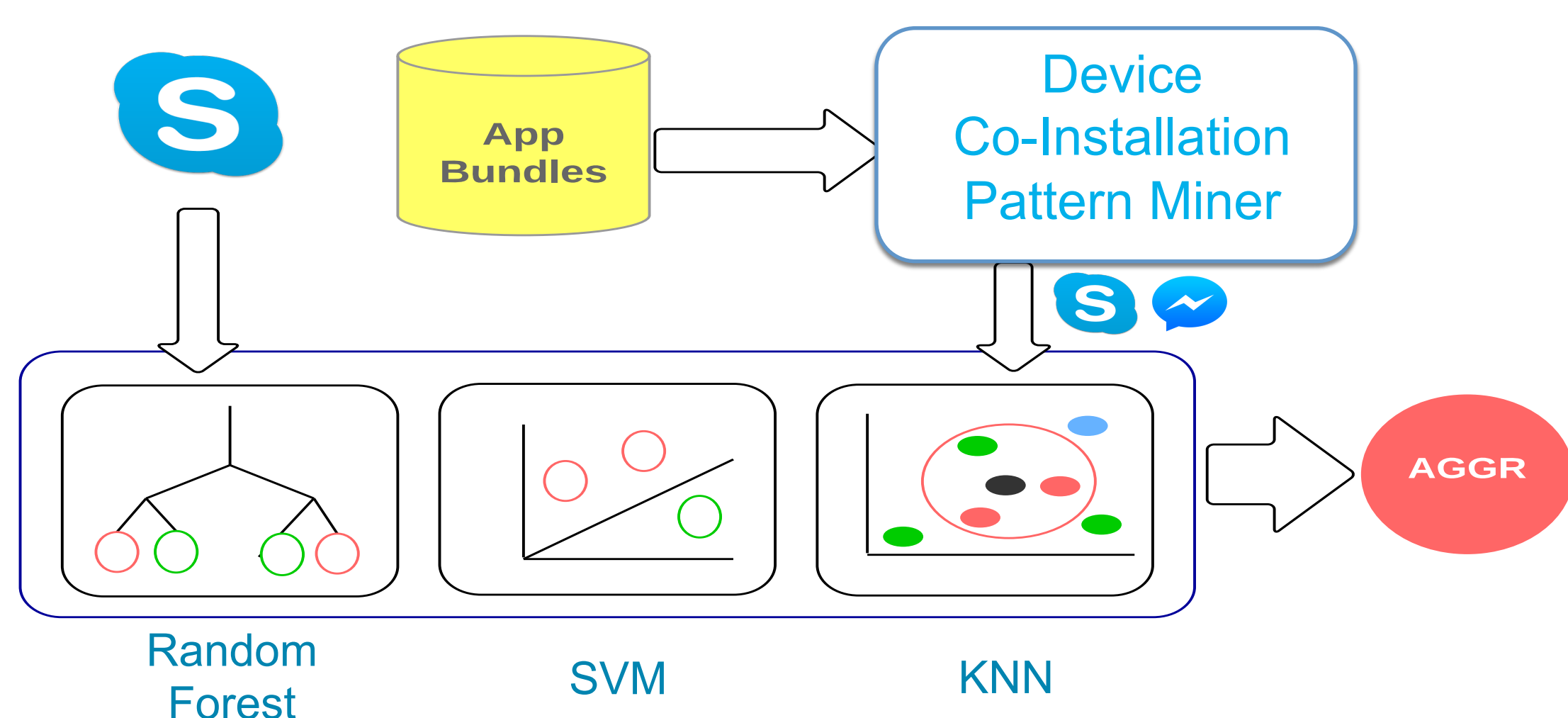
- **Risk Score:** (higher values indicate higher privacy risk)

X: set of data point values, α : app,

$$\text{App score } z_{\alpha} = \frac{x_{\alpha} - \bar{\min}(X)}{\sum_{i=1}^n x_i - \min(X)}$$

CATEGORY	APP TITLE	AVG # INSTALL/S	RISK SCORE [0 - 10]
MEDICAL	Depression CBT Self-Help Guide	100K – 500K	8.14
MEDICAL	Prognosis : Your Diagnosis	500K – 1M	6.31
HEALTH & FITNESS	Dream Body Workout Plan	100K – 500K	7.33
HEALTH & FITNESS	myCigna	100K – 500K	5.62

Pluto: out-app exposure discovery



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.

