

The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads

MAR 21ST, 2016

[论文下载](#)

Introduction

- 作者分析了Android平台上广告库的隐私泄露问题，指出广告库不仅能推测个人感兴趣的消息类别，还可能推测出用户的生活情况，例如收入，宗教信仰等，并用于针对性的投放广告。
- 在分析方法上，作者使用了真实用户作为测试对象，在不影响广告库正常运行的情况下，对广告库的个性化广告投放行为进行分析。
- 文章还讨论了如果APP开发者是恶意的，仅通过显示的广告内容来分析出用户隐私的可行性。

Methodology

定义了两类隐私信息：个人的兴趣信息以及个人的生活情况

TABLE I: Demographics distribution of subjects.

Gender		Political Affiliation		
Female	Male	Independent	Democrat	Republican
95	122	108	80	29
43.78%	56.22%	49.77%	36.87%	13.36%
Parental Status		Income		
Not a parent	Parent	< \$30K	\$30K-\$60K	> \$60K
128	89	107	67	43
58.99%	41.01%	49.31%	30.87%	19.82%
Religion				
Atheist		Non-Christian	Christian	
82		47	88	
37.79%		21.66%	40.55%	
Age				
18-24	25-34	35-44	45-54	55+
45	106	47	14	5
20.74%	48.85%	21.66%	6.45%	2.30%
Marital Status				
Single	Married	Separated, divorced or widowed		
124	73	20		
57.14%	33.64%	9.22%		
Ethnicity				
Other	Hispanic	Asian	African American	Caucasian
8	12	12	23	162
3.69%	5.53%	5.53%	10.60%	74.65%
Education				
High school or less		Associates	Bachelor	Master or higher
78		50	71	18
35.94%		23.04%	32.72%	8.30%

- 作者制作了一个APK，使用AdMob生成100次广告，每次请求都会经过VPN转发，用来过滤IP地址无关参数（测试的时候关闭地理信息服务）。测试完成时会让被试者填写自己相关信息的问卷。
- 收集所有广告指向的URL，分析meta data,分析url参数，重放了一些重定向的url。

- 对于URL对应的兴趣类别，使用google的preference接口，(对一个空白账号，重复访问某个广告链接10次)

Result

依据兴趣类型进行广告投放，结果显示，超过79%的用户有至少21%的满意度

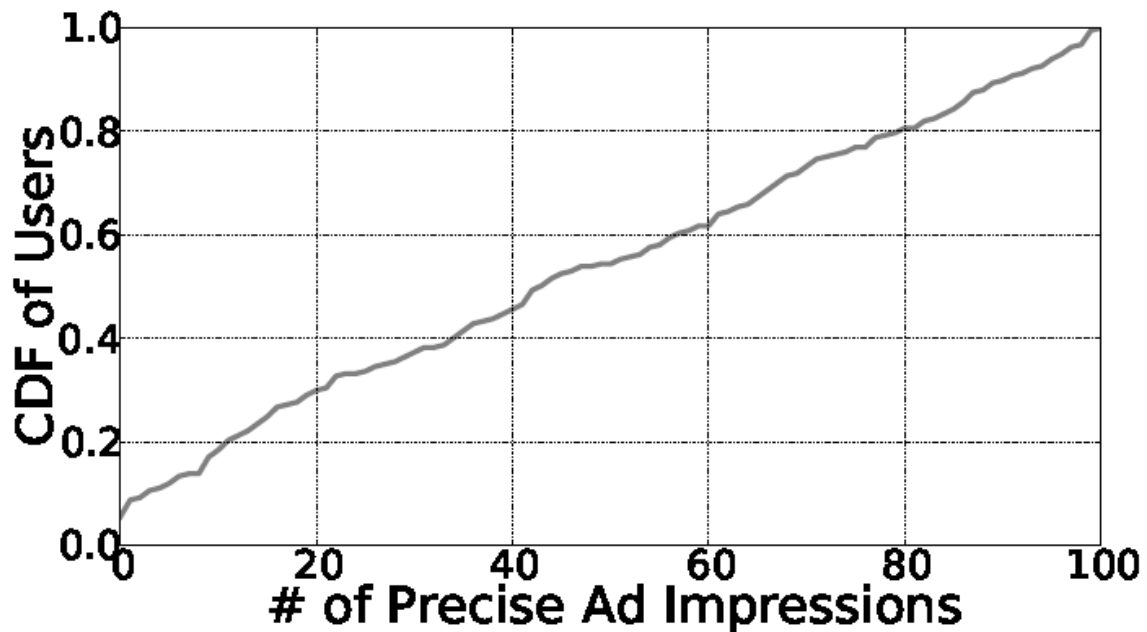


Fig. 9: Number of precise ad impressions of users.

用户的生活信息和广告投放量是否相关，结果显示，92的用户被依据生活状况而投放了广告

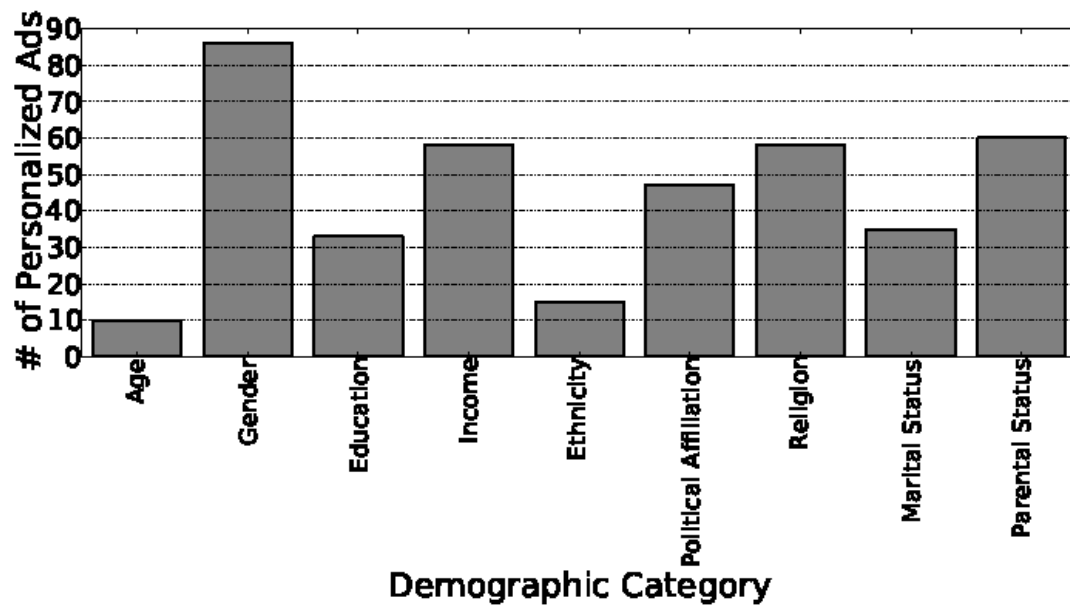


Fig. 10: Number of unique ads that are personalized based on demographics.

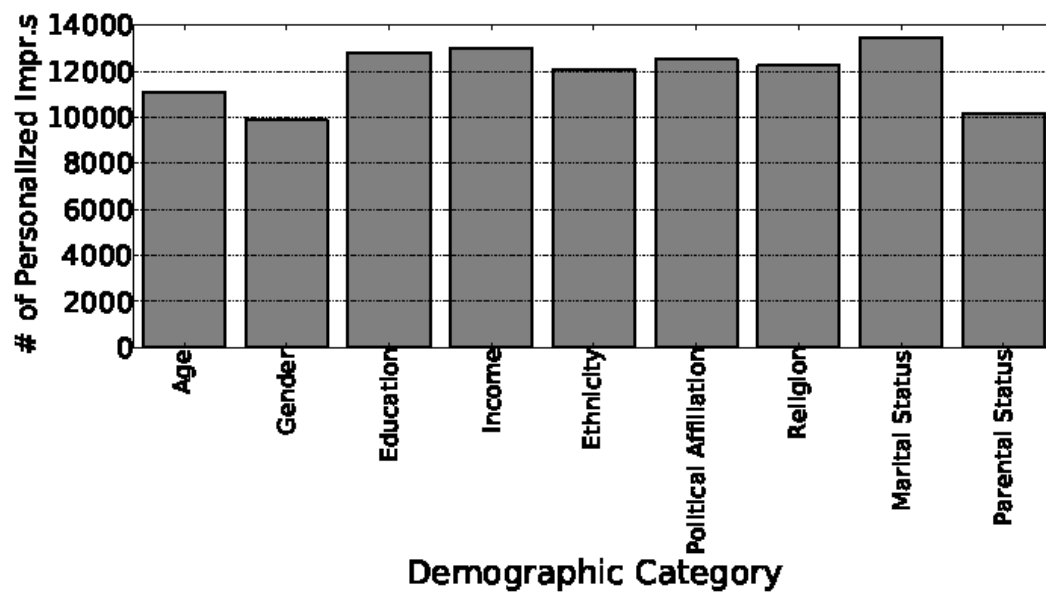


Fig. 11: Number of ad impressions that are personalized based on demographics.

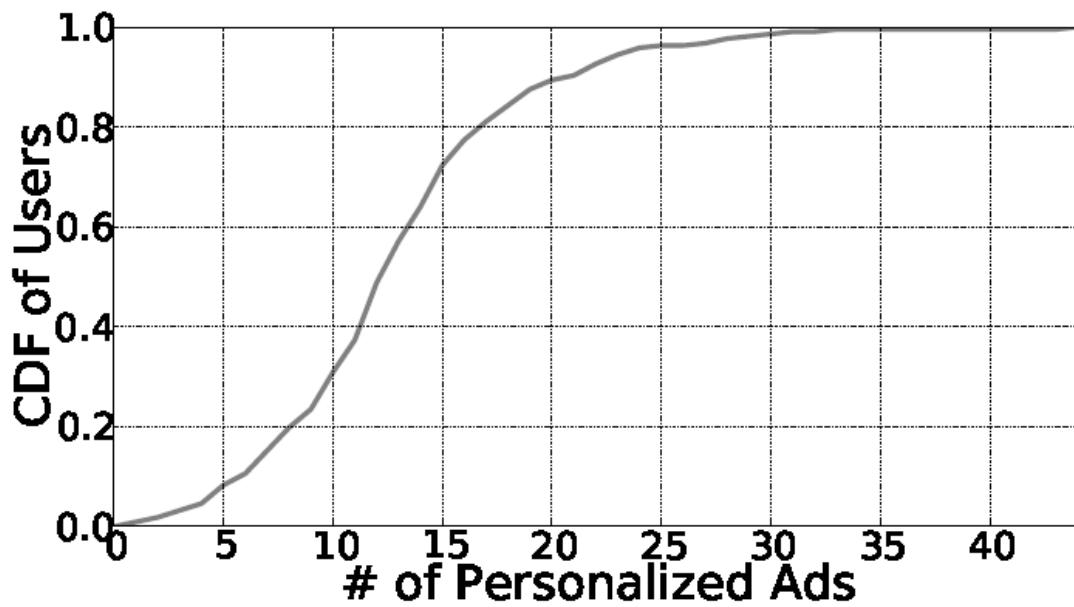
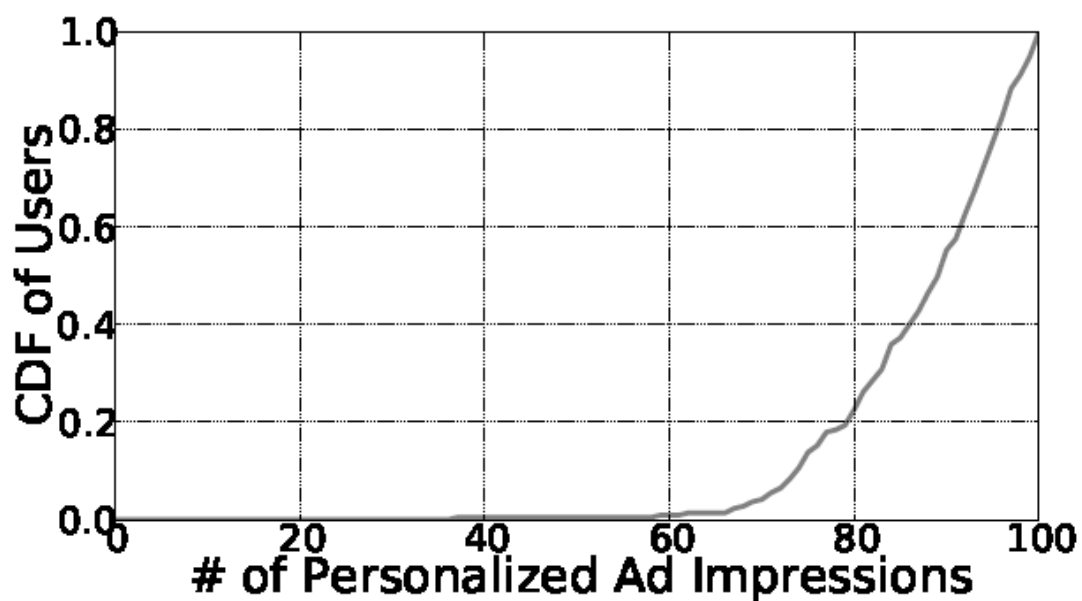


Fig. 12: Number of unique ads that are personalized based on demographics across users.



Evaluation

作者采用聚类的方法，分析能否在某类广告与某类生活情况之间建立联系。

TABLE II: Accuracy of classifiers of demographic categories.

	Age	Education	Ethnicity	Gender	Income	Marital Status	Parental Status	Political Affiliation	Religion
Decision Tree	0.51	0.30	0.76	0.64	0.47	0.62	0.62	0.50	0.35
Logistic Regression	0.38	0.37	0.72	0.73	0.45	0.62	0.65	0.50	0.39
Multinomial NB	0.37	0.35	0.61	0.73	0.36	0.49	0.65	0.41	0.43
K-Nearest Neighbors	0.39	0.34	0.75	0.65	0.45	0.47	0.59	0.45	0.40
Random Forest	0.39	0.36	0.68	0.67	0.43	0.59	0.58	0.46	0.41
SVM	0.49	0.40	0.75	0.74	0.47	0.59	0.66	0.49	0.46
Dummy	0.20	0.25	0.20	0.50	0.33	0.33	0.50	0.33	0.33