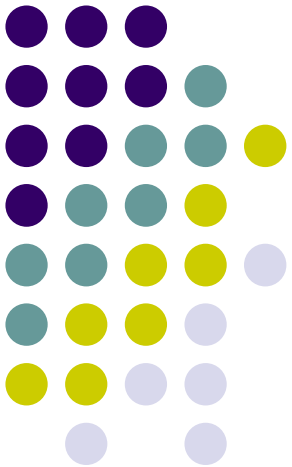
Ubiquitous and Mobile Computing
CS 528: *Information Leakage through Mobile Analytics Services*

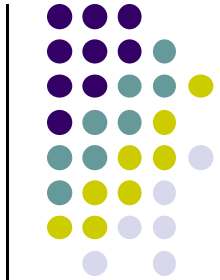
Amit Srivastava

Computer Science Dept.

Worcester Polytechnic Institute (WPI)

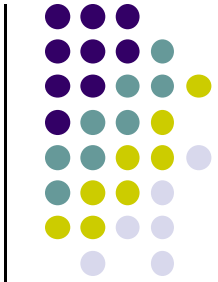


This paper is about..



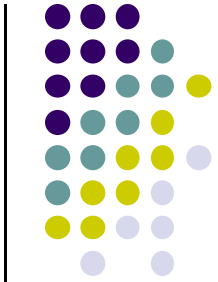
- Analytics
- User profiles and Analytics
- Profile theft /misuse
- Experimental setup
- Results
- Conclusions

Did you ever wonder ..



- Why does Facebook have Insights?
- Why did Google buy Admob?
- Why did Yahoo buy Flurry ?
- What is Adobe doing, hawking analytics tools ?

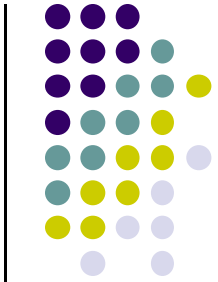
Analytics, Advertising and Revenue



- Generate insights to drive performance improvements
- Increase conversion i.e. metrics to insights, insights to actions
- Notable players in the mobile analytics include- Adobe, Apsalar, Flurry, Google, IBM, ForeSee, comScore, WebTrends*
- Collect usage data, user profile etc
- Advertisement - the only successful revenue model for mobile outside e-commerce

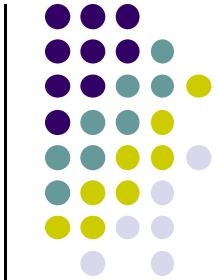
*Source: Forrester Research, Inc

User Tracking and its Dangers



- Analytics vendors create user profile based on app usage
- This profile can be compromised, as shown by the paper
- Privacy issues seen in Flurry and Google AdMob
- Compromise user identity – targeted attack

Spoofing

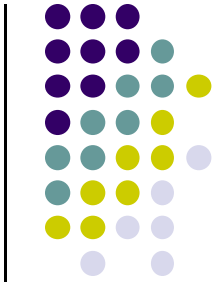


- Capturing the device-id

1. Capture usage reports from analytics vendor message over the network and extract device id
2. Or install an app for just this purpose (REALLY ?)

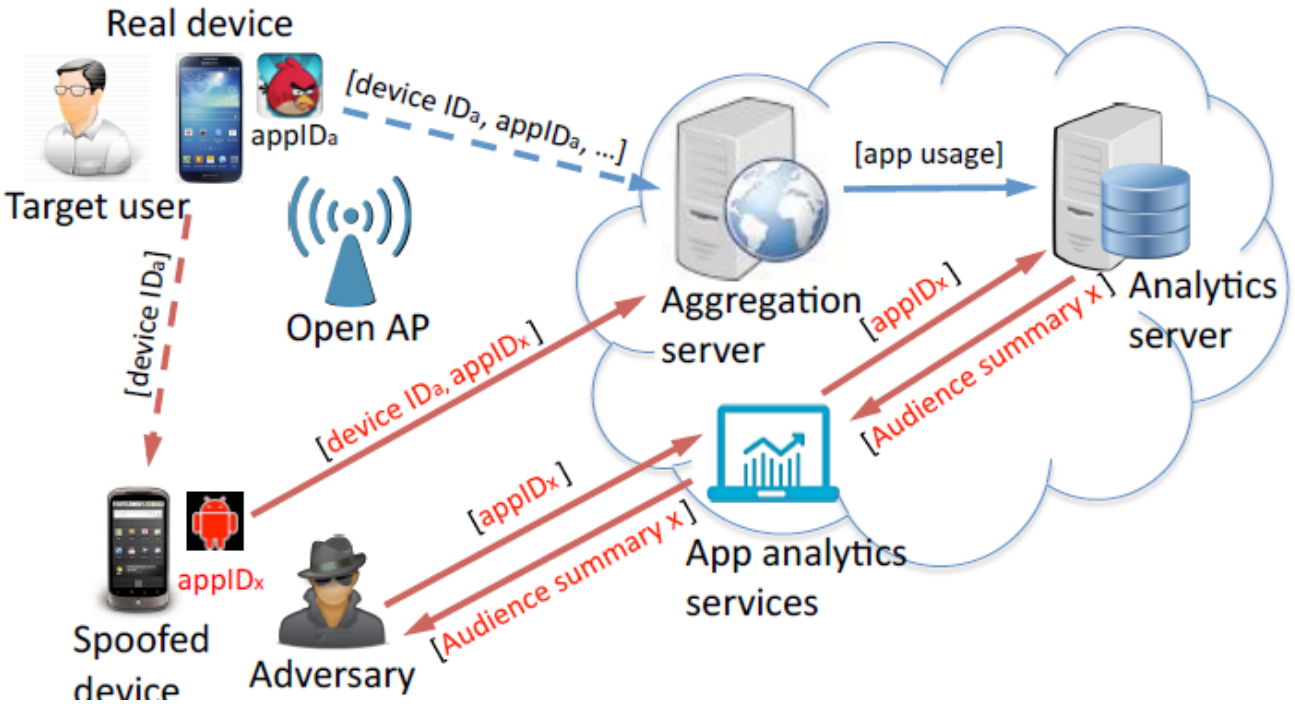
- Google hashes the device-id but other third party vendors may not do this
- Device-id access does not require user permission in an app

User Profile Extraction

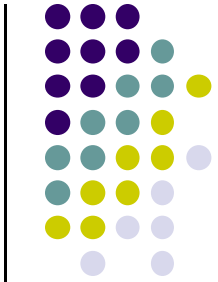


- Adversary spoofs a target device* (capture device-id)
- Uses an emulator or another device along with it
- Install apps and change usage behaviour
- Manipulates usage statistics
- Retrieve profile based on Android Id – Google AdMob
- Install a new app that uses flurry and access user profile through it

Spoofing and Device ID

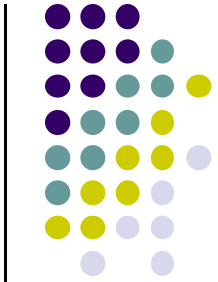


Experimental Validation



- 44 participants in 4 countries
- A custom app developed to fetch App id
- 80% users did not have Google profile
- 84% had Flurry profile
- Possibly Flurry is more widely used in apps or maybe user had more furry based apps

Ad response to hacked Profiles



- Based o App usage the analytics services create or update profile
- Verify an attack by showing high degree of certain kind of apps
- Or change usage to effect profile and hence the ads
- Flurry updated the profile in a wekkly manner
- Google updates frequently, in 6 hrs approx.
- Flurry hides ad traffic (why?) Google does not

Ad response to hacked Profiles

- Categories: Games (GA), Business(BU), Books(BO), Media(ME), Productivity (PR), Social (SO)
- A and B are unique set of ads
- Google has less unique ads but
- Compare similarity of ads shown in different categories using Jaccard index

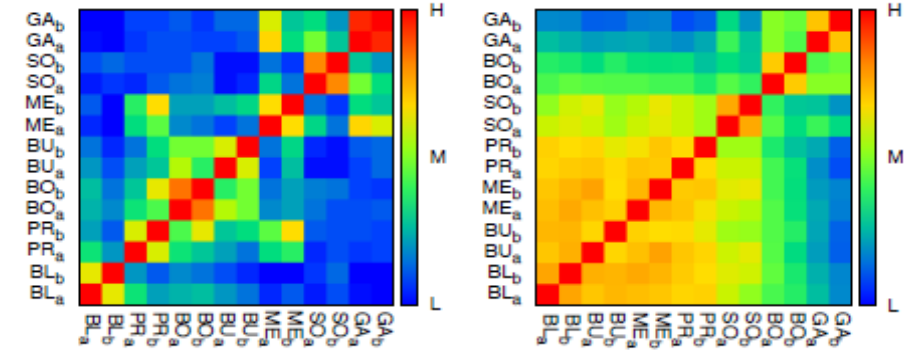


Figure Unique ads similarity between profiles, sorted by Jaccard index v.s. blank profile. (H - high, M - moderate and L - low)

Profile Category	Code	avg # unique ads (Google)	avg # unique ads (Flurry)	Jaccard index vs. BL (Google)	Jaccard index vs. BL (Flurry)
Blank	BL	42.5	212	0.645	0.92
Books & Refs	BO	106	260.5	0.32	0.704
Business	BU	148	219.5	0.275	0.8752
Games	GA	148	219.5	0.1825	0.608
Media	ME	166.5	220.5	0.235	0.8705
Productivity	PR	110	215	0.3325	0.8435
Social	SO	176	181.5	0.235	0.793

Measuring ads received by different profiles

Ad response to hacked Profiles

- 6 app categories – games, business, books, media social and productivity
- Train 2 profiles in each category, by 24 hour usage
- Collect ads from all devices
- Try Game profile to Business profile shift

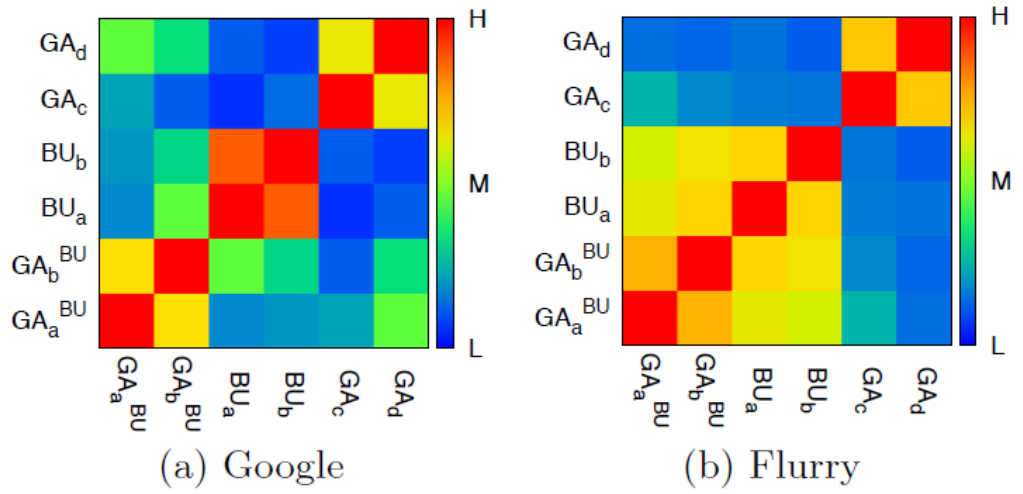
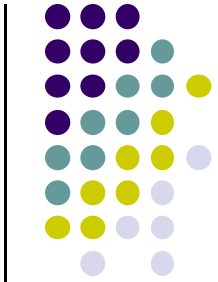


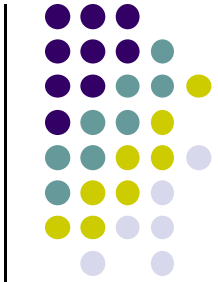
Figure 6: Unique ads similarity before and after profile perturbation. (H - high, M - moderate and L - low)

What is Jaccard Index



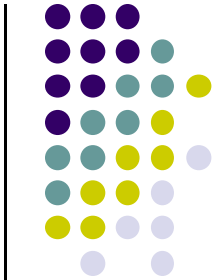
- The **Jaccard index**, also known as the **Jaccard similarity coefficient** (originally coined *coefficient de communauté* by Paul Jaccard), is a statistic used for comparing the similarity and diversity of sample sets. T
- The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets: $J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$ 0 <= J(A,B)<=1

Potential Countermeasures



- MockDroid – an android sandbox platform to test app behavior
- Pdroid –allows fine grained control over your private data usage by user/system apps
- Android and iOS should deprecate device id

Conclusion



- Ads are good for developers but bad (potentially) for you, data leakage exists
- It will take a lot of effort to impact too many people for targeted attack

References

- *https://developer.yahoo.com/analytics/*
- *https://www.youtube.com/watch?v=AewnM85Bxic*
- *https://www.forrester.com*
- *http://en.wikipedia.org/wiki/Jaccard_index*