Investigating User Privacy in Android Ad Libraries

Ryan Stevens, Clint Gibler, Jon Crussell, Jeremy Erickson, Hao Chen

20 November 2016
Research Presentation Demo

Overview

Problem Statement:

Android ad libraries are unexplored with respect to privacy. We must investigate how they work. What permissions do they use? what data do they send? what risks do they pose?

Goals:

- Examine how Android ad libraries behave with respect to privacy
- Determine whether Android ad libraries behave worse than browser ads
- Improve privacy in face of ads

Solution path:

- Download and install popular ad libraries
- Create mock apps which use them
- Examine data sent and received by ad libraries
- Examine what permissions are used by ad libraries
- Examine how ad code interacts with user code
- Offer improvements to protect privacy

Results:

- Many ad libraries use app permissions to send out private data
- Some ad libraries open an attack vector
- Some ad libraries silently grab data
- Many ad libraries enable tracking across apps and time using unique ids

Analysis of Problem

Things we know:

1. What an ad is
2. What network data is
3. App developers use ads to make money
4. App developers don't know how ad libraries work
5. Ad networks make money by selling targeted ads from advertisers
   - Age, gender, location, profiling → tracking
6. Network communication can be sniffed and tampered with
7. Ad networks must balance income (targeting) with invasiveness
8. Users can't tell how apps behave

Things we want to know:

1. How do mobile ads really work?
2. What data do they really send when getting ads?
3. What can an advertiser do with the data it gets?
4. What can a third party do with data sent?
5. How do ad libraries behave when include in apps?
6. Are ad libraries evil? Which?

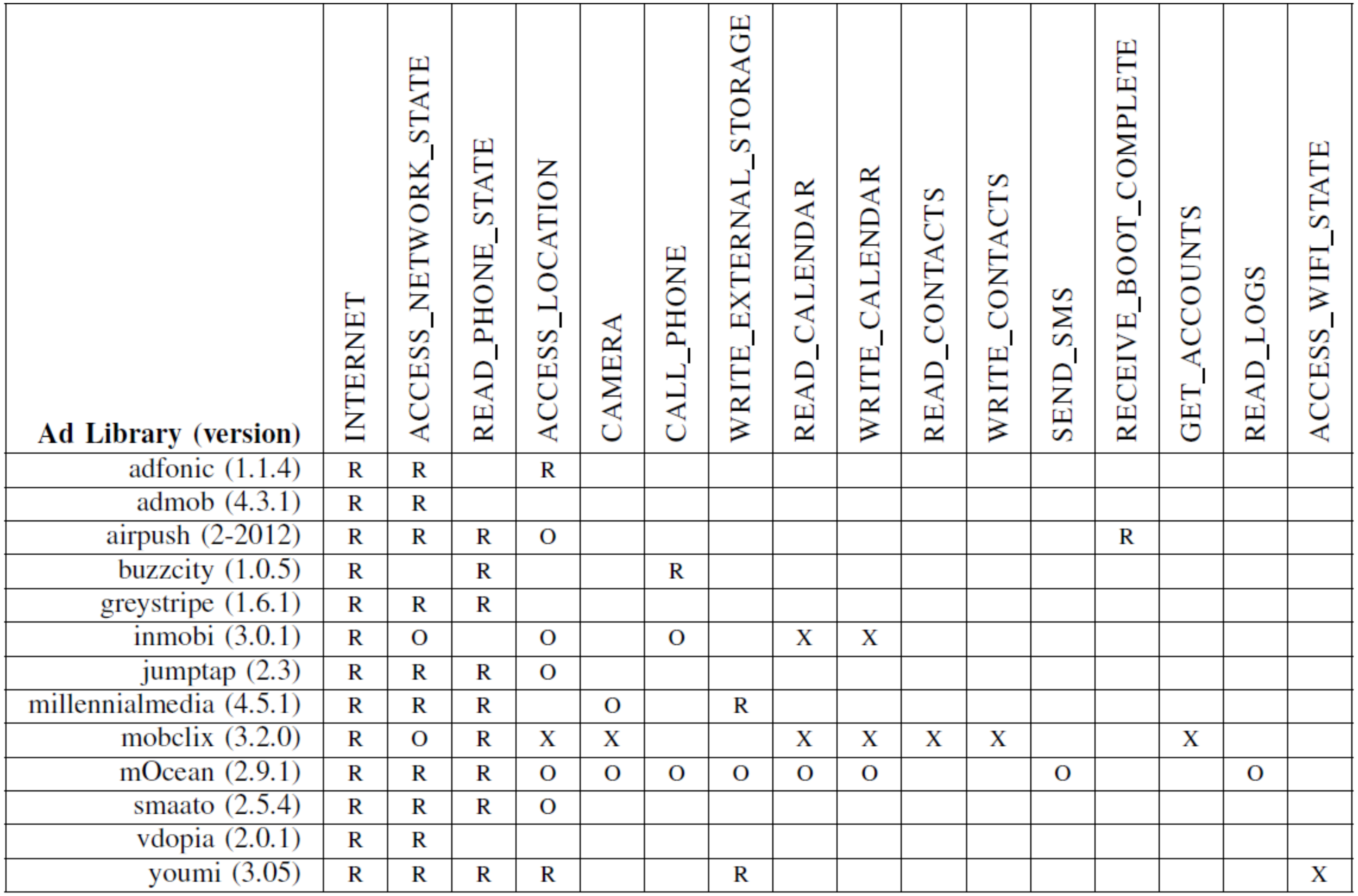
Things we want to do:

1. Ads won't go away. Let's make them safer privacy wise.
2. Prevent attackers from being able to use ad traffic for evil.

Research Methods

- On ad libraries
  - Found 13 most popular ad libraries
    - By top 500 apps on Android Market
    - By network traffic from an ISP
  - Analyzed how each worked
    - Examined documentation
    - Used Stowaway to tell what permissions the ad libraries used
    - Used a network router to track what data went out and in
    - Used *backsmali* to decompile one interesting one
  - Looked for attacks
    - Examined four ad libraries for attack vectors
- Analysis
  - Table of permission usage, highlighting bad cases
  - Table of which libraries use which unique identifiers (UIUD) and how
  - Table of what data is sent by which library
  - Offered thoughts on improving use of UIUD (hashes)

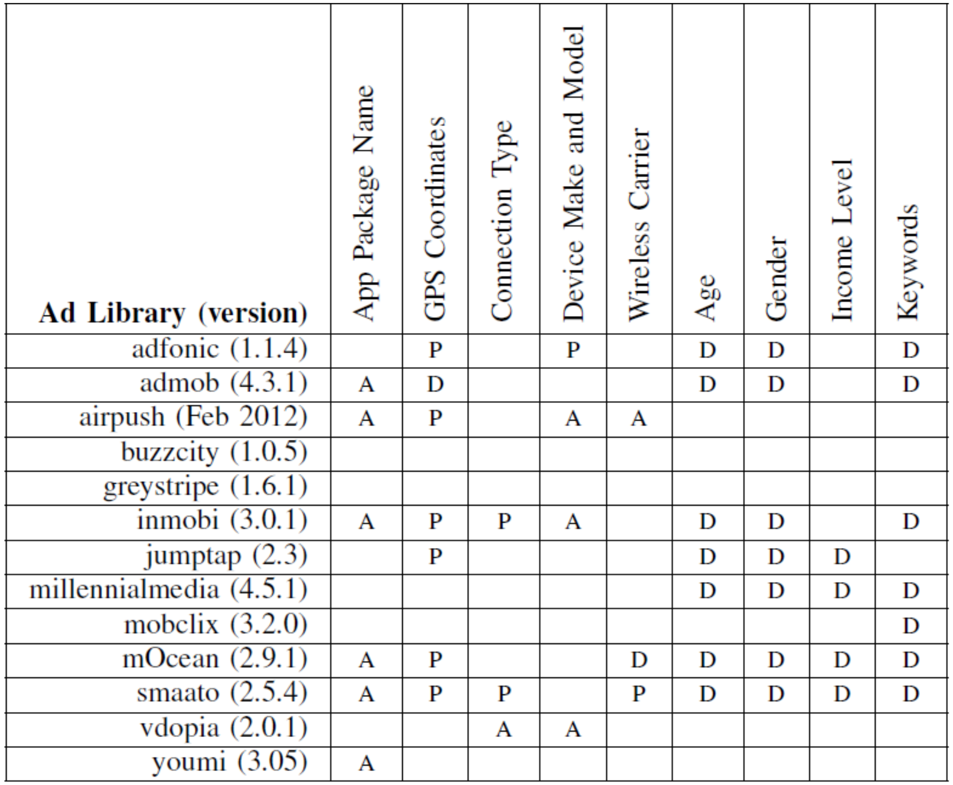
Results Table



Ad Library (version)	INTERNET	ACCESS_NETWORK_STATE	READ_PHONE_STATE	ACCESS_LOCATION	CAMERA	CALL_PHONE	WRITE_EXTERNAL_STORAGE	READ_CALENDAR	WRITE_CALENDAR	READ_CONTACTS	WRITE_CONTACTS	SEND_SMS	RECEIVE_BOOT_COMPLETE	GET_ACCOUNTS	READ_LOGS	ACCESS_WIFI_STATE
adfonic (1.1.4)	R	R		R												
admob (4.3.1)	R	R														
airpush (2-2012)	R	R	R	O									R			
buzzcity (1.0.5)	R		R			R										
greystripe (1.6.1)	R	R	R													
inmobi (3.0.1)	R	O		O		O		X	X							
jumptap (2.3)	R	R	R	O												
millennialmedia (4.5.1)	R	R	R		O		R									
mobclix (3.2.0)	R	O	R	X	X			X	X	X	X			X		
mOcean (2.9.1)	R	R	R	O	O	O	O	O	O			O			O	
smaato (2.5.4)	R	R	R	O												
vdopia (2.0.1)	R	R														
youmi (3.05)	R	R	R	R			R									X

TABLE I: *Ad SDK Permission Usage* As specified in their online documentation, ad libraries may require a permission (R) or declare it optional, but use it if available (O). Worryingly, some ad libraries check for and use undocumented permissions (X).

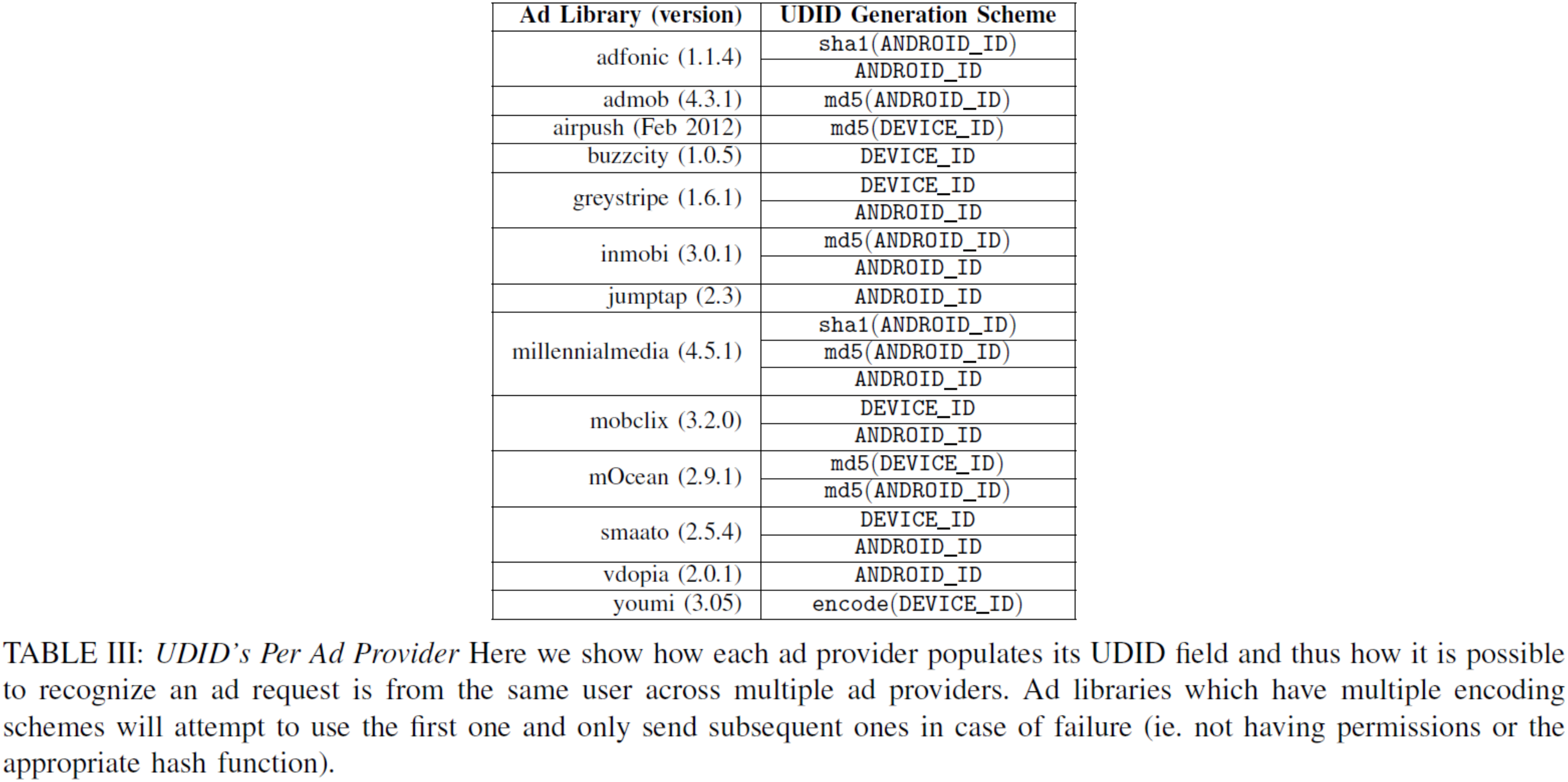
Results Table



Ad Library (version)	App Package Name	GPS Coordinates	Connection Type	Device Make and Model	Wireless Carrier	Age	Gender	Income Level	Keywords
adfonic (1.1.4)		P		P		D	D		D
admob (4.3.1)	A	D				D	D		D
airpush (Feb 2012)	A	P		A	A				
buzzcity (1.0.5)									
greystripe (1.6.1)									
inmobi (3.0.1)	A	P	P	A		D	D		D
jumptap (2.3)		P				D	D	D	
millennialmedia (4.5.1)						D	D	D	D
mobclix (3.2.0)									D
mOcean (2.9.1)	A	P			D	D	D	D	D
smaato (2.5.4)	A	P	P		P	D	D	D	D
vdopia (2.0.1)			A	A					
youmi (3.05)	A								

TABLE II: *Private Data in Ad Requests* Some fields ad libraries will always populate in ad requests (A) while others it will populate only when the application has the appropriate permissions (P), both automatically. Alternatively, some ad libraries choose to only populate fields when the developer explicitly passes the value to the library (D).

Results Table



Ad Library (version)	UDID Generation Scheme
adfonic (1.1.4)	sha1(ANDROID_ID)
	ANDROID_ID
admob (4.3.1)	md5(ANDROID_ID)
airpush (Feb 2012)	md5(DEVICE_ID)
buzzcity (1.0.5)	DEVICE_ID
greystripe (1.6.1)	DEVICE_ID
	ANDROID_ID
inmobi (3.0.1)	md5(ANDROID_ID)
	ANDROID_ID
jumptap (2.3)	ANDROID_ID
millennialmedia (4.5.1)	sha1(ANDROID_ID)
	md5(ANDROID_ID)
	ANDROID_ID
mobclix (3.2.0)	DEVICE_ID
	ANDROID_ID
mOcean (2.9.1)	md5(DEVICE_ID)
	md5(ANDROID_ID)
smaato (2.5.4)	DEVICE_ID
	ANDROID_ID
vdopia (2.0.1)	ANDROID_ID
youmi (3.05)	encode(DEVICE_ID)

TABLE III: *UDID's Per Ad Provider* Here we show how each ad provider populates its UDID field and thus how it is possible to recognize an ad request is from the same user across multiple ad providers. Ad libraries which have multiple encoding schemes will attempt to use the first one and only send subsequent ones in case of failure (ie. not having permissions or the appropriate hash function).

Related Work

- For browsers: Same Origin Policy
- On ads: Ad Split
- On making ads trust things more: Quire

Conclusions

- Some ad libraries behave badly
  - Silently taking user data and sending it out
  - Ads can perform anything the app can (no same origin policy)
  - Some ad libraries can do really bad stuff
  - Some ad libraries open attack vectors to let others do bad stuff
  - Tracking users is really easy

- Remedies
  - Use encryption! (how?)
  - Use some hashing tricks to make UIUDs more obscure
  - Separate ads from apps

Thoughts

- Hard parts:
  - Figuring out exactly what they did
  - What was their intended goal? Did they come with preconceived notions? What’s privacy?
- Interesting bits:
  - Compiling the tables comparing ad libraries
  - Digging deeper into Youmi (Chinese ad library with some pseudoencryption)
  - Showing that some ad libraries appear evil at code level
    - The relevant ad companies were defensive
- Too bad:
  - Encryption takes too much time – why would ad networks use it? They punt.
  - They blur privacy (my name, device id) with security (attackers and sniffing)
  - Don’t really offer a solution – just describe the status quo