An Exposure of Android Social Media Apps and the Data ISPs Can Collect

Daniel E. Krych, Meghan Riegel, Charles Sestito, Patrick McDaniel, dek5156@cse.psu.edu

Problem:

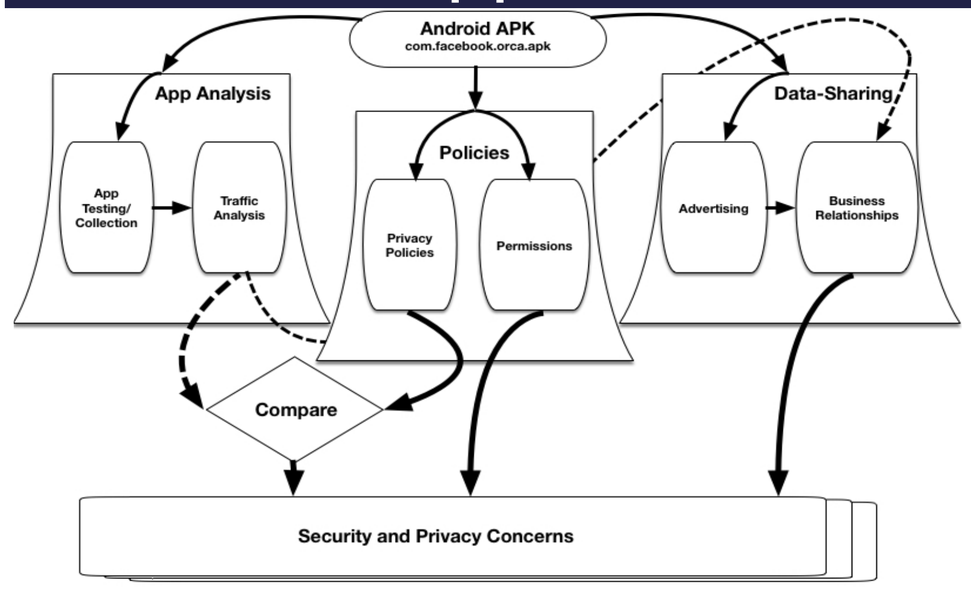
- Android apps continue to share data with third parties and transmit data unencrypted
- Social media applications use and handle sensitive/private data in their nature
- **ISPs collect and sell sensitive user data to ad companies and third parties**
- Privacy Policies have been absent, and those present lacked detail, especially in their security methods; several state data security is hard & cannot be guaranteed
- Inconsistencies between app policies and actions found through static code analysis, and dynamic analysis, but studies lacked depth or context of application

Hypothesis: Social Media apps will continue to show the trend of the collection and direct/inadvertent exposure of privacy sensitive information and lack proper disclosure.

Related Publications

- Federal Trade Commission et al. Mobile privacy disclosures: Building trust through transparency. *USA: Federal Trade Commission*, 2013.
- J Graves. An exploratory study of mobile application privacy policies. *JoTS,* Oct 2015
- J Zang et al. Who knows what about me? a survey of behind the scenes personal data sharing to third parties by mobile apps. *JoTS,* Oct 2015.
- Rocky Slavinet al. Toward a framework for detecting privacy policy violations in android application code. *IEEE/ACM ICSE,* 2016.
- Sebastian Zimmeck et al. Automated analysis of privacy requirements for mobile apps. *AAAI Fall Symposium Series*, 2016.

Technical Approach



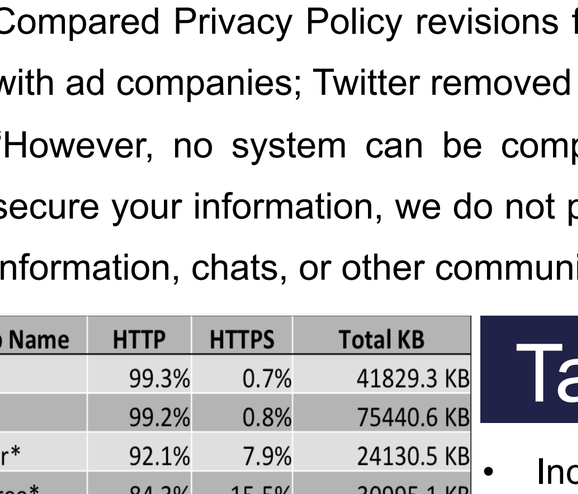
- Dynamic analysis of traffic to reveal transmission practices
- Leverage the Platform for Privacy Preferences (P3P) to analyze & compare behaviors of apps to their Privacy Policies
- Investigate permissions, advertising/analytics libraries used, and business relationships
- Determine extent of user data leakage to ISPs, third parties, and traffic sniffers
- Propose unique testbed environment / tools for testing apps

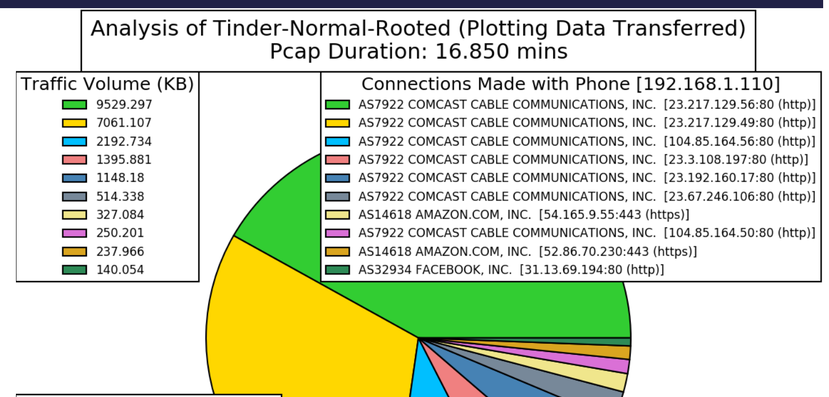
Testbed Environment:

- Android Marshmallow (6.0)
- Android Studio Emulator
- Rooted Google Nexus 6P
- Google's Android SDK
- AVD, ADB, `Monkey`
- `tcpdump`, Android `tcpdump`
- Normal, No-Login, and Login tests
- ~5-15 min each test

Analysis Tools:

- Python, `matplotlib`
- `Dshell` (USARL)
- P3P
- `Dare`, `Soot`, `AAPT`





Results

Application	Behaviors Observed Over HTTP	P3P Category	P3P Purpose	Inferred	For Ads	Disclosed	File types
Imgur	User-agent field w/ device info	6	A	✖	✖	✔	
	Referer info	7	A	✖	✖	✖*	jpeg, png, mp4, css, txt
	Content (pics) user viewed	7 OR 10	A	✖	✖	✖	
	Code files	10 OR 17	A	✖	✖	✖	
	HTTP cookies / ETag	11	A, C	✖	✖	✔	
	Leaks user interests	14	NA	✔	✖	✖	
Pinterest	User/real names used in profile pic filenames	2	A	✖	✖	✖	
	UID, unique identifier	3	F/H	✖	✔	✔	jpeg, png, ico, gif, css, json, js, php, html, txt, woff, woff2
	Device info - OS/make/model/build	6	A, C	✖	✔	✔	
	User-agent field w/ device info	6	A, C	✖	✔	✔	
	Referer info	7	A, C	✖	✖	✔*	
	Content (pics) user viewed	7 OR 10	A	✖	✖	✖	
	Code files	10 OR 17	A	✖	✔	✖	
	HTTP cookies	11	A, C, H	✖	✔	✔	
	Leaks user interests	14	NA	✔	✖	✖	
Textfree	Phone numbers/contacts	1	A, C	✖	✖	✔	
	First/last name(can be nickname), sender name(name, or number)	1 OR 3 OR 9	A, C	✖	✖	✔	
	UDID, unique identifiers	3	A, C	✖	✔	✔	
	Notification email, forgot password email	3	A, C	✖	✖	✔	
	User ID number	3	A, C	✖	✖	✔*	
	*Facebook/Google+ IDs and token	3	A, C	✖	✖	✔	
	Username	3 OR 9	A, C	✖	✖	✔	
	Oauth consumer key, nonce, sig, method, APNS notif. token	3 OR 17	A	✖	✖	✖	
	Device info - OS/make/model/build, public/private IP addresses	6	A, C	✖	✔	✔	jpeg, png, gif, m4v, html, js, jsp, json, do, php, php5, ashx, txt
	User-agent field w/ device info	6	A, C	✖	✔	✔	
	Referer info	7	A, C	✖	✔	✔	
	Account/message/call logs (msg/call ID, R/U, codec, creditBalance)	8	A, C	✖	✖	✔	
	Birthday, age, gender, language, country	9	A, C	✖	✔	✔	
	User's Profile Picture	9 OR 10	A	✖	✖	✔	
	SMS/MMS messages, message signature	10	A, C	✖	✖	✔	
	Code files	10 OR 17	A	✖	✔	✖	
	HTTP cookies	11	A, C, E	✖	✔	✔	
	Potential leak of interests	14	NA	✔	✖	✖	
	Location tracking, coordinates	15	A, C	✖	✔	✔	
	Location tracking, Wi-Fi mapping	15	A	✖	✖	✖	
	Acc Settings: isPasswordSet, ShowAds, textfreeNotificationPrivacy msgStatusPrivacy, textfreeIntercept, textfreeInterceptPhone	17	A	✖	✖	✔	
Tinder	IDFA, unique identifiers	3	C, F	✖	✔	✔	
	Unique IDs per user within pic names	3	A	✖	✖	✖	
	Device info - OS/make/model	6	C, F	✖	✔	✖*	
	Potential leak of matches	7 OR 8	NA	✔	✖	✖	
	Content (pics) user viewed	7 OR 10	A	✖	✖	✖	
	AWS (Server: AmazonS3) request ID and ID-2	8 OR 17	A	✖	✖	✖*	jpeg, html
	City, zipcode	9	A, C	✖	✔	✖*	
	Code files	10 OR 17	C, F	✖	✔	✖	
	HTTP ETags	11	A, C	✖	✖	✖*	
	Health info - sexual orientaiton	13	NA	✔	✖	✖	
	Potential leak of interests	14	NA	✔	✖	✖	
	*Location Tracking - city/zipcode	15	A, C	✖	✔	✖*	
Vine	User-agent field w/ device info	6	A, C	✖	✖	✔	
	Content (videos) user viewed	7 OR 10	A	✖	✖	✖	jpeg, mp4.jpg, png, mp4
	AWS? (Server: ECAcc) request ID, version ID, ID-2	8 OR 17	A	✖	✖	✖*	
	HTTP ETags	11	A, C	✖	✖	✔	
	Potential leak of interests	14	NA	✔	✖	✖	
Tumblr	Unique identifiers	3	A, C	✖	✔	✔	
	User-agent field w/ device info	6	A, C	✖	✔	✔	
	Referer info	7	C	✖	✔	✔	png, js, html, css
	Code files	10 OR 17	A	✖	✔	✖*	
	HTTP cookies	11	A, C	✖	✔	✔	
Twitter	User-agent field w/ device info	6	A, C, F/H	✖	✖	✔	N/A
	HTTP cookies	11	A, C, F/H	✖	✖	✔	
Messenger	User-agent field w/ device info	6	A	✖	✖	✔	.ogg
	.ogg audio file - failed call occurred	17	A	✖	✖	✖	
Facebook	User-agent field w/ device info	6	A	✖	✖	✔	N/A

- None of the apps disclosed which data was sent encrypted vs. unencrypted
- Most sent a large portion of the content viewed unencrypted: leaking usage trails, interests
- Most leaked user data (e.g. interests) further through inference attacks
- Confirmed previous studies that found evidence of the over asking of permissions and the use of a variety of advertising libraries − Bolts, Flurry, Google Ads/AdMob, Fabric
- Compared Privacy Policy revisions from testing to now − Tinder/Twitter now working more with ad companies; Twitter removed age restriction; Textfree no longer complies w/ COPPA
- "However, no system can be completely secure. Therefore, although we take steps to secure your information, we do not promise, and you should not expect, that your personal information, chats, or other communications will always remain secure." ~Tinder

App Name	HTTP	HTTPS	Total KB
Vine	99.3%	0.7%	41829.3 KB
Vine*	99.2%	0.8%	75440.6 KB
Tinder*	92.1%	7.9%	24130.5 KB
Textfree*	84.3%	15.5%	30995.1 KB
Pinterest	84.2%	15.7%	23744.2 KB
Pinterest*	75.7%	24.3%	27337.2 KB
Imgur	53.4%	46.6%	16352.1 KB
Textfree*◆	43.7%	53.7%	8026.4 KB
Textfree	23.0% 8.5%	74.1% 89.2%	4164.8 KB 7218.0 KB
Twitter	15.6%	84.3%	12250.0 KB
Tumblr	11.9%	88.0%	24513.6 KB
Messenger	5.8%	93.6%	1183.0 KB
Facebook	<0.1%	98.8%	5442.3 KB

*** App Traffic collected on rooted Google Nexus 6P**
◆ Same APK and version retested after 10 months

Takeaways

- Inconsistencies found between social media app's Privacy Policies and the app's behaviors
- Policies still don't detail how data is transmitted or secured
- Every unencrypted packet can be collected at any hop
  → ISPs, third parties gathering and selling user data
- Several Privacy Policies now have their own definition of 'Personal Information' or 'Sensitive Data'
- A few apps leak a lot of sensitive data, others use more encryption and protect user data, none detail security

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). Additionally sponsored by NSF Grant Nos. CNS-1228700 and CNS-1064900