Assessing Privacy in Social Media Aggregators

Gaurav Misra
Security Lancaster
School of Computing and Communications
Lancaster University, UK
Email: g.misra@lancaster.ac.uk

Jose M. Such
Department of Informatics
King's College London, UK
Email: jose.such@kcl.ac.uk

Lauren Gill
School of Computing and Communications
Lancaster University, UK
Email: l.gill2@lancaster.ac.uk

***Abstract*—Social Media Aggregator (SMA) applications present a platform enabling users to manage multiple Social Networking Sites (SNS) in one convenient application, which results in a unique concentration of data from several SNS accounts in addition to the user's mobile phone data available to them. We describe a three-step methodology to assess how privacy is considered in these applications: 1) We inspect the mobile data and social media data; 2) we study any privacy policies and their compliance with respect to distributor's vetting policies; and 3) we perform a qualitative assessment of traceability between privacy policies and the actual transparency and control mechanisms offered to users by the apps' interfaces. We then present the results we obtained for 13 popular SMAs from 3 app stores, showing a variation in data accessed by the individual applications, an absence of privacy policies for 5 of the SMAs evaluated, and a lack of traceability between privacy policies and transparency and control of interface operations. After this, we report our experiences using the methodology and the lessons learned, together with potential future work to improve the methodology and its potential to also assess privacy in other mobile applications that also connect with social media.**

***Index Terms*—Social Media Aggregators, Social Media Privacy**

I. Introduction

It is evident that our engagement with Social Networking Sites (SNS) is becoming ever more ingrained in our daily lives. This has been, in part, facilitated by the spectacular growth of mobile social networking, which has a worldwide penetration of 23% (1.7 billion). This proliferation of mobile devices have enabled the users to access social media accounts with more ease and convenience. This is demonstrated by the huge surge in usage of social applications on mobile platform to the extent that an estimated 80% of time spent on social media is using mobile applications[1].

This shift towards the mobile platform for social media activity has led to the development of Social Media Aggregators (SMAs) which enable users to access all of their social media accounts from a single application. This is partly driven by the fact that users are often found to have accounts on multiple Social Networking Sites (SNSs)[2]. It can be quite attractive to users to use SMAs, a single application for all social media accounts, compared to installing separate applications for all their social media sites. An additional attraction of installing a single SMA replacing all social media applications is also related to better utilization of the often limited resources (RAM, CPU power and battery) of the mobile phone itself. Indeed, many SMAs clearly convey this to potential customers as an advantage and a selling point[3].

While it is clear that SMAs can be beneficial for users, they also potentially introduce severe privacy risks for users. Users are meant to use SMAs to combine multiple social media accounts and all the activity is routed through a single SMA. This is different from using separate applications for different social media accounts as a user's Facebook application, for example, cannot access their Twitter activity unless an explicit link is made by the user. Such a link between various social media profiles is implicit in the case of SMAs. Moreover, this information about social media activity is augmented with mobile device data such as GPS location, contact lists, camera, etc. Given this potential threat to the privacy of social media users, it is essential to take a closer look at the transparency and control mechanisms offered by these applications. This understanding will help further in-depth analysis of gaps in policy and technology which are required to be overcome in order to safeguard user privacy and enable appropriate usage of SMAs.

In this paper, we describe a three-step methodology to assess how privacy is considered in these applications. We begin by looking at the *Data Permissions* requested by SMAs. This includes both mobile data as well as social media data of the user. We then check whether the SMAs have relevant *Privacy Policies* or other related documentation which explain the collection, usage and purpose of the user data being collected by them. Then, we qualitatively analyze the privacy policies and perform a *Traceability Analysis* where we evaluate whether the interface provided to the users are congruent with documented policies to evaluate how transparent data collection is and whether users have a control over the amount and nature of data being collected.

We report the results we obtained for 13 popular SMAs from 3 app stores, showing: a variety in the data accessed, especially when it comes to mobile data; a partial lack of privacy policies (5 out of the 13 SMAs do not have privacy policies); and that a substantial proportion (45%) of SMAs show *Broken* traceability between policy documentation and

[1]http://marketingland.com/facebook-usage-accounts-1-5-minutes-spent-mobile-171561

[2]http://www.pewinternet.org/2013/12/30/social-media-update-2013/

[3]https://play.google.com/store/apps/details?id=com.friends.socialnetworkingsites

interface operation whereas *Complete* traceability is observed in about 19% of the cases. We also report our experiences using the methodology, together with lessons learned and potential future improvements to the methodology.

II. Methodology

We begin by listing the various SMAs we have considered in our research along with their sources. We have surveyed 13 popular SMAs for this research. We studied the 6 most popular SMAs (in terms of reviews and installs) each from *Google Play Store* and *iTunes*. Additionally, we included a *Cydia* SMA to account for the variation between SMAs with different levels of adoption as well as between different app stores that have different vetting procedures or policies (e.g., Cydia only works on rooted iOS devices and does not have a vetting process in place). The SMAs are listed with their platform, number of times they have been rated and the number of times they have been downloaded (wherever available)[4] in Table I. Note that number of reviews was not available for Social Butter and Social hub as there were not enough reviews for iTunes to publish the number.

TABLE I: The 13 SMAs evaluated,the app stores they belong to, number of reviews and downloads when available.

SMA	Platform	No. of Reviews	Installs
iSocial	Cydia	–	–
Hootsuite	Google Play	80760	1000k - 5000k
Buffer	Google Play	24948	500k - 1000k
Social Networking all in one	Google Play	18336	1000k - 5000k
Social Media all in one	Google Play	11106	1000k - 5000k
Everypost	Google Play	4502	100k - 500k
Social Media	Google Play	1392	100k - 500k
Hootsuite	iTunes	4865	–
Buffer	iTunes	1150	–
Everypost	iTunes	138	–
Social Media Vault	iTunes	12	–
Social butter	iTunes	N/A	–
Social hub	iTunes	N/A	–

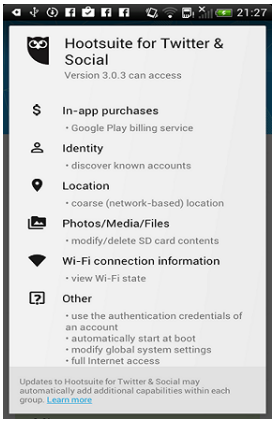


Fig. 1: Mobile data access permissions required by Hootsuites on Android device

A. Examining Data Permissions

The first step of our analysis requires us to identify exactly which SMAs request permissions to access personal data from the user. All mobile applications are required to request permission for the data they access on the user's phone. We compare the permissions requested by the 11 SMAs included in our analysis. It is important to note here that applications asking for permissions of any data from the user does not mean they are actually accessing it. However, it means that this data is available to them with the consent of the user (demonstrated by granting the access permission while using the application).

Most applications have a "permissions screen" which is shown to the user to communicate the list of data access permissions requested by the application (refer to Fig. 1). However, for the analysis, in addition to the permissions screens, we also looked at the phone settings section for the individual permissions the applications were using. Both Android and iOS display the data access permissions for each application installed on the mobile phone. We also checked the permissions granted to individual SMAs by using "Permissions Manager" application on Android devices. We examine the social network data (such as profile information, communication, lists, etc.) that are accessed by the SMAs separately. This helps us understand exactly what information each SMA will try to have access for each of the SNS the user will associate to the SMA. To look at this, we created social media accounts and then authorized the individual SMAs. We then checked the social media site to see what permissions the SMA had been granted. The permissions can also be checked by the user when the SMA is used to log in to a particular social network account for the first time. Only permissions which were specified explicitly in either the permission screen or the phone settings (or seen using "Permissions Manager" on Android SMAs) were included in our results.

B. Privacy Policies

The next step in our analysis was to examine the privacy policies of the individual SMAs. In some cases, the relevant document was titled differently (such as "Terms of Service") but we refer to all privacy related documentation as privacy policies for simplicity. The aim of this evaluation was to check for compliance with distributor vetting policies.

The 3 app stores included in our research are:

1) **Cydia**: It does not have an official vetting process for its applications.
2) **iTunes Store**: It has a vetting process which reviews all applications.[5] Personally identifiable information may not be collected or used without the user's consent. More generally, privacy policies are required if an application stores, shares or uses personal data.

[4]These figures were found from the respective app stores and are accurate as of 9th February, 2017. Please note that Apple does not publish official statistics about number of downloads for individual iOS applications so this information is absent from the table. Statistics for iSocial could not be found as well.

[5]https://developer.apple.com/app-store/review/guidelines/

3) **Google Play Store**: It has a vetting process which looks at app permissions[6] and outlines the application provider agreement to protect the privacy and legal rights of users.[7] If an application accesses registration or personal information, users must be made aware of this, and an adequate privacy policy must be provided in appliance with the law.

C. Mapping Traceability

Finally, we performed a qualitative analysis of the privacy related documentation to facilitate the traceability analysis with transparency and control interface operations. Previous research has identified a methodology for analysing software requirements from privacy policies [1]. Concepts, categorized as a commitment, privilege or right, are attained from statements by identifying helping verbs, and used to produce a set of software requirements. Similarly, we use content analysis to identify action statements through verbs that we then categorize into privacy implications, which are split into categories by way of answering the following questions:

1) What information is **collected** by the application?
2) What is the **purpose** of collection?
3) Who can **access** this information?
4) How long is information **retained**?

These privacy implications help us in contextualizing the traceability analysis. In particular, we map the extent to which application features and controls match expectations set out to users as data actions in privacy policies or application interfaces. By measuring the traceability of privacy policy implications in application content, we can assess the extent to which data transparency and control are delivered to the user.

For those applications with privacy policies, information provided in these documents present a means of gathering expectations for this analysis. A method for traceability analysis of SNS is presented by Anthonysamy, et al. [2] where action statements identified in privacy policies are mapped to those in interface operations by way of assessing the extent to which data actions are controllable by users. We applied a similar methodology to SMAs and extended it to consider mobile phone data and the transparency of interface operations. In Anthonysamy's methodology, privacy implications found in policies are matched to corresponding operations available through interfaces during installation and use of the application. We have defined actions of privacy policies as privacy implications, and define features and controls of an application as its operations. Also, and extending upon Anthonysamy's methodology, our study aims to identify the traceability of data privacy implications through interface awareness mechanisms. Therefore we assess the transparency of data actions through interface operations, as well as controls.

For SMAs with privacy policies, transparency of data usage is analyzed, mapping information provided in the privacy policy, to that presented through application operations. Traceability between data actions and the extent to which we control each privacy implication is the second aspect for analysis. In this way we map privacy implications to data transparency and control operations for SMA applications with privacy policies, by carrying out the following steps.

For each privacy implication identified:

1) Identify a corresponding interface operation by matching terminology of data actions.
2) Assess the transparency of data actions made visible to the user through interface operations, *contrasting data actions in privacy policies*.
3) Assess the extent of user control on data actions through operations, mapping data visible in the previous step (2) with control operations.

We measure the extent to which privacy implications are transparent and controllable through user interfaces against three main categories; complete, partial and broken in a similar way as in Anthonysamy, et al. [2], but specifying the categories both for transparency and control:

Complete mappings signify complete transparency of information presented to the user, through both transparency and control operations. Information presented to users is unambiguous; with unmistakable meaning and appropriate detail. For transparency, complete traceability can be achieved by providing accurate information to the user through the user interface. An example is when a user is accurately informed about all data being accessed by an app through the permission screen. The control operation is mapped as complete when the user can regulate this list and can choose to withhold certain items of information.

Partial mappings involve ambiguous information provided in privacy documentation or data operations. For example, vague terms like 'personal information', which are not explicitly defined, make mapping data operations difficult. Access permissions are partial data operations because they do not inform users of all data collected. Hootsuite collects location and traffic data, much like most other applications. Although we are prompted for permission regarding location access, the application does not provide any information on the user of traffic data collection. Control over a privacy implication is found to be partial when incomplete, with some control provided but not all data collected have associated controls. Taking Everypost as another example, we find partial control operations are evident for traffic data collected. Everypost's privacy policy[8] states that cookies used by third parties may be opted out of, as is apparent through interface operations. However, collection of traffic data for internal usage such as analytics does not match any control operations.

Broken mappings occur when there is a disconnect between privacy implication expectations and application operations. Control operation mappings are broken when documented expectancies and/or data transparency operations do not have a matching control. Detachment from policy expectations is

[6]https://support.google.com/googleplay/answer/6014972?hl=en-GB&ref_topic=6046245

[7]https://play.google.com/about/developer-distribution-agreement.html

[8]http://everypost.me/privacy-policy/

TABLE II: Mobile data accessed by each SMA

SMA	Identity	Photos /Media	Location	Contacts	Wi-Fi	Camera	Mic	Device ID & Call info	SMS	Phone	Network Access	In App Purchases	USB Storage
iSocial	✓	–	–	–	–	–	–	–	–	–	✓	–	–
Hootsuite	✓	✓	✓	–	✓	–	–	–	–	–	✓	✓	✓
Buffer	✓	✓	–	–	–	✓	–	–	–	–	✓	✓	✓
Social Networking all in one	–	–	✓	–	✓	–	–	–	–	–	✓	–	–
Social Media all in one	–	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
Everypost	✓	✓	✓	✓	✓	–	–	–	–	–	✓	✓	–
Social Media	–	–	✓	–	–	–	–	–	–	–	✓	–	–
Hootsuite	✓	✓	✓	–	✓	–	–	–	–	–	✓	✓	✓
Buffer	✓	✓	–	–	–	✓	–	–	–	–	✓	✓	✓
Everypost	✓	✓	✓	✓	✓	–	–	–	–	–	✓	✓	–
Social Media Vault	–	–	✓	–	–	✓	–	–	–	–	✓	–	–
Social butter	–	✓	✓	–	–	✓	–	–	–	–	✓	–	–
Social hub	–	✓	✓	✓	–	–	–	✓	–	✓	✓	–	–

Key: **Yes**: ✓ **No**: –

apparent among privacy implications such as advertising and aggregation. These purposes for data collection are expressed in privacy policies but no corresponding information is provided through application data or control operations. Likewise implications of age restriction in concern to data retention are expressed in policies with disconnect to interface operations.

There are many cases in which there is an absence of a clear traceability mapping between privacy implications and interface operations. We have classified these applications as **Unknown** and represented them in our analysis.

Apart from the above 4 classifications, there are some cases where the privacy implication was not applicable to a particular SMA. In such cases, we have represented this as **N/A** in our analysis. The detailed results of our analysis is presented in section 4.3.

III. Results

A. Data Access Permissions

1) Mobile Data Access Permissions: As can be seen from the results in Table II, most applications require access to photos/media, location, identity, which refers to any user accounts on the phone accessed by the application, and network access. In addition, many application require access to the USB storage as well. These findings confirm that personal data of the user is accessed by most of the application that were analyzed. An interesting observation is that permissions seem consistent for the same SMA developers across app stores. However, for different SMAs we observe a wide variety in the mobile data being accessed. While this could be attributed to different functionality being provided, it may also be a sign of some SMAs asking for more permissions than required [3], as arguably one of the most mature and used SMA (Hootsuite) seems to use a relatively smaller set of permissions when compared to other SMAs. An interesting case is that of Social Media all in one, which seems to access everything except Identity (which could be retrieved from the SNSs accessed anyway).

2) Social Media Data Access Permissions: SMAs are different from other mobile applications as they can access a user's social media data as well. We have summarized the data permissions requested by SMAs while a user logs into their social media accounts in Table III. We have used general terms such as "Activity" and "Lists" in this table to simply convey the meaning as each social media site uses different names for such features. For example, "posts" on Facebook and "tweets" on Twitter as well as inbox messages are classified under "Activity". Similarly, "Lists" refers to groups or lists that the user might have created (or used by default) to organize their contacts on various social media sites.

We can find in Table III that 5 SMAs, namely, iSocial, Social Networking All in One, Social Media all in one, Social Media and Social Media Vault are marked with a ' * ' sign and are shown to access all social media data. This is to highlight the fact that these applications do not disclose what social media data they access to function as they just provide an interface for either the social media apps (such as Facebook, Twitter) already installed on the user's phone or to the web link of the social network via the web browser. As all the social media activity goes via these applications, they have the potential to access all communication. Moreover, these applications do not require to be authorized by the user with their Facebook account so the user cannot regulate the permissions by logging into their Facebook account as is possible with other Facebook applications. For the other SMAs, we find that many of them access almost all social media activity such as posting on walls/tweeting, access the friend or contact lists, update the profile on the users' behalf, post on their behalf, access to inbox messages or the email ID which was used to create the account. Needless to say, all this information may be classified as personal and sensitive to the user and we find that most applications who disclose the permissions access this information.

B. Application Privacy Policies

Applications that collect personally identifiable information are required to produce a privacy policy in order to comply with the previously discussed distributor vetting policies. Table IV shows that 8 out of the 13 SMAs that we evaluated were found to include this documentation. The lack of privacy policies among the other 5 SMAs seems to suggest a vio-

TABLE III: Social media data accessed by each SMA

SMA	Activity	Lists	Update Profile	Post	Messages	Email ID
iSocial*	✓	✓	✓	✓	✓	✓
Hootsuite	✓	✓	✓	✓	✓	✓
Buffer	✓	✓	✓	✓	✓	✓
Social Networking* all in one	✓	✓	✓	✓	✓	✓
Social Media* all in one	✓	✓	✓	✓	✓	✓
Everypost	–	✓	–	✓	–	✓
Social Media*	✓	✓	✓	✓	✓	✓
Hootsuite	✓	✓	✓	✓	✓	✓
Buffer	✓	✓	✓	✓	✓	✓
Everypost	–	✓	–	✓	–	✓
Social Media Vault*	–	✓	–	–	–	✓
Social butter	–	✓	–	–	–	✓
Social hub	✓	–	–	–	–	✓

Key: **Yes**: ✓ **No**: –

TABLE IV: Whether privacy policies are provided by each SMA provider

SMA	Privacy Policies
iSocial	✓
Hootsuite	✓
Buffer	✓
Social Networking all in one	–
Social Media all in one	–
Everypost	✓
Social Media	–
Hootsuite	✓
Buffer	✓
Everypost	✓
Social Media Vault	–
Social butter	–
Social hub	✓

Key: **Yes**: ✓ **No**: –

TABLE V: Traceability mappings represent transparency and control of privacy implications respectively, or collectively.

	Social hub	Hootsuite	Hootsuite	Buffer	Buffer	Everypost	Everypost	iSocial
Collection								
Mobile Data	◐●	◐	◐●	◐	◐●	◐	◐●	●
Social Media Data	◐	◐	●◐	–	–	◐	◐	●
Traffic Data	✖	✖	◐	✖	✖	◐	◐	✖
Purpose								
Services	✖◐	●	●	●	●	●	●	●
Internal use	✖	◐	◐	✖	✖	✖	✖	✖
Asset transfer	?	✖	✖	●?	●?	✖	✖	✖
Advertising	◐●	✖◐	–	–	–	–	✖◐	✖
Aggregation	✖	✖	✖	✖	✖	–	–	?
Access								
Service Provider	◐●	◐●	◐●	◐●	◐●	◐●	◐●	✖
3rd party by user	◐	●✖	●✖	●✖	●✖	●	●	●
3rd party by provider	✖◐	✖	✖	–	–	?	?	✖
Legality	✖	✖	✖	✖	✖	✖	✖	✖
Retention								
Age Restriction	◐	✖	✖	✖	✖	✖	✖	✖
Information	✖	✖	✖	–	–	✖	✖	–

Key:
Complete: ● **Partial**: ◐ **Broken**: ✖ **Unknown**: ? **N/A**: –

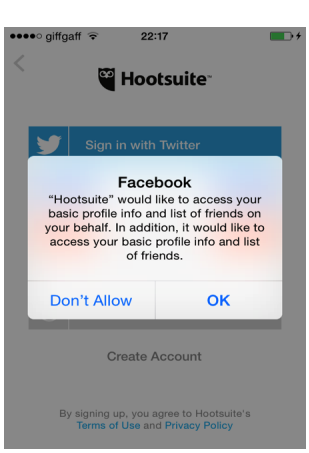
lation of the distributor vetting policies which mandate such documentation for all applications which process personal data from users. We did find in Table II that the SMAs without a privacy policy do not access "Identity", so technically they may argue they do not access personally identifying information. However, they are found to be able to access most of the social media data, photos, location, etc., which can be classified as personal information.

C. Traceability for Transparency and Control

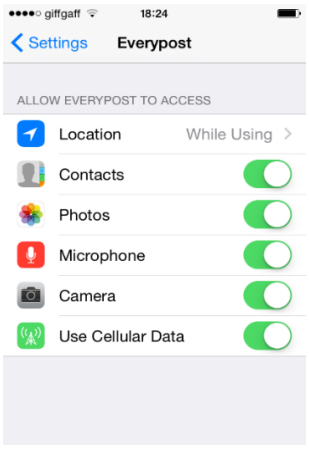
Common data actions have been categorized to form 14 privacy implications seen in the left column of Table V. Privacy implications fall under further categories by way of answering our privacy questions set out in section 3.3; **collection**, **purpose**, **access** and **retention** of data. Operations refer to features provided by SMA providers or distributors which inform us of data collection and use as well as providing us with control over data actions. Each symbol in the table provides a mapping to the degree of traceability offered by transparency and control operations respectively. Data operations refer to the extent to which transparency of data actions is presented to the user through interfaces, these include access permission prompts and other mechanisms which detail privacy implications. Control operations refer to features and mechanisms presented through interfaces which enable control over some data action, these include device settings, accept/decline button options etc. If the same degree is found for both transparency and control operations assessed, then only one symbol need be provided in representation. If a different degree of traceability is found, the first symbol in the particular cell of the table corresponds to transparency operations and the second symbol corresponds to control operations. In the resulting table, we refer to content as the social media data collected shown in Table III. Other privacy implications and results will be further explained and justified in the following subsections.

1) Complete: All SMAs provide control over some data collection through access permissions. iSocial does not specify any such method of informing the user of data collected through the requirement to accept access permissions. iSocial's terms and conditions specifies privacy implications; "Any site registration information is used only by the website and is not sold or given out to others", likewise users may provide an email address for the service provider to provide support. Complete transparency for **collection** can be found when an SMA communicates the data its going to access to the user through the interface operations. Fig 2a shows Hootsuite's permissions screen which tells the user about the social media data that will be accessed by it. Complete traceability mapping for control operations are when a user can regulate the access permissions through interface operations (such as Fig. 2b which shows Hootsuite for iOS).

Users have control over content provided for use by services, through accepting access permissions and the posting of information. Sharing information intentionally with SNS involves sharing this with these third parties by users, the transparency of third party access is completely apparent to the user in this case. Some applications offer settings which enable the user a level of control over who accesses information posted to SNS, and the restriction of data access to particular accounts. Controls offered are as found on common SNS;



(a) Notification of Social Media data access by Hootsuite



(b) iOS device settings which enable users to restrict access permissions

Fig. 2: Transparency and control operations

share with only friends or everyone. Asset transfer refers to personally identifiable information being transferred as businesses buy and sell assets.

2) Partial: The transparency of privacy implications through access permissions maps only partially to expectations provided by SMA privacy policies. An example of which is partial content collection made visible and controllable to the user. SMAs with privacy policies commonly state their rights to collect all information provided to the site, including shared with associated SNS. Google Play's Hootsuite provides a 'Send usage data' setting; the user is informed anonymous data is collected which is used to help improve Hootsuite. Partial transparency and control over internal use is apparent, with an ambiguous description collection and purpose, along with control over 'anonymous data' but no matching control for all data collected as specified in the privacy policy, such as content posted.

3) Broken: Internal use of data includes analytics used to improve or better understand services. It is common for servers to automatically collect usage information; "Server logs may include such information as a mobile device identification number and device identifier, web requests, IP address, browser type, browser language, referring/exit pages and URLs, platform type, number of clicks, domain names, search terms, landing pages ...", the list goes on and on. This type of information collected is referred to as the traffic data privacy implication, and may be shared with third parties on an aggregate basis for advertising and analytic purposes. We can see that both transparency and control for this example are broken in most SMAs, leaving users unaware in their normal use through the interface of the collection of this data and without a way of controlling that in any shape or form.

4) Unknown: Analyzed traceability mapping of data use as specified in privacy policies has shown us not to expect applications to inform users about the passive collection of non-identifiable information. We are aware that providers are likely to use and share traffic or aggregate data with third parties, for the purpose of analytics and advertising. We are unable to determine whether an application without a privacy policy passively collects such non-identifiable information. Therefore, for some SMAs, data disclosure to 3rd parties by the provider are shown to be unknown.

TABLE VI: Summary of traceability mappings for transparency, control and overall traceability of all privacy implications analyzed. Figures rounded to the nearest whole number.

		Complete	Partial	Broken	Unknown	N/A
Cydia	Transp.	29%	0%	57%	7%	7%
	Control	29%	0%	57%	7%	7%
	Total	29%	0%	57%	7%	7%
Android	Transp.	17%	24%	43%	2%	14%
	Control	17%	19%	45%	5%	14%
	Total	17%	21%	44%	4%	14%
iOS	Transp.	14%	27%	45%	4%	11%
	Control	22%	17%	45%	5%	11%
	Total	18%	22%	45%	4%	11%
Overall	Transp.	17%	23%	45%	4%	12%
	Control	21%	16%	46%	5%	12%
	Total	19%	19%	45%	4%	12%

5) Summary: Table VI summarizes our results, presenting rounded percentages of privacy implications found to be complete, partial, broken, unknown or not applicable. We provide a breakdown for each of the 3 app stores. The overall traceability of transparency and control are also provided.

We find a general **lack of transparency** across SMAs with 45 percent of SMAs revealing broken transparency mappings. Privacy implications offering complete transparency of data involve collection of personal information made visible to the user through in some way (e.g. showing the access permissions required). In order to consider current guidelines for user privacy as adequate, we must rule out mistrust between the user's expectations and reality of how SMAs treat their information by making them aware, either through privacy policies or through other awareness mechanisms, of any data collected, how it will be used, whom it will be shared with, and how long it will be retained.

We also find that users have a **lack of control** as less than a quarter of the results indicated complete control over privacy implications. In order to give more control to users, developers could work to increase application functionality while restricting access to data. Settings should enable control over all data collected, including information perceived as non-identifiable. Research has shown that pragmatic approaches of providing privacy related intervention, where users are shown the effect of exposures of their data, work well [4].

IV. Discussion and Lessons Learned

In this paper, we inspected how SMAs handle privacy and looked at it from three different angles. Evaluating the permissions requested by the SMAs was fairly straightforward. The SMAs communicate permissions to the user directly and the user also has the opportunity to verify social media permissions by checking their social media account and authorizing the applications. While there are many tools that enable the user to automatically check the mobile data permissions requested by apps, checking of social media permissions

is slightly more complex. The process may potentially be automated by simulating an authorization of the SMA to a dummy social media account (like a "guest" account, possibly built-in to the SMA), to reveal the permissions to the user, before they use the SMA with their own social media account. The larger problem here seems to be the lack of understanding that users have about the permissions requested by mobile applications. Greater awareness is desirable where users are informed about the implications of the permissions they are granting.

While looking for privacy related documentation, we found a fair degree of ambiguity. Not only do different application providers have different names for such documentation ("privacy policy", "terms of service", etc.), there is an absence of consistency in the content of these documents as well. This inconsistency makes it difficult to construct any expectations from the users' perspective of what they should be looking for in order to educate themselves about the privacy implications of using a particular app. Moreover, we found 5 SMAs which do not provide this documentation at all. This is, as pointed out earlier, in clear disagreement with the vetting policies of both Google Play and iTunes app stores. A possible mitigation may be found in automated solutions like "AutoPPG" which is an automatic privacy policy generator for Android applications [5]. It simply identifies the important privacy issues emanating from the usage of the application by conducting a static analysis of the application's source code. Automated solutions such as these may enable development of a consistent structure and terminology in such privacy policies which would enable easier traceability analysis. Furthermore, such mechanisms may also encourage SMA and other application developers to include privacy policies without putting in too much effort.

The qualitative analysis of privacy policies and analyzing traceability with interface controls was a comparatively less objective part of our methodology. Such analysis is harder due to the relative inconsistencies in privacy related documentation across apps as mentioned earlier. Moreover, the interfaces for each individual SMAs have different operations which necessitate a case-by-case analysis. This is the most costly part of the methodology in terms of time and effort. It is possible to automate the traceability analysis if the privacy documentation is standardized and the privacy implications are clearly defined. It is an interesting future direction in which research can progress where such an automatic tracaeability analysis might be used to certify SMAs. Any such efforts can rely on the analysis methodology shown by similar work in the area of social media sites and indeed the work done in this paper.

The methodology proposed in this paper may also be extended to other apps which provide users with the opportunity to link their social media accounts (such as gaming apps). It would be interesting to see whether the problems highlighted in this paper are specific to SMAs or whether other similar apps, which let the users post to multiple social media accounts, portray similarly low traceability. Future attempts at using this methodology may consider using multiple researchers to conduct the traceability analysis and look for a consensus based approach or provide inter-rater reliability between multiple researchers. This would potentially enhance the objectivity of the traceability analysis.

V. Related Work

A. Analysis of Mobile Data Access Permissions

Mobile applications generally are explicit in disclosing the data access permissions they require to the users. There is generally a screen which is shown to the user at the time of installation which tells them the data that the particular application will be allowed to access. The major issue is the "all or nothing" nature of mobile applications [6]. The user is required to grant the requested permissions to the application for them to use it. This is a problem as it has been shown that mobile applications often introduce risk vectors by asking for more permissions than required [3]. The problem is that the applications are somewhat hamstrung in this regard and have to request for permissions that they envisage using at any time during execution. There have been some solutions put forth to detect and possibly prevent malicious mobile applications by using anomaly detection to detect applications behaving maliciously and in a deviant manner from normally expected behavior [7]. The idea is to use static analysis to create profiles of applications' expected behavior and detect anomalies at runtime to secure mobile applications. This is similar to the work of Hussain et al. which looks at detecting malicious database applications [8]. Another proposed approach, "PrivacyGuard" uses the VPN service of Android devices to intercept network traffic of mobile applications to detect information leakage [9]. It also provides mechanisms of tricking the malicious applications by manipulating the leaked information. We found that most of the previous work in this area only looks at leakage of mobile data and not social media data which SMAs have access to as well.

B. Analysis of Privacy Policy Traceability

There is previous work which shows that control over data disclosure can affect decisions made by users [10]. Greater transparency about data being shared often acts as a mitigating factor against erroneous decisions being made. Our work looks at the traceability for transparency and control by looking at the interface operations and how closely they match with privacy policies. Qualitative analysis of documented policies and analyzing traceability with interface features is an extensively researched topic in software engineering. More recently, this technique has been used to analyze whether the privacy policies outlined by SNSs are congruent with the interface controls provided to the users. Anthonysamy et al. demonstrated that SNSs themselves suffer from a lack of traceability between data actions defined in privacy policies and corresponding data operations apparent to users through interfaces [2], [11]. Our work extends this methodology to perform a privacy analysis for SMAs by performing an analysis of the mobile phone data and social media data accessed by the SMAs in addition to a traceability mapping

which considers the transparency of interface operations and the control provided to the user.

VI. Conclusions

In this paper, we described a three-step methodology to examine the privacy issues posed by SMAs by examining the data (both mobile and social media) permissions requested by them, checking whether they provide the user with privacy related documentation and analyzing traceability between privacy implications identified in the privacy policy with the interface operations provided to the user. We used this methodology to evaluate 13 popular Social Media Aggregators (SMAs) from 3 app stores and found that the majority of the SMAs we evaluated accessed users' personal information including their social media activity. However, we also found that 5 of the 13 SMAs did not provide any privacy related documentation which is in clear conflict with the vetting policies of the app stores. Our results show that 45% of SMAs show *Broken* traceability between privacy documentation and interface operations while *Complete* traceability is observed in only 19% of the cases. These results highlight the need for major improvements to ensure that the usage of SMAs does not compromise user privacy. The methodology described in this paper can be reused for further investigation of SMAs or be extended, with certain improvements, to examine similar applications which enable the user to link their social media activity.

References

[1] J. D. Young, A. Antón *et al.*, "A method for identifying software requirements based on policy commitments," in *Requirements Engineering Conference (RE), 2010 18th IEEE International*. IEEE, 2010, pp. 47–56.

[2] P. Anthonysamy, P. Greenwood, and A. Rashid, "Social networking privacy: Understanding the disconnect from policy to controls," *Computer*, no. 6, pp. 60–67, 2013.

[3] P. H. Chia, Y. Yamamoto, and N. Asokan, "Is this app safe?: a large scale study on application permissions and risk signals," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 311–320.

[4] A. Kapadia and A. J. Lee, "Improving privacy through exposure awareness and reactive mechanisms," in *CHI 2016 Workshop on Bridging the Gap between Privacy by Design and Privacy in Practice*. ACM, 2016.

[5] L. Yu, T. Zhang, X. Luo, and L. Xue, "Autoppg: Towards automatic generation of privacy policy for android applications," in *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 2015, pp. 39–50.

[6] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang, "Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 501–510.

[7] E. Bertino, "Securing mobile applications," *Computer*, vol. 49, no. 2, pp. 9–9, 2016.

[8] S. R. Hussain, A. M. Sallam, and E. Bertino, "Detanom: Detecting anomalous database transactions by insiders," in *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*. ACM, 2015, pp. 25–35.

[9] Y. Song and U. Hengartner, "Privacyguard: A vpn-based platform to detect information leakage on android devices," in *Proc. of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 2015, pp. 15–26.

[10] S. Patil, R. Schlegel, A. Kapadia, and A. J. Lee, "Reflection or action?: How feedback and control affect location sharing decisions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 101–110.

[11] P. Anthonysamy, P. Greenwood, and A. Rashid, "A method for analysing traceability between privacy policies and privacy controls of online social networks," in *Privacy Technologies and Policy*. Springer, 2014, pp. 187–202.