Are we trusting social networks too much?

Authors
Minerva Hoessl, MSc (Royal Holloway, 2016)
John Austen, ISG, Royal Holloway

Abstract

User privacy and commercial profit are in an on-going battle. Social networks engage in extensive information collection, analysis and storage. Users have no control over such data collection practices. Most social network users will never read the policies provided by social networks, yet are showing an increased desire for better privacy. This article examines the main flaws in the privacy policies of Facebook, Twitter, LinkedIn and Google+/Google that may put user privacy at risk. To some, these flaws may present the appearance that ever-increasing profits override concerns for privacy.[a]

[a]This article is published online by Computer Weekly as part of the 2017 Royal Holloway information security thesis series `http://www.computerweekly.com/ehandbook/Are-we-trusting-social-networks-too-much`. It is based on an MSc dissertation written as part of the MSc in Information Security at the ISG, Royal Holloway, University of London. The full MSc thesis is published on the ISG's website at `https://www.royalholloway.ac.uk/isg/research/technicalreports/technicalreports.aspx`.

Here is the bad news: "*If an Internet service is free, you are not the customer, you are the product*". This quote by Matthew Bailey, author of The Complete Guide to Internet Privacy, Anonymity & Security, certainly is a wake up call for many social network users, and sadly it is the truth.

User privacy and commercial profit are in an on-going battle. Cyber crime is on the rise. Information obtained through social networks is often used to facilitate such crimes. LinkedIn, for example, is used frequently to gain information about companies in order to attempt spear phishing and other types of attacks.

Data collection and its analysis is now at a massive scale, allowing social networks and search engines to build comprehensive profiles of the individuals using their services. Whether the entertainment value of social networks will override a person's need for privacy remains to be seen.

We analyse the privacy policies of Facebook, Twitter, LinkedIn and Google+/Google to determine their adequacy in protecting user privacy. The most disquieting revelation and source of a significant threat to user privacy came from the fact that these social networks use their own definitions of what constitutes *personal* data versus *public* data.

Why do social networks collect and share all this information?

Advertising revenues are the main reason for much of the information sharing practices of social networks. It is how these websites amass vast yearly profits. Facebook, for example, earns billions of pounds each year from advertising revenues. Corporate profit appears to be prioritised over user privacy. Users either agree to all terms of such sites or must choose not to use the service, which nowadays could have social repercussions. A voluntary reduction of profits in order to provide a more privacy friendly environment cannot reasonably be expected.

Lawmakers are well aware that revenue losses, as well as international business relations, have to be a consideration when changing and implementing new privacy legislation. Because of this, the relevant pieces of legislation reviewed were found to be of limited effectiveness. Generally speaking, such legislation favours free flow of information for commercial purposes over increased user privacy. The best effort to improve user privacy was seen in the General Data Protection Regulation (EU) 2016/679

(GDPR). The GDPR aims to improve the quality of communication between social networks and users. Social networks will be obligated to disclose more concise and easy to understand information, to put a maximum limit on the length of time data may be stored and to use a common definition of what constitutes *personal* data. However, it is seen to favour commercial aspects by encouraging data anonymisation and by providing support for the transfer of personal data to a third country or an international organisation. Furthermore, the GDPR will only protect EU residents.

How do Facebook, Twitter, LinkedIn and Google fail us?

Content is public by default

Content submitted to social networks is deemed *public* by default. In addition, social networks use their own definitions of *personal* versus *public* data. This enables social networks to share a vast amount of information with third parties for profit, much of it information its users may assume to be personal.

As per Twitter's Privacy Policy, "*public information includes the messages you Tweet; the metadata provided with Tweets, such as when you Tweeted and the client application you used to Tweet; the language and time zone associated with your account; and the lists you create, people you follow, Tweets you mark as likes or retweet, and many other bits of information that result from your use of the Twitter Services*". As is shockingly apparent, Twitter leaves itself a lot of room by stating "*and many other bits of information*" and fails to provide users with a clear list of all the information it deems *public*. The Privacy Policy also advises its users that they "*should think carefully about what you are making public*". No doubt, this is sound advice not only for Twitter, but for all social networks.

Facebook fails its users equally. Facebook defines "personally identifiable information", the American equivalent of the European notion of *personal data*, as "*information like name or email address that can by itself be used to contact you or identifies who you are*".

Definitions of personally identifiable information tend to vary within the United States. According to the U.S. Department of Commerce Office of the Chief Information Officer and as defined by the OMB Memorandum M-07-16, the definition of personally identifiable information is "*information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc.*". The California State Online Privacy Protection Act of 2003, Business and Professions Code, Chapter 22 Internet Privacy Requirements, Section 22577, on the other hand, only defines personally identifiable information as "*individually identifiable information about an individual consumer collected online by the operator from that individual and maintained by the operator in an accessible form*".

Worried about your privacy ...

... on Facebook, Twitter and LinkedIn? Simple steps you can take:

1. Check all settings and change them to suit your needs.
2. Don't post private information. Keep communications between you and your friends private. Send messages or create private groups instead of communicating using posts.
3. When friends tag you in a photo, remove the tag as soon as possible and advise friends not to tag you in the future.
4. Applications available on Facebook and other sites should be seen as a possible threat. These applications use their own privacy policies. Check what information the application wishes to access and how such information is used and shared by that application.

The Irish Data Protection Act and the U.K. Data Protection Act both agree that *per-*

sonal data is data that can identify an individual on its own or, as stated in the U.K. Data Protection Act 1998, "*from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller*". Even more detailed is the definition given in the newly adopted General Data Protection Regulation (EU) 2016/679. It defines personal data as any information that can identify a person "*directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*".

Overall, Facebook's definition of personal data, as stated in their Data Policy, does not appear to accommodate user privacy rights from outside of the state of California. This poses a definite threat to many users' privacy as Facebook may be sharing information that such users would naturally assume to be personal data.

Excessive data collection

The Internet is a massive data collection machine. Ever-new technologies are developed to collect and store more information. No legislation is in place to protect Internet users from such excessive data collection.

Facebook is a prime example. It utilises new technologies and features disguised as user experience enhancing additions, such as, for example, the new options presented when *liking* a post where users can now express how they *feel* about it. Such features were not solely developed for user enjoyment, but appear to have been developed for the purpose of gathering more detailed information about its users. Information collected by Facebook includes not only content submitted by users but also information such as:

- Actions of the user, including content viewed and interacted with.
- Content and information provided by other Facebook users about another user, such as tagging a photo or sending a message to that user.
- Extensive device information such as "*operating system, hardware version, device settings, file and software names and types, battery and signal strength, and device identifiers*".
- Location information, as can be ascertained using GPS, Bluetooth or Wi-Fi signals as well as Internet service provider or mobile operator names, browser types, mobile phone number and IP addresses.

Information collected by Twitter is very similar and includes the user's interaction with the site, as well as their "*IP address, browser type, operating system, the referring web page, pages visited, location, mobile carrier, device information (including device and application identifiers), search terms, and cookie information*". Every time the user interacts with any Twitter Service, including simply visiting a website that uses a Twitter button or widget, all available information is collected and processed by Twitter.

LinkedIn operates much the same as Facebook and Twitter. It collects information whether the user is a LinkedIn member or just a visitor. When a user is not logged in, information collected is associated with a user's account if LinkedIn can positively identify the user by other means. This occurs when a cookie can make a positive identification based on device or other information that is collected and logged.

Google collects information from any and all available sources including search queries, language preferences, types of browsers used, extensive device information and location information. Location information is not only gathered using a variety of sources such as IP address, Wi-Fi access points and mobile network towers, but also from search queries by using locations that the user is showing an interest in. This includes any information that infers the user is either interested in a particular location or might be at that location. Information is collected both while the user is signed in and also when not signed in and can include contacts the user added, calendar events, photos, videos or documents

uploaded. Every move made on Google is tracked, analysed and associated with a user's account whenever possible.

Data collection has clearly grown to excessive proportions. There is no legal premise that would minimise such collection.

Information sharing and disclosure

Two significant problems were encountered: the inconsistent definitions of "personal data", and the fact that users will never know exactly with whom their information is shared.

Facebook Facebook's weak definition of *personal data* brings uncertainty as to what information is being shared. Facebook shares information with a wide variety of entities from companies that are owned by Facebook to third-party partners and customers.

Twitter Information deemed *public* by Twitter is shared with third parties including analytics websites such as Twitonomy and Followerwonk, allowing these to analyse and profile users and expose such information on a global scale.

LinkedIn Even upon request, LinkedIn does not supply a list of entities with whom user information is shared.

Google Google shares information with Google affiliates and other trusted partners. It is unclear how far reaching the sharing practices are.

The definition of *personal data* or *personally identifiable information* is of immense importance and thus should be consistent across all social networks. It must also take into consideration users from outside of the United States and allow users to obtain a detailed list of what type of information is shared and with what entities. Organisations should not be able to refuse such a request since such entities are a part of the privacy policy agreement made with users.

Data retention

As many users well know, deleting an on-line account is not always an easy task. On-line accounts seem to purposely make it difficult for users to delete accounts. Some user information is being stored and processed even after the user requests account deletion. Little information is given to clarify what information is kept and why.

LinkedIn was the worst offender, using vague and ambiguous wording. The Privacy Policy states: "*If you close your account(s), your information will generally be removed from the Service within 24 hours. We generally delete closed account information and will de-personalize any logs or other backup information through the deletion process within 30 days of account closure, except as noted below.*" However noted "below" is simply an explanation stating that LinkedIn will "keep your information for as long as your account is active or as needed". Because of the wording used, users will be unable to determine LinkedIn's exact data retention practices.

Facebook does not state the maximum period of time user data is stored for and the information provided is generally unclear, while Twitter gives no reason to its users for why continued storage is necessary. Google, on the other hand, actually informs its users that the information submitted may be stored indefinitely, even when a user discontinues using Google services. Deleting a Google+ account is also a complex task and not user-friendly.

In many cases, anonymisation of data is used as a crutch to allow user data storage. It is however widely known that effective anonymisation is hard to achieve. Without strict enforcement of appropriate types of anonymisation techniques, social networks are able to use any techniques they deem appropriate and thus it is unknown as to how effective the anonymisation actually is.

Deletion of on-line accounts should mean the complete deletion of all user information. Unfortunately, there is no legislation in place that forces social networks to delete all user information. However, thanks to the new General Data Protection Regulation (GDPR), the maximum retention time will at least need to be clearly indicated.

Data transfer

Social networks are not transparent when it comes to sharing data transfer information with their users. In Google's case for example, its Privacy Policy simply informs users that: "*Google processes personal information on our servers in many countries around the world. We may process your personal information on a server located outside the country where you live*."

Currently, it is impossible for users to know exactly where their data is transferred. A list of locations, or at the very least of the countries, should be available to users. It should also be made clear to users, that governments of the listed countries may be able to gain access to user information. This may be an important consideration that could dissuade users from using some social networks. The newly adopted EU-U.S. Privacy shield was shown to still present significant inadequacies in terms of data transfer to the United States. It does not look to be in the best interest of European users, but still heavily favours the surveillance needs of the United States.

How do we improve user privacy?

A first step would certainly be to improve the quality of the information provided to users by social networks. Information should be clear and easy to understand so that users can make an educated decision whether they still wish to use the service or not.

Even though the GDPR addresses better communication between social networks and users, it fails to include information that should also be made accessible such as a detailed list of entities with whom user information is shared and a list of all locations where data is transferred and processed. At the very least such information should be made available to users upon request. Without legally enforcing the right to request such information, organisations are not likely to cooperate.

Current and new pieces of legislation do attempt to find a balance between user privacy and the commercial flow of information. However, commercial aspects continue to be favoured over providing increased user privacy. Social networks are revenue driven; user privacy will never be their top priority.

Worried about your privacy ...

... on Google?

1. Visit the Google Dashboard and "My Activity" to control what information Google stores about your browsing habits and to turn off personalised advertisements.
2. Clear your browser cookies regularly. Cookies and similar technologies are used to save user preferences. This includes the websites you visit, how long you were on them and what you clicked on.
3. When deleting a Google+ account read the information provided by Google carefully so that you are aware of items that will actually not be deleted. For example, any photos added to the Google+ account will still be available on Google's Album Archive.
4. Read the Google Privacy Policy.

Users must not forget that social networks need them to survive. Boycotting social networks may be necessary in order to force change. Another possible option would be to establish social networks that operate on a paid-for membership basis. This would eliminate the need for user data collection for the purpose of revenue generation, as well as the nuisance of excessive advertisements. If users

choose instead to accept any and all policies provided to them by social networks, such companies will continue to use this lack of involvement to their benefit. Social networks are aware that the majority of their users will never read the provided policies.

Ultimately, user privacy is in the hands of the users. It is time for users to take responsibility for both their activities on the Internet and their lack of interest in policies. Just because the Internet is accessed from the comfort of our homes and privately owned devices does not mean that the Internet itself is a private place. The Internet is a public place. To expect a high degree of privacy is simply unreasonable.

Biographies

Minerva Hoessl is a 2016 MSc (Distinction) graduate of the Royal Holloway Information Security Masters programme. She specialises in open source intelligence gathering. She has developed and taught a Level 3 Open Source Intelligence gathering course (osinttraining.co.uk) to clients including the Security Industry Authority, the Foreign and Commonwealth Office, the NHS and specialised UK military intelligence units. She has also trained staff from private organisations such as British Airways, Credit Suisse, Burberry, Jack Wills, Camp America, Apple, the Environmental Investigation Agency (EIA) and many more. Her interests are in cyber crime investigation, investigations that expose crimes against wildlife and the environment, and educating the public on cyber threats and privacy issues on the Internet.

John Austen (BA, MSc, FBCS, NEBSS) is a consultant lecturer in the Information Security Group at Royal Holloway University of London. He is a specialist in cyber-crime and investigation techniques, international law, and organisational security. After studying at the FBI National Academy in Quantico, Virginia he became the founder and Head of The Computer Crime Unit New Scotland Yard from 1984 to 1996, an operational specialist unit and the forerunner of the National Hi-Tec Crime Unit and the e-Crime Unit. He was the first Chairman of the Interpol Computer Crime Committee, from 1990 to 1996, which was responsible for the worldwide standardisation of Police procedure and international training in the area of cyber-crime investigation and digital forensics. He was the 2003/4 President of the U.K. Chapter of I.S.S.A. (Information Security Specialists Association).