Facebook User’s Data Security and Awareness: A Literature Review

Johanna P. Calbalhin
Student, College of Graduate Studies, Samar State University, Philippines
jp_calbalhin@yahoo.com

Abstract: Facebook has become more powerful now more than ever. As it expands its network, more and more information becomes available for them to use for various purposes for an improved experience of a Facebook user. The social capital to which people have invested in the social media platform made it more difficult for people to quit. The review of the literature revealed that awareness level is very variable, making reports from one publication to the other conflicting. It is however very evident that even if people who are aware of the risks involved still patronize Facebook. Some may have tinkered with their privacy setting to improve privacy control but their actions/activities on the net make that extra precaution useless. As the internet become more of a necessity than choice, more and more information is readily available for data mining through Facebook or other platforms capable of phishing valuable information such as passwords. The adult population who has more at stake have taken extra precaution compared to the young population who has the tendency to self-disclose.

Keywords: data privacy, phishing, identity theft, data risk, Facebook quiz

1. Introduction

The society has benefited so much from the advances in information and communications technologies. In the past, people need to visit libraries in best universities or monasteries where precious manuscripts were hoarded (Hean, 2000). The development of digital computers opens its way to networked communications which has revolutionized information sharing. In 1969, the first computer-to-computer link was established on ARPANET or the Advanced Research Projects Agency Network, the precursor to the internet (UCLA, nd). The increased reliance on networked systems has also exposed users to various risks such as identity theft or misuse of private data. As of June 2018, a total of 4.21 billion or about 55.1% people are connected to the internet; about 1,066% increase in 18 years (www.internetworldstats.com, 2018). Penetration rate is very high in North America where 95% of their people are connected to the internet and Africa has the lowest with only36.1% of its people have access to the internet (ibid). One famous internet phenomena are Facebook. The social media giant has about 58 million active users in 2007 and this reach to about 2.27 billion active users in 2018 (Facebook Newsroom, nd). India, US, Indonesia, Brazil, Mexico and Philippines are the top countries with the most number of Facebook users. In the US, about 68% and 73% of its people are users of Facebook and YouTube respectively (Smith & Anderson, 2018).

According to the number of linking root domains, Facebook.com, Twitter.com, Google.com, Youtube.com, and Instagram.com are the top five websites in the world (Moz, nd). Facebook is an
especially important website to people as it is where users store their life story, their identity. Facebook generally collect information in various ways. Every time a user visits a website with like button, the browser used sends a signal to Facebook (Facebook, nd). On the other hand, even if a user is logged-out from his/her account but visited a site with a like button or another social media plug-in, the browser used also sends the information about the clients' visit (ibid). This is performed by Facebook to customize what is the users’ preference and provide the user with the maximum benefit it can offer. Google is also collecting information about its user’s preferences, where they have been, what the users have accessed and searched. It also gathers the information like what the applications used, all YouTube history was. Google can even access the users’ webcam and microphone and they even have everything you have deleted (Curran, 2018). In a Google mail, you can even add your other emails but by doing, the user has allowed Google to peek into your accounts.

Facebook and Google are just two of many websites that keeps information about the users. However, most of the user is not aware of the volume of information they are providing. Ideally, software companies must share responsibility about the products and services they offer in a manner that both helps protect customers from potential harm (Kutterer, nd.). But this may not be the case to free services such as Google and Facebook. These free services need resources to exist continuously. Data is the currency of free social media. Facebook for example has generated profits at a rate of about US$20.21 per user in 2017 (Glum, 2018). They sell the very behaviour that a user shares or expresses in Facebook to marketers. In 2016 around 98 data points are used to be sold to interested companies such as user’s gender, age, political position, relationship status and many more (Glum, 2018). Just like Facebook, Google earns much through online advertising. Both use the user’s profile and internet activity for advertisement purposes (Rosenberg, 2018).

The utilization of vital information shared by the user in their internet accounts is covered by the agreement the users signified when they created their accounts. The Facebook policy stated the kinds of information the company collects which includes things the users and others provide, device information to access Facebook and information from Facebook partners. Everything a user does in Facebook is collected, processed and shared to friends and user’s network. It is also shared to a certain level to Facebook third-party partners, sometimes are taken advantage by them.

The huge amount of information is stored in Facebook's database that when hacked or through technical glitch may be exposed accidentally. Early on Facebook life, it faced serious privacy breaches. About two years after Facebook's inception, users’ passwords were still being sent without encryption which can be easily intercepted by a third party (Jones & Soltren, 2005). This is just one of the several glitches Facebook experienced, specifically when it introduces new features to the system (Schonfeld, 2008; Boyed, 2008; Singel, 2007; Perez, 2007; Romano, 2006; Gross & Acquisti, 2005). Aside from these issues within the Facebook system, phishing is another form of accessing data from Facebook users. Phishing takes advantages of software and security weaknesses on both the client and server sides (Wilson, 2005).

Undeniably Facebook is very popular with more than half of the world
population is connected (www.internetworldstats.com, 2018). How aware are its users regarding the information they share, what do they do and in what ways this information is mined?

2. Objective

This literature review focuses on the level of Facebook user’s privacy awareness, control, and risks.

3. Methodology

This review of literature focused on Facebook privacy issues, specifically in terms of user’s privacy awareness, and privacy control. In gathering the needed data, the researcher firstly, define a research question, and then locate and select relevant previous research studies, technical papers, press releases and news regarding the topic. Included in the review are reports published from 2005 to date.

There were more than 120 articles (scientific publications, magazines, news reports, press releases and many others) considered and later reduced to 53. These include scientific articles from journals, technical review of experts in the area of internet and privacy, commentaries including news articles and government reports. Topics in privacy and awareness on the internet especially Facebook, data sharing and third-party phishing approaches used to mine Facebook data were collated and reviewed. It used publications from 2005 to 2018.

4. Results and Discussion

Facebook provides users with a free social networking site. In return, users can provide basic demographic information through their likes, interests, and activities shared in their accounts. All of this information is very valuable to marketers who want to pay Facebook to place their ads on the target consumer’s accounts. The information Facebook shares to third-party companies are all enshrined in its sign-up page. The users are made to agree to lengthy and complicated social networks terms of service (ToS) where 30% don’t even dare to read with only 17.56% always read the ToS (Morrison, 2015). More than half of those who said they don’t read ToS says it’s useless primarily because you need to agree to sign-up even if you do not like the terms (ibid). As Facebook continuously evolves, it encounters glitches that expose user’s information.

4.1 A Timeline of Facebook Privacy Issues

Facebook was launched in 2004 and since then it has evolved to be the most popular social media platform to date. The foregoing are several glitches that were reported. Hacking incidents were not included in the list shown in Table 1.

4.2 User’s Privacy Awareness

In 2005, a year after Facebook was launched, Govani and Pashley (2005) studied how student shares their information despite the many privacy issues in the new system. They found out that about 84% of Facebook users were aware of the privacy settings but only half of them used it. These numbers are almost similar as reported by Jones and Soltren (2005). It was also found out that adult and young Facebook users have similar behaviour when it comes to privacy settings. Five years after, the picture is still the same significantly smaller proportion from older and younger users of Facebook actually used privacy settings (Christofides et al., 2010). There are however reports showing different numbers. It might be attributed to other
Factors such as knowing how the information generated from Facebook accounts can be used against the user.

Table 1: Timeline of Facebook Major Privacy Issues

<table>
<thead>
<tr>
<th>Year</th>
<th>Privacy Glitches</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>News Feed: The new feature allowed every post of the user to appear to friends Facebook wall. Privacy controls were introduced after about 1 million users protested. The feature later becomes one of the major parts of its success.</td>
</tr>
<tr>
<td>2007</td>
<td>Advertisement: The feature allowed the company to track purchases of by Facebook users and notify their friends what was bought even without user’s consent.</td>
</tr>
<tr>
<td>2013</td>
<td>Bug exposes private contact information: About 6 million users were revealed to anyone who had some connection to the person or if they have at least one contact information.</td>
</tr>
<tr>
<td>2014</td>
<td>Mood-manipulation experiment: It involved more than half a million randomly selected Facebook user for the experiment. The result of the experiment was published and later removed due to ethical issues.</td>
</tr>
<tr>
<td>2015</td>
<td>Cuts off apps from taking any data from Facebook: An app downloaded by user A can allow extracting user’s A friend’s data. Even after the stoppage, third-party users were known to have still used the previously collected data.</td>
</tr>
<tr>
<td>2018</td>
<td>Privacy Bug: About 14 million users were affected. They may have unknowingly posted private information to the public.</td>
</tr>
<tr>
<td>2018</td>
<td>87 million user’s data: A researcher had sold Facebook data collected via a personal quiz was revealed.</td>
</tr>
</tbody>
</table>

Sources: Newcomb, 2018; Rodriguez et al., 2018; Lee, 2018

Many argue that the lengthy, complicated and technical aspects of privacy setting discouraged users from reading it and taking extra precaution in using it. There is a need to simplify the ToS of Facebook for user’s greater protection. This action however may be counterproductive to Facebook and they are unlikely going to do something serious about it. This is where the need for government intervention to secure the privacy of people’s information and for institutions to make users literate on the use of the social networking site (Nyoni & Velempini, 2018).

A study in The Netherlands reveals that almost 6 in every 10 of their Facebook users are not comfortable with the level of exposure (Yazici, 2017), but they remain to use it for the benefit they get from it. Facebook users have been giving away personal for more than a decade. Those who have exhibited a higher level of privacy awareness are the ones who have negative social network site experiences (ibid). The number of aware about privacy settings since 2005 to date is ever changing.

Table 2. User’s Awareness on Facebook Privacy Setting

<table>
<thead>
<tr>
<th>Relevant Studies</th>
<th>Aware of the Privacy Settings</th>
<th>Not Aware / Have not read FB Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Govani and Pashley (2005)</td>
<td>84%</td>
<td>16%</td>
</tr>
<tr>
<td>Jones and Soltren (2005)</td>
<td>74%</td>
<td>26%</td>
</tr>
<tr>
<td>Acquisti &amp; Gross (2006)</td>
<td>22%</td>
<td>78%</td>
</tr>
<tr>
<td>Yong (2011)</td>
<td>48%</td>
<td>52%</td>
</tr>
<tr>
<td>Centre for the Advancement of Social Sciences Research (2013)</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>Yong (2016)</td>
<td>6%</td>
<td>94%</td>
</tr>
<tr>
<td>Nyoni &amp; Velempini (2018)</td>
<td>12%</td>
<td>88%</td>
</tr>
</tbody>
</table>
Table 2 may appear unbelievable as there is more aware Facebook user in 2005 than 2018. The difference is likely because of different settings today than before, today’s Facebook is more complicated than when it was launched in 2004. It is also because the literature list shown in Table 2 refers to different communities from different countries. The lower awareness level is perhaps because of the recent glitch that Facebook encountered. They thought they had made everything private from other users but only to found out, their profile information was harvested through other means. This means that even if the information was set to private, some entity might be able to access its data.

4.3 User’s Privacy Control

Facebook has continually made it possible for users to control their privacy settings to protect their personal data and limit that has access to this information; however users are not aware or do not always employ these safeguards. As one in control, Facebook can play around everyone’s privacy settings and the users can do so little about it (Brewster, 2018). For every privacy glitches Facebook encounters, they try to correct it. The amount of information available in the Facebook database has increased through other electronic media platforms (Smith, 2018). This is because everything that the user does to his/her account, Facebook (and other platforms like Google) collects it. This includes users’ instant messaging services, newspapers, blogs, and other Facebook and third-party services (Debatin et al., 2009).

One thing is very clear, people are aware of the data that they have provided through Facebook but data showed that not all had taken serious action over it. In a study in 2013 reveals that most teen social media users say they are not very concerned about other’s using their data, almost a quarter is not at all concerned (Madden et al., 2013). Some may have done something about its privacy level, but its only up to making their posts for their friends only or to a select group of people including those who can tag, share, and communicate with (Madden, 2013; McDunnigan, nd). This concept of privacy is flawed as a user can be seen at a friend's posts in which a user is tagged making it impossible to be really private in social media (Haney, 2016; Strachan, 2015; McDunnigan, nd). Other information derived from the users’ activity on Facebook and other websites with like buttons also collect data (Debatin et al., 2009; McDunnigan, nd). Facebook is also permitted by the user (part of its ToS) to use user’s data in some of its development activities (McDunnigan, nd). Many times, the development activities sometimes lead into glitches that exposing users at different levels.

A Gallup poll showed that more people are very concerned about the invasion of privacy in 2018 than eight years ago (Jones, 2018). This perhaps is the reason why the numbers of people and organizations who have taken action over security issues have increased. Some have gone to the extent of taking government actions. The US Senate conducted inquiry over the data breach that exposed more than 87 million of its user’s information to a data firm (Griffen, 2018; Lapowsky, 2018). This data breach has made 74% of American users to change their privacy settings; 42% have taken a break for weeks or more and more than 26% deleted their Facebook applications from their phones (Perrin, 2018). Europe has enacted a law on data protection which Facebook complied in January 2018 (Newcomb, 2018). Belgian courts ordered Facebook to delete all data it
collected from their users and others which may have landed on its pages (ibid). Belgian court also told Facebook to stop tracking people across the entire internet but Facebook appealed the court ruling (ibid).

Table 3: User’s Privacy Control

<table>
<thead>
<tr>
<th>Relevant Studies</th>
<th>Take Control /Adjust their Privacy Settings</th>
<th>Do not use Privacy Settings/Engage in Risky Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross and Acquisti (2005)</td>
<td>1.5 %</td>
<td>98.5 %</td>
</tr>
<tr>
<td>Govani and Pashley (2005)</td>
<td>40 %</td>
<td>60 %</td>
</tr>
<tr>
<td>Jones and Soltren (2005)</td>
<td>64 %</td>
<td>36 %</td>
</tr>
<tr>
<td>Campbell &amp; Kraan (2005)</td>
<td>48 %</td>
<td>52 %</td>
</tr>
<tr>
<td>Debatin et al. (2009)</td>
<td>69 %</td>
<td>31 %</td>
</tr>
<tr>
<td>Yong (2011)</td>
<td>61 %</td>
<td>39 %</td>
</tr>
<tr>
<td>O’Brien &amp; Torres (2012)</td>
<td>78.3 %</td>
<td>21.7 %</td>
</tr>
<tr>
<td>Centre for the Advancement of Social Sciences Research (2013)</td>
<td>37 %</td>
<td>63 %</td>
</tr>
<tr>
<td>Madden et al. (2013)</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>Perrin, (2018)</td>
<td>74%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Christofides et al. (2010), found that adults were more likely to control their information than were youth, but this difference could not be fully accounted for by differences in knowledge about privacy settings. However, a more recent survey revealed that younger users had taken a stronger stance on Facebook privacy control versus the older respondents (Perrin, 2018). They even go to the point of deleting their Facebook app (ibid). Women and girls were more likely to control their information than were men or boys (Christofides et al., 2010; Madden et al., 2013). For both youth and adults, the strongest predictor of information control on Facebook was a greater awareness of the consequences of sharing information. In addition, youth with higher self-esteem and lower levels of trust were more likely to control their information on Facebook.

Smith (2018) enumerated 18 changes that must be acted upon right away. These are privacy settings available for every user to manipulate, all of which can be seen at privacy settings. The user can control who sees his/her posts, who sees apps activity, hide user's personal information, hide a post from other people, hide a post from user's timeline, restrict people to share photos and posts, control how public profile will look like. The user can also control who can see his/her old posts, limit who can send friends requests, can block anyone, limit who can find you through any contact information, to keep off from Google. The user can also hide his/her own real name, protect user's location, prevent and remove from a tag, tag friends but control who can see the post and finally a user can make his/her Facebook profile completely private.

4.4 Internet Meme, Facebook Quizzes.

Facebook has access to everything the users does within the environment of Facebook. Facebook data use policy states that they do not share user’s information unless the user has given permission (McDunnigan, nd). The 87 million user's information that was illegally used was extracted from a third-party application, a personality quiz (Rodriguez et al., 2018).

Many companies/organizations even individuals specifically advertisers use Facebook services for their advantages.
These Facebook services include social plugins, Facebook Login, Facebook Analytics and Facebook ads and measurement tools. When a user visits a site or app that uses Facebook services, the company receives information even if the user is not online or even if the user does not have a Facebook account (Facebook, 2018). Many apps available are intertwined with Facebook accounts. Apps are software that allows people to play games and share a common interest to other users (Steel & Fowler, 2010). The web of information available is so vast that Facebook database is one of best targets for data theft. The building of information starts when a user registers an account with basic profile information such as name, gender, date of birth, email or mobile number. Once logged-in, the user is tempted to add more features, installed some applications and interacts with other users and many other activities done on the internet (inside or out of Facebook environment). Every ad a user’s click on, every IP address that the user used in logging the account, every friend in the network and the interactions made; basically everything is recorded (Korosec, 2018). Risks for data theft are increased when generic passwords are used for various online accounts. This is where memes become very useful for hackers to take advantage of hacking not only Facebook but other online accounts including mobile banking data. Many of Facebook user’s login information are sold at “dark web” for little more than a few dollars (ibid).

A simple test of some Facebook quiz reveals how data are collected by third-party applications such as personality tests and many other tests or fun games. Ceukelaire (2018) tried a Facebook quiz entitled Which Disney Princess Are You? In his examination, it was revealed that the quiz taker information available at Facebook was mined by the application. The third-party application can be used to steal someone’s information and use it to access the user's account and see private information (ibid). The said flaw however has been corrected by Facebook after it was reported.

In 2015, there was a Facebook app that illustrated the most frequent word used through a word cloud. About 17 million Facebook users participated in this game. It was revealed later that the app was able to collect information about the user including his/her entire friend's list and everything the user has liked (Wakefield, 2015; Patterson, nd; Karcz, 2015). Jongwa Kim, the CEO of Vonvon Inc., the maker of the word cloud quiz has stated that they never store data nor sold it to anyone (Karcz, 2015). Facebook allows third-party developers to use some of the user's data to function. This information may include name and profile picture, gender, schools, workplace username and ID, age range, language and country (Patterson, nd). Another permission request from the third party developer is reviewed by Facebook staff (ibid). Facebook privacy has improved after the Cambridge Analytica incident (ibid).

4.5 Information Security and Disclosure Policy

Facebook has provided people an immense advantage because of the amount of information it has from its users. Most people have adopted the technology into nearly every aspect of life. Devices like phones, laptops (where Facebook is accessed) become more “smart” but they become more hackable (OSAC, 2014). The technology adaption is surpassing the ability to secure it (ibid). Cybersecurity is not anymore a personal dilemma but of the entire organization, the country (ibid) or even entire humanity.
Report shows that a considerable number of Facebook users are aware of their data easily accessible to others both knowingly and unknowingly. But despite the awareness, but not all have taken major steps to protect their accounts (Perrin, 2018). Some social media account holders feel they are secured but in reality they are not as there is no truly private account as Facebook themselves mines user's data, third-party developers, etc. (Steel & Fowler, 2010; Patterson, nd). Even non-Facebook users are made public thru the Facebook’s friends (Newcomb, 2018).

Unless the user is making up a pseudo or fake account, user’s need to reveal a certain detail about him/herself online to prove he/she is who he/she claims to be, an accurate representation of the offline context online (Lampe, 2007). The more information a user share the easier it is to predict some of user’s most private information using algorithms. In the US for example, a person’s social security number may be revealed or predicted at a certain level of accuracy by using combination of hometown and birthdate (Acquisti & Gross, 2009). The privacy-seeking behaviour of Facebook user have increased and the amount of personal data shared have decreased (Gross & Acquisti, 2013).

Disclosure of private information is high and is variable depending on many factors. Some reports says that the amount of disclosed information is 60% (Campbell et al., 2005), 55% (Yong, 2011), 97.1% (O’Brien et al., 2012), and 72.4% (Center for the Advancement of Social Science Research, 2013). Though it was pointed out that users are now reducing what information they shared but still the amount of shared data is still high. This information does not only include those that are allowed by the user to be revealed from Facebook itself but also from other applications attached to it. Facebook has close relationships with several corporations and they integrate their marketing efforts into the site by giving them special “Groups” for interested account users. This disclosure is legal, and users are receiving the use of an extremely useful and popular site for free in exchange for it. However, not all users understand the terms of the bargain: 46% of Facebook users believed that Facebook could not share their information with third parties (Grubb, 2011). Facebook, although hit with a lot of privacy suits recently, is more or less doing the best it can under the current ambiguity of privacy laws or merely lack thereof altogether. Many of the issues the social network has been sued over haven’t had clear cut laws to back up the suits. Many times Facebook has settled out of court, not out of fear for losing, but wanting to keep their customers happy and to move on from the issue. Every time there has been a major complaint about a privacy feature, Facebook has moved to correct the issue, even if what they were doing was not illegal—it wants to keep its users happy, and it wants to grow.

The benefits Facebook is providing to its users are dependent on the extent of information its user’s share. With smaller information available, some of the features Facebook offers may be restricted or will not fully function well (Cross, 2013). Perhaps, the created social capital by sharing personal information including the creation and maintaining of interpersonal relationship and friendship online is the most important benefit of online networks (Ellison et al., 2007). Online networks like Facebook uses this intimate information traded and exchanged for the service (ibid), a ‘complicit risk’ of personal information (Ibrahim, 2008). There are three main issues
with privacy and trust over the Internet: visibility, accountability, and scale. A lot of what happens on the Internet is basically invisible. When ISPs, websites, or third parties collect data, that activity is usually hidden from the provider of that data. Second, when data travels from one computer to another or is combined with other data, information instructing how the data should be or not be used is not included. Data doesn’t have a tag on it saying which user agreement policies the provider consented to. If people or companies use the data in a way not agreed upon, there is usually no one who is automatically accountable. Third, the Internet connects people and organizations that can be hard to identify and from all parts of the world—the scale of the information sharing is intimidating.

The proportion of people protecting their data is very varied depending on the location and age, some are contradicting. Christofides, et al. (2010) says its adult who likely tinker with their privacy setting on the other hand Madden et al. (2013) is the opposite. Extent of data sharing or protection is more likely related to the level of risk and knowledge about privacy settings.

5. Conclusion and Recommendation

Privacy awareness does affect lightly user’s privacy control activities. Facebook users are anxious who sees their posts but does little serious to limit access of their private data. User does try to make settings of their posts into private and limit the viewing to friends or specific people or group. However, this private information (picture or other posts) leaks through their friends whom without approval from people partly owning the data/information. Privacy and control awareness increases every time a glitch or an issue is raised against Facebook.

Without knowing it, users very private activities within and outside the Facebook environment such as pictures/videos/posts clicked and viewed, ticked like (and other reaction buttons), private conversations, and many more are recorded.

Facebook has very rich web of very crucial personal data derived from various activities within the system’s environment and even outside it. These data sets are very important source of revenue for Facebook as this is used for targeted advertisements. As such, Facebook is very good avenue for data mining activities.

Facebook memes including quizzes are ways to harvest data sets for purposes only the data miner knows.

6. Bibliography


Facebook (nd). What information does Facebook get when I visit a site with the like button. https://www.facebook.com/help/186325668085084 October 15, 2018


Facebook Newsroom (nd.) Stats. https://newsroom.fb.com/company-info/ Accessed October 3, 2018


