

A Study of Current State of Privacy Policies in Social Networks

Pranay Garg*, and Chethana R Murthy

Department of Information Science and Engineering RV College of Engineering, Bangalore-560059, Karnataka, India pranaygarg2013@gmail.com, chethanamurthy@rvce.edu.in

Abstract

Objective: The study aims to bring forward the hardship in understanding the current scenario of privacy policies in social networks from a user/customer's perspective. **Methods:** To understand how privacy policies have evolved through time, two of the biggest and most influential social networks, namely 'google.com' and 'twitter.com' have been studied. Their privacy policies have been analyzed from the time of inception, carefully taking each change into account. Other metrics like word and technical jargon usage has also been considered. **Findings:** Policies are getting harder to read with time. This includes increasing amount of ambiguous statements like, 'we may', more technical language, decreasing number of examples, and exponentially increasing length of the policies. There's also the practice of splitting the privacy policy over multiple pages that social networks are using in order to make their policies small and readable, which has a negative effect. **Applications/improvements:** The study restarts an open discussion on how policies should evolve with evolving inclusion of technology in people's lives. As people rely more on technology than ever, our study shows that rethinking the way we share privacy policies is required.

Keywords: Open Social Networks, Privacy, Privacy Policy, Security, Terms of Service, User Awareness.

1. Introduction

Internet has changed the way humans interact and exchange information in an astonishing way. Interacting with a person in real time verbally, living across the oceans sounded magical if not insane and unrealistic just few decades ago. Fast forward to today, we now live in a world where electronic interaction has overtaken conventional methods of communication, such as physical interaction¹. The idea of a future where everyone is electronically connected to one another and share their ideas in real-time does not seem to be unrealistic anymore. Today Open Social Networks (OSN) has become the medium of mass interaction between people. More than 30% of world's population actively uses one or the other social media platform². It enables us to interact and make impact on a global level.

1.1. What is OSN?

OSN, just like air is a medium that transfers knowledge, but unlike air we can make global impact here. As indicated by multiple studies, 'people influence people'. With the global outreach that OSNs provide, they have undoubtedly become the most used and influential source of information or misinformation in history and the visible future.

OSNs are used for multiple purposes which can span from sharing personal information with family and friends, all the way up to sharing political views with the public. All the top used OSN are free to use, with most charging nothing for full features³. One may wonder, then where does the money come from? Inarguably, the answer for most OSNs is customized advertisements and user data. People, who think they are using the product, might just be the product themselves.

*Author for correspondence

The customer information is collected by OSN itself, third party applications (apps) or by application providing some service free of charge. The exchange of this sensitive information is governed by the Terms of Service (TOS) or privacy policies which users are mostly unaware of, or which they simply ignore. In most of the cases, users have no idea of the data that they are mined of and how it's used. Many of the smaller applications have no policies bound to them and customer data is exploited as per the developer's will. Thus, customers have little to say no on what they want to share and what they don't. People are left on the mercy of the companies who vaguely express the policies and exploit the data with little to no consent, the way they see fit.

In the current reality however, it's more important than ever for people to have a say and contribute to the policy's structure in a way that companies can leverage. This is all because, OSNs like Facebook, Twitter and Google have indeed become an intricate part of our lives, without which the 'Normal' would no longer be possible. Studies have proven that with OSNs, opinions of individuals can be influenced as per likings, which have resulted in a grey area when it comes to the involvement of the owners in such practices⁴. User data as anybody would guess plays a significant role in this. OSNs collect user data for known reasons like customized advertising and user satisfaction with prior consent. This paper discusses the problem that lies in the way policies and terms are disclosed, specifically the language and the medium.

In the following sections the subject mentioned above is discussed in detail and we try to point out the facts which the current system is lacking. Section 2 discusses the need of privacy policies and who sets the standards for them. Section 3 details the few of the recent statistics of data sharing from the customer point of view. Section 4, discusses few possible steps of remediation, followed by a conclusion in Section 5.

2. Policies

Terms of Services (TOS) are a set of rules or directives a user must abide to use a service provided by another party⁵. TOS can impose directives and rules that must be followed while using the service, not following which the use can be termed as 'unethical', giving the party to terminate the service at the least. Privacy policy is an intricate part of TOS. It details the personally identifiable informa-

tion which is being collected from the user⁶. Collection of personally identifiable information is always been a topic of tender discussion. Personally identifiable information can be used in host of commercial ways, which may be a breach of one's privacy⁷.

Major social networks like Facebook (www.facebook.com) have become a hub of personal information, with information of people over the globe and across societies. It's always been speculated that personal information on these mediums can be exploited for political agendas⁸. This is no longer science fiction with the Cambridge Analytica Fiasco in 2018. This a massive political campaign powered by the wealth of personal information⁹. It has undoubtedly become a hot area of active discussion.

Without a doubt, we need set of rules in our country's constitution which mandates proper disclosure of the data that is being collected and how exactly it's being disclosed and used with other parties. For this very reason, California's 'California Online Privacy Protection Act (CalOPPA)' mandates that any website collecting any kind of personally identifiable information, must have privacy policy containing specifications of the information being collected in their terms of services that is publicly available¹⁰. Indian IT Act of 2008 explicitly states the requirement of privacy policies on the company's website¹¹. Lastly, General Data Protection Regulation (GDPR) which was recently introduced by European union emphasis on the requirement of privacy policies¹². Most of the countries have rules specific to this area in their constitution.

One question that we might have is, if the privacy policies are indeed available and required, how did something like 'Cambridge Analytica Fiasco' happen?¹³. Why do people feel cheated when everything is in the privacy policy?

3. Case Studies and Results

As per the regulations norm, privacy policies have come a long way through the years. As the OSNs matured, the regulation around them have evolved as well. OSNs like Google (www.google.com) and Facebook (www.facebook.com) are much more diverse than they used to be. New OSNs like Twitter (www.twitter.com), Instagram (www.instagram.com) and SnapChat (www.snapchat.com) have come up in the last decade with the advancement of smartphones.

As per the survey conducted by Pew Research Centre (Table 1), most of the Americans used one or the other OSN daily. It was also noted that 74% of Facebook users say they visit the site daily, while 51% said they visited the sites multiple times a day¹⁴. As we see, all the top OSN have a privacy policy attached to them. Among these OSNs, the median privacy policy is around 3900 words long, which would take a person 15-16 minutes to read with a pace of around 250 words per minute. That doesn't seem to be very long, but as we will see, it's only part of the picture.

Table 1. Analysis of OSN privacy policies

| Open Social Network | Adoption (% of people in US)* | Length of Policy (in approx. words)** |
|---------------------|-------------------------------|--|
| Youtube | 73% | 4,008 (Google Privacy Policy – redirect) |
| Facebook | 68% | 5,252 (Privacy policy + Data Policy) |
| Instagram | 35% | 7174 (Privacy policy + Data Policy) |
| Pinterest | 29% | 3789 (Privacy policy) |
| SnapChat | 27% | 3844 (Privacy Policy) |
| LinkedIn | 25% | 6010 (Privacy Policy) |
| Twitter | 24% | 3838 (Privacy Policy) |
| Whatsapp | 22% | 3499 (Privacy Policy) |

* as revealed by 'Pew Research Center 2018'

** The word count belongs the home page of the policy. We did not count the words in the children pages of the policy.

Privacy policies and TOS are being read only by an aging few. This is evident from a survey conducted by Axios and SurveyMonkey in February of 2019 on four thousand American people. The survey revealed that over 56% of people always accept the privacy policy without reading them¹⁵. At the same time, study also revealed that about 87% say it's either very or somewhat essential to

read them. The question arises why is there such a gap between the emotion and the behaviour?

Another study in 2008 revealed that it would take 76 Work Days leading all of the digital privacy policies they agree to in the span of a year¹⁶. This number is from 2008. Without a doubt, the length and complexity of the privacy policies have risen over time.

Since our lives are so bound to OSNs in this connected age, it is of utmost importance that we understand the root cause phenomenon and solve the problem at its point of inception. In our journey to understand why are privacy policies not being read, we take a look at how privacy policies have evolved over time through inception of two giant OSNs. In the cases studies below, we would try to highlight the features which take the policies especially undesirable to read and hard to comprehend. We now take a look at the analysis of case studies that we did for Google and Twitter privacy policies.

3.1. Google Case Study

Every one of us undoubtedly uses Google and its services. It's logical to argue that the involvement of Google in our lives has undoubtedly become a new normal. There are currently 7.7 billion people living on the Earth, and Google's Android platform itself has 2 billion active users¹⁷. Thus, we studied the privacy policies that Google published on its website. Google keeps archived every version of its privacy policies, which we will be using to compare, contrast and understand how the policies have changed or evolved over the years¹⁸.

3.1.1. Ambiguity

From the very early privacy policies to the very latest one, all seem to have intentional ambiguity in them. For example, Google's January 2019 policy says, "Remember, when you share information publicly, your content may become accessible through search engines, including Google Search." This statement may be justified as, the user's data will only be accessible if he has not changed the default setting to disallow such behaviour. But the importantly, it's seldom that an unknowing user knows where to look for the settings controls.

At the same time, another example, "We may share non-personally identifiable information publicly and with our partners — like publishers, advertisers, devel-

opers, or rights holders”, notes that information which cannot identify the user may be shared. However, we couldn’t find any publicly available information on what data is being shared and any guidelines that it is following to term it as “Non-Identifiable”. While their explanations have improved overtime and have become specific, but the use of phrase “we may” have grown through time and this worries us.

3.1.2. Examples

Examples are a critical part of the privacy policy. Privacy policies at times cover highly specific topics like, technically what personally identifiable data is being collected for an instance. It’s often that many normal users would require examples rather than definition of technical terms in order to comprehend the type of data being collected.

For instance, a cookie can tell us, ‘This is the same computer that visited Google two days ago,’ but it cannot tell us, ‘This person is Joe Smith’ or even, ‘This person lives in the United States,’ was an example used in the initial phases of the privacy policies to describe how the cookies might be used. However, this example is no longer present in the latest revision of the policy. It’s logical to argue that the models and architecture have advanced quite a lot over time and such examples are no longer valid. However, we have to understand that some technical information about cookie doesn’t help much. In the latest iteration, Google has even made use of videos to ease the process, which seems to be a step in the right direction¹⁹.

It is important to understand that at times, people might not be able to comprehend the significance of the data being collected and the uses that can come out of it. This makes the existence of examples way more relevant since it enables common users to understand the implications and uses of the data they share. This is something which we still find lacking in the system.

3.1.3. Length & Links

Length of the policy would be another interesting parameter to look at. As studies indicate, people don’t read very long articles of text²⁰. This is a significant factory why people do not and in reality cannot read the privacy policies of all the services they use¹⁶. With the ever growing reliance of OSNs on user data, it shouldn’t be a surprise,

if with time, the size of the privacy policies also grew. In (Figure 1), we see the growth of number of words in the privacy policy page of Google over time. The increase seems to be mostly linear over time until recently when it started to fluctuate a little. This peaked our interest. How can the privacy policy size reduce so drastically when OSNs reliance on user data is always increasing?

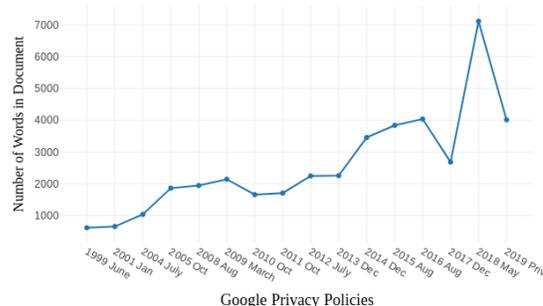


Figure 1. Word count over Google privacy policies.

As we were able to figure out, the actual size of the policy is increasing, but parts of policy have been now split into multiple links spread over the webpage. Each little term now has a popup explaining more about it. As shown in (Figure 2), for the latest privacy policy, the base page is 4000 words long and it keeps increasing as we traverse through the links. Expanding only few sections of the policy, we see that the numbers sharply rise, and the policy becomes unreasonably long.

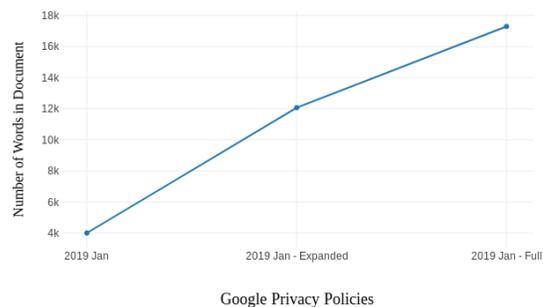


Figure 2. Google privacy policy 2019 links expanded.

While first small page may decrease the overall initial burden on the user reading the policy, we believe that it does more harm to the user than good. As studies indicate, people seldom open more than few links on a webpage²¹. Going by this matrix, we believe that from the small fraction of users who read the policy, even smaller fraction read the entirety of it.

3.2. Twitter Case Study

Twitter is one of most active social platforms on the Internet. It is arguably the most influential of all OSNs. It is often that influential as that of people and politicians talk about what they are thinking on the networks and get criticism and at times, recognition for the same²². Many people share what they see and believe on Twitter, at time this could be fake or real. Researchers have shown that the data people share in real time often turn into news and worthy information²³. When you have so much of inflowing data, which could be ground zero information, it is important on how you handle that information.

Since the model used by Twitter is largely public, they also provide API to scrape the data which its users are generating. This provides for a small and easy to understand privacy policy. The structuring of the privacy policy²⁴ provides simple brief information on the use of private data.

However as in case of Google, a significant chunk of how location, cookies and other parameters tied to personal data are used, are still split across multiple links, which we believe should be a part of the main privacy policy page²⁵. Twitter's policy page's length seems to have been linearly growing in length as seen in (Figure 3).

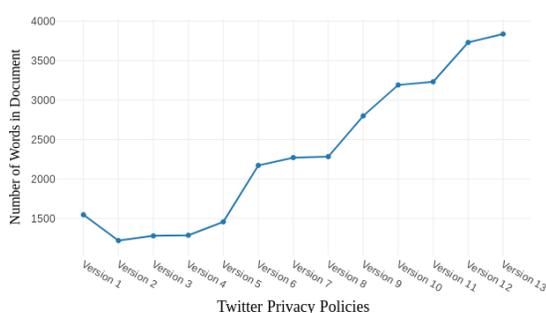


Figure 3. Twitter privacy policies word count

From the discussion above, there's no doubt in the fact that there are gaps in the system. While the major discussion has been using the case study of Google and Twitter along with other major OSNs, in no way we are targeting the flaws in their policy. The reason to pick them is simple, they are the major players. They have the most impact on people's lives and thus play an important role. There's no doubt that as OSNs expand, their use of private data will increase for any number of reasons. No matter how good of a job OSNs do, the policies are bound to grow in size and complexity. Hence, maybe it is time for us to change the way we think, handle and communicate privacy policies.

3.3. Disclaimer

The methods used to calculate the statistics above are very open to interpretation as one would argue. All the scripts used to generate the statistics along with a dump of the scraped data will be available on GitHub for open review²⁶.

4. Discussion and Future Scope

Now that we have seen the problems that exist in the space, the next logical question that arises is "how do we solve the problem?". From all the study above, it is evident that even though all the effort that has been put to create informative policies, they are not serving their intended purpose. This leads us to believe that maybe the medium used to convey the policies isn't apt anymore. Though companies like Google have tried to cover parts of their policies in animations and videos, that model seems to be ineffective as well.

4.1. Future Scope

We believe that a new standard, regulated by government has to be created in-order to make sure that the privacy policies meet a set of standards. This can be something very similar to hallmark or ISI mark but something that people can relate to and use to understand what's happening with their data.

Though similar feats have been tried by organizations like <https://tosdr.org/> which try to identify the quality of the privacy policy along with the main points. Since economy is largely becoming data driven[S42], it's now

more important than ever for the government to know what data is being collected about its citizen, and how it is being used. Right now, no medium exists through which customer can plea and tell the companies what data they want to share and what they don't. While being a gateway to transfer citizen's will, the government organization can also help simplify the privacy policy to a level that all citizens can understand. A system with levels of data access would be good too.

5. Conclusion

The goal of this study was to analyze the privacy policies of the major OSN from inceptions and identify the trends which are pushing the readability of the policies down. The discussion intern allowed us to deep dive into the negative trends that privacy policies are facing today. These trends (length for example) are decreasing the number of people who read the policy. Effectively killing the policy they were made to enlighten. Our in-depth analysis re-enforced the known fact that, length of privacy policies is growing out of control and shows no sign of stopping until and unless a fundamental change is made in the way to disclose and share privacy policies. Our study also exposed the policy splitting technique that social networks are using to make their policies look smaller. We were able to connect this trend with habit of the users to conclude that splitting the policy is not going to work. People who rarely open more than 3 links will seldom read the spread over privacy policy.

6. References

- Pierce T. Social anxiety and technology: Face-to-face communication versus technological communication among teens. *Computers in Human Behavior*. 2009; 25(6): 1367–1372. <https://doi.org/10.1016/j.chb.2009.06.003>.
- Number of social media users worldwide 2010-2021 [internet]. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. Date accessed: 2019.
- Global social media ranking 2019 | Statistic [internet]. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. Date accessed: 2019.
- Kuan K, Zhong Y, Chau P. Informational and Normative Social Influence in Group-Buying: Evidence from Self-Reported and EEG Data. *Journal of Management Information Systems*. 2014; 30(4): 151–178. <https://doi.org/10.2753/MIS0742-1222300406>.
- AOL dumps new member policy [internet]. <https://www.cnet.com/news/aol-dumps-new-member-policy/>. Date accessed: 29/07/1997.
- Privacy Policy [internet]. https://en.wikipedia.org/wiki/Privacy_policy. Date accessed: 09/05/2019.
- Graham JP. *Privacy, Computers, and the Commercial Dissemination of Personal Information*. 1986.
- Wills D, Reeves S. Facebook as a political weapon: Information in social networks. *British Politics*. 2009; 4(2): 265–281. <https://doi.org/10.1057/bp.2009.3>.
- Kozłowska I. Facebook and Data Privacy in the Age of Cambridge Analytica [internet]. <https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/>. Date accessed: 30/04/2018.
- Consumer Federation of California [internet]. <https://consumercal.org/>. Date accessed: 2019.
- Ministry Of Law, Justice And Company Affairs [internet]. <https://meity.gov.in/content/ministry-law-justice-and-company-affairs-legislative-department-0>. Date accessed: 09/06/2000.
- The Data Privacy of KPMG Portugal [internet]. <https://home.kpmg/pt/en/home/misc/privacy.html>. Date accessed: 22/04/2019.
- EPIC - In re Facebook - Cambridge Analytica [internet]. <https://epic.org/privacy/facebook/cambridge-analytica/>. Date accessed: 2019.
- Social Media Use 2018: Demographics and Statistics [internet]. <https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>. Date accessed: 01/03/2018.
- Hart K. Privacy policies are read by an aging few [internet]. <https://www.axios.com/few-people-read-privacy-policies-survey-fec3a29e-2e3a-4767-a05c-2cacdcbaecc8.html>. Date accessed: 28/02/2019.
- The cost of reading privacy policies [internet]. <http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf>. Date accessed: 2008.
- World Population [internet]. https://en.wikipedia.org/wiki/World_population. Date accessed: 09/08/2019.
- Privacy Policies [internet]. https://en.wikipedia.org/wiki/Privacy_policy. Date accessed: 09/05/2019.
- Cookie [internet]. <https://en.wikipedia.org/wiki/Cookie>. Date accessed: 03/03/2019.

20. Shaikh AD, Chaparro B. A Survey of Online Reading Habits of Internet Users. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2004; 48(5): 875–879. <https://doi.org/10.1177/154193120404800528>.
21. Lee J. No. 1 Position in Google Gets 33% of Search Traffic [internet]. <https://searchenginewatch.com/sew/study/2276184/no-1-position-in-google-gets-33-of-search-traffic-study>. Date accessed: 20/06/2013.
22. Kahne J, Bowyer B. The Political Significance of Social Media Activity and Social Networks. *Political Communication*. 2018; 35(3): 470-493. <https://doi.org/10.1080/10584609.2018.1426662>.
23. The New Media's Role in Politics [internet]. <https://www.bbvaopenmind.com/en/articles/the-new-media-s-role-in-politics/>. Date accessed: 2018.
24. Twitter Privacy Policy [internet]. https://twitter.com/en/privacy/previous/version_13. Date accessed: 20/04/2019.
25. Our use of cookies and similar technologies [internet]. <https://help.twitter.com/en/rules-and-policies/twitter-cookies>. Date accessed: 2019.
26. Privacy_Init [internet]. https://github.com/pranaygarg/Privacy_Init. Date accessed: 04/03/2019.