... I will pay with my privacy

Gopikrishna Rathinavel
rgopikrishna@vt.edu

Gaurav Dixit
gdixit@vt.edu

Abstract

Extensive use of email services and social networks offered by private companies has become a norm in today's contemporary world. Not many realize that these companies are private and what it means. The authors of this paper conducted a survey and found that most of the participants have not completely read the SLAs for these services. Yet most participants store their private data in these cloud-based services. This raises ethical issues. Current paper studies this intricately balanced ecosystem and point out current and future potential ethical issues. The paper then further comments on various aspects of these services and their related stakeholders. Finally, the paper suggests possible remedies to this ethical dilemma.

1 Introduction

In modern times, email and digital connectivity has become as essential as any other essential thing. People need internet connectivity as much as they need housing and food supplies. Internet outages become major news cen, and is treated with same urgency as a disaster. The email usage has risen considerably in last decade. Smart phones have added to this tremendous growth in number of users of email services. Companies like Google, Outlook and yahoo are among key organizations in this field.

Most of email service companies provide free email services. So much so that many people believe email is inherently a free service. This forms the root of an ethical dilemma. Individuals consider Terms of Service, a nuisance. In the survey conducted for this paper, the authors found that most individuals miss the "got you" clauses. Such widespread ignoring of such Terms of Service for common commercial email services and social networks can potentially lead to serious privacy and ethical issues.

The users of such services store some of their very private conversations and pictures in these services. All this is done without carefully reading the Terms of service gma yah out for these services . As per their inherent nature, these services are cloud based. Once some data is uploaded to the cloud, it is practically impossible to take it down from there. There are backups that happen everyday, every hour, every month, in multiple locations and in multiple forms. Even if company claims that the data has been deleted, there is no way to practically ensure complete deletion of data.

Another problem is, training of machine learning algorithms using the private conversations of users. Using sophisticated mechanisms and algorithms the identity of users can sometimes be deciphered. And, as a matter of fact, all these companies do read the conversations and train their machine learning algorithms over this private data. Thus, even if the data was deleted from database, there is a distinct possibility that in future an algorithm can connect different data in a trained machine learning system to discover the identity of original user. In essence, such advanced algorithms can reconstitute the data which was supposed to have been deleted.

There are many other aspects which are important and must be understood by users from the Terms and Conditions documents. For example,

- How long will the data persist after deletion?

- Will the company share my data with other entities? Government? Third party companies?
- Will my data be used for purposes that I do not approve of? Violence? War? Discrimination?
- Will these companies who now own my data, ask me money for retrieval of my data by myself? in future, after 10 years?
- Is there a mechanism for me to ensure that data that I deleted has actually been deleted? Are there any audit records publicly available?

There are many more questions like these. The questions are not for a particular company, but rather for email industry in general.

Further, the terms are very misleading sometimes. For instance, some of these services outright announce that the users are owners of their data. But in "rights of the company" sections practically all uses of data have been implicitly granted permission to. Thus, it becomes very important that the entire document must be studied thoroughly to avoid any misunderstandings on users' part. Since, if such a service reserves the right to display your data, in reality most of them do, there is not much meaning left to ownership of data by an individual user.

Another aspect that authors noticed is, studies like this are not easily available to the public as most search engines are also controlled by these private companies and have possible influence over results. Like any other industry, mishaps happen in data industry too, but no such spotlight has been observed in common mass media outlets.

Thus, skipping reading of the complete Terms of Service and not understanding in detail the related implications, creates potential very difficult to resolve ethical issues.

2 Related Work

Authors found very little related work in this field of ethical issues arising due to public use of private digital services. There have been few antitrust proceedings against these big names in digital communications domain. One good research paper is Obar and Oeldorf-Hirsch [2020] titled "The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services". It surveys on a fictitious social networking service, NameDrop. It investigated privacy policy (PP) and ToS reading behavior of the people. The findings were interesting. About three quarters (74%) skipped reading the PP and performed quick-join. Users read a 15-17 mins long ToS in 50 secs. Most (93%) agreed to ToS without understanding it in full text. Almost all (98%) missed important clauses regarding data sharing. The authors of Paine et al. [2007] talk about privacy concerns and privacy actions. Then there are research papers Xiao and Xiao [2012] and Ion et al. [2011] which talk about privacy concerns in cloud. The focus of research in many papers has been the possible loss of privacy in cloud platforms. Though there is need for research in ethical issues arising due to extensive use of cloud resources but lack of public awareness of privacy issues.

3 Experimental Setup

Authors conducted a short survey to understand people's understanding of privacy policies and terms of service. The participants consisted of students from Virginia Tech, friends from USA and from India. The survey consisted of 13 questions of which 6 questions were related to information on the individual filling the survey and the rest regarding our topic of interest. Authors have tried to obtain maximum information from minimum number of questions. The survey was designed in such a way that the questions answer some of the most important points when it comes to privacy policies.

Normalized value counts ratio are calculated as follows. Value counts of each item in a column, for example gender is calculated and normalized. After a filter is applied to a question, for example, participants who responded 'Yes, I read and understood the documents completely' for the reading terms and conditions question, in this population gender value counts are calculated and normalized. This normalized counts are divided by the normalized value counts of the entire population found previously. This value is termed as normalized value counts ratio.

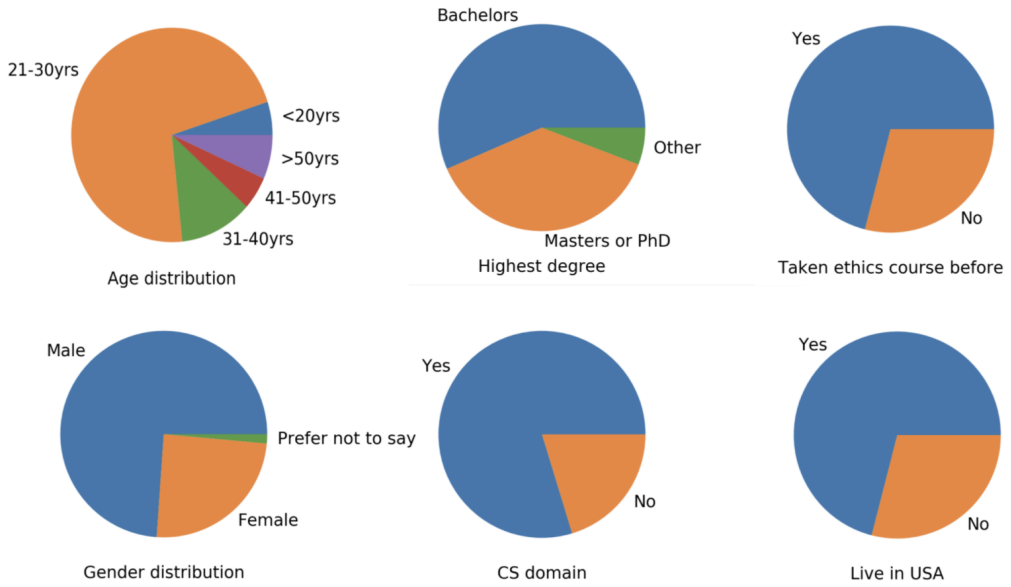

Figure 1: Demographics of the survey participants.

4 Results

The participants consisted of 72 people. Figure 1 shows the demographics of the participants of the study. Most participants were in the age of 21-30yrs. Almost 75% of the participants were male and 24% were female and one person did not want to reveal their gender. 55% of the participants have Bachelor's and 37.5% have Master's or PhD as their highest degree. 72% people live in US and 28% live outside USA (mainly in India). The two most important features we are interested in are if a participant is from a computer science domain and if a participant has taken an ethics course previously. 80% are people from the computer science domain and 73% participants have taken at least one ethics course previously.

Our questions can be split into two parts. One set of questions regarding terms and conditions and another set regarding email services and policies. We have analysed the results of the survey in two directions. One is to directly count what everyone has answered for each question. The other one a one vs other analysis where we contrast an answer from a group of individuals from another group.

During analysis of the survey we make a very important assumption that the participants were fair in answering the questions.

4.1 Reading terms and conditions

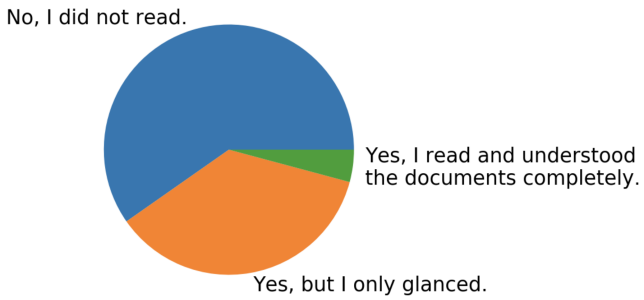

Figure 2: Reading terms of conditions

Figure 2 shows the answer for the question asking if the participant read the terms of service while creating email service accounts and if yes, if they read it partially or completely. More than 60% of the

participants as seen have responded that they did not read the terms and conditions and 36% answered that they only glanced through the terms of service. The remaining 4% participants answered that they read and understood the terms of services fully before creating accounts. This 4% which consisted of 3 individuals seemed interesting and we wanted to check if there is any commonality in those participants. The only common feature that we were able to identify is that all the three were male participants, but since a much larger number of males answered the survey this result might not be representative of entire population.

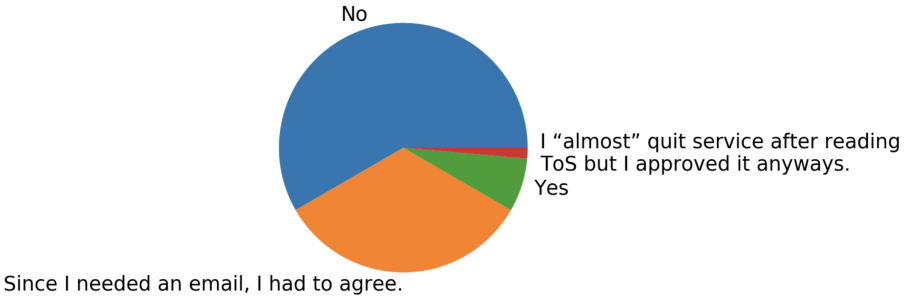

Figure 3: Declined due to Terms of Service.

We then asked if participants declined creating email accounts due to some disagreement or something they did not like in the terms of service. Figure 3 is a pie chart of the count of various answers. The answers for this question agree with the answers from the previous question as most participants chose not to read or only glanced. **It is interesting to note that lot of participants chose the answer "Since I needed an email, I had to agree". We can infer from here that email services are a necessity in current world and one simply does not have a choice but to accept such terms and conditions. Another important point to notice here is that 2 out of the 3 participants who answered "Yes, I read and understood the documents completely.", have declined to create those accounts.**

4.2 Email services

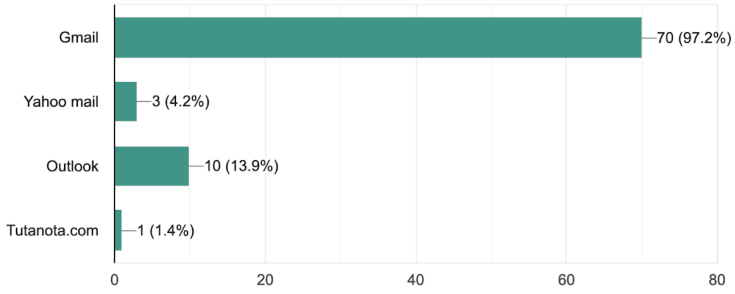

Figure 4: Commonly used email service.

Table 1: Normalized value counts ratio for participants who did not believe email services use their data

Questions	Normalized value counts ratio
I “almost” quit service after reading ToS but I approved it anyways.	None
No	0.43
Since I needed an email, I had to agree.	1.50
Yes	3.60

From Figure 4, we can see that most people have Gmail as their email service. Figure 5 shows the distribution of answers for the question "Do you know if your email service (Gmail, Yahoo, etc) uses

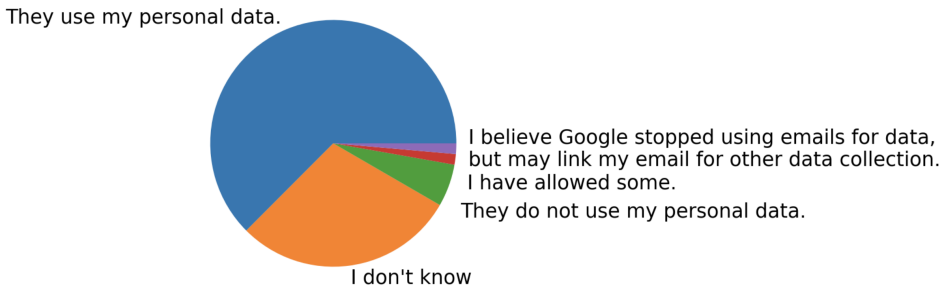

Figure 5: Do they use personal data?

your personal data (emails, chats, photos, files, etc)?". For participants who gave the answer "They do not use my personal data.", we normalized the value counts of the question regarding declining email accounts. It was interesting to notice from this analysis that large portion of people who answered that they declined creating email services after reading ToS, answered that the email services do not use their personal data (Table 1). **We can infer from here that, even people who decline who after reading ToS are not aware that the email services use their personal data.**

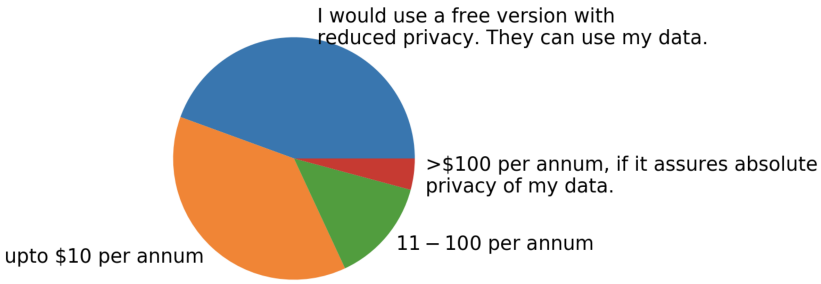

Figure 6: Will you pay for email services?

Table 2: Normalized value counts ratio of age for participants who said they would pay $11-$100 per annum for email services with privacy protection.

Questions	Normalized value counts ratio
21-30 yrs	0.82
31-40 yrs	1.60
41-50 yrs	2.40
<20 yrs	None
>50 yrs	1.80

Table 3: Normalized value counts ratio of answers to declined due to Terms of Services question for participants who said they would pay $11-$100 per annum for email services with privacy protection.

Questions	Normalized value counts ratio
I “almost” quit service after reading ToS but I approved it anyways.	NaN
No	0.685714
Since I needed an email, I had to agree.	1.200000
Yes	2.880000

The figure 6 shows the proportion of answers for the "If Gmail or other email service offers its service in exchange for money and preserves your privacy, how much would you be willing to pay?". It can clearly be seen that most participants either wanted the service for free or they are willing to pay $10

per annum for email services which offer data protection. However, it becomes important to analyse which demographic of the participants are willing to pay more money for the email service. **From Table 2, we can see that participants of the age 41-50 yrs are more willing to pay more money ($10-$100). Same group of participants also answered that they are willing to pay more than $100 per annum.** Another interesting finding is that participants who are willing to pay more money for email services have also declined creating email accounts due to ToS.

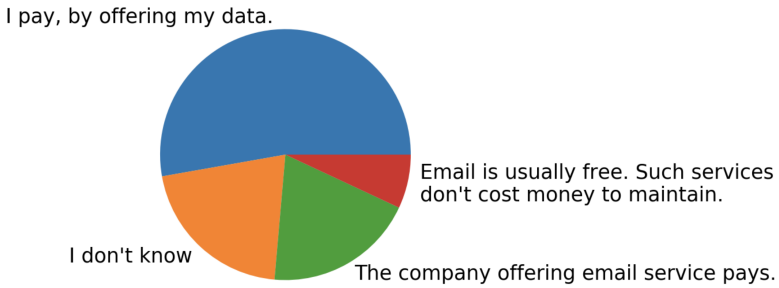

Figure 7: Who pays for email services?

We then asked the participants who do they think pays for the email services. In reality, we are not offered the service for free. The companies use our personal data in many ways for their advantage. They can even contract such data to third party. Figure 7 shows the proportion of responses. More than 50% of the participants think they pay by offering their data. An important, rather amusing observation from the value counts analysis of age of participants similar to above but table not included) is that participants with age <20yrs have responded that they think email services are free. The normalized value count value for this observation is 4.8. **This indicates the lack of awareness in younger participants.**

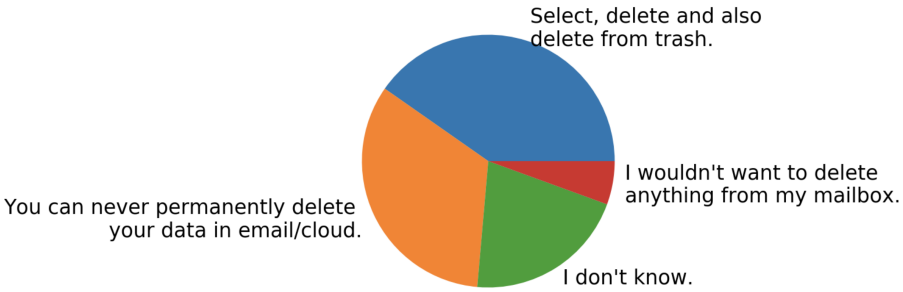

Figure 8: How to permanently delete an email?

The final question was asked if we can actually delete emails permanently. Figure 8 is a pie chart of the responses. To our surprise, almost equal number of participants responded that they can "select and delete and then delete from trash" and that "emails can never be permanently deleted".

4.3 Company privacy policies

Just as a sample, authors discuss a portion of one such company's Terms of service. Again, authors trust that company the most, the discussion here is for benefit of reader, as an example of industry practices.

The ToS says: The company has right to host or distribute the data from users, to provide better service. The company has rights to create derivatives of your data.

Though this is said in relatively simple language, a variety of meanings can be assumed from these statements. The users at large might be completely unaware of these statements. The users even if they read such statements might not be fully capable of understanding the implications.

5 Discussion

5.1 The company

This study is to bring out potential ethical issues due to avoidance of reading ToS for email services. The study is not about a specific company but rather about the existing practices in current times. The authors found that most of the companies in the email services business behave similarly and have similar rules. For example none of the company explicitly lists how they delete the data from all their backups once the user deletes it. They have very similar other rules. This study tries to bring out the ethical issues related to email service.

Having said that the authors appreciate these companies for providing a valuable service in a form which is currently acceptable to all users. The companies and the entrepreneurial efforts are an important part of economy.

Companies should be promoted for making economic activity happen. From their perspective, they are in business to make profits in a way which is acceptable to the society at large. They are thus fulfilling their obligations towards their owners and equity-holders. Authors bring out a point that maybe this industry grew organically. This led to a growth in practices which were never reviewed. There is potentially a time now to review the practices.

Email provider companies(gmail, outlook, yahoo, others): Authors believe they are just another corporate companies. These enterprises exist to make money for their stakeholders.There is nothing bad about working hard to make profit. The investors of these companies trust these entities to work for them. These companies hire some of the best and brightest minds to create algorithms and create relevant software for them. It just so happens that the business they are in is yet to get understood by society in general.

Authors are sympathetic to the businesses. In absence of any clear indication against these businesses, authors trust these companies with their own data. Since these companies research at the forefront of the technology, the users or the government might have inadequate information and tools to supervise them. Thus, these companies have to be trusted for now. Having said that, if at a later time such companies were found to be willfully damaging user's privacy, this would constitute a breach of trust. Again, large companies in this domain (like gmail and outlook) are well trusted by most well educated people.

5.2 The User

A wide range of users use such systems. There are scientists, data scientists and data engineers that use such services, who understand most aspects of the data industry by virtue of their training. Then there are teenagers, manufacturing sector employees and elderly people, who have not been given any training in these domains. The rich sections of society have time and resources to discuss and decide if they should pay for such services. On the other hand, there are unemployed youth who need a working email-ID and cannot afford to pay for such services. There are many other ways to group users. Each section of the users have their own requirements, and they are vulnerable in their own ways. Authors recommend a concerted effort to educate users. Also, the different user groups should be provided with different classes of services. For example, a teenager would have a very different requirement in email services as compared to an octogenarian.

5.3 The government and the laws

Authors believe the laws have not been formulated for the newer technologies. It would take many years before a complete set of the laws relevant to data services come into existence. The authorities might need to be a bit compassionate towards large sections of the society, if they commit any minor mistakes, since large portions of the society do not have a very detailed understanding of the privacy and related policies. Authors would like to point out that authorities should take into cognizance the fact that most people are not well aware of their own data privacy and its implications.

6 Conclusion and further studies

There is a clear conclusion that most of the users were not aware of the inner working of the digital services and the potential issues that can arise from saving personal data in cloud. The ethics course and studies of computer science in their current format helped a bit, but more is required. Authors of this paper felt the need of focused education for students in this aspect of computer science.

6.1 Recommendations

Authors have following recommendations:

1. Companies should make public the exact procedure they follow to dispose off the data. Authors note from the results of the survey that only a few people understand the complexity of deletion of data from cloud.
2. The users, when they are creating new account should be shown a video detailing dangers of sending personal data to cloud. As seen from the survey results, the users lack understanding of data usage by the service providers.
3. The companies should clearly mention how and why paid and unpaid accounts differ in services. The terms should be in explicit and lucid wordings which common people can understand.
4. Course on ethics should be mandatory with every computer science degree. As is evident form the survey, the understanding of these ethical issues must be imparted along with relevant degrees.
5. The large companies which offer digital services, are very similar to utility(gas, electric) companies. They inherently have responsibilities, which they should be explicitly made aware of.
6. More such projects, and at PhD thesis research level, should be undertaken to understand the possible impact of inadvertent leakage of information from such data platforms.
7. Different services should be provided for different segments of society to better meet their needs and to reduce their vulnerability. For example, School students should have a different email service product than office employees, since they have different requirements and different levels of understanding of the digital world.
8. Quantization of privacy levels is also required. So that people can choose what level of privacy they want at what cost. For example, say at level 3, the user data will never leave the company's premises and no external data analytics will happen on user data, but users will have to pay 6 dollars per month.

References

Centurylink outage led to a 3.5% drop in global web traffic. URL https://www.zdnet.com/article/centurylink-outage-led-to-a-3-5-drop-in-global-web-traffic/.

Google privacy and terms. URL https://policies.google.com/terms?hl=en-US.

Verizon media terms of service. URL https://policies.yahoo.com/us/en/yahoo/terms/index.htm.

Microsoft services agreement. URL https://www.microsoft.com/en-us/servicesagreement.

Jonathan A Obar and Anne Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1):128–147, 2020.

Carina Paine, Ulf-Dietrich Reips, Stefan Stieger, Adam Joinson, and Tom Buchanan. Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6):526–536, 2007.

Zhifeng Xiao and Yang Xiao. Security and privacy in cloud computing. *IEEE communications surveys & tutorials*, 15(2):843–859, 2012.

Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. Home is safer than the cloud! privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, pages 1–20, 2011.