An In-depth Review of Privacy Concerns Raised by the COVID-19 Pandemic

Jiaqi Wang

Penn State University

jqwang@psu.edu

Abstract

COVID-19 has hugely changed our lives, work, and interactions with people. With more and more online activities, people are easily exposed to privacy threats. In this paper, we explore how users self-disclose on social media and privacy concerns raised from these behaviors. Based on recent news, techniques, and research, we indicate three increasing privacy threats caused by the COVID-19 pandemic. After that, we provide a systematic analysis of potential privacy issues related to the COVID pandemic. Furthermore, we propose a series of research directions about online user self-disclosure and privacy issues for future work as well as possible solutions.

Introduction

COVID-19 has spread across the world and affected how people work, live, and interact with each other. People are recommended or required to work remotely, quarantine at home, and keep social distance. Under these circumstances, people expect more interactions with others via social media platforms, which has led to a huge increase of social media usage (Holmes 2020). Based on a study (Kanter 2020) of 25,000 consumers across 30 markets published on April 3rd, 2020, WhatsApp has seen a 40% increase in usage; in the early phase of the pandemic usage increases 27%, in mid-phase 41% and countries in the late phase of the pandemic see an increase of 51%; Facebook usage has increased 37%. China experienced a 58% increase in usage of local social media apps including Wechat and Weibo. Another study of 4500 *Influenster* community members, most of respondents agreed that their social media consumption (72%) and posting (43%) have increased during the pandemic. Moreover, TikTok, one of new social media platforms, was used by the largest share of teenagers (48%), overtaking even Instagram (47%) from March, 2020 to April, 2020 (Perez 2020).

One possible reason is that people are searching for alternative approaches to interact with others to stay mentally healthy. People generate content, comment content, forward content, and communicate with others on social media platforms. To increase a sense of intimacy with others, people share details of their lives with text, pictures, videos,

live video streaming, etc. To a great extent, the content can reveal personal private information including age, gender, location, race, etc. Compared with interactions in the real world, self-disclosure information can more easily be propagated, searched, saved, and even processed on social media. The increasing and more abundant self-disclosure may cause unpredictable and unacceptable privacy disclosure to users online. Furthermore, a recent research shows that people's mental health problems are prevalent because of social media exposure (Gao et al. 2020) itself, which means the expected results might be on the contrary to the mental health cure.

However, the pandemic is changing people's sensitivity and attitude to privacy including what and how personal information can be disclosed (Nabity-Grover, Cheung, and Thatcher 2020). Discussion about COVID-19 may include basic personal information, travel schedule, test results, symptom description, and medicine in use. These acts of self-disclosure reveal a lot of sensitive information that people are not willing to share previously (Kordzadeh and Warren 2017). For example, health status and detailed description of individual body information are shared to ask for comparison, suggestions or pre-diagnosis. Some communities even encourage people to share more personal information related to COVID-19 in the name of society responsibility without clarifying the boundary of gathered information and how to use the collected data. Based on the observation, users would sacrifice personal information to a unprecedented degree to help the society back to the expected normal status. Recent work (Blose et al. 2020) provides early evidence that the situational factors caused by COVID-19 may affect people's self-disclosures and privacy calculus. There is another issue we need to pay attention to. Along with the COVID-19 pandemic, 2020 the United States presidential elections started from February and ends in November. Noting that the date when United States officially declared the COVID-19 pandemic as a national emergency is March 13 and the first statewide "stay-at-home" order was issued at California is March 16. That time is approximately only one month later than the early voting in February. During the whole process of the presidential election, people are isolated at home and keep social distance in essential activities at most time. People have participated extensively in political discussions, and actively engaged in
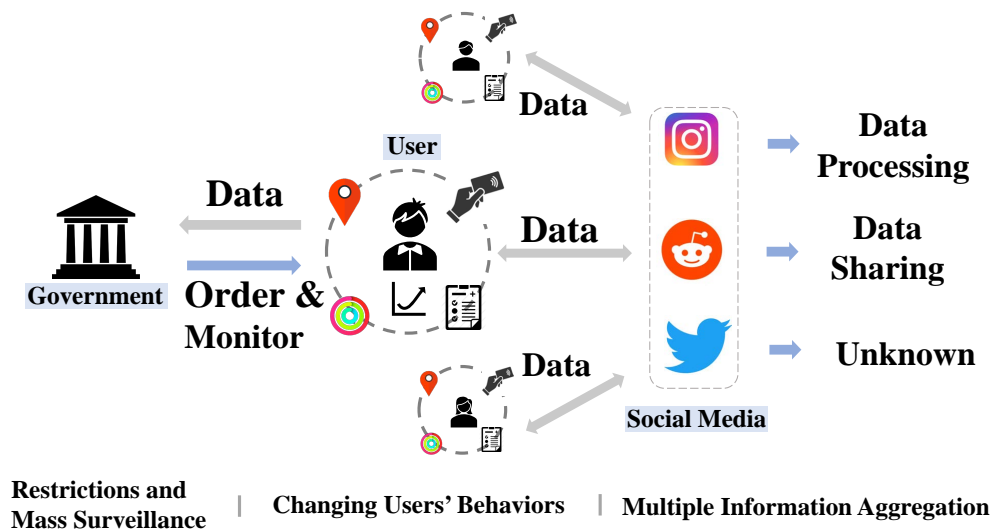
Figure 1: A Systematic Overview of Privacy Threats from Multiple Domains Related to the COVID-19 Pandemic

social media pushed by a highly divisive environment. This is likely linked to users disclosing sensitive information including but not limited to political stand, home address, and family relative information. The potential privacy harms to users in the context of political debates have been studied before (Rubinstein 2014). However, this election has introduced even additional situational factors, as it happened in the middle of a pandemic.

Information sources across multiple social media may cause serious user privacy issues and unclear self-disclosures under the chaotic interactions with natural and social environment. Advanced machine learning and data mining techniques investigate non-obvious relationships and search hidden data patterns, which can provide insights to the data owners and external parties for unknown analysis (Chamikara et al. 2020).

In the following, we first summarize and analyze emerging privacy threats triggered by or enhanced by the COVID-19 Pandemic. Based on our findings, we provide a high-level comprehensive analysis of privacy from multiple domains, propose related potential research directions,and conclude implications for future online public privacy in crisis.. Finally, we discuss possible solutions of proposed research questions.

Increasing Privacy Threats due to the COVID-19 Pandemic

Mass Surveillance

There is an ongoing public conversation about whether and under what circumstances the United States should embrace a surveillance program for COVID-19 (Ram and Gray 2020). Here, we focus on what tools the government and companies are leveraging from the phenomenon perspective. There is increasing surveillance over people's daily behaviors from the government and companies during the COVID-

19 pandemic in the name of monitoring and tracing the virus spread (Hussein et al. 2020). Many countries and companies are leveraging people's personal data (location, body temperature, facial information, etc.), which is collected by cell phones, traffic cameras, and other sensors, to track human mobility, identify individuals with risk, and monitor the disease spread (Singer and Sang-hun 2020). In the United Kingdom and India, smart city infrastructure has been re-used to monitor the people's social distance. In China, people can download a cell phone application that can tell whether they have been exposed to COVID-19 by analyzing the collected location data and local infection situation (BBC 2020). In the United States, Apple and Google provided a contact tracing application for their mobile users as well with bluetooth specification (Apple and Google 2020a) and cryptography specification (Apple and Google 2020b). However, as a key part of the extension of the surveillance state, researchers stated that the anonymized data is not always anonymous and location data can exacerbate inequality. (Frith and Saker 2020).

Data Usage across Multiple Platforms

During the COVID-19 pandemic, people spent extensive time online communicating, generating content, and engaging in other activities. With the development of data science techniques, people have more computational power and various channels to collect, process, and share data. There have already a lot of released open datasets focusing on different aspects related to the COVID-19 (Blose et al. 2020; Chen, Lerman, and Ferrara 2020; Pepe et al. 2020; Cohen et al. 2020; Cheng et al. 2020; Dong, Du, and Gardner 2020). Many social media platforms provide APIs for people to acquire data, such as Twitter [1] and Reddit [2]. Those APIs lower

---

[1]https://developer.twitter.com/en/docs/twitter-api
[2]https://www.reddit.com/dev/api/

the barrier to access social media data. However, we can not fully prevent malicious usage of the collected data.

At the same time, more digital records and accounts containing sensitive information are being created online, for example, online shopping accounts (Brough and Martin 2020) and other services that are brought online. Online users may not be fully aware of the fact their private information can be collected, shared, and used in an unexpected way (Malandrino et al. 2013). Many users may have more than one accounts on social media. How to measure privacy disclosure score based on the information across multiple social networks has been discussed (Aghasian et al. 2017) extensively. Zola et al. explored a cross-source cross-domain sentiment analysis with training data from Amazon and Tripadvisor and testing on the data from Facebook and Twitter (Zola et al. 2019).

Change of Individual Privacy Calculus

Another observed phenomenon and potential concern is the change of individuals' perception to self-disclosure and privacy. Individual-level behavior during the pandemic is a result of voluntary and government-enforced behavioral change (Farooq, Laato, and Islam 2020). From the individual perspective, people are calibrating their behavior between information acquisition and privacy loss. Users may have different attitudes and sensitivity to their privacy and self-disclosure during the pandemic (Fahey and Hino 2020). People would more easily sacrifice their private health status information to get suggestions, pre-diagnosis, or contribute to what the government appeals during the COVID-19 pandemic, especially in Asia (Cha 2020). Discussing personal health status, symptom, and test results on social media has become more common. Governments and companies provide convenient tools for people to update their personal information and implicitly convince people that the behaviors are a contribution to the public good (Nabity-Grover, Cheung, and Thatcher 2020). However, to my best knowledge, there are not enough official files to remind people about individual privacy issues or broadcast basic knowledge of data usage for people during the COVID pandemic.

A systematic overview of privacy issues from different aspects during the COVID-19 Pandemic is shown in Figure 1.

Post-pandemic Potential Privacy Risks

Over-collected Data Abuse

The COVID-19 pandemic has promoted the development of e-commerce, online education, social media platforms, smart phone applications, and related virtual service. Due to the health emergency, many countries relax the regulation restrictions or cooperate with companies to put the public security in the first place by collecting and analyzing data to support governmental prevention decision making. The governments could leverage contact tracing information to monitor and analyze citizens' behaviors, e.g. LGBT people identification in South Korea (Fahey and Hino 2020). Some countries will put pressure on their companies to release the collected data and provide data analysis on the involved

users. The European Commission has invited telecommunications companies to make their metadata available (Turner 2020).

Tech companies, including Instagram, Twitter, Facebook, and etc., can abuse this detailed data sets of individually, by selling, processing it to derive sensitive information, or sharing it inappropriately. Relying on powerful computational resources such as GPU clusters, a huge amount of data, and advanced data processing techniques, users behaviors can be described, modelled, and predicted accurately without any consideration for users' privacy. For example, an example of user behavior identification and prediction across multiple social media platforms is shown in Figure 2. Moreover, people share content via text, pictures, video, live streaming, and other formats, which can provide comprehensive information of users. Online interactions, e.g., "Follow", "Hashtag", "Mention", "Reply", can even reveal users' friends and relatives and create their social network structure. That would cause other related users' the privacy loss and over-disclosure and the propagation of the threat across the whole social media.
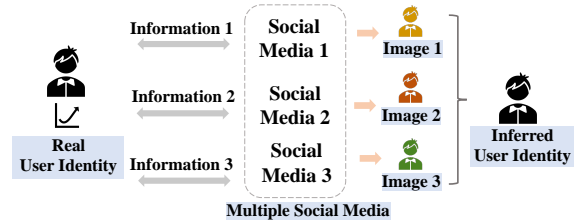

Figure 2: Users Potential Privacy Risks: User Identity Inference based on Multiple Social Media. For each social media, one user would self-disclose part of personal information, for example, Information 1, Information 2, and Information 3. According to the disclosed information, one user can be treated as fuzzy image with released and limited inferred information on one social media, for example, Image 1, Image 2, and Image 3. However, given multiple social media data of one user and advanced across-platform data processing techniques, data can be aggregated to infer a more accurate user identity with detailed personal information.

Public Privacy Concern and Social Trustworthiness

As the COVID-19 pandemic carries on, debates and laws surrounding surveillance capabilities are at the forefront of many minds (ROSS 2020). However, a majority of Americans said that they were concerned about how their personal data would be used by data collectors and they knew extremely little about the laws or regulations to protect their data privacy(Auxier 2020). Many governments gather or even over-collect people's data during the pandemic via different approaches. There is a great possibility that they will not delete the collected personal data or even continue collecting the data without informing users. Another survey result in (Auxier 2020) shows that 69% U.S. adults thought they should have the right to have the medical data permanently deleted after necessary and legal usage. While peo-

Table 1: Possible Research Directions and Questions about Privacy Issues and Self-disclosure related t Crisis On Social Media

Research Directions	Research Questions
Self-disclosure Interaction and Propagation	• How and to what extent users' self-disclosure behaviors can affect other related users on social media? • How the self-disclosure behaviors propagate on the social media? • To what extent the crisis would affect the user self-disclosure behaviors? • How to find the balance point between the privacy preserving and self-disclosure to get enough and appropriate information in crisis? • How to quantify self-disclosure across multiple social media and provide a varying evaluation considering situational factors?
Public Privacy Concern and Attitude Tracing	• How to trace the public privacy attitude change to their current status? • How to design an appropriate data-driven mechanism and regulation to gather appropriate data and decrease the public privacy concern? • How to model the complex and dynamic observations considering users' privacy concern, users' behaviors, and the pandemic crisis?
Mental Health in the COVID-19 Pandemic	• How to find a balance between keeping mental health and privacy during the pandemic? • How the mental health status, self-disclosure, and privacy concern affect each other? Certain self-disclosure can help users keep a good mental health, while it takes private concerns to users as well. • During the health emergency crisis, considering users with different physical health status, would there be any differences of their mental health and online behaviors?
Prevention, Prediction, and Protection	• How to design a comprehensive mechanism to prevent over self-disclosure and privacy-disclosure according to complicated scenarios in crisis? • How to predict public behavior and provide appropriate suggestions with limited access of data during the pandemic? • How to protect users' provided data, protect the stability on social media, and establish social trust?

ple enjoy the benefit of pandemic tracking and controlling via the data-driven approach, it also raises public concerns for their individual privacy. Kye and Hwang argued that the government actions do have a huge impact on social trust and government Trustworthiness. The temporal over-disclosed data and privacy data disclosure is gradually causing a stronger public privacy concern and challenging the government social trust.

Potential Research about Pandemic-related Privacy Issues on Social Media

Based on previous work and our discussion, we propose a set of related research directions (shown in Table 1) to understand and explore further privacy issues at time of COVID. They include: (i) self-disclosure interaction and propagation; (ii) public privacy concern and attitude tracing; (iii) mental health; (iv) prevention, prediction, and protection in the COVID pandemic. For each research direction, we provide several related specific research questions in the table 1 as well for future exploration.

Conclusion

The COVID-19 pandemic has generated a lot of practical problems and research questions related to privacy issues in online settings. In this paper, we describe how the COVID-19 affects user behaviors on social media. After that, we discuss three increasing privacy threats due to the pandemic including mass surveillance, data usage across multiple platforms, and change of people's privacy calculus. Furthermore, we introduce possible privacy risk after the pandemic. Finally, we propose a set of related research topics for further study. There could be several possible research directions: (i) appropriate and adaptive approaches to quantify self-disclosure and privacy combining peoples' comprehensive behaviors in multiple scenarios; (ii) mathematical and statistical models of privacy and human behaviors rather that can complement data-driven approaches ; (iii) study the interactions between people's awareness and sensitivity of privacy and self-disclosure considering the changes of environment. Different people may have different initial attitudes towards their personal information and decide how much information they feel comfortable to self-disclose. The exploration of the hidden relation between privacy attitudes, self-disclosure behaviors, and the reaction got from the environment can help us understand humans' privacy-related behaviors better and provide comprehensive suggestions for privacy-preserving mechanism design.

References

Aghasian, E.; Garg, S.; Gao, L.; Yu, S.; and Montgomery, J. 2017. Scoring users' privacy disclosure across multiple online social networks. *IEEE access* 5: 13118–13130.

Apple; and Google. 2020a. Contact Tracing Bluetooth Specification. URL https://www.blog.google/documents/58/Contact_Tracing_-_Bluetooth_Specification_v1.1_RYGZbKW.pdf.

Apple; and Google. 2020b. Contact Tracing Cryptography Specification. URL https://www.blog.google/documents/56/Contact_Tracing_-_Cryptography_Specification.pdf.

Auxier, B. 2020. How Americans see digital privacy issues amid the COVID-19 outbreak. URL https://www.pewresearch.org/fact-tank/2020/05/04/how-americans-see-digital-privacy-issues-amid-the-covid-19-outbreak/.

BBC. 2020. China launches coronavirus 'close contact detector' app. URL https://www.bbc.com/news/technology-51439401.

Blose, T.; Umar, P.; Squicciarini, A.; and Rajtmajer, S. 2020. Privacy in Crisis: A study of self-disclosure during the Coronavirus pandemic. *arXiv preprint arXiv:2004.09717* .

Brough, A. R.; and Martin, K. D. 2020. Consumer Privacy During (and After) the COVID-19 Pandemic. *Journal of Public Policy & Marketing* 0743915620929999.

Cha, V. 2020. Asia's COVID-19 Lessons for the West: Public Goods, Privacy, and Social Tagging. *The Washington Quarterly* 1–18.

Chamikara, M. A. P.; Bertók, P.; Liu, D.; Camtepe, S.; and Khalil, I. 2020. Efficient privacy preservation of big data for accurate data mining. *Information Sciences* 527: 420–443.

Chen, E.; Lerman, K.; and Ferrara, E. 2020. Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372* .

Cheng, C.; Barceló, J.; Hartnett, A. S.; Kubinec, R.; and Messerschmidt, L. 2020. Covid-19 government response event dataset (coronanet v. 1.0). *Nature human behaviour* 4(7): 756–768.

Cohen, J. P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T. Q.; and Ghassemi, M. 2020. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988* .

Dong, E.; Du, H.; and Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 20(5): 533–534.

Fahey, R. A.; and Hino, A. 2020. COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management* 55: 102181.

Farooq, A.; Laato, S.; and Islam, A. N. 2020. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. *Journal of medical Internet research* 22(5): e19128.

Frith, J.; and Saker, M. 2020. It Is All About Location: Smartphones and Tracking the Spread of COVID-19. *Social Media+ Society* 6(3): 2056305120948257.

Gao, J.; Zheng, P.; Jia, Y.; Chen, H.; Mao, Y.; Chen, S.; Wang, Y.; Fu, H.; and Dai, J. 2020. Mental health problems and social media exposure during COVID-19 outbreak. *Plos one* 15(4): e0231924.

Holmes, R. 2020. Is COVID-19 Social Media's Levelling Up Moment? URL https://www.forbes.com/sites/ryanholmes/2020/04/24/is-covid-19-social-medias-levelling-up-moment/?sh=5fa080dd6c60.

Hussein, M. R.; Shams, A. B.; Apu, E. H.; Mamun, K. A. A.; and Rahman, M. S. 2020. Digital Surveillance Systems for Tracing COVID-19: Privacy and Security Challenges with Recommendations. *arXiv preprint arXiv:2007.13182* .

Kanter. 2020. COVID-19 Barometer: Consumer attitudes, media habits and expectations. URL https://www.kantar.com/Inspiration/Coronavirus/COVID-19-Barometer-Consumer-attitudes-media-habits-and-expectations.

Kordzadeh, N.; and Warren, J. 2017. Communicating personal health information in virtual health communities: an integration of privacy calculus model and affective commitment. *Journal of the Association for Information Systems* 18(1): 1.

Kye, B.; and Hwang, S.-J. 2020. Social trust in the midst of pandemic crisis: Implications from COVID-19 of South Korea. *Research in social stratification and mobility* 68: 100523.

Malandrino, D.; Petta, A.; Scarano, V.; Serra, L.; Spinelli, R.; and Krishnamurthy, B. 2013. Privacy awareness about information leakage: Who knows what about me? In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*, 279–284.

Nabity-Grover, T.; Cheung, C. M.; and Thatcher, J. B. 2020. Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. *International Journal of Information Management* 55: 102188.

Pepe, E.; Bajardi, P.; Gauvin, L.; Privitera, F.; Lake, B.; Cattuto, C.; and Tizzoni, M. 2020. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data* 7(1): 1–7.

Perez, S. 2020. TikTok Engagement Among Kids Surges During the Pandemic. URL https://www.marketingcharts.com/demographics-and-audiences/teens-and-younger-113749.

Ram, N.; and Gray, D. 2020. Mass surveillance in the age of COVID-19. *Journal of Law and the Biosciences* 7(1): lsaa023.

ROSS, C. R. 2020. Will we give up privacy for security after Covid-19? URL https://www.statnews.com/2020/04/08/coronavirus-will-we-give-up-privacy-for-security/.

Rubinstein, I. S. 2014. Voter privacy in the age of big data. *Wis. L. Rev.* 861.

Singer, N.; and Sang-hun, C. 2020. As Coronavirus Surveillance Escalates, Personal Privacy Plummets. URL https://www.nytimes.com/2020/03/23/technology/coronavirus-surveillance-tracking-privacy.html.

Turner, J. 2020. Privacy vs. Security in the Post-Pandemic World. URL https://www.natlawreview.com/article/privacy-vs-security-post-pandemic-world.

Zola, P.; Cortez, P.; Ragno, C.; and Brentari, E. 2019. Social media cross-source and cross-domain sentiment classification .