Eight months into the COVID-19 Pandemic- Do Users Expect less Privacy?

TANUSREE SHARMA*

University of Illinois at Urbana-Champaign

HUNTER A. DYER

University of Illinois at Urbana-Champaign

MASOODA BASHIR

University of Illinois at Urbana-Champaign

Abstract-- While mobile apps provide immense benefits for the containment of the spread of COVID-19, privacy and security of these digital tracing apps are at the center of public debate. To understand users concerns and preferences for using a COVID-19 app, we conducted 2 surveys 3 months apart to determine what kind of privacy protections users were seeking in these apps and if those expectations change over time. Our survey results from participants N1: 2294 and N2: 2140 indicate that, trust plays a vital role in their adoption decision of such apps. Additionally, users' preference for certain privacy protection and transparency revealed a disconnect between what technologists' and users' expectation. In this paper, we present our survey results to formalize design guidelines for app designers, providers and relevant authority. We recommend three important mechanisms of trust, preferences, and transparency that could greatly influence users' adoption of such apps and can provide critical design components that can satisfy users' expected privacy protections.

CCS CONCEPTS • Security and Privacy • Human and Societal aspects of Security and Privacy • Privacy Protection

Additional Keywords and Phrases: Mobile Apps, Privacy & Security, Human-Centered, COVID-19

1 INTRODUCTION

A number of mobile phone application have been built and deployed to combat COVID-19, offering various services to the users including virus information, contact tracing and symptom monitoring. From January 30th , 2020 when the World Health organization (WHO) declared Coronavirus as a public health emergency to date, there are many COVID-19 apps that has been developed within a very short amount of time. In contrast to some countries, in United States, COVID-19 cases continue to rise and the number of deaths due to the pandemic is at 220740 as of this writing and active coronavirus case are at 8084411 [1]. Among many other factors, it is believed that one main factor contributing to the lack of containment of the virus in the US might be due to people not accepting to adopt contact tracing apps [2]. While there could be numerous reasons influencing this desist such as, people's political leanings, economic disparity [18] etc., one consistent reason that is often raised in this debate is around individual privacy and how personal information collected by contact tracing apps may violate it [4]. From the onset of this pandemic, there have been many studies that have specifically focused on privacy and security aspects of these apps and have recommended privacy and security components for contact

tracing apps [2,3,4]. In addition, there has been studies examining public opinion and concerns for COVID-19 apps [5]. However, to the best of our knowledge, majority of these user studies does not provide design strategies or specific user preferences regarding privacy protections that can be implemented in COVID-19 apps in order to increase adoption and alleviate privacy concerns.

To address this gap, we conducted a study at two different times to investigate users' attitudes and preferences towards COVID-19 apps' and their privacy concerns. We have formalized our survey in two different timelines (3 months apart) to determine if people's attitudes and preferences towards privacy may have changed from the onset of the pandemic to six months into the pandemic. We believe that by conducting the study at two different time series, we can capture a more accurate and consistent sense of users' privacy concerns that may have been more/less heightened at the beginning of the pandemic. Hence, it is our expectation that our user attitudes are more stable and can provide critical and timely design recommendations that can be easily implemented and meet users' expectation of trust, preferences and transparency when it comes to COvID19 apps.

Our online survey included N1= 2294 and N2= 2140 participants (N1: population for survey 1 conducted in April-May and N2: population for survey 2 conducted in July). The survey contained 200 questions that covered all aspects of COVID19 pandemic including social, psychological, COVID-19 experiences, COVID-19 apps, and General Technology. For this scope of this study, we only considered users' privacy concern and preferences related to using COVID-19 apps as well as their general privacy and security attitudes. In our analysis, we have analyzed in two different ways: 1) descriptive and distributive statistical analysis, 2) supervised learning algorithm for clustering open ended questions. Our entire study will answer the following questions:

RQ: Does trust, transparency and preferences influence COVID-19 app adoption?

RQ1: What aspects of the app influence users' trust in adopting COVID-19 app?

RQ2: What kind of protections do users want when providing personal information in COVID-19 apps?

RQ3: What kind of data access do users consider privacy invasive practices in COVID-19 apps?

We incorporated privacy principles and design specifications to formalize our user study findings into actionable control design. Users' opinion was used as decision metrics in design implementation.

2 BACKGROUND

This section briefly presents a review of recently identified design shortcomings and users' privacy risk related scenarios to advocate the need for user-based privacy model in COVID-19 apps.

2.1 Recent COVID-19 apps' shortcomings

For responding to COVID-19 pandemic, there has been a large number of apps for different purposes, including contact tracing and information/ health assessment. Since many of these apps are developed in a short time around the world under emergency situations, several of them lack sufficient privacy protections [20]. Recent research study analyzed Singapore's OpenTrace app and its' use of Google firebase services to manage users' data and deployment of reversible encryption and found that such method can be vulnerable to secret key disclosure [7], [19]. Subsequently, some research suggests that decentralized data management approach can be helpful in this situation. National Vulnerability database's, Common Vulnerabilities and exposure (CVE) from NIST showed that COVIDSafe for iOS allows a remote attacker to crash the

apps and interfere with COVID-19 contact tracing via Bluetooth advertisement containing manufacturer data which presumably caused by erroneous OpenTrace manuData.subdata call [8]. Some of the other COVID-19 tracing apps holds these same vulnerabilities.

2.2 Using Privacy Design Principles in Developing Human-Centered Technologies

Users' security and privacy concerns are widely considered as one of the most important decision-making criteria for successful implementation of modern technologies [9]. To better deploy mitigation strategies, some researchers present concepts of a privacy-preserving Trusted Computing environment [10], smart home technology implementation [11, 12], and so on. Some earlier studies have specifically examined users' perceptions, beliefs, and attitudes regarding Android apps [13]. Other studies have utilized crowdsourcing as a privacy evaluation technique to better understand mental models of users [14]. Our work is informed by all these human-centered approaches for privacy and security to design our survey. Again, there exists different regulatory approaches around the world to preserve users' privacy and security. For example, in the United States, the Privacy Act of 1974 establishes "a code of fair information practices" in order to govern how individual's information is collected, used, and disseminated by federal agencies [15]. Furthermore, National Institute of Standards and Technology (NIST) developed a Privacy Framework, which aims to improve privacy through optimized use of personal data [16]. Outside the US, different levels of international privacy principles and standards are implemented in different regions. Motivated by these frameworks, we decided to form our findings into privacy and security design components. We believe, these can be an actional design approach for app providers and developers in their development process and functional requirement assessment.

3 DATA COLLECTION AND METHODOLOGY

In this section, we discuss our research model for data collection, sampling and assessment strategy.

3.1 User Survey

To assess users' attitudes toward using digital tracing apps and the factors that might contribute to either positive or negative attitudes towards these technologies, we launch user studies in two different time series. The online survey is conducted by Center for Social and Behavioral Science (CSBS) at a midwestern University. The first survey was conducted in between April-May 2020 at the beginning of the Pandemic and then the same survey was conducted again in July 2020. For the explaining purpose, we will use T1 for the first survey and T2 for second survey. Participants recruited for this survey were from the United States and they were all adults above the age of 18. The survey link was distributed through REDcap survey platform and the survey data was collected by Dynata, LLC. It should be noted that REDCap is HIPAA compliant and the survey proposal was approved by the Institutional Review Board (IRB). The survey includes 200 questions, however, we only considered COVID-19 apps and Tech privacy questions for our analysis in this paper. The online survey solicited different types of responses in the form of multiple choices, likert scales, Yes/No and open-ended questions. We had a total of 2294 responses from first survey and 2140 responses from second survey.

3.2 Data Analysis Approach

We performed descriptive and distributive statistics as well as linear model analysis on the context of users' privacy concerns, preferences and trust in using COVID-19 apps. Further, we captured the high-level themes from participants' open-ended explanations by performing a thematic content analysis. From the initial findings, we further decided to utilize topic modeling algorithms to cluster the overarching theme. Specifically, we used latent Dirichlet allocation (LDA) to cluster users' responses. While there are many other ML methods and approaches could analyze this data, we believe this particular method is a best-fitted strategy when dealing with such an exploratory and qualitative data analysis. Furthermore, our open-ended question includes a mixture of topics from users' direct feedback within a given question belonging to several topic and so, each question can be represented as a vector of proportions that denote what fraction of the words belong to each topic [17].

3.3 Sample Demographics

Table 1 includes sample of demographic that can give an idea about our participants in both of the survey.

Table 1: Demographic of Survey Data

Survey Data 1 (N1: 2294)			Survey Data 2 (N2: 2140)		
Data	Types	Percentage	Data	Types	Percentage
Age	18-30	26.9%	**Age**	18-30	43.3%
	31-40	19.9%		31-40	17.9%
	41-50	15.7%		41-50	13.1%
	51-60	16.2%		51-60	13.1%
	61+	21.3%		61+	12.5%
Gender	Female	47.9%	**Gender**	Female	66.6%
	Male	51.5%		Male	32.6%
	Nonbinary	0.5%		Nonbinary	0.9%
Education	High School	3.6%	**Education**	High School	5.6%
	Diploma/GED	16.5%		Diploma/GED	24.8%
	Vocational	5.5%		Vocational	4.5%
	College	23.7%		College	28%
	College-Graduate	30.6%		College-Graduate	25%
	Graduate Degree	19.3%		Graduate Degree	12.1%
Political	1: Extremely liberal 5: Moderate 9: Extremely conservative	Mean: 3.524	**Political**	1: Extremely liberal 5: Moderate 9: Extremely conservative	Mean: 4.613

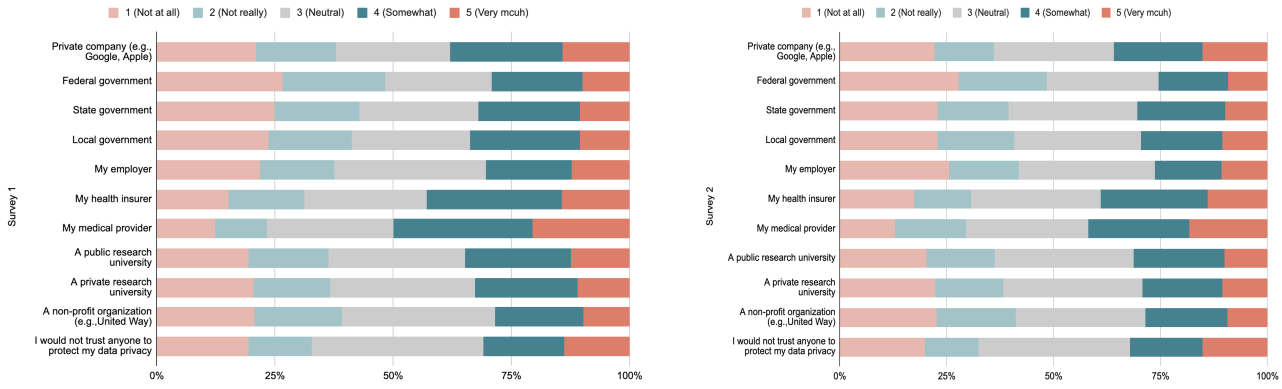
4 FINDINGS

In the following subsections, we discuss findings from our two survey results on how users trust, their preferences and concerns increase/decrease their adoption of COVID-19 apps. Within that, in every section, we will be offer recommendation and actionable suggestions for App design which comprise users' perception into account.

4.1 User's Trust

One question that assessed if users' trust in technology providers increase/decrease their motivation in using COVID-19 apps. Our survey results showed that users are more likely to use a COVID-19 app if the apps were offered by entities they trust. The majority of study participants explicitly indicate that factors such as who are offering the apps and who had access to their personal information were important app adoption criteria. While there is a change in study results related to these factors from T1 to T2, there is still a clear influence. For example, 68.2 % of the participants responded Yes to the question "would it matter to you who offered the app and controlled your data" in T1 and this percentage went down to 61.6% in T2.

Furthermore, we have found that participants trust certain entities more than other in protecting their privacy. In case of trusting a certain entity in providing privacy protection, participants showed trust towards medical provider, health insurer and public research universities in protecting their privacy which is consistent in this two-time period from T1 to T2. However, in T1, survey 1 showed that participants don't trust Federal govt, state and local govt at all and in T2, their distrust towards state government decreased overtime. We also derived relationship between users' trust in particular entities in providing COVID-19 apps which is related to general security concerns and have significant relationship with p value less than 0.05.



(a) (b)

Figure1: If such an app were offered, how much would you trust the following groups to protect your privacy?

Design Recommendations: Our results showed that Participants are more likely to use apps if those are provided by certain entities than others. In our results, the trusted entities are medical provider, health insurer and public research universities. It should also be noted that users in the US consistently trust the Government the least. But, in times of medical emergencies, they are typically the entity that would develop or publish such apps and thus it is important to consider other approaches that can address this predicament. For example, perhaps the apps that are developed by the government can be validated or approved by a trusted party that users trust.

4.2 User's Privacy Protection and Transparency Preferences

The majority of COVID-19 apps ask for different permissions during installation and run time for photos and media, Bluetooth, contact information, internal and external storage, network state, etc. To understand user's preferences for data protection and what they considered to be personal and sensitive. one of the questions in our survey asked participants "Mobile apps often collect information for them to function. Please indicate if do not want COVID-19 app collecting any of the following information… ". Users' rated photos, browsing history, contacts, and email address as their top sensitive and personal information followed by username, machine address, geographical location, device operating system and screen size. These results were consistent in both timelines of the survey.

Our survey results reveal an interesting finding and dilemma. Many of the recent literature on privacy preserving contact tracing focuses on privacy protections for location data by providing protocols that can either make user's exact location confidential [3] or utilizing Bluetooth token exchange in certain distances to avoid using GPS location [4]. For example, some researchers proposed construction of Bluetooth tokens, utilizing those tokens in a privacy preserving manner [2]. However, our survey results regarding location data collection from both T1 and T2 indicate that users do not consider this to be sensitive or personal. Thus, it is clear that there remains an information asymmetry between users and technologist when it comes to privacy protections that is sought or considered important. This disparity in users' priority is also concerning because it does not follow what the COVID-19 literature considers to be invasive when our study participants rated their geographical data much lower than other data types.

Further, we found that one of the main thing users would like to know is how their information is collected and why, how information will be shared, how data will be protected and be assured that data will only be used for the pandemic. From their response presented in Figure 3, we can clearly see the three main data practices they were seeking: transparent data collection, data sharing, data processing for only primary purpose (use of data limitation). Their responses are consistent both in T1 and T2. Also, from another survey question on transparency showed that users would more likely to use COVID-19 apps if they are guaranteed that they could completely delete all the data from the app any time. Their response towards this high towards "yes" in both T1 (63.3%) and T2 (56.4%) timeline.

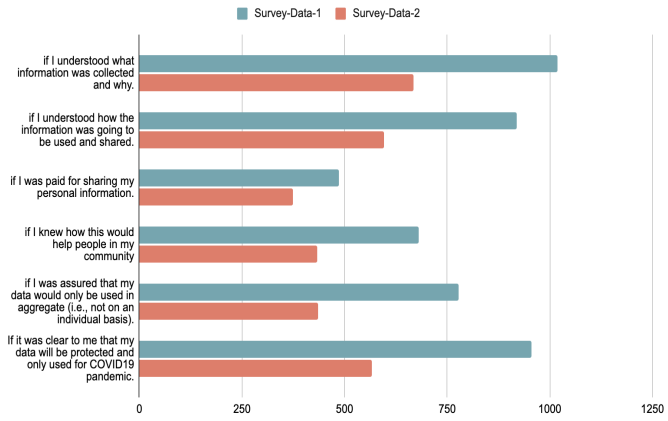


Figure 2: Providing personal information in certain condition

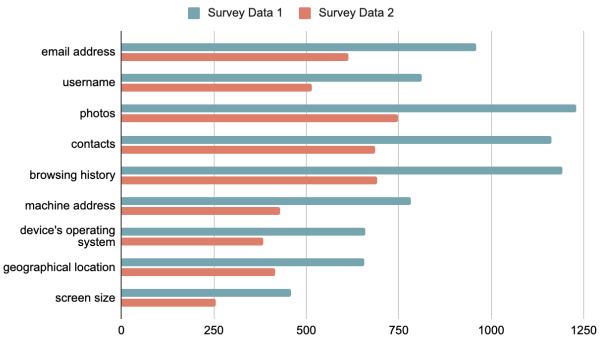


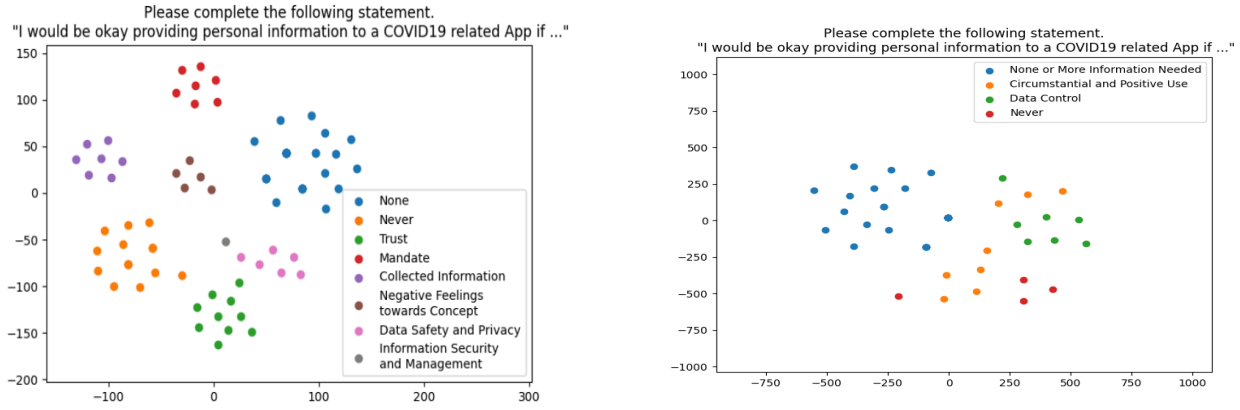
Figure 3: Preferences of privacy for data types

Suggestion for Design Components: This is a critical factor in any app design since the success of any app depends on its users, and the particular features they seek in a given app. When developing COVID-19 apps, we recommend that developers also consider cultural and regional differences in their design. Based on our survey results from a midwestern US state it is clear that users have a strong preference for privacy and security protections if they were to use such an app. Mainly, they prefer to have transparent data life cycle within their apps. Developers might consider transforming these principles into activities for data inventory mapping. Additionally, there needs to be a proper consideration in app development process to include users' prioritized data types to deploy security components within. Which means, development process needs to understand the level of severity from users' point of view.

4.3 User's Privacy Concerns

From our survey, we have seen that, user's security and privacy concern slightly changed from the beginning of the pandemic to T2 where 74.9% participants were concerned about security vulnerabilities in COVID-19 apps in survey 1 and 71.4% are concerned during survey 2. Also, from the survey, it showed that, their wiliness to use tracing apps slightly changes from T1 (39.8%) to T2 (36.75%). From the beginning of the pandemic to T2 time period, respondents were concerned consistently about using mobile phone apps.

For better understanding user's privacy concerns, we have studied open-ended question for understanding the overarching theme of users' privacy in using COVID-19 apps. We have modeled topics within the open-ended questions LDA in order to assess the overarching themes within each question's response. Topics were extracted from open-ended questions using a common topic-modelling process, latent Dirichlet allocation (LDA). Answers were preprocessed using the NLTK Python module, utilizing nouns, adjectives, verbs, and adverbs. Answers were also filtered with lemmatization and the standard stop word list in the module. The number of topics chosen for each application of LDA was iteratively tested through 15, and the final number of topics were chosen based on a mix of conciseness (avoiding repetition of topics) and the UMass coherence score. We then utilized the topic weights from LDA to produce a t-SNE embedding (t-distributed stochastic neighbor embedding) which allows for a high-level view of the response weight structure and magnitudes for a given question.



(a) (b)

Figure 4: Condition on which users will be okay to provide personal information via COVID-19 apps

In Figure 4**,** we presented one of the results from our topic modeling where they were asked how they will be comfortable in providing information via COVID-19 apps. Our results indicate that most participants would be okay sharing their personal data via COVID-9 apps if privacy protections were provided for the information that is being collected and security measures were put in place for the information storage. The other insight from our topic modeling is that users would never provide personal information or would only do so if there wasn't a mandate. The topics that covered negative feelings towards this particular scenario were populated by statements that had negative sentiment towards tracking apps and COVID-19 itself.

In participants' own words:

"Data can be very easily manipulated and tracked and having such personal data there for people to see does not sit right with me. There is no guarantee everything will completely delete forever even if that were offered an option."

"Don't like this idea of the government knowing my business. how would it even work? will the app release the person's name? I am concerned that data collected could be used for other reasons."

"I am just worried about the app and if it will tract my internet footprint. Because I don't want to use an app that tracks my online movement."

"I don't really like any opportunities for the government or companies to get my info because that's mine. I'd be less suspect before the whole NSA thing came out."

These quotes from respondents represents much of distrust towards government and concerns that their data will be used for other purposes than pandemic and their movement will be tracked. There are responses that touched on concerns about privacy and security of the information collected as well as what they prefer to have better data policy in those apps in terms of using those. In participants' word:

"I need my privacy protected. I would need to need to know about privacy and usage policy."

"If I have assurance that the information cannot now or ever be used to deny health insurance, employee benefits, medical services or hospitalization. That I have full authority to deny or grant permission for use of my personally identifiable data"

5 DISCUSSION

While some countries have been successful in using apps to tackle the pandemic, the use of such apps in the US and some regions of the world remains unclear and there is much public debate about not only the efficacy of such apps but also the protection of individual privacy rights. To provide some insight into this national debate, we conducted two time series survey in the US to learn more about people's attitudes towards such innovative technologies, their privacy/security concerns, and their preferences for privacy protections.

Our results indicate that most of our survey participants regard privacy/security protections as one of the main features they seek in these types of apps. They prefer transparent data practices to be implemented around their data collection, sharing, processing and usage in COVID-19 apps. Interestingly, our study participants prioritize their privacy protections for personal data in the following order with the highest protection offered to photo library, browsing history followed by contacts, email, username, machine address, geographical location, and so on. This is a surprising and yet illuminating discovery because this data type is one of the main data components that is needed for the functionality of contact tracing and yet most our participants did not rate this as a top privacy vulnerability. Perhaps some of this disconnect maybe attributed to participants' lack of knowledge about location data and the privacy vulnerabilities that it may introduce or perhaps users prefer or may trade off other data types in order to share their location. We assessed participant's trust in using COVID-19 apps and our results reveal that when it comes to tracing app, most participants would put their trust in medical providers while they trust least the federal government. More in-depth analysis of the open-ended question show a similar pattern where participants would trust private insurance and health care provider the most for developing such apps.

In conclusion, our research findings suggest that in order to design and develop a privacy inclusive COVID-19 app, technologists must plan for the appropriate data protection mechanisms that considers three important factors related to its users namely trust, transparency, and their preferences. Especially if we are aiming to increase timely adaptation and our lives and well-being may depend on it.

6 LIMITATION

This survey was conducted in the United States and therefore the results may not be applicable to other parts or people of the world. Further research studies that includes different populations and regions of the world needs to be conducted to determine if these results apply to a global context. In addition, factors such as social desirability may have influenced our results because participants may have been primed to value privacy protections based on the questions. Furthermore, our survey was online, hence, we only reached participants who are familiar with using technologies. Therefore, their response on privacy/security concerns involving COVID-19 technologies can be biased in some ways.

Considering above limitations, we plan to conduct more in-depth analysis to implement the three user-based mechanisms in formal engineering method with a focus of technological aspects. Furthermore, we will further consider factors such as political ideology and level of education and many other factors that may have influenced our findings.

REFERENCES

[1] Worldometer. https://www.worldometers.info/coronavirus/country/us/. Accessed on Oct 13th, 2020.

[2] Cho, H., Ippolito, D., & Yu, Y. W. (2020). Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511.

[3] Troncoso, C., Payer, M., Hubaux, J. P., Salathé, M., Larus, J., Bugnion, E., ... & Barman, L. (2020). Decentralized privacy-preserving proximity tracing. arXiv preprint arXiv:2005.12273.

[4] Reduce the spread of covid-19 without increasing the spread of surveillance."

[5] accessed: 23-06-2020. [Online]. Available: https://covid-watch.org/

[6] Simko, L., Calo, R., Roesner, F., & Kohno, T. (2020). COVID-19 Contact Tracing and Privacy: Studying Opinion and Preferences. arXiv preprint arXiv:2005.06056.

[7] Leith, D., & Farrell, S. (2020). Coronavirus Contact Tracing App Privacy: What Data Is Shared By The Singapore OpenTrace App?

[8] CVE-2020-12717, National Vulnerability Database. https://cve.mitre.org/cgibin/cvename.cgi?name=CVE-2020-12717

[9] Sharma, T., & Bashir, M. (2020, July). Privacy apps for smartphones: An assessment of users' preferences and limitations. In *International Conference on Human-Computer Interaction* (pp. 533-546). Springer, Cham.

[10] Qiu, M., Gai, K., Thuraisingham, B., Tao, L., Zhao, H. (2018). Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. Future Generation Computer Systems, 80, 421-429.

[11] Haney, J. M., Furman, S. M., Acar, Y. (2020, July). Smart Home Security and Privacy Mitigations: Consumer Perceptions, Practices, and Challenges. In International Conferenceon Human-Computer Interaction (pp. 393-411). Springer, Cham.

[12] Winkler, Thomash, & Rinner, Bernhard. 2010. A systematic approach towards user-centric privacy and security for smart camera networks. In Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC '10). Association for Computing Machinery, New York, NY, USA, 133–141.

[13] Felt, A. P., Chin, E., Hanna, S., Song, D., Wagner, D. (2011, October). Android permissions demystified. In Proceedings of the 18th ACM conference on Computer and communications security (pp. 627-638).

[14] Lin, J., Liu, B., Sadeh, N., Hong, J. I. (2014). Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In 10th Symposium On Usable Privacy and Security (SOUPS 2014) (pp. 199-212).

[15] The Fair Information Practice Principles (FIPPs). https://www.nationalpublicsafetypartnership.org/

[16] Lefkovitz, N., Boeckl, K. (2020). NIST Privacy Framework: An Overview (No. ITL Bulletin June 2020). National Institute of Standards and Technology.

[17] Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

[18] Mbunge, E. (2020). Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(6), 1631-1636.

[19] Sharma, T., & Bashir, M. (2020). Use of apps in the COVID-19 response and the loss of privacy protection. *Nature Medicine*, 1-2.

[20] T. Sharma, J. C. Bambenek, and M. Bashir. (2020). Preserving Privacy in Cyber-Physical-Social Systems: An Anonymity and Access Control Approach. [Online]. Available: https://www.ideals.illinois.edu/handle/2142/106049