Thesis Portfolio

Health Modeling Using Smart Device Data
(Technical Report)

Physiological Data Privacy in the Digital Age
(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Aldrick Johan
Spring, 2021

Department of Computer Science

Physiological Data Privacy in the Digital Age

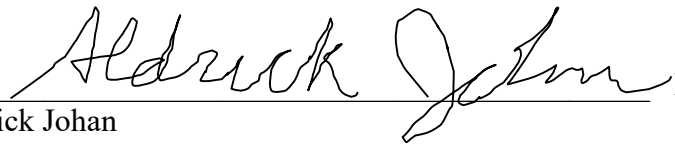
A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Aldrick Johan
Spring, 2021

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature  Date: 05/13/2021
Aldrick Johan

Approved _____ Date _____
Sharon Tsai-Hsuan Ku, Department of Engineering and Society

Physiological Data Privacy in the Digital Age

Introduction

The COVID-19 virus has surged throughout the world, forcing governments to employ various methods to slow its spread. One of these methods is using contact tracing apps to track which people have been in contact with each other. To achieve this, data is required from people's smart devices. The collection of this data has many social ramifications. It normalizes the monitoring of the general public, while also desensitizing them from losing their privacy. The collection of the physiological data would serve to contribute to this and could lead to a culture that believes it is normal to be monitored.

While this idea is alarming from the perspective of American society, other societies view this data collection as a necessity. As the world gets further into this digital age, the concept of privacy is evolving. Almost every person owns smart devices that collect information on themselves. Some societies believe this is a positive change because the data can be used for the public good, especially during a health crisis such as the COVID-19 pandemic. It can be used to detect the illness in individuals and can identify at-risk individuals. As the number of cases across the world increases, this could be one of the many possible solutions to slowing these cases down.

Research Question

The question of data privacy in regards to physiological data collection does not have a simple answer. Many aspects of the problem will have to be explored. To what extent should this data be collected? Who should have access to this data? What exactly should the data be used for? What defines privacy, especially during this digital age? I will study these questions by

exploring what data is already being collected and how that data is being used. I will do research on the prevalent attitudes towards privacy and data collection from the perspective of different cultures. This will allow me to explore how the concept of privacy is constructed and how it is changing.

Literature Review

I examined prior works of literature to provide context regarding the topic of physiological data collection. I began by reading articles that discussed how the COVID-19 pandemic impacted physiological data privacy. Then I explored the different viewpoints on the topic and why they exist. Finally, I outlined my data collection plan.

One of the major topics in the literature was how data collection should be handled during a global health crisis. The most prevalent school of thought on this problem is to accept the tradeoff between personal privacy and information, in the name of public health. This ideology posits that the cost of not collecting this data is far greater than the loss of some personal privacy (Cho et al., 2020). They argue that although some privacy is lost, the amount of personal information that is collected is minimal and is only revealed to authorized personnel. This ideology acknowledges that the current methods are not perfect, but they will improve as the pandemic progresses and that the information gained will be very valuable in terms of public health. The flow of the data would be as secure as possible using modern data transfer techniques and would anonymize as many parties as possible (Beskorovajnov et al., 2020). Using these methodologies would make it harder for harmful entities to attack them and serves to minimize the risk of potential attacks. These methods would also ensure that not much personal information would be revealed even if an attack was successful (Liu & Sun, 2016). The data will be collected from users' devices then would be sent to secure databases using the described

technologies. Once it reaches these databases, it can be processed and used to trace the spread of the virus.

Proponents of this ideology also argue that collecting this data would not cause significant changes to personal privacy, because it has already been impacted by the rise of big data (Perera et al, 2015). The collection of large amounts of data on the public, also known as big data, has been occurring since Internet of Things (IoT) devices have become more commonplace. These devices include cellphones, smart watches, medical sensors, and other devices that can connect to the internet. These devices collect metrics from their users and it is argued that true personal privacy is no longer possible due to them. When big data first became popularized many people took issue with it because of weak infrastructure and security (Terzi et al., 2015). However, many of these concerns have been addressed and infrastructures are now secure enough to prevent leaking of the data. Furthermore, the majority of the data will be useless to attackers as the data will be encrypted within the database.

Another recurring theme in the literature was the difference in global viewpoints when it came to data sharing and privacy. Privacy seemed to have a different definition depending on where people were from. Although many governments across the world already collect information on their citizens, the lack of global data sharing standards prevented mass cooperation amongst multiple governments (Allam & Jones, 2020). Many countries have “smart cities” where there are copious IoT networks. This allows the government to collect detailed information on their citizens. During a global crisis, it would be mutually beneficial for countries to share relevant data with each other. Unfortunately, there is no global consensus on how to achieve this. The World Health Organization (WHO) implemented some policies to solve this problem, however there are many doubts whether the data shared with WHO is accurate.

Another aspect that is holding back global participation with these programs is the difference in viewpoints concerning personal privacy. In the United States, personal privacy is considered to be important to the people (Bellman et al, 2010). This is due to the individualistic culture that has developed within the country. People from the United States have always been more focused on their own personal rights and have been wary of the government. This culture has led to citizens wanting the most secure data regulation while also collecting as little as data as possible (Ballman et al, 2004). Since the United States is an influential nation, other countries have also followed suit with restrictive data regulations including the United Kingdom. These policies make it difficult for global collaboration to grow especially during a pandemic and prevent the implementation of systems that would help prevent future outbreaks.

An additional trend I noticed while reading the literature was that many people still bought and used smart devices, knowing that it would affect their personal privacy. Consumers know that these devices have sensors and can connect to the internet without their permission (Perez & Zeadally, 2018). The sensors allow the producers of these devices to collect personal information on the users. The internet functionality of the devices allows them to transmit user data to other parties. While this may be alarming to some, others said that they had "nothing to hide" and were not concerned with their data being collected (Udoh & Alkharashi, 2016). This mentality was surprising as the data could be used to track a person's location in real time and to get a detailed look into their lives. This carefree attitude towards data collection may stem from these people growing up with devices and perceiving the data collection as inevitable. Fortunately, there are new data collection methods that manage to preserve privacy while getting the necessary data (Kim et al, 2019). These methods can help assuage fears about loss of privacy and can lead to the normalization of data collection.

STS Framework

There are various factors that both affect and are affected by the technologies that are relevant to physiological data privacy. These factors, known as actors, have varying levels of influence on the other actors. This influence is known as agency in the Actor Network Theory (ANT) and any actor with agency is known as an actant. By using the ANT framework, the various actants and their agency can be analyzed to see how wearable smart devices reached their current state.

The major actors in this network are governments, public policies, wearable device producers, the wearable devices themselves, the collected data, and the users of the devices. In this network the governments are connected to public policies because they create the laws. The policies are then applied to the remaining actants. This provides the governments with agency over the public policies and the public policies with agency over the producers, the devices, the data, and the users. The producers are connected to the wearable devices and the data collected by them. The producers have significant agency over both of these actants as they design the devices and how the data will be collected and stored. The devices have agency over the data and the users of the device because they interact directly with them. The data has ties with both the users and the government. The data has some agency over governments because if this data leaked it could be potentially be used as intelligence against the government. It has agency over users because users want their data to be private and would be upset if this data was leaked. The users have agency over the producers of the devices and the government. The designers utilize user feedback to improve their product going forward. Users have agency over the government as they can influence what topics are addressed in new laws. Users can also vote for new leaders in many countries so the government officials want to keep those users' content. The actants in

this complex web interact with each other, indirectly causing many different outcomes that eventually lead to major changes within the network.

The ANT framework also includes a concept known as translation. This occurs when actants are displaced and transformed to fit into the network. A successful translation in this network occurs between the wearable devices and the users. Before the translation the actors are two independent parties: a wearable smart device and potential consumers in the market. The wearable smart device then recruits the potential consumers and makes them users of the device. When this occurs, the consumers become a data collector who impact the network. As a data collector they have agency over the collected data because they are enabling the collection of the data. This alters the overall dynamic of the system by pushing the focus from the means of collecting the data to how the data should be used. This will eventually bring in new actors and actants to the system who will also alter the system. This concept of translation can also be used to describe the transformation of the concept of privacy. Privacy, in this context, refers to the accessibility of one's data by other parties. As more smart devices enter the market, people become more accustomed to the daily use of such devices. This reality makes more personal information available to the public, forcing society to adapt to having relatively less personal privacy and forcing regulators to adjust their policy making. Based on the culture of a society, the actors in their respective networks will displace and transform the concept of privacy until a definition that is acceptable to the society is reached.

Method

For the data collection, a variety of research methods were utilized. These methods consisted of utilizing surveys, interviews, and document review. The surveys were used to

collect the general public's thoughts about privacy. This involved presenting participants with questions that pertained to the topic of personal privacy and forcing them to challenge their own biases by carefully wording the questions. The survey questions presented the participant with a statement about privacy and asked them to indicate their level of agreement with the statement. The statements either supported a traditional view of privacy or a modern view of privacy. A traditional view of privacy consists of holding one's privacy in high regard and believing that privacy should not be compromised in any scenario. On the other hand, a modern view of privacy still values privacy, but does not believe that privacy has to be absolute. The modern view of privacy supports the use of some personal information to improve people's lives. The participants' responses to these statements were used to ascertain which view of privacy they supported more. The survey was distributed through email, direct messaging, and online discussion boards. Because of this, the participants consisted of Americans coming from a diverse range of ages and locations.

The interviews were used to understand the reasoning behind the participant's stances. The owner of a digital marketing company was interviewed, along with two college students: one raised in the United States and the other raised in China. Although these three interviewees all shared a modern view of privacy, they had different reasons for having that stance. Each one of them had a different experience with personal privacy and their answers provided context to the survey answers. Document review was also used to collect data on the topic of privacy. The document review consisted of studying privacy legislation in the United States, China, and the European Union to understand the role of legislators in privacy.

Data Collection and Analysis

The data collected for this STS study, when paired with the previously discussed ANT Framework, provides valuable context to the topic of privacy in regards to smart devices. The data collection methods of surveying, interviewing, and document review were used to acquire information on the public's opinion on privacy and current legislation on privacy. The surveys and interviews revealed that people largely had one of two views on privacy: a traditional view of privacy or a modern view of privacy. The document review was able to complement the other information by providing legislators' perspective on privacy.

The survey responses were able to indicate whether an individual had a traditional view of privacy or if they had a modern view of privacy. The survey responses showed that participants who were at least 30 years old were more likely to have a traditional view of privacy and that participants younger than 30 years old were more likely to have a modern view of privacy. Both groups indicated that they strongly valued their privacy, but the younger participants showed a greater acceptance of the use of personal data to enhance their own experience or to help the general public. The ANT framework can be used to explore why this is the case. People from the older age group did not have many technologies as they were growing up. During that time, the concept of privacy was close to absolute. The only actors that could impact what personal information became public was the individual, their close friends and families, and the media in rare cases. When new technologies emerged as new actants and started making more personal information available to the public, these older people were alarmed by the shifting concept of privacy. Because of this, the older population used these new smart devices less than the younger population. This led the older population to having smaller online social network and having less interests that require the use of smart devices. Consequently, the

older generation has a lesser understanding of what information is collected by these devices and their view of privacy has not shifted greatly from a traditional view. On the other hand, the younger people have always had smart devices and are accustomed to the consequences of having them. These younger people have vast social networks via the internet and consequently have been molded to have a different view on privacy. They understand what information is being collected on them and the benefits of this data collection. Therefore, they have a tendency to have a modern view of privacy. If both groups' ideologies are represented as actants that influence another actor, the public's view on privacy, then it is clear the impact these two groups have on the larger network that defines privacy.

The interviews that were conducted only served to strengthen the idea that the younger population has a better understanding of what their data is being used for. Three people were interviewed for this study: the CEO of a digital marketing company, a college student raised in America, and a college student raised in China. The CEO of the marketing company stated that the majority of personal information that is collected is harmless. He claimed that his experience as a marketer has shown him that health and phone data is limited and do not pose a threat to the user. He said that companies do not intend to be malicious with their data handling and ultimately use it to tailor the user experience. The student raised in the United States agreed with this sentiment and believed that the use of personal data was not inherently harmful. Since absolute privacy is not practically possible, he prefers for any data use to be transparent and non-malicious. The student from China also agreed, but he believed that individuals should be able to opt-out of sharing their information. One may have expected the Chinese student to have a radically different view of privacy compared to the American student, but it seems most young people have similar thoughts on privacy.

The document review revealed the role of legislators in the network that defines privacy. Policy makers influence other actants in the network by defining bounds on what information can be collected. By placing limits on the collection of the data, they displace other actants in the network, forcing a shift in the concept of privacy. In the United States, there are federal laws the protect people's financial data, health data, and children's data. These policies forced data collectors to collect less information than what they collect without any restrictions. This caused a ripple effect across the network, forcing actants such as device manufacturers and the general public to adjust. Device manufacturers adapted to these restrictions by excluding the capabilities to collecting that information. The general public adjusted by shifting their view on privacy. The majority of society agreed with the sentiments expressed within the laws, and actually may have served as the impetus behind the ideas in the laws being brought up. The data collected in this study shows that this web of actants influencing each other leads to changes to the concept of privacy.

Conclusion

In conclusion, the STS study provided some interesting results in regards to the perception of privacy in the digital age. The study found that the overall view of privacy is evolving, especially to the younger demographic. Due to the growing role of technology in the daily lives of many, the participants felt that privacy has been forced to change. As expected, different groups of people had different opinions on the extent of this change. People from a younger demographic were more likely to support more liberal use of collected data, while those from an older demographic were reluctant to accept some of the more intrusive uses of the data. This result is likely due to the older generation lacking information on how the data may be used

and what data is actually collected. The younger generation, having grown up with technology, is more educated on the topic. Surprisingly, the origin of the participant did not have a substantial effect on their views on privacy. It was expected that participants raised in societies with differing social structures would have jarring differences in their views on privacy. However, this factor did not seem to have a great effect and the participant's age was a stronger indicator of their views. Regardless of the diverse views expressed by the participants, almost everyone agreed that technology has substantially changed privacy.

This scenario can be further rationalized through the lens of the Technology Mediation theory, a framework that posits that society and technology mutually shape each other through their interactions. Applying this theory to the concept of privacy will allow us to see how the emergence of new technologies has contributed to the evolution of privacy.

Considering how smart devices impacted society reveals the interplay between the technologies and society. Once smart devices were introduced and online social networking became popular, individuals saw that less of their private lives was still private. Private conversations between friends were now preserved in databases owned by corporations, while a generation ago they would have been gone with the wind. This is one reason that older people feel wary when using smart devices.

Many technology companies intended to use this data to finetune the user's experience. However, the public had to choose whether to utilize this technology at the cost of losing autonomy of their personal data and at the risk of their data. Due to the convenience provided by smart devices, society chose to embrace the new technology regardless of the privacy costs. This choice by many, progressed the concept of privacy to include companies holding their personal data. As long as these companies properly handled their data, people felt as if their privacy was

intact. This forced the concept of privacy to evolve into an idea that was more accepting of the large amount of data being collected about people on a daily basis.

The impact of technology on privacy has been consistently observed using multiple research methods. As the use of technology grows, privacy will continue to evolve and governments, and their citizens, should be ready for it.

Bibliography

Allam, Z., & Jones, D. (2020). On the Coronavirus (COVID-19) Outbreak and the Smart City Network: Universal Data Sharing Standards Coupled with Artificial Intelligence (AI) to Benefit Urban Health Monitoring and Management. *Healthcare*, *8*(1), 46. https://doi.org/10.3390/healthcare8010046

Bellman, S., Johnson, E., Kobrin, S., & Lohse, G. (2004). International Differences in Information Privacy Concerns: A Global Survey of Consumers. *The Information Society*, *20*(5), 313-324. https://doi.org/10.1080/01972240490507956

Beskorovajnov, W., Dörre, F., Hartung, G., Koch, A., Müller-Quade, J., & Strufe, T. (2020). *ConTra Corona: Contact Tracing against the Coronavirus by Bridging the Centralized-Decentralized Divide for Stronger Privacy* [Ebook]. Retrieved 25 October 2020, from https://eprint.iacr.org/2020/505.pdf.

Cho, H., Ippolito, D., & William Yu, Y. (2020). *Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-Offs* [Ebook]. Retrieved 26 October 2020, from.

Kim, J., Lim, J., Moon, S., & Jang, B. (2019). Collecting Health Lifelog Data From Smartwatch Users in a Privacy-Preserving Manner. *IEEE Transactions On Consumer Electronics*, *65*(3), 369-378. https://doi.org/10.1109/tce.2019.2924466

Liu, J., & Sun, W. (2016). Smart Attacks against Intelligent Wearables in People-Centric Internet of Things. *IEEE Communications Magazine*, *54*(12), 44-49. https://doi.org/10.1109/mcom.2016.1600553cm

Perera, C., Ranjan, R., Wang, L., Khan, S., & Zomaya, A. (2015). Big Data Privacy in the Internet of Things Era. *IT Professional*, *17*(3), 32-39. https://doi.org/10.1109/mitp.2015.34

Perez, A., & Zeadally, S. (2018). Privacy Issues and Solutions for Consumer Wearables. *IT Professional*, *20*(4), 46-56. https://doi.org/10.1109/mitp.2017.265105905

Terzi, D., Terzi, R., & Sagiroglu, S. (2015). A survey on security and privacy issues in big data. *2015 10Th International Conference For Internet Technology And Secured Transactions (ICITST)*. https://doi.org/10.1109/icitst.2015.7412089

Udoh, E., & Alkharashi, A. (2016). Privacy risk awareness and the behavior of smartwatch users: A case study of Indiana University students. *2016 Future Technologies Conference (FTC)*. https://doi.org/10.1109/ftc.2016.7821714