5-2020

Digital Privacy: Personal Data Collection Methods and the Myth of Online Privacy

Ryan Salner
rsalner@gmail.com

Digital Privacy: Personal Data Collection Methods and the Myth of Online Privacy

by

Ryan Salner

A Starred Paper

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Information Assurance

May, 2020

Starred Paper Committee:
Abdullah Abu Hussein, Chairperson
Hazem Farra
Akalanka Bandara Mailewa

Abstract

Mobile devices offer users a constant connection to information and entertainment. Our society has become hyperconnected. We have unprecedented access to information at any time of the day. Mobile devices have the potential to make people more efficient and productive or more distracted and negatively influenced. The use of applications or apps on mobile devices brings with them unparalleled access to intimate information about the users of mobile devices. Corporations have been quick to provide apps that make life easier for, or entertain, the end-users. But the entertainment and access come at a price. That price is incredibly detailed information about the users, and it is being used and sold on the internet. Companies are requiring users to allow mobile applications access to far more detailed information than is necessary, and the end-user is unaware of just what the price they are paying is. This paper will explore the permissions that mobile apps request, a company's terms of service, and third-party relationships to determine if software manufacturers are honest with their stated permissions or if apps are overreaching in their efforts to collect information about their users. An examination of application permissions and analysis of the data transmissions to and from the device on behalf of the application will be performed. This work aims to provide users with more insight into how to protect their confidential data and to improve users' perception of privacy.

Keywords: Digital Privacy, Social Media, Personal Information, Mobile Apps

Acknowledgments

I would like to thank my committee, Dr. Abdullah Abu Hussein, Professor Hazem Farra, and Professor Akalanka Mailewa, for their time and input on this project. I also want to thank my wife and children for their patience and support.

Table of Contents

Chapter Page

Chapter Page

List of Tables

Table Page

List of Figures

Chapter I: Introduction

Introduction

The proliferation of mobile devices around the globe has created a society of hyper-connected people willing to share extremely personal information in exchange for convenience and entertainment. Social media services such as Facebook, YouTube, Snapchat, Twitter, and the like provide the pulse of society and allow people from every corner of the globe to connect in a myriad of ways. While these services cost nothing to the consumer, the budgets to support the digital infrastructure of these companies is astronomical. YouTube's annual operating costs alone are around $6.3 billion and produces only 6% of Google's ad sales revenue. Through the use of apps on mobile devices, many companies collect information from users while they interact with their app as a way to generate revenue by selling that information. However, some apps request more access to a device than is necessary for the tasks performed by the application. Other third-party applications are also common and collect data for monetization on behalf of the primary application or corporation.

Divided amongst various categories of applications and depending on the software platform of Apple or Android, there are hundreds of thousands of third-party tracking applications in the mobile app ecosystem. Unfortunately, the process of collection and monetization of personal data through the use of apps on mobile devices is largely not understood by the people who use those devices (Keng, 2016). Rajasegaran, Karunanayake, Gunathillake, Seneviratne, and Jourjon (2019) also exposed over 1,500 mobile applications that required unnecessary permissions to users' mobile devices and another 1,400 apps that utilized at

least five software libraries directed at the collection of information for the purposes of delivering targeting marketing.

By examining the permissions requested by an app, the apps terms of service, and the actual data transmitted from the devices, it can be determined if the apps are taking liberties with user's data. From these results, cross-examination can be performed to determine where the data is being sent and what companies are benefitting from the excessive data collection. Lastly, when a device owner decides to upgrade or trade-in their old devices, a forensic examination of what happens to the existing data can be done to determine if a factory reset actually removes a user's data or is it a false sense of security when it comes to personal data stored on a mobile device?

Until the users of mobile devices and applications choose to protect against the excessive and unregulated collection of personal data, they will continue to be manipulated by corporations through the virtually unfettered access to personal information.

Problem Statement

Widespread and daily use of mobile devices has created an economy of data that flows from the user to the corporation. Much of the data that is collected is done without the users' knowledge or even their active participation. A simple "I Agree" button is all it takes for a person to be a more valuable source of revenue than the products or services that a company sells. Applications are potentially understating the system permissions granted to the app, by the user, allowing for more access to data then a user expects. Some app vendors allow third-party marketing companies to "tag-along" on a user's device and collect even more personal information, including sensor data, geolocation information, and other unintentional user data.

When a company or service vendor application requests certain permissions to the user's device, those permissions are intended to benefit the user of the app by allowing the app to perform some function in tight relation to the functionality of the app. However, when an application that has no need for access to the device's camera, requests such access, the user is faced with the choice of allowing the access or not using the app. In addition, the terms of service for the application may or may not provide any additional protections for the end-user and may in fact be even more liberal with the data that is collected.

Nature and Significance of the Problem

By the year 2025, it is estimated that 72% of internet usage will be consumed through a mobile device (Handley, 2019). When corporations offer no-cost services to end-users, where does the revenue needed for that company to operate come from? According to Facebook's 10-K Security Exchange Commission filing (a report that gives a comprehensive summary of a company's financial performance), in 2012, each user of the service contributed $5.32 purely by using Facebook (United States Security and Exchange Commission, 2012). That number has grown to $6.09 in September of 2018. While the number may seem insignificant, consider when that amount is multiplied by the total user base equates to approximately $55.83 billion dollars worldwide (FourWeekMBA, 2019). The notion that one company can earn billions of dollars from the passive browsing of its users leads to a larger issue of applications that take more liberal and direct advantage of end-users by collecting information about location, activities, device usage and internet habits to manipulate what the user sees, shares or even believes.

In the documentary, *The Great Hack,* Amer and Noujaim (2019) exposed the use of personal data collected by social media companies to influence United States citizens during the

2016 election. By using data analytic and marketing company Cambridge Analytica, political strategists utilized data collected by Facebook, indicating that there are approximately 5,000 points of personal data that is tracked for every U.S. citizen. These pieces of information are then used to target messages, group suggestions, and news articles to the people that Cambridge Analytica deemed as "The Persuadables" in an effort to sway them toward their political client's side (Amer & Noujaim, 2019).

If end-users were more aware of what information companies were collecting about them and what those companies were doing with that data in an effort to influence the purchasing or promotion of products they may be less inclined to click "I agree" without careful consideration to what they might be agreeing to share with social media companies and their marketing partners. To that end, this study will examine those processes and connections to determine if what users agree to and what is being collected are equivalent or if the end-user is a victim whose personal information is being used to manipulate their online experience.

Objective of the Study

The objective of this study is to determine if the terms of service and permissions required of a mobile application are relevant and related to the operation of the application. Additionally, this paper intends to examine the attitudes and perceptions of the end-users to determine if the exchange of personal information for entertainment is a concern or not.

Study Questions

For this research, four essential questions were devised. These questions will help us identify potential data abuses and help define our end-user survey.

1. Are mobile applications collecting data unrelated to the functionality of the app? Is there cross-application data sharing between applications from different companies? Why does the application need the data it collects? Does the data serve a purpose in relation to the application, or is it collected purely for the benefit of the company?
2. What data is transmitted to and from a device once the application is installed? Are the applications utilizing the devices' sensors, wireless network connections, and cellular data transmission? Are the services being accessed beyond what is needed for the application to perform its function?
3. Does performing a "factory reset on a device remove personal data artifacts? What other means do end-users have when trying to protect their information?
4. What is the perception from the end-user's perspective of the exchange of personal information and digital privacy for entertainment?

Limitation of the Study

Due to the restricted accessibility to raw cellular data, the research being done on mobile devices will be limited to the Wi-Fi transmissions. Data transmitted from applications should remain relatively unchanged in content and purpose as the devices are designed to operate similarly across both mediums.

Definition of Terms

Android Application Package (APK): A compressed digital package of directories and libraries used to distribute applications on the Google Play Store.

App: An app or application is computer software, or a program, most commonly a small, specific one used for mobile devices. The term app originally referred to any mobile or desktop

application, but as more app stores have emerged to sell mobile apps to smartphone and tablet users, the term has evolved to refer to small programs that can be downloaded and installed all at once.

Average Revenue Per User (ARPU): ARPU is a measurement that companies use to indicate how much money a single user is worth to the company. This amount can be regional, national, or global in averages. For example, the ARPU for a North American user of Facebook is $27.61, but a global ARPU for a Facebook user is $6.09 (McFarlane, 2019).

EXIF: An exchangeable image file format is a data format standard for images and sound, which allows for additional data tags to be used to provide metadata such as geolocation information, device settings, and technical information.

Fiddler: Fiddler is an HTTP protocol debugging tool that allows the user to log and analyze web traffic between devices.

Data Point: A specific piece of information about an internet user. Collection of users "likes," "shares," posts, and locations, all create data points that allow companies to target specific experiences and messages to the end-user.

HTTP/HTTPS: Hypertext Transfer Protocol is a communication protocol designed for the formatting and presentation of internet website traffic processed by internet web browsers.

iOS: Apple Inc. operating system found in mobile devices created by the Apple corporation. Found on iPhone, iPod, and iPad devices.

Metadata: Descriptive information about a piece of data. Metadata can include creation date, user information, geo-location information, timestamps, file permissions, and the like.

Packet Capture: The process of intercepting or "capturing" network traffic across various mediums and devices in order to analyze and monitor communications between systems.

SD/MicroSD Card: Small removable device for storing data on computing devices. Micro SD cards are typically the size of a fingernail and fit into a small slot on a mobile device allowing for additional storage of files and media.

Sensor: An electronic component designed to capture input from the environment to provide feedback to applications and systems within the device.

Social Context: Social context is information that highlights a friend's connections with a brand or business. Facebook utilizes social context when selling marketing opportunities to advertisers and marketing groups to better target future ads and influencing articles.

Social Contract: The social contract in regard to privacy, social media, and mobile devices is defined as the agreement between the user and the social media company. Often the social contract includes some protection of the user and, in the case of digital privacy, an expectation of what is being done with the users' data.

Packet Capture: Packet capture is the process of intercepting a data packet that is crossing or moving over a specific computer network and saving that information for future analysis.

Public Key Infrastructure (PKI): A set of processes, policies, software, and hardware used to create an exchange of keys used to encrypt and decrypt data on the internet using digital certificates that are unique to each user or device.

WireShark: Wireshark is an industry-standard network packet analyzer. It allows the user to examine data sent across a communication medium such as network cable or wireless transmission.

Summary

The first part of this paper presents the issues at hand regarding personal information, and the access companies gain thorough mobile device application permissions in an effort to monetize the end-user. The questions presented will lead to a better understanding of how advertisers gain access to and collect information about the user's habits and activities. The terms explained allow readers of various technical knowledge to process the research with a better understanding of the technical issues. The next chapter of this research paper will be an examination of existing research and information regarding how user data is collected and what permissions apps have to a user's device. Additional review of research related to how social media's partnerships with third-party advertising companies might affect decisions into what permissions an application requests and what is done with such information.

Chapter II: Background and Review of Literature

Introduction

This chapter provides a review of existing literature and research in the area of mobile device permissions, digital privacy in relation to mobile applications, and the collection of user information by mobile devices and applications. This information will provide a thorough examination of the various pieces of the puzzle that mobile app permissions and potential overreach of those apps and their parent companies are participating in. In addition, how information regarding privacy ratings is presented in the various application stores influences user decisions to install and trust the app or not.

Background Related to the Problem

To better understand the problem, background on the issues, and motivations for companies to collect user data needs to be addressed. According to Cuofano (cited in FourWeekMBA, 2019), 98% of Facebook's revenue comes from highly targeted advertisements. These advertisements are finely tuned based on information collected about social media users (FourWeekMBA, 2019). Less than 2% of the revenue earns by Facebook comes from payments that developers provide to allow the processing of payments through the Facebook payment infrastructure (retrieved from http:fourweekmba.com). Facebook has experienced annual revenue increases at 37%, and ad revenue of over $55 billion. One of the primary factors of their growth is the better engagement of advertising campaigns, which leads to the heart of the problem of this paper. Facebook uses user information to create tuned advertising and sell those advertising opportunities to companies at a higher and higher rate because of the ability to target individuals

based on any number of identifiable metrics. These metrics provided to Facebook purely based on what users share by using the app (United States Security, 2012).

While Facebook is a clear and common focus for much of the research and criticism, other companies have learned from Facebook how best to utilize the population using their apps to generate additional revenue. The big five, Facebook, Amazon, Apple, Netflix, and Google, all participate in the collection, sales, and profit from end-user data collection. Because of the massive amount of digital touchpoints, a single person has in today's hyper-connected world, it is easy to feel like there is an invasion of privacy taking place. With the increased proliferation of mobile devices, our digital footprint becomes even larger, and many companies are taking advantage of users to finely tune their marking machines through overreaching permissions on mobile devices.

Recently, former CIA contractor Edward Snowden explained that it's not just the data collected from applications on your phone that help in identifying and tracking what you do. He discusses, in length, the ability for mobile devices to utilize the globally unique identification numbers of wireless access points, cellular towers and other signal transmission devices to triangulate a user's location even when not actively engaged in using any specific application (Rogan, 2019). Mr. Snowden emphasized the notion that social media and mobile devices influence the options and position of society.

Literature Related to the Problem

Shilton and Greene (2019) poured over numerous mobile software developer forums to discover what kind of conversations developers were having in regard to ethics in data collection, private conversations, and other "ethical deliberations." Shilton and Greene (2019)

explored the discussions that were had and how developers defined and justified their concerns regarding mobile software development. While the privacy conversations were considerably different between iOS and Android developers, the authors found that the developers had very similar justifications for including privacy options. Most of these justifications were based on the moral positions of the developers, "cautionary tales" and technical rationalization to legitimize privacy features (Martin & Shilton, 2016). The research paper explored the thoughts and concerns of the software developers when considering what information they feel is personal or privileged and should be treated as such. That research applies to this paper because it provides an insight into the consideration of privacy-first or process first and whether the user is an important piece of the development model.

Martin (2015 cited in Sarabia-Sánchez et al., 2019) researched how a person's experience with a product or industry related to their perception and expectation of privacy. The comparison between individuals with more experience against those with less experience determined whether or not an application's use and collection of personal information met the user's privacy expectations. The creation of "social contracts" in relation to what information people deem personal and why that information should be viewed as personal or private. Also, why a company might want to collect or track that information. Martin researched how privacy norms are developed through the lens of a social contract and to define what a privacy violation is given the link between the social contract and end-user expectations. The research helps define the role of the end-user and the role of the software provider or social media company in defining what constitutes a privacy norm or expected use and protection of the end-user's data. Martin includes the notion that users are more willing to share information within a specific community. For

example, participation within a particular Facebook group allows a user to enter into a state of mind that what they share information within that group more freely because of the false notion that the information shared is contained within that group. When in reality, the information shared within the social groups and pages of social media sites is just as accessible and open to the company that runs the site than it would be if it were posted or shared directly to the main feeds of the site. The user wrongly believes that they are being discrete and discriminate in the information they are sharing. Martin's paper provides evidence to the effect that users are often unaware or lead to a false sense of security when sharing information based on how the service provider presents the social construct.

The Amer and Noujaim (2019) documentary, *The Great Hack*, dug deeply into the use of Facebook's massive data point collection of its millions of users. The digital traces that users create as they utilize the applications provide marketers with extremely detailed information about users' likes, dislikes, and relationships. This data is more valuable than oil in 2019 and, when a company gains access to this incredible resource, it can apply psychological profiling to cultivate a narrative that has the ability to influence entire populations of people (Amer & Noujaim, 2019). In the documentary, Cambridge University professor Aleksander Kogan explained that app users agree to the terms of an application based, not on what it collects about them, but what the app does for them. When a user clicks on a silly survey, they are allowing that survey to reach through their profile and connect to every one of their contacts or friends. These companies are using this data in ways that the end-user does not understand and does not care about, mostly because they do not see how it affects them. There are currently no laws that

protect the end-user from the collection of personal information in this manner because the user agrees to share it.

The unaware user and the terms of service will be examined in this paper to determine if the permissions granted by the user have been clearly and thoroughly represented. An effort will be made to track and measure the various data pulls and pushes an app performs to verify what data an application is sending and storing.

In their exploration of what they term the "privacy paradox," researchers Sarabia-Sánchez, Aguado, and Martínez-Martínez (2019) produced an argument that there is no rational connection between the emotional response to social media stimuli and the users' discretion when it comes to managing applications and social media services permissions to their personal information. Sarabia-Sánchez et al. referenced the Cambridge Analytica incident and how Facebook sold over 80 million users' data and the fact that the fallout of that event has had virtually no impact on the millions of users affected and influenced by Cambridge Analytica. The authors also posited that "while privacy is a primary concern for end-users, they are just as easily distracted from that concern in exchange for insignificant rewards or time-saving convenience" (p. 2). The research shows that there are a number of proposals that attempt to explain why users are so willing to forgo basic data privacy practices in regard to social media access. Some of these explanations include the increased need for instant gratification or psychological compensation for future rewards (Do, Martini, & Choo, 2014). Additionally, some of the indifference to digital privacy is the notion that privacy in the modern age of the 'omnipresent internet' no longer actually exists. While these notions are logical from a human emotional perspective, they are confusing when the user is suddenly concerned about a data

breach in which similar or less intrusive information is taken when they have clearly demonstrated that they are willing to provide such information freely. The research done by Sarabia-Sánchez and Martínez-Martínez provides a foundational understanding of why end-users might not understand or even care that a particular application is misrepresenting its permissions or access to the user's data.

Polykalas, Prezerakos, Chrysidou, and Pylarinou (2017) stated right in the title of their research paper that "when the service is free, the product is your data." In their paper, the authors examined the Google Play free app landscape to determine what the cost of a free app really is. Their research mentioned the free apps often provide the user terms of service page similar to the type of End-user License Agreement or EULA that users of desktop software are used to seeing, and just as quickly dismissing. They also discovered that the free apps often do not include any way to modify the permissions that the application claims once the app has been installed. Since the end-users do not read the full contents of the agreement, they are arbitrarily choosing to accept the terms and give little thought of the access these applications have to their data once they start using the app. Additionally, their paper examined over 500 free applications and categorizes them by the types of access, how many times they were installed, and what ratings users have given these apps. The results of this research showed that the permission requests were vague, stating things like "Access to Wi-Fi" but, when the authors explored the definition of such permissions, they discovered that the extent of that access was much more than the average users really understood. From a security perspective, access to the Wi-Fi connection means that all data traversing that connection could be captured. The most common permission requested was access to the users "Photo/Media/Files" which ultimately allowed the software to

access every file on the mobile device. This research directly relates to this paper in that this paper intends to examine similar claims in both the Android and Apple software repositories.

In a related paper, Sleeper et al. (2018) approached the issue of digital privacy from a slightly different but still very relevant angle. In her paper, she examines the effect of hardship on an end user's perception of how important privacy is and what challenges those with economic difficulties face. Through a focused qualitative study of residents of transitional homeless shelter, Sleeper et al. explored what the issues were regarding digital security and privacy. Their findings showed that four major factors came to light that affected the user's positions on digital privacy. Those issues included financial resources, unreliable devices, personal relationships, and stress. The financial resources issue presented itself as a failure to be able to purchase reliable or up to date mobile devices. This leads to older protocols, and other inherent security issues with the device itself as well as limitations to software updates as older devices are often unable to run the latest versions of operating systems. The issue of personal relationships presents the issue of partner abuse and the higher ratio of online stalking and harassment.

The importance of these issues should not be ignored in the greater examination of digital privacy as they expose the audience of these papers to issues that are potentially unfamiliar yet still important pieces of the overall conversation.

The Economics of Privacy by Acquisti, Taylor, and Wagman (2016), takes a look at the issue of the "value and regulation" of personal data and its use by corporations in a myriad of ways. Acquisti et al. pointed out that personal information is not just the height, weight economic status of an individual but also their digital experiences such as each mouse click, media post,

and photo upload. How all of these provide companies with abundant economic value. The problem arises when the user suddenly feels as if the personal benefits of using an application in exchange for personal information is no longer equitable. At one point in the paper, they made a rather poignant point, “Privacy is not the opposite of sharing - rather, is control over sharing.”

This tends to be the heart of the matter, and where much of the disagreement on personal data collection occurs. As much of the research has shown, there is a perception that most users have that is used to determine if they are sharing enough, too much, or being taken advantage of.

Literature Related to the Methodology

Wu, Chen, and Clarke (2014) produced a paper titled *Sensitive Data Protection on Mobile Devices*, which addressed how sensitive data is stored and can leak from mobile devices. Wu et al. described the various data transfer methods such as computer to computer, mobile to mobile and mobile to the server. For the purposes of this paper, the transmission from mobile to the server (or service provider) will be examined. Wu et al. continued to describe the types of data a mobile device can store, such as personal information, device sensor, and GPS information as well as metadata related to the creation and modification of data on the device. The authors presented three possible ways to protect the data on the device, which includes backing up the information to a secondary location, apply encryption to the information, and securing the device with a “lock and wipe code.” The lock and wipe code would force the device to lock after failed attempts to access the device and then perform a factory reset of the device. The section on “lock and wipe” directly applies to the research that will be done in this paper as we explore the functionality of the “factory reset” to determine if the personal information and

metadata of the device are truly deleted from the device or if any of it can be recovered by third-party software.

Arapinis, Mancini, Ritter, and Ryan (2017) examined mobile devices from an attacker/target perspective. The applied research identified that mobile device permissions and security policies need to be reworked to include the multitude of sensors and data that the device itself produces. Often the device is generating this information without the input or knowledge of the end-user. As a result, end-users rarely, if ever, acknowledge that there is a privacy issue. Arapinis et al. proposed a formal verification methodology to better define a set of standard permissions, or at the very least, a level of access that all end-users might agree upon that software creators should abide by. Which he admits would be difficult since users share information differently with different people and in different situations. The authors proceeded to diagram and describe the various methods of attack on a mobile device to further illustrate the difficulty in securing mobile devices. Lastly, using WireShark, the authors captured transmissions from mobile devices to take a more detailed look at some of the security protocols. What they discovered was that even one of their fundamental security standards, a particular algorithmic value that should be updated frequently, went unchanged for extended periods of time, including almost three days at one point in their research, as shown in Figure 1.

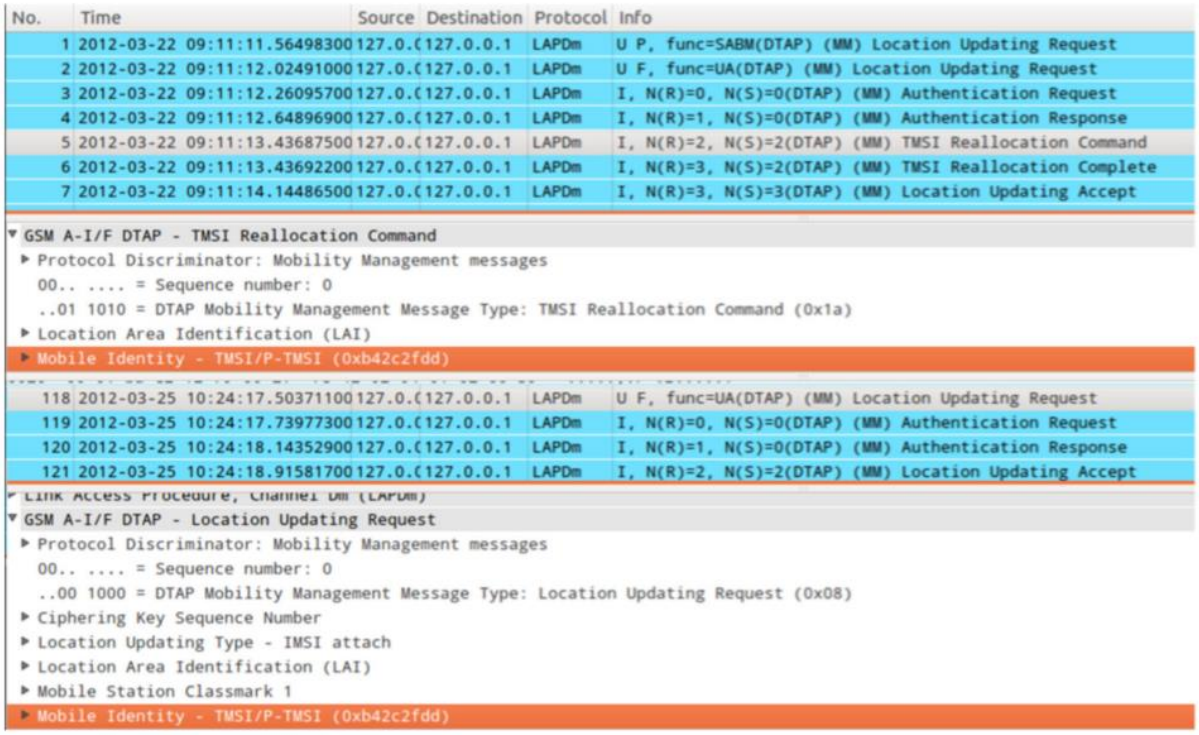

Figure 1. Demonstration that the TMSI Value Has Failed to be Updated for Three Days (Arapinis et al., 2017)

The point is made that their research is not only theoretical but also demonstrates that it is possible for a malicious actor to manipulate the vulnerabilities to breach the user's privacy. Ultimately the authors' several fixes to the problems they exposed include but not limited to PKI at various layers and fixes to the mobile device identification algorithm.

Much of the research in mobile data security focuses on the Android environment. Slavin et al. (2016) presented a tool to help software developers identify coding deficiencies in relation to privacy standards and then recommend potential solutions, thus helping curb the unnecessary overreach of some applications when requesting permissions on mobile devices. In their research, they developed a tool that performed two major functions. The first function was to create an ontology of security-related phrases for the applications. The second function of their

application was to then map the phrases to functions within the Android API. Once their application was complete, they could run any Android app through the software to compare the byte code of the software to the privacy policy of the app to create a list of violations the ranged in severity. Figure 2 shows the structure of the application, as created by Slavin et al (2016).

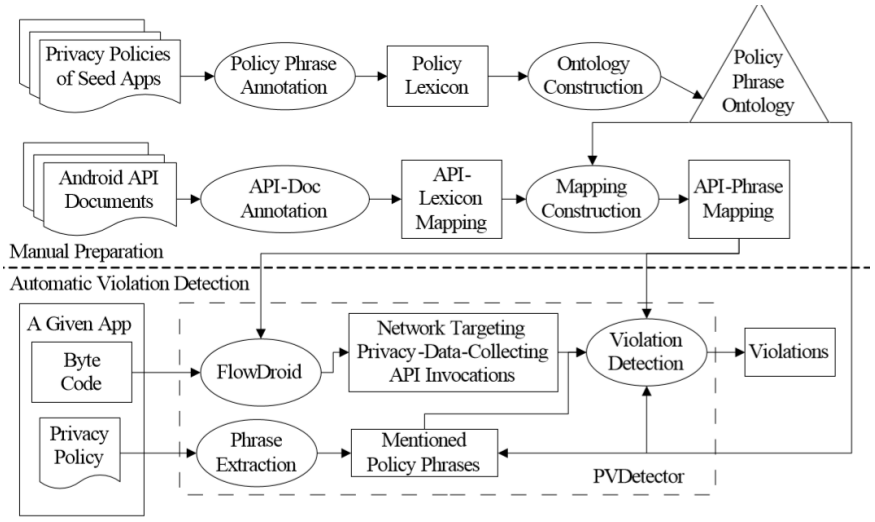

Figure 2. Diagram of PVDetector application (Slavin et al., 2016)

Choe, Jung, Lee, and Fisher (2013) approached the problem of mobile apps overreaching for permissions by proposing a visual representation of an application's privacy score. By collecting information on over 100 apps in the Android and Apple stores, they created a rating system in an attempt to help the end-user visually recognize an app that had the potential to collect more personal data than what was advertised or presented. With this information made available to the end-users, the second phase of the study explored what the result of having such knowledge made on the user's choice of applications to install on their mobile device. By applying a visual of a privacy rating, or privacy rating and a user rating, Choe et al. surveyed 332 users to determine if users' decisions to load an app would be influenced based on the scales presented. Their results showed some significant results when the privacy values were

represented in high and low privacy ratings, but little change when the apps displayed high and medium values. Moreover, depending on how the information was framed (either negatively or positively), users had differing opinions. As shown in Table 1, if the privacy score of an application is presented in a negative format, users were more inclined to install that app than if the score was represented in a positive format. However, the authors mentioned that results were the opposite when the framing between positive and negative were done in text descriptions rather than graphical images (i.e., plus signs vs. minus signs).

Table 1

Results of Survey Representing Privacy in a Negative or Positive Format, With or Without Recommendation Data Support.

Privacy Rating	Answer	INSTALL Question			RECOMMEND Question		
		Positive Framing	Negative Framing	p-value[a]	Positive Framing	Negative Framing	p-value[a]
Low Privacy Rating App	Yes	3 (2.8%)	9 (8.6%)	.06‡	2 (1.8%)	10 (9.5%)	.02*
	No	106 (97.2%)	96 (91.4%)		107 (98.2%)	95 (90.5%)	
Medium Privacy Rating App	Yes	13 (11.9%)	25 (23.8%)	.02*	10 (9.2%)	17 (16.2%)	.12
	No	96 (88.1%)	80 (76.2%)		99 (90.8%)	88 (83.8%)	
		No User Rating	User Rating = 3	p-value[a]	No User Rating	User Rating = 3	p-value[a]
High Privacy Rating App	Yes	88 (82.2%)	77 (72.0%)	.07‡	83 (77.6%)	69 (64.5%)	.04*
	No	19 (17.8%)	13 (28.0%)		24 (22.4%)	38 (35.5%)	

The information collected by Choe et al. (2013) ultimately shows that how users are exposed to the information about the level of privacy that an app does or does not provide can mean a world of difference when it comes to the user making an informed decision about downloading sed app. For the purposes of this paper, the questions asked of the users will purposely avoid any ambiguity in the survey questions.

There are several ways of mitigating or controlling over-privileged applications, but in a paper by Do et al. (2014), one suggestion was to reverse engineer the software coming from the

Google Play Store. Specifically, the authors suggest an approach that would examine an application APK and the AndroidManifest.xml file where the permissions for the application are stored. Figure 3 shows a generic APK, and Table 2 shows a list of specific permissions that each social media apps are known to request from mobile devices. Do et al. devised a process of unpacking the APK and modifying the XML file used to apply the permission for the application to the mobile device the repackage the application before the mobile device installs it. In their testing, they were careful not to remove permissions that would break the overall functionality of the app, such as internet access for social media applications or camera access for photo-sharing apps like Instagram.

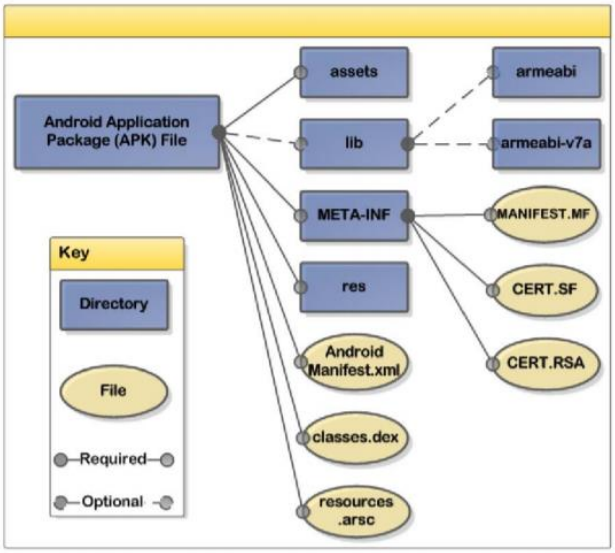

Figure 3. Android APK Structure

Table 2

Common App Permissions

Permission	Facebook	Twitter	Instagram	Tango Text	Pinterest	LinkedIn	Tumblr
ACCESS_ FINE_LOCATION	X	X	X				
ACCESS_ NETWORK_STATE	X	X		X	X	X	X
AUTHENTICATE_ ACCOUNTS	X	X		X		X	X
CAMERA	X		X	X			
GET_ACCOUNTS	X	X		X	X	X	X
INTERNET	X	X	X	X	X	X	X
MANAGE_ ACCOUNTS	X	X		X		X	X
READ_CONTACTS	X	X	X	X		X	X
READ_ PHONE_STATE	X			X		X	
READ_ SYNC_SETTINGS	X	X		X		X	X
VIBRATE	X	X		X		X	
WAKE_LOCK	X	X	X	X	X	X	X
WRITE_CONTACTS	X	X		X		X	
WRITE_EXTERNAL_ STORAGE	X	X	X	X	X	X	X
WRITE_ SYNC_SETTINGS	X	X		X		X	X

The most difficult issue with this approach is that it could be possible for permissions to be so intertwined with the code of the application for the permission to be removed and still have app functionality. Additionally, as soon as an app requires an update, the entire process would have to be repeated, and given the frequent nature of mobile app updates, this would cause a major challenge. The remainder of the paper explored the testing of removing specific permissions, one at a time, and recording the results of attempting to run the app. Some

permissions allowed for the app to operate as expected until a function of the app needed access to the data permission provided, such as READ_ CONTACTS, when Facebook attempts to find friends. The results cause the app to crash. While Do et al.'s (2014) work in removing permissions was a manual endeavor, it leads the way for the possibility of automating the process in the future. The research is relevant to this paper because of the need for the potential curbing of apps, possibly over-reaching in terms of privilege and a need to provide end-users some possible control of their own digital environments.

Lastly, the application of sensor data, in addition to the regular usage of social media applications, has been used extensively to research mental health issues (Saeb et al., 2015). Using GPS/location data, usage statistics such as duration and features, and location variance helped clinicians identify habits and early indicators of depression and other mental illness issues. This type of research is important as it shows the positive application of user data to benefit society as a whole; however, when used for marketing, the same positives can be turned into negative, especially when the users are unaware of the collection and use of their data.

Summary

This chapter examined the broad range of subject matter associated with the problems of social media data collection and end-user compliance. The literature includes research into the software developers' conversations regarding user privacy and what liberties users allow in exchange for a social media provider's service, issues with devices using stagnant data, and software used to flag potential overreaching privileges in applications. The next chapter will begin with the design of this paper's research into what information the social contract between

end-users and service providers details and what information is collected. The methodology of the research to follow will help determine if there is a breach in the social contract.

Chapter III: Methodology

Introduction

The first two research questions of this paper examine the data communications of applications installed on a mobile device. The devices in this research are two cellular phones. However, capturing cellular data is only possible from a cellular service provider or a law enforcement agency. As such, the focus of the research was placed on the capture and analysis of wireless data to and from the two mobile devices. An examination of the terms of service of each of the five application being examined in this research was done to evaluate the intent and purpose of the application in relation to how the application is used by the end-user, the access to device resources such as sensors, accelerometers, global position systems, and system idle processes.

The third research question of "Does performing a factory reset on a device removes personal data artifacts?" is performed by capturing a forensic image of the systems storage device to determine the contents of the device after being used by a user for a period of time and then after a factory reset is completed.

Lastly, a survey of mobile device users was conducted to determine if the attitudes and understanding of digital privacy and mobile apps coincides with the mobile app and social media usage. Attitudes towards acceptance of third-party data sharing in exchange for entertainment or usability were also measured to evaluate where the trade-off of privacy for service occurs.

Design of the Study

The research for this paper was done using a qualitative examination of mobile device applications and transmissions to determine whether appropriate and relevant data was being

accessed based on the permissions expected and granted by the end-user. To answer Research Question 4, the devices were factory reset using the operating software settings and then imaged to view existing file structures and device content. Finally, an additional quantitative, online survey was performed to measure end-user attitude and understanding of mobile device permissions and application privileges.

Data Collection

In order to accurately answer Research Questions 1 and 2, data from the two devices was captured and analyzed in a controlled environment. The capture of the application and operating system over Wi-Fi was performed using a specialty device called a Wi-Fi Pineapple. The devices are marketed as a penetration testing device for performing man-in-the-middle attacks, advance reconnaissance, and open-source intelligence gathering. The device was connected to a Dell G7, Intel processor-based laptop via USB connection. Figure 4 diagrams the lab design. Data transmitted from the mobile devices to the network was captured using the software application Wireshark. The laptop computer was connected to the internet via an ethernet connection to an Arris brand cable modem, which was a connection to the researcher's ISP.

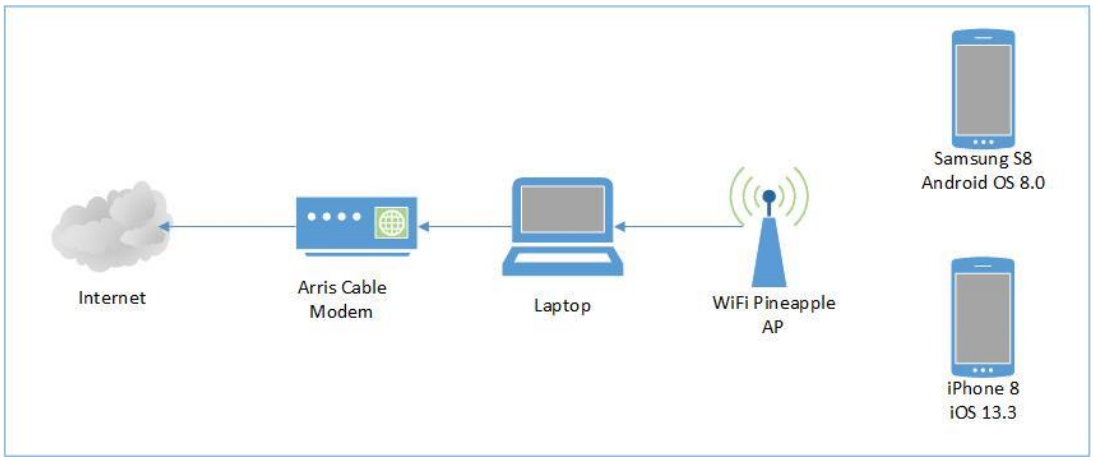

Figure 4. Research Lab Hardware Layout

Because research is conducted in the researcher's home location for several social media accounts, an additional step of installing a VPN service was taken. The VPN service allows for the originating location of the application data to appear to come from a different location. Data collection was done with both VPN enabled and disabled to evaluate what, if any, communication occurred differently when presenting from a different location.

To collect data in an as standardized state as possible, both mobile devices were factory reset and configured with as near-identical settings as possible between the two different operating systems. Both phones were joined to the research SSID (named research). From this point on, each device was operated individually and never at the same time as the other to ensure that data collection was clear on which device was sending and receiving data and what data was being sent or received.

Social media applications used for the research include Facebook, Snapchat, Instagram, LinkedIn, and MyFitnessPal. Except for Instagram being owned by Facebook, the other applications have different parent companies. This is important since data from two different applications from the same company may or may not be accessing different remote IPs. For Research Question 4, "Does performing a 'factory reset' on a device remove personal data artifacts?," both devices were decommissioned from the research and the phones operating systems functions were used to perform a 'factory reset' placing the devices in a "like new" state. Before being decommissioned for reset, after being used for the duration of the research, the devices were imaged to provide a capture of the state of their respective storage medium. Upon the completion of the imaging, both forensic images were examined using Oxygen Forensic Detective. No additional set up or configuration was done to the devices, and both

devices were powered off until the forensic image was captured, after which the same software was used to determine if any previous user data could be recovered from the device.

Tools and Techniques

The tools used in this research included two modern mobile devices from two of the most prolific device vendors. The first device was a Samsung Galaxy S8 cellular/wireless phone. It was running version 8.0 of the Android operating. The second device was an Apple iPhone 8 running version 13.3 of the iOS (see Figure 5). Both devices had the cellular service disconnected, and all transmissions were monitored through the Wi-Fi adapter. Due to limitations in collecting the cellular transmissions, we have decided to focus on the signals we have more control over capturing without violating FCC laws.

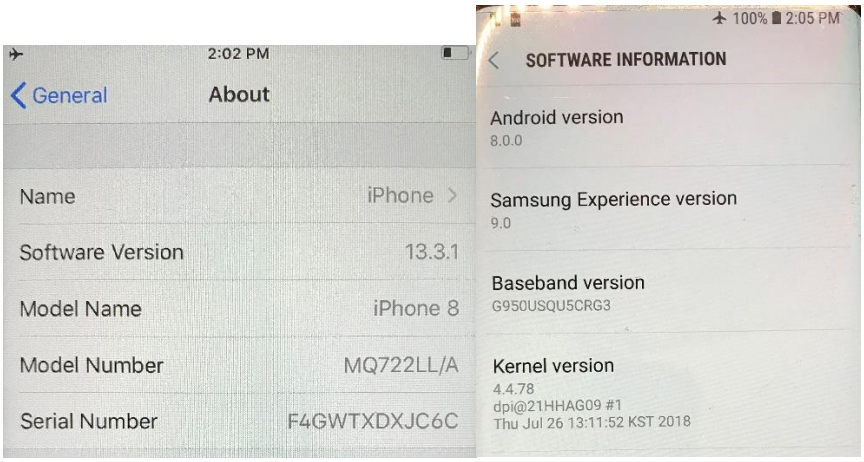

Figure 5. Mobile Phone Configurations

All software not run from the mobile devices was run from a Dell G7 laptop running Windows 10 Professional edition. Several device libraries were installed to ensure compatibility between the mobile devices and forensic software. Wireshark used to capture network traffic between the mobile device and the internet, was installed on the laptop. Additionally, an

application called Fiddler was installed to assist with capturing and analyzing any HTTP/HTTPS protocol data traveling between the internet and the mobile devices as well.

To answer some of the questions proposed in this paper, the capture of data to and from each device was necessary. Efforts were taken to ensure that the data being captured was only from our specific devices. The network, as referenced in Figure 4 earlier in the paper, was used and segmented from all other network and internet traffic. The two mobile devices were then used individually on the research network to avoid any sort of crosstalk between devices. Each device was started, joined to the research wireless access point then used in accordance with the operation of the social media applications being tested. The Samsung S8 was the first device used for testing. The android operating system connected to the research AP and promptly began transmitting requests for software updates to several companies. This data was captured, and the system was updated as needed before the testing of the social media applications began. Once the system and network traffic were idle, the device was used to access the Google Play Store to download and install the first social media application to be tested. Only one application was tested at a time to make discerning which application was communicating with which remote system during use clearer than if all applications were installed at the same time. The reason for this extra step was because initial testing with multiple applications showed that all of the applications sent and received data even when the application was not actively being used on the device. That initial data was discarded but did provide some insight into the operation of the applications during the devices' idle state and prompted additional testing to be discussed later in this section.

Each social media application was used for a period of two hours and then allowed to be idle for 24 hours with a follow-up use for another hour. Activities using the application include standard usage, such as scrolling through the news/activity feed. “Like”-ing other people’s postings and creating some posts. Testing avoided clicking on “sponsored” links and targeted marketing links. At the conclusion of the active testing time, the application was “minimized” by pressing the home button of the device, and the home screen was displayed until the device turned off the screen based on the default setting of the operating system. The device was left to idle for 24 hours. Idle was defined as no actively open applications, and the screen was off and untouched for the entirety of the 24 hours. No user input or influence on the device during that time.

Forensic images of the mobile device’s storage were done using several different software applications. A post-usage capture of each device was performed using Oxygen Forensic Detective to provide a deep level inspection of the storage medium. The same software was used to compare the post-usage state to the factory-reset state to determine if any files could be recovered post-wipe. Additional tools were employed to ensure due diligence in the research. Using Android’s SDK, the Android storage system was examined and indexed. To examine the Apple iOS, the software application iDevice Manager was used to index existing files and explore application folders for user data.

The quantitative research was performed using a survey created for deployment to college students in the Network Administration program at St. Cloud Technical and Community College. Forty-six students participated in the online survey between the ages of 19 and 28 years of age. The selection of college-age information technology students was decided upon for a

couple of reasons. The first, IT students were expected to have a deeper understanding of technology, data communications and basic security awareness. Secondly, most if not all these students use multiple social media applications, so they are familiar with the use of the applications in the research. The survey asked users to verify that they have experience using a variety of mobile applications, including but not limited to the applications studied in this paper. The questions looked for answers to how users felt about their trust in the way companies used their data, the benefits of sharing their data, and their opinion on the cost of privacy in terms of control over accessibility. The survey also evaluated the knowledge of users regarding the privacy controls available to them.

Hardware and Software Environment

The research done in this study was performed, as indicated previously. The first device will be a Samsung Galaxy S8 running the Android operating system version 8.0.0 with a kernel version of 4.4.78. The second device will be an Apple iPhone version 8 running an operating system version 13.3. These mobile devices were used to run applications and measure application privileges, access, and transmission. In order to capture and process the information from the mobile devices, this study incorporated a wireless access point designed to allow for the capture of data transmissions related to the operation of the mobile devices. The Wi-Fi Pineapple from Hak5 provided the research to be conducted by passing all wireless communication from the mobile device through the access point, which then passed through the ethernet adapter on the laptop computer to be captured and processed before being sent along to the internet. All returning data is passed back through the same route. Examination of the data transmissions was captured using the WireShark application on a Dell G7 laptop computer running the most current

version of Windows 10. A forensic examination of the mobile device was done at the end of the permissions tests to capture the state of the systems, including files and metadata captured during the use of the devices. This information will be cataloged, and the systems will then be “factory reset.” After each device is reset, the same forensic examination was done to determine if the factory reset left any personal data behind. The examination of the storage media was done using two methods of examination.

Summary

The process of collecting forensic data from a mobile device has substantial challenges. With current encryption standards and system sandboxing, methods system files and non-application data can be difficult to extract from the devices. Federal restrictions on monitoring cellular data also limit data collection methods. However, mobile devices still provide complete data communication through the IEEE wireless standard 802.11 and its extensions. We were able to capture data as it is passed from the device through an access point and passed along to a wired connection. This chapter laid out the methodology of the research to provide an understanding of how the data would travel from the mobile device through the collection computer and on to the internet. Software running on the workstation would provide a glimpse into the sources and destinations of communications from each of the applications. The next chapter will explore the data that was collected and what it means in relation to each of the research questions posed in this research document.

Chapter IV: Data Presentation and Analysis

Introduction

The research questions proposed at the beginning of this paper laid forth three major sections of data collection and analysis. The first, an examination of the privacy policies of five different social media applications. These policies were examined for expected protections and access to various device resources such as cameras, storage, and location information. This information was then compared to the information that was sent to the applications various URLs and IP addresses to determine if the applications were abiding by not only the data collected but to whom the data was sent. In addition, data was examined to determine if the applications were sending data that did not coincide with the purpose or needs of the application. The third research question addresses what happens to all of the data that applications store on the device itself. This chapter will explore the storage medium of the mobile devices before and after a user has operated the devices for a period of time. Pre-usage, post usage, and post factory reset status of the devices' digital storage will be analyzed. The last research question will provide insight to what end-users value in regard to digital privacy and examine the willingness to exchange various levels of privacy for convenience or entertainment.

According to research done by Zang, Dummit, Graves, Lisker, and Sweeney (2015), there are several ways to examine the privileges and permissions requested by mobile device applications. The first is permissions analysis, where the terms and policies of each company are viewed and compared. The second approach is static code analysis, which involves the decompiling of an application. While effective requires several tools unavailable to this author. Lastly, the dynamic analysis of an application captures what is happening while an application is

being used (Zang et al., 2015). This paper used the first and third methods suggested by Zang et al. to evaluate the research questions and determine if the applications are abusing the trust and expectations of the users to collect and profit from their personal information.

Data Presentation

In today's social media and mobile device environment, the installation of an application by an end-user typically involves that user clicks on a checkbox or button indicating that they are agreeing to both the Terms of Service and the Privacy Policy of whichever company is providing the application. However, very few people spend more than a moments glance at the actual terms presented, and those documents are often considered too difficult to understand or too obscure to evaluate by many users (Mcdonald & Cranor, 2008).

The United States has no fewer than eight federal or state laws dictating that any company collecting information that can be used to identify a person must include a Privacy Policy (TermsFeed, 2020). These pieces of information include, but are not limited to birthdates, first and land names, billing addresses, and email addresses. Personally Identifiable Information (PII), in the United States, is regulated by the Federal Trade Commission (FTC), and some states are taking action to create more transparency between the policies, companies, and end-users. Starting January 1, for example, the state of California implemented the California Consumer Privacy Act (CCPA) to require businesses that reach California residents to provide them with a Privacy Policy to promote transparency and to provide end-users with more control of their personal information and how it is used. For the purposes of this research paper, we will explore the Privacy Policies of the five applications used to explore the research questions. Those five applications are Facebook, Instagram, Snapchat, LinkedIn, and MyFitnessPal.

The following section will summarize the policies of each of the applications to provide a baseline of what should be expected by using each application in regard to the user's personal information.

Facebook

Facebook's Data Policy (Facebook, 2020) lays out the details of the information they collect and how they use that information. Basically, any data the user enters or accesses while using the application are a fair game for Facebook to access. This includes "Things you and others do and provide," which means that from the time you create an account to the moment you shut down the application, Facebook can use your information. It also means that Facebook can use anything that anyone you interact with, enters into the application as well. "Our systems automatically process content and communications you and others provide to analyze context and what's in them" (Facebook, 2020). If you share something and someone likes that post, you are now linked to that person and are creating a network of people, preferences, and connections that Facebook will use in a variety of ways, which will be discussed later in this section. A major focus of this paper is the information not provided by the end-user but rather by the end-user device. Facebook Data Policy (Facebook, 2020) contains a section addressing what information the company collects. This includes computers, phones, connected TVs, and any internet-connected device that uses a product provided by the Facebook company. The information that the policy says they collect includes Device attributes such as operating system, hardware and software versions, battery levels, signal strength, and file names. Additionally, the policy lists device operations, device signals, and settings and network connections. In the Device operations section, Facebook mentions the collection of foreground and background useable of the

application, mouse movements. However, there is no mention of other specific sensor data collection. An important aspect of this research includes the use of sensor data associated with the use of the application our research will address this issue later in the paper. GPS information, signal information, including Wi-Fi access points and cellular towers accessed by the device, are also collected and fair game for Facebook to use. Included in that information is your mobile device phone number as well as "other devices that are nearby or on your network" (Facebook, 2020). All this information is fair game for Facebook to use to identify, target, and influence users.

In addition to the information that Facebook collects directly from its application, Website, and its users, they also collect information from any app developer who uses Facebook's social plug-in such as the Like button, or if an app allows users to create accounts/log in using their Facebook account. All of these things allow Facebook to track, link any consolidate users' actions and activities to develop a robust profile of the user. Lastly, Facebook has a program called Facebook Pixel which allows web developers to add a single transparent pixel to their websites that allow Facebook with "data about what millions of people read, shop for, and watch online as they move around the web." (John, 2018). If you have ever left an item in the online shopping cart of a website only to see an ad for that or a similar item on Facebook, that is what Pixel is doing.

Instagram

In 2012, Instagram was purchased by Facebook and presented users with a notice indicating that they, the users, would still have control over who views what they share. However, while the end-user can control who can see the images they post to the application and

site, they still are bound by Facebook's terms and privacy policy. So anything not covered directly by Instagram's policies are still covered by Facebook's policies. Instagram does directly mention that they "When you use a mobile device like a tablet or phone to access our Service, we may access, collect, monitor, store on your device, and/or remotely store one or more 'device identifiers' "(Instagram, 2017). The modern image capturing devices, like mobile phones, include a great deal of information when taking a picture or screenshot. This information is called Metadata and, in its raw format, provides information such as GPS location, time and date information, and device settings. If users are unaware of this information, they could potentially expose personal information to the public. This information can have global implications. For example, a Russian soldier posted selfies while in Ukraine. Despite the fact that no Russian soldier should have been in the area or the country at the time. The Russian military denied the soldier was ever in the location; however, the metadata from the image posted shows information counter to that claim (Gallagher, 2014). Instagram, however, does not enable this option by default, and examination of images uploaded with default app configurations do not provide metadata. While Instagram may scrub such information when a user posts to their site, it does still get transmitted to Instagram for their use. Figure 6 below shows the same image in its raw format and then after being posted to Instagram and downloaded for examination.

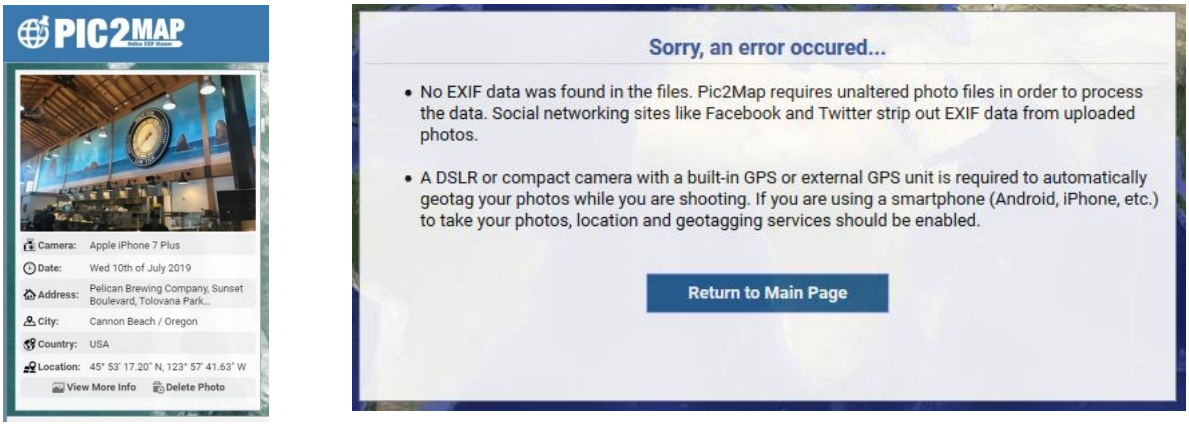

Figure 6. EXFIL/Metadata Results of Original Versus Instagram Posted Image (Pic2Map, 2020)

Notice that the original image on the left contains metadata, including the longitude and latitude of where the image was taken while the second image has that information removed. The policy is fairly light on how the application uses the data, and that is partially due to the fact that they are owned by Facebook and fall under their policy umbrella, but it does address a few things directly. Instagram uses user data to "provide personalized content and information" as well as "monitor metrics" for visitor counts and site usage. Noticeably absent from the privacy policy is any mention of device information short of the device identifier data that is stored.

Snapchat

Snapchat is similar in service to Instagram. The parent company of Snapchat is Snap Inc. They provide a lengthy privacy policy similar to the other companies, but they also provide a link to a toned-down, easier to read policy summary titled "Your Privacy, Explained" which leaves out the technical details and states that it is "blissfully free of the legalese that often clouds these documents" (Snap Inc., 2019).

In a similar fashion, Snapchat collects the information you provide, such as user profile information, as well as information collected when you use the service. Information collected

when using the service includes EXFIL information (metadata) from images as well as any text/emoticons used and any picture "lenses," which apply cute, silly, or unusual overlays to the pictures taken. These lenses can be geofenced to allow advertisers to create custom and temporary lenses people can use to promote an event or product with a Snapchat image posting. These advertisers become one of the many "partners" social media company mentions in their privacy notices. Snap Inc. defines usage information as any time you interact with the application, including but not limited to snaps and chats with other users, exchanged messages, and when you open a message or view someone else's snaps. They are also clear that they vie wall metadata of images provided to the service. Lastly, about Snapchat, they are the clearest about the collection of device sensor data, "such as accelerometers, gyroscopes, compasses, microphones, and whether you have headphones connected" (Snap Inc., 2019). Much of the information listed in Snap Inc.'s privacy policy is clear as they outright state that they will use your GPS, wireless, and cellular locations and that they have access to your images.

LinkedIn

As with the previous applications and services, the LinkedIn Privacy policy highlights what we've come to expect. However, LinkedIn is the first to provide a link to address California's specific consumer privacy law CCPA (California Consumer Privacy Act), where it clearly states several times that "We do not sell personal information" (California Privacy Disclosure/LinkedIn, 2020). In addition, through the online Policy Agreement, LinkedIn includes links to the settings that allow users to opt-in or out of certain data collection options such as the tracking of visits to other websites even when not logged into LinkedIn and additional cross-website tracking as seen in Figure 7 below.

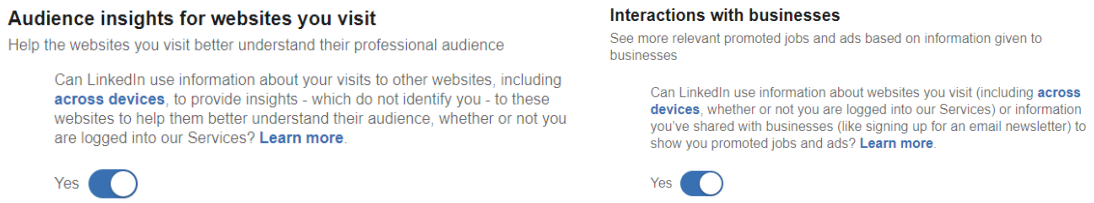

Figure. 7. LinkedIn Data Privacy Opt-Out Settings (Facebook, 2020).

One interesting take away from the examination of the opt-out options is that several of the settings are very similar in stated purpose and by failing to opt-out of all of them, you essentially open yourself to the same accessibility of data as you would if you had not opted out of any of them. LinkedIn has the most robust and customizable data sharing settings of any of the applications researched in this paper with specific pieces of data having the option of whether or not to be shared, such as Salary, Search history, Connections, and even some third-party data. Given its professional nature, and less of an entertainment social media, the ability to share and define what others see is paramount to the brand; however data about user usage of the service is still highly collected and used for advertising revenue and marketing partnerships.

MyFitnessPal

MyFitnessPal is a health and wellness application that is owned by the Athletic apparel company Under Armour. As a result of being a part of a multifaceted company not solely built around social media, the privacy policy for the application is a bit more difficult to read through. One section that could use a legal expert to interpret is the section of "Does Under Armour "Sell" my personal data?" to which the companies written response is that they do not exchange user's data for money, rather they provide sed information to other companies for advertising purposes. Basically, stating that they do give your information to third parties so they can target advertisements to users, for which Under Armour is compensated. This research is not novel and

is not the point of this paper but the notion that all these companies use terminology similar to this to justify taking user data and, in effect, selling it to third parties to use drive to the heart of what concerns end-users when the topic privacy in social media arises. As for the MyFitnessPal app, the website containing the Privacy Policy is difficult to navigate with each section of the policy linked to a different web page. However, one of the first items listed indicates that free version of the software collects and keeps data for two years. This data is only accessible to the end-user for two years, though, the company makes no claim as to how long it holds on to the data. Premium members could go back farther in their data, exposing a potential longer-term storage capability. Regarding this paper's research specifically, the policy does state that it uses GPS location and network location information. MFP also stated that they "collect or infer such data from mobile device sensors," including mobile phones or computing devices but also heart rate sensors, fitness trackers, and other interconnected devices. This seems logical given the nature of the application as a Health and Wellness app and would seem reasonable to the average person until you consider that the data collected by those devices could be provided to health insurers or medical facilities without user knowledge. Nowhere in their terms of service does it explicitly indicate that they do not share that information with those types of entities. The privacy policy links to an Under Armour listing defining the various types of companies they share information with, a listing that is 13 different vendors in length. This is the only company that so clearly lists the advertising and social media business partners. The amount of data available to them in the form of health and fitness trackers is, however, must more personal and private in nature to most uses.

Summary of Privacy Policy Review

Much of what was identified in one service privacy policy was also seen in the others. All the companies were very clear in that they provide user data to "partners" with whom they do business with. Companies who advertise using the services application to reach customers are provided the most access. Additionally, those companies who also utilize the social media services to track activity from their website back to Facebook, for example, have access to the widest range of user's data to allow for the most specific targeting marketing campaigns. Based on the research done for this paper, the social media service user data also includes any mobile device sensor data collected by the application. The applications initially prompt a user for access to common sensors such as the camera, microphone, call activity, and messaging and users do have the opportunity, through the operating systems of the mobile devices to disable access to various system resources. However, limiting access to things like the camera or microphone of an image capture social media applications like Snapchat or Instagram defeats the purpose of the application. As expected the privacy policies of these applications are clear enough to give a person the impression they are protected and that the user's data is not going anywhere except to the social medial applications company and its associates and still leave room for legal loopholes that allow the company to do what they like with the data. Based on the polices viewed in this research, end-users have some control of the information in a stored state on the service providers' servers. Table 3 shows a summary of user control in the five social media apps and Figure 8 lists whether or not a user has the ability to control or request the information that a provider has collected on the user of the use of the application.

Table 3

Summary of User Control in the Five Social Media Apps.

	Users Can Request All Data	Specify Third-Party Data Partners	Third-Party Data Limitations	Uses Can Delete Account and Data
Facebook	Yes	No	No	Yes
Instagram	Yes	No	No	Yes
Snapchat	Yes	No	No	Yes
LinkedIn	Yes	No	Yes	Yes
MyFitnessPal	Yes	No	No	Yes

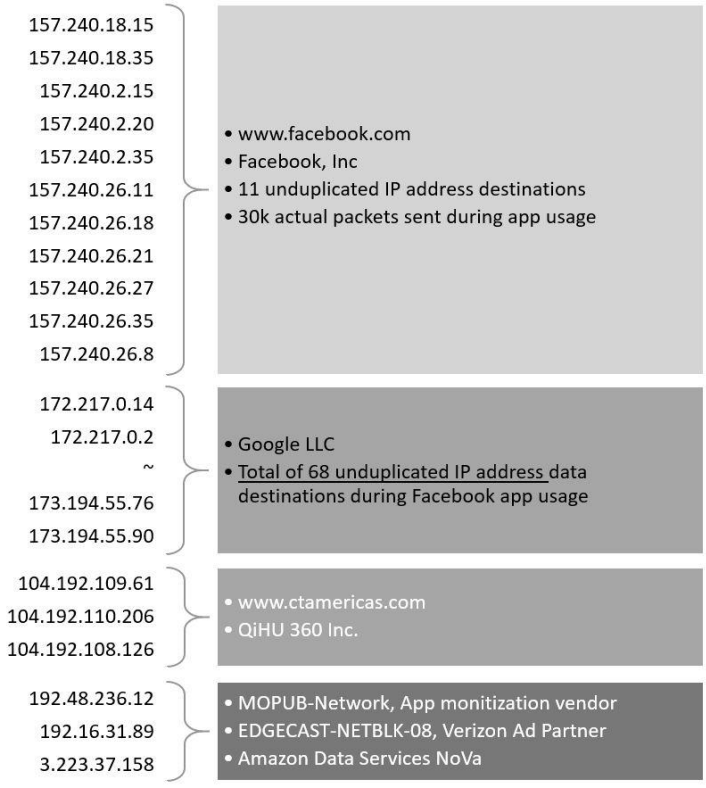

Figure 8. Example of Data Destination IP Addresses.

In all of the social media applications examined, similar terms were used to label personal data. Words like trackers, fingerprints, and footprints are all terms used to encompass personal data and identity tracking across social media platforms and advertising markets. Once

again providing evidence that the policies are intended to be easy to read across companies any yet vague enough to lead consumers to gloss over the true meaning.

Device Usage

During the testing, each device was provided a full battery charge to ensure that the device did not shut down or go into any low-power mode during the tests. The usage of the device proceeded, as indicated in the methodology portion of the paper. During the time of testing, all network transmissions were collected using a man-in-the-middle process. Packet captures were analyzed to discern the source and destination IP address of communications during use. While this shows us where the application and the device are connecting to, the encryption inherent in today's applications and secure communication protocols makes it nearly impossible to determine what the content of those messages is. Dynamic analysis of the applications used in this research provided a lot of connections to various servers; however, the widely adopted use of cloud computing resources made identifying individual advertising or third-party data destinations very difficult. While using each application, it was clear that data was going to three major IP address blocks, with each app sending additional data to multiple other sites but with much less frequency. For example, while using Facebook, IP addresses associated with Facebook Inc. were the destination of most of the communication. However, examining the addresses not associated with Facebook, we can see that a large segment of the communication was to Google and the Google cloud services hosts and Amazon servers and the AWS environment. Figure 9 does not reflect the 15 Amazon AWS IP associations but does show a truncated listing of the 68 unduplicated destinations.

During the operation of each of the social media applications, the results were similar. The bulk of the communication, as expected, was to the applications parent company. However, there were substantial transmissions from applications during times when the devices were idle. Idle time for the devices meant that no application had priority over the device, and the system was either displaying the home screen, or the screen was off. The default settings for most applications is to allow for background data transmission to allow the device to provide user notifications. During testing, this setting was configured both ways to evaluate if the number of transmissions was impacted. The results of the five applications indicated that the changing of this setting had little impact on the communication of the application. When analyzing this data, it was discovered that most of the traffic was from the device operating system. For example, on the Apple device, most idle traffic was traced to itunes.apple.com, www.icloud.com, and cl5.apple.com. By filtering out those records, we can see that the devices were still busy sending data while idle, as Figure 9 shows.

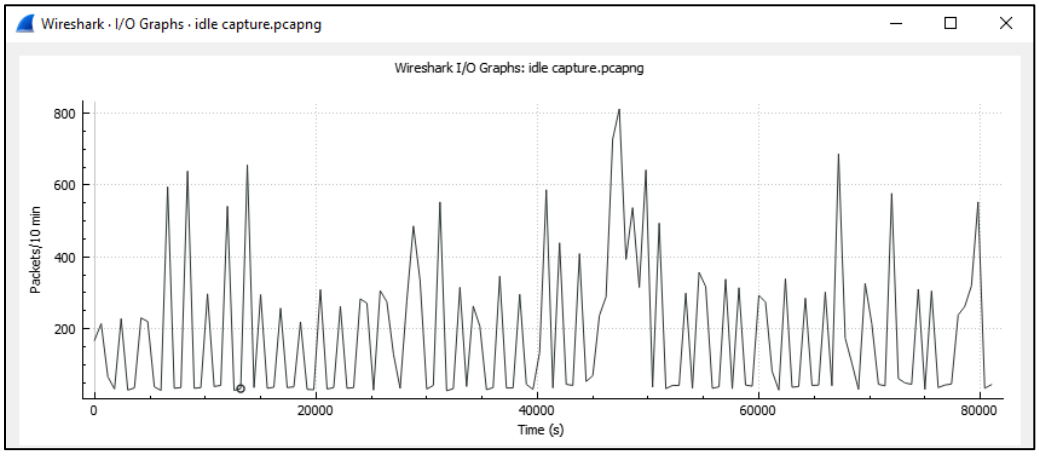

Figure 9. Idle System Application Transmission Levels.

Sensor Data

Our devices are not just palm-sized computers displaying information in a static environment. Modern mobile devices contain several sensors that provide a wide variety of information about the environment and use of the device to provide detailed interaction with applications and services (Mehrnezhad & Toreini, 2019). Each of our devices tested contained sensors that provided data about everything from ambient light altitude, orientation, motion as well as audio and video inputs. Except for location (GPS) and audio/video sensors, most users are not keenly aware of the role, and impact sensors have on their interaction with modern mobile devices. To discern if applications were collecting and using this data, the two devices were connected to their respective software development kits to monitor changes in sensor status while using the applications. According to Android Developers, there are 13 different sensors that a device may have. The type and purpose of each sensor can be viewed in Table 5 in Appendix A of this paper. Reviewing the technical specifications of the two devices in this research, we see that the Samsung device has more sensors than the Apple device does, as indicated in Table 4 below.

Table 4

Sensors by Device

Samsung S8	Apple iPhone 8
Accelerometer	Touch ID Fingerprint Sensor
Barometer	Barometer
Fingerprint Scanner	Three-Axiz Gyro
Iris scanner	Accelerometer
Gyroscope	Ambient Light Sensor
Hall Sensor	
Proximity Sensor	

Since the privacy policies of all the social media applications admit to collecting device data, it was decided to focus on the frequency of sensor access the applications were using and try to determine what sensor data is being transmitted. Unfortunately, without the ability to decompile the applications, it is impossible to know exactly what and when the app accesses sensor data. As a result, a kind of workaround was set up to view both sensor logs from the device as well as capture network transmissions. The Samsung device was connected to the computer using a USB-C to USB-A cable and attached to the Android SDK software. Device Logging was enabled and monitored to see when sensors were actively reporting data. Next, Wireshark was started to collect all network traffic from the mobile device, and the social media application was started.

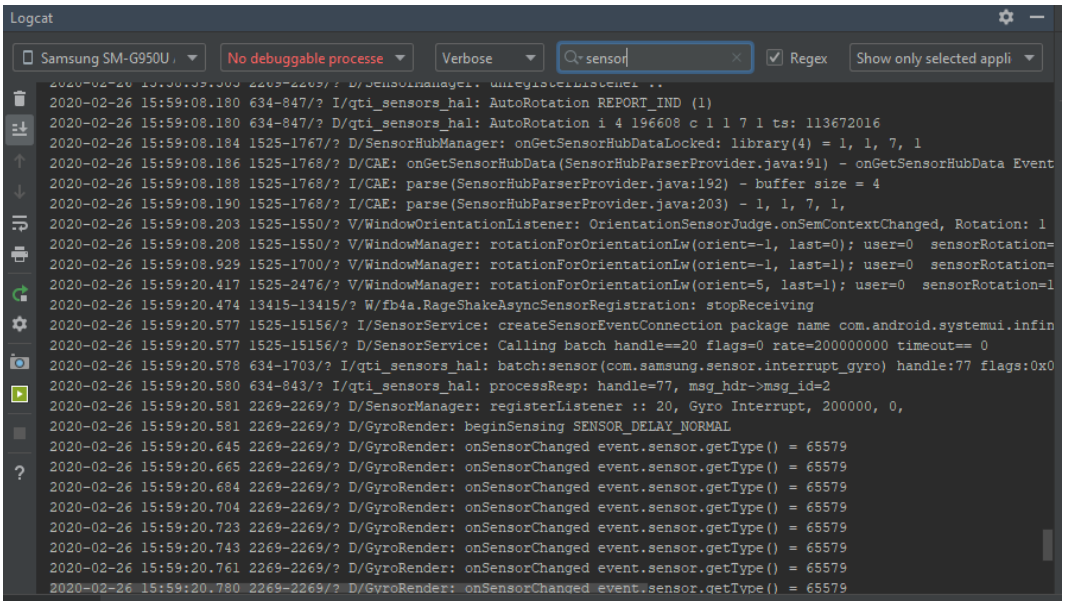

Figure 10. Android Log Displaying Sensor Activity

Additional research indicates that even though each sensor has a specific purpose, those sensors have the ability to do more than users or even developers intend. One such example is the patent for a system that allows the accelerometer to detect a user's voice activity. These kinds

of developments pave the way for companies to obscure what type of access they have to user's information. Researches developed an application capable of using the gyroscope sensors as a crude microphone allowing for the application to pick up conversations (Greenberg, 2014). However, this paper's research is not about the multitude of threats from third-part applications but the common and more trusted social medial applications. Data collected in the research of Question 3 provides a correlation of sensor readings and application data transmissions. Each time an application was started, a sensor reading of phone orientation was captured. This is a reasonable step to orient the display in a manner consistent with how the user is holding the device. Other tests included using the device to post an image, checking into a location, and taking a "selfie." From examining the two sources of information, it was clear that the application used the sensor data in a manner consistent with the operation of the software. Figure 11 attempts to show the time and process correlation of the log data next to the network data.

Android SDK Log Data

```
22020-02-10 15:36:2.283 959-21671/? I/mm-camera: <SENSOR>< INFO> 3735: companion_process: is_supported = 0x001F

22020-02-10 15:36:21.386 959-21670/? E/mm-camera: <SENSOR><ERROR> 497: module_sensor_offload_init_config: Success

22020-02-10 15:36:21.583 959-21668/? I/mm-camera: <SENSOR>< INFO> 282: port_sensor_handle_stream_on: H/W revision = 12(12),
Criterion ver = 12

2020-02-10 15:36:22.110 959-21668/? I/mm-camera: <SENSOR>< INFO> 284: port_sensor_handle_stream_on: Reference Value :
CAMERA_CAL_CRC = 0x1FF, COMPANION_CAL_CRC = 0x1F, CAMERA_CAL_CRC_FRONT = 0xFF

2020-02-10 15:36:22.113 3474-3474/? I/LauncherAppWidgetHostView: updateLastOrientation, orientation: 1, activityOrientation:
1, fromRemoteViews: true, widget:
ComponentInfo{com.google.android.googlequicksearchbox/com.google.android.googlequicksearchbox.SearchWidgetProvider}

2020-02-10 15:36:22.130 3474-3573/? D/Notification.Badge: LauncherModel:onNotificationFullRefresh() not need to update,
updatedBadges.remove[com.facebook.katana], count : [9

2020-02-10 15:36:23.160 2269-2269/? D/GyroRender: onSensorChanged event.sensor.getType() = 65579

2020-02-10 15:36:35.517 3474-3573/? D/Notification.Badge: [filtered] shouldBeFilteredOut().isGroupHeader[true],
missingTitleAndText[true], [sbn : StatusBarNotification(pkg=com.facebook.katana user=UserHandle{0} id=2147483647
tag=ranker_group key=0|com.facebook.katana|2147483647|ranker_group|10208|ranker_group:

2020-02-10 15:36:35.530 959-18222/? I/ShotBeauty: CAMERA_CMD_STOP_FACE_DETECTION : 0

2020-02-10 15:36:35.530 1525-11861/? D/ConnectivityService: filterNetworkStateForUid() uid: 10083 networkInfo: [type: WIFI[]
WIFI, state: CONNECTED/CONNECTED, reason: (unspecified), extra: "Whatever1", failover: false, available: true, roaming:
false, metered: false]
```

Wireshark Data

No.	Time	Source	Destination	Protocol	Length	Info
5688	3:36:35 PM	172.16.0.216	157.240.26.27	TCP	54	52941 → 443 [ACK] Seq=20017 Ack=2573889 Win=1055488 Len=0
5691	3:36:35 PM	172.16.0.216	157.240.26.27	TCP	54	52941 → 443 [ACK] Seq=20017 Ack=2576689 Win=1055488 Len=0
5702	3:36:35 PM	172.16.0.216	157.240.26.27	TCP	54	52941 → 443 [ACK] Seq=20017 Ack=2590689 Win=1055488 Len=0
5704	3:36:35 PM	172.16.0.216	157.240.26.27	TCP	54	52941 → 443 [ACK] Seq=20017 Ack=2592089 Win=1055488 Len=0
5727	3:36:35 PM	172.16.0.216	157.240.26.27	TCP	54	52941 → 443 [ACK] Seq=20017 Ack=2614019 Win=1055488 Len=0

Figure 11. Data Correlation Between Device Log and Network Traffic

That is not to say there are not applications out there taking advantage of this hidden data. There are numerous reports and research indicating that creative developers and less scrupulous companies can potentially take advantage of sensor data and device permissions to collect additional information about their users.

One additional interesting observation was, while the iPhone was connected to the research network, the connected clients were being monitored to ensure that no unwanted devices appeared in the stream of data. Even though it had not been explicitly allowed to connect to the research network, an Apple watch associated to the iPhone via Bluetooth automatically connected and appeared in the connected clients list. The act of a device joining a wireless access point simply because another device it is paired to is a dangerous practice. This is another example of the devices do things that the typical end-users are completely unaware of and could lead to an erosion of user security.

Storage Utilization

Mobile devices are designed to store and transmit data. Each device handles the storage of user data differently and has different storage options. For example, a device typically has non-removable internal storage. A device might also have a removable SD/MicroSD card to increase storage capacity. Research Question 3 asks if the data from typical usage by an end-user is removed during the processes involved in a "factory reset." Neither of the devices used for this research contained additional storage in the form of an SD card, so all testing was done on the internal storage medium of each device. Both devices report having a 64 GB internal drive. After the initial factory reset of each device and before any non-system or non-service provider applications were installed, the devices operating system reported that approximately

15.4 GB of storage space was used. Examination of the file hierarchy indicated that there were 6,612 directories and over 567,411 files on the "clean" Samsung device. These counts are all taken before the operating system of the device could update the software or any pre-installed applications. During testing, there were five applications installed and operated in a standard user expected fashion. Social media sites were visited, postings were made and viewed, pictures were taken and uploaded. Upon completion of the research into application permissions and sensor data collection, the devices were once again connected to the forensics software to create another disk image. Usage of the five social media applications created an additional 516MB of files on the Samsung device's internal storage and 473MB of data on the iOS device. The examination of the images produced little of interest. Files discovered include cached files from social media usage and other data that would be useful if we were performing an investigation of user activities but not enough to examine post "factory reset."

Both devices were once again reset using the settings within the operating system. A final, post-wipe image was captured, and attempts were made to discover any remaining data. As expected and previous research indicates, the file structure is destroyed and rebuilt. Since all of the data on the device is encrypted, the reset destroys the keys associated with the encryption process, so even if we were able to identify data at the bit level, the encryption process and destruction of the keys has made it impossible to access the data. A quote from Apple's support pages states, "The "Erase all content and settings" option in Settings obliterates all the keys in Effaceable Storage, rendering all user data on the device cryptographically inaccessible" (Apple Support, 2020).

End-user Survey

The final part of this paper's research was to conduct an end-user survey of perceptions of privacy both before and after exposure to how the social media applications access, process, and utilize personal data. The initial survey was given to 46 college students in the Central Minnesota area to evaluate the students' opinions on privacy and to measure the change in responses after providing additional information on the subject matter. The survey was based on five observations of the research. The first being that even when idle, the device continued to transmit data not related to the operation of the applications. Second, companies now use tracker IDs as a way to obscure users' personal identity from third parties, but the use of these trackers across devices and applications allows for highly detailed profiles to be created, making the identification of users rather simple. Third, privacy policies are written as to appear simple and straight forward; however, they allow of substantial leeway and loopholes for companies to maintain control of user's data. Fourth, most of the data transmitted while actively using a social media application go to servers hosted by that company. There are still significant packets that travel to third-party advertising affiliate sites that collect data in association with the application vendor. Lastly, permissions requested by applications to user data provide much more access to personal information than users have been led to believe. Survey takers were presented with ten questions, and the results were evaluated to determine the level of user concern regarding their personal data, how it is accessed, and the access of social media applications. Questions were posed to measure the user's awareness of the corporate use of their data and assess their acceptance of that usage. Based on the results, the research also intended to discover opinions on the tradeoff of convenience or entertainment in exchange for access to their personal data.

Table 5 (see Appendix A) shows the results of the survey asking which social media applications used in this paper's research were used. 100% of the users used the Facebook application. 64.52% of respondents also used Instagram, which is owned by Facebook and falls largely under the same privacy policies. Survey takers indicated that 41.94% used MyFitnessPal, which was interesting in the fact that the UnderArmour corporation is the most open to sharing and selling of user data and that fitness tracking applications collect the most personal health information of all the applications.

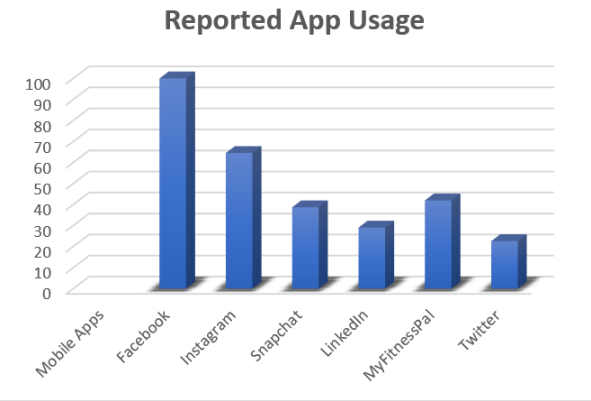

Figure 12. Survey Results: App Usage

When asked if users felt that the privacy policies provided adequate protection against a company using personal data against their wishes, 19.35% of respondents indicated "yes." The overwhelming balance is indicating that they did not believe that the privacy policies protected them. As our research indicated earlier, there are virtually no federal regulations on safeguarding consumer digital privacy in the United States. If this research had been conducted in some other countries, we might see different results given the implementation of GDRP (General Data Protection Regulation) enacted in May of 2018. "The GDPR puts digital consent, privacy, and control front and center" (Stewart, 2018). When asked to rate several statements about user

privacy and protections, 38.71% of users responded that they felt “neutral” about whether social media companies were concerned about user data privacy. An equal amount fell into the Disagree to Strongly Disagree range, with only 22.58% felt that the companies were interested in consumer protections. Interestingly, 29.04% of survey takers also responded that they did not have any data they felt required protections and so were comfortable with the social media applications using their data. Observations on this dichotomy will be discussed later in this document.

When asked if a company should be financially penalized for data breaches or data misuse of user data, a majority (83.87%) indicated they strongly agreed. The sentiment indicating that even though many users do not feel that their data is really in need of protection or important enough to be concerned with how a company might use it, they still feel that a company should be penalized if that data is accessed unlawfully.

Users were asked to share what data they felt was acceptable for a social media application to share with its partners and advertisers. Many uses opted to indicate that none of their data is acceptable; however, the majority indicated that any information a user enters into their profile if the fair game for the company to utilize. Figure 13 shows the breakdown with only one user, indicating that just about any data is acceptable. If we were to drop this user from the survey, a clear line emerges that anything within the application is suitable, but other files such as files saved on their device or contacts not part of the application should be more protected from access.

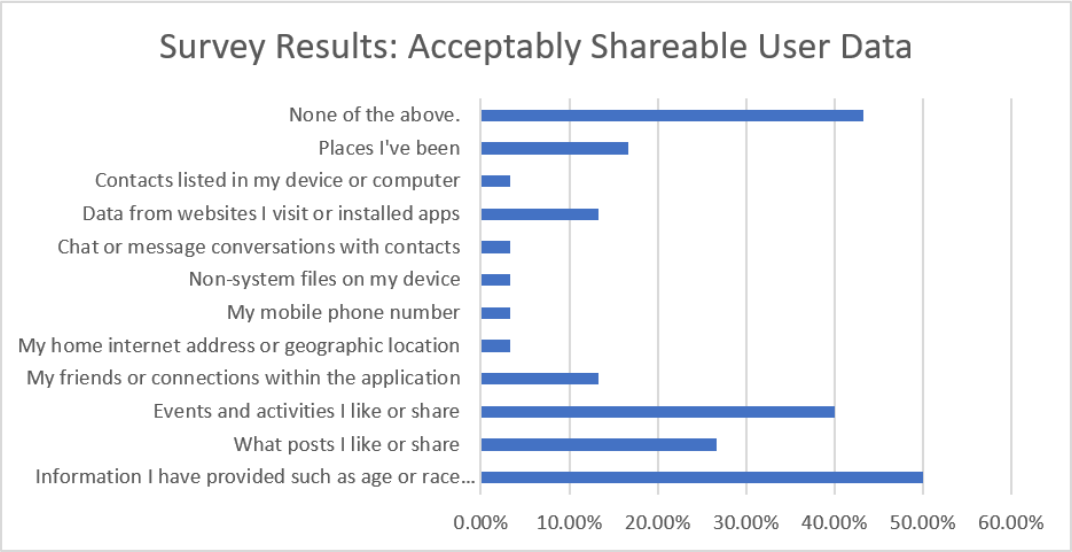

Figure 13: Survey Results: Acceptably Shareable User Data

Currently, most of the data that is collected by social media is provided through the standard use of the applications. However, as experienced in our testing, new users are prompted several times during the installation and setup of user accounts to allow the application to access user contacts stored in the device in order to make connections with other users of the application/service who might already have an account. Other access requests include access to stored images to allow a user to upload a profile picture. However, granting access to the stored images allows for much greater access to data than the user is typically aware of. Most service providers view this granting of permission as an Opt-In of the user agreeing to allow ongoing access to data stored on the device. This survey asked the users if there should be more clear opt-in procedures to protect user data, and 87% responded "yes."

Lastly, after explaining what data companies have access to and how users are tracked across applications and devices, users were asked one last question. If all the data listed in the question about sharing acceptability was collected by the application's company, "Would you

still use it?" Even with the knowledge that the application collects all the data listed, including information not used in the application, 58.6% of users still said they would use the application.

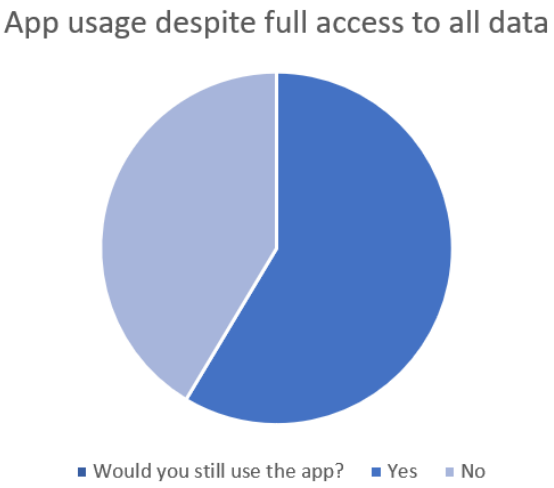

Figure 14. Apps Usage with Full Access to User Data

Observations

The research began by evaluating the public privacy policies of each of the applications. While all the policies were easy to read and described in very straightforward terms what, and how they used user data, they were very vague in specific access methodologies. Almost all the applications list "business partners" as those who use the service to target and reach a specific audience. Individual companies do not have to include tracking data collection outside of the primary application. In other words, companies do not have to have their own servers to collect data across the internet. It is all provided by the host application. User interactions with online advertisements still initiate additional data on the network, but ad viewer count and exposures are still captured within the applications. While the data is not "sold," it is provided to companies who "partner" with the host application. A partnership that costs money to maintain. The use of the term "sold" is clearly defined differently between end-users and service providers.

During the collection of network traffic, devices sat idle for 24 hours between usage. This allowed us to capture transmissions from applications even when the user was not operating the device. Much of the data collected was related to the operation of the devices operating system; however, regular transmissions to vendor-specific domains indicated that the application was pinging the servers. Limitations to accessing the payload of the images due to encryption prevented deeper analysis, but the fact that even when idle the application is reaching out to the servers indicates that the device is not as inactive as a user might believe.

Kerry (2018) made an astute observation about the current state of digital privacy, and the challenges of keeping track of and control of the vast amounts of information users are producing. “It’s a losing game both for individuals and for our legal system” (Kerry, 2018). During traffic analysis, about three dozen non-service-related IP ranges were identified as receiving or sending traffic while using the various applications. The use of tracker files and profile analysis by social media companies allow for much of the user data transactions to take place within the data centers run by the various companies. Additionally, the rise of cloud computing for many companies makes identification of various domains much more challenging as blocks of IP address can be assigned to a company, but the global WHOIS databases only register the host, which is often Amazon (AWS) or Google Cloud services. Because of the massive amount of data created and accessed very minute of every day, it’s easy to see how much of a challenge state and federal governments will have if data privacy becomes a greater issue in the future.

The final observation during this research was the extent of access that applications request when they are first installed and how little the average user understands about those

permissions. As stated earlier in the paper, when providing app access to your photos for the sake of posting images to a social media site, you are allowing that vendor full access to all of the data each photo. This includes the EXFIL data, such as location the picture was taken and when. And while this information is stripped out by the company before the image is posted, there is no policy or law that dictates what a company may or may not do with that data. Sensor data collection is limited only to what the application programmers want to collect. Simple gyroscope sensors allow the app to determine if it should display in landscape or portrait mode, but accelerometers can measure speed, estimate mode of travel, and, along with GPS sensors, provide a precise app location and travel details. All of which users are readily providing insurance companies in exchange for potentially lower rates or rebates (Shilton & Green, 2019).

To measure the knowledge and comfort level of users in regard to their privacy online, the survey was given to 32 college students as a graded assignment. The small sampling size and focused group of students used to collect the information made it clear that IT students have a keener view of privacy than many end-users. The students are taking the survey information technology students, some of which were specializing in cybersecurity. Even with the knowledge of how data is collected and used, the majority still stated they would continue to use social media applications. In conversations with some of the students, the question of “why still use the app?” was asked. The responses indicated that they didn’t mind the targeted apps, or they used ad blocking software, so they really didn’t see the effects of the data collection. This thinking provides another glimpse into the average user as we are overwhelmed by advertising is all of its various forms to the point that targeted ads just provide us with things that might be more interesting than general ads for things we might never use.

Summary

This chapter covered the majority of our research, including live app and idle system data transmission collection, device sensor usage by the applications, and transmissions to servers other than those operated by the social media service providers. Thanks to the continued development of encryption standards and whole device encryption being more commonplace in mobile devices, access to user data post-factory resets is virtually impossible in more modern devices. Survey results show us that even though users are uncomfortable with the level of access applications have to their data; they are still willing to freely exchange that data for convenience or reward.

Chapter V: Results, Conclusion, and Future Work

Introduction

Over the course of this research, several different areas of privacy have been examined, from the policies and promises of social media companies to the data transmissions from devices to servers. At every step of the research, there have been subtle deviations from the expected operation of social media applications. Little pieces of information that lead to an overall conclusion. This section will review the results of the data collection, the system resets, and survey results. The conclusions of which will provide insight into what is the current status of digital privacy and digital privacy literacy and what recommendations can be made for future users.

Results

This research was intended to provide some insight into the type and frequency of data collected by social media applications during the regular and expected use of the software and to test whether or not data created by end-users would be persistent on the device even after the use of a factory reset typically used to clear a device back to an "out of box" state. Based on those observations, a survey was created to measure what end-users' positions were in regard to privacy prior to, and then after, being enlightened to the results of the research. During the research, four questions were asked to discover if our data is being collected or used outside of the bounds laid out in user agreements and privacy policies.

The first research question was whether social media applications are collecting data unrelated to the functionality of the app? To answer this question, new devices were deployed with no prior user data, and new accounts were created. Each tested application was run on both

an Android and Apple device, and each application could run as the only installed application. Data captured indicated that the majority of the data being sent and received by the device was related to the application; however there were enough packets sent to domains outside of the application vendor to question what the third party might be collecting additional information. Attempts were made to examine packet payloads, but encryption obfuscated the data, so only source and destination information was accessible. Attempted man-in-the-middle access resulted in most applications refusing to connect to their hosts. Despite the limitations of deep packet inspection, it was noticed that data, unrelated to the user's operation of the application, did occur as regular data transmissions were sent while the device was idle. While this idle data transfer is likely tied to end-user notification capabilities in the application, end-users awareness of such exchanges is rather limited.

The second research question is bound to the first question. What data is transmitted to and from a device once the application is installed? The research into the payloads of data transmissions was limited by current encryption standards, but other information could be correlated to produce some results. In addition to the data collection, sensor data was observed using software APKs while using the applications. Logs of the data transmissions coordinated with the packet captures indicate that sensor data is transmitted to the application vendor. As indicated in the existing literature, sensor data usage by applications has been known to be an issue for years; it is still an unresolved problem given the users' desire for ease of use and a willingness to allow it despite the security implications.

According to the privacy policies of the five applications tested in this research, only one failed to outright state that they do not sell user data. The four that made such claims protected

themselves by stating that user data is only provided to “business partners” to provide better or more focused user interactions. It’s this kind of legal speak that lulls average users into a false sense of security about what they share online. While the wording is mundane, it leaves much room for interpretation. As this paper’s research showed, providing access to the images stored on a user’s device gives the companies access to so much more information. Facebook does not have any specific details about the EXIF information it collects, and while the data is removed when an image is posted, the data is still collected by Facebook and used in unknown ways (James, 2011). Once again, end-users do not receive a full understanding of what providing access to certain device sensors or software directories really means and, as a result, open themselves to all matter of personal privacy abuses.

The third research question of; “Does performing a factory reset on a device remove personal data artifacts?” For this part of the research, the two devices that were used for the previous data usage testing were both imaged using Oxygen Forensic Examiner software. Before testing and creating user profiles for any of the applications, a forensic image was created of the device. After testing, the devices were images a second time. An examination of the social media applications provided some interesting data in the forensic software, but data stored on the device still used by an end-user was outside of the scope of this research. A factory reset of each device was then performed, and the data in the images was examined. On the Android phone, there were several image files created by using the device’s camera to upload to Facebook, and Instagram were still recoverable. A couple of system logs retained some data post-reset as well. However, with the Apple device, the images were compared, and only system files were

recognizably recoverable. Apple's tight control of the operating system and app environment allow for cleaner system wipes.

The final question: "What is the perception from the end user's perspective of the exchanging personal information, and thus sacrificing digital privacy, in exchange for entertainment/social interaction?" the survey polled 40 college students ranging between 20 and 30 years of age. The results of which indicate that even though they know that data provided while using the applications is provided to an unknown number of business affiliates, they would still use the app. In addition, the lack of knowledge of how much access permissions really give an app, such as access to images, leaves users vulnerable to privacy abuses. The lack of transparency in the overall reach of privileges and a lack of user knowledge in what they agree to when creating an account with a social media app is what allows companies to abscond with user data unquestioned. Unfortunately, even when users are aware of the misdeeds, they refuse to give up their social media applications. The convenience and entertainment value is more important than privacy for the vast majority of people.

Conclusion

In conclusion, this research has provided a look into the flow of information from mobile devices to the application servers in such a way as to determine if our data is true to the usage intended by installing and running social media applications. Through the use of packet captured data in a controlled environment, the research shows that not all of the data being sent is dose so through user interactions. Idle communication is taking place on a regular basis providing the applications and the device operating system additional information about the user. The use of factory reset on a device has mixed results with a more controlled environment such as Apple's

iOS providing better end-user protections when wiping their device. Lastly, user understanding of the data processed by today's mobile devices is insufficient to effectively protect users from manipulation by companies with access to the vast amount of data. Modern profile tracking technology far exceeds users' concern for privacy, especially when it comes in conflict with their desire for entertainment.

Future Work

Digital privacy, especially regarding mobile devices, personal freedoms, and government overreach/censorship, is a major concern in the world today. As more people become untethered by desktop or even laptop computers, the importance of having privacy and control of personal data will be ever-present. Demands for transparency and accountability will continue to expand as users slowly become more aware of the methods and processes used by social media companies to make a profit from sharing user information. The abuse of information, such as that of Cambridge Analytica, is just the beginning of what will happen if users continue to hide their heads in the sand for fear of losing some level of convenience or entertainment.

The United States continues to drag their feet when dealing with personal data protection. Some states are making progress, like the European Union, but the corporations have virtually unfettered access and use of our information. Future research on the processes used to create and track users across sites, devices, and companies might open opportunities for user education. More educated users will produce more secure users, especially if they are provided more granular control over their digital environment and the data that they share. The idea of "virtual walls" where users have more control over their digital footprint has a potentially positive impact on user privacy.

References

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442-492. doi.org/10.1257/jel.54.2.442

Amer, K. (Director), & Noujaim, J. (Director). (2019). *The great hack* [Motion picture, Documentary]. California: Netflix. Retrieved from https://www.thegreathack.com/

Apple Support. (2020). Encryption and data protection overview. Retrieved from https://support.apple.com/guide/security/encryption-and-data-protection-overview-sece3bee0835/1/web/1

Arapinis, M., Mancini, L., Ritter, E., & Ryan, M. (2017). Analysis of privacy in mobile telephony systems. *International Journal of Information Security*, *16*(5), 491-523. doi.org/10.1007/s10207-016-0338-9

California Privacy Disclosure|LinkedIn. (2020). *California consumer privacy act*. Retrieved from https://www.linkedin.com/legal/california-privacy-disclosure

Choe, E. K., Jung, J., Lee, B., & Fisher, K. (2013). Nudging people away from privacy-invasive mobile apps through visual framing. In P. Kotze, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler (Eds.), *Human-Computer Interaction—INTERACT 2013* (pp. 74-91). New York, NY: Springer. https://doi.org/10.1007/978-3-642-40477-1_5

Do, Q., Martini, B., & Choo, K.-K. R. (2014). Enhancing user privacy on android mobile devices via permissions removal. *2014 47th Hawaii International Conference on System Sciences*, 5070-5079. doi.org/10.1109/HICSS.2014.623

Facebook. (2020). *Data policy*. Retrieved from https://www.facebook.com/policy.php

Fauerbach, T. (2017, December 21). Data reigns in today's data economy. *The Northridge Group*. Retrieved from https://www.northridgegroup.com/blog/more-valuable-than-oil-data-reigns-in-todays-data-economy/

FourWeekMBA. (2019). *How does Facebook make money? Facebook business model in a nutshell*. Retrieved from https://fourweekmba.com/how-does-facebook-make-money/

Gallagher, S. (2014, August 4). Opposite of OPSEC: Russian soldier posts selfies—from inside Ukraine. *Ars Technica*. Retrieved from https://arstechnica.com/tech-policy/2014/08/opposite-of-opsec-russian-soldier-posts-selfies-from-inside-ukraine/

Greenberg, A. (2014, August 14). The gyroscopes in your phone could let apps eavesdrop on conversations. *Wired*. Retrieved from https://www.wired.com/2014/08/gyroscope-listening-hack/

Handley, L. (2019, January 24). *Nearly three quarters of the world will use just their smartphones to access the internet by 2025.* Retrieved from *2025*. https://www.cnbc.com/2019/01/24/smartphones-72percent-of-people-will-use-only-mobile-for-internet-by-2025.html

Instagram. (2017). *Privacy policy.* Retrieved from https://help.instagram.com/402411646841720

James, J. (2011, December 11). How Facebook handles image EXIF data. *ITProToday*. Retrieved from https://www.itprotoday.com/strategy/how-facebook-handles-image-exif-data

John, A. S. (2018). How Facebook tracks you, even when you're not on Facebook. *Consumer Reports*. Retrieved from https://www.consumerreports.org/privacy/how-facebook-tracks-you-even-when-youre-not-on-facebook/

Keng, J. C. J. (2016). Automated testing and notification of mobile app privacy leak-cause behaviours. *2016 31st IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 880-883.

Kerry, C. F. (2018, July 12). Why protecting privacy is a losing game today—And how to change the game. *Brookings*. Retrieved from https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/

Martin, K., & Shilton, K. (2016). Why experience matters to privacy: How context-based experience moderates consumer privacy expectations for mobile applications. *Journal of the Association for Information Science and Technology*, *67*(8), 1871-1882. doi.org/10.1002/asi.23500

Mcdonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *Journal of Law and Policy for the Information Society*, *4*, 543-568.

McFarlane, G. (2019, October 19). How Facebook, Twitter, Social Media make money from you. *Investopedia.* Retrieved from https://www.investopedia.com/stock-analysis/032114/how-facebook-twitter-social-media-make-money-you-twtr-lnkd-fb-goog.aspx

Mehrnezhad, M., & Toreini, E. (2019). What is this sensor and does this app need access to it? *Informatics*, *6*(1), 7. Retrieved from https://doi.org/10.3390/informatics6010007

Pic2Map. (2020). *Pic2Map photo retriever*. Retrieved from https://www.pic2map.com

Polykalas, S. E., Prezerakos, G. N., Chrysidou, F. D., & Pylarinou, E. D. (2017). Mobile apps and data privacy: When the service is free, the product is your data. *2017 8th International Conference on Information, Intelligence, Systems Applications (IISA)*, 1-5. Retrieved from https://doi.org/10.1109/IISA.2017.8316392

Rajasegaran, J., Karunanayake, N., Gunathillake, A., Seneviratne, S., & Jourjon, G. (2019). A Multi-modal neural embeddings approach for detecting mobile counterfeit apps. *The World Wide Web Conference*, 3165-3171. Retrieved from ttps://doi.org/10.1145/3308558.3313427

Rogan, J. (2019, October 24). *Joe Rogan Experience #1368—Edward Snowden.* [Video file]. Retrieved from https://www.youtube.com/watch?v=efs3QRr8LWw

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research, 17*(7). Retrieved from https://doi.org/10.2196/jmir.4273

Sarabia-Sánchez, F.-J., Aguado, J.-M., & Martínez-Martínez, I. J. (2019). Privacy paradox in the mobile environment: The influence of the emotions. *El Profesional de La Información, 28*(2), 111. doi.org/10.3145/epi.2019.mar.12

Shilton, K., & Greene, D. (2019). Linking platforms, practices, and developer ethics: Levers for privacy discourse in mobile application development. *Journal of Business Ethics*, *155*(1), 131- 146. doi.org/10.1007/s10551-017-3504-8

Slavin, R., Wang, X., Hosseini, M. B., Hester, J., Krishnan, R., Bhatia, J., Breaux, T. D., & Niu, J. (2016). PVDetector: A detector of privacy-policy violations for Android apps. *2016 IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, 299-300. https://doi.org/10.1109/MobileSoft.2016.069

Sleeper, M., Acquisti, A., Cranor, L. F., Kelley, P. G., Munson, S. A., & Sadeh, N. (2015). I would like to..., I shouldn't..., I wish I...: Exploring behavior-change goals for social networking sites. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1058-1069. https://doi.org/10.1145/2675133.2675193

Snap Inc. (2019). *Privacy policy*. Retrieved from https://www.snap.com/en-US/privacy/privacy-policy

Stewart, E. (2018, April 5). *Why you're getting so many emails about privacy policies*. Vox. https://www.vox.com/policy-and-politics/2018/4/5/17199754/what-is-gdpr-europe-data-privacy-facebook

TermsFeed. (2020). *Privacy policy URL for Facebook app*. Retrieved from https://www.termsfeed.com/blog/privacy-policy-url-facebook-app/

United States Security and Exchange Commission. (2012). *FB-12.31.2012-10K*. Retrieved from https://www.sec.gov/Archives/edgar/data/1326801/000132680113000003/fb-12312012x10k.htm

Wu, F., Chen, C., & Clarke, D. (2014). Sensitive data protection on mobile devices. *International Journal of Advanced Computer Science and Applications, 5*(9), 38-41.

Zang, J., Dummit, K., Graves, J., Lisker, P., & Sweeney, L. (2015). Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps. *Technology Science*. Retrieved from https://techscience.org/a/2015103001

Appendix A: Android Sensor Listing

Sensor	Type	Description	Common Uses
TYPE_ACCELEROMETER	Hardware	Measures the acceleration force in m/s2 that is applied to a device on all three physical axes (x, y, and z), including the force of gravity.	Motion detection (shake, tilt, etc.).
TYPE_AMBIENT_TEMPERATURE	Hardware	Measures the ambient room temperature in degrees Celsius (°C). See note below.	Monitoring air temperatures.
TYPE_GRAVITY	Software or Hardware	Measures the force of gravity in m/s2 that is applied to a device on all three physical axes (x, y, z).	Motion detection (shake, tilt, etc.).
TYPE_GYROSCOPE	Hardware	Measures a device's rate of rotation in rad/s around each of the three physical axes (x, y, and z).	Rotation detection (spin, turn, etc.).
TYPE_LIGHT	Hardware	Measures the ambient light level (illumination) in lx.	Controlling screen brightness.
TYPE_LINEAR_ACCELERATION	Software or Hardware	Measures the acceleration force in m/s2 that is applied to a device on all three physical axes (x, y, and z), excluding the force of gravity.	Monitoring acceleration along a single axis.
TYPE_MAGNETIC_FIELD	Hardware	Measures the ambient geomagnetic field for all three physical axes (x, y, z) in μT.	Creating a compass.
TYPE_ORIENTATION	Software	Measures degrees of rotation that a device makes around all three physical axes (x, y, z). As of API level 3 you can obtain the inclination matrix and rotation matrix for a device by using the gravity sensor and the geomagnetic field sensor in conjunction with the getRotationMatrix() method.	Determining device position.
TYPE_PRESSURE	Hardware	Measures the ambient air pressure in hPa or mbar.	Monitoring air pressure changes.
TYPE_PROXIMITY	Hardware	Measures the proximity of an object in cm relative to the view screen of a device. This sensor is typically used to determine whether a handset is being held up to a person's ear.	Phone position during a call.
TYPE_RELATIVE_HUMIDITY	Hardware	Measures the relative ambient humidity in percent (%).	Monitoring dewpoint, absolute, and relative humidity.
TYPE_ROTATION_VECTOR	Software or Hardware	Measures the orientation of a device by providing the three elements of the device's rotation vector.	Motion detection and rotation detection.
TYPE_TEMPERATURE	Hardware	Measures the temperature of the device in degrees Celsius (°C). This sensor implementation varies across devices and this sensor was replaced with the TYPE_AMBIENT_TEMPERATURE sensor in API Level 14	Monitors device temperature.

Appendix B: Research Survey Questions

Data Privacy and Personal Information Survey

PAGE 1:

Social media apps provide us with many ways to interact with each other. They also collect a lot of data about what we like, who we interact with and where we go. The following questions will measure your comfort with the current way apps use our data.

Q1: Which of the following mobile apps do you use? Check any or all that you use.

- Facebook
- Instagram
- Snapchat
- LinkedIn
- MyFitnessPal/Other health or fitness app
- Twitter

Q2: Do you believe the privacy policy that companies provide protect your personal information from unwanted collection by other companies?

- Yes
- No

Q3: Do you feel that social media apps provide adequate ways for you to control access to your data?

- Yes
- No

Q4: Do you believe that access to your information is a fair exchange for a free service? (i.e. Facebook or Instagram)

- Yes
- No

Q5: Please rate the following statements:
Scale: Strongly Agree……Agree……Neutral……Disagree……Strongly Disagree

- I believe social media apps are concerned about my privacy.
- I have nothing to hide so it does not matter if they use my data.
- I would exchange my information for a discount on goods or services.
- A company that misuses or loses my data to a breach should be financially penalized.

PAGE 2:

Data Privacy Survey - Part 2

It has been proven that social media companies track users across devices and even applications. Your mobile devices collect location information in various ways and your images all have location information stored with them. By allowing apps to access photos you provide them with much more information than intended. Privacy policies are written to be easy to read yet leave massive loopholes for data collection and resale.

“We never sell your data” means that the company does not sell data directly but does provide all data to their “partners.” Partners is merely a term for other companies that pay to have access to data collected by the app. Applications track every “like”, post, friend/relationship, “check-in” and message. Others collect sensor information directly from your device including location, recent contacts, acceleration, even camera and microphone status.

Based on this information please answer the questions below.

Q6: Privacy policies of the top five apps state they do not sell users data. Do you believe that using the term “Business Partner” is a deceptive way to try to ease end users worries about profiting from user data?

- Yes
- No

Q7: When allowing apps to access your images (required to post images on sites) you are allowing access to the location data included in those images. Are companies taking advantage of users who are unaware of technical capabilities?

- Yes
- No

Q8: What data do you believe is acceptable for an app to use for profit? Check those you would allow a company to sell.

- Information I have provided such as age or race (user profile information)
- What posts I like or share
- Events and activities I like or share
- My friends or connections within the application
- My home internet address or geographic location
- My mobile phone number
- Non-system files on my device
- Chat or message conversations with contacts
- Data from websites I visit or installed apps
- Contacts listed in my device or computer
- Places I've been
- None of the above.

Q9: If you knew that all the data listed in question 8 was collected by your most used application, would you still use it?

- Yes
- No

Q10: By using more services/applications you are agreeing to allow them the use of your data. Should companies be required to ask users to "opt-in" for different data access?

- Yes
- No