The Data Collection Capabilities of Third-Party iOS Developers

Kunal Srivastava

Abstract

The world of technology is changing. It used to be one where users could post on social media, order clothes online, and play mobile games without a care in the world. As time progressed, companies eventually realized that they could collect small pieces of information from a user's technological habits, and use that to form a profile of the user of some sort. Such a profile could be used for marketing, grouping, or even just to sell-off. I conducted a study of mobile device application developers, specifically relating to iOS third-party applications. This is more of a hidden loophole in the "security" of Apple's iOS. This is important to the field of data privacy because many users use third-party applications without the knowledge of what data is being taken and when it is being taken. I used a packet-monitoring service to listen in on third-party applications and wrote a script to filter out relevant pieces of data that were being sent back to an application's server. The findings of this study concluded that third-party iOS applications have unmonitored power to access pieces of information like contacts, location, and emails.

1 Introduction

As more and more technology becomes introduced into this modern-day world, data privacy continues to become an increasingly important topic of discussion. But more importantly, data privacy proves to be an increasingly important problem to solve. In early 2017, data surpassed oil in value, deeming it the most valuable resource in the world [7]. In sum, data is money. With data, companies and data collectors can make several logical assumptions about a user; this furthers their depth of such a user's digital profile, which could consist of details like sleep schedule, shopping interests, location, etc.

According to GSMA real-time intelligence data [6], there are now over 5.13 billion people in the world with smartphones, and there are 1.2 billion more mobile connections than humans on this earth. Forrester's Research [8] claimed in their annual American survey that 73 percent of Americans use their mobile phones as their most-used device. With mobile phones being so popular, the issue of data privacy specifically about smartphones (mobile devices) must be raised. Android's Google Play Store and iOS' Apple Store both prove to be somewhat secure and transparent when it comes to data transparency; this means that the user is notified about what data each application collects in some way.

Third-party iOS Applications can bypass the limitations that a traditional App Store imposes. These applications are popular in the sense that they are not allowed in the classic App Store, meaning that a user must go to their mobile browser to download the app. These types of apps offer certain advantages to their traditional App Store counterparts. For example, it might contain certain cheats to a mobile game, or special insights to a social media profile. Third-party apps are in demand as they offer an advantage over regular apps to their users.

In this paper, I reveal the extent to which third-party iOS application developers can go to when collecting data. The general approach is to study some of the most popular third-party applications, monitoring/decrypting the packets they send and to which/what domains. The significance of this study is that while the traditional App Stores are under heavy scrutiny regarding data privacy, third-party apps are disregarded. They pose an equally significant threat to data security in the mobile scope. The results of this study will hopefully raise awareness of the dangers of having third-party apps downloadable.

Similar works have addressed the privacy concerns in both the App Store and the Google Play Store, but none have gone into the topic of third-party apps, and how their actions are unverifiable. In the end, the user will have to trust the developer, rather than a traditional App Store having to do it.

The remainder of this paper is structured as follows. In the Prior Related Work, I will provide work that corroborates and adds further knowledge to my own. In the Research Plan, I will provide an in-depth outline of the research process. In Method I will specifically explain my data and also begin to analyze it. In Findings and Insights, I will provide an overview of my findings and conclusions on the topic. In Discussion, I will bring up my personal opinions and further questions about my process. Finally, I will conclude the paper with an overarching lesson and a general answer to my research purpose.

2 Related Work

While Third Party iOS App Development is still somewhat an undiscussed topic, one can learn from previous works in the realm of mobile device privacy, data, and iOS developers.

Many works relate to the topic of mobile device privacy. Firstly, a work by Keith et al. [3] is a work regarding the potential dangers of mobile devices, and what information can be disclosed. It assumes that consumers are willing to accept privacy risks for relatively negligible benefits. Under that assumption, the authors conduct a fairly smooth and accurate study around the realtime dangers of mobile devices. The study operated while maintaining a high degree of experimental control and internal validity, and demonstrated a way to collect real information regarding mobile devices. The result was the generation of more accurate and practical conclusions. A work by Minch et al. [4] is a work specifically centered around the location aspect of mobile device information. Although this paper is more of a policy paper, it gives the reader an idea of how and when mobile devices can track a user's specific location. The study concludes that regulation from higher organizations is the best way to combat this problem. This could be from governments, industry groups, market forces, and many more groups.

When it comes to data, it is vital that one understands the potential danger of giving it all away. Just like a social security number, or your home address, it is a right to keep one's information private. The work by Boyle et al. [1] is a study centered around the protection of data in mobile devices. This statistic-centered study found many surprising and actionable pieces of information regarding the data that mobile devices collect, and even more importantly, the actions humans take to stop their data from being taken. This can include clearing application history, turning off location services, and wiping browsing history. The study found surprising statistics, like twelve percent of users feel their data privacy has been invaded. Mobile device data is something that must be protected.

iOS Applications have been around ever since the birth of the Apple OS in 2007. The first selected work, by Mohamed et al. [5], is a comparative study between the security of iOS and Android. These are the two main brands of mobile device operating systems, so a comparative study into their functions is something that is well-needed and important. The work continues to compare these two operating systems, on factors like application permissions, encryption mechanisms, and application isolation. Another work, by Egele at al. [2], is about the privacy threats that iOS applications pose to a user. In this study, the authors examine privacy threats from mobile applications, via a tool developed that allows them to analyze applications for possible leaks of important information, such as location or contacts. Of the 1,400 iOS applications analyzed, the researchers concluded that most developers respect user privacy. The bad applications, however, leaked information like the device hardware ID, and even location in some instances.

All these works give us a premise to begin this study off of. With information about mobile devices, data, and iOS applications, I am ready to conduct my study.

3 Method

The plan for this study is to deeply understand the potential of third-party iOS developers, meaning understanding if there is any difference between authorized iOS applications and third-party applications. If there is a difference, the goal is to understand the true potential of these apps. Is a third-party application able to track my location? My email? My contacts? These are the types of questions I hope to answer.

A generic framework to my research plan is to take the top forty third-party applications, download them on a dummy (fake) device, and use a VPN/Packet Monitoring service to intercept packets and download them on a PC. I will then use trust certificates to decrypt each packet, giving me the power to read each packet in a python script and look for keywords like "Mountain View, California", "dummy542@gmail.com", or third-party services like "amazon.com" or "nike.com". This will accurately give me information about what apps are giving away certain pieces of information. I originally chose to manually read each packet, and it worked for the first five applications. Eventually, automation became a more efficient way.

The top forty third-party apps proved to be somewhat fun. The applications provided cool and insightful additions to social media apps, and very useful cheats on some of the most popular games in the modern era, including Clash Royale and Pokemon Go. It is clear to see why these applications are becoming increasingly popular; they offer attractive additions to popular authorized applications.

My process started with the downloading of the top forty third-party iOS applications. After downloading many games, social media, and even "better app stores" (leading to even more apps), the most surprising thing I noted was the absence of a privacy notification. Usually, when downloading an authorized application, a message will pop up like "Application name would like to have access to the camera". This gave the user the option to essentially give the application their data. In this case, it would be access to his/her camera. The lack of this already gave the hint that something was up.

My data model was somewhat simple; it consisted of the number of packets sent out in two minutes, the number of POST requests, GET requests, total requests, addresses/domains, and packet size for each application. After playing/using each application for two minutes precisely, I used a service to intercept and read the data. Charles in a highly polished HTTP proxy that enables a developer to view all of the HTTP and SSL traffic between their device and the internet. Charles proved to be extremely useful to collect all my data, being both efficient and accurate in the data collection process. With exactly eighty minutes of packet information (two minutes of forty sessions each), I loaded each session onto my PC. At first, the packet contents were encrypted. After some research, I used a trust certificate to allow the PC to decrypt the information in each packet.

After collecting the data, I loaded the decrypted text of each of the forty sessions into a single text file. I used a python script to scan the large-sized file and look for keywords that specifically pertained to my name, location, email, phone number, contacts, and other pieces of data that could be extracted on a mobile device. I organized all the data, including the findings for each packet, into a data table.

The contents of the collected data proved to be surprising. There was a large amount of GET/POST packets (the packets that send out information, usually data), compared to the total number of packets. Also, I saw both familiar and unfamiliar addresses. The meaning of this will be later discussed. All the packet size remained around the same. There was also something in the description column of the table that I expected — that each type of application collected around the same type of data. Social media applications took my email, name, and other personally identifiable information. Other applications like games took information that would pertain to them succeeding, like contacts.

Overall, I am fairly happy with my research design. The only change in the design was to make the process of interpreting the data automated. Reading each session individually proved to be a challenge, especially to a sole researcher like myself. I am happy with the planning and execution of my research plan.

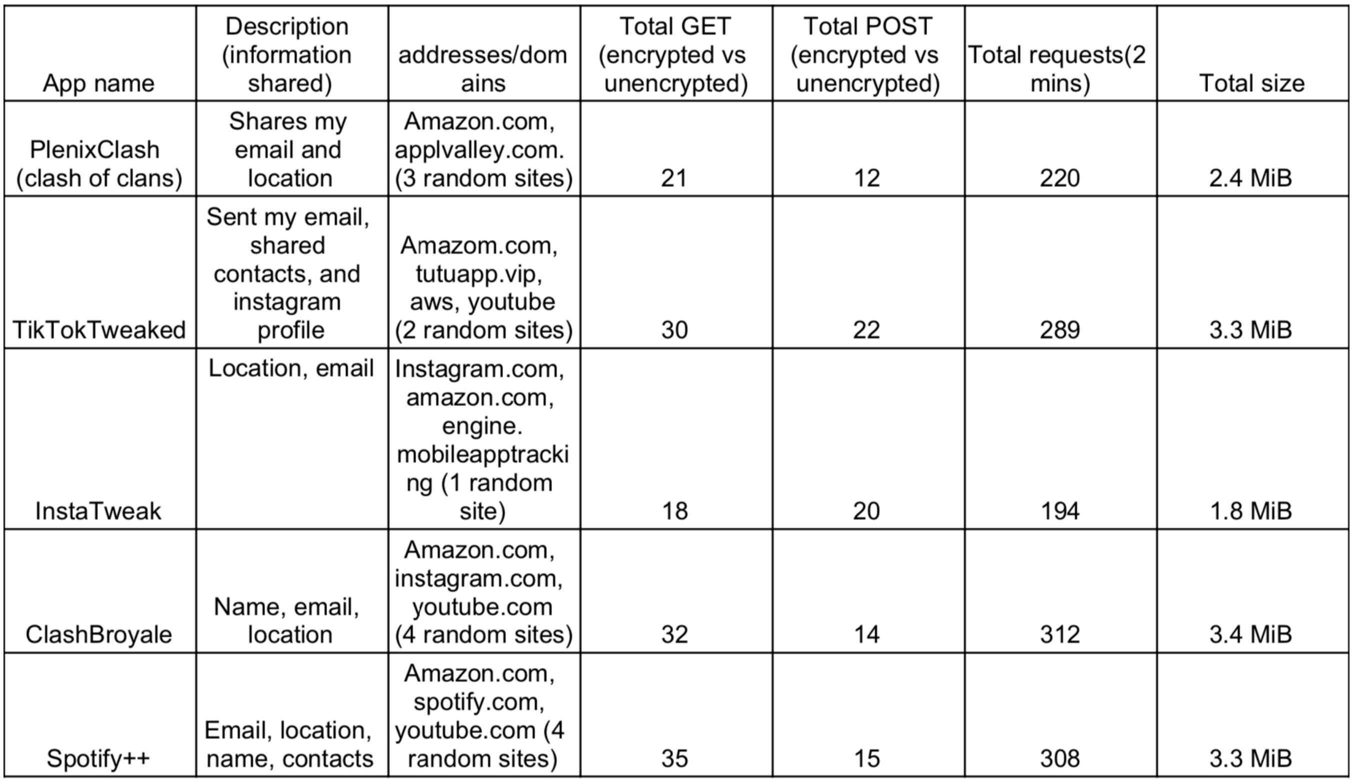
4 Findings and Insights

The findings of the study were both expected and surprising. Out of the forty third-party applications studied, sixteen of them showed some sort of suspicious activity. Here are five of the most surprising elements of certain applications and their details:

Honestly, while collecting this data, I was scared. Luckily, my entire mobile device (including the information on it) was a dummy. This meant that the data, including my name, email, and location, was fake. While looking through these data sessions collected by Charles, I constantly kept seeing the applications send my dummy name and email back to their server. The domain/address would be a totally random site, nothing like amazon.com or youtube. It is fair to assume that sites like amazon and youtube are collecting the dummy data for advertisement purposes. This is similar to how a user could browse amazon, specifically shopping for a certain item, and then see the same advertisement on their Instagram feed minutes later. This goes to show the power and value of a strong data collection system. Taking a closer look at this small set of the collected data, one can see a strong positive correlation between the total POST/GET requests, and total size of each session. This means that the more one uses a "tweaked" application, the more data is collected and sent back to the application's servers. It is clear that these attractive, manipulative applications are sending information that a user is not aware of. Let it be known that not once did I receive a notification or prompt that requested the application's access to certain pieces of data, whether that be the phone's hardware ID, a user's name, or location. After looking at this collected data, it is clear to see how third-party iOS applications can prove to be manipulative when stealing user data. My findings from this study complement the findings found in the Related Works section. The majority of the related works cited ended with a warning to the power potential of certain data-specific technologies, whether it be location services or iOS compatibility.

5 Discussion

Many additional topics and concerns have arisen during my research process. One specific question that has come up is whether there is an actual limit to the data that third-party iOS application developers can extract from users. Although that was the exact purpose of this study, my results showed that although there were common trends in the data points collected, certain applications showed no limit to their data collection capabilities. Clearly this raises some pretty large concerns about citizens and their privacy concerns. It is in my opinion that eventually, the ultimate solution to the entire problem of data privacy will have to be a policy/law that limits operating systems like iOS and Android to certain protocols when collecting data. I think that it is such a big problem right now because of how early and new data collection



App name	Description (information shared)	addresses/domains	Total GET (encrypted vs unencrypted)	Total POST (encrypted vs unencrypted)	Total requests(2 mins)	Total size
PlenixClash (clash of clans)	Shares my email and location	Amazon.com, applvalley.com. (3 random sites)	21	12	220	2.4 MiB
TikTokTweaked	Sent my email, shared contacts, and instagram profile	Amazom.com, tutuapp.vip, aws, youtube (2 random sites)	30	22	289	3.3 MiB
InstaTweak	Location, email	Instagram.com, amazon.com, engine. mobileapptracking (1 random site)	18	20	194	1.8 MiB
ClashBroyale	Name, email, location	Amazon.com, instagram.com, youtube.com (4 random sites)	32	14	312	3.4 MiB
Spotify++	Email, location, name, contacts	Amazon.com, spotify.com, youtube.com (4 random sites)	35	15	308	3.3 MiB

Figure 1: Specific Data Related to Selected Applications

is. Laws and policy just haven't had time to adjust. Regardless of this, user's have a right to keep their data private.

If I had to do something differently next time, I would find a way to survey and analyze even more applications. I felt like the most popular forty applications were a solid start, but it really comes down to the whole third-party community as a whole if I want to gauge accurate and potential dangers. I think that more data will prove to show a little bit more quantitative data rather than the few points of qualitative data that I provided.

One main challenge came up as I was executing the research process. When I was interpreting the data from the forty sessions, it was really hard to extract the relevant information, whether it be in raw text, hex, or anything else. To solve this, I wrote a python script that added each of the forty sessions to a text file, and then ran through the file, both converting all of it to raw decrypted text, and then looking for keywords that pertained to certain data points like location, name, or email. This severely improved my accuracy and efficiency when it came to reading my data and turning it into something usable.

6 Conclusion

In conclusion, this study investigated the power and capabilities of third-party iOS developers. In a world of technology, data is one of the most valuable aspects to an individual. Each user's data must be protected, or the user should at least know where their data is going. The findings of this study concluded with the finding that these unverified third-party applications theoretically have unlimited power, with access to pieces of data that include location, contacts, email addresses, and many more. The lesson learned after conducting this study is that there must be something done with third-party IOS apps for the data to be truly secure and safe.

References

[1] Jan Lauren Boyles, Aaron Smith, and Mary Madden. “Privacy and data management on mobile devices”. In: *Pew Internet & American Life Project* 4 (2012).

[2] Manuel Egele, Christopher Kruegel, Engin Kirda, and Giovanni Vigna. “PiOS: Detecting Privacy Leaks in iOS Applications.” In: *NDSS*. 2011, pp. 177–183.

[3] Mark J Keith, Samuel C Thompson, Joanne Hale, Paul Benjamin Lowry, and Chapman Greer. “Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior”. In: *International journal of human-computer studies* 71.12 (2013), pp. 1163–1173.

[4] Robert P Minch. “Privacy issues in location-aware mobile devices”. In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the.* IEEE. 2004, 10–pp.

[5] Ibtisam Mohamed and Dhiren Patel. “Android vs iOS security: A comparative study”. In: *2015 12th International Conference on Information Technology-New Generations*. IEEE. 2015, pp. 725–730.

[6] “The Mobile Economy 2019”. In: *GSMA* (2019).

[7] “The world’s most valuable resource is no longer oil, but data”. In: *The Economist* (2017).

[8] Giselle Tsirulnik. “Mobile phone ranked most used electronic device: Forrester”. In: *mobilemarketer.com* (2017).