VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

A Systematic Study on Mobile App Privacy in Software Engineering Research

Course Slot : A1+TA1

Course Instructor : Dr. B. D. DEEBAK

Project Associates:

1. Uday Sonawane (19BCE2373)

2. Sudhanshu Gupta(19BCE2674)

3. Riya Kumar (19BCE2694)

OBJECTIVE

With increasing digitalization and dependency of users on mobile applications for various day-to-day activities and the emerging trend for personalized services, the security of users' personal data is at stake. Most of the users are unaware about technical details of data collection and distribution with the help of third party APIs. Even when the permission to access the data is given by the users themselves, there is limited restriction as to where the accessed data may be shared and how it may be used. With the help of this study, we aim to conduct a theoretical research on a) the importance of mobile app privacy, b) the influence of operating systems on the extent to which privacy is ensured, c) the current measures implemented to enhance user privacy and d) the scope of future research.

Type of research: Qualitative

PRIOR RESEARCH

In a paper by Yu et al., an automated approach for establishing semantic correlations between apps' privacy policies and their code has been introduced by the name of TAPVerifier. This approach relies on text analysis to automatically establish and match the semantic meaning of the app's privacy policy, its bytecode, app store description, and permissions [1,2]. In addition to this, Yu et al. has also presented PPChecker, a tool for assessing the trustworthiness of Android apps' privacy policies that revealed that about 74% apps have at least one kind of identified privacy problem.[3] In a paper on framework for detecting privacy policy violations in Android application code by Slavin et al., A semi-automated framework has been proposed that helps app developers detect inconsistencies between their privacy policies and codes.[4] VisiDroid is a visual configuration interface for end users to understand and configure the way their private information is utilized is introduced by Aydin et al.[5]. Apart from these, an approach to automatically detect app privacy policy violations based on user input by mapping each Graphical User Interface (GUI) input element of the app to ontology concepts, which in turn are matched with text from the privacy policy has also been proposed by Wang et al. in his paper on Tracing privacy policy claims [6].

SIGNIFICANCE

These articles have given us a better grasp of mobile app privacy and the criteria we use to determine how secure an app or software is. They also assisted us in analysing the various privacy features available on Android and iOS, which greatly aided our literature review. We learned about various methods by which apps and software attempt to violate our privacy, as well as methods to avoid these things.

Overall, these papers have brought to light a lot of issues that, most of the time, go overlooked when it comes to privacy. As a result, these articles have made significant contributions to the study of mobile app privacy.

Mobile app ecosystems have been exposing a growing number of APIs through their software development kits. Many of these APIs involve accessing sensitive functionality and/or user data and require approval by users. According to the paper published by Liu, B., Lin, J. and Sadeh, N., Android for instance allows developers to select from over 130 possible permissions. [7] It has become unrealistic for common users to review and adjust settings related to so many permissions. Most of the time, the users as well as the developers are unaware about libraries that use third party API's to employ IAMs and track the applications downloaded by the user. ". IAMs usages appear to be considerably more common in commercial apps, with a total of 4,214 apps employing them, amounting to 30.29% of the total.[8] This is used mainly for the purpose of customized advertisements. Apart from this, many apps are found asking for unnecessary permissions from users including reading data from their mobile phones. These are the main sources of data leak and privacy breaches.

INTRODUCTION

Data-driven service can be a boon or a bane for users as the perceived benefits of personalization are often more visible than the perceived risks of sharing personal data. Examples such as the Facebook Cambridge Analytica scandal in which a political data analytics firm was able to access data belonging to about 87 million people and use this data to influence the outcome of the 2016 US elections display a weak link in the protection of users' personal data and privacy. Recent studies have concluded that mobile apps are an even bigger threat than facebook when it comes to privacy.

While studying the permission usage among VPN apps available on Play Store, it was observed that more than 60% apps ask for permissions that are not needed for their functioning and thus pose a risk to users' privacy. It is commonly seen that when a user searches for a particular product on one application, he/she is immediately flooded with advertisements of a similar product on other applications and platforms. Such incidents raise a concern amongst the user as to what extent is their data being used and how it is being circulated without the users' knowledge.

With increasing cases of privacy breaches and data leaks and increasing awareness among the users, it has become an urgent need to study the cause of such incidents and to introduce measures to prevent such incidents in future. As developers, it is our responsibility to explore the field of mobile app privacy to build an understanding of the drawbacks of personalized user experience and to study and compare the privacy policy of various platforms to conclude the methods that can be used to improve the protection of users' data.

METHODOLOGY

We conducted a systematic study about mobile app privacy, we studied research papers based on the year in which they were written and what they had to offer. From these papers we were able to find out about various methods and ways in which privacy is being compromised.

Our research compares privacy between IOS and Android, we try to identify the differences between privacy policies, and among them which provides us with better privacy and shows concern related to the user's personal data. We have done this by answering following Research Questions:

-RQ1: Privacy as a feature between IOS and Android?

Under this question, We try to find whether there are any specific features between IOS and Android. This will help us understand how well our privacy is protected without our supervision. Over Here our objective is to find efforts done on ensuring security of users and steps taken by them.

-RQ2: Privacy as a User control in IOS and Android?

Under this question, We try to compare privacy functions which can be accessed by users, and modified. Does this help with any better security of our personal/private information.

LITERATURE REVIEW

A Comparative Study between android and iOS

A comparative study of two most prevalent mobile app platforms, namely Apple's iOS and Google's equivalent, Android, shows that Android and iOS are not only competitors in the market but are also fundamentally different environments for both the developers and the consumers. While Apple acts as a gatekeeper to control the market access of the applications available on its App Store, Google is one of the main data-collecting adversaries. To publish an application on Apple's App Store, developers are required to comply with Apple's content, privacy and security policies in order to gain approval for publishing it. However, Android does not have an approval process for the applications. To publish an app in the Play Store, the only requirement is to agree to Google developers Distribution Agreement which states that Google does not "Undertake an obligation to monitor the products or their content.". Thus, while Apple undertakes the responsibility to ensure that the applications available on the app store are cross-verified for their content as well as their privacy measures, Google does not take responsibility for the apps available on Play store[9].

Privacy in iOS

According to iOS, privacy is considered to be a legitimate design concern. Apple's gatekeeper role plays a significant role in defining the scope of privacy. For iOS developers, privacy is generally defined by Apple's review process. iOS promises to keep all data private, meaning that sharing of data beyond the app company is prohibited. Therefore, sharing of data beyond the application will lead to violation of Apple's privacy expectations and thus may lead to the removal of the application from the App Store. This forces developers to limit the data sharing and thus improve the security of the users' data. One of the most discussed privacy definitions in iOS is transparency with users in the form of consent and notice. Most kinds of data collection are allowed on iOS as long as the user provided their consent. However, Apple does not state a clear boundary as to what "counts" as privacy and Apple's approval process is the only obstacle to restrict the apps from accessing data without prior permissions from the user.

Privacy in Android

Android lacks the approval process. It discloses privacy in the form of a choice for the users in a mobile ecosystem that contains many potential threats, including google itself. Privacy is defined by android in two major ways:

1) Privacy as a feature:

Android application developers mostly categorize privacy as a feature rather than as a necessity. They develop applications that ensure users' privacy in order to distinguish their application from the other apps that are available on the Play Store.

2) Privacy as user control:

Android also defines privacy as users' individual control over personal data. Emphasis on "End-to-end encryption" by WhatsApp is one such example of spreading awareness about the importance of users' control over the full mobile stack. Including user control under privacy also led to the development of many applications that allow the user to block certain other apps from accessing the data at any particular time. However, there was no guarantee that these applications that blocked the data for other apps were not, indeed, accessing the data themselves.

CONTRIBUTIONS

From this study, we can conclude that the collection and usage of users' data significantly depends on how privacy is defined by the platform. While iOS presents that strict control and regulation by the platform is necessary to ensure the privacy of users, android displays a free market model to encourage open-source development while compromising with privacy related issues. Because of the regulations and restrictions in usage of the majority of third party APIs in iOS, it acts as a more secure platform than android. However, privacy is a yet more complex

notion than what is defined in Apple's current policies and can not be ensured only by blocking third party APIs or regulating the app approvals.

Considering the fact that privacy still remains to be a major issue independent of the platform chosen, there was a need to study further as to how privacy leaks can be detected and prevented. According to the surveys conducted by Ebrahimi et al. in the base paper, It was found that the majority of work done in this field is still limited to introducing new and effective methods to detect privacy leaks. The methods found were mostly static models to analyse the codes of the application (code, API calls, GUI components, etc) that might be the cause of leakage of information. However, these static means are not enough to tackle the privacy issues and there is a need for applying dynamic analysis along with the static analysis in order to track the execution traces of apps which would help in obtaining a detailed view of the app's behaviour. Some of the limitations found in the authors' work are based on inclusion and exclusion of literature based on the author's preferences and the lack of focus on methods to improve privacy after detecting the cause of such leaks.

KEY ISSUES

1) Weak or no encryption:

Encryption algorithms are the first line of defence when it comes to blocking attacks on user phones and even the servers. But with the rise in technology, these encryptions also need to be upgraded. There have been apps that stored user information using basic encryption techniques which can be decrypted or hacked easily. Areas where encryption is not possible: Encryption can be implemented in some fields such as texting apps like whatsapp, telegram, etc. But encryption cannot be used in the search engines.

2) Insecure Data Storage

Data storage is one of the important concerns which can cause privacy issues. Nowadays most of our personal information is stored within a social networking app. So if somehow the data related to these apps is being compromised, it is similar to making our information public to everyone.

3) Data Leakage

Data leakage is the unauthorized transmission of data from within an organization to an external destination or recipient. A “leaky” app is defined as one that can easily be tapped to retrieve users’ personal details such as age, location, gender, etc. There are many causes of data leakage from a mobile app such as application vulnerabilities, Overly complex access permissions, Malwares, third-party APIs, etc. Some of the well known data breach issues were faced by Facebook(due to Insecure web app features),WhatsApp(due to Malware injection through

insecure call function), Under Armour(due to mobile security breach that revealed user details and passwords). Another major source of users' privacy breach are poorly developed applications which may have vulnerabilities that reveal data and locations of its users. Data Leakage by applications such as healthcare mobile apps can even reveal a patient's personal data including their medical reports and medical test details which may lead to violation of medical security regulations.

FURTHER RESEARCH

After the paper analysis and research, we found that iOS is tackling the issue of security better than Android. Like iOS, every operating system should test an application for these key issues. We analysed the use of virtual machines or applications that would detect the risk in using the particular application, if any. Now our concern is how efficient the application is to detect the risk. Accuracy is a factor which can be improved at any point of time and it leaves a scope for further research and implementation. But this also leaves us behind with a thought: *What if the application used for testing other applications misuses our data?*

Another difficulty was the inability to encrypt the information that we feed on search engines. Since we cannot encrypt, the application can use the information for commercial purposes or any other. We can make extensions to prevent the ads getting displayed on different websites or applications we use. *Could we stop them from getting our search feeds?* No. This can only prevent applications from commercializing, which according to our research was the main cause of misusing our data. This again leaves scope for further research on how efficiently we can block ads on different operating systems and the methods by which we can prevent applications to store our search feeds.

The applications that take access to our phone's storage, may misuse them. In our research we found some methods on how we can determine if the application is misusing them. We can also think of assigning different directories to the applications. The user can transfer or make a copy of files into that directory and the application gets access to only that. *Is it storage and cost efficient?* The concern arises with storage as we are using multiple copies of a file. This gives us a scope of future research on how the storage problem could be solved. *What can be a possible solution to it?* We can make some mapping from the application to the files that we want to give access for. By this the application can access only permitted files. Again this leaves us with an idea that can be researched upon in future.

REFERENCES

Base Paper: *Ebrahimi, F., Tushev, M., & Mahmoud, A. (2020). Mobile App Privacy in Software Engineering Research: A Systematic Mapping Study. Information and Software Technology, 106466.*

Reference Papers:

[1] L. Yu, X. Luo, C. Qian, S. Wang, and H. Leung, "Enhancing the description-to-behavior fidelity in Android apps with privacy policy," IEEE Transactions on Software Engineering, vol. 44, no. 9, pp. 834–854, 2018.

[2] L. Yu, X. Luo, C. Qian, and S. Wang, "Revisiting the description-to-behavior fidelity in Android applications," in Inter. Conf. on Software Analysis, Evolution,

and Reengineering, 2016, pp. 415–426.

[3] L. Yu, X. Luo, J. Chen, H. Zhou, T. Zhang, H. Chang, and H. Leung, "PPChecker: Towards assessing the trustworthiness of Android apps' privacy policies," IEEE Transactions on Software Engineering, 2018.

[4] R. Slavin, X. Wang, M. Bokaei, J. Hester, R. Krishnan, J. Bhatia, T. Breaux, and J. Niu, "Toward a framework for detecting privacy policy violations in Android application code," in Inter. Conf. on Software Engineering, 2016, pp. 25–36.

[5] A. Aydin, D. Piorkowski, O. Tripp, P. Ferrara, and M. Pistoia, "Visual configuration of mobile privacy policies," in Inter. Conf. on Fundamental Approaches to Software Engineering, 2017, pp. 338–355.

[6] X. Wang, X. Qin, M. Bokaei, R. Slavin, T. Breaux, and J. Niu, "Guileak: Tracing privacy policy claims on user input data for Android applications," in Inter. Conf. on Software Engineering, 2018, pp. 37–47.

[7] Liu, B., Lin, J. and Sadeh, N., 2014, April. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help?. In Proceedings of the 23rd international conference on World wide web (pp. 201-212).

[8]Scoccia, G.L., Kanj, I., Malavolta, I. and Razavi, K., 2020, July. Leave my apps alone! a study on how Android developers access installed apps on user's devices. In Proceedings of the IEEE/ACM 7th International Conference on Mobile Software Engineering and Systems (pp. 38-49).

[9] Greene, Daniel, and Katie Shilton. "Platform privacies: Governance, collaboration, and the different meanings of "privacy" in iOS and Android development." new media & society 20.4 (2018): 1640-1657.