THE MEASUREMENT OF THE COMPLETENESS OF MOBILE APPLICATION PRIVACY POLICIES IN THE GOOGLE PLAYSTORE USING A MANUAL RANDOM SAMPLING APPROACH

Ollscoil Teicneolaíochta an Oirdheiscirt
South East Technological University

C00346748 Luke Raeside

MSc in Cybersecurity, Privacy and Trust, Supervisor: Dr Jason Barron

Abstract

Modern mobile phones (smart phones) provide end users with technology that provides for personalized and customized software. Applications (Apps) span many different categories including, gaming, personal organization, health, fitness and finance. To provide these personalized applications, mobile App developers often need to collect, store and process personal identifiable information (PII). The nature of the data processed within these Apps is can be highly sensitive and requires special attention by the developers.

The Google Playstore and Apple App store are the most popular providers of Apps, and they provide guidelines to developers to help protect sensitive user data. End users implicitly place a lot of trust in the App developers to treat their data with the appropriate level of security and privacy. Privacy policies are the only formalized method for end users to assess the level of privacy and security their data receives when considering the use or purchase of an App. Several researchers have developed automated methodologies and tools to review the completeness of privacy policies. The tools are primarily based on natural language processing (NLP) and subsequent automated analysis techniques. However, given the complex nature of the language used in the policies, the brute force approach to analyzing App privacy policy files may not be as effective as random sampling techniques and human analysis.

This research project will carry out a randomized manual analysis of a sample set of privacy policies in the Playstore and compare the results of the manual analysis to the results achieved using automated analysis. The results will provide insights into the effectiveness of the human analysis when compared to the automated analysis. The results of this study may assist in formulating the most effective techniques to ensure users can have trust in mobile App personal data processing.

Research Questions

The three driving research questions for this research study are:

- How do the results of a manual randomized sample analysis of the completeness of mobile App privacy policies in the Google Playstore compare with recent automated privacy policy analyses?
- Expanding the definition of privacy policy completeness to the completion of related duties: what proportion of the random sample of mobile Apps from the Google Playstore respond to queries relating to the privacy policy via the named data controller, within a reasonable amount of time (14 working days).
- Expanding the definition of privacy policy completeness to include their effectiveness in achieving their stated purpose: what proportion of a random sample of mobile applications users are aware of, have read, accessed or sought further details from an Apps privacy policy.

.

Introduction

- The Apple Store provides almost 2 million mobile applications (Apps), and the Google Playstore provides almost 4 million Apps
- Data Safety information is generally provided by the developer of the App prior to purchase\download
- The Google Playstore provides basic guidelines for the completion of privacy policy files
- Developers, however, are responsible for the documentation and compliance of the privacy policies within the Playstore
- Developers are obliged, by regulators, to ensure that regulations are complied with wherever the App is available
- Regulatory compliance is not represented by one standard there are numerous international regulations and standards (Figure 1)
- Given the complexity of the privacy policy language and the variety of information and data collected it is possible that human-based assessment of the completeness of privacy policies is the most effective approach (even though much of the research suggests that automated analysis is the way forward)
- This research will carry out analyses of existing App privacy policies using a human-based approach and the output of these analyses will be compared with the findings of recent automated analyses
- This research will help to scope the most effective way to assess completeness of privacy policies

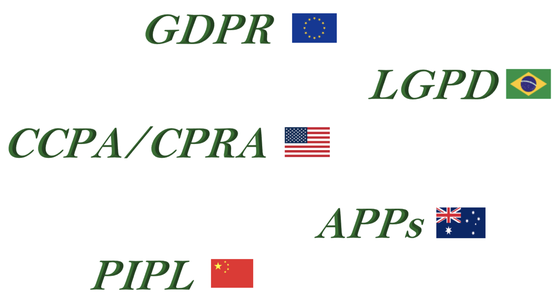

Figure 1: Some International Regulation and Compliance for Data Privacy

Methodology

The project will be split into three distinct phases. The first phase will involve the collection of a stratified random sample of privacy policies from the Google Playstore. Stratification will be used because to gather a representative sample different categories of Apps will be downloaded. The second phase will extract the contact information from the privacy policies and attempt to make contact and await a response. Simultaneously, the sample set of policy files will be read and measured against the criteria for completeness. The last part of the second phase will research end user attitudes toward privacy and awareness of policy files: a questionnaire may be conducted based on the research finding. Finally, all the quantitative data will be summarized and analyzed, and conclusions and suggestions will be made based on the data. The flowchart below (Figure 2) summarizes the planned research methodology.

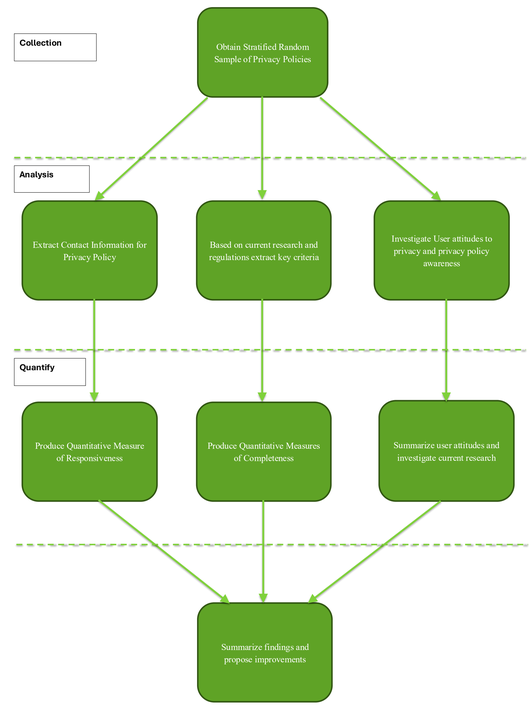

Figure 2: Overview of Project Methodology

Conclusion

Mobile smartphones are the number one access medium for the internet. Mobile Apps are the primary mechanism that smartphones utilize for accessing online services. Many of these services collect sensitive data from the mobile users. It is very important that user privacy is respected by the developers. We must devise mechanisms to inform users when Apps are clearly not providing adequate protection, without placing too high a burden on the end users to self-police completeness and compliance.

References

Alkhattabi, K., Yue, C. (2024). Completeness Analysis of Mobile Apps' Privacy Policies by Using Deep Learning. In: Zhao, J., Meng, W. (eds) Science of Cyber Security. SciSec 2024. Lecture Notes in Computer Science, vol 15441. Springer, Singapore. https://doi.org/10.1007/978-981-96-2417-1_5

Amaral, O, Abualhaija, S, Torre, D, Sabetzadeh, M & Briand (2022), 'AI-Enabled Automation for Completeness Checking of Privacy Policies', IEEE Transactions on Software Engineering, vol. 48, no. 11, pp. 4647-4674

CCPA (2018). *California Consumer Privacy Act (CCPA)*. [online] State of California - Department of Justice - Office of the Attorney General. Available at: https://oag.ca.gov/privacy/ccpa. Accessed 29 October 2025.

CovidTracker.ie. (n.d.). *The Covid Tracker app*. [online] Available at: https://www.covidtracker.ie/. Accessed 31 October 2025.

Fan, M., Yu, L., Chen, S., Zhou, H., Luo, X., Li, S., Liu, Y., Liu, J., & Liu, T. (2020). An Empirical Evaluation of GDPR Compliance Violations in Android mHealth Apps. 31st International Symposium on Software Reliability Engineering (ISSRE), Software Reliability Engineering (ISSRE), 2020 IEEE 31st International Symposium on, ISSRE, 253–264. https://doi.org/10.1109/ISSRE5003.2020.00032

GDPR (2018). *Complete guide to GDPR compliance*. [online] GDPR.eu. Available at: https://gdpr.eu/. Accessed 29 October 2025.

Google (2019), Privacy Policy Guidance, Google Developers, https://developers.google.com/assistant/console/policies/privacy-policy-guide. Accessed 25 Oct. 2025.

Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K.G. and Aberer, K., (2018). Polisis: Automated analysis and presentation of privacy policies using deep learning. In 27th USENIX Security Symposium (USENIX Security 18) (pp. 531-548).

Rodriguez, D., Yang, I., Del Alamo, J. M., & Sadeh, N. (2024). Large language models: a new approach for privacy policy analysis at scale. Computing, 106(12), 3879–3903. https://doi.org/10.1007/s00607-024-01331-9

Richter, F. (2023). *There are now more mobile phones than people in the world*. [online] World Economic Forum. Available at: https://www.weforum.org/stories/2023/04/charted-there-are-more-phones-than-people-in-the-world/. Accessed 29 October 2025.
TrustArc (n.d.), *TrustArc The Leader In Privacy Management Software*. [online] Available at: https://trustarc.com/. [Accessed 29 Oct. 2025].

websitepolicies.com (2025). *Breadcrumb from Facebook*. [online] WebsitePolicies. Available at: https://www.websitepolicies.com/create/privacy-policy [Accessed 29 Oct. 2025].

S Wilson, F Schaub, AA Dara, F Liu, S Cherivirala, PG Leon, MS Andersen, S Zimmeck, KM Sathyendra, NC Russell, TB Norton, E Hovy, J Reidenberg, N Sadeh (2016) 'The creation and analysis of a Website privacy policy corpus', in 54th Annual Meeting of the Association for Computational Linguistics Acl 2016 Long Papers. doi:10.18653/v1/p16-1126.

.