Evaluating the data privacy of mobile applications through crowdsourcing

Ioannis CHRYSAKIS [a,b], Giorgos FLOURIS [a], George IOANNIDIS [c], Maria MAKRIDAKI [d], Theodore PATKOS [a], Yannis ROUSSAKIS [a], Georgios SAMARITAKIS [a], Alexandru STAN [c], Nikoleta TSAMPANAKI [a], Elias TZORTZAKAKIS [a], and Elisjana YMERALLI [a],

[a] *FORTH, Institute of Computer Science, Greece*
[b] *IDLab, Dept. of Electronics and Information Systems, UGent, imec, Belgium*
[c] *IN2 Digital Innovations GmbH, Germany*
[d] *FORTH, PRAXI Network, Greece*

Abstract. Consumers are largely unaware regarding the use being made to the data that they generate through smart devices, or their GDPR-compliance, since such information is typically hidden behind vague privacy policy documents, which are often lengthy, difficult to read (containing legal terms and definitions) and frequently changing. This paper describes the activities of the CAP-A project, whose aim is to apply crowdsourcing techniques to evaluate the privacy friendliness of apps, and to allow users to better understand the content of Privacy Policy documents and, consequently, the privacy implications of using any given mobile app. To achieve this, we developed a set of tools that aim at assisting users to express their own privacy concerns and expectations and assess the mobile apps' privacy properties through collective intelligence.

Keywords. data privacy, mobile apps, GDPR, crowdsourcing, collective intelligence

1. Introduction

We experience a massive increase in personal information utilised by smartphone applications (apps), whose invasive nature for harvesting personal data has been demonstrated in many studies. This trend is continuing, despite the recently-established legislation for personal data protection, such as CCPA (California), LGPD (Brazil) and GDPR (Europe). In fact, studies have shown that the level of compliance of organizations and businesses to GDPR is low[1]. Although tracking and data access by apps is often legitimate, users are unaware of the related privacy risks, because apps describe their privacy behavior in a vague Privacy Policy (PrP) document, which is typically written using legal language and terminology [1], in long and frequently changing documents[2], making it hard for users to read and understand the critical aspects related to their privacy. Thus,

[1] See: `https://gdpr.report/news/2019/07/22/almost-a-third-of-eu-firms-still-not-gdpr-compliant/`, `https://symantec-enterprise-blogs.security.com/blogs/expert-perspectives/gdpr-turns-1-many-companies-still-not-ready`

[2] `https://www.varonis.com/blog/gdpr-privacy-policy/`

it comes as no surprise that the typical consumer is not investing time in studying such documents before agreeing, thus unintentionally granting permission to apps to access, use, and share a wealth of personal information, in a manner unknown to the user.

In this paper, we present the CAP-A H2020 project[3], which aims to *support users in the daunting task of understanding the content of a PrP document and to be aware of the privacy implications of using any given mobile app*[4].

Our position is that technical solutions and legal regulations are necessary but not fully sufficient for accomplishing a paradigm shift; at the heart of our solution is the hypothesis that data protection can also be powered by the society itself. By mobilising consumers to become active players, we can harness our collective power, leading to a more ubiquitous adoption of the technical and regulatory frameworks. To protect privacy adequately, society needs awareness, but also consensus about privacy protecting measures and processes that generate norms, with which service providers will voluntarily comply because it is profit maximising [2]. Exploring this knowledge is also of value to social scientists to better understand the community dynamics involved, as well as to policy makers to design more accurate and timely policies.

Along these lines, CAP-A deploys ICT tools that facilitate community interaction and co-creation in various ways that improve users' privacy awareness, and support a more efficient interaction among developers and end users; the latter will lead to a new innovation model that will allow consumers to collectively express their concerns, and developers to adopt more privacy-friendly practices and to better respond to market needs. CAP-A will also help in identifying and highlighting differences in opinions (i.e., norms), in a way that will be beneficial for users, developers, social scientists and policy makers.

2. The CAP-A portal and mobile app

The CAP-A portal is a responsive web page, whereas the mobile app offers additional functionalities adapted for small screens. They both rely on the same backend (which uses data stored using semantic technologies) and are available for public use in: `https://www.cap-a.eu/tools`. Due to space restrictions, we provide a brief description of the most important functionalities of the CAP-A portal and mobile app below[5].

Expectations. Through CAP-A, users can *express expectations*, i.e., whether they consider (or not) reasonable a certain data request on behalf of the developer. Each expectation is related to a certain privacy-related process, such as "access to camera", "minimisation of data collected" etc (called *Privacy Policy Practice* or *PPP* for short).

PrP annotator. CAP-A allows users to *annotate PrP documents* of apps, by marking a block of text in the PrP document and stating the relevance of this block to a certain PPP. Annotations are meant to highlight the important blocks of text in a PrP document and how they are related to PPPs, thereby simplifying the task of understanding its content.

Sharing evidences. Users can *share evidences* related to an app, which may be online articles, grounded claims by people who tested the app, or official documentation regarding its privacy properties. The credibility of such evidences is assessed by users.

[3] `https://cap-a.eu/`, funded by NGI_Trust, and implemented by the authors

[4] In the context of this paper, the term PrP refers to any type of Privacy Policy, Terms of Use, Consent Form etc document prescribing legally binding obligation on behalf of a developer concerning a particular app.

[5] Similar info, with screenshots, can be found in [14].

The mobile app. The *CAP-A mobile app* is a native Android app, which is not just a mobile-friendly version of the portal, but also allows users to conduct an "audit" of their installed apps, which allows targeted retrieval of information from Google Play.

Gamification and rewarding. *Gamification features* based on *rewarding mechanisms* are a well-known tool to support sustaining communities and for motivating contributors [10]. The CAP-A rewarding mechanism was developed using a general-purpose ontology [13], which captures various common features of diverse reward schemes. It encapsulates well-known gamification principles ([11]) and employs both intrinsic and extrinsic rewards ([12]).

App ratings. Each app in CAP-A is associated with two *privacy-related ratings*. These ratings are the *Satisfaction of Community's Expectations*, which measures how close the privacy expectations regarding the app (as expressed by the users) are to what the app is requesting, and the *Privacy Friendliness* rating, whose computation takes into account privacy-related best practices, such as easy-to-understand PrP documents. The calculation of an app's ratings is based on a set of weighted functions and parameters that aim to ensure an intuitive and fair behaviour.

Browsing apps. An easy-to-use *search and browsing facility for apps* is provided to allow users to access the app-related information (e.g., expectations, annotations, app ratings, evidences etc.). For legal reasons, only public information is shown. Moreover, not all apps found in Google Play have been downloaded; instead, the system automatically downloads data on the apps most relevant for its users.

The Privacy Dashboard. In the *Privacy Dashboard*, users can find visual representations of aggregated information about users and apps, as well as an aggregation of users' behavior in the form of privacy norms. For example, we can determine whether certain age ranges tend to adopt a certain privacy stance towards specific categories of apps.

The role of developers. CAP-A is not only addressed to consumers, but also developers, who can *claim the development of a certain app*, giving them special privileges.

Mini-tours. An important feature is the concept of the *mini-tours*, which allow newcomers to get a grasp of the main CAP-A functionalities, through step-by-step tutorials.

3. Related Work

Various works aim to improve privacy awareness, but using different methods than CAP-A. In [3,4,5] various techniques (textual summarization, NLP, semantic text matching etc) are used to support users in understanding the content of PrP documents. Similarly, in [6], a remodelling of PrP documents is proposed, as well as an Annotator for visualizing them using semantic metadata. Tools for improving privacy awareness using visual techniques have appeared in [7,8], whereas [9] presents an app which enables users to behaviourally analyse the privacy aspects of other installed apps.

4. Conclusion and Future Work

We presented CAP-A, a socio-technical solution aiming to improve privacy awareness and users' understanding of the privacy implications associated with the use of any given

mobile app. Our solution is based on crowdsourcing and collective intelligence measures. Despite the existence of a 1000-strong user base (partly through the sister initiative CAPrice[6]), only internal evaluation has been carried out for CAP-A so far; a large-scale evaluation through several pilots is currently planned. We also consider the incorporation of a debating/chatting tool (e.g., along the lines of our previous work, APOPSIS [15]), that will allow users, experts, and developers to express opinions on privacy-related aspects, share individual experiences, or justify viewpoints (e.g., on annotations).

Acknowledgement

This work has been supported by the EU H2020 programme under the NGI_TRUST grant agreement #825618.

References

[1] Anton, A.I., Earp, J.B., Bolchini, D., He, Q., Jensen, C., Stufflebeam, W.: The lack of clarity in financial privacy policies and the need for standardization. In: IEEE Security and Privacy, vol. 2, (2004).

[2] Sloan, R.H., Warner, R.: Unauthorized Access: The Crisis in Online Privacy and Security. CRC Press, Inc., 1st edn., (2013).

[3] Tesfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S. and Serna, J. PrivacyGuide: towards an implementation of the EU GDPR on internet privacy policy evaluation. In the 4^{th} ACM International Workshop on Security and Privacy Analytics (2018).

[4] Wilson, S., Schaub, F., Ramanath, R., Sadeh, N., Liu, F., Smith, N.A. and Liu, F. Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really Work?. In WWW-16, (2016).

[5] Oltramari, A., Piraviperumal, D., Schaub, F., Wilson, S., Cherivirala, S., Norton, T.B., Russell, N.C., Story, P., Reidenberg, J. and Sadeh, N. PrivOnto: A semantic framework for the analysis of privacy policies. Semantic Web, 9(2), (2018).

[6] Pandit, H.J., O'Sullivan, D. and Lewis, D. Personalised Privacy Policies. In European Conference on Advances in Databases and Information Systems, (2018).

[7] Angulo, J., Fischer-Hübner, S., Pulls, T., Wästlund, E. Usable transparency with the data track: a tool for visualizing data disclosures. In 33^{rd} Conference on Extended Abstracts on Human Factors in Computing Systems (2015).

[8] Raschke, P., Kupper, A., Drozd, O. and Kirrane, S. Designing a GDPR-compliant and usable privacy dashboard. In IFIP International Summer School on Privacy and Identity Management, (2018).

[9] Hatamian, M., Kitkowska, A., Korunovska, J. and Kirrane, S. 'It's Shocking!': Analysing the Impact and Reactions to the A3: Android Apps Behaviour Analyser. In IFIP Conference on Data and Applications Security and Privacy, (2018).

[10] McGonigal, J.: Reality is broken: Why games make us better and how they can change the world. Penguin (2011).

[11] Morschheuser, B., Hamari, J. and Koivisto, J., 2016, January. Gamification in crowdsourcing: a review. In HICSS-16, (2016).

[12] Kavaliova, M., Virjee, F., Maehle, N., Kleppe, I.A.: Crowdsourcing innovation and product development: Gamification as a motivational driver. Cogent Business & Management 3(1), (2016).

[13] Chrysakis, I., Flouris, G., Patkos, T., Dimou, A. and Verborgh, R.: REWARD: Ontology for reward schemes. In 17^{th} Extended Semantic Web Conference: Posters and Demos (2020).

[14] Chrysakis, I., Flouris, G., Ioannidis, G., Makridaki, M., Patkos, T., Roussakis, Y., Samaritakis, G., Stan, A., Tsampanaki., N., Tzortzakakis, E., Ymeralli., E.: CAP-A: a Suite of Tools for Data Privacy Evaluation of Mobile Applications. In 32^{rd} JURIX 2020, Demo session, (to appear).

[15] Ymeralli, E., Flouris, G., Patkos, T. and Plexousakis, D.: APOPSIS: A Web-based Platform for the Analysis of Structured Dialogues. In "On the Move to Meaningful Internet Systems" (2017).

[6] `https://www.caprice-community.net`