Enhancing Privacy Using Crowdsourcing Mechanisms

Alexandros Kostopoulos,
Ioannis Chochliouros,
Ioanna Papafili
Hellenic Telecommunications Organization
S.A. (OTE)
{alexkosto, ichochliouros,
iopapafi}@oteresearch.gr

Andreas Drakos
VELTI
adrakos@velti.com

ABSTRACT

Personal data have become a merchandisable asset encouraging various stakeholders to collect data and trade them without the end-user's awareness and acceptance. Privacy Flag combines crowd sourcing, ICT technology and legal expertise in order to enable citizens to monitor and control their privacy with a user friendly solution provided as a smart phone application and a web browser add-on. In this paper, firstly, we focus on collective protection frameworks and tools that intend to address the arising challenges with respect to citizen awareness over data protection. Then, we present the Universal Privacy Risk Area Assessment Methodology, as well as the main functionalities of two Privacy Flag tools we developed; the *browser add-on* and the *smartphone application*.[1]

KEYWORDS

Privacy, data protection, browser add-on; smartphone application

1 INTRODUCTION

Personal data have become merchandisable asset encouraging various stakeholders to collect such data and trade them without the end-user's awareness and acceptance. In response, personal data protection is becoming a challenge both in terms of privacy and exploitation in 5G and IoT environment. The European Union (EU) has taken the lead in adapting the legal framework to better protect the citizens' rights and interests. However, the extent of the Internet and smart phone applications, the fact that data can be retrieved without the owner's knowledge and the vast majority of those applications are developed from outside the EU jurisdiction, strongly limit the possibility to effectively impose a privacy-protection framework globally with a conventional approach. Moreover, privacy norms are perceived as "complex" by many citizens.

Privacy is a complex and evolving concept. The perception of privacy may vary from one society to another, from one period of time to another, and from one individual to another [1]. Under a broader consideration, privacy is the right to respect for a person's private and family life, his home and his correspondence. Several researches have "highlighted" the wide and multidimensional nature of privacy concerns (such as, for example, the concepts proposed in [2], [3]). Privacy has essentially a certain level of uncertainty: It is simultaneously a universal concern combined with different understandings, which may vary from one country to another, as well as from a domain of activity to another. This "duality" is reflected by a rather large number of international and regional conventions protecting privacy as well a certain level of heterogeneity among the national laws.

The rise of new businesses, new architectures and new technologies through 5G (and IoT in particular [4]), can lead to a multiplicity of important challenges for security and privacy protection. In order to assure the necessary compliance, any related requirements may also impose strong constraints on networks and service platforms. 5G aims to help European citizens to manage their personal data, tune their exposure over the Internet and protect their privacy. Moreover, modern 5G design works should guarantee a high flexibility and expected to be driven by a service-like approach. The network should be flexible and quick to adapt to a broad range of usage requirements and offer converged services preserving security and privacy across a versatile architecture with unified control of any type of ICT resources [5]. It is worth mentioning that among the fundamental KPIs proposed in the framework of modern 5G-PPP enabled research activities is also the option of enabling advanced user controlled privacy [6].

The Privacy Flag project combines the potential of crowdsourcing, ICT technologies and legal expertise to protect citizens' privacy when visiting websites, using smartphone applications, or living in a smart city. It enables citizens to monitor and control their privacy with a user friendly solution made available as a web browser add-on, and a smart phone application,- all connected to a shared knowledge database. It aims to provide a new paradigm of privacy protection combining: (i) "endo-protection" with locally deployed privacy enablers protecting the citizen's privacy from unwanted external access to their data, and; (ii) "exo-protection" with a distributed and crowd-sourced monitoring framework, able to provide a collective protection framework together with increased citizens' awareness and implicit pressures on companies to improve their privacy compliance.

In this paper, we focus on collective protection frameworks and tools that intend to address the arising challenges with respect to citizen awareness over data protection. In particular, we

PCI 2017, September 28–30, 2017, Larissa, Greece

ACM ISBN 978-1-4503-5355-7/17/09…$15.00
https://doi.org/10.1145/3139367.3139405

provide a literature overview on frameworks considering the degree of personalization and the type of information that is used, as well as tools for maintaining anonymity while browsing (Section II). We briefly present the Universal Privacy Risk Area Assessment Methodology (Section III), as well as the main functionalities of two Privacy Flag tools we developed; the *browser add-on* (Section IV) and the *smartphone application* (Section V).

2 PRIVACY-CONTROL MECHANISMS AND TOOLS

Various techniques based on reverse-engineering have been employed to shed some light on the degree of personalization and the type of information that is used. Bobble [7] is a Chrome browser extension allowing users to see how the search results that Google returns to them differ from the results that are returned to other users, by executing search queries from various world-wide vantage points under different conditions. This study reports that location-based factors introduce more inconsistencies than other profile-based factors do (i.e., search history, OS). Hannak *et al.* [8] investigate popular general e-commerce retailers and travel retailers sites. It analyses the results of searches performed by real-world users (recruited from Amazon's Mechanical Turk), as well as for synthetically generated fake accounts. Taking into account factors, such as web browser, OS, account log-in, click history, and purchase history, they reveal numerous instances of price steering and discrimination. However, it cannot be systemically proved how the specific characteristics of users trigger the personalization. Similarly, the browser extension $heriff [9] identifies price variations in a set of online stores. The prices are obtained by a set of test users surfing the web for products of their interest and are correlated based on a variety of parameters, such as location, product type, log-in, personal budget-related information, OS, and web browser. The paper concludes that prices are mainly affected by the location, product type and log-in, but the analysis does not provide a concrete explanation of how price discrimination is actually applied. On the other hand, [10] focuses on whether airline companies (not travel aggregators or brokers) practice price discrimination on the tickets sold through their websites. Although a numerous set of different profiles were investigated based on OS, web browser, personal budget-based information, cookies and location, no any clear evidence of price discrimination is found. The aforementioned approaches focus on whether personalization is applied or not by investigating a set of websites. A slightly different approach is followed in [11]. In particular, XRay tool is not used for the identification of whether a web site applies personalization; instead it is a personal Web data tracking system, which correlates designated data inputs with data outputs results; each personal account is compared with a number of shadow accounts, which contain similar (not identical) subsets of data inputs. X-ray informs users how services (e.g., Gmail, Amazon, YouTube) use their data (by sending google ads, recommended products or videos) once they have it (e-mails, products in wishing lists, visited videos). The aforementioned tools aim to make users aware of how their personal data are actually used by online services. However, even if we assume such awareness, the collected data are not always used in favor of the user, since the user has no control on how personalization is designed.

In order to keep maintaining anonymity while browsing, end users tend to adopt several obfuscation approaches. ShareMeNot [12] is a browser extension designed to prevent third-party buttons (such as Facebook's "Like" or Twitter's "tweet" button) embedded by web sites from tracking users. Tools such as Surfhidden [13] and [14] provide rotating IP address proxy servers; users send requests which are randomly routed through different proxy IP servers, in order to avoid IP-based tracking by the web sites. A different approach for avoiding data-profiling by search engines is chosen by the TrackMeNot tool [15]. TrackMeNot is a browser extension for Chrome and Firefox, which periodically submits random search queries to major search engines (AOL, Yahoo!, Google, and Bing). By submitting fake queries to confuse search engines, the user's real queries are actually covered. Tor [16] is another obfuscation technique that directs Internet traffic through a worldwide distributed network consisting of thousand relays to conceal users' location from network monitoring and traffic analysis. Tor network routes packets from a given source to a destination by establishing random paths through several relays. However, techniques like Tor were criticized due to their requirement of large number of volunteers and poor anonymity [17]. In response, [18] and AnonymousCloud [19] proposed the Cloud-based Onion Routing (COR), which employs moving onion-routing services to the cloud in order to leverage the large capacities, and robust connectivity inherent to datacenters. Tor Cloud [20] is a cloud-based anonymization service for commercial use, which implements a full-scale Tor system running on the Amazon EC2 cloud computing platform. Providing cloud-based Tor-like services for anonymization induces several technical, as well as market challenges, which are not well investigated so far.

3 UNIVERSAL PRIVACY RISK ASSESSMENT METHODOLOGY

The main backbone component of the Privacy Flag project is UPRAAM, the Universal Privacy Risk Area Assessment Methodology. The UPRAAM plays an important role in the framework of the full Privacy Flag project. It constitutes the "pivotal enabler" to translate the legal and technical risks into a methodology to be applied and implemented by the technical enablers. Thus, UPRAAM defines the way to "enable" the user to assess the level of risk that an application, a website or an IoT deployment would breach personal data protection norms. It structures the way end-user can assess this level of risk by gathering inputs for the evaluation of privacy risks.

In order to "address" the various targeted use cases in the Privacy Flag, the UPRAAM must be: (i) Rather generic & universal, in order to be successfully applied to diverse objects, including websites, smart phone applications and IoT deployments; (ii) Reliable and trustable, and; (iii) Accessible & user-friendly, with a certain level of "user centricity" in the design and fine-tuning of the methodology. As complementary requirements, we can also mention: (i) the need to encompass and address both legal and technical risks, and (ii) an optimized

"Dynamicity", with a process enabling to identify and assess any relevant risk, as well as to help mitigating it [21].

Taking into consideration UPRAAM, the Data Valuation Tool sets a way for users to evaluate their data by calculating a user score based on a set of questions regarding privacy protection. The Data Valuation Tool aims from the one side to re-use outcomes of UPRAAM and, *in parallel*, provides an input to UPRAAM regarding the user's data valuation.

The main element in the Privacy Flag Data Valuation Tool to create awareness upon data value is the comparison of ones' perspective with the average user as calculated by the crowd. To achieve this, users are asked to first fulfil a questionnaire and then compare their answers with those of the average user. Going one step further and in relation with other valuation tools (e.g. Klout – see, for example [22] and [23]), a scoring framework is introduced to calculate a user-specific score based on the answers. In the end, both a final score and a question-specific score is provided as a scaled number from 1 to 100. This score also represents a person's influence in different social media. The *Klout Score* is calculated as an aggregation of signals across several dimensions for each user (from social media, Wikipedia, etc.), creating a large feature set containing over two thousand features. Using additional weights obtained from models, the *Klout Score* generates a network or community scores and the score represents how important or less important a user value his or her data.

While the score by itself is meaningless, the score is used to create a comparison value between the user and the rest of the crowd. For the calculation of the score the following concept is used:

- *Positive score*, if the user response is towards privacy protection (e.g. will not share sensitive data), values: +1 to +4.
- *Zero score*, if the user response has no negative or positive feedback (e.g. will share his name or surname).
- *Negative score*, if the user puts himself in privacy risks (e.g. will share openly sensitive data) values: -1 to -4.

In addition to the above, and in order to create also an economic approach of the data value, users are asked to provide their perspectives regarding the economic value of their data. Both the score and especially the economic value are then passed to the system and are being used as "input" to the UPRAAM.

As a final result, the user gets a visualization of both his/her generic score in comparison with the average user, as well as insights on specific questions (including the economic valuation) to create his/her own conclusions. In addition, the user has the option to submit his/her data in Privacy Flag, through an anonymized way.

4 PRIVACY FLAG BROWSER ADD-ON

4.1 Main Functionality Description

The Privacy Flag web browser add-on is a tool that allows users to get information about potential privacy risks when browsing throughout the Internet. The add-on informs users whether a web site is considered safe or not based on the analysis conducted by the Privacy Flag back-end system; an analysis which includes both input gathered by technical enables and by exploiting the power of crowdsourcing data from end users incorporating the UPRAAM methodology. The Privacy Flag web browser add-on is one of the main points of interaction between end-users and the Privacy Flag project.

The browser add-on communicates with the Privacy Flag back-end through a web service to exchange information as for example the site a user is visiting or the evaluation he or she is providing. The back-end tackles the evaluation of the web site (through both the UPRAAM assessment and the automatic assessment) and feeds back the web browser with the information on the risks involved with a specific website. The whole communication with the back-end needs to take into account the anonymization of the user.

4.2 Functionality Workflow

Following the above high level description, the workflow for the functionality of the web browser can be summarized as follows:

Step 1: While the user browses through the sites, the browser add-on informs him of the current evaluation of the site by changing the color of its icon. The evaluation of the site comes from the existed information on the Privacy Flag back-end and contains both the automated evaluation (through the implementation of the Threat List Matrix) and the UPRAAM evaluation. If no evaluation exists, the add-on informs the user accordingly. In parallel the browser add-on starts calculating automatically possible threats (e.g. http and 3rd party cookies) and stores the information locally.

Step 2: When the user opens the pop-up menu he gets a visualization of the site classification (Privacy Friendly/Not Friendly) and has the option to provide his own evaluation by answering the UPRAAM defined questions.

Step 3: When the user submits his evaluation, a JSON file is created that contains the user answers, the automated calculated threats on the browser add-on and a unique identifier.

Step 4: Crowdsourcing Evaluation Tool experts evaluate manually the web site and submit their evaluation to the database the Local Crowdsourcing Evaluation Tool Score

Step 5: The back-end performs various calculation based on the threat matrix and in combination with forecasting epidemiology models calculates the browser add-on Local Threat Level Score.

Step 6: PF back-end decides based on:

- The browser add-on Local Threat Level Score,
- The Local Crowdsourcing Evaluation Tool Score,
- Mean Threat Level Score,
- Mean Crowdsourcing Evaluation Tool Score.

4.3 Threat Evaluation and GUI

The evaluation of the website is based on the top 25 threat matrix. For the evaluation to take place a number of the threats are implemented on the browser add-on while others are executed remotely on the PF back-end. For the implementation of the threat a number of javascript libraries are used, including chrome APIs.

The following list of threats presents the ones implemented in the add-on:

- Does the website provide data encryption (SSL/TLS)?

- What information does the website/server directly learn about a user (using forms)?
- Which communication parties is data transferred to?
- Does the website use HTML cookies?
- Does the website use third party cookies?
- Does the website use HTML5 Web SQL database?
- Does the website use LSOs?

Figure 1 depicts the GUI of the Privacy Flag browser add-on.

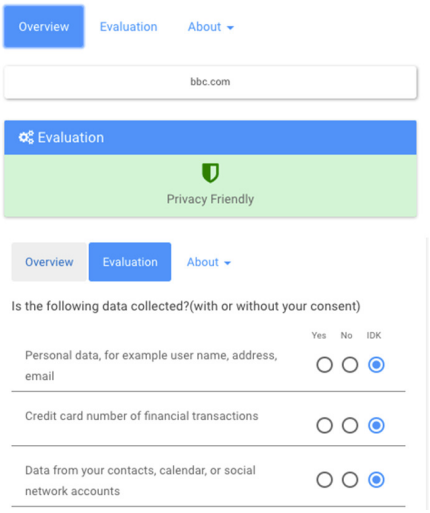

Figure 1: The Privacy Flag Browser Add-on GUI

5 PRIVACY FLAG SMARTPHONE APPLICATION

5.1 Main Functionality Description

The Privacy Flag smartphone application allows users to get information on potential privacy risks from installed applications in their Android-powered mobile phones and tablets. The application informs users whether installed software is considered privacy friendly or not friendly based on the analysis conducted by the Privacy Flag back-end system, an analysis which includes both input gathered by technical enables and by exploiting the power of crowdsourcing data from end users incorporating the UPRAAM methodology. Combined with the Privacy Flag web browser add-on, the smartphone application is one of the main points of interaction between end-users and the Privacy Flag project.

Similarly to the browser add-on, the smartphone application communicates with the Privacy Flag back-end through a web service to exchange information related to the already installed applications. The back-end tackles the evaluation of the apps (through both the UPRAAM assessment and the automatic assessment) and feeds back the application with the information on the risks involved with a specific application.

5.2 Functionality Workflow

Following the needs of the project, the functionality of the application was updated as follows:

Step 1: When first installed and opened, the application asks users to submit for a first and only time their preferences in regards to privacy and the user permission. The user's answers and main information are passed to the backend through a JSON format file, in order to include this in the following calculations.

Step 2: The smartphone application reads the user's installed applications (and checks for new or updated ones if this is not the first time the user uses the PF application) and sends the list of applications to the backend in order to retrieve their evaluation (privacy friendly, not friendly or not evaluated).

Step 3: The users browse among the installed applications and can view its evaluation and the user set permissions that have been given to the app.

Step 4: The smartphone evaluations contributors manually evaluate applications and submit their evaluation to the database. This is called the Crowdsourcing Evaluation Tool score. In addition the user set permissions of the application are also transmitted to the backend

Step 5: The Privacy Flag backend performs various calculations based on Artificial Intelligence and Machine Learning algorithms. It also employs advanced statistical and epidemiological models to detect outliers (applications with vastly different Threat Level Scores) which indicate possible data leakage. The outputs of the database calculations are the Mean Threat Level Score and the Mean Crowdsourcing Evaluation Tool Score.

Step 6: The PF backend decides based on:

- The smartphone application's Local Threat Level Score,
- The Local Crowdsourcing Evaluation Tool Score,
- Mean Threat Level Score,
- Mean CET Crowdsourcing Evaluation Tool Score.

5.3 Threat Evaluation and GUI

The automated evaluation is related to the user set permissions of an application. Those may include body sensors, calendar, camera, contacts, location, microphone, phone, SMS, and storage.

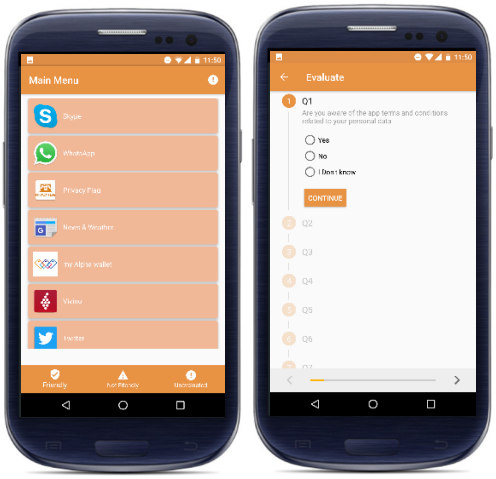

Figure 2: The Privacy Flag Smartphone Application viewing specific applications (left) and providing user evaluation (right).

Each time an application evaluation is submitted, the permissions that are given to this application are also passed on to the backend in order to be included in the evaluation process.

Figure 2 depicts the GUI of the Privacy Flag smartphone application

6 CONCLUSIONS

Crowdsourcing is a model capable of aggregating talent, leveraging ingenuity while reducing the costs and time formerly needed to solve problems. Additionally, the rise of crowdsourcing correlates with the rise of the Internet and web technologies and it is enabled only through the technology of the web, which is a creative mode of user interactivity, not merely a medium between messages and people. The above-mentioned activity is an emerging and promising model for information gathering and problem solving that is already transforming industry and scientific practices, allowing researchers to access to human resources in a scope that was not possible before. As the use of crowdsourcing is increasing in many sectors, it opens up new horizons for scholars in the field of communications, technology and other sciences.

The actual Privacy Flag EU-funded research effort is built on a crowdsourcing model to design and select the best solution to protect their privacy and data ownership. Crowd sourcing dramatically increases the number of potential "*good ideas*" provided by a large number of sources and to filter solution that would not be accepted by the end-users. Privacy Flag should enable the "*wisdom of the crowd*" to improve privacy and data ownership. Furthermore an effective set of tools should be promoted to identify risks and prevent data misuse using viral dissemination to produce increases in awareness, thus serving project's specific aims.

In this paper, we investigated collective protection frameworks and tools intending to address the arising challenges with respect to citizen awareness over data protection. Additionally, we presented the main functionalities of the two Privacy Flag tools we developed; the *browser add-on* and the *smartphone application*.

ACKNOWLEDGMENTS

This research activity has been funded by the EU-funded "Privacy Flag" project, H2020, Grant Agreement (GA) No.653426.

REFERENCES

[1] K. Renaud K, G´a andlvez Cruz D., "Privacy: Aspects, definitions and a multi-faceted privacy preservation approach". Information Security for South Africa (ISSA), pp. 1-8, 2010.

[2] W. Hong, and J. Y.L. Thong, "Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies", MIS Quarterly, vol.37(1), pp.275-298, 2013.

[3] A. Calder, "Cyber Security: A Critical Business Issue". IT governance green paper, 2013 (August).

[4] J.H. Ziegeldorf, O.G. Morchon, and K. Wehrle, "Privacy in the Internet of Things: Threats and Challenges". In Proceedings of the Security and Communications Networks 2014 Conference. Cornell University Library, 2015.

[5] European Commission and 5G-PPP, "5G Vision - The 5G Infrastructure of Public Private Partnership: the next generation of communication networks and services", 2015 (February).

[6] 5G – Public Private Partnership (5G-PPP), "Advanced Network Infrastructure for the Future Internet", 2013 (May).

[7] Xing, W. Meng, D. Doozan, N. Feamster, W. Lee, A. Snoeren, "Exposing Inconsistent Web Search Results with Bobble". 15th International Conference PAM, 2014.

[8] A. Hannak, et al., "Measuring Price Discrimination and Steering on E-commerce Web Sites". IMC '14, 2014.

[9] J. Mikians, L. Gyarmati, V. Erramilli, N. Laoutaris, "Crowd-assisted Search for Price Discrimination in E-Commerce: First results". 9th CoNEXT, 2013.

[10] T. Vissers et al., "Crying Wolf? On the Price Discrimination of Online Airline Tickets". 14th PETS, 2014.

[11] M. Lecuyer, et al., "xRay: Enhancing the Web's Transparency with Differential Correlation". 23rd USENIX Security Symposium, 2014.

[12] ShareMeNot website, http://sharemenot.cs.washington.edu/.

[13] Surfhidden website, http://www.surfhidden.com/.

[14] HideMe website, http://hideme.be/.

[15] D.C. Howe and H. Nissenbaum, "TrackMeNot: Resisting surveillance in web search". Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society. 2009.

[16] R. Dingledine, N. Mathewson, P. Syverson, "Tor: The Second-Generation Onion Router". Proceedings of the 13th conference on USENIX Security Symposium (SSYM '04), 2004.

[17] A. Panchenko, L. Niessen, A. Zinnen, T. Enge, "Website fingerprinting in onion routing based anonymization networks". 10th ACM WPES '11, 2011.

[18] N. Jones, et al., "Hiding Amongst the Clouds: A Proposal for Cloud-based Onion Routing". USENIX 2012.

[19] S.M. Khan, K.W. Hamlen, "AnonymousCloud: A Data Ownership Privacy Provider Framework in Cloud Computing". 11th IEEE TrustCom, 2012.

[20] Tor Cloud project website, https://cloud.torproject.org/, 2012.

[21] S. Ziegler, I.P. Chochliouros, and L. Ladid, "Privacy Flag – Collective Privacy Protection Scheme Based on Structured Distributed Risk Assessment". In Proceedings of the IEEE World Forum on Internet of Things (WF-IoT), Milano Italy, (December 14-16, 2015).

[22] A. Rao, N. Spasojevic, Z. Li, and T. Dsouza, "Klout Score: Measuring Influence across Multiple Social Networks". In: Proceedings of the 2015 IEEE International Big Data Conference Workshop on Mining Big Data in Social Networks, pp. 2282-2289. Santa Clara, CA, USA, 2015.

[23] M. Schaefer, Return on influence: The revolutionary power of klout, social scoring, and influence marketing". McGraw-Hill Professional, London, 2012.

[24] 5G-PPP framework website, https://5g-ppp.eu/, 2016.

[25] European Commission and 5G-PPP, "5G Vision: The 5G-PPP Infrastructure Private Public Partnership: The Next Generation of Communication Network and Services", 2015.

[26] H.C. Pöhls, V. Angelakis, S. Suppan, K. Fischer, G. Oikonomou, E.Z. Tragos, R.D. Rodriguez, T. Mouroutis, "RERUM: Building a Reliable IoT upon Privacy- and Security- enabled Smart Objects". In Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW), pp. 122-127. Istanbul, Turkey, 2014.

[27] R. Roman, J. Zhou, J. Lopez, "On the Features and Challenges of Security and Privacy in Distributed Internet of Things". Computer Netwoks, vol. 57(10), pp. 2266-2279, 2013.