Transparency through Contextual Privacy Statements

Denis Feth

denis.feth@iese.fraunhofer.de
Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern, Germany

Abstract

Privacy policies are the state of the practice technique to achieve data transparency. However, they have a variety of issues in practice: They are presented in a non-prominent way, are typically quite lengthy, and not written in the users' language. Additionally, they are quite abstract, as privacy policies are generic documents that do not relate to the current activity of the user but give a high level overview on the overall system. In this paper, we present our idea of "contextual privacy statements" that overcome the shortcomings of state of the practice privacy policies. Instead of having one generic privacy policy that has to fit every use case and every user group, contextual privacy statements provide concrete information about privacy and data protection in a specific use case or activity. We aim for better understandability of privacy policies, resulting in an increased transparency and user acceptance.

1 Introduction

1.1 Problem

Modern IT systems and services are getting more and more customized and aligned to the user. However, this comes along with massive collection, processing, and potentially sharing of sensitive data. For example in Smart Homes, one household can easily produce more than 15,000 (potentially sensitive) data points every day (Federal State Commission2015). Because of that, many users have privacy concerns when using online and cloud services (IControl Networks2015).

For industry, these privacy concerns and the lack of trust users have in their services are a major problem. According to (Wirtz et al.2007), "increased concern resulted in higher power-enhancing responses such as the fabrication of personal information, use of privacy-enhancing technologies and refusal to purchase." To pose it simple: Even if a system is

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2017 in
M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.):
Mensch und Computer 2017 – Workshopband, 10.–13. September 2017, Regensburg.
 https://doi.org/10.18420/muc2017-ws05-0406

highly secure and privacy-preserving, we also need to convince the users. Thus, transparency is already perceived as unique selling point for many service providers.

In addition to the users increasing demand for privacy, data protection laws (e.g., EU GDPR) are changing as well. They demand the implementation of aspects like data sovereignty, explicit consent and transparency—all of which will only be achievable (respectively meaningful) if users understand how their privacy is protected.

As a state of the practice solution, almost every service we are using provides a privacy or data policy in order to achieve a certain level of transparency (as demanded by legal obligations). These privacy policies are legally demanded and contain privacy statements, describing the way a service provider gathers, uses, disclosed or manages user data. However, in their current form, privacy policies typically have major shortcomings:

1. They are not presented in a prominent way (Ermakova et al.2015).
2. They are lengthy (Milne et al.2006). In a study by (Waldman2016), the average word count was 2,716 words. According to (Cranor2012), users would need 244 hours per year in average to read the privacy policy of every website they visit.
3. They are not written in the users' language. Instead, they are mostly written by and written for lawyers (Waldman2016).
4. The statements in the policy are not verifiable by users. In the end, the user has to decide whether he believes that the provider adheres to the privacy policy.
5. They are abstract and generic. There is only one single policy for the whole service. It remains the users' task to map this policy to the current activity or data he is dealing with.

All of these issues result in a high mental load, which users are typically unable or unwilling to spend. Even if they are willing to spend high effort, some aspects are still not possible to understand. For example, if the policy states that "data is shared with third party providers", it is not possible to tell whether a specific data field (e.g., an address) entered in a form is affected. An example for that is Android's permission system that presents the app permissions (which are similar to privacy policies to a certain extent) at installation time. At this time, the user cannot assess the permissions as all the mentioned shortcomings apply. As a consequence, only 17% of the users paid attention to this information and only 3% understood the information that was presented (Felt et al.). In newer Android versions, permissions are (in addition) requested directly in the context of use (e.g., when an app accesses the microphone).

Summarizing, current privacy policies are not designed in a way so that they are suited for users to achieve transparency and meet requirements demanded by law (such as the EU General Data Protection Regulation). In a survey by Obar and Oeldorf-Hirsch (Obar and Oeldorf-Hirsch2016) it is stated that 74% of users skipped the privacy policy completely. For the remaining 26% of the users, the average reading time was only 73 seconds.

1.2 Contribution

Based on the problem stated above, we aim for "Contextual Privacy Policies" that overcome the major shortcomings listed above. We present our ideas for a privacy policy meta model, the alignment of privacy policies to user activities and the integration and representation of contextual privacy statements in software.

2 Contextual Privacy Policies

2.1 Idea

In general, a privacy policy is a collection of (privacy) statements that describes how a party (e.g., a service provider) gathers, uses or discloses a client's data. We define a privacy policy to be contextual, if the following conditions apply:

- The privacy statements are explicitly relating to the current activity of the user. A bad example would be: "We share certain data with third parties". A good example is: "The address data entered in this form is shared with advertisement provider XYZ in order to regularly send you info material"
- The privacy statements are shown "in context", i.e., specific privacy statements are shown to the user in situations, when data is entered, processed or presented. The contextual display of privacy statements does not force the user to pause his current activities in order to research about affected privacy statements in the current context. In the bad case, one abstract policy document is linked on the bottom of a website. In a good example, concrete information about data usage shown closely to the interaction (e.g., to a form).
- The abstraction level depends on the user group, in order to make the statements understandable. For user A, the following statement is suited: "Data transfer is protected by SSL 2.0, passwords will be hashed and salted with SHA1 and data stored AES512 encrypted". For user B, this phrasing might be too technical, and he prefers: "We handle your data confidentially. All data you entered in the form will be encrypted during transit from your device to our servers and in our databases".

By binding privacy statements to the current activity, we kill several birds with one stone. The privacy statements can be more specific and explicitly relate to the current activity (fifth shortcoming). We assume that this increases understandability. In addition, the individual privacy polices presented to the user become shorter, as only currently relevant information is shown (second shortcoming). This increases the likelihood that users take time to read the privacy statements. At the same time, we need to find a good middle way so that contextual privacy policies are not too intrusive, but still noticeable enough (first shortcoming).

2.2 Example

Imagine, you are using an online/cloud service “example.com” that provides free e-books that can be read online.

2.2.1 As-Is Situation

First visit: When you visit the website, you are notified that cookies are used to provide the services. This notification does not give you control, it just informs you that cookies are used. In many cases, cookies are already set when user’s see the popup for the first time. These cookie popups become more and more common—but are still used in less than 20% of the websites (Waldman2016).

Registration: To use the service, you need to register. Therefore, you enter personal data, like name, email address and so on. Before confirming the registration, you have to accept the terms and conditions and the privacy policy. This can be done by ticking the corresponding check boxes. Via a link, you can read the complete privacy policy document in advance.

Usage: At usage time, the privacy policy is accessible via a link on the bottom of the pages. Besides that, you are not informed about privacy-related activities anymore.

2.2.2 To-Be Situation

If privacy-related data (personally identifiable information) is processed in a certain activity the contextual privacy policy is shown and explicitly relating to the current activity.

First visit: The information about the usage of cookies is shown before the cookie is set. Also, the content and purpose of each cookie is explained to the user.

Registration: For example, during registration time, the user is informed what happens with his name, email address and password. Additionally, he is informed that his IP address was collected. At this point, it is for example not important that reading times will be logged (cf. Figure 1).

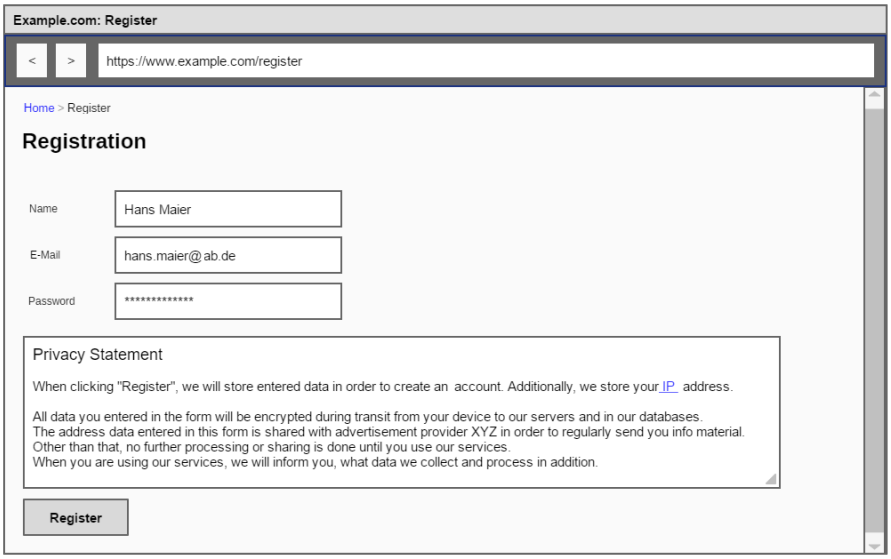

Figure 1- Contextual Privacy Policies – Registration

Usage: Instead of a static link to the generic privacy policy, the page displays the contextual policy, explaining that title, reading time and end-device type was collected. Additionally, he sees that the title is shared with advertisement providers (cf. Figure 2).

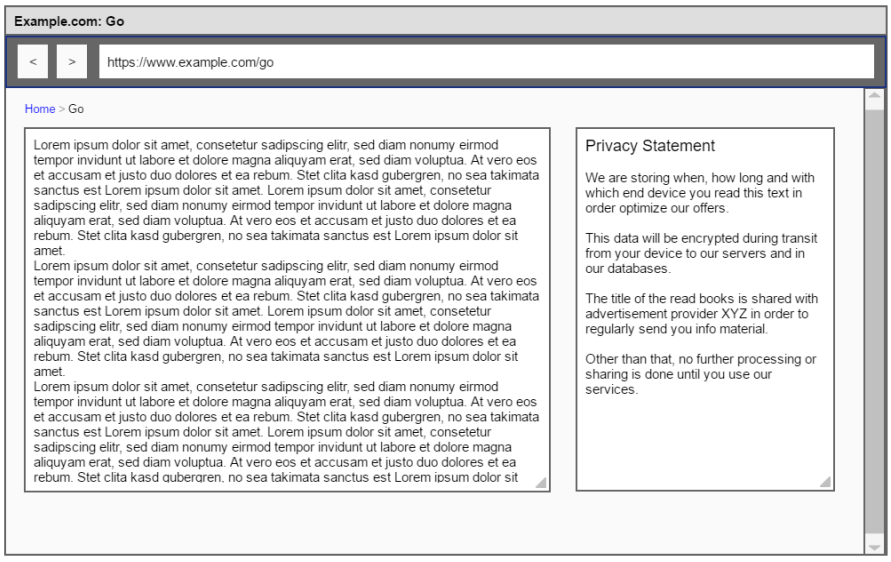

Figure 2- Contextual Privacy Policies – Usage Time

2.3 Implementation

There are three major challenges that have to be solved in order to implement contextual privacy policies. For each of the topics, we describe the fundamental research question and the rationale behind. Then, we describe our solution approach and evaluation plans.

2.3.1 Privacy Policy Meta Model

Question: Which elements, relations, constraints, etc. are relevant for a privacy policy?

Rationale: Although privacy policies contain domain and application specific information, the basic elements and relations are similar. Thus, we can build a privacy policy meta model. This helps us to understand the content of state of the practice privacy policies and is the basis for automated processing at runtime. For example, privacy policies can be tailored according to the current situation or the current user group using the instantiated models.

Approach: There is a variety of different stakeholders interested in privacy policies, especially the users, but also lawmakers and the service providers themselves. For each of the stakeholders, or stakeholder groups, the privacy policy has to be understandable. We build a privacy policy meta model that can be used to instantiate all aspects that are relevant for mapping and describing privacy protection in a system. We do that by extracting information from a set of real privacy policies (e.g., Facebook). Additionally, we consider preliminary work about privacy policy perception and categorization, e.g., on www.usableprivacy.org. Finally, laws and regulations (e.g., EU-GDPR) are analyzed to identify obligatory elements required by lawmakers.

Evaluation: While formal proofs are rather impossible for a meta model like the proposed, we evaluate our meta model empirically. By checking whether privacy policies that were not in the "learning set" can also be instantiated with the meta model, we show completeness. To show correctness, we compare the model with the initial privacy policy and show that all statements in the model instance are also part of the policy and no additional statements are included.

2.3.2 Contextualization of Privacy Policies

Question: How can privacy policies be tailored to the current activity (i.e., made contextual)?

Rationale: To overcome the problem of generic and abstract policies, contextual privacy policies are shown in context of the current activity. The hypothesis is, that they drastically decreases the user's effort (compared with reading a complete state of the practice privacy policy), while increasing transparency at the same time.

Approach: First, we need to define what privacy relevant activities are. As this can be very generic ("everything that triggers the execution of personal data processing"), or limited to a specific scenario and technology. Second, the meta-model need to reflect identified activities and it must be possible to tailor down the instantiated model to the current activity. Third, we

map abstract activities to concrete privacy-relevant behavior in the code. For that, we provide an SDK that allows developers to tag these parts.

Evaluation: The evaluation is done mainly with respect to completeness and correctness of the mapping. For the current activity, the relevant model parts needs to contain all relevant information for the activity (completeness). On the other hand, it should not contain additional, irrelevant information (correctness). We validate that the mapping function includes only related and relevant elements. In addition, we can verify that users have all relevant information included using different case studies.

2.3.3 Integration and Representation

Question: How can the machine-readable model be transformed into a form that is understandable by all user groups?

Rationale: The model is instantiated manually or semi-automatically and reflects relevant elements and relationships. However, it is not understandable by users. There need to be some form of transformation and/or abstraction step in order to have a suitable representation for the end user.

Additionally, we cannot assume that user groups are homogeneous. Depending on the user group (which needs to be identified), the abstraction level of the privacy policy has to be different in order to be useful for the user. For example, IT security experts expect more technical details and another terminology compared to inexperienced users. As we cannot assume to know the skills of a particular user in advance, a staging approach can be used. We start presenting policies in the highest abstraction level. If a user revealingly wants to learn more about policies, subsequent policies are directly shown in more detail.

In addition to the abstraction level, also the visualization is an important aspect. On the same abstraction level, there exist a variety of possibilities to visualize the information. Text, diagrams, or pictograms are examples. However, within an abstraction level, different visualizations should transfer the same information, i.e., they form an equivalence class in that respect. For example, in its privacy check[1], Google uses pictograms to support the user to understand what a section is about. For contextual privacy statements, the major challenge with respect to visualization is to find an appropriate way to show the policies, while not annoying or interrupting the user. For example, if a user repeats a certain activities many times and the user read the policies already, they should be less prominent.

Approach: To find the right form of representation, we analyze different design patterns from literature and existing privacy policies and setting pages. With user studies, we can evaluate which pattern fits best for typical user groups. Technology-wise, we integrate the representation module in an SDK, which enables developers to easily integrate privacy

[1] https://myaccount.google.com/privacycheckup

information screens / dialogs. We focus on web and cloud services and will tailor our SDK to these technologies.

Evaluation: Feasibility is shown in form of case studies using the SDK. Understandability of the provided information has to be shown in user studies. Questionnaires can be used to test if the users understood the privacy policies.

3 Related Work

Since the mid-1990s, huge efforts have been made to develop approaches for aligning usability and security. Unfortunately, the number of security incidents caused by unusable security measures or usable, but insecure systems is still high (Furnell2007). In (Garfinkel and Lipford2014) Garfinkel and Lipfort summarize the history and challenges of the "usable security" domain.

Existing literature on usable security shows that the user is an important and active part of modern security chains. The research field of usable security and privacy has been approached both in a theoretical fashion and in the form of case studies. Famous case studies analyze the usability of email encryption with PGP (Whitten2004, Whitten and Tygar1999), of file sharing with Kazaa (Good and Krekelberg2003), and of authentication mechanisms and password policies (Choong and Theofanos2015, Eljetlawi and Ithnin2008, Inglesant and Sasse2010). However, case studies are specific to one system, system class, or application domain and can hardly be generalized. On the other hand, theoretical work (Adams and Sasse1999, Cranor and Garfinkel2005) is typically more abstract and hard to apply in practice.

This gap is closed by design principles for usable, yet secure systems (Garfinkel2005, Good and Krekelberg2003, Whitten and Tygar1999). These principles focus on the development of usable security systems by supporting developers and emphasizing the importance of considering the user. However, they do not adopt the user's viewpoint or active involvement of users in the development process. However, it is crucial considering both the user's viewpoint and to involve users in the development process.

The acceptance of privacy policies by end users, as well as the consequences of missing acceptance have been analyzed in different surveys (Tsai et al.2007, Obar and Oeldorf-Hirsch2016, Symantec2015). These studies showed that privacy policies are mostly unsuited for users in practice. To improve the situation, there exists work targeting readability (Ermakova et al.2015, Milne et al.2006), understandability (Reidenberg et al.2014) and design (Waldman2016) of privacy policies. These are important aspects, but all of these works consider a privacy policy to be a large monolithic document, which is in contrast to our considerations. Tools like the Platform for Privacy Preferences P3P Project[2] targeted

[2] https://www.w3.org/P3P/

transparency of web service privacy, but were not well accepted for usability reasons. In contrast to P3P, we aim to integrate contextual privacy policies into the applications (e.g., web application), rather that the platform (e.g., browser), which allows for a better tailoring, visualization and support of different browsers.

4 Summary

Privacy policies are an established approach for achieving transparency. However, the policies are hardly usable in practice. Due to their length, language and abstraction level, the mental load required to read understand them is simply not appropriate.

In this paper, we recommend to extend state of the practice privacy policies by providing contextual privacy statements. This means that concrete information about collection, using and sharing of information is shown at the time a user performs a certain activity. However, there exist a variety of legal requirements for privacy policies, e.g., the German Telemediengesetz. It remains to be verified, whether our contextual privacy policies can replace traditional policies, or if they are an add-on. However, we expect a major increase in transparency. This is primarily desirable for users, but also for most service providers, as a reduction of concerns and an increase of trust will lead to a higher acceptance of the systems or services.

5 Acknowledgements

The research presented in this paper is supported by the German Ministry of Education and Research (BMBF) project Software Campus (grant number 01IS12053). The sole responsibility for the content of this document lies with the authors.

6 References

Adams, A. and Sasse, M. A. (1999). Users are not the enemy. *Communications of the ACM*, 42(12):40–46.

Choong, Y.-Y. and Theofanos, M. (2015). What 4,500+ People Can Tell You - Employees' Attitudes Toward Organizational Password Policy Do Matter. In Tryfonas, T. and Askoxylakis, I., editors, *Human Aspects of Information Security, Privacy, and Trust*, pages 299–310. Springer International Publishing.

Cranor, L. (2012). Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal of Telecommunications and High Technology* , 10:273–307.

Cranor, L. and Garfinkel, S. (2005). *Security and Usability*. O'Reilly Media, Inc.

Eljetlawi, A. M. and Ithnin, N. (2008). Graphical password: Comprehensive study of the usability features of the recognition base graphical password methods. *Proceedings - 3rd International Conference on Convergence and Hybrid Information Technology, ICCIT 2008*, 2:1137–1143.

Ermakova, T., Fabian, B., and Babina, E. (2015). Readability of Privacy Policies of Healthcare Websites. *Wirtschaftsinformatik*.

Federal State Commission (2015). IoT Privacy & Security in a Connected World.

Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., and Wagner, D. Android Permissions: User Attention, Comprehension, and Behavior.

Furnell, S. (2007). Making security usable: Are things improving? *Computers & Security*, 26(6):434–443.

Garfinkel, S. and Lipford, H. R. (2014). Usable Security: History, Themes, and Challenges. *Synthesis Lectures on Information Security, Privacy, and Trust*, 5(2):1–124.

Garfinkel, S. L. (2005). Design Principles and Patterns for Computer Systems That Are Simultaneously Secure and Usable. *Gene*, 31:234–239.

Good, N. S. and Krekelberg, A. (2003). Usability and privacy. *Proceedings of the conference on Human factors in computing systems - CHI '03*, (5):137.

IControl Networks (2015). 2015 State of the Smart Home Report. Technical report.

Inglesant, P. and Sasse, M. (2010). The true cost of unusable password policies: password use in the wild. pages 383–392.

Milne, G. R., Culnan, M. J., and Greene, H. (2006). A Longitudinal Assessment of Online Privacy Notice Readability. *Journal of Public Policy & Marketing*, 25(2):238–249.

Obar, J. A. and Oeldorf-Hirsch, A. (2016). The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. In *The 44th Research Conference on Communication, Information and Internet Policy 2016*.

Reidenberg, J. R., Breaux, T., Carnor, L. F., French, B., Cranor, L. F., Grannis, A., Graves, J. T., Liu, F., Mcdonald, A., Norton, T. B., Ramanath, R., Russell, N. C., Sadeh, N., and Schaub, F. (2014). Disagreeable Privacy Policies: Mismatches Between Meaning and Users' Understanding. *Berkeley Technology Law Journal*, 30.

Symantec (2015). State of Privacy Report 2015.

Tsai, J., Egelman, S., Cranor, L., and Acquisti, A. (2007). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study.

Waldman, A. E. (2016). Privacy, Notice, and Design.

Whitten, A. (2004). Making Security Usable. *Computers Security*, 26(May):434–443.

Whitten, A. and Tygar, J. (1999). Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, page 14. USENIX Association.

Wirtz, J., Lwin, M. O., and Williams, J. D. (2007). Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management*, 18(4):326–348.

7 Author

Feth, Denis

Denis Feth holds a Master in Computer Science from Technical University of Kaiserslautern. He is doing research on data usage control at Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern. Additionally, he is a PhD student at Technical University of Kaiserslautern. In his PhD topic, he is researching on usable security and privacy mechanisms with a focus on transparency and understandability. To this end, he is also participating in the German UPA work group "Usable Security and Privacy".