Privacy Risk Analysis to Enable Informed Privacy Settings

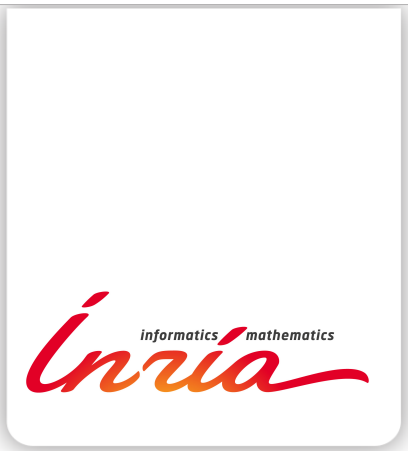
Daniel Le Métayer, Sourya Joyee De

Privacy Risk Analysis to Enable Informed Privacy Settings

Sourya Joyee De, Daniel Le Métayer

RESEARCH REPORT
N° 9125
November 2017
Project-Team Privatics

ISSN 0249-6399 ISRN INRIA/RR--9125--FR+ENG



Privacy Risk Analysis to Enable Informed Privacy Settings

Sourya Joyee De, Daniel Le Métayer

Project-Team Privatics

Research Report n° 9125 — November 2017 — 24 pages

Abstract: The work described in this paper is a contribution to enhancing individual control over personal data which is promoted, *inter alia*, by the new EU General Data Protection Regulation. We propose a method to enable better informed choices of privacy preferences or privacy settings. The method relies on a privacy risk analysis framework parameterized with privacy settings. The user can express his choices, visualize their impact on the privacy risks through a user-friendly interface, and decide to revise them as necessary to reduce risks to an acceptable level.

Key-words: Privacy risk analysis, privacy setting, fitness tracking system, harm tree, personal data

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Analyse de risques pour guider les choix de paramètres de protection de vie privée

Résumé : L'objectif du travail décrit dans cet article est de renforcer le contrôle des individus sur leurs données personnelles. Nous proposons une méthode permettant d'effectuer des choix de paramétrage de politiques de protection de vie privée informés par le résultat d'une analyse de risques. L'utilisateur peut exprimer ses choix et visualiser leurs conséquences en termes de risques, puis les réviser si nécessaire dans un processus itératif, jusqu'à obtention d'un niveau de risque acceptable.

Mots-clés : Analyse de risques en vie privée, politique de vie privée, donnée personnelle, contrôle, arbre d'attaque

1 Introduction

Users reveal a lot of personal data to various websites and service providers. Even if data controllers must, in most cases, obtain their consent before collecting their data, it is well known that this consent is more a formal right than a true protection. The main reason is that data subjects do not have the time and the expertise (and sometimes even the motivation) to read and understand the general terms of use or privacy policies of the data controllers. A first step towards more balanced relations between data controllers and data subjects is to make it possible for data subjects to express their own privacy requirements. Different languages have been proposed to express privacy policies and some services offer a great deal of flexibility for users to define their privacy settings. For example, major online social networks allow users to choose whether the information they post or share should be visible to only friends, friends of friends or the public. Some of them also offer the possibility of drawing up a customized list of audiences for different posts. Smartphone operating systems also provide ways to grant or deny to applications permissions to access different types of data (such as contact lists, call lists). In the same spirit, various fitness tracking products have features, such as heart rate tracking, which can be turned off by the user [7]. When faced with such options, users may not fully understand the implications of the choices they make as they have limited expertise and no means to compare their effects [8].

Ideally, users' choices should be based on a clear appraisal of the risks and benefits of the available options. Users can generally have a more or less precise idea of the benefits, which may include for example the possibility for friends (or friends of friends) to be aware of some of their activities, or the possiblity for themselves to receive information about special offers from online trading sites. Privacy risks, however, are more difficult to assess. The advent of the Internet of Things (IoT) and the development of quantified self, smart homes and smart cities, further aggravate this problem. Transparency enhancing technologies (TET) aim to address these difficulties by providing data subjects with information on how their data is stored, processed, shared and used [21]. However, to our best knowledge, existing tools do not help users assess the impacts of their choices in terms of privacy risks.

The EU General Data Protection Regulation (GDPR) [13] emphasizes control over personal data[1] and states that data subjects should be made aware of the risks related to personal data processing[2]. It also mandates data protection impact assessments (DPIA) for certain categories of personal data processing[3]. Privacy risk analysis (PRA) methodologies proposed so far are primarily meant to be used by data controllers to assess the privacy risks of their products or services. In this paper, we propose a method, based on privacy risk analysis, to help users to understand the privacy risks that may result from their choices of privacy settings. Our work relies on a privacy risk analysis methodology proposed in [11,12]. The core of the approach is the construction and analysis of harm trees derived from information about the system, the personal data involved, the relevant risk sources, the feared events and their impacts in terms of privacy. The methodology is extended to take into account the privacy settings of the users and analyze their impact on the likelihood of the privacy harms.

To illustrate our approach, we use as a case study a quantified self application. Quantified self is chosen both because of its fast growth and for the various privacy risks that such systems may pose to their users [14, 16, 19, 25, 26]. Fitness tracker devices (e.g., Fitbit) allow their users to track their number of steps, strenuous activities, heart beats and location. They also

[1] For example, Recital 7 states that "Natural persons should have control of their own personal data".

[2] For example, Recital 39 states that "Natural persons should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing."

[3] Article 35: Data protection impact assessment.

provide users different types of derived information such as sleep patterns, calories burnt or goals achieved through a comprehensive dashboard. For the rest of this work, we consider a high-level specification of a fitness tracking system inspired by existing products, but we focus on a limited subset of functionalities for the sake of conciseness.

One of the desirable features of TETs targeted at data subjects is that they should be very user-friendly, with an easy-to-understand presentation of information about the privacy implications of different actions and possible choices [18]. To address these needs, we also propose a user interface through which the user can easily communicate to the service provider his preferences and can visualize and rationalize the impact of his choices on the likelihood of the privacy harms.

We begin with a description of the basic background on privacy risk analysis (illustrated with our case study) in Section 2, followed by a discussion of what we mean by user privacy preferences in Section 3. In Section 4, we discuss the design of a user-friendly, interactive interface that enables users to express their preferences and understand the privacy risks resulting from them. In Section 5, we "lift the hood" and present the engine used to compute privacy risks. Finally, we discuss related works in Section 6 and conclude with perspectives in Section 7.

2 Preliminaries

In this section, we present the terminology used in the rest of the paper and illustrate it with our case study. We underline that the technical terms and notions presented in this section (including harm trees) are useful to the reader but do not have to be known by the users. Users interact with the system only through the interface presented in the next section, which hides all technicalities.

2.1 Definition of the System

A fitness tracking service consists of a fitness tracking device TD for each user i. This device collects fitness data fit_i and location data loc_i. This data is then forwarded to the service provider to be stored and processed. Apart from owning the device itself, the user generally needs to create a personal user account UA where he must provide identification ID_i and other information. The user must authenticate himself using his identification (ID_i) and password (pwd_i) to access his account. The fitness device owned by the user is linked to UA.

The system provides comprehensive information about the level of fitness of the user through a personalized dashboard accessible through the user account. The service provider uses fit_i and loc_i to derive this fitness related information $dfit_i$ (e.g. calories burnt, sleep pattern, active minutes and distance covered). Users can also maintain a list of contacts, share his data ($dfit_i$ and loc_i) with them and see the data shared by the contacts.

For simplicity, we assume that the service provider manages the application server AS where all data processing takes place and the database server DS where all data are stored.

2.2 Definition of data

We assume that some data such as fit_i and $dfit_i$ may or may not be associated with the identity ID_i and we use the notation $\overline{x_i}$ to denote the pair (x_i, ID_i) for the sake of conciseness. For example, the service provider may always store and process $\overline{fit_i}$ but give access to only fit_i to a third party.

The database server DS, which is a persistant storage, stores most types of data for user i in an encrypted format ($\overline{efit_i}$, $\overline{edfit_i}$, $\overline{eloc_i}$). It also stores the cryptographic keys k, k' and the

passwords pwd_i. Password-protected data ($\overline{pdfit_i}$ and $\overline{ploc_i}$) are accessible through UA. Since the user knows the password, he can access $\overline{dfit_i}$ and $\overline{loc_i}$ through UA. He can access data that dates back to one year or one week (depending again on the retention time).

Data processing takes place in the application server AS. The tracking device TD usually stores data for a short period of time (for e.g., seven days) before it starts losing older data. In the sequel, this type of storage is called *transient*. In both AS and TD, data is stored in encrypted format for a short duration.

The service provider ensures that all data are suitably protected by appropriate encryption and access control mechanisms. He also ensures the protection of cryptographic keys and passwords.

2.3 Definition of the risk sources

Risk sources[4] either intentionally or unintentionally, legally or illegally cause privacy breaches [11]. We consider the following risk sources for our case study: the system owner or service provider (A.1), friends of the user (A.2), hackers (A.3), the general public (A.4) and business partners of the service provider[5] (A.5).

2.4 Definition of the privacy harms

Fitness service providers may sell identifiable (or improperly de-identified) data to third parties such as health insurance providers who may use the user's fitness data to increase health insurance premiums (H.1). User's personal habits or health conditions may also become accessible to the public (H.2) due to hackers or via other means. We refer to such negative impacts on the data subjects as privacy harms [11]. Other harms are also possible (e.g., undesired disclosure of personal habits to friends), but we do not discuss them here because of space limitations.

2.5 Definition of the feared events

Harms result from the combination of one or more feared events [11] which are technical events of the system occurring due to the exploitation of the data. Generally speaking, we distinguish three types of feared events resulting from, respectively, the access to personal data (FE.3), the use of personal data (FE.1), and the disclosure of personal data (FE.2).

2.6 Construction of the harm trees

A harm tree represents the relationships among privacy harms, feared events and the exploitation of personal data. The root node of a harm tree denotes a privacy harm. Leaf nodes represent the exploitation of data by the most likely risk source (for the root harm). They are represented as triples (personal data, system component, risk source). Intermediate nodes are feared events caused by the risk sources. They can be seen as intermediate steps of privacy attacks. Children nodes are connected by an AND node if all of them are necessary to give rise to the parent node and by an OR node if any one of them is sufficient. A harm tree can be associated with either an individual risk source or a group of risk sources, colluding or not, depending on the interactions needed to exploit the data.

As an illustration, the harm trees pictured in Figure 1 and Figure 2 are assumed to result from a risk analysis[6] conducted for our case study (for example in the context of an enhanced Data

[4] Attackers or adversaries in IT security.
[5] For example, insurance providers or marketing companies.
[6] Hints about the privacy risk analysis are presented in Section 5.

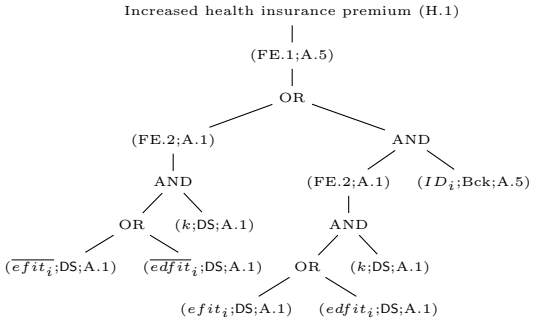


Figure 1: Harm tree for the harm increased health insurance premium (H.1)

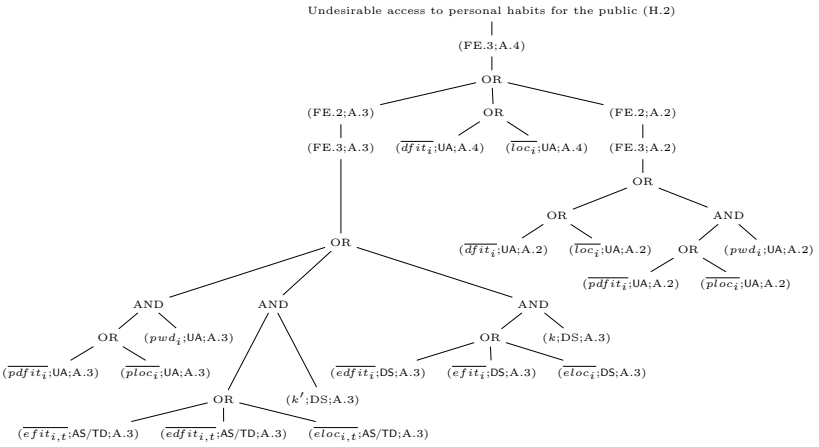


Figure 2: Harm tree for the harm undesirable disclosure of personal habits to the public (H.2)

Protection Impact Assessment). The harm tree in Figure 1 shows that the harm increased health insurance premium (H.1) can be caused by the service provider disclosing to health insurance providers (FE.2) fitness related data (which may be done by disclosing either fitness data fit_i, or other data that can reveal fitness data such as $dfit_i$ in identified or de-identified form). Health insurance providers may use this data to increase health insurance premium (FE.1) for users who they deem unfit. To exploit de-identified data, the health insurance provider (A.5) must have access to identification (ID_i) information of the users as background information. Similarly, Figure 2 pictures the harm tree for H.2.

Although theoretically possible, some combinations of risk sources and exploitations are very unlikely in practice. For example, a member of the public (A.4) is very unlikely to try to find out the password of the user (if they do, they can be classified as hackers (A.3)). Similarly, friends (A.2) of the user are very unlikely to attack servers to get access to the data. These combinations are thus left out of the harm trees.

ID_i may be obtained by a risk source either from a system component or as background information. The abbreviation "Bck" in the harm trees denotes background information. We assume that all other data elements can be obtained only from a system component (they are unlikely to be known as a background information by a risk source).

3 User Privacy Preferences

In this work, we assume that data subjects can specify their privacy preferences or (privacy settings) through *privacy parameters*. For the sake of conciseness, we consider only four *privacy parameters* for our case study:

- The *retention duration* (Ret) of fitness ($dfit_i$, fit_i) and location data (loc_i) at the service provider's database (DS) and in the user account (UA). It can have two values: one year (L) and one week (S). The default value is: one week (S). The value of Ret for the components TD and AS are always short.

- The *visibility* (Vis) of derived fitness ($dfit_i$) and location data (loc_i) from the user account (UA). These data can be made visible to the public (Pu) or friends (F) or kept private (Pr). The default value is: private (Pr).

- The *recipients* (Rec) of fitness ($dfit_i$, fit_i) and location data (loc_i) from the service provider. The service provider may choose to disclose these data only to his sub-contractors (DA) essential to provide the service or to any third party (All) for different incentives. The default value is: sharing only with sub-contractors (DA).

- The *form* (Form) in which the service provider discloses fitness ($dfit_i$, fit_i) and location data (loc_i) to their recipients. The service provider can disclose these data in an identified form (Id) or disclose only de-identified data ($deId$). The default value is: disclosure of de-identified data ($deId$).

A *user privacy preference* is a conjunction of the values assigned to the *privacy parameters*. For example, the conjunction $(\mathsf{Ret} = L) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$ is a *user privacy preference*. The likelihood of the privacy harms may be affected by these preferences. In this paper, for simplicity, we have assumed that the user sets the same value for each parameter for all data elements. In reality, however, the user may set different values for the same parameter for different data elements, resulting in a different user preference for each data element. For example, the user may set $\mathsf{Vis} = F$ for derived fitness data ($dfit$) but $\mathsf{Vis} = Pr$ for location data (loc).

The default values of the *privacy parameters* are chosen such that they constitute the most protective privacy preference which is given by $(\mathsf{Ret} = L) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$.

4 User Interface Design

The objective of this section is to show how users can be informed about the consequences of their privacy settings in a simple and intelligible way. Therefore, we focus on the user interface here and leave the presentation of the actual computation of the risks to the next section.

Users can express their *privacy preferences* through their account UA. The choices can be made when they initially open their account and set up their fitness tracking device.

Figure 3 shows the interactive screen through which users can set their *privacy preferences*. *Privacy preferences* are displayed on the left pane, which we refer to as the *privacy preference pane*. For the different *privacy preferences* selected by the user, the screen displays (prominently, on its right hand side) the risks that the users may face. We refer to this pane as the *privacy risk pane*. Below, we discuss the design of each of these panes and how they interact with each other.

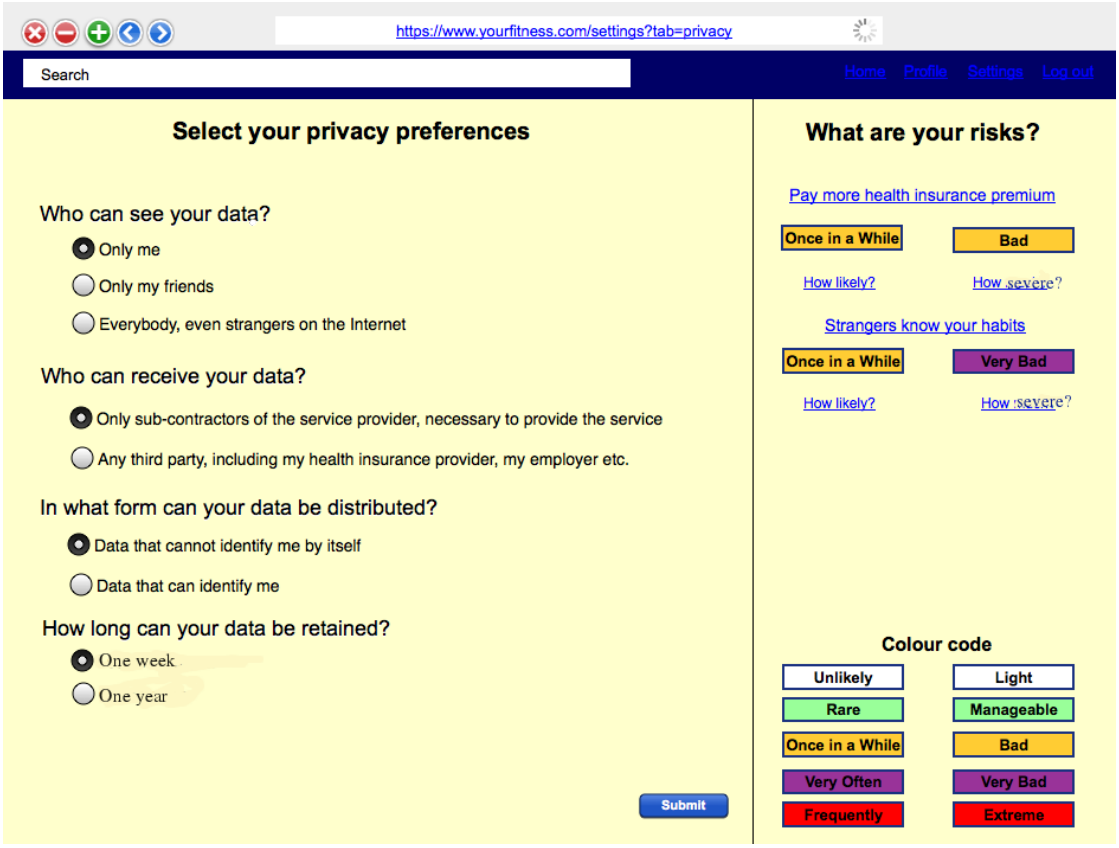


Figure 3: First level screen showing the default *user privacy preference* $(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = deId)$ and its risks

4.1 The Privacy Risk Pane

The task of the right pane (see Figure 3) is to show to the users the risk levels corresponding to each harm. Each harm is presented using a short phrase that can be easily understood by the user. For example, the harm H.1 is presented as "*Pay more health insurance premium*".

Below each harm, the risk level is presented in terms of the likelihood of the harm and the severity of the harm presented using coloured buttons. The likelihood of the harm is dependent on the *user privacy preferences*. In contrast, the severity results only from the nature of the harm and is not derived from the *user privacy preferences*. To explain to the user what these buttons mean, we colour and label them using very short text, both indicative of their meaning and also caption them with phrases like "*How likely?*" (referring to the likelihood) and "*How severe?*" (referring to the severity).

Each button is coloured according to a colour code presented briefly at the bottom of the privacy risk pane. The colour coding is also explained in more detail in another screen shown in Figure 8 (see Section 4.4.4). When a user changes his privacy setting, the colours and texts inside the button representing likelihood will also change (see Section 4.3).

4.2 The Privacy Preference Pane

On the left pane of the screen, users are asked a series of questions to determine their *privacy preferences*. The questions are followed by alternatives which the users can select (by clicking on the corresponding radio buttons). The default selection is the most privacy preserving one. The following questions and answer alternatives are presented to the user. As Figure 3 shows, the most privacy preserving alternative is presented first.

1. Who can see your data? (Vis)[7]
   - Only me (Pr)
   - Only my friends (F)
   - Everybody, even strangers on the Internet (Pu)
2. Who can receive your data? (Rec)
   - Only sub-contractors of the service provider, necessary to provide service (DA)
   - Any third party (All) including my health insurance provider, my employer etc.
3. In what form can your data be distributed? (Form)
   - Data that cannot identify me on its own ($deId$)
   - Data that can identify me (Id)
4. How long can your data be retained? (Ret)
   - One week (S)
   - One year (L)

[7]The parameters (i.e., Vis, Rec, Ret and Form) corresponding to each question and the values (e.g., All, Pu, Pr, L etc.) corresponding to each answer alternative are shown here for the sake of clarity to the readers of the paper. The screen does not show these parameter and values.

Whenever the question or the answer alternatives involve term(s) that the user may be unfamiliar with, suitable but short explanations and examples are used. For example, the user may not fully understand who a "third party" is or what the phrase "any third party" refers to. Therefore, suitable examples are provided to make the user aware that a third party may mean his health insurance provider or even his employer. Links to more detailed explanation for important terms (e.g., "third party", "processing", "data", "identify" etc.) can also be added.

4.3 Interaction between the panes

Initially, the privacy risk pane of the screen displays the risks to the users for the default alternatives. Whenever the user inputs a preference that is different from the default option, the risk pane displays the changes in the risk levels resulting from this choice. The severity level of the harms are independent of the *user privacy preferences*. They are presented to the user for a complete understanding of the risk.

The interactive screen allows the user to observe how any change he makes in the *privacy preferences* has an impact on the risk level of the harms. Based on these risk levels, the user can ultimately arrive at a decision as to which *privacy preference* is the most acceptable to him, given the risk levels. After the user is satisfied with his selection, he can press the "*Submit*" button to communicate his pereference to the service provider.

4.4 Links for more information

The primary or first level screen leads to several linked or second level webpages. There are four types of links, all from the privacy risk pane: 1) from the privacy harms ("*Pay more health insurance premium*" and "*Strangers know your habits*"); 2) from the likelihoods ("*How likely?*"); 3) from the severities ("*How severe?*") and 4) from the coloured buttons denoting the level of likelihood and severity. In all the second level webpages, we still retain the privacy risk pane so that the user gets the opportunity to see the risk levels as he learns more about the different terms and colour codes. A "Back" button on the top left of these webpages allow the users to go back to the first level screen. In the rest of this section, we discuss the design of the second level screens.

4.4.1 Links from privacy harms

The privacy harms on the privacy risk pane are linked to further screens that explain what these harms mean. Figure 4 shows the screen that is obtained by clicking the link on the text "*Pay more health insurance premium*" on the privacy risk pane. Each of these screens states the harm in a full comprehensible sentence and also answers potential questions that the user may have after reading it. For example, in Figure 4, we see the explanation "*Your health insurance provider may charge you more premium based on your fitness data.*" along with two red bubbles that answer the following important questions that the user may ask: 1) why a health insurance provider could charge more premium based on fitness data and 2) how the health insurance provider could obtain such data.

To substantiate our claims that the harms considered are indeed legitimate scenarios and to educate the user further, we also provide links to relevant news articles at the bottom of the screens. These news articles either present a scenario where the privacy harm has happened or discuss the possibility of the occurrence of such harms. For example, the news articles on the screen shown in Figure 4 all discuss about the possibility of the disclosure of fitness data to health insurance providers whereas the link on the screen shown in Figure 5 presents an actual

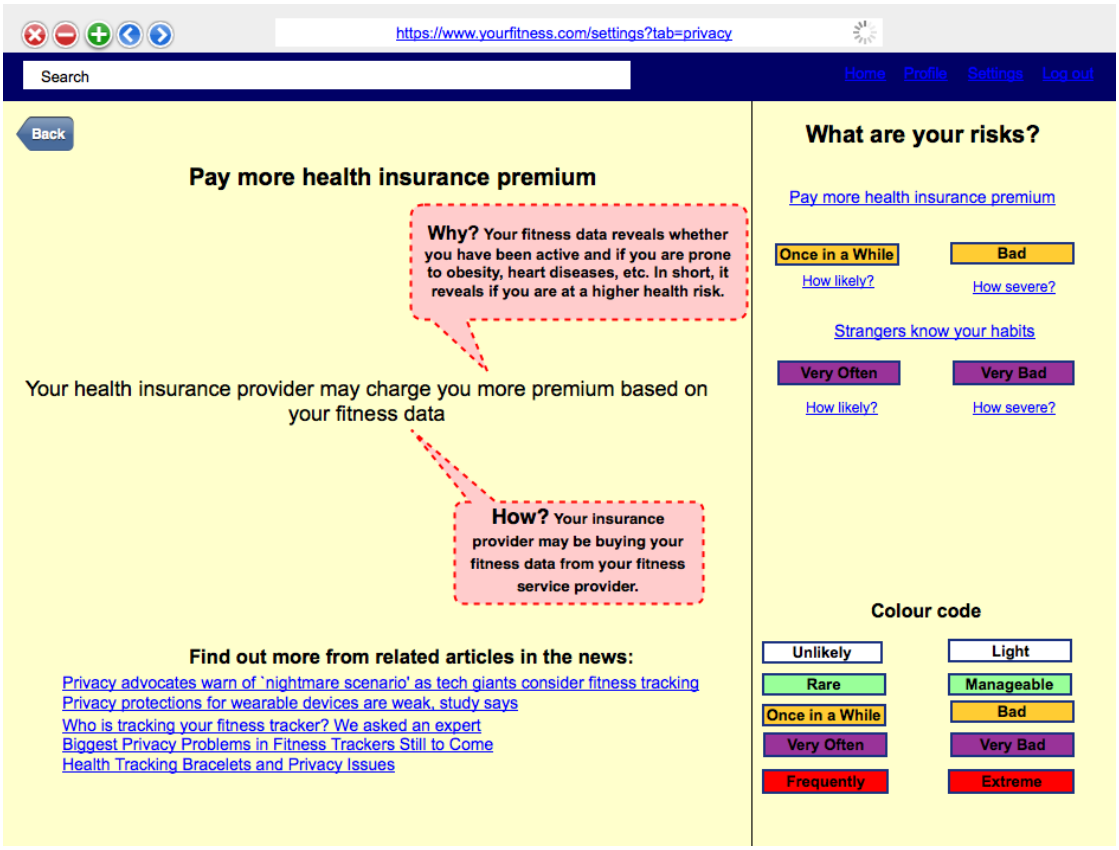


Figure 4: Second level screen linked from the harm "*Pay more health insurance premium*" for the user privacy preference $(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$

incident where sensitive personal information had been disclosed on the Internet by a fitness service provider.

4.4.2 Links from likelihoods

The privacy risk pane shows the likelihoods of the different privacy harms. The likelihood of each harm is labelled as "*How likely?*" which also links to a screen that shows what leads to the current level of likelihood. Figure 6 shows this screen corresponding to the likelihood of H.1. In this screen, we highlight to the user what contributes to the likelihood of the harm, based on our analysis using the harm trees. For example, Figure 6 shows that the likelihood of the harm H.1 is only "*Once in a While*" mainly because of two positive (indicated by green button with "+") factors: 1) the service provider can only disclose de-identified data to his sub-contractors (since $\mathsf{Form} = deId$ and $\mathsf{Rec} = DA$) and 2) the sub-contractors are bound legally by the service provider not to re-identify the data disclosed to them.

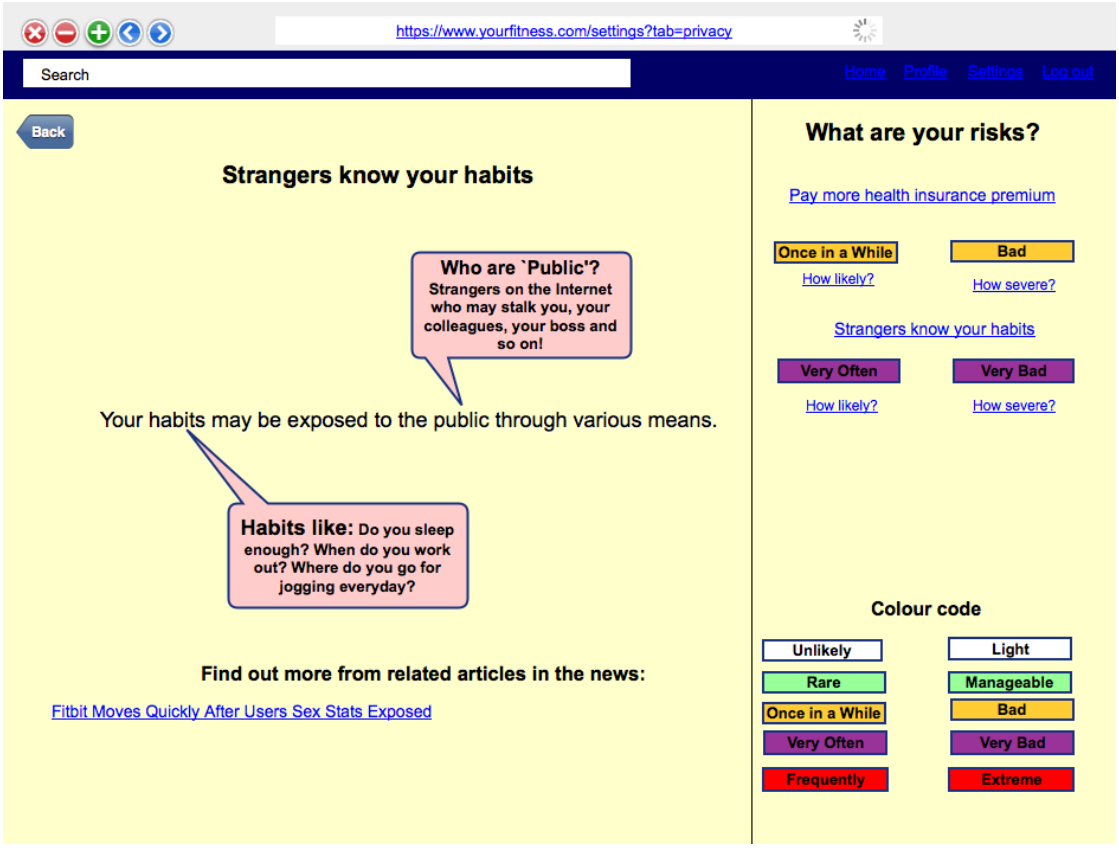


Figure 5: Second level screen linked from the harm "*Strangers know your habits*" for the user privacy preference $(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$

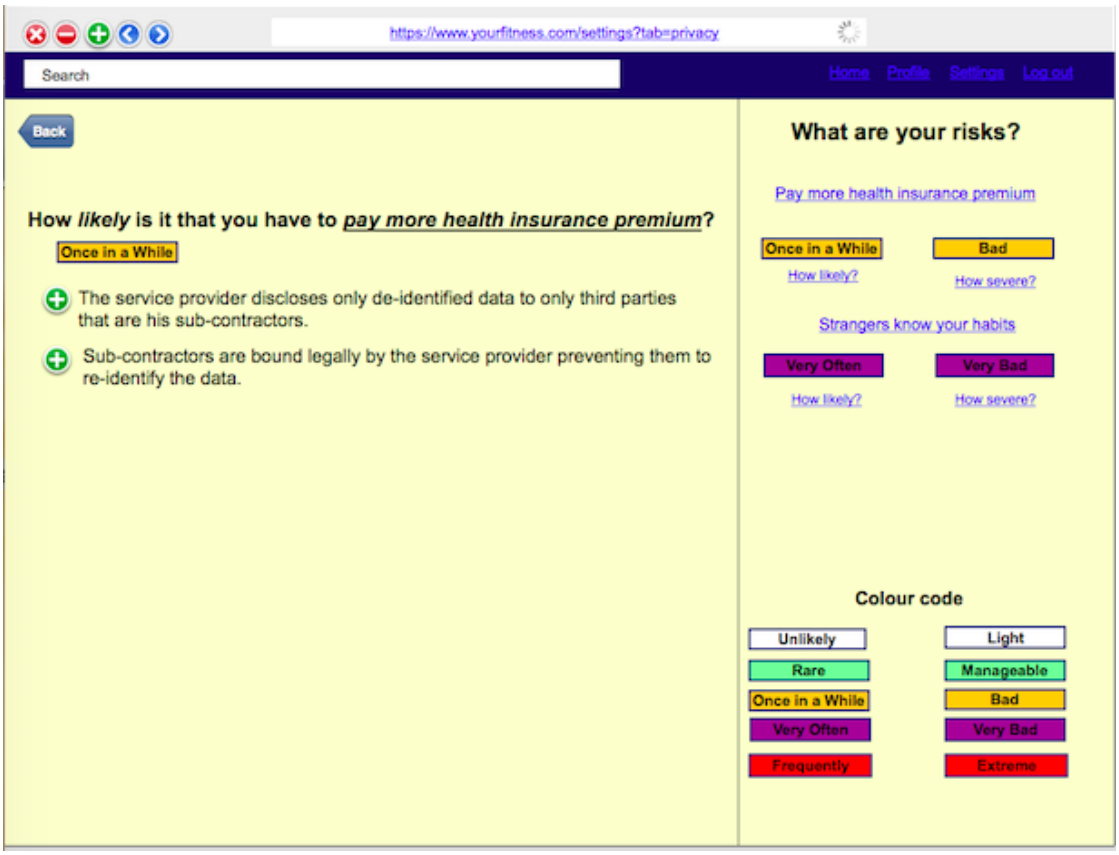


Figure 6: Second level screen linked from "*How likely?*" for H.1 for the user privacy preference $(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$

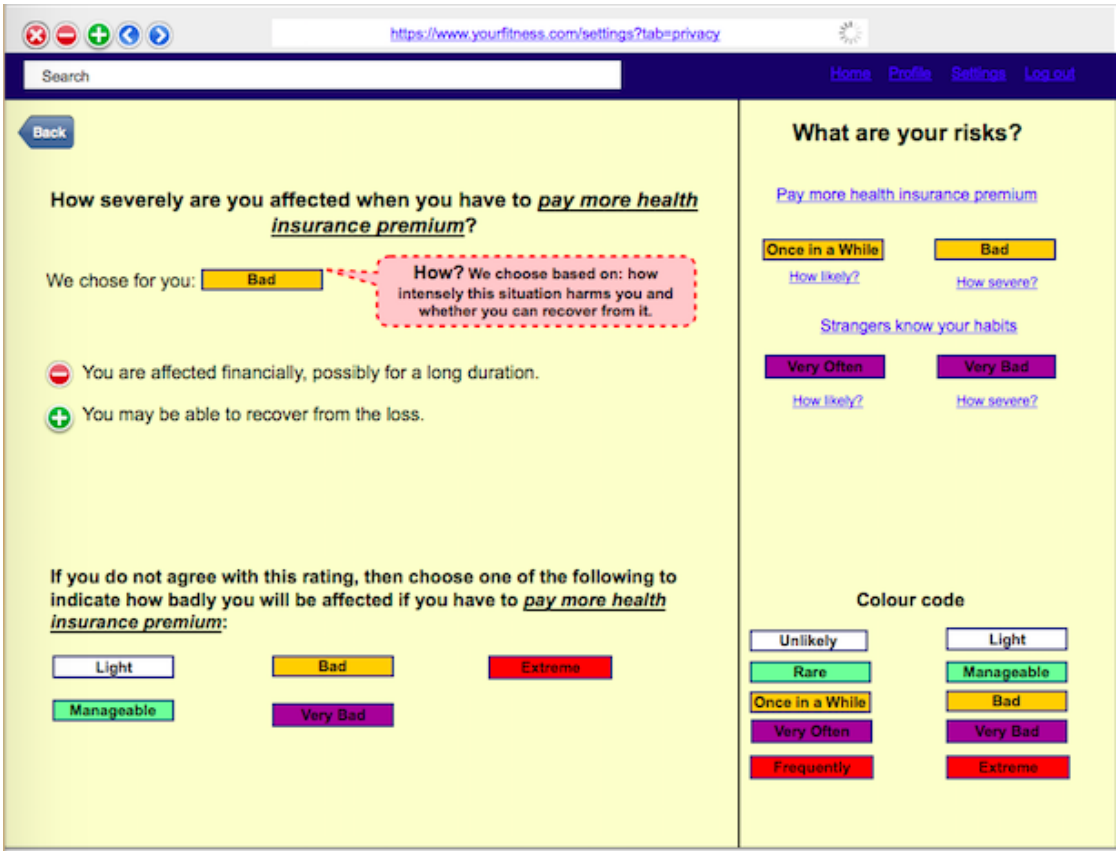


Figure 7: Second level screen linked from "*How severely?*" for H.1 for the user privacy preference $(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$

4.4.3 Links from severity

The privacy risk pane shows the severities of the different privacy harms. The severity of each harm is labelled as "*How severe?*" which also links to a screen that shows what leads to the level of severity. Figure 7 shows this screen corresponding to the severity of H.1.

We assign the severity values to harms based on two factors which are also used by the CNIL [9]: 1) how much inconveniences or difficulties are faced by the data subject and 2) whether or how easy it is for the data subject to recover from the harm. We remind the user of these two factors in a red bubble following the level of severity (see Figure 7).

We also describe to the user our reasons for assigning a certain level of severity to a harm. For example, Figure 7 shows that the severity of the harm H.1 is "*Bad*" mainly because of one positive (indicated by green buttons with "+") and one negative (indicated by red button with "−") factor: 1) the user will be affected financially, possibly for a long time but 2) it may be relatively easy for the user to recover from the harm.

The user may find that the severity level assigned to a harm does not match his expectations. This may happen due to different contextual factors (for e.g., for an employed person, financial loss may not be as severe as for a college student without a job). In such cases, the user has the

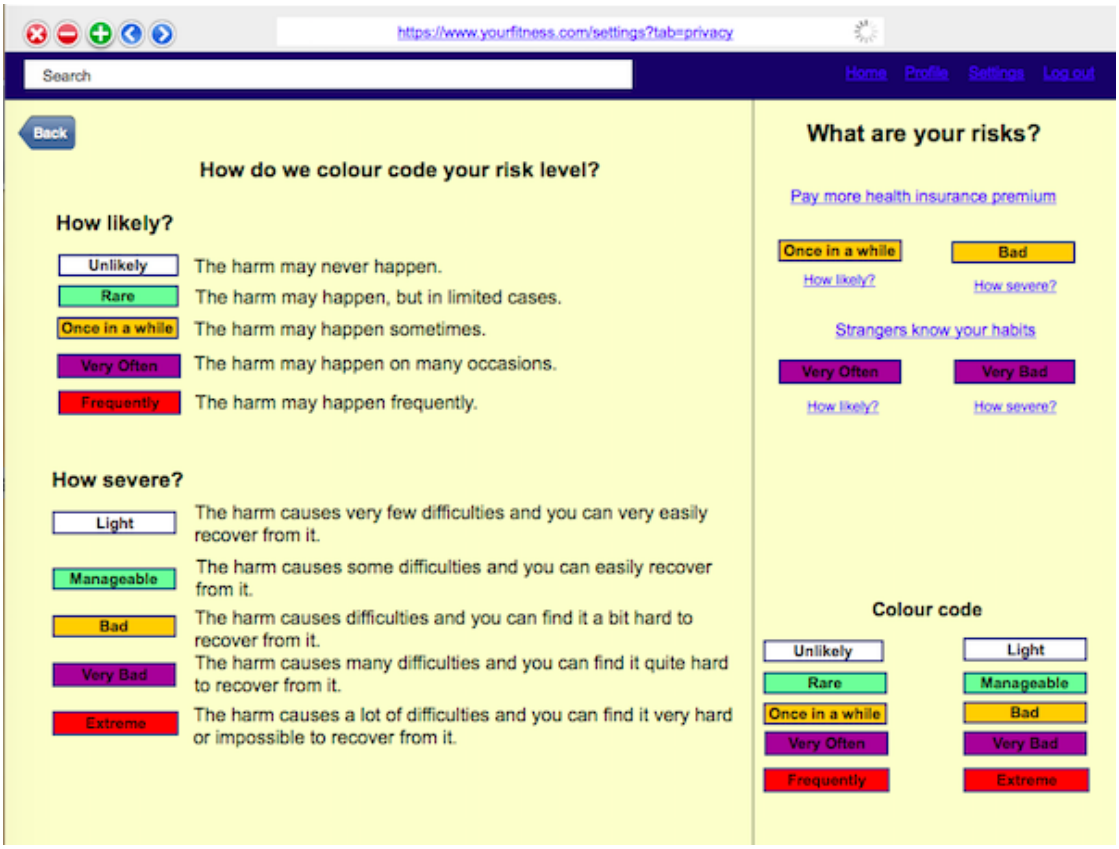


Figure 8: Second level screen linked from the likelihood and severity level buttons

opportunity to select his own level of severity (from the same scale). This severity level is then displayed in the privacy risk pane.

4.4.4 Links from coloured buttons

The coloured buttons on the privacy risk pane of the first level screen indicating the level of likelihood also link to another screen which explains in detail the colour scheme and the texts inside the buttons. Figure 8 shows the meaning of each coloured button indicating the level of likelihood in a way which is comprehensible to the user.

4.5 Usability features

Great care has been taken to enhance the usability of the user interface presented in this section. In particular, we have:

- Used illustrative examples and avoided the use of technical terms.
- Presented the privacy harms in a simple language so that users can relate them to harmful consequences in their own lives.

- Ensured readability by restricting the amount of information presented in each screen and using appropriate fonts and colours.

- Ensured that the process of selection of privacy preferences is not confusing or too much time consuming to the user by including only one basic window for the purpose and limiting the degree of freedom (few questions and few options) of the user.

- Allowed the users to choose the severity level of the privacy harms if they do not agree with the default choice, but kept this as an option only at a later stage to avoid confusion and too much time consumption for the average user who may just agree with the default values.

- Presented more information through hyperlinks about the privacy harms and their severity and likelihood and the colour codes for inquisitive users who may want to educate themselves further. We have included news articles related to each privacy harm so that the user can refer to them to improve their awareness.

- Kept the default privacy preference to be the most privacy preserving one, so that the users who skip this selection step can still enjoy the highest level of privacy.

5 Risk Analysis with Privacy Parameters

In Section 4, we have presented the interactions with users to allow them express their privacy preferences and to inform them about privacy risks, but without explaining the actual computation of these risks. In this section, we focus on the risk analysis itself, which is based on the methodology proposed in [11, 12], enhanced with facilities to deal with privacy parameters.

The primary objective of a risk analysis is to identify the privacy harms for a system in a given context and to assess the associated risks, which are generally measured in terms of likelihood and severity.

Several factors can influence the likelihoods of the privacy harms. The *exploitability* of personal data can be characterized by the resources (e.g., technical resources, access rights, background knowledge) needed by a risk source to exploit them. The dual notion is the *capacity* of a risk source which is defined by its resources (e.g., technical resources, access rights, background knowledge). The *motivation* represents the incentives and disincentives of a risk source to cause a feared event or a harm. The values of the *privacy parameters* influence the values of *exploitability* of data, the *capacity* and the *motivation* of risk sources in certain cases. In the next subsections, we study this influence and show how it can be taken into account in the privacy risk analysis process.

5.1 Exploitability of Data

Some data may be accessible to certain risk sources legitimately. It may be either because the risk source controls a component storing or processing the data or because the risk source has legitimate access rights to the data (e.g., when $\mathsf{Vis} = F$ or $\mathsf{Vis} = Pu$). The control over data allows a risk source to use that data in any way. In our case study, the service provider (A.1) has full control over the database server (DS) and the application server (AS) and hence can access the data in these components without having to attack them[8].

[8]Since the service provider can already access all necessary data from AS and DS, it does not need to attack UA.

Component	UPref	Data	Exploitability
DS	Ret $= L$	$edfit_i$, $efit_i$, $eloc_i$	Transience
DS	Ret $= S$	$edfit_i$, $efit_i$, $eloc_{i,t}$	Persistance
UA	Ret $= L$	$pdfit_i$, $ploc_i$	Transience
UA	Ret $= S$	$pdfit_i$, $ploc_{i,t}$	Persistance
AS/TD	$\times$ (always stored for short time)	$edfit_i$, $efit_i$, $eloc_i$	Persistance

Table 1: Exploitability values of data affected by Retention (Ret)

Some risk sources, under certain circumstances, may have control over some data, rather than a component. For example, when the user chooses to make their personal data visible to their friends (Vis $= F$), then the friends (A.2) have legitimate access over this data (in other words, control) but not over the component UA itself.[9]

Risk sources that have no control over a piece of data have to exploit (or attack) it, persistently or transiently. To this aim, they need resources that may or may not be available to them. By transient exploitation, we mean an exploitation for a short period of time or infrequent exploitation; by persistant exploitation we mean an exploitation for a long period of time (e.g., for several days or months). When the retention duration of a data is long, i.e., a year (Ret $= L$), it becomes vulnerable to transient exploitation. In contrast, if the retention duration is short, i.e., a week (Ret $= S$), it is only vulnerable to persistant exploitation. For example, data (such as $\overline{edfit_i}$ and $\overline{pdfit_i}$), if retained for a short duration in DS or in UA, require persistant exploitation. On the other hand, when retained for a long duration, transient exploitation is sufficient. We assume that data is stored in AS and TD for a short time. Hence, these data require persistant exploitation.

Cryptographic keys and passwords are securely stored by the service provider. So, they require control on the component storing them, in this case, DS, to be exploited (this is the highest level of protection or the lowest level of exploitability). We also assume that the service provider has taken enough security measures so that the user's password cannot be disclosed through the user account[10]. So, control is required to exploit any password-protected data by obtaining pwd_i from UA.

The exploitability values of the different data types for different components and for different retention durations are shown in Table 1[11]. The exploitability values that are not affected by retention durations are shown in Table 2.

Background information is not a part of the data stored in the system. So, it does not have any exploitability value.

[9] In practice, the user may allow his friends to see *dfit* but not *loc*. In such cases, the differentiation between control over data and the control over a component becomes important.

[10] For example, the friend of a user may try to guess the user's password from previously known information about the user such as date of birth, passwords used for other services etc. He may also try to obtain the password by social engineering. Here, we assume that the service provider has taken sufficient measures to prevent these, for example, by spreading awareness among the user about not disclosing passwords, locking the user account after a few unsuccessful sign-in attempts, ensuring the use of strong passwords etc.

[11] In Table 1, the exploitability value of data stored in TD or in AS are affected by the retention time. However, in our case, we have assumed that the service provider has reasonably decided to store the data in these components for a short duration. If he had made the other choice or different choices for these two components, the exploitability values would have been different.

Compo-nent	UPref	Data	Exploitability
UA	× (irrespective of retention time)	$dfit_i$, fit_i, loc_i	Control
DS	× (irrespective of retention time)	k, k', pwd_i	Control
UA	× (no retention time)	pwd_i	Control

Table 2: Exploitability values of data not affected by Retention (Ret)

5.2 Capacity and Motivation of Risk Sources

A risk source can possess the capacity for transient or persistant exploitation or may control one or more data elements or one or more components. The highest capacity of any risk source with respect to a data element or a component is to have control over that data element or component. For example, the service provider (A.1) controls AS and DS. The least capacity of any risk source is the inability to perform any exploitation as in the case of the friend (A.2) of the user when Vis = Pr and the public (A.4) when Vis = Pr or when Vis = F. A.3 and A.5 are assumed to have persistant and transient capacities respectively.

The control over data can be influenced by the value of *visibility*. For example, under the default value for *visibility* (i.e., Vis = Pr), a friend (A.2) of the user or the public (A.4) have no control over any data and no technical resources to exploit them. However, when the user allows his data to be visible to his friends (i.e., Vis = F), they gain control over this data (similarly for the public in general when Vis = Pu).

The availability of background information to some risk sources is also considered as a part of their capacity. We consider that a risk source has a "high" capacity if it possesses the background information relevant for an exploitation, and "low" otherwise. The only background information considered here is ID_i. As shown in Figure 1, A.5 must be able to exploit $dfit_i$, fit_i and/or loc_i which can be provided by A.1 in de-identified form. We assume that A.5 has a "low" chance of possessing this background information.

The *privacy parameters recipients* and *form* influence the value of the motivation of the service provider (A.1) to perform an exploitation. We assume that the motivation of the service provider to comply with the privacy settings of the user is always "high". If the user specifies a less privacy preserving option, then the motivation of the service provider to choose the option which gives him more incentive is always "high" and that of the option which gives him less incentive is "low". For example, the motivation of the service provider to disclose identified data is "low" when the user limits this disclosure to deidentified data, due to the fear of legal sanctions and the loss of consumer trust. Otherwise, the motivation of disclosing identified data is "high" due to financial incentives.

For other risk sources, the motivations (see Table 3) are not affected by the privacy parameters. The motivation of a friend (A.2) to access (FE.3) and disclose the user's personal data (FE.2) when it is legitimately accessible to him (i.e., Vis = F) is "medium" because such an access is generally easy but the disclosed data may negatively affect friendship relationships or lead to embarrassments. On the other hand, the motivation for a friend (A.2) to access (FE.3) and disclose personal data (FE.2) through an attack to know the password[12] is "low" because

[12]We assume that the friend of the user is not an expert hacker (A.3). At most, he may be a hacker with very limited resources. The friend of the user may try simple measures to get the password from the user such as social engineering or try several passwords that use known information about the user such as his date of birth, name etc. In any case, the only way the friend can attack the system is by compromising the user account (UA) through very simple measures. He will not attack the database server (DS) to get access to any data.

Data	Risk Source	Feared Event	Motivation
$\overline{dfit_i}$, $\overline{fit_i}$, $\overline{loc_i}$	A.2	FE.2 and FE.3	Medium
$\underline{pwd_i}$, $\underline{pdfit_i}$, $\underline{ploc_i}$	A.2	FE.2 and FE.3	Low
$\overline{dfit_i}$, $\overline{fit_i}$, $\overline{loc_i}$, $\underline{pwd_i}$, $\underline{pdfit_i}$, $\underline{ploc_i}$	A.2	FE.1	×
$\underline{dfit_i}$, $\underline{fit_i}$, $\underline{loc_i}$	A.4	FE.3	Medium
$\underline{dfit_i}$, $\underline{fit_i}$, $\underline{loc_i}$	A.4	FE.1, FE.2	×
$\overline{dfit_i}$, $\overline{fit_i}$, $\overline{loc_i}$, $dfit_i$, fit_i, loc_i, ID	A.5 (other than DA)	FE.2 and/or FE.1	High
$\overline{dfit_i}$, $\overline{fit_i}$, $\overline{loc_i}$, $dfit_i$, fit_i, loc_i, ID	A.5 (DA)	FE.2 and/or FE.1	Low
$\overline{dfit_i}$, $\overline{fit_i}$, $\overline{loc_i}$, $dfit_i$, fit_i, loc_i, ID	A.5	FE.3	×
pwd_i, $\overline{pdfit_i}$, $\overline{ploc_i}$, k, k', $\overline{edfit_i}$, $\overline{efit_i}$, $\overline{eloc_i}$	A.3	FE.1 and/or FE.2 and/or FE.3	High

Table 3: Motivation of risk sources other than the service provider (A.1)

it may lead to a significant loss of trust in the relationship. Hackers, on the contrary, have a "high" motivation to access (FE.3) and disclose (FE.2) any data. Any other member of the public (A.4) has a "medium" motivation to access the user data (FE.3) (they may seek quick monetary gains, but also fear getting caught). Third parties, other than sub-contractors of the service provider, have a "high" motivation[13] to disclose and use data for unauthorized purposes (FE.2, FE.1) and also to make use of any identifying information (ID) that may be available to them as background information. Sub-contractors of the service provider, however, have "low" motivation as they are legally bound by the service provider not to disclose, misuse or re-identify data disclosed to them by the service provider.

Some combinations of feared events and risk sources do not make sense. The corresponding rows are marked with '×' in Table 3. For example, the friends (A.2) of the user are not given access to data for any particular purpose. Similarly, the public (A.4) is not provided access to data with any specific purpose nor is there any intention of the user to hide some data when he allows its disclosure to the public. Third parties (A.5) do not perform unintended access to data (FE.3) because this would qualify them as hackers (A.3)[14].

The motivation of business partners of the service provider (A.5) to use background information is "high" due to potential financial incentives.

5.3 Computation of Likelihoods

The computation of the likelihoods of the harms based on the harm trees shown in Section 2.6 can be carried out in two steps. The first step is the assessment of the likelihoods of the leaves of the harm trees (likelihood of exploitation of personal data) from the *motivation* and the *capability* of the relevant *risk sources* using Table 4. The capability of the risk source to perform an exploitation is derived by comparing the value of the exploitability of the data and the capacity of the risk source. A risk source has a "high" capability when its capacity satisfies the desired conditions (w.r.t. control, persistant and transient access) for exploitability, otherwise it has a "low" capability. This assessment is based on Section 5.1 and Section 5.2. To be consistent with other leaf nodes, the leaf nodes corresponding to background information (for which there are no exploitability) are directly assigned a likelihood value based on a "high" capability (since background information, when available, is easily usable by risk sources) and the motivation of

[13] To err on the safe side, we consider worst case scenarios here.

[14] We assume that third parties are not looking for quick monetary gains, unlike some members of the public.

Likelihood of exploitation	Risk source capability	Motivation
Negligible Limited	Low High	Low
Negligible Significant	Low High	Medium
Limited Maximum	Low High	High

Table 4: Measurement rule for likelihood of exploitation

Scale used for computation	How likely?
Negligible	Unlikely (white)
Limited	Rare (green)
Intermediate	Once in a While (amber)
Significant	Very Often (purple)
Maximum	Frequently (red)

Table 5: Mapping of scales for likelihood

the risk source to use it.

The second step is the computation of the likelihood of each harm according to the following rules (applied bottom-up), where P_i is the likelihood of the ith child node:

R1. AND node with independent child nodes: $\prod_i P_i$.

R2. AND node with dependent child nodes[15]: $Min(P_i)$, i.e., minimum of the likelihoods of the child nodes.

R3. OR node with independent child nodes: $1 - \prod_i (1 - P_i)$.

R4. OR node with dependent child nodes[16]: $Min(1, \sum_i P_i)$.

To perform the computations of the second step, it is necessary to translate the symbolic likelihood values of Table 4 into numerical values. This transformation has to be made by the privacy expert in collaboration with the owner and should be documented. In this paper, we use as an illustration the following correspondance for the likelihood values (p): 1) *Negligible (N)*: $p < 0.01\%$; 2) *Limited (L)*: $0.01\% \leq p < 0.1\%$; 3) *Intermediate (I)*: $0.1\% \leq p < 1\%$; 4) *Significant (S)*: $1\% \leq p < 10\%$; 5) *Maximum (M)*: $p \geq 10\%$.

5.4 Choice of Privacy Preferences

Table 6 shows the effect of the various values assigned to different privacy parameters on the likelihood values of the harms H.1 and H.2. Table 5 shows how the colour coding and the texts used in the screens for *"How likely?"*in Section 4 map to the scale used to measure likelihood in Section 5. Not surprisingly, the default user setting leads to the lowest level of risk, i.e., the likelihoods of both H.1 and H.2 are "intermediate". We also observe that changing the default values of Ret to L does not increase harms (i.e., their likelihood values remain unchanged). Changing the default values of Rec to *All* or Vis to Pu or Form to Id make the harms riskier (i.e., their likelihood values increase).

[15] In order to err on the safe side in terms of privacy protection, we consider dependent nodes such that one node may imply the other nodes.

[16] In order to err on the safe side in terms of privacy protection, we consider dependent nodes such that each node may exclude the other nodes.

User Preference	Likelihood for H.1	Likelihood for H.2
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = deId)$ (Default preference)	Intermediate	Intermediate
$(\mathsf{Ret} = L) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = deId)$	Intermediate	Intermediate
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = All) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = deId)$	Significant	Intermediate
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = F) \wedge (\mathsf{Form} = deId)$	Intermediate	Significant
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = Pu) \wedge (\mathsf{Form} = deId)$	Intermediate	Significant
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = DA) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = Id)$	Significant	Intermediate
$(\mathsf{Ret} = S) \wedge (\mathsf{Rec} = All) \wedge (\mathsf{Vis} = Pr) \wedge (\mathsf{Form} = Id)$	Significant	Intermediate

Table 6: Summary of likelihoods of harms for different user preferences

The results of the previous sections can help the user to decide upon an acceptable likelihood for each harm, given their severity. Based on Table 6, and the acceptable threshold, he can then decide which values for the privacy parameters he prefers. Let us assume that the user decides that the acceptability threshold for a harm with "very bad" severity (for example H.2) is "intermediate" and that of a harm with "bad" severity (for example, H.1) is "significant". Then, he may choose any privacy preference (from Table 6) other than the ones in which $\mathsf{Vis} = F$ or $\mathsf{Vis} = Pu$.

The choice of values for *privacy parameters* by the user may depend on factors other than privacy. For example, when setting $\mathsf{Vis} = F$, the user may be emphasizing on the utility of the social networking aspect of the fitness tracking system (e.g., sharing fitness data with friends and/or competing with them for a fitness level may act as a motivation to the user).

6 Related Works

The communication of privacy policies to users in a comprehensible form has been an important focus of privacy research. The ToS;DR project [27] aims to classify, through a transparent and peer-reviewed process, the terms of service and the privacy policies into six colour-coded classes. Privacy icons [20] or privacy policy icons [15] are simplified pictures used to visualize elements of privacy policies. Inspired by nutrition labeling, drug facts, energy information and the efforts to standardize financial privacy notice, Kelley et al. [22, 23] propose the privacy nutrition label to improve the accessibility, readability and understanding of privacy policies among users. Poor, confusing interface design, permissive default settings, limited visual feedback etc. can often lead to the under-utilization of available privacy options [17, 24].

Different types of Transparency Enhancing Tools (TETs) have also been proposed to allow users to express their privacy choices or to inform them about the privacy policies of service providers. PrivacyBird [1] can compare user privacy preferences with P3P policies and help the user decide whether to reveal data to the website [10]. Other similar TETs provide insight about the privacy implications of intended data revelation from the user's side (Mozilla Privacy Icons [2], Privacyscore [21]) or of data already revealed by the user (PrimeLife's Privacy Dashboard [3], Google Dashboard [4]) [21]. Still others provide insights into third party tracking (Lightbeam [5]) or promote privacy awareness and education about privacy problems among users and (Me & My Shadow [6]). In a different context, Zhou et al. [29] design a privacy mode for smartphone users with a flexible and fine-grained control on the types of personal information that can be accessed by an application. The work described in this paper is complementary to the above proposals. Unlike previous work in this area, our objective is to help users in the definition of their privacy settings, based on a privacy risk analysis.

7 Conclusions and Future Directions

The concepts presented in this paper can form the basis of a full fledged privacy tool enabling data subjects to make informed choices about their privacy settings. It could also form the core of an education tool to increase awareness about privacy. The need for this kind of tool has been identified by other authors such as Schaub and his co-authors in [28], emphasizing that *"Communicating risks, for instance with examples, supports an individual's assessment of privacy implications, especially when data practices are complex or abstract."* An avenue for further research is the extension to dynamic risk analysis to take into account the personal data disclosure history of the user. It would be useful, for example, to analyze the impact on privacy risks of the disclosure of new personal data to a third party that has already collected data on the subject. Another interesting research direction is the integration of alternative actions (such as disclosure of anonymized data, less precise data, or even fake data) and their consequences. A better understanding of these options would further enhance the control of individuals on their personal data.

References

[1] http://www.privacybird.org. Accessed: 2017-01-05.

[2] https://disconnect.me/icons. Accessed: 2017-01-06.

[3] http://primelife.ercim.eu/results/opensource/76-dashboard. Accessed: 2017-01-06.

[4] https://support.google.com/accounts/answer/162744?hl=en. Accessed: 2017-01-06.

[5] https://www.mozilla.org/en-US/lightbeam/. Accessed: 2017-01-06.

[6] https://myshadow.org. Accessed: 2017-01-06.

[7] Able To Turn Off Surge Heart Rate Monitor? https://community.fitbit.com/t5/Surge/Able-To-Turn-Off-Surge-Heart-Rate-Monitor/td-p/659645, 2015. Accessed: 2017-02-08.

[8] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer, 2006.

[9] Commission Nationale de l'Informatique et des Libertes (CNIL). Privacy Impact Assessment (PIA) Tools (templates and knowledge bases), 2015.

[10] L. F. Cranor, P. Guduru, and M. Arjula. User Interfaces for Privacy Agents. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):135–178, 2006.

[11] S. J. De and D. Le Métayer. PRIAM: A Privacy Risk Analysis Methodology. In *11th International Workshop on Data Privacy Management (DPM)*. IEEE, 2016.

[12] S. J. De and D. Le Métayer. Privacy Risk Analysis. In *Synthesis Series*. Morgan & Claypool Publishers, 2016.

[13] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016.

[14] E. Field. Biggest Privacy Problems in Fitness Trackers Still To Come. http://www.law360.com/articles/686145/biggest-privacy-problems-in-fitness-trackers-still-to-come, 2015.

[15] S. Fischer-Hübner, J. Angulo, and T. Pulls. How Can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used? In *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, pages 77–92. Springer, 2013.

[16] E. Gratton. Health-tracking bracelets and privacy issues. http://www.eloisegratton.com/blog/2014/12/20/health-tracking-bracelets-and-privacy-issues/, 2014.

[17] R. Gross and A. Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.

[18] H. Hedbom. A Survey on Transparency Tools for Enhancing Privacy. In *IFIP Summer School on the Future of Identity in the Information Society*, pages 67–82. Springer, 2008.

[19] K. Hill. Fitbit Moves Quickly After Users' Sex Stats Exposed. Forbes, 2011.

[20] L.-E. Holtz, K. Nocun, and M. Hansen. Towards Displaying Privacy Information with Icons. In *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, pages 338–348. Springer, 2010.

[21] M. Janic, J. P. Wijbenga, and T. Veugen. Transparency Enhancing Tools (TETs): An Overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*, pages 18–25. IEEE, 2013.

[22] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A Nutrition Label for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

[23] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 1573–1582. ACM, 2010.

[24] H. R. Lipford, A. Besmer, and J. Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.

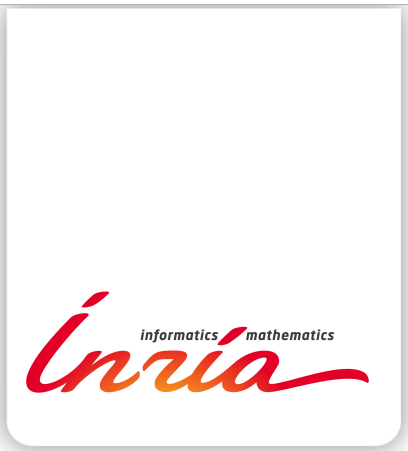
[25] I. Oskolkov. Your fitness is their business. Nothing personal. https://blog.kaspersky.com/fitness-trackers-privacy/6480/, 2014. Accessed: 2016-02-12.

[26] S. R. Peppet. Regulating the Internet of Things: First Steps toward Managing Discrimination, Privacy, Security and Consent. *Tex. L. Rev.*, 93:85, 2014.

[27] H. Roy, d. M. Jong, J.-C. Borchardt, I. McGowan, J. Stout, and S. Azmayesh. Terms of Service Didn't Read. https://tosdr.org, 2012. Accessed: 2017-02-10.

[28] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17. USENIX Association, 2015.

[29] Y. Zhou, X. Zhang, X. Jiang, and V. W. Freeh. Taming Information-Stealing Smartphone Applications (on Asndroid). In *International Conference on Trust and Trustworthy Computing*, pages 93–107. Springer, 2011.



RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399