# **Data-Driven Privacy Indicators**

Hamza Harkous EPFL, Switzerland hamza.harkous@epfl.ch Rameez Rahman EPFL, Switzerland rrameez@gmail.com Karl Aberer EPFL, Switzerland karl.aberer@epfl.ch

#### **ABSTRACT**

Third party applications work on top of existing platforms that host users' data. Although these apps access this data to provide users with specific services, they can also use it for monetization or profiling purposes. In practice, there is a significant gap between users' privacy expectations and the actual access levels of 3rd party apps, which are often over-privileged. Due to weaknesses in the existing privacy indicators, users are generally not well-informed on what data these apps get. Even more, we are witnessing the rise of inverse privacy: 3rd parties collect data that enables them to know information about users that users do not know, cannot remember, or cannot reach [2]. In this paper, we describe our recent experiences with the design and evaluation of Data-Driven Privacy Indicators (DDPIs), an approach attempting to reduce the aforementioned privacy gap. DDPIs are realized through analyzing user's data by a trusted party (e.g., the app platform) and integrating the analysis results in the privacy indicator's interface. We discuss DDPIs in the context of 3rd party apps on cloud platforms, such as Google Drive and Dropbox. Specifically, we present our recent work on Far-reaching Insights, which show users the insights that apps can infer about them (e.g., their topics of interest, collaboration and activity patterns etc.). Then we present History-based insights, a novel privacy indicator which informs the user on what data is already accessible by an app vendor, based on previous app installations by the user or her collaborators. We further discuss future ideas on new DDPIs, and we outline the challenges facing the wide-scale deployment of such indicators.

#### 1. INTRODUCTION

Nowadays, we are witnessing the rise of the tech platform economy [8]. A lot of successful software services reach a point where their further growth necessitates opening their products to third party developers via APIs. By creating this ecosystem of 3rd parties, they essentially instigate the network effect: the more the users of those apps, the more valuable the underlying platform becomes. This has been the motivation in the case of mobile platforms (e.g., Apple's App Store and Google Play Store), cloud platforms (e.g., Google Drive's and Dropbox's app ecosystems), and social apps platforms (e.g., Facebook, Twitter, and LinkedIn). These 3rd party apps obtain access to users' data via permission interfaces. Such permissions serve as the main privacy indicators that users view before deciding on whether to authorize an

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

Workshop On Privacy Indicators, at the Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016, June 22–24, 2016, Denver, Colorado.

app to access their data. However, these permissions are not always easy to comprehend, which reduces their effectiveness in deterring users from installing over-privileged apps. In addition, due to continuous exposure to the same permissions interface, users are prone to notice fatigue and habituation effects [6].

By demonstrating the potential of Data Driven Privacy Indicators (DDPIs) in the cloud apps scenario, our aim is to further contribute to transforming the static permissions interface into a dynamic one that helps the user make more informed decisions while minimizing the habituation effect. The goal of DDPIs is not only to personalize the interface based on the user's background or privacy preferences as suggested with personalized privacy notices [5]. They rather aim to incorporate visualizations based on the user's data in the permissions interface itself. For example, these visualizations indicate to the user what the app can concretely extract about her or whether the app vendor already has access to her data.

In the following section, we further explain the 3rd party cloud apps' ecosystem. In that context, we then present two realizations of the DDPI paradigm in Sections 3 and 4. We discuss in Section 5 other possible scenarios where a similar approach can help. Finally, in Section 6, we explain the challenges that DDPIs might face and their possible shortcomings.

#### 2. THIRD PARTY CLOUD APPS

Over the previous years, the apps that interact with users' cloud files (on Dropbox, Box, Google Drive, etc.) have seen traction at both the individual and the enterprise level. Such apps provide a variety of services for most types of files (document signing, PDF annotation, image styling, video editing, etc.). On a high level, there are three entities that interact in the third-party cloud app system: (1) a vendor which is responsible for programming and managing a third-party cloud app (or shortly a cloud app), (2) a user who uses that app for achieving a certain service, and (3) a cloud storage provider (CSP) hosting the user's data.

With their data in the hands of these new potential adversaries, users are under increased risk of data leakage and data-based profiling. Users' files expose the topics they are interested in, the entities they are mentioning the most, and their sentiments towards various subjects. Their photos expose the faces of people they are frequently with, the locations they have been to, and what activities they have been doing. The hundreds of millions of users who might trust Google Drive or Dropbox with their data do not necessarily trust all these 3rd party apps' developers. This becomes even more critical in the context of enterprises. In a lot of cases, enterprises have special agreements with Google or Dropbox on how the data is handled, where it is stored, and how long it is retained. This is not necessarily the case with 3rd party apps, which might not even display a privacy policy

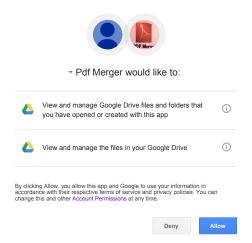


Figure 1: Example of the current permissions interface of Google Drive

to begin with. Employees installing 3rd party apps are then exposing the enterprise data to potentially unaccountable entities. Based on the above, there is a growing need for more effective privacy notices in the context of 3rd party cloud apps. In the next two sections, we discuss our experience with multiple privacy indicators in the context of Google Drive, one of the most popular 3rd party ecosystems. For reference, Figure 1 shows the current permissions interface of Google Drive. In this ecosystem, we differentiate on a high level between two types of access that apps can request: (1) full access, where the app can view all the files in the user's account and (2) partial access, where the app can only view the files that the user has opened or created with the app itself.

# 3. IMMEDIATE AND FAR-REACHING INSIGHTS

In this section, we describe the design and evaluation of Immediate Insights and Far-reaching Insights, two privacy indicators that help users concretize what the adversary can potentially infer from their data. An extended version of this section appeared in our recent work [4]. However, our goal here is to summarize the high-level findings which fit the discussion on DDPIs.

#### 3.1 Motivation

In a previous work, we have shown that almost two-thirds of the featured Google Drive apps on Google Chrome store are over-privileged: they request more permissions than they need to function [4]. An app that is meant to crop a single photo might request access to all the users files. The current model has no indicator of whether the app is over-privileged. Moreover, users might not necessarily understand the real risk of exposing their files. Hence, we were motivated to develop new privacy indicators in the form of two new permission models.

# 3.2 Description

#### 3.2.1 *Immediate Insights*

The first model is based on the following hypothesis:

"When users are shown samples of the data that can be extracted from the unneeded permissions granted to apps, they are less likely to authorize these apps." We call this model *Immediate Insights (IM)*, and we show an instance of it in Figure 2. On the left of the figure, we distinguish the needed from the unneeded permissions<sup>1</sup>. On the right, we have the *Insights Area*, where we show a question that says: "What do the **unneeded permissions** say about you?", followed by an answer in the form of a visual with short explanatory text. In Figure 2, the Insights Area visualizes the map **Location** of a randomly chosen user photo. We have also developed three more variations, where the Insights Area shows a random **Image**, a random excerpt from a **Text** file, or a random **Collaborator**.

#### 3.2.2 Far-reaching Insights

The second model we introduced was based on the following hypothesis:

"When the users are shown the far-reaching information that can be inferred from the unneeded permissions granted to apps, they are less likely to authorize these apps."

These are insights that go beyond direct data examples and include what can be inferred by running more involved algorithms. Hence, we denote this model by Far-reaching Insights (or shortly FR Insights). The interface layout is the same as that of Figure 2, but with the Insights Area containing an FR insight instead of an immediate insight. In this work, we have designed 6 types of FR insights that can be extracted from users' data. For textual files we had (1) an insight showing the Entities, Concepts, and **Topics (ECT)** extracted (Figure 3), (2) an insight showing the **Sentiments** towards entities (Figure 4), (3) and an insight showing the Shared Interests between the user and various collaborators (Figure 5). For image files, we had (1) an insight, Faces with Context, showing the most frequent faces along with the objects appearing with them (Figure 6) and (2) another insight, Faces On Map, showing a map of where these faces and objects were captured, based on their geotags (Figure 7). Finally, we had an additional insight that showed the user's **Top Collaborators** (appears in Figure 12 of the Appendix).

# 3.3 Experimental Setup

In order to test the efficacy of both Immediate and Farreaching insights, we conducted an online user experiment. For brevity purposes, we will provide a high-level description of the setup and methodology. We refer the reader to our full paper for the full description of the setup, methodology, apps used, and data handling [4].

#### 3.4 Setup and Methodology

We recruited users online via our university's mailing list and website. The experiment could be completed from any user's web browser (i. e., without supervision). We required that users sign with their Google Drive account after reading the privacy policy. Only users who had at least 10 files containing text or 20 images were allowed to continue. This is to ensure that they possess at least a minimal level of experience with Google Drive. As an incentive, random participants were rewarded by gift cards for app stores of their choice. The first goal of the experiment was to investigate the efficacy of the two new permission models by comparing them to the existing Google Drive permission model. The second goal was to perform micro-comparisons among the different types

<sup>&</sup>lt;sup>1</sup>The details of how the unneeded permissions are determined are available in [4].

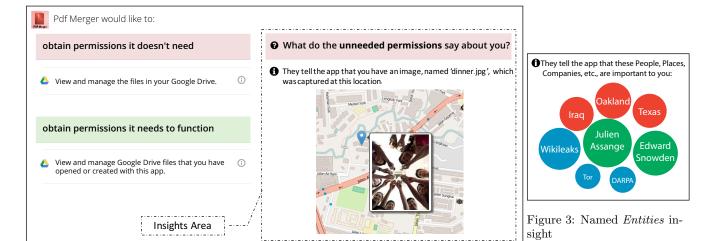


Figure 2: Example of Immediate Insights interface; the same layout is used for Far-reaching insights, with the insights area content changing accordingly



Figure 4: Sentiments insight

Figure 5: SharedInterests insight

of IM and FR. Accordingly, we went for a mixed betweensubject and within-subject design. As far as these permission models are concerned, we had three experimental groups:

- Baseline group (*BL* group): Users in this group were presented with a clone of the original interface that Google shows upon installing the app (shown in Figure 1). This group serves as the control group, and we refer to it as *BL*.
- Immediate Insights Group (*IM* group): Users in this group were presented with the modified interface of Figure 2, with the Insights area containing one of the *IM* insights of Section 3.2.1.
- Far-reaching Insights Group (FR group): Users in this group were presented with the modified interface, of Figure 2, with the Insights Area containing the FR insights described in Section 3.2.2.

The experiment itself was composed of a series of app selection tasks. In each task, the user was required to choose an app with a specific purpose from a list of apps. We show this interface in Figure 9 of the Appendix. Only one app on the list satisfied the indicated purpose. When the user clicked on the app, she was presented with the permission dialog corresponding to its group (i.e., that of Figure 1 for the BL group and that of Figure 2 with a randomly selected visual for the IM or FR Insights groups). The user was then presented with a question that says: "Based on permissions

Figure 6: FacesWithContext Figure 7: FacesOnMap ininsight

below, would you be likely to install this app?". She could choose between "Permissions are too invasive" (accept) and "I'm OK with these permissions" (reject). We worded the question so that we avoid all users rejecting the installations of all apps. We rather aimed that users would reject apps whose permissions they consider as too invasive, thus allowing us to do within-subject comparisons. After answering the question, the user was directed to the next task with another app, until she completed the whole set of tasks. Figures 10, 11 and 12 of the Appendix show screenshots of the interface for the BL, IM and FR experimental groups. The apps were presented to the users in randomized order to compensate for the effects of learning and fatigue. We also avoided using apps from popular vendors to avoid the bias resulting from users being influenced by famous brands.

#### 3.5 Findings

We had 160 users in total who successfully completed this part of the experiment. Out of them, 55 were in the BL Group, 54 in the *IM* Insights group, and 51 in the *FR* group. The metric we used in our evaluation was the *Acceptance Likelihood AL*, defined as:

$$AL = \frac{\#(Accepts)}{\#(Accepts) + \#(Rejects)},\tag{1}$$

where Accepts denotes the cases where users were fine with the permissions, and Rejects denotes the cases where they found them too invasive. Accepts and Rejects were aggregated across users and tasks for the permission model under consideration. The lower the AL, the better the performance in deterring users from installing over-privileged apps. To evaluate the significance of the AL differences among the interface types, we fit a generalized linear mixed model (GLMM) with the user's decision (Accepting/Rejecting the app installation) as the binary response variable and the interface type as the fixed effect. The most important outcomes from our experiment can be summarized as follows:

Inefficacy of Baseline Permissions: We found that the baseline interface had a significantly higher AL (i.e. p-values  $\leq 0.05$ ) than all the insights, except for the Collaborator insight. This highlights the fact that showing well-selected insights will result in deterring more users compared to the case of not showing any insights.

The Power of Relational Insights: We found out that users are significantly more deterred by *Relational Insights*, characterizing their relationships with other people (*Faces With-Context*, *Top Collaborators*, and *Shared Interests* insights), than by *Personal Insights*, characterizing the users themselves (*Image, Text, ECT*, and *Sentiments* insights).

Impact of Face Recognition: We also noticed that showing examples of user's images (AL = 0.21) is significantly less effective than showing the important faces and listing the concepts in the image (AL = 0.08). This supports the hypothesis that we need to go beyond simple examples to achieve a deterring effect on users.

Influence of High-Level Textual Insights: Contrary to the case of images, in the case of textual documents, showing the high-level entities or concepts extracted from the text did not have a significant difference as compared to simply showing direct excerpts from the text (p-value= 0.94). Only when the relationship factor is introduced does the AL significantly decrease (as in the case of SharedInterests).

Superiority of Far-reaching Insights: By aggregating the results over all the experiments with FR Insights, we obtained a significantly lower AL value compared to IM Insights (AL=0.161 and 0.226 respectively). We also confirmed that the AL difference is significant with a pairwise comparison p-value= 0.004). Overall, these results demonstrate the superiority of our novel approach of FR Insights. Nevertheless, IM Insights are still significantly better than the BL model.

#### 4. HISTORY-BASED INSIGHTS

In this section, we present History-based Insights (shortly HB Insights), another privacy indicator that realizes the DDPIs paradigm.

#### 4.1 Motivation

Users of cloud storage services typically share files among each other so that multiple people can view or edit these files. When a user installs 3rd party cloud apps, those apps' vendors can get access to her files. Similarly, when her collaborators install apps, the corresponding vendors get access to files shared between the user and those collaborators. If the user is subsequently given a choice between an app whose vendor already has access to (part of) her data and an app from a totally new vendor, it makes sense, from a privacy angle, to go with the first. This minimizes the number of potential adversaries that get access to her data. It is particularly useful when the user wants to choose among apps with similar functionalities and permission levels. However, the current privacy indicators are not equipped to help users take such

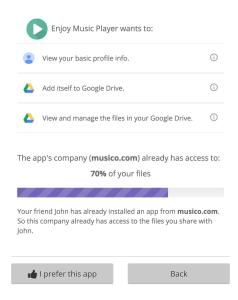


Figure 8: History-based Insights interface

decisions when they make sense. The permissions interface is totally independent of the current shareholders of a user's data. For that reason, we introduce *HB* Insights, which is a new privacy indicator that informs the user about how much of her files is already accessible by the app vendor.

#### 4.2 Description

We continue to consider Google Drive as a case study, and we show this indicator in the context of Google Drive apps' permissions in Figure 8. Compared to the current interface provided by Google (Figure 1), we added a new part to highlight the percentage of user files accessible by the vendor (due to user's or collaborators' previous decisions). Following the best practices in privacy indicators' design [6], our indicator was multilayered, with both textual and visual components. The wording of the main textual part was brief and general enough to hold for both data percentage exposed by friends and that exposed by the user. We used a percentage value rather than a qualitative measure to facilitate making comparisons among apps based on this value. The visual part showed the percentage as a progress bar with a neutral violet color. The bottom textual part was added in a smaller font to provide further explanation for those interested.

In order to evaluate the new permissions interface, we performed an online user study. Again, we will be giving a high-level overview of the experiment, and we leave the full details to an upcoming paper. The hypothesis we wanted to test was the following:

"Introducing the new privacy indicator will significantly increase the probability that the user chooses an app with a minimal incremental privacy loss."

#### 4.3 Setup and Methodology

We recruited users via CrowdFlower's crowdsourcing platform. In our study, we restricted participation, via the platform's filtering system, to the highest quality contributors (Performance Level 3). We also geographically targeted countries where English is a main language as our interface was only in English. In order to further guarantee quality responses, each user was rewarded a small amount of \$0.5 for merely completing the survey and an additional amount of \$1.25 that is manually bonused for those who do not enter irrelevant text in the free-text fields.

Participants were first presented with introductory instructions that explained the context of the study (i.e., cloud storage services and 3rd party apps that can be connected to them). They were asked to only continue if they had good familiarity with cloud storage services (e.g., Google Drive, Dropbox, etc.). The introductory survey was then front-loaded with questions about cloud storage services (several of which required free-text input) in order to discourage users who had not used these services from continuing to the actual study. We did not explicitly require that participants have experience with 3rd party cloud apps. However, we educated them about such apps throughout the instructions, particularly showing them two examples of 3rd party apps in action (PandaDoc for signing documents and iLoveIMG for cropping photos). These apps were displayed via animated GIFs that play automatically and do not rely on the user clicking. We used limited deception by neither mentioning the focus of the study on participants' privacy nor giving hints about selecting apps based on the installation history. The advertised purpose was to "check how people make decisions when they install 3rd party apps."

Next, users could proceed to the study page. We used a split-plot design in the study. Participants were randomly assigned to one of two groups:

- Baseline Group: where the permissions interface used is that currently provided by Google Drive (Figure 1).
- HB Insights Group: where the new permissions interface of Figure 8 is used.

In each group, the study consisted of 3 modules, which cover the main conditions that can occur when users desire to install a cloud app.

- Module 1 (Self-History Scenario) tests whether the user is more likely to select an app from the same vendor she has just installed from before (Figure 14 of the Appendix).
- Module 2 (Collaborator's Vendor Scenario) tests the likelihood that the participant selects the same app (or an app from the same vendor) as her collaborator (Figure 15 of the Appendix).
- Module 3 (Multiple Collaborators Scenario) given collaborators  $f_{more}$  and  $f_{less}$ , where the user shares much more data with  $f_{more}$ , this scenario checks the likelihood of the participant authorizing an app that  $f_{more}$  has installed (Figure 16 of the Appendix).

In each module, the user reads the instructions (Figure 13 of the Appendix) and proceeds to a set of tasks. One of those tasks involves selecting between two apps. The metric we used in our evaluation was the likelihood of selecting the more privacy preserving app (i.e., the app requiring the minimal additional access).

#### 4.4 Findings

We had 141 users who successfully completed the experiment. Out of them, 72 were in the Baseline group and 69 in the HB Insights group. The outcomes of our experiment can be summarized as follows:

Superiority of the HB Insights: In the three modules, the Baseline group witnessed an almost even split between the app whose vendor has access to zero (or small) percentage of the user's files on one side and the app from a vendor who has access to all (or most) of the user's files on the other side. The percentage of users who favored the former ranged from 41.7% to 55.6%. On the other hand, participants in the HB Insights group were significantly more likely to install the more privacy preserving app than those in the Baseline group. The percentage of users who did so ranged from 75.4% to 88.4%. On average, the new privacy indicator has increased the likelihood of users choosing more privacy-preserving apps by 30%. This validates our hypothesis that introducing a DDPI, based on the user's installation history, can serve to minimize the privacy loss resulting from an increasing number of data shareholders.

User Motivations: Privacy was not the only motivation for users to favor apps from vendors with existing access to their data. A significant fraction of users in Module 1 has mentioned cross-app compatibility, interface familiarity, and satisfaction with the previous vendor as justifications for their decisions. In Module 2, users added that they were motivated by their friends' choices (i.e., took them as recommendations). Still, 17% of the participants in the HBInsights group explicitly mentioned privacy as the reason behind their choices. Interestingly, the HB Insights interface has indirectly made users think about various positive effects, which lead them to make more privacy-preserving decisions. This might also be achievable in other DDPI instances, where users' data can be used to highlight the other advantages of taking privacy-aware decisions. For example, the user can be encouraged to select more conservative privacy preferences by showing her that protecting her data might have financial benefits, such as avoiding price discrimination.

Willingness to Participate: At the end of the user study, users were presented with a final set of questions. We asked them whether they would like to be notified when a friend installs an app that gets access to their shared files. Around 92% of users in the Baseline group and 90% of users in the HB group agreed. We further asked the participants whether they are fine with a collaborator being notified when they install applications that access files shared with that collaborator. The percentage of people who agreed dropped to 75% in the Baseline and 78% in the HB group. The difference between the answers to these two questions highlights that only a minority of users is not willing to make the trade-off of contributing to the overall system. Such users can be given the option to not use privacy indicators based on their friends' decisions.

#### 5. FURTHER APPLICATIONS OF DDPI

So far, we have presented three instances of DDPIs that we have tested in the context of cloud apps privacy. Immediate Insights, Far-reaching Insights, and History-based Insights all communicate a privacy message to the user by leveraging her actual data. Moreover, all of them are shown as a part of the app authorization process. In this section, we discuss how DDPIs can be extended to other platforms and how they can be used as privacy indicators *after* the app is installed.

**Previous Works in Other Ecosystems:** To begin with, there have been a few works that also followed the DDPI paradigm in other ecosystems. Most notably, Harbach et

al., proposed to integrate examples from user's data in the permissions request screen to expose the data apps can get access to [3]. This involved showing random pictures, call logs, location, and contacts from user's data that correspond to each permission. Another related work in the context of Facebook is that by Wang et al. [7], who introduced the "Privacy Nudges" technique to aid users while posting statuses to Facebook through showing random profile pictures of friends who can see the post or showing the post sentiment. With Far-reaching Insights, our aim was to show that well-crafted visuals extracted from users' data can be more effective than randomly selected data. We also showed through pairwise comparisons among the insights themselves that the choice of the displayed insight highly affects the interface's effectiveness.

Extensions of FR and HB Insights: Looking forward, it is obvious how a direct extension of FR Insights and HB Insights can work in mobile or social networking platforms. These same indicators can also work with browser extensions that unnecessarily request access to a user's web history. The only difference will be the type of user's data that is visualized. Moreover, in the scenario of 3rd party apps' installation, we can introduce DDPIs that visualize the network of 4th party entities that can get access to the user's data. These can be advertisement providers, data brokers, or even governmental agencies. Another DDPI can be used to visualize the location where the user's data will be hosted and accordingly the jurisdictions governing the data handling.

New DDPIs: There are other scenarios where new types of DDPIs can be introduced. One scenario is helping users visualize the consequences of privacy settings. When the user is presented with the option to encrypt her disk or phone, a special privacy indicator can show her how others view her data as a result. Another example is in the context of selective encryption. A lot of solutions allow users to selectively encrypt some sensitive folders before syncing them to cloud storage services. However, encryption comes with a trade-off: most apps do not typically operate on encrypted files. DDPIs can be used to visualize which of the user's apps still operate after she encrypts the folder in question.

Post Installation Scenario: Furthermore, since DDPIs depend on the user's data, they are not exclusively valid at app installation time. For example, a cloud provider can show the user far-reaching insights that a 3rd party app can glean from files it has actually downloaded (as opposed to what it can potentially access). This also deters the apps from accessing unneeded files as the user would be better equipped to discover such events. One notable realization of post-installation DDPIs was the work by Almuhimedi et al., where users are notified about how frequently their location data is being queried by mobile apps [1].

#### 6. LIMITATIONS

For DDPIs to be deployed on a wide scale, the privacy motivation is not sufficient. There are several issues that can stand in that way.

The Business Case: First, there should be a business advantage for the platform owner to engage in such change. On the surface, providing the users with interfaces that deter them from installing certain 3rd party apps might seem as a growth-curtailing step for the platform. However, the fact

that privacy itself has recently become a major selling point for several companies might serve as a business motivation for the app platforms.

The Economic Cost: From an economic perspective, DDPIs might result in an extra computational and development overload. This is due to the various data analysis and machine learning techniques that might be needed in the privacy indicator. In a lot of cases, however, the platform already analyzes the user's files for other purposes, such as improving search (e.g. Google Photos analyzes user photos for finding objects and faces). Hence, the privacy indicators can build on existing data that is readily extracted by the platform, without additional cost.

Usability Challenges: Introducing DDPIs should not negatively impact the usability of the platform. Evidently, bombarding users with several privacy indicators at once and allowing indicators to have conflicting messages will result in user frustration. A key point in successfully deploying these indicators is to deliver a succinct privacy-related message in a concrete way that is tailored to the user. Hence, an important area of potential research is how to prioritize the message to deliver to the user in order to give her the best privacy level within a minimal attention span.

Trusting the Provider: Finally, it is important to note that in several scenarios, the user is assumed to trust the platform itself to perform the data analysis and construct the privacy indicator. Nevertheless, this assumption is not an additional requirement imposed by DDPIs. It is rather an existing assumption that users make whenever they host their data on the platform itself.

#### 7. FUTURE WORK

In this paper, we have presented our work on Data-Driven Privacy Indicators (DDPIs) for 3rd party cloud apps. In a nutshell, DDPIs communicate privacy messages in a language the user understands most: her own data. We presented three instances of DDPIs: Immediate Insights and Far-reaching insights from our previous work [4] and History-based insights which we have newly developed. Despite their limited spread in real world systems, DDPIs have the potential to reduce the habituation effect of privacy notices due to their inherent dynamicity. With the rising trend of personal assistants delivering virtually any service, privacy personal assistants might soon become commonplace. Such trusted privacy "coaches" can benefit from research around DDPIs in order to better relay the consequences of data collection or of specific privacy settings. This is especially useful in voice-based assistants, where the goal is to relay a short, understandable message to the user. As research around DDPIs shows, involving users' data in these messages will create a greater user inclination towards more privacy-preserving options.

In the future, we plan to continue our work on DDPIs in the cloud scenario and to further introduce new ways of assisting users in that context. So far, we have developed PrivySeal, a privacy-focused assistant which uses Far-reaching insights to deter users from installing over-privileged cloud apps. PrivySeal is available for public at https://privyseal.epfl.ch and has been used by over 1500 registered users. We further aim at complementing DDPIs with data-based, actionable recommendations too (e.g. suggesting privacy-preserving apps to use). That way, users are not only informed but are also guided towards the most privacy-preserving options.

#### 8. REFERENCES

- [1] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of* the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 787–796. ACM, 2015.
- [2] Y. Gurevich, E. Hudis, and J. M. Wing. Inverse privacy (revised). Technical Report MSR-TR-2015-37, May 2015.
- [3] M. Harbach, M. Hettig, S. Weber, and M. Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the 32Nd* Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pages 2647–2656, New York, NY, USA, 2014. ACM.
- [4] H. Harkous, R. Rahman, B. Karlas, and K. Aberer. The Curious Case of the PDF Converter that Likes Mozart: Dissecting and Mitigating the Privacy Risk of Personal Cloud Apps. In *Proceedings on Privacy Enhancing* Technologies (PoPETs), volume 2016, 2016.
- [5] N. Sadeh, A. Acquisti, T. D. Breaux, L. F. Cranor, A. M. McDonalda, J. R. Reidenbergb, N. A. Smith, F. Liu, N. C. Russellb, F. Schaub, et al. The usable privacy policy project. Technical report, Technical Report, CMU-ISR-13-119, Carnegie Mellon University, 2013.
- [6] 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, Ottawa, July 2015. USENIX Association.
- [7] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor. Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13 Companion, pages 763–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
- [8] I. Wladawsky-Berger. The Rise of the Platform Economy - The CIO Report - WSJ. http://blogs.wsj.com/cio/2016/02/12/the-rise-of-the-platform-economy/.

### **APPENDIX**

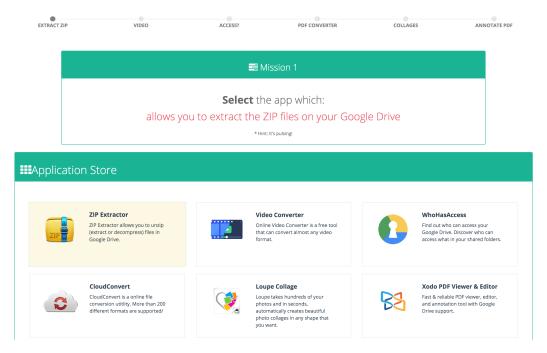


Figure 9: Task interface presented for the users in the experiment, where they had to select the app satisfying the given purpose (already highlighted for them)

Based on **permissions** below, would you be likely to **install this app**?

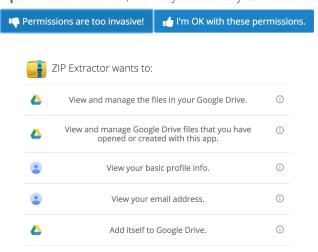


Figure 10: Example interface shown to users of the BL group, with the decision dialog on top

Based on **permissions** below, would you be likely to **install this app**?

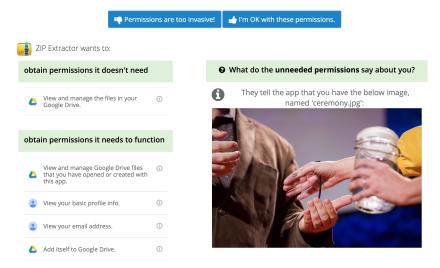


Figure 11: Example interface shown to users of the IM group, with the decision dialog on top

Based on **permissions** below, would you be likely to **install this app**?

They tell the app that these People are your top collaborators

Obtain permissions it doesn't need

○ What do the unneeded permissions say about you?

○ View and manage the files in your ○ Collaborators

Obtain permissions it needs to function

○ View and manage Google Drive files ○ Charles of that you have opened or created with this app.

○ View your basic profile info. ○ Collaborators

○ View your email address. ○ Collaborators

○ View your cased on the files in your ○ Collaborators

○ View your basic profile info. ○ Collaborators

Figure 12: Example interface shown to users of the FR group, with the decision dialog on top

- Take a breath.
- You will now play a role of someone who has a Google Drive account and has already stored files on it (like
  your images from your trips with the family, some official documents, music files, etc.).
- At some point, you'll be asked to choose some apps to connect to your Google Drive account. Although no
  apps will be actually installed, we ask you to think as if they were real apps and that this is a real Google Drive
  account
- Once you install an app, assume it is still installed throughout this experiment.
  - For example, if the first tasks says: "you have installed an application from the company pandadoc.com", this application is still there when you go to the next task. So if the second task says: "who has access to your data?", the answer would be "pandadoc.com company".
- Similarly, whenever you are informed that your friends have installed applications, consider that these
  applications are still connected to their Google Drive throughout this experiment.
- When this experiment ends and you move to the next experiment, you will start from scratch (i.e. with no
  apps installed).

Figure 13: Instructions for each module experiment of Section 4.3

Task:

As explained, we now start from scratch. Consider that this is the first app you will install. Please install any application from the company: **thetimetube.com**. (Only one such app exists, and you can click on the app to view its info.)

Task:

You now need an app that allows scanning your Google Drive files for viruses.

Two such apps exist below. Check them both by clicking on them. Then choose the one that you prefer to install.

Be prepared to give a reason for that choice.

Figure 14: The 2 tasks for Module 1 (Self-History Scenario)

Task:

Google Drive allows you to share files with friends. You decided to share all your photos on Google Drive with your friend John . Up till now, who is the friend who has access to your data?

○John ○Lisa

Task:

Your friend  ${\sf JOhn}$  has installed an application called <code>Photo</code> <code>Editor</code> and has given its company

access to all his files (including shared files). Write below the name of the company that owns this application. (You can click on the app to view its info.)

Task:

You now need an app that allows converting all audio files on Google Drive to MP3 format. Two such apps exist below. Check them both by clicking on them. Then choose the one that you prefer to install.

Be prepared to give a reason for that choice.

Task:

Your friend John has installed an application called Online Player and has given its company access to all his files (including shared files). Write below the name of the company that owns this application. (You can click on the app to view its info.)

Task:

You now need an app that allows playing your music files that are saved on Google Drive. Two such apps exist below. Check them both by clicking on them. Then choose the one that you prefer to install.

Be prepared to give a reason for that choice.

Figure 15: The 5 tasks of Module 2 (Collaborator's Vendor Scenario)

some of your photos. Who has more files from you in their Google Drive?

○ John ○ Lisa

Task:

Your friend Lisa

has installed an application called PDF Mergy and has given its company access

to all her files (including shared files). Write below the name of the company that owns this application. (You can click on the app to view its info.)

Task:

Your friend John has installed an application called PDF Files Merger and has given its company access to all his files (including shared files). Write below the name of the company that owns this application. (You can click on the app to view its info.)

Task

You now need an app that allows you to merge multiple PDF documents on your Google Drive into a single PDF file.

Two such apps exist below. Check them both by clicking on them. Then choose the one that you prefer to install.

Be prepared to give a reason for that choice.

Figure 16: The 4 tasks of Module 3 (Multiple Collaborators Scenario)