APRIM: An Account and PeRsonal Information Manager

Christian Hartlage
Fraunhofer FKIE
Friedrich-Ebert-Allee 144
53113 Bonn, Germany
hartlage@cs.uni-bonn.de

Delphine Reinhardt
University of Bonn and
Fraunhofer FKIE
Friedrich-Ebert-Allee 144
53113 Bonn, Germany
delphine.reinhardt@cs.uni-bonn.de

CHI'17 Extended Abstracts, May 6–11, 2017, Denver, CO, USA.
ACM ISBN 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053125

Abstract

Most online services require the creation of a user account and the disclosure of personal data, such as addresses or phone numbers. With the multiplication of these services, the digital footprint left by the users rapidly grows. In order to increase the users' awareness and support them in the management of these data, we propose a new account and personal information manager called APRIM. Our solution logs disclosed personal information and presents it to the users in the form of a matrix. We have designed and implemented a proof-of-concept and evaluated the concepts and the usability of APRIM by means of a user study involving 18 participants. The results confirm that APRIM is a usable solution to keep track of users' digital footprint.

Author Keywords

Privacy; Digital Footprint; Transparency; Usability; User study

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction

Many existing websites require users to create an account prior to fully access their online services. To create such an account, users usually have to at least provide an e-mail address and/or a username and choose a password. Depending on the nature of the services provided by the website, additional personal information must sometimes be disclosed, such as a users' date of birth or credit card number. Consequently, users increase their digital footprint, i.e., the amount of personal information disclosed to these parties which each newly created account.

In order to help users in managing their account information and hence protect their personal information, different approaches including password managers have been proposed and adopted by the users [4]. In this paper, we therefore build on the concept of password managers to allow users not only to manage their account information, but also their personal information. By doing so, we aim at reducing the memorization efforts demanded from the users, while simultaneously catering for transparency and awareness of their digital footprint. To this end, our contributions can be summarized as follows.

We present the design of a new account and personal information manager called APRIM. Our solution builds on an existing open source password manager FPM2 [9] and extends it in two dimensions. In addition to store login information, APRIM includes (1) the personal information disclosed by the users when creating the corresponding account and (2) temporal information about the creation and the last access of each account as well as the last password update. The latter dimension aims at helping users in identifying unused accounts, so that they can delete their account and linked personal information to reduce their digital footprint. Additionally, it allows users to easily recognize passwords which may need to be updated. In order to reduce the users' interactions to the minimum, APRIM is further based on a browser extension that seamlessly gathers the personal information entered by the users and transmits it to a local database. Users can access the stored information via an interface to get an overview about the personal information of their online accounts.

In addition to presenting APRIM, we evaluate it by conducting a user study with 18 participants in a lab setting. In our study, the participants tested a mock-up of APRIM and answered two questionnaires. The first one aims at analyzing their experience with existing solutions, while the second one focuses on their experience with APRIM and the evaluation of its usability. The results show that most participants do not have a good overview of their current digital footprint and that APRIM would help them to have both a better control over it. Overall, the participants rated the usability of APRIM as good.

Related Work

Our solution bridges the gap between existing password managers and auditing tools. By utilizing a combination of a browser extension and a local database, it benefits from the advantages of both online and offline password managers in terms of user experience and exposure to online threats. Moreover, APRIM integrates a temporal component and proposes a novel compact visualization of disclosed information.

Existing password managers can be divided into two categories: online and offline password managers. In the former category, password managers, such as Lastpass [11], or 1Password [1], directly save and retrieve passwords via the users' web browsers. By doing so, users do not need to manually enter them during log-in. In these solutions, the login information is however stored online by the service providers. This means that the security of these passwords can be threatened by potential vulnerabilities as demonstrated in [12] for five well-established password managers.

Requirements:

Reduction of memorization efforts: Our solution should support users in remembering login credentials and the associated personal data.

Transparency: It should further allow users in visualizing and auditing previously disclosed personal data.

Non-disruptiveness: It should not interrupt users' primary tasks, i.e., login or creating a new account.

Versatility: It should be applicable independently of the used operating systems, browsers, and websites.

Autonomous: It should not require the collaboration of the websites' owners.

Privacy-friendly: Personal data should not be disclosed to third parties.

In addition to passwords, 1Password [1] also manages users' identities, shipping details, or credit cards, and synchronizes them between devices for a monthly subscription fee. As compared to APRIM, the underlying motivation of 1Password is different. In 1Password, personal data are collected to prevent users from manually fulfilling the same web forms, so that the associated overhead is reduced. By doing so, the disclosure of personal information is eased, as users can do it in just one click. As shown in [13], automated entry functions can lead to over-disclosure of personal information in web forms since optional fields may be automatically completed. As a result, APRIM and 1Password follow two opposite goals, as we aim at raising the users' awareness about their digital fingerprint and helping them to manage it.

In contrast to online password managers, users' account information is locally stored in offline password managers like, e.g., Keepass [17]. Users hence avoid exposing their login information and thus personal data to online threats. Simultaneously, offline managers do not support synchronization between devices, and thus require more efforts from the users, as these still need to memorize or manually report them when used with other devices. To address this issue, Versipass proposed in [16] combines the concepts of password managers and cued graphical passwords into one system. Instead of storing the passwords, it helps users in generating and remembering the passwords based on the stored passwords cues.

In addition to password managers, solutions have ben proposed to help users in monitoring their digital footprint. For example, *Data Disclosure Log* [8] and *MozPETs* [6] log and display personal information disclosed to web services. While both solutions are integrated into the Mozilla Firefox browser, MozPETs only stores the data, but leaves their analysis and visualization to the users. In contrast, Data Disclosure Log offers different graph-based visualizations. Additionally, *Data Track* [2] has similar visualization options, but relies on a larger framework for collecting and processing personal data. However, the collaboration of the providers of the web services is required to enable its full functionality. In the latest version of Data Track, the disclosed information is presented in strace view, which connects the different services sharing the same information item based on colored lines.

Requirements

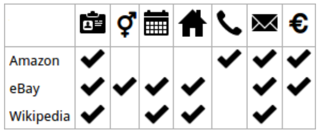
Based on our analysis of related work, we identify the requirements listed in the left margin, which guide the design of our solution.

Design and Implementation

In order to meet the aforementioned requirements, our solution relies on a browser extension and a local database completed by a GUI.

Browser Extension

It automatically detects the creation of a new account by searching for `<form>` elements. In this case, an icon is displayed next to the websites' addresses to notify the users. If a password field is found, we ask the users to confirm that they want to include this website as new entry in their APRIM. Assuming their agreement, the users can leverage the browser extension to generate a new password as already experienced in current browsers. The new password is then stored in the local database along with the users' identifier. As compared to existing solutions, we further store the date of the account creation and also search for additional entry fields, potentially requesting personal information during the registration process. Without loss of generality, we focus on the user's first name, last name, gender,



	🪪	⚥	📅	🏠	📞	✉	€
Amazon	✓				✓	✓	✓
eBay	✓	✓	✓	✓		✓	✓
Wikipedia	✓		✓	✓		✓	

Figure 1: Example of APRIM entries. Icons from [10] 🅭🅯

age, date of birth, address, payment information, e-mail address, and phone number within the scope of this paper.
To support the different terminologies used to describe the same entry fields in HTML forms, we have created a dictionary that maps them. This means that we have created a list including all <input>- und <textarea> elements. We have completed our list by further considering nearby elements such as help texts, potentially revealing the nature of the collected data. For each recognized text element and HTML attribute, a score is attributed. The entry fields are then ordered by the total number of obtained scores and their type is determined as the highest one. As a result, the identified personal data are transmitted and stored to the local database.

Once an entry has been created in APRIM for a website, our browser extension monitors further accesses to this website. The latest access date is stored in the local database and displayed to the user. Depending on their privacy preferences, users can manually deactivate this function in the browser extension or it is automatically deactivated when they use the incognito mode of their browsers. Selective deactivation based on, e.g., the website types, could be envisaged, but is considered as future work. Furthermore, our browser extension keeps track of potentially new or updated personal data entered by the users after the initial registration.

Our proof-of-concept browser extension is implemented using JavaScript, jQuery and HTML5 and is supported by most current web browsers, i.e., Chrome, Firefox, and Safari [15, 18].

Local Database and Graphical User Interface (GUI)
In contrast to most existing solutions, we have decided to store the monitored information in a local database, instead of relying on browser-based or web-based storage. This allows us to (1) guarantee a secure and encrypted storage and (2) provide a solution compatible with the utilization of multiple browsers. In addition to the information monitored by the browser extension, users can manually add or modify entries. Our prototypical implementation builds upon the existing open source password manager FPM2 [9]. Note that the access to the local database is protected by a password.

Users access the information stored in the local database via a dedicated GUI. Six users have tested a preliminary version of the developed GUI. Their feedback and comments have been integrated in the final design. The proposed GUI allows users visualizing their disclosed personal information based on a matrix as illustrated in Fig. 1. Each row corresponds to a specific website, while each column displays whether a particular information item as been disclosed. Using this compact view, users can hence have an overview of their digital footprint. Moreover, users can access a detailed view, in which the entered data are available. In this view, users can also see the date of the account creation, the time elapsed since its last use as well as the time since the last change. Based on this temporal information, users can decide to delete unused accounts or update the corresponding passwords.

Note that related work have adopted different representations. For example, a chronological visualization is leveraged in [2, 8], thus requiring users to chronologically browse through all communication with the website to infer the disclosed information. Alternatively, a tree-based graph with different node types is proposed in [8], but has been shown to be difficult to understand by potential users in [8]. In a recent version of *Data Track* [2], users can select the icon of a particular website to visualize the corresponding information. Colored lines link the icons of the different websites sharing the same personal items. For users with numerous accounts disclosing many information items, the resulting colored graph may become complex to analyze.

Tested tasks:

1. Create a new account in a given website by:
   (a) Fulfilling the registration fields with given data,
   (b) Creating a new password by leveraging the function offered by the browser extension,
   (c) Verify the created account in APRIM,
2. Find an existing account in APRIM containing a given address and phone number and update this account,
3. Identify the account showing the least recently password and update this password,
4. Find the least recently used account and open the corresponding website.

In summary, our solution aims at helping users in visualizing personal information they may have disclosed online using a novel matrix-based representation. In addition, temporal information about the account creation, its last use, and the last update of the corresponding password are made available to the users. By utilizing a browser extension combined with a local database, the information disclosed by different browsers can be merged and secured in a single location, without requiring users to manually enter them.

Evaluation

We have evaluated our approach by means of a user study involving 18 participants. At the beginning, we have informed the participants about the goals and conditions of the study. After their agreement, they answered a preliminary questionnaire, which aims at investigating their current experience with existing solutions. Next, the participants tested our solution in a lab setting, before answering a final questionnaire to assess their experience with our system. Note that both questionnaires were in German mainly because test subjects were also German. In the test phase, the participants have first obtained a textual description of the main features of APRIM and could freely interact with a mockup of APRIM. The HTML and JavaScript-based mockup shares the same features and design as the original system. It does, however, not store any entered information, but instead logs the number of users' interactions as well as the average duration of each task. The participants were able to ask the supervisor of the study in case of difficulties. After having autonomously discovered APRIM, our participants performed the same tasks listed in the left margin. In average, the completion of the study took approximately 30 min per participant. No incentives were provided. Note that our institutions currently do not have ethical boards for reviewing user studies. We have, however, limited the data collection to the minimum and conducted it anonymously. Participants were informed that they could opt out at any time and that their data would be removed.

Demographics and Attitude towards Privacy

In our sample, a majority of the users are female (67%, n/a: 6%) and younger than 24 (>34:16.7%, n/a: 5.6%). Three participants are studying or working in the fields of natural sciences, computer sciences, or geography. All participants use a computer at least one hour daily. To measure the participants' attitude towards privacy, we have applied the *Concern For Information Privacy* (CFIP) scale [14]. The results show that most participants are concerned about their online privacy when considering the three CFIP-scale dimensions, i.e., collection (79%), unauthorized secondary use (95%) and improper access (91%) (SD=18%). Our sample is therefore not representative for the whole population, but builds a set of potential users of our solution.

Experience with Current Solutions

28% of our participants use a password manager, 44% do not use any, and 22% do not know what a password manager is. Among the participants using a password manager, two indicated using KeePass, one Bitdefender Wallet, and one Apple's keychain. 17% estimated to have disclosed personal information to more than 20 websites, while a majority (33%) indicated to have done the same to 5 to 9 websites. To secure the access to these accounts, 11% use the same password, while only two participants use a specific password for each account. The remaining 72% reuse different passwords across different accounts, thus confirming existing studies [7]. The quartiles of the number

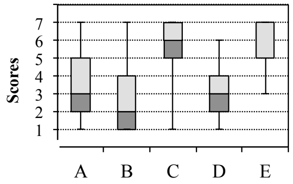


Figure 2: Extrema and quartiles attributed to the statements: (A) I know which websites know which kind of personal information about me, (B) I am satisfied with the current control I have over my digital footprint, (C) It bothers me not to know which websites know personal information about me, (D) I have a good overview on the digital footprint that I leave, (E) I would like to have more control over my digital footprint

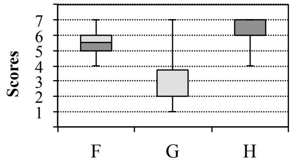


Figure 3: Extrema and quartiles attributed to the statements: (F) The system would help me to have more control on my digital footprint, (G) I would dislike that the system would store my personal data, (H) I would have a better overview over which websites know my personal data using the system

of reused passwords per user are Q_1=3, Q_2=4, Q_3=5. We further asked the participants to indicate their degrees of agreement to different statements by using a seven-point Likert scale (1 corresponds to a full disagreement, 7 to a full agreement). The extrema and quartiles of their ratings are compiled in Fig. 2. The results show that a majority of our participants indicated not to know which personal information they have previously disclosed to websites (question A) and they confirmed that they do not have a good overview of the digital footprint they leave in the Internet (D). Additionally, most of them also expressed their dissatisfaction with the control they are currently able to exercise over their digital footprint (B) and would like to have more control over it (E). This lack of control overall bothers them (C).

Experience with APRIM
We have measured the perceived usability of APRIM using the *System Usability Scale* (SUS) [5]. As a result, APRIM was rated with a SUS-score of 85%, which reflects a good usability [3]. We have further submitted specific statements about APRIM to the participants. The corresponding results are displayed in Fig. 3 and show that most participants agree that APRIM would help them to have both a better control over their digital footprint (F) and a better overview over which websites know their personal data (H). For most of them, it would also not be an issue that APRIM will store and have access to their personal data (G).

Conclusions and Future Work

We have proposed an account and personal information manager called APRIM. In addition to support users in remembering their login information, users can visualize which websites have access to which personal data items and hence, provide them an overview of their digital footprint. Our solution is based on a browser extension, a local database, and a new compact matrix-based GUI. As compared to existing solutions, we have further introduced a temporal component by displaying to the users the time elapsed since the first disclosure, the last access to the website, as well as the last update of the corresponding password. The results obtained in a lab study with 18 participants confirm the insufficiency of existing solutions. Moreover, the usability of APRIM has been rated as good by our participants and their feedbacks is encouraging.

In the future, we plan to further refine APRIM. Its current version allows users to audit manually entered personal data. We plan to develop additional solutions, so that users could not only review them, but also directly take action by, e.g., erasing their digital footprint. By doing so, users would regain control over their data. To reach this goal, further work is needed as the current state-of-the-art requires users to manually ask for data erasure and have no guarantees that it is done in practice. Moreover, the current version of APRIM focuses on users' self-disclosed information, i.e., data manually entered by the users. Their digital footprint, however, also include information collected by visited websites. These data should also be displayed to the users to further increase their awareness. While increasing users' awareness is a first step to help users in better controlling their digital footprint, additional efforts should be conducted to assist users in better protecting their online privacy. To further evaluate APRIM, we plan to deploy it in a long-term user study, in which a representative set of participants could test APRIM in real-world conditions.

Acknowledgements

We would like to thank the participants of our user study. Christian Hartlage was partially supported by the University of Bonn.

References

[1] *1Password*. Online: https://1password.com (accessed in 2.17).

[2] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures. In *Proceedings of the 33rd ACM Annual Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*. 1803–1808.

[3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4, 3 (2009), 114–123.

[4] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *Proceedings of the 33rd IEEE Symposium on Security and Privacy (S&P)*. 553–567.

[5] John Brooke. 1996. SUS - A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.

[6] Lars Brückner and Marco Voss. 2005. MozPETs - A Privacy Enhanced Web Browser. In *Proceedings of the 3rd Annual Conference on Privacy and Trust (PST)*. 1–4.

[7] Shirley Gaw and Edward W. Felten. 2006. Password Management Strategies for Online Accounts. In *Proceedings of the 2nd Symposium on Usable Privacy and Security (SOUPS)*. 44–55.

[8] Jan Kolter, Michael Netter, and Günther Pernul. 2010. Visualizing Past Personal Data Disclosures. In *Proceedings of the 5th International Conference on Availability, Reliability, and Security (ARES)*. 131–139.

[9] A. Koval. *Figaro's Password Manager 2*. Online: http://als.regnet.cz/fpm2/ (accessed in 11.16).

[10] J. Kovarik. Online: http://glyphicons.com (accessed in 2.17).

[11] Lastpass. *Password Management Service*. Online: https://www.lastpass.com (accessed in 11.16).

[12] Zhiwei Li, Warren He, Devdatta Akhawe, and Dawn Song. 2014. The Emperor's New Password Manager: Security Analysis of Web-based Password Managers. In *Proceedings of the 23rd USENIX Security Symposium (USENIX Security)*. 465–479.

[13] Sören Preibusch, Kat Krol, and Alastair R. Beresford. 2012. The Privacy Economics of Voluntary Overdisclosure in Web Forms. In *Proceedings of the the 11th Workshop on the Economics of Information Security (WEIS)*.

[14] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. 1996. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly* (1996), 167–196.

[15] StatCounter. 2015. *StatCounter Global Stats*. Online: http://gs.statcounter.com/ (accessed in 11.2016).

[16] Elizabeth Stobert and Robert Biddle. 2014. A Password Manager that Doesn't Remember Passwords. In *Proceedings of the ACM Workshop on New Security Paradigms*. 39–52.

[17] KeePassX Team. *KeePassX - Cross Platform Password Manager*. Online: https://www.keepassx.org/ (accessed in 11.16).

[18] E. Zachte. *Wikimedia Traffic Analysis Report*. Online: http://stats.wikimedia.org/archive/squid_reports/2015-06/SquidReportClients.htm (Accessed in 11.16).