MoiPrivacy: Design and Evaluation of a Personal Password Meter

Ankit Kariryaa
University of Bremen
Bremen, Germany
kariryaa@uni-bremen.de

Johannes Schöning
University of Bremen
Bremen, Germany
schoening@uni-bremen.de

ABSTRACT

Passwords commonly contain personal information. However, there is limited awareness about its detrimental effect on the user's online security. Current password meters do not take into account personal information and, therefore, their users are susceptible to targeted password guessing. In this paper, we present the MoiPrivacy password meter, that extends a neural network- and heuristic-based approach and considers a user's personal information, while calculating the password strength and feedback. To do so, we analyzed the type of personal information used in passwords through an online survey (n = 62). We conducted a second user study (n = 49) for evaluating the MoiPrivacy browser extension. Our results show that MoiPrivacy significantly limits the inclusion of personal information in passwords.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**.

KEYWORDS

Personal information, passwords, password meter, social media, browser extension

ACM Reference Format:
Ankit Kariryaa and Johannes Schöning. 2020. MoiPrivacy: Design and Evaluation of a Personal Password Meter. In *19th International Conference on Mobile and Ubiquitous Multimedia (MUM 2020), November 22–25, 2020, Essen, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3428361.3428397

1 INTRODUCTION

In the last decades, social media has become ubiquitous with widespread internet access on mobile and desktop devices [1, 23]. Facebook currently boasts over 2.41 billion active users [27], while other social media platforms have hundreds of million users. Among various other purposes, social media is used to communicate and interact with friends, family, and colleagues. During these interactions, people often share various kinds of personal information about their daily activities and life [14].

While usage of social media could potentially cause negative consequences, ranging from denial of job applications [42] to rejection of visa applications [36], it also implies a big risk for online security, as personal information is commonly used in passwords [3, 28]. Former studies have reported that as high as 90% of all passwords are based upon personal information [2]. While this ratio has reduced in the last years, recently conducted studies also report that more than one-third of the passwords contain basic and sensitive personal information, such as name, date of birth, and phone number of the user [28]. In targeted attacks on individuals, attackers often exploit this knowledge and use personal information obtained through social media and other sources in their attack. Research has shown that having access to someone's personal information increases the success of password cracking in the 20 attempts by more than 200% and in the first 100 attempts, it is increased by more than 600% [19].

More recently the focus has been on using personal information datasets for cracking leaked passwords. In one of the first works in this direction, Castelluccia et al. [3] used 3,140 leaked passwords of Facebook users to quantify the effect of personal information in password cracking. They found that 35% of the passwords had some similarities with personal information attributes that they collected from Facebook profiles of those users. Additionally, their password cracking techniques gained up to 30% when using personal attributes. Another important finding of their work was that in a small, yet a significant number of cases, the username and password were very similar. In their study, they considered the effect of Firstname, Lastname, Username, Friends, Edu/Work, Contacts, Location, Birthday, and Siblings.

Password meters can be used to inform users about the strength of their passwords and have gradually increased in accuracy and complexities over the past few years [40]. Many password meters provide feedback about the use of dictionary words or warn about re-using a password from another account [30]. However, no previous work has tackled the issue of personal information in passwords.

In this paper, we present the MoiPrivacy password meter. MoiPrivacy extends the current state of the art in password meter research by including personal information in the password strength estimation and textual feedback. To the best of our knowledge, MoiPrivacy is the first password meter that aims to improve the online security of users by using personal information from their social media profiles.

Our contributions are two-fold; (1) we surveyed for the type of personal information used in the passwords (n=62) and found that information likely to be used in password is available on social media for the majority of the participants. (2) We present the design and evaluation of the MoiPrivacy password meter that provides personalized feedback and strength estimations. We evaluated

MUM 2020, November 22–25, 2020, Essen, Germany

ACM ISBN 978-1-4503-8870-2/20/11.
https://doi.org/10.1145/3428361.3428397

MoiPrivacy through a user study (n=49) and found that feedback about the use of personal information in passwords significantly limits its use in the passwords.

2 RELATED WORK

As relevant prior work, we firstly summarize the current state of password creation strategies, focusing on password semantics. We then present the state of the art in password meters. Lastly, we reflect on more general work using social media for improving online security.

In 2004, Brown et al. [2] asked 218 participants to describe the type of information they used in PINs and passwords for services ranging from email accounts to gym locker codes. Two thirds of the passwords were based upon the names of the user and 90% of the passwords had some kind of personal information including information regarding themselves, their relatives, lovers, friends, pets, and particular products, locations, organizations, activities and celebrities. Their research provides an important insight into the process of password creation. However, these results may not be applicable in full anymore, due to the increasing awareness about the threats and the continuously evolving password generation strategies.

The question of password composition can also be investigated through the semantic lens. Here, instead of asking the user to describe their passwords, researchers study a large corpus of leaked passwords to understand the constituents of the passwords. Veras et al. [33] studied the semantic patterns in RockYou leaked passwords dataset, comprising of 200 Million passwords. They found, that the most probable semantic categories were: "first name", "city", "surname", "s.be.v.01", "s.love.v.01", "s.love.n.01", "s.baby.v.01", "month" etc. It is quite likely that first name, city and surname represent personal information and even that the noun in "s.love.n.01" is personal. Their study provides a comprehensive overview of what passwords are, or at least a vast majority of them.

To understand the use of personal information in password another strategy is to study the leaked password and personal information datasets. Li et al. [19] studied the role of personal information in 131,389 leaked passwords from the official Chinese Railways website (12306.cn). For their study they considered *name, email address, phone number, account name* and *government-issued ID number*. They found that 10.5% of passwords perfectly matched to exactly one type of personal information and about 60% of the passwords contained personal information. By using personal information, their system was able to crack 4.8% of passwords within the first 5 guesses. And in 100 guesses, they were able to crack 634% more passwords than the original PCFG approach on which they based their algorithm.

In a later study, Su and Zhu [28] studied 200 million leaked passwords and 20 million personal information records from users in China. They found that 37.2% of passwords contained personal information including name, cell phone number, date of birth and email address. By matching personal information on various leaked passwords datasets they were also able to study multiple passwords of some individuals. 17.25% of passwords were from multiple accounts of some users and the users had 54.89% password reuse probability in their dataset. Overall, they were able to improve the password guessing efficiency by 12.41% using personal information. The password reuse probability was similar to one found by Shay et al. [26], who conducted a survey on 470 University students, staff and faculty members and showed that 60% used one password with slight changes for different accounts. While most of these approaches extended the PCFG algorithm [38, 39] to include categories related to personal information, Houshmand and Aggarwal [15] demonstrated a general technique to extend the PCFG grammars without changing the PCFG algorithm. They also studied the effect of old password patterns for targeted attacks. On a limited test set, their algorithm achieved significant improvement in cracking passwords modified from earlier passwords.

With the exception of Castelluccia et al. [3] who explicitly obtained personal information from Facebook, most previous studies have relied on leaked datasets containing limited personal information. Thus, they only report the lower bound of the use of personal information in passwords and the actual value is likely to be much higher.

2.1 Password Meters

Password meters are tools that provide real-time feedback on a password and, therefore, help people choose secure passwords. They are a common feature on most widely used websites. Password meters differ widely in their complexity and accuracy [40]. While initial password meters mainly used a count of various character classes to estimate the strength of the password, over the years, the approaches for calculating the strength and providing feedback have improved significantly. In the recent years, password meters based upon a range of approaches including context-free grammars [35], Markov models [4], heuristics [40], and neural networks [21] were proposed.

A common feature of password meters is a colored bar that represents the strength of a password. Research has indicated that feedback through a colored bar led people to create significantly stronger passwords. However, the effect of the bar varied for websites of different importance and depended upon the context [8]. Moving beyond the *coloured bar*, zxcvbn password meter [40] also incorporated textual feedback. They were also instrumental in using an advanced heuristic to estimate the strength and provide textual feedback to the user. Their heuristics also consider the name and email address of the user and allow for user inputs. A different approach to password meters was presented by Kamundari et al. [17]. Instead of providing feedback on the strength of the password, their password meter termed Telepathwords predicted the next character in the password. Indicators above the password field were placed to tell the user how many characters were guessed correctly, in addition to showing the best guesses for the next character. The main idea of this approach was to help users to create stronger passwords by demonstrating that the system could easily guess the next characters in case of a weak password. They found that using Telepathwords people created a far stronger password than commonly used character composition policies.

In a recent study, Ur el al. [30] developed and evaluated a data-driven password meter. They used a recurrent neural network modeled for adversarial password guessing and combined it with 21

heuristics to calculate the strength of the passwords. These heuristics were based upon keyboard patterns, dictionary words, names, and common passwords among other things. In an extensive online study with 4,509 participants, they evaluated 18 conditions related to password-composition policy, type of feedback, and stringency. One of their key findings was that the text feedback led to more secure passwords than a colored bar alone.

While accuracy and feedback of password meters have improved significantly in the past, none of them have directly tackled the issue of frequent use of personal information in the passwords.

2.2 Using social media for improving online security

Researchers have proposed various applications for improving online security using social media. For mitigating forgotten password attacks, in the patent "Techniques for mitigating forgotten password attacks", Gauvin [12] proposed to use the public information associated with the user for verifying the identity of the user. Dunphy et al. [7] used social media to study experience-centered insights into security practices. They found that social media posts could be used as a resource of naturally generated reflections on security practices and workarounds. Using qualitative content analysis they found that on social media people shared specific difficulties in password creation or retrieval. Among other practices, people discussed password sharing and revoking scenarios. Franchi et al. [11] reviewed social attacks that are conducted on social networks. They suggest, that the user's carefree attitude in sharing information, sub-par security measures and high value of the published information are the main influencing factors for such attacks. Queiroz et al. [24] proposed a system for using social media for finding information related to software vulnerabilities. Using the support vector machine their system achieved 94% accuracy in classifying tweets related to software vulnerabilities into relevant and non-relevant. These examples indicate that social media can play multiple roles in improving user security.

While tremendous progress has been made in using personal information for cracking individual passwords, the role of personal information in password meters remains largely unexplored. Personal information is frequently shared on social media, however, it is not clear to what extend is the same information used in passwords. To fill this gap in the literature, we created a survey that sheds light on the diversity of personal information available on social media and its use in passwords.

3 ONLINE SURVEY

In the previous section, we described the extent of personal information in passwords. To further improve this knowledge, we conducted an online survey to understand

(1) how likely it is today that personal information is used in passwords,
(2) which types of personal information are used in passwords and
(3) whether the information is available on social media.

The purpose of this survey was to inform the design and development of our MoiPrivacy password meter by highlighting which particular types of information should MoiPrivacy use for password feedback and strength estimations. In the later sections, we refer to it as the InfOnPWD survey.

To protect the privacy of the participants, in this survey we did not ask for their actual passwords. Instead, we asked them to imagine that they were creating a password for their primary email provider that they use for everyday and official communication. After the participants had imagined the password, they rated the likelihood of a certain type of personal information to be used in that password on a five-point scale from "Not at all likely" to "Extremely likely". In a multiple-choice grid, they were asked to rate 23 types of personal word categories ranging from the first name to name of the favorite restaurant and 18 types of personal number categories ranging from date of birth to number from an address in the neighborhood. The rows of the grid were randomized. Next, for each of 41 personal categories, we asked the participants if this information was explicitly available, if it could be determined or if it was not available on their social media profile. The 41 personal word and number categories were based upon a review of personal information reported to be in previous studies.

In the last part of the survey, we asked the participants if they used personal information in their real passwords and to what extend. And whether they would use a password meter that could warn them about the use of personal information in their passwords.

3.1 Participants

We advertised the survey on our university's mailing lists and personal social media pages. The study ran for one week and 62 people (39 male, 18 female, 5 undisclosed) participated in the study. The mean age of the participants was 27 years (std 7.7). 20% had a high school diploma, 24.2% had a Bachelor's degree, 33.9% had a Master's degree, 16% had a Doctorate, and the rest did not disclose their highest level of educational qualification. Participants were not compensated for participating in this survey.

3.2 Results

We found that the user's date of birth, followed by the date of an important life event and date of birth of the significant other was the most likely to be used in a password. Numbers, containing personal information, were generally more favored over words containing personal information. While Brown et al. [2] had reported that more than two-thirds of the passwords were based upon names, in our survey *Firstname* and *Lastname* were less likely than many other categories such as date of birth, name of an animal and term from a hobby.

We constructed a score that indicates the likelihood of a certain type of information in a password. For doing so, we weighed each type of personal information by its likelihood, extremely likely received the weight of 4 and not at all likely received the weight of 0. Figure 1 shows the 20 types of personal information that are most used in a password with their scores. The mean total weighed personal information (TWPI) per participant was 22.0 with a standard deviation of 11. Min TWPI was 0 and max TWPI was 50. We did not find a significant correlation of TWPI with age, gender or education. Participants reported on average 4.8 items (SD= 5) that they were at least somewhat likely to use in a password. Majority of the participants reported that the information which

	Information Type	Score		Information Type	Score
1	Date of birth	1.67	11	The Name of a family member	0.77
2	Date of an important event in life	0.98	12	Name of a character, creature or thing from an online game	0.77
3	Date of birth of the significant other	0.97	13	Name of your significant other	0.70
4	Prominent number from a book, movie or series	0.97	14	A number from your address	0.70
5	A name or term from your hobby	0.93	15	Name of a pet	0.66
6	Name of a character, creature or thing from a book, movie or series	0.90	16	Your phone number	0.66
7	Date of birth of a family member	0.85	17	Your lastname	0.62
8	Name of an animal	0.82	18	Your account name from any online service	0.62
9	Name of a mythical creature	0.79	19	Date or number related to a famous historical event such as independence day	0.62
10	Your firstname	0.77	20	A name or term from a sports team	0.57

Table 1: Results from our InfOnPWD survey. The table shows 20 personal information types that are most likely to be used in a password. We found that dates and number are more likely to be used than names and words, as indicated by the higher mean likelihood score (min 0, max 4).

was likely to be used in a password was available on their social media profile.

We had also asked the participants if their actual passwords contained personal information. They could answer this question in the following five categories; 0%, 1-25%, 26-50%, 51-75% or 75-100%. Most of the participants reported that 25-50% of their passwords contained personal information. The majority of participants strongly agreed that a password meter which could warn them about the use of personal information in passwords would be helpful. The results of the InfOnPWD survey supported our theory that information likely to be used in passwords is available on social media and that social media can be used for personalization of the password meter.

4 MOIPRIVACY

MoiPrivacy, the personal password meter, is implemented as a browser extension. It provides feedback on the strength of the password and suggestions for improving the password with a tooltip, as shown in Figure 1. The extension can be personalized by adding information from a Facebook profile. It is important to note, that the personal information stays with the users and is not uploaded to any external service. By doing so, the password meter can also give feedback on various types of personal information in passwords as derived in the InfOnPWD survey.

4.1 Implementation

We developed MoiPrivacy as a Firefox browser extension using the WebExtensions API [22]. WebExtensions API is a cross-browser system for developing extensions and therefore MoiPrivacy can be easily extended to support Chrome and Opera. Similar to most standard extensions, MoiPrivacy has two main parts; the content scripts and the background scripts. The content scripts are used to manipulate the view of the active browser web page and for handling user interaction. The background scripts are used for computationally intensive tasks, e.g. for calculating the strength of the passwords, storing the user data and the state of the extension. Besides these, the MoiPrivacy extension offers an options page that gives general information about the extension and provides functionality to control the behavior and to upload data to personalize the extension.

4.2 Password strength estimation

We extend the work of Ur et al. [30] for estimating the strength of passwords. Ur et al. used a recurrent neural network (RNN) model in parallel with a heuristic-based method for estimating the strength of the passwords. Since there are no established neural network methods for personalized estimation of password strength, we decided to use a non-personalized RNN model. We used the RNN model trained and published by Melicher et al. [21] for predicting the strength of a password under the 1class8 composition policy. The 1class8 policy implies a minimum of eight characters. It is a widely used but relatively lax password-composition policy. We extended the heuristics used by Ur et al. to handle the cases when the passwords contained personal information as described below.

We asked the users to personalize the MoiPrivacy extension with information available on their Facebook profiles. Our decision to use Facebook profiles for personalization was based upon the findings of the InfOnPWD survey that information which was likely to be used in a password is available on social media profiles. Facebook allows its users to download all their data as a zip file [9]. Users downloaded their Facebook data using this feature and then upload it to the extension. It is important to note, that the Facebook data is processed and stored locally in user's browser. Based on the feedback we received from our InfOnPWD survey, we categorized the information into the primary and secondary. Primary information was extracted from the user's profile page

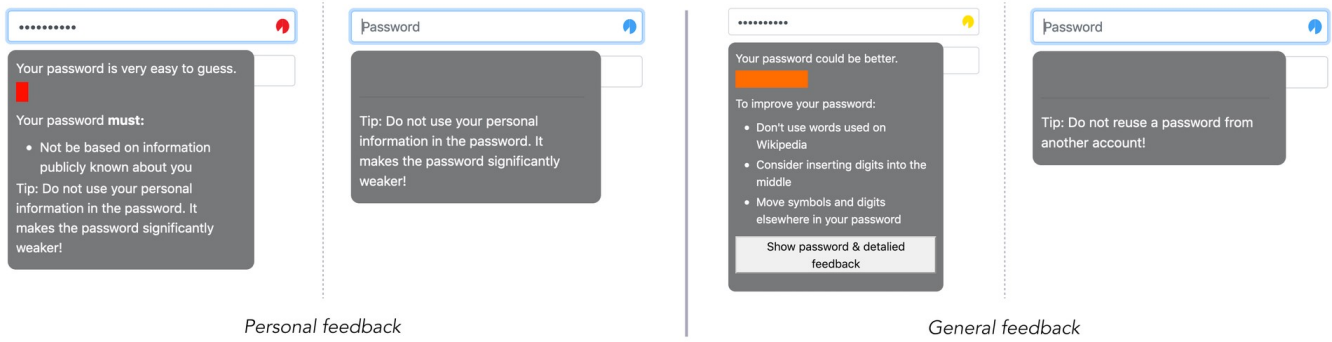

Figure 1: Study conditions: MoiPrivacy with personal information (PI, left) and without personal information (GI, right).

(https://www.facebook.com/fbusername/about) and friend page (https://www.facebook.com/fbusername/friends). The profile page of a Facebook user can include meaningful information such as *family and relationships information*, *life events with dates*, *contact and basic info*, *places* including hometown and current city, *workplace* and *school*. A profile page also contains interests categorized into *sports, music, movies, books, events and check-in*. The friends page contains the name of all the user's Facebook friends. We extracted the information from these pages and removed all non-alphanumeric characters. Furthermore, we extended this primary information set with their 4+ length sub-string, limited to sub-strings starting at the beginning of the words. Secondary information consisted of text extracted from all other Facebook pages (still removing all non-alphanumeric characters). We performed part-of-speech tagging on the secondary information with the *posjs* JavaScript library [37] and we removed all other parts of speech except for nouns and cardinal numbers.

For calculating the heuristic score we extended the 21 heuristics used by Ur et al. [30] with two additional heuristics for primary and secondary information. 21 heuristics of Ur et al. considered length, number of character classes, presence of dictionary word and presence of a blacklisted password string among others. To determine the weight of each heuristic, Ur et al. used regression on the score of heuristics and score from CMU's Password Guessability Service (PGS) [29]. If the password contained a blacklisted password (e.g. "password" or "123456") they calculated the heuristic score after removing the blacklisted word from the user's password. The per-character maximum negative weight was assigned to length of matched dictionary words (-0.55), followed by length of matched common passwords excluding the blacklisted passwords (-0.39). If the password contained primary information, our first novel heuristic calculated its score after removing the primary word. And if the password contained secondary information our second heuristic gave it a score of -0.45 relative to the length of the matched word. The score summed up for all the 23 heuristic gave us the predicted strength on a scale of 0 to 100.

4.3 Visual Design

The design of MoiPrivacy was inspired by publicly available password managers such as 1Password (https://1password.com/) and Psono (https://psono.com/). By default, the extension tooltip was hidden and it would show up on the screen the first time the user would click or focus on the password field. Users could also toggle the visibility of the tooltip through a small icon in a password field.

The main screen displayed the strength of the password and provided detailed feedback to help the user in improving the password (see Figure 1). If the password field was empty, the main screen would display a general tip to the user (see the second and fourth panel in Figure 1). Once the user would start typing a password, the main screen would show the password strength and textual feedback. We provided information about the strength of the password in three places. Firstly, the color of the MoiPrivacy extension icon, appended to the password field, changed from red to yellow to green depending upon the strength of the password. Changing icon color was helpful to show password strength even when the main screen was hidden. Secondly, we employed a colored bar [6, 8, 31] that fills up and changes color from red to green to indicate the strength of the password. Thirdly, we provided text feedback about the strength of the password along with text feedback about which specific aspects of the user's password could be improved. The color bar and text feedback were similar to the one used by Ur et al. [30] with the exception that we also provided detailed feedback if the user used personal information in the password. If the password did not meet the composition requirement such as it was less than eight characters or contained personal information or a blacklisted password, then we would show the requirement that the password must fulfill (see Figure 1 panel one). We showed feedback about other specific problems once all the requirements were fulfilled, or if more than two-thirds of the bar was full. At this stage, the user could see the detailed feedback and the password by clicking on the "Show password & detailed feedback" button.

In addition, the extension had an options page located in the extension menu of the browser settings. Here we provided details about the extension, option to upload the Facebook data as a zip file to personalize the password meter and other settings. We also displayed the number of words the extension had extracted from the Facebook profile on this page. Figure 2 shows the Facebook data upload option on the options page.

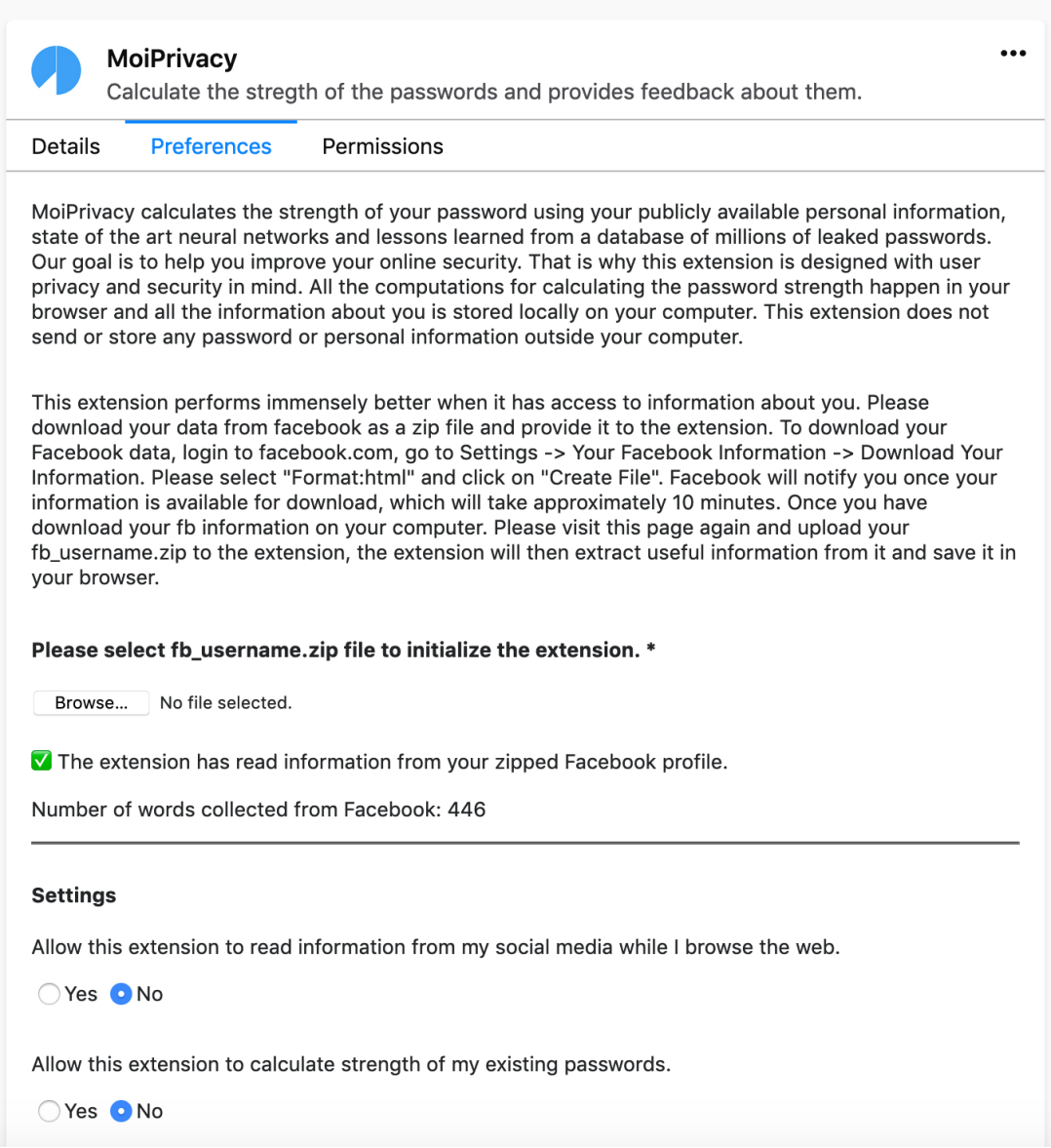

Figure 2: Facebook data upload options on the options page of the MoiPrivacy extension.

5 STUDY

We performed a between-subjects study with the MoiPrivacy extension to test the following hypothesis:

H1: Users who are given personalized password strength estimation and feedback are less likely to use personal information in their passwords.

We also investigated the effect of personalized feedback on password strength, password creation, and behavioral measures such as modifications and time taken during creation. In this section, we describe our methodology and present our results.

5.1 Conditions

We evaluated two conditions: First, the MoiPrivacy extension with personal password strength estimation and feedback as described earlier. We refer to this as the PI condition (personal information feedback). Second, a baseline condition without personal feedback and strength estimation termed as GI condition (general information feedback). In GI condition we only provided general feedback and password strength estimation (similar to the work of Ur et al. [30]). We conducted a between-subjects online study and randomly assigned participants to one of two conditions at the start of the study.

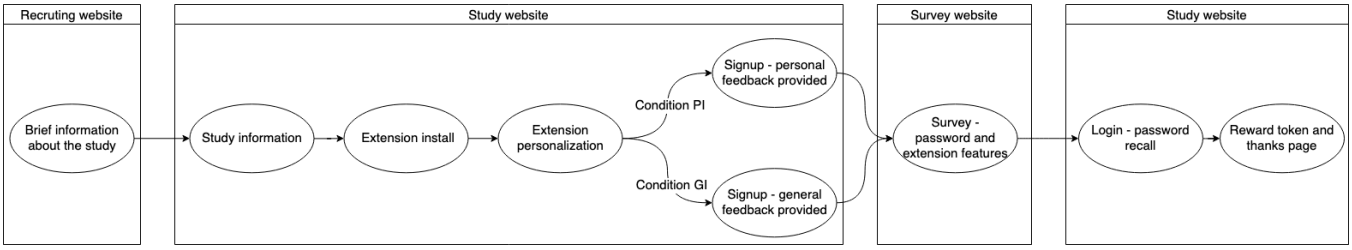

Figure 3: Setup of the user study for evaluating the MoiPrivacy password meter.

5.2 Task & Procedure

Figure 3 shows the study setup. On the introductory page, participants were informed about the tasks and they were provided information about the type of data that was collected. Upon their consent and additional acknowledgment that they were at least 18 years of age, participants were forwarded to the extension install page, where they could download the extension, and read text instructions or watch a video to install the extension. The extension was signed by Mozilla Firefox for self-distribution. Upon installation, the extension would send a confirmation message to the study website, and participants were only allowed to proceed after they had installed the extension.

The participants were then asked to personalize the extension with their Facebook data. We provided text and video instructions on how to download the Facebook data zip and upload it to the extension. The participants were once again informed that their personal information will stay on their computers and not on an online platform. Figure 2 shows the options page of the extension where participants would upload their data (video instructions were provided only on the study website). Upon successful personalization, the extension would send a confirmation message to the study website after which the participants were allowed to proceed to the signup page. During the trial runs, we observed that on average the personalization step took approximately 10 minutes. The majority of this time was consumed by Facebook for preparing the data zip file. We informed the study participants about it and they were asked to return to the study website once they had downloaded the data from Facebook. We required participants from both conditions to personalize the extension to ensure that both conditions were equivalent, and that there was no systematic variation due to the extra time spend in personalization.

On the signup page, participants were asked to create a username and password and re-enter the password. They were advised to create a unique username and a strong password as they would normally do for an important email provider. Participants were also informed that they would need the username and password again in the future. The installed extension provided feedback to the participants while they created the password, but not when they re-entered the same password. The participants who were assigned the PI condition received personal, as well as general feedback, while the participants who were assigned the GI condition only received the general feedback (see Figure 1). The calculated password strength also depended on the assigned condition. On this page, we also recorded keystrokes and mouse clicks. Upon successful signup, they were forwarded to the external survey website.

The survey was divided into two sections. The first section focused on the password creation process and the second section focused on the usability of the MoiPrivacy extension. In the first section, the participants were asked if they reused an existing password, modified an existing password, or created a fresh password. We also asked if they stored or wrote down the created password anywhere. Next, they were asked to describe their password in plain text, e.g. name of my pet followed by the year of my birth. In a separate binary question, they were asked if the password contained personal information.

In the second section, we also asked their opinion on various features of the password meter and the feedback. The shown features depended upon the assigned condition. The participant would be shown an image of the feature in question and asked if they noticed the feature, if it was informative, if they modified their behavior due to it and if it helped them to create a stronger password.

After that, the participants returned to our study website and were required to login with the password that they had created during signup. This was done in order to test the recall of the password created earlier. The username on the login page was autofilled and if they forgot their password we allowed the participants to proceed after five incorrect attempts. Upon login, participants received their completion token.

5.3 Participants

We used two approaches to recruit participants for our study. Firstly, we advertised the study on personal social media pages. Secondly, we recruited participants on Amazon Mechanical Turk in order too engage an overall diverse set of users. A precondition of the study was that they should use Firefox and do not use a password manager.

A total of 56 people participated in the study. We filtered out the participants who used a password manager or stored the password somewhere. Thus, we were left with 49 valid responses. The mean age of the participants was 27 years (std 7.7). 57% did not have either a job or degree in computer science or related field. The participants recruited on Mturk received $2.50 for participating in the study.

6 RESULTS

Our findings are summarized in Figure 4. We used Pearson's chi-square test to determine the statistical significance in difference between the frequency of personal information use under the two

	Personal Information in Password		Password guessed (upto 1e14 guesses)		Password creation time (seconds)		Modifications during creation (#characters)		Password length (#characters)		Login Attempts	
	Percentage		*Percentage*		*Mean (std)*		*Mean (std)*		*Mean (std)*		*Mean (std)*	
Personal Information (PI) (n=24)	16.67%		50%		72.3 (81.6)		3.87 (5.26)		12.1 (3.2)		1.54 (1.44)	
General Information (GI) (n=25)	36%	$\chi^2 = 7.05$*	32%	z = -1.20	63.7 (72.4)	t = 0.38	3.24 (5.86)	t = 0.39	12.9 (2.87)	t = -0.9	1.2 (0.7)	t = 1.05
	*p<.05											

	Response to the password creation process (Agreed or Strongly agreed)				
	Creating the password was fun	*Creating the password was annoying*	*Creating the password was easy*	*Creating the password was difficult*	*Creating the password was informative*
Personal Information (PI) (n=24)	7	6	8	6	15
General Information (GI) (n=25)	12	6	15	9	10

	Tip (Agreed or Strongly agreed)			Suggestion (Agreed or Strongly agreed)		
	Tip was informative	*Tip helped create a strong password*	*Created a different password because of the tip*	*Suggestion was informative*	*Suggestion helped create a strong password*	*Created different password because of the suggestion*
Personal Information (PI) (n=24)	19	16	13	9	14	10
General Information (GI) (n=25)	19	19	16	14	9	8

Figure 4: Evaluation of the MoiPrivacy extension. The top table shows the impact of personal information and general information on personal information in password guessability, creation time, modification during creation, length, and login attempts in the recall task. The middle table shows the sentiment of user during password creation. The bottom table demonstrates the usefulness of the personal and general tip and suggestion.

conditions. For all statistical methods, except where explicitly mentioned, we used the implementation available through Scipy python library v1.2.1 [34]. We found, that participants with personal feedback, are less likely to use personal information in the password. $X^2(1,N= 49)=7.05$, $p < .05$. Out of the 49, 24 had received the PI condition and 25 had received the GI condition. 36% of the participants in the GI condition, while only 16.67% of the participants in the PI condition used their personal information in the password.

To measure the strength of the participant's password we used the CMU's PGS [29], which calculates minimum attempts needed to guess the password over multiple guessing attacks. It is considered a conservative estimate of an experienced attacker [32]. In both conditions, users were able to create relatively secure passwords. For comparing password guessability, we used the Cox Proportional-Hazards Model, which was developed for survival analysis and used in the past for comparing password guessability [20, 30]. We used a right-censored model available through the *lifelines* python library [5]. The passwords created under the PI condition were found to be less strong than the passwords under the GI condition (*coef* = 0.55, *exp(coef)* = 1.73, *se(coef)* = 0.46). However, the two conditions did not differ significantly (p=0.23). Overall, in GI condition 32% and PI condition 50% of participant's passwords were guessed through PGS after 10^{14} attempts.

Besides these two main metrices, we compared the length, creation time and modifications of participant's password and the login attempt count in the password recall step. To test the differences between the two conditions, we performed Welch two sample t-tests. We did not find any significant difference in these measures. Participants in PI (M = 72.30 sec, SD = 81.66) took slightly longer than participants in GI condition (M = 64 sec, SD = 72.45). This also could be the effect of participants modifying their passwords, while creating them, reacting to the advice of the MoiPrivacy extensions. The mean modification count was 3.87 (SD = 5.26) in PI and 3.24 (SD = 5.86) in GI. The number of deletions during the password creation step was used to measure modification. Also there was no significant effect on the length of the passwords, t(49) = 0.9, despite participants in PI (M = 12.1, SD = 3.2) using slightly shorter passwords than participants in GI condition (M = 12.9, SD = 2.87). Participants in PI condition (M= 1.54, SD = 1.44) also took more attempts to login into their accounts than in GI condition (M = 1.2, SD = 0.7) but the difference was found to be non-significant.

Besides, these behavioral measures we collected several self-reported measures and qualitative feedback. In general, more participants reported that creating the password was more informative with the PI, than the GI condition (Figure 4 middle row). But fewer participants said that creating the password was fun and easy in the PI condition compared to the GI, which could be the effect of more frequent editing the password while creating it. Nevertheless, as we expected, we did not find any significant differences between both conditions on our qualitative scales.

We also evaluated the usefulness of personal tips and suggestions and compared them to the general tips and suggestions (Figure 4 bottom row). In both conditions, participants reacted positively to the *Tip* and the textual suggestion. 80% and 74% of the participants

in the GI and PI condition respectively said that they noticed the tip. Most people agreed or strongly agreed that the personal suggestion helped them create a strong password and that they created a different password due to the suggestion. Furthermore, we also asked the participants if they learned something new from the feedback. Participants in the GI condition learned tricks to create strong passwords such as "to not use dictionary words" and "put the symbol earlier in the sequence." In the PI condition, the participants learned the "importance of NOT using personal info in passwords" and "not using personal info. longer password is better (though I like my passwords short nonetheless)."

Overall we can conclude that the users of MoiPrivacy password meter created secure passwords with significantly less personal information as compared to the baseline condition. We could observe in the behavioral data, that the password creation process was affected by the password meter, as users more often changed their password in the creation phase.

7 DISCUSSION

A good password has two key characteristics, it is hard to guess and easy to remember [2, 41]. Meaningful information, such as the name of a loved one, is easier to remember [13] and probably, for this reason, personal information is commonly used in the passwords. While this problem has been highlighted various times in the past, it remains a major challenge to online security (see Section 2).

The rise of the Internet and social media has made the personal information of a large population publicly available. As Ronald Rivest noted in 2001, the digital revolution reverses defaults; *"What was once hard to copy is now trivial to duplicate. What was once forgotten is now stored forever. What was once private is now public. What was once simple and secure is now complex and insecure"* [25].

In the last years, password meters have improved significantly. Recent password meters have also explored some ideas of limiting personal information in the password. For example, zxcvbn [40] considered it through the optional feature of adding user input and Ur et al. [30] through the use of common names and dictionary words in strength estimations. However, none of them directly tackle the issue of widespread use of personal information in the passwords.

In this paper, we present the design and evaluation of MoiPrivacy personal password meter that limits users from using personal information in the passwords. MoiPrivacy is personalized with the user's information from their social media profile. This information is then used to better assess the strength of the user's passwords. Traditionally password meters are integrated at the password creation step of online platforms. However, this approach limits the possibility of providing personal feedback. To overcome this problem, we developed MoiPrivacy as a browser extension, which integrates it with the browser rather than a specific online platform. MoiPrivacy runs independently in the user's browser, therefore keeping sensitive personal information on their machines rather than an online platform and in this way, it provides standardized strength calculations and suggestions for all online platforms. It is an important feature since online password meters are notorious for their varied password strength feedback [40].

We considered various options to personalize the MoiPrivacy extension. These options included a questionnaire about the user's personal information, collecting information through social media APIs, information scraping during web browsing, and requesting the user to upload a complete profile. We decided to present accurate and complete personal feedback based on most of the publicly available personal information about the user. In our case, the completeness of information outweighed the ease of collection. Thus, we decided to ask the user to upload their social media profiles, even though it was not the most convenient option. The decision of using social media for personalization was motivated by our InfOnPWD survey, where we studied the likelihood of various types of personal information in the password and found that, for the majority of the users, information likely to be used in the passwords is available on social media. While personalization was necessary for the user study, MoiPrivacy can be deployed without it and allowed to collect useful information over time.

We evaluated the MoiPrivacy extension through a user study (n=49). We found that users receiving personalized feedback are less likely to use personal information in the password, thus conforming to our initial hypothesis. People, who received personal feedback, more frequently modified their password. Since people commonly use personal information in their password, it could be that personal feedback nudged them away from this behavior thus leading to more frequent modifications. It could also explain why participants considered the condition with personal feedback to be slightly less fun than the baseline.

While MoiPrivacy lead to significantly less percentage of personal information in the password, its impact the password's length, strength, or short term recall was minimal. In a non-targeted attack simulated through the CMU's PGS service, the password in the PI condition were found to be less strong then the passwords in the GI condition, however, the difference was not significant. We think that in PI condition, due to feedback about the personal information, more people used dictionary words and other common patterns, which might make them slightly more exposed in a non-targeted attack (i.e. when personal information of the user is not used in the guessing attack). Thus, removing personal information should not be seen as a generic way to improve security. Instead, it could be seen as a way of limiting the impact of targeted password guessing attacks.

Figure 5 shows the personal information in passwords over the years. The percentage of personal information has decreased over the years due to increasing awareness about online security. We found out that with the help of a personal password meter we can limit it further.

Lately, there has been growing awareness about the need for strong and unique passwords. While more and more people are nowadays using password managers, there is still a need to create passwords outside a password manager. Even when using a password manager, it is desirable to retain full control of important accounts, such as identity and financial accounts. In addition, password managers are a single point of failure, so if the master password of a password manager is compromised, it would also compromise all of a user's online credentials. We argue that password meters are still relevant in age of password managers. The participants of our study reported on meaningful lessons that they

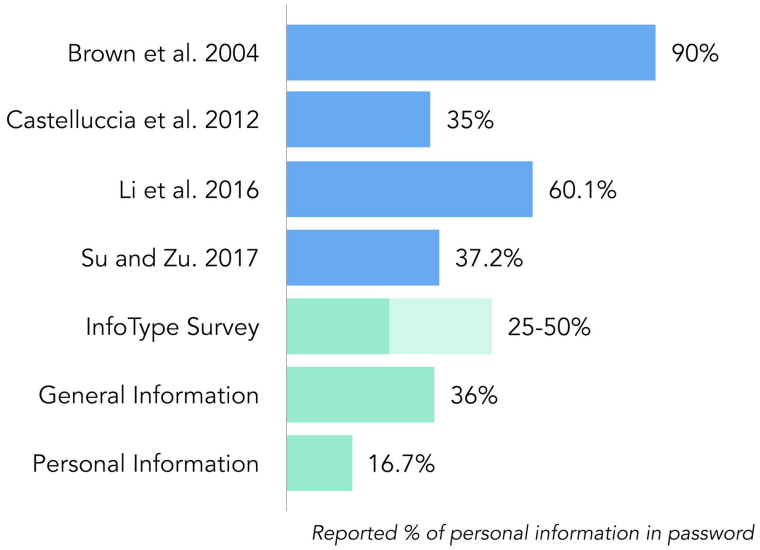

Figure 5: Personal information in passwords over the years. The percentage of personal information has decreased over the years and with the help of a personal password meter, we can limit it significantly. Note that Castelluccia et al. used a stricter matching metrics than the others.

learned through use of MoiPrivacy password meter, so it is conceivable that a good password meter can help change the long term behavior of a user. The lessons learned from the feedback of a password meter can be applied while creating a password for offline accounts such as for a personal computer or a desktop-based application.

In the future, we envision the MoiPrivacy password meter would be more playful to the users, e.g. making puns on what one had posted on Facebook "last summer".

7.1 Ethical Principles

We took multiple steps to minimize the impact of our personalized password meter and study on the privacy of the users. Firstly, we implemented our password meter as a browser extension, so that the sensitive personal information of the users stayed with the users instead of an external service. This increased the deployment efforts, but we considered it necessary for preserving the user's privacy. Secondly, we advised our participants in the introduction as well as signup step to create unique usernames and passwords, so that their information could not be misused. Besides the username and password of the users, we did not collect any personal information from the users. Lastly, we informed the participants and asked for their consent that their passwords would be stored in plain text. We needed the passwords in plain text for determining their strength using the CMU's PGS service [29].

7.2 Limitations

Our study design was based upon previous research on the evaluation of password meters [16, 18, 30, 40]. Similar to other studies, we asked the participants to create a strong password as they would normally do for a primary email account. However, as noted by Egelman et al. [8], it heavily depends on the use case whether a strength meter has any measurable effect. Thus, while this method of password meter evaluation is considered reasonable [10], the ecological validity of our study is limited. Additionally, we only tested the short term recall of the password, and our password meter's impact on long term memorability is unknown.

8 CONCLUSION

In this paper, we describe the design and evaluation of the MoiPrivacy password meter. MoiPrivacy extends a state of the art neural network - and heuristic-based password meter. It works as a browser extension and it can be personalized with information available on social media for personal feedback and strength estimations. In a user study (n = 49), we evaluated MoiPrivacy and found that with it users created secure passwords with significantly less personal information.

ACKNOWLEDGMENTS

We would like to thank the participants of our studies for their time, our colleagues Gian-Luca Savino and Daniel Diethei for the discussions on this topic, and Daria Vladimirovna Soroko for the help with the figures. This research was supported in part by the Volkswagen Foundation through a Lichtenberg Professorship and by the Federal Ministry of Education and Research of Germany (BMBF) through the Wintermute project (award number 16KIS1127).

REFERENCES

[1] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*. 47–56.

[2] Alan S Brown, Elisabeth Bracken, Sandy Zoccoli, and King Douglas. 2004. Generating and remembering passwords. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 18, 6 (2004), 641–651.

[3] Claude Castelluccia, Abdelberi Chaabane, Markus Dürmuth, and Daniele Perito. 2013. When privacy meets security: Leveraging personal information for password cracking. *arXiv preprint arXiv:1304.6584* (2013).

[4] Claude Castelluccia, Markus Dürmuth, and Daniele Perito. 2012. Adaptive Password-Strength Meters from Markov Models.. In *NDSS*.

[5] Cameron Davidson-Pilon. 2019. lifelines: survival analysis in Python. *Journal of Open Source Software* 4, 40 (2019), 1317.

[6] Xavier de Carné de Carnavalet and Mohammad Mannan. 2014. From very weak to very strong: Analyzing password-strength meters. In *Network and Distributed System Security Symposium (NDSS 2014)*. Internet Society.

[7] Paul Dunphy, Vasilis Vlachokyriakos, Anja Thieme, James Nicholson, John McCarthy, and Patrick Olivier. 2015. Social Media As a Resource for Understanding Security Experiences: A Qualitative Analysis of #Password Tweets. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 141–150. https://www.usenix.org/conference/soups2015/proceedings/presentation/dunphy

[8] Serge Egelman, Andreas Sotirakopoulos, Ildar Muslukhov, Konstantin Beznosov, and Cormac Herley. 2013. Does my password go up to eleven? The impact of password meters on password selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2379–2388.

[9] Facebook. 2008. Accessing and Downloading Your Information. Website. https://www.facebook.com/help/1701730696756992/.

[10] Sascha Fahl, Marian Harbach, Yasemin Acar, and Matthew Smith. 2013. On the ecological validity of a password study. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. 1–13.

[11] Enrico Franchi, Agostino Poggi, and Michele Tomaiuolo. 2015. Information and password attacks on social networks: An argument for cryptography. *Journal of Information Technology Research (JITR)* 8, 1 (2015), 25–42.

[12] William Gauvin. 2013. Techniques for mitigating forgotten password attacks. US Patent 8,555,357.

[13] Richard Hébert. 2001. Code Overload: Doing a Number on Memory. *APS Observer* 14, 7 (2001).

[14] Al Amin Hossain and Weining Zhang. 2015. Privacy and security concern of online social networks from user perspective. In *2015 International Conference on Information Systems Security and Privacy (ICISSP)*. IEEE, 246–253.

[15] Shiva Houshmand and Sudhir Aggarwal. 2017. Using Personal Information in Targeted Grammar-Based Probabilistic Password Attacks. In *IFIP International Conference on Digital Forensics*. Springer, 285–303.

[16] Jun Ho Huh, Seongyeol Oh, Hyoungshick Kim, Konstantin Beznosov, Apurva Mohan, and S Raj Rajagopalan. 2015. Surpass: System-initiated user-replaceable passwords. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 170–181.

[17] Saranga Komanduri, Richard Shay, Lorrie Faith Cranor, Cormac Herley, and Stuart Schechter. 2014. Telepathwords: Preventing Weak Passwords by Reading Users' Minds. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 591–606. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/komanduri

[18] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. 2011. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the sigchi conference on human factors in computing systems*. 2595–2604.

[19] Yue Li, Haining Wang, and Kun Sun. 2016. A study of personal information in human-chosen passwords and its security implications. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[20] Michelle L Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring password guessability for an entire university. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 173–186.

[21] William Melicher, Blase Ur, Sean M. Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 175–191. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/melicher

[22] Mozilla. 2015. Browser Extensions. Website. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions.

[23] Esteban Ortiz-Ospina. 2019. The rise of social media. Website. Retrieved January 12, 2020 from https://ourworldindata.org/rise-of-social-media.

[24] Andrei Queiroz, Brian Keegan, and Fredrick Mtenzi. 2017. Predicting Software Vulnerability Using Security Discussion in Social Media. (2017).

[25] Ronald Rivest. 2001. Whither Information Security? Retrieved January 12, 2020 from http://web.archive.org/web/2007*/http://wean1.ulib.org/Lectures/Distinguished%20Lectures/2001/03.0%20Ronald%20L%20Rivest/6SLIDES/security.ppt.

[26] Richard Shay, Saranga Komanduri, Patrick Gage Kelley, Pedro Giovanni Leon, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2010. Encountering Stronger Password Requirements: User Attitudes and Behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (Redmond, Washington, USA) *(SOUPS '10)*. Association for Computing Machinery, New York, NY, USA, Article 2, 20 pages. https://doi.org/10.1145/1837110.1837113

[27] Statista. 2019. Number of monthly active Facebook users worldwide as of 2nd quarter 2019 (in millions). Website. Retrieved January 12, 2020 from https://www.statista.com/statistics/264810.

[28] Chen Su and Yuesheng Zhu. 2017. Using personal information to aid in guessing passwords of Chinese webs. *2017 IEEE International Conference on Communications (ICC)* (2017), 1–6.

[29] Carnegie Mellon University. 2015. Password Guessability Service. Website. https://pgs.ece.cmu.edu/.

[30] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, et al. 2017. Design and evaluation of a data-driven password meter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3775–3786.

[31] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation. In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12)*. USENIX, Bellevue, WA, 65–80. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/ur

[32] Blase Ur, Sean M Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L Mazurek, William Melicher, and Richard Shay. 2015. Measuring real-world accuracies and biases in modeling password guessability. In *24th USENIX Security Symposium (USENIX Security 15)*. 463–481.

[33] Rafael Veras, Christopher Collins, and Julie Thorpe. 2014. On Semantic Patterns of Passwords and their Security Impact.. In *NDSS*.

[34] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* 17, 3 (2020), 261–272.

[35] Ding Wang, Debiao He, Haibo Cheng, and Ping Wang. 2016. fuzzyPSM: A new password strength meter using fuzzy probabilistic context-free grammars. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 595–606.

[36] Rinchen Norbu Wangchuk. 2019. Applying For a Visa? Here's Why You Should be Careful About Your Social Media Posts! Retrieved January 12, 2020 from https://www.thebetterindia.com/153762/applying-visa-passport-social-media-posts/.

[37] Percy Wegmann. 2017. PosJs. Git. https://github.com/dariusk/pos-js.

[38] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. 2010. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 162–175.

[39] Matt Weir, Sudhir Aggarwal, Breno De Medeiros, and Bill Glodek. 2009. Password cracking using probabilistic context-free grammars. In *2009 30th IEEE Symposium on Security and Privacy*. IEEE, 391–405.

[40] Daniel Lowe Wheeler. 2016. zxcvbn: Low-Budget Password Strength Estimation. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 157–173. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/wheeler

[41] J. Yan, A. Blackwell, R. Anderson, and A. Grant. 2004. Password memorability and security: empirical results. *IEEE Security Privacy* 2, 5 (Sep. 2004), 25–31. https://doi.org/10.1109/MSP.2004.81

[42] Mile Zivkovic. 2018. The Dos and Don'ts of Social Media Screening in the Hiring Process. Retrieved January 12, 2020 from https://toggl.com/blog/social-media-screening.