Int. J. Human-Computer Studies 72 (2014) 860–874

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Hierarchy of users' web passwords: Perceptions, practices and susceptibilities ☆

S.M. Taiabul Haque [a],*, Matthew Wright [a], Shannon Scielzo [b]

[a] *Department of CSE, University of Texas at Arlington, Arlington, TX 76019, USA*
[b] *Department of Psychology, University of Texas at Arlington, Arlington, TX 76019, USA*

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form
21 July 2014
Accepted 25 July 2014
Communicated by Karen Renaud
Available online 6 August 2014

Keywords:
Security
Usability
Password
Survey

ABSTRACT

In this study, we propose a hierarchy of password importance, and we use an experiment to examine the degree of similarity between passwords for lower-level (e.g. news portal) and higher-level (e.g. banking) websites in this hierarchy. We asked subjects to construct passwords for websites at both levels. Leveraging the lower-level passwords along with a dictionary attack, we successfully cracked almost one-third of the subjects' higher-level passwords. In a survey, subjects reported frequently reusing higher-level passwords, with or without modifications, as well as using a similar process to construct both levels of passwords. We thus conclude that unsafely shared or leaked lower-level passwords can be used by attackers to crack higher-level passwords.

1. Introduction

The World Wide Web (WWW) has changed the way people trade money, maintain social relationships, and pursue pastimes. Creating a unique identity lies at the core of today's Web experience. A user needs to create a unique identity on security-insensitive sites to create and use a profile, post messages and comments, and get content tailored to her interests, such as local news and weather. Maintaining a unique identity of this type typically requires remembering a unique username and a password for that site, much like the authentication procedures on more security-sensitive sites, such as those for banking, stock trading, email, and online social networks.

Advances in technology and design have led to various other forms of user authentication, such as biometrics (Roddy and Stosz, 1997; Marino et al., 2006; Kim, 1995), graphical passwords (Goldberg et al., 2002; Brostoff and Sasse, 2000; Wiedenbeck et al., 2005b), and token-based authentication (Corner and Noble, 2002; Syta et al., 2010). Still, password-based authentication remains the most popular due to its simplicity and cost effectiveness.

Unfortunately, password-based authentication is by no means a panacea as far as usability is concerned. As formulated by Wiedenbeck et al., a good password needs to satisfy two conflicting requirements at the same time: being “easy to remember” and “hard to guess” (Wiedenbeck et al., 2005a). Naturally, passwords that are easy to remember are short single words found in dictionaries or wordlists, or slight variations. Choosing such words as passwords makes them vulnerable to dictionary attacks.

This password management problem becomes more pronounced when a user needs to maintain multiple accounts that require passwords (as many as 25 for an average user Florêncio and Herley, 2007). Several Single Sign-On (SSO) systems, such as Facebook Connect and Sign-in with Google, have been used to alleviate this problem. SSOs, however, represent a single point of failure and a possible privacy risk. Indeed, a recent study has shown that users have several trust, security, and privacy concerns that hinder the wide deployment of SSOs (Sun et al., 2011).

Numerous studies have been conducted to better understand users' strategies for managing passwords between multiple accounts. Notoatmodjo and Thomborson (2009) confirmed that users mentally make a classification of all of their accounts and tend to make stronger passwords for accounts that they consider more important. Such user behavior is justified because users have different levels of incentive to protect their different accounts. If a user's webmail account is hacked, the potential damage is massive because every web account associated to that account could also be compromised. On the other hand, if her online news account is hacked, it does not cost her much.

Apart from password strength, the privacy levels of different passwords of a user also vary to a great extent. While users maintain the highest level of privacy for their financial passwords, they willingly share passwords in some cases (Kaye, 2011). When multiple persons share the same wireless Internet connection in

☆This paper has been recommended for acceptance by Karen Renaud.
* Corresponding author. Tel.: +1 972 704 6815.
E-mail address: eresh03@gmail.com (S.M. Taiabul Haque).

http://dx.doi.org/10.1016/j.ijhcs.2014.07.007
1071-5819/© 2014 Elsevier Ltd. All rights reserved.

an apartment, a common password is shared by all so that each roommate can have access. Another widespread use of password sharing prevails in paid subscription sites. Paid subscription sites like Netflix[1] or Hulu Plus[2] do not forbid password sharing explicitly.

Apart from sharing passwords, users also reuse passwords across different sites. This phenomenon of password reuse can be explained from the viewpoint of cognitive capacity of users. An average user with 25 password-protected accounts should not be expected to be sufficiently equipped on a cognitive level to deal with 25 different passwords (Florêncio and Herley, 2007). In fact, Adams and Sasse (1999) reported that a typical user can be expected to cope with at most four or five passwords effectively. As a result, users reuse passwords across different websites, with little or no modification.

Although the user password reuse habit has been investigated in the current literature, what still remains quite unclear is the pattern of reuse across websites of different categories. Reusing passwords between a news site and a weather site is of little consequence. On the other hand, reusing a password from an important site, like a banking or webmail website, on an untrusted site, puts the more important account at risk.

In this study, we seek to find out to what extent users reuse a password for an important account to create password for an unimportant account or a shared account. Our other important contribution is that, in contrast to most of the related works, we do not observe the degree of reuse only; rather, we observe the *degree of similarity* among passwords of different categories of a user. Our main objective is to test how vulnerable the more important passwords of a user would become if one of her less important passwords is leaked or one of her passwords for shared accounts is abused. Therefore, in addition to reuse without modification, we also analyze in detail to what extent users reuse passwords with some modifications, and to what extent they use a similar process to create passwords for websites of different categories.

To these ends, we first propose a theoretical model of a user password hierarchy based on their perceived importance. We then analyze the degree of similarity among passwords used at different levels of the hierarchy. To verify our model and confirm our research hypotheses, we administered a survey with 80 students at the University of Texas at Arlington (UTA). The complete study was approved by the local Institutional Review Board (IRB).

The responses of the participants suggested that, although users conceptually believe that banking and webmail/social networking passwords are more important to them, these important passwords have a great degree of similarity with their less important passwords. If one of those less important passwords is shared unsafely or leaked, it can be used effectively by an attacker to crack some of the user's banking and webmail/social networking passwords.

Rather than relying on what the users said, we also sought to observe and analyze their password creation behavior. We asked them to construct new passwords of different categories and examined how vulnerable the higher-level passwords would become if the lower-level passwords could be compromised. For each participant, we combined the participant's lower-level passwords with a comprehensive wordlist to form a dictionary, which was subsequently used to perform dictionary attacks on the higher-level passwords of the same participant. Each word in the dictionary was mangled to generate other possible combinations and matched against each of the target higher-level passwords. Almost one-third of the higher-level passwords of the participants could be cracked by this method. We reported some of our preliminary password-cracking results in Haque et al. (2013). These password-cracking statistics and the survey responses confirmed that many users maintain a poor strategy for maintaining multiple password-protected accounts.

The rest of the paper is organized as follows. In Section 2, we discuss related works. In Section 3, we propose a user password hierarchy based on the perceived importance of the passwords and discuss in more detail about passwords of different categories. We present our hypotheses in Section 4 and our experimental methodology in Section 5. We present our findings in Section 6. Finally, we discuss the implications of our findings and future research directions in Section 7.

2. Related work

A number of researchers have examined issues of password reuse, classification of accounts by importance, and password sharing. We now discuss these and relate them to our present study.

The phrase “domino effect of password reuse” was first coined in the work of Ives et al. (2004), who speculated that a domino effect might occur as a result of one site's password file falling prey to a hacker. They conjectured that the hacker might exploit that file to try infiltrate other systems as well, and the habit of password reuse across different sites would certainly make the job of the hacker easier. They predicted that with the proliferation of password-protected accounts, users would reuse passwords more across different sites and the scenario might get worse.

Notoatmodjo and Thomborson (2009) surveyed a group of users and found out that they mentally group their accounts. They identify the factors based upon which these groupings are made and also show that password reuse rate is greater for accounts that are considered less important than accounts that are considered more important.

Their classification of user accounts in terms of perceived importance level is vague, since it consists of only two groups—“less important” and “more important” accounts, and they only focus on password reuse without any kind of modification. We propose a more concrete classification of user accounts and examine other significant issues, such as reuse with some modifications and reuse using a similar thought process. Notoatmodjo and Thomborson also argue that password reuse is a good strategy for less important accounts, reserving mental capacity for more important accounts. While we agree with this notion, we find that users exhibit both partial and complete password reuse between less and more important accounts, creating a serious increased risk for the user.

Some security experts advocate the use of longer passphrases consisting of multiple words (Porter, 1982). The purported advantages of passphrases are twofold: resistance against brute-force attacks and increased memorability. However, the results of a 12-week experiment conducted by Keith et al. (2007) demonstrate that passphrases do not offer significant improvement over passwords with regard to rate of unsuccessful logins due to memory recall failure.

Several researchers have advocated the use of graphics for user authentication. Goldberg et al. (2002) studied the usability of hand-drawn doodles, Brostoff and Sasse (2000) explored the potential of human face based authentication, while Dhamija and Perrig proposed an authentication mechanism based on images (Wiedenbeck et al., 2005b). However, graphical passwords

[1] In their terms of use, Netflix specifies that if the primary account holder shares her account with other people, she takes full responsibility for the actions of those people. It does not, however, forbid sharing an account with others.

[2] In their Frequently Asked Questions section, Hulu Plus specifies that only one simultaneous stream is allowed, but they do not explicitly forbid account sharing.

have yet to replace textual passwords as the common authentication mechanism on the Web.

Preibusch and Bonneau (2010) used a game theoretic model to explain the password schemes used by security-indifferent and security-concerned websites. In another work, they performed a large-scale comparative analysis of password implementation strategies of websites of different categories (Bonneau and Preibusch, 2010). Komanduri et al. (2011) conducted a user study to examine how different password-composition policies for different websites actually affect the users.

The only prior work we found on password similarity was that of Zhang et al. (2010), who studied password expiration and the relationship between users' previous passwords and their new passwords. They examined over 7700 accounts of a Single Sign-On system of a university and their results demonstrated that old passwords are effective predictors of new passwords. We examine similarity of passwords from different accounts of different classes.

Kaye (2011) investigated password sharing practices of users through a self-report measure, in which one-third of the participants reported that they shared their personal email password, while a quarter reported that they shared their Facebook password, mainly with partners and close friends. A part of our current study is devoted to investigating shared passwords, but we mainly explore the hidden consequences of password sharing by investigating the extent to which shared passwords are reused elsewhere. For example, when users share their Netflix passwords with friends, they are consciously doing it, but if they reuse the same Netflix password for their personal email accounts, they inadvertently create a potential breach in the privacy of their personal email accounts. Kaye investigated the former case while we focus on the latter.

Chiasson et al. (2009) conducted a laboratory study to compare the recall success rate of textual passwords with that of graphical passwords. Their experimental methodology involved constructing textual passwords for six different kinds of accounts: bank, email, instant messenger, library, dating, and work. These passwords conceptually belong to the higher level of our hierarchy. They performed a visual inspection of these higher-level passwords and observed that most of the participants constructed passwords following a common pattern across their accounts. Their results also demonstrated that, compared to graphical passwords, recall success rate was lower for textual passwords. By conducting all these tests, they highlighted the weaknesses of textual passwords and advocated the effectiveness of graphical passwords.

Although our experimental methodology has some similarities with Chiasson et al.'s study, their experimental hypotheses were completely different. They focused on comparing between textual and graphical passwords, whereas we focus on comparing between textual passwords of different importance levels. The spectrum of constructed passwords in their experiment did not contain any lower-level passwords. On the other hand, we asked our participants to create passwords of different importance levels and exploited the lower-level passwords of a specific user to crack that user's higher-level passwords.

Adams et al. (1997) conducted a web-based survey with 139 participants to investigate usability issues in password systems. They found that the memorability of a particular password is significantly correlated with its frequency of use. Memorability is also significantly correlated with automaticity or the ability to recall a password spontaneously without conscious thinking. They argue that these findings are consistent with cognitive theory principles like *encoding specificity* and *explicit vs. implicit* memory models. They further conduct semi-structured interview sessions with 30 participants to examine a few important issues more deeply and use grounded theory from social sciences to analyze the responses and build a model of users' password behavior.

In contrast, we first propose a model of user password hierarchy based on our observations and then verify the model by collecting data from users. We also analyze the similarity between the lower and higher level passwords in our hierarchy. The preliminary evidence of such similarity is indicated by another important finding of their study which highlights that almost half of the users have a common theme for all or most of their passwords (Adams et al., 1997).

Many other novel methods have been used for understanding user password habits and attitudes. Hayashi and Hong (2011) used a diary study, Florêncio and Herley (2007) installed an opt-in component of the Windows Live Toolbar in users' machines, Shay et al. (2010) capitalized on the opportunity of a Carnegie Mellon University (CMU) password policy change, while Gaw and Felten (2006) gathered feedbacks from users after they had made actual login attempts in different websites.

Our work differs from all these works in two major ways. First, we propose a concrete categorization of password-protected sites and present our hypotheses and research questions based on this categorization. We tested these hypotheses by both collecting passwords from users and reviewing their responses to a questionnaire. Second, we do not observe the degree of reuse only, rather we observe the degree of similarity among passwords used at different levels of our proposed hierarchy. Therefore, when designing our survey, we considered all possible similarities (both syntactic and semantic) among multiple passwords of a user and prepared our questionnaire accordingly.

3. User password hierarchy

As we are interested in examining the degree of similarity among passwords of different importance levels of an individual user, we first propose a user password hierarchy based on the perceived importance of the site. In this paper, we use the term "importance" to indicate for a site the level of effort a user provides to protect the security and privacy of a password for that site. This includes activities such as constructing a strong password that would be hard to guess, not writing the password on a piece of paper, and not sharing it with others. Since users do more to protect the passwords that they consider more important, this hierarchy is also a password hierarchy based on the level of privacy that the users give their passwords.

First, we classify all the password-protected accounts of a user into five broad categories. Whereas the first three of them are adopted from the categorization of Bonneau and Preibusch (2010), we further add two more categories. We now present a brief description of each category.

Identity accounts: In today's digitalized world, the virtual identity of a user has become more and more important. A user's official webmail account acts as the medium of her professional correspondence, her social networking account personifies herself among friends and family members, and her blogging account represents her voice about different issues. A user creates online identities in these sites that act on her behalf. She builds a long-term reputation of trust in her professional and personal life through them. In short, a user has significant incentive to protect the security of these accounts.

Financial accounts: For online banking and investing, users need to create online accounts for various kinds of financial transactions and bill payment. Users are always concerned about the security of these financial accounts because they represent the user's access to their money and credit. Compromise of these accounts may reveal credit card information and other financial credentials. We consider online banking accounts and accounts in all kinds of merchant sites as financial accounts. Users try to ensure maximum protection for these financial accounts.

Content accounts: Users create accounts in some websites only to customize the contents of those sites. In these accounts, users do not have significant interactions with other users nor any financial transactions. For example, if a user wants Weather.com to show the weather of her location when she visits the site, she might need to create an account there. News and other informational websites also belong to this category. If an account of this kind is hacked, it has at most a moderate impact for the user, such as a loss of privacy or the theft of profile and contact information that could be leveraged for spear phishing attacks.

Sketchy accounts: It is unlikely that all of user's password-protected accounts belong to a category of well-recognized identity, financial, or content sites as described above. In our study, we consider users' accounts in any kind of unrecognized websites as sketchy accounts. This category includes unfamiliar sites that claim to have various kinds of deals or coupons and little-known online forums or content provider sites. Users create accounts in these sites for superficial purposes and often they maintain anonymity by providing a false name or age. As a result, compromise of any account of this kind usually does not create any breach to their privacy. Users have the least incentive to protect the security of these accounts.

Shared accounts: Sharing accounts among multiple users is a common practice. In most cases, apartment mates share the same wireless Internet account. Accounts are also shared to a large extent on paid subscription websites that offer paid access to premium content. Users who want to save money split the subscription fees of these sites and share their member accounts. The password associated with a shared account is known to all of the members who share the account. Passwords are shared for both identity (e.g., colleagues may share the password of a work email account) and financial accounts (e.g., spouses may share the password of a bank account) (Kaye, 2011).

3.1. Hierarchy

If we consider the privacy of passwords for these accounts, it appears that users maintain maximum privacy for their financial and identity passwords. There exists no empirical work in the current literature that shows which of these two types of account is more important to users. A financial account (online banking account, for example) is certainly of great importance, but having access to an identity account (email account, for example) also often means getting access to other accounts that are linked to that identity account (by sending a password reset request mail to the linked email account). We therefore consider financial and identity passwords to be equally important. If these passwords are leaked, serious consequences may result.

On the other hand, passwords for shared accounts are constructed for the purpose of sharing them with others. Generally, passwords are shared among close friends or family members, and people who share passwords also share a level of trust. We predict that users do not create these passwords with the thought that they would share them outside their close circle friends or family members in the future. They create these passwords according to their own criteria, and then sometimes share them due to expedience or circumstance.

For example, a user shares a wireless Internet password with her trusted long-term next-door neighbor. Suddenly, the neighbor moves on and a new neighbor comes to the apartment. The user might ask the new neighbor whether she is willing to share Internet connection (for splitting the subscription fee). If she agrees, she gets the password from the user. Now the new neighbor is just a casual acquaintance whom the user barely knows but the neighbor already knows the shared wireless Internet password and other credentials of the user (e.g., the email ID that the user uses to forward the bill payment receipt to the neighbor). If the neighbor has malicious intent, she may try to hack the email account of the user by using the email ID and the wireless password.

Similarly, passwords for sketchy accounts can also be exploited for compromising other important accounts. A sketchy account is created on an unrecognized website, which may be purposefully designed as part of a social engineering scheme. This scheme exploits the fact that users tend to create accounts on new websites somewhat indiscriminately. An attacker could thus create a website that provides a simple web service and recommend that a new user should create an account on the site to gain full access to that service. In this way, the attacker collects users' passwords and other credentials like usernames and email addresses. Subsequently, the attacker tries to hack accounts on other common financial or identity sites by using this information. If a user reuses the same combination on those important websites, her accounts on those websites would be compromised by the attacker.

In the light of the above discussion, it can be seen that financial and identity passwords are the target passwords that an attacker would like to crack, while shared and sketchy passwords are the passwords that an attacker might want to exploit. Unlike shared or sketchy passwords, content passwords are not readily available to a potential attacker. As mentioned before, content passwords are created in well-recognized trusted websites (New York Times, for example). However, compared to financial and identity websites, content websites do not require a high password security level because these sites do not protect sensitive personal or financial information for their users. As opposed to a financial or identity password, the potential harm that is caused by the leakage of a content password is nominal. Thus, neither the user, nor the site authority, has much incentive to protect the privacy and security of a content password.

We therefore propose a user password hierarchy by placing financial and identity passwords at the higher level, and content and sketchy passwords at the lower level. Since shared passwords can belong to multiple categories, we do not include them in our hierarchy. However, in our user study, we add questions about shared passwords in order to explore the hidden consequences of password sharing. Fig. 1 illustrates this hierarchy.

4. Hypotheses and research questions

One important objective of our study is to test the validity of our proposed hierarchy regarding financial, identity, content and sketchy passwords.

Hypothesis 1. Users mentally classify their passwords into different levels according to the perceived importance of the site, where financial and identity passwords sit at the top level of hierarchy,

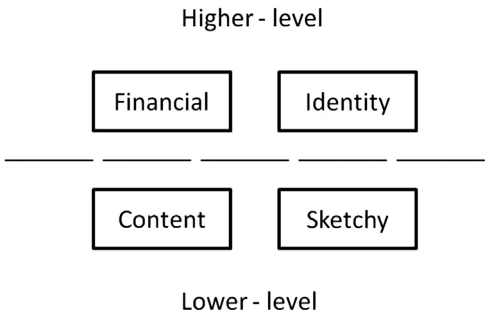


Fig. 1. User password hierarchy based on the perceived importance level.

while content and sketchy passwords sit at the bottom level of hierarchy.

The next and the most important objective of our study is to examine how the knowledge of a password of a lower-level account (content or sketchy account) could increase the chance to crack a higher-level account (identity or financial account) based on similarity to the lower-level password.

Due to the prevalence of content and sketchy sites, users frequently encounter these sites while surfing the Internet. Therefore, we hypothesize that most users maintain a fixed set of passwords for these unimportant sites so that they do not need to create and remember a new password each time they create a new account on these sites. These users are perceived to be more careful and we hypothesize that they usually do not reuse this fixed set of passwords in their financial or identity accounts.

Hypothesis 2a. Most of the users (a) use a fixed set of passwords for sketchy sites and (b) of those that do, they usually do not reuse this fixed set of passwords in their important financial or identity accounts.

Hypothesis 2b. Most of the users (a) use a fixed set of passwords for content sites and (b) of those that do, they usually do not reuse this fixed set of passwords in their important financial or identity accounts.

On the other hand, users who do not maintain a fixed set of passwords for content or sketchy sites need to create new passwords frequently. As discussed before, the cognitive capacity of a typical user restricts the user from constructing a new random password from scratch every time because it is not possible for the user to remember so many random passwords. We therefore predict that when creating the new password, users either reuse one of their existing passwords (with or without modifications), or they use a similar process as they have used before to create one of their existing passwords.

However, what remains quite unclear is to what extent they reuse their important financial or identity passwords (with or without modifications), or to what extent they use a similar process as they have used before to create one of their existing financial or identity passwords. We address this as an open research question that we try to answer through our user study.

Research Question 1. *When creating a new password for a sketchy account or a content account*, (a) *to what extent users reuse one of their financial or identity passwords, without any modification*, (b) *to what extent users reuse one of their financial or identity passwords, with some modifications, and* (c) *to what extent users use a similar process as they have used before to create one of their existing financial or identity passwords*?

Another objective of our study is to explore the degree of similarity between shared passwords and higher-level passwords. We seek to learn the extent to which users reuse their higher-level (financial or identity) passwords for creating a shared password. This leads to the formulation of our second research question.

Research Question 2. *When creating a password for a shared account*, (a) *to what extent users reuse one of their financial or identity passwords, without any modification*, (b) *to what extent users reuse one of their financial or identity passwords, with some modifications, and* (c) *to what extent users use a similar process as they have used before to create one of their existing financial or identity passwords*?

5. Methodology

We conducted a computer-based two-phase laboratory study with 80 UTA students to test our research hypotheses and answer our research questions. In the first phase of the study, we asked the participants to construct new passwords for websites of different categories. This phase was hosted on the secure web server run by the Information Security (iSec) Lab at UTA. Once this phase was completed, each participant was redirected to www.surveymonkey.com[3] for the second phase. In this phase, we had the participants answer some questions regarding their password behaviors for multiple accounts.

Although a larger number of participants could have been drawn from an online survey, we preferred a laboratory study because our pilot study ($n=12$) showed that a laboratory study would produce more consistent responses, especially in the first phase, where the students would be asked to create passwords for eight different websites. Students were assigned partial course credit in exchange for their participation. The complete study was approved by the UT Arlington Institutional Review Board (IRB). With prior approval from the IRB, electronic informed consent was obtained from the participants in lieu of written informed consent. After analyzing the passwords constructed by the participants, we encrypted and stored them in a disk disconnected from any kind of network.

5.1. Study administration

We administered the study through the research pool of the department of psychology, UTA. The department of psychology at UTA maintains the pool for assigning partial course credits to the students who enroll for the course "Introduction of Psychology" and for some other advanced elective courses that offer extra credits. Researchers who collaborate with the department of psychology can post a brief description about their studies to the pool. Students in the research pool can view all the studies and sign up for those that interest them.

The main advantage of conducting a study through the pool is that it can draw a wide range of participants from various departments, because most of these courses are offered for majors from all departments. However, before the beginning of the study, we explicitly informed each participant that the study was being conducted by the Information Security Lab. This was done so that no participant would confuse our study as an experiment for measuring the psychological aspects of people through their constructed passwords.

5.2. First phase

The main objective of the first phase of the experiment was to capture multiple passwords of a user so that we could examine the degree of similarity among them. For this purpose, we designed a PHP script that prompted the users to create passwords for their new accounts for eight different websites in four different categories:

- Financial website: Chase and Wells Fargo
- Identity website: Yahoo! Mail and Facebook
- Content website: NY Times and Weather.com
- Sketchy website: Dreamdeals.com and Justchill.com (hypothetically constructed sites)

[3] SurveyMonkey is a website for administering and analyzing online surveys.

We selected Chase and Wells Fargo as representatives of banking/financial websites because these two banks should be familiar to UTA students due to the prevalence of their ATMs on the campus. Facebook and Yahoo! Mail were selected as identity websites, mainly because of their popularity as a social networking site and a webmail site, respectively. For content websites, we selected the NY Times website and Weather.com, because these two sites readily present a clear distinction between identity sites and content sites, without us needing to explicitly label them as content sites.

During the first phase, we did not want to give the participants any clue about our experimental motive because we expected them to spontaneously construct new passwords, exactly in the same way as they do in real life. Therefore, for all the six real sites, we designed the interfaces so that they would look similar to the original sites. For the two hypothetical unfamiliar sketchy sites, we gave their interfaces an informal appearance so that they would appear to the participants as real-world sketchy websites.

We selected six specific conditions corresponding to six different orders of websites and each participant was randomly assigned to one of the six conditions. For example, one participant created passwords in this order: identity1 – content1 – financial1 – content2 – sketchy1 – identity2 – financial2 – sketchy2. Another participant created in this order: sketchy1 – financial1 – content1 – sketchy2 – identity1 – financial2 – content2 – identity2.

Thus, although the orders of the sites were not completely random, the way we designed the conditions counterbalanced the orders in which the participants created passwords for sites of different categories.

5.2.1. Password construction

In our study, all instructions were presented on the computer screen. We did not read the participants a script or provide them any written instructions. For ethical and security reasons, we explicitly told the participants through warnings in our interface not to provide any of their existing passwords. For each website, we provided a brief introduction and presented a real-life scenario to the participants. The scenario was created in such a way that it resembled a real-world application as much as possible. For example, for Weather.com, the participants were presented with the following scenario:

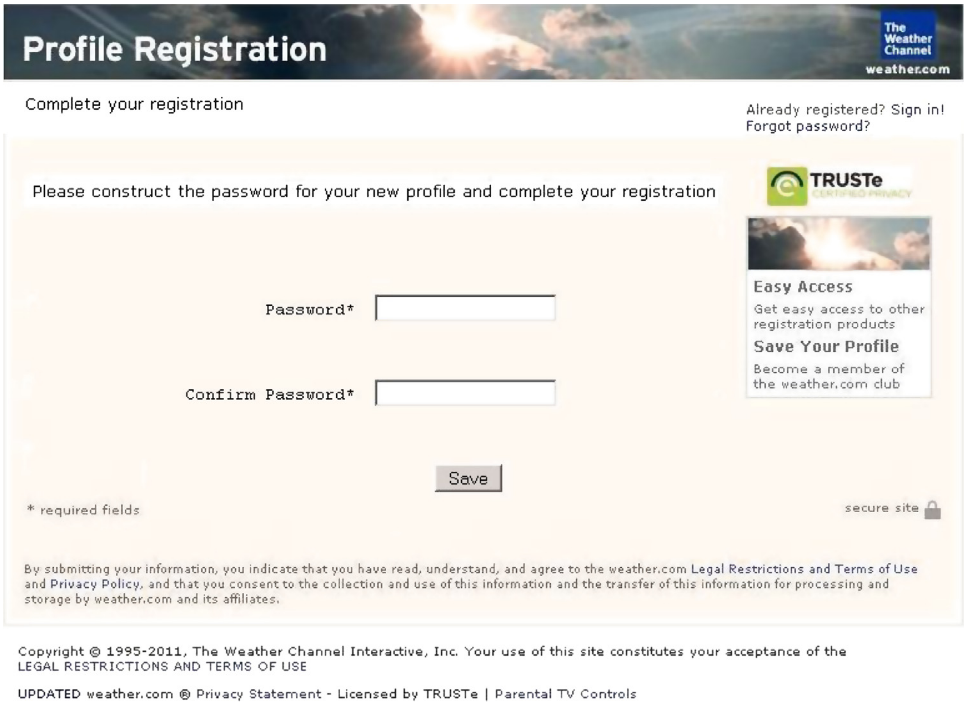
> Weather.com provides the latest weather forecasts, maps, and alerts. You want Weather.com to show weather for Arlington, TX when you go to the site. To do that, you need to register an account on Weather.com so that you can customize your location. Imagine that you are registering a new account on Weather.com. You have reached the final step of registering your new account, and you need to input a password. Proceed to the next page to input your new password.

Once the user clicked the link, our mock password construction page for Weather.com appeared. Fig. 2 shows the interface for the mock password construction page for Weather.com. We note that the URL of the web pages (the URL of our secure web server that hosted the code) appeared in the browser and the participants were aware that it was not the actual Weather.com password construction page. We also anticipated the fact that some participants might consider Weather.com as a sketchy site (due to their unfamiliarity with Weather.com). We believe that our short descriptions of the sites helped to reduce this kind of misconception.

Similarly, for Dreamdeals.com, the participants were presented with the following scenario:

> Imagine that you are doing a Google search to find discount coupons for the Six Flags amusement park in Arlington. While browsing multiple search results' pages, you come across a site called Dreamdeals.com that offers coupon codes for Six Flags and also deals, discounts, and cash backs for other purchases. Dreamdeals.com requires you to register an account in order to gain access to the coupons and discount codes. Imagine that you are registering an account on Dreamdeals.com. You have reached the final step of registration, and you need to input a password. Proceed to the next page to input your new password.

As mentioned before, we created an informal looking interface for Dreamdeals.com, as shown in Fig. 3.



Complete your registration

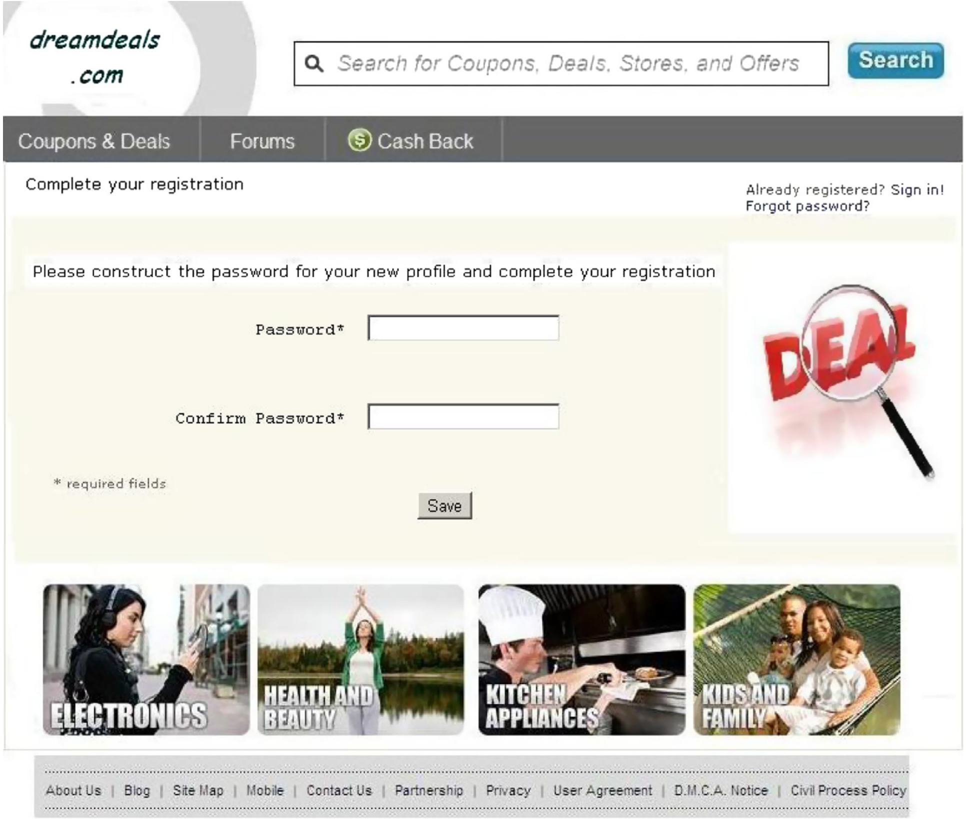
Please construct the password for your new profile and complete your registration

By submitting your information, you indicate that you have read, understand, and agree to the weather.com Legal Restrictions and Terms of Use and Privacy Policy, and that you consent to the collection and use of this information and the transfer of this information for processing and storage by weather.com and its affiliates.

Copyright © 1995-2011, The Weather Channel Interactive, Inc. Your use of this site constitutes your acceptance of the LEGAL RESTRICTIONS AND TERMS OF USE

UPDATED weather.com ® Privacy Statement - Licensed by TRUSTe | Parental TV Controls

Fig. 2. Mock password construction page for Weather.com.


dreamdeals
.com
Search for Coupons, Deals, Stores, and Offers
Search
Coupons & Deals
Forums
Cash Back
Already registered? Sign in!
Forgot password?
Password*
Confirm Password*
* required fields
Save
DEAL
ELECTRONICS
HEALTH AND BEAUTY
KITCHEN APPLIANCES
KIDS AND FAMILY
About Us | Blog | Site Map | Mobile | Contact Us | Partnership | Privacy | User Agreement | D.M.C.A. Notice | Civil Process Policy

Complete your registration

Please construct the password for your new profile and complete your registration

Fig. 3. Mock password construction page for Dreamdeals.com.

5.2.2. *Password policy*

For all six real websites, we enforced exactly the same password policies as they are enforced in those sites. For example, Wellsfargo requires any password to be 6–14 characters long, with at least one letter and one digit. We designed our script in such a way so that the participants had to conform to this policy. For the two hypothetical sites, we ensured that the participants' passwords were at least five characters long. Like the original sites, participants were also required to reconfirm their passwords in a second box, which prevented them from typing some random characters as their passwords.

In this way, we implicitly tried to trigger the real life password creation mechanisms of users for websites of different categories. In designing the interfaces and providing the introduction for each site, we were careful about not revealing to the participants that our main objective is to categorize their constructed passwords based on our categories. We believe that this helped to avoid any kind of experimental bias that is associated with "demand characteristics" (Orne, 1969).

5.3. *Second phase*

In the second phase of our experiment, we asked the participants to answer a survey. In this phase, we were relatively overt about our categorization of passwords. We asked users to contemplatively respond to some questions about their password sharing habits and password reuse habits, with and without modifications, across websites of different categories.

We were aware of the fact that it would not be a straightforward task for the participants to exactly recall the construction processes of passwords they had set up some time ago. Moreover, users are not actively aware of the processes carried out during password construction since password construction is not a user's primary task, rather it is one of a series of subtasks required for completing the primary task (opening a bank account, for example). For this reason, instead of asking them to provide open-ended responses for some questions, we gave them a series of options and asked them to select those options that are related to their password construction processes. The options were carefully selected from prior studies. We believe that these options provided cues for the participants to recall the processes they usually carry out during their password construction activity.

We also asked most of the questions as Likert-type questions. This, in turn, allowed the participants more flexibility in choosing their answers (for example, to choose the option "seldom" instead of "never" whenever they could not exactly recall their construction processes). We give a brief overview of our survey questions here.

5.3.1. *Rating of sites*

We asked the participants to rate the importance of their passwords on a 5-point Likert-type scale for all eight websites that were presented in the first phase. We were aware of the argument on whether responses from Likert scales should be considered as ordered-categorical data or interval-level data (Clason and Dormody, 1994). To treat the scale as an interval-level scale, anchors were only included on the bipolar ends of the scale (1 = "not important", 5 = "important"), and the middle point (3 = "moderately important").

5.3.2. *Shared passwords*

Next, we gave the participants two real-word examples of password sharing: one about sharing Netflix password with friends and the other about sharing a wireless Internet password with apartment mates. We asked them whether or not they share their passwords for cases like these. Those who responded that they do share were then asked how they create passwords for these shared accounts. We first asked them to what extent they reuse a password they have used elsewhere, without any modification. We used a 4-point Likert scale (1 = "never", 2 = "seldom", 3 = "sometimes", 4 = "often") for this and the subsequent questions. Those who responded that they reuse a password without any modification were further asked which kind of password they reuse.

We then asked them the extent to which, when creating a shared password, they reuse a password they have used elsewhere, with some modifications. More specifically, we asked them "When creating a password for your shared account, how frequently do you reuse a password that you have used elsewhere, with some modifications (for example, abc→abc1, abc→abd etc.)?". The examples were intended to limit the different ways that different users might interpret "modifications" to mainly focus on minor changes, such as adding or changing just a single character. Those who responded that they reuse a password with some modifications were further asked which kind of password they reuse with modifications, and what kind of modifications they make. Finally, we asked them to what extent they use a similar process as they have used before when creating shared passwords. Those who responded that they use a similar process were further asked which kind of password would be most similar to the new shared password, and how they would be similar.

5.3.3. *Sketchy passwords*

In this part, we first revealed to the participants that the two sites Dreamdeals.com and Justchill.com that were presented before belonged to our category of unfamiliar sketchy sites. Then we asked the participants whether, for this kind of site, they use a fixed set of passwords or create a new password each time they open a new account for each site.

The participants who responded that they use a fixed set of passwords were further asked whether they reuse this fixed set of passwords elsewhere, especially in their identity/financial websites. On the other hand, those who responded that they create a new password each time were further asked the same questions as they were asked about shared passwords. They were asked the extent to which, when creating a new sketchy password, they reuse a password they have used elsewhere, use a similar process as they have used before, and so on.

5.3.4. *Content passwords*

As with sketchy passwords, we first revealed to the participants that the two sites NY Times and Weather.com that were presented before belonged to our category of familiar content sites. Then they were asked the same questions about their content passwords as they were asked for sketchy passwords.

5.4. *Survey analysis*

To test our hypotheses, we analyzed the responses of the participants from both phases. We first thoroughly reviewed their responses to the questionnaire that was provided in the second phase of the study. Then we collected the passwords that the participants constructed in the first phase and grouped passwords of similar kinds together: financial passwords in one group, identity passwords in one group, and so on. We analyzed each group separately to find the frequency of using capital letters, digits and special characters.

Finally, we tried cracking the higher-level (financial and identity) passwords with the help of lower-level (content and sketchy) passwords. We used the John The Ripper (JTR) password cracker for this purpose, using JTR's *wordlist* mode combined with the *single crack* mode.

Wordlist mode cracking is basically a dictionary attack where every word in a wordlist is tried against the candidate password until a match is found. If word mangling rules are enabled, each word in the wordlist is modified or mangled to generate other possible combinations. The single crack mode is the default cracking mode of JTR in which a large number of word mangling rules are applied to a very small dictionary to perform a dictionary attack. As the default set of word mangling rules is very small in the wordlist mode, we modified the JTR configuration file so that it would be possible to apply the large set of word mangling rules of the single crack mode while performing cracking in the wordlist mode.

For each participant, we combined the participant's lower-level passwords with a comprehensive wordlist and tried to crack the higher-level passwords of the same participant by using JTR in our modified wordlist mode.

6. Results

In this section we present all the major findings based on our analysis of 80 surveys.

6.1. *Demographics*

Overall, 47 female and 33 male students participated in our survey. Students from diverse majors, including Psychology (13), Nursing (12), Kinesiology (6), Biology (5), Engineering (5), Business (4), Education (4), Social Work (4), and others (27) participated in our survey.

6.2. *Perceived importance of passwords*

Our first hypothesis was that users mentally classify their passwords into different levels according to their perceived importance. By analyzing the ratings provided by the participants (Section 5.3.1), we found evidence to support this hypothesis. The identity and financial passwords were perceived to be more important (had higher mean and median ratings) than the content and sketchy passwords. Fig. 4 summarizes the ratings of the participants for all eight sites.

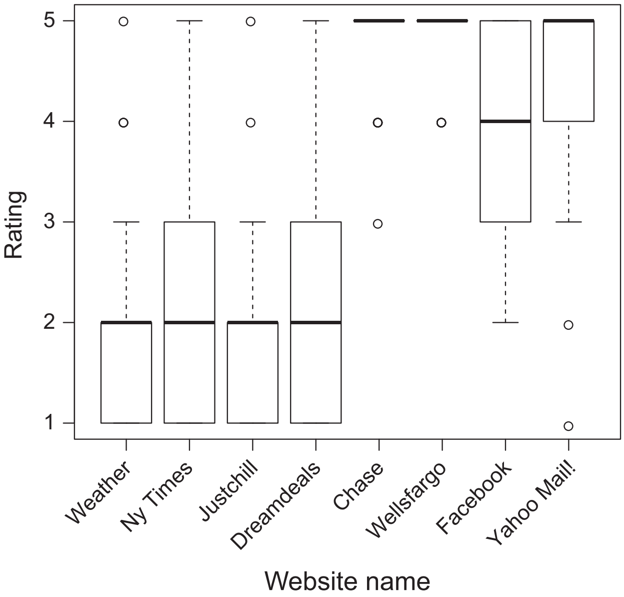


Fig. 4. Box plot of participant ratings on a 1–5 scale (1 being "not important", 3 being "moderately important", 5 being "important") about the perceived importance of passwords of different websites.

Table 1
Summary of post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction applied. For brevity, all entries with $p<.00179$ have been omitted.

Pair	Z	Asymp. Sig. (2-tailed)
Wellsfargo-Chase	−1.342	.180
Dreamdeals.com-Justchill.com	−2.238	.025
Dreamdeals.com-NY Times	−.256	.798
Dreamdeals.com-Weather.com	−2.753	.006
Yahoo! Mail-Facebook	−2.339	.019
Justchill.com-NY Times	−2.207	.027
Justchill.com-Weather.com	−1.207	.228

Given the skewness evident in Fig. 4, we decided to conduct a nonparametric repeated measure statistical test to properly examine Hypothesis 1. Accordingly, we conducted Friedman's test. The perceived importance of passwords differed significantly across the eight sites, $\mathcal{X}^2(7)=448.017$, $p<.001$.

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p<.00179$. All the banking and identity sites had significantly higher ratings than all the content and sketchy sites, $p<.00179$ for all cases. The banking sites also had significantly higher ratings than the identity sites, $p<.00179$ for all cases. The content site Weather.com had a marginally significant lower rating from the content site NY Times, $p<.00179$. The differences between other pairs of content and sketchy sites were statistically insignificant. The difference between the webmail site (Yahoo! Mail) and the social networking site (Facebook) was also statistically insignificant. Table 1 summarizes the results for the post hoc analysis.

The above results suggest that, although financial and identity passwords sit at the top level of the password hierarchy according to their perceived importance, financial passwords are perceived to be significantly more important than identity passwords. On the other hand, content and sketchy passwords sit at the bottom level of the hierarchy. However, there was not enough evidence to make any clear distinction between content sites and sketchy sites.

The hierarchy of users' Web passwords, therefore, turned out to be a three-level one (as opposed to a two-level hierarchy as we hypothesized), where financial passwords sit at the top level and identity passwords sit at the next level, while content and sketchy passwords sit at the bottom level of the hierarchy.

6.2.1. *Password characteristics*

We now analyze the passwords constructed by the participants in the first phase to help validate the findings from the survey in the second phase. Our findings are also consistent with Hypothesis 1.

We calculated the length of the passwords and the frequency of using capital letters, digits, and special characters for passwords of different categories. The length and the frequency values decreased as the perceived importance of the sites decreased. Figs. 5 and 6 summarize our analysis.

We begin with an analysis of password length, shown in Fig. 5. Passwords are longer for financial sites and then shorter for identity, content, and sketchy sites, in order. We note that the minimum password length requirement was not the same for all the sites (seven for Chase, five for NY Times, Justchill.com, and Dreamdeals.com, six for all others), and this may have affected the password lengths.

The frequency of using capital letters also decreased from higher-level passwords to lower-level passwords (Fig. 6). Unlike length, no confounding effect existed in this case because the participants were not required to use capital letters in any of the sites. They spontaneously used more capital letters while constructing their higher-level (financial and identity) passwords.

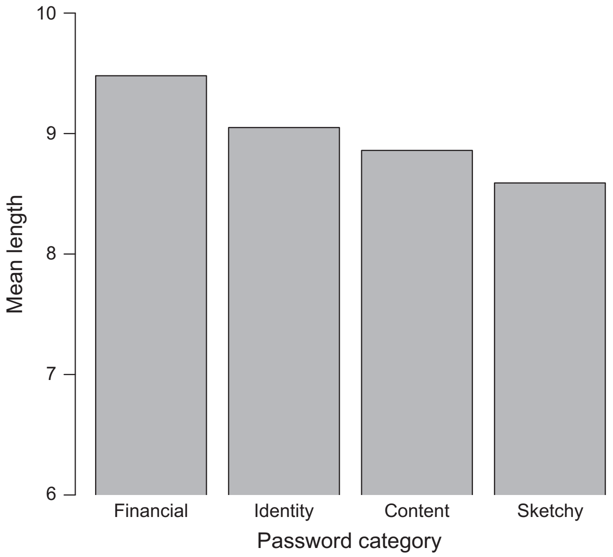


Fig. 5. A comparison of mean lengths.

Among the eight sites, only the two financial sites required their passwords to contain at least one digit. Yet the frequency of using digits decreased from identity sites to sketchy sites (Fig. 6). The same is true regarding the use of special characters. None of the identity, content, or sketchy sites had any kind of requirement or restriction for using special characters. Still the frequency of using special characters decreased from identity sites to sketchy sites (Fig. 6). Since Chase does not allow special characters for their website and we followed the password policy of the actual websites, we did not allow any special characters to be included in the passwords constructed for Chase. Therefore, the frequency of using special characters was less in the financial sites than the identity sites (Fig. 6).

We also calculated the percentage of passwords that contained only lowercase letters, with no capital letters, digits, or special characters. The percentage increased from identity sites to sketchy sites (Fig. 6), although the password policy was same for all the identity, content, and sketchy sites regarding the use of capital letters, digits and special characters.

6.3. *Sketchy password*

We report our results about sketchy passwords by relating them to relevant hypothesis/research question.

Hypothesis 2a: As suggested in the first part of Hypothesis 2a, it was found that most of the users use a fixed set of passwords for sketchy sites. Specifically, 52 of our 80 participants (65%) reported using a fixed set of passwords for sketchy sites whereas 28 reported that they create a new password each time they open a new account at an unfamiliar sketchy site. The chi-square analysis results demonstrated that the difference was significant, $\mathcal{X}^2(1)=7.20$, $p<.01$.

For the second part of the Hypothesis 2a, it was proposed that of the participants that do use this fixed set of passwords that they usually do not reuse this fixed set of passwords in their important financial or identity accounts. To assess this, we compared participants who reported “never” to “seldom” reusing against those that “sometimes” to “often” reused (Table 2 shows the descriptive statistics of participants' responses for this question). This part of the hypothesis was also supported, 37 of these 52 participants reported “never” to “seldom”, whereas only 15 participants

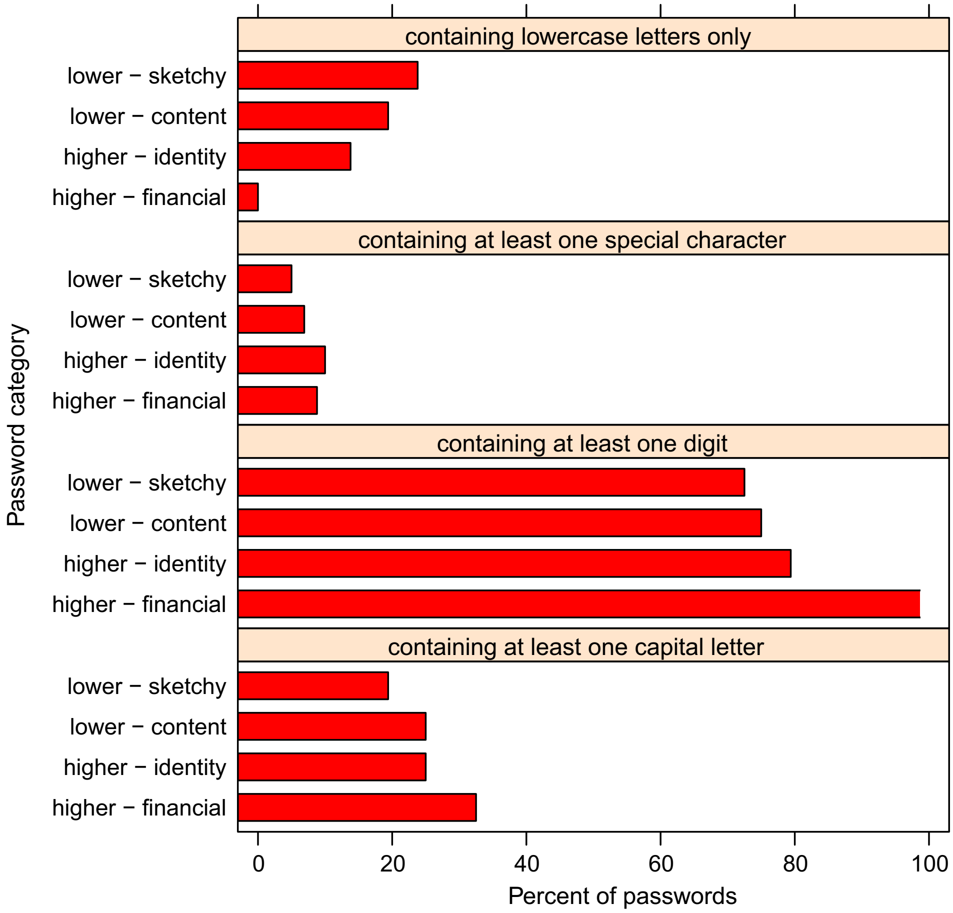


Fig. 6. A comparison among passwords of different categories.

Table 2
Descriptive statistics of participants' responses (1="never", 2="seldom", 3="sometimes", 4="often") for the question "You have a fixed set of passwords for unfamiliar sketchy sites like Dreamdeals.com or Justchill.com. How frequently do you reuse this specific set of passwords elsewhere, especially in familiar banking/webmail/social networking sites?".

Mean	Median	SD
1.98	2	1.04

Table 3
Type of modifications made when reusing an existing password to create a sketchy password.

Modification	Never	Seldom	Sometimes	Often	Total
Add/delete 1–2 characters	4	5	8	6	23
Add/delete more than 2 characters	6	7	7	3	23
Replace 1–2 characters with others	4	5	7	7	23
Add special symbols	5	2	10	6	23

reported "sometimes" to "often" reusing these fixed passwords for their financial or identity accounts, $\mathcal{X}^2(1) = 9.31$, $p < .01$.

Thus, Hypothesis 2a turned out to be true.

Research Question 1: Overall, 28 of our 80 participants reported that they create a new password for a sketchy site. As mentioned in Section 4, it was not quite clear how carefully these people construct the new sketchy password. Therefore, we asked these 28 participants in detail about their strategy of creating the new sketchy password. Their responses provided answers to our Research Question 1.

Reuse without modification: Almost half of these participants (13 out of 28) said that when creating a new sketchy password, they "never" reuse a password they have used elsewhere without any modification. We compared participants who reported "never" to "seldom" reusing their identity or financial passwords against those that "sometimes" to "often" reused them. Overall, 24 of these 28 participants reported "never" to "seldom", whereas only 4 participants reported "sometimes" to "often", $\mathcal{X}^2(1) = 14.29$, $p < .001$.

Reuse with modification: The reuse with modification rate is comparatively high among the participants. Only 5 out of 28 participants (18%) said that when constructing a new sketchy password, they "never" reuse a password they have used elsewhere, with some modifications. Specifically, 13 of these 28 participants reported "never" to "seldom" reusing their identity passwords (with some modifications), whereas 15 participants reported "sometimes" to "often" reusing them, $\mathcal{X}^2(1) = .14$, $p = .706$. For financial passwords, 24 reported "never" to "seldom" reusing them (with some modifications), whereas 4 reported "sometimes" to "often" reusing them, $\mathcal{X}^2(1) = 14.29$, $p < .001$.

These participants also answered what kind of modifications they make. Their responses are shown in Table 3.

Reuse similar process. These 28 participants also indicated that they use a similar process as they have used before when creating a new sketchy password. Only a single participant (4%) reported that he/she "never" uses a similar process. In particular, 18 of these 28 participants reported "never" to "seldom" using a similar thought process of constructing an identity password when constructing a sketchy password, while 10 participants reported

Table 4
Type of similarity between an existing password and the new sketchy password.

Type of similarity	Never	Seldom	Sometimes	Often
Both are inspired by common source	4	7	10	6
Both are dictionary words/minor variations	11	10	5	1
Both are English phrases	10	7	5	5
Both are related to friend/family	11	5	8	3
Both are personally meaningful words	6	5	10	6
Both are personally meaningful numbers	6	5	9	7

Table 5
Descriptive statistics of participants' responses (1=“never”, 2=“seldom”, 3=“sometimes”, 4=“often”) for the question “You have a fixed set of passwords for familiar content sites like Weather.com or Nytimes.com. How frequently do you reuse this specific set of passwords elsewhere, especially in familiar banking/webmail/social networking sites?”.

Mean	Median	SD
2.41	3	1.08

Table 6
Descriptive statistics of participants' responses (1=“never”, 2=“seldom”, 3=“sometimes”, 4=“often”) for the question related to password sharing.

Mean	Median	SD
2.95	3	.93

“sometimes” to “often” doing so, $\mathcal{X}^2(1) = 2.29$, $p = .131$. On the other hand, 27 of these 28 participants reported “never” to “seldom” using a similar thought process of constructing a financial password when constructing a sketchy password, while only a single participant reported “sometimes” to “often” doing so, $\mathcal{X}^2(1) = 24.14$, $p < .001$.

Finally, we asked these participants how the new sketchy password would be similar to their existing passwords. Their responses are shown in Table 4.

We consider two types of similarity: semantic similarity and syntactic similarity. If two passwords are similar based on their meaning or semantic content, we denote this as semantic similarity. For example, two passwords are semantically similar if they are inspired from a common source (literature, film, music etc.) or if both are personally meaningful words or numbers. On the other hand, two passwords are syntactically similar if one is a minor variation of another. Examples of minor variation are capitalizing a letter, adding digits/special characters, and replacing one letter with another.

Option 2 referred to syntactic similarity (both passwords are dictionary words, or minor variations of those words), whereas options 1, 5 and 6 were related to semantic similarity. The responses of the participants clearly showed that they follow options 1, 5 and 6 more frequently than option 2 (Table 4). This demonstrated that the semantic similarity is much more evident than the syntactic similarity. This issue is discussed further in Section 7.

6.4. Content password

We report our results about content passwords by relating them to Hypothesis 2b.

Hypothesis 2b: As suggested in the first part of Hypothesis 2b, it was found that most of the users use a fixed set of passwords for content sites. Specifically, 66 of our 80 participants (83%) reported using a fixed set of passwords for content sites whereas 14 reported that they create a new password each time they open a new account at a familiar content site. The chi-square analysis results demonstrated that the difference was significant, $\mathcal{X}^2(1) = 33.80$, $p < .001$.

For the second part of the Hypothesis 2b, it was proposed that of the participants that do use this fixed set of passwords that they usually do not reuse this fixed set of passwords in their important financial or identity accounts. To assess this, we compared participants who reported “never” to “seldom” reusing against those that “sometimes” to “often” reused (Table 5 shows the descriptive statistics of participants' responses for this question). This part of the hypothesis was not supported, 31 of these 66 participants reported “never” to “seldom”, whereas 35 participants reported “sometimes” to “often” reusing these fixed passwords for their financial or identity accounts, $\mathcal{X}^2(1) = .24$, $p = .623$.

Thus, Hypothesis 2b turned out to be partially true.

Only a minority of participants (18%) said that they construct a new password each time they create a new account in any content site. We do not report in detail on their strategies for creating the new content passwords, but their responses were consistent with the responses for sketchy passwords. For example, the responses also suggested that the semantic similarity is much more evident than the syntactic similarity.

6.5. Password sharing

The responses of the participants showed that almost all of the participants share passwords with others for maintaining a shared account. Only 6 out of 80 participants (8%) reported that they “never” share a password with others for maintaining a shared account. Table 6 shows the descriptive statistics of participants' responses for this question.

We compared participants who reported “never” to “seldom” sharing against those that “sometimes” to “often” shared. Only 24 of our 80 participants reported “never” to “seldom”, whereas 56 participants reported “sometimes” to “often”, $\mathcal{X}^2(1) = 12.80$, $p < .001$.

Research Question 2: We asked the 74 participants (participants who reported “seldom”, “sometimes”, or “often” sharing) in detail about their strategy of creating a shared password. Their responses provided answers to Research Question 2.

Reuse without modification: We first asked them the extent to which they create a shared password by reusing an identity password they have used elsewhere, without any modification. We compared participants who reported “never” to “seldom” reusing against those that “sometimes” to “often” reused them. Overall, 43 of these 74 participants reported “never” to “seldom”, while 31 participants reported “sometimes” to “often”, $\mathcal{X}^2(1) = 1.95$, $p = .163$.

We also asked them the extent to which they create a shared password by reusing a financial password they have used elsewhere, without any modification. Overall, 58 of these 74 participants reported “never” to “seldom”, while only 16 participants reported “sometimes” to “often”, $\mathcal{X}^2(1) = 23.84$, $p < .001$.

Reuse with modification: We next asked these participants the extent to which they create a shared password by reusing an identity password they have used elsewhere, with some modifications. Among 74 participants, 40 reported “never” to “seldom”, whereas 34 reported “sometimes” to “often”, $\mathcal{X}^2(1) = .49$, $p = .486$. For financial passwords, 59 out of these 74 participants reported “never” to “seldom”, whereas only 15 reported “sometimes” to “often”, $\mathcal{X}^2(1) = 26.16$, $p < .001$.

Table 7
Type of modifications made when reusing an existing password to create a shared password.

Modification	Never	Seldom	Sometimes	Often	Total
Add/delete 1–2 characters	8	13	22	19	62
Add/delete more than 2 characters	15	18	19	10	62
Replace 1–2 characters with others	6	10	28	18	62
Add special symbols	20	8	20	14	62

Table 8
Type of similarity between an existing password and the new shared password.

Type of similarity	Never	Seldom	Sometimes	Often
Both are inspired by common source	11	9	23	24
Both are dictionary words/minor variations	30	25	8	4
Both are English phrases	22	16	14	14
Both are related to friend/family	22	10	17	17
Both are personally meaningful words	9	5	27	26
Both are personally meaningful numbers	10	8	23	26

These participants also answered what kind of modifications they make. Their responses are shown in Table 7.

Reuse similar process. The responses of these 74 participants indicated that they use a thought process similar to the one used for creating an identity password when they create a shared password. Specifically, only 29 of these 74 participants reported "never" to "seldom", whereas 45 participants reported "sometimes" to "often", $\mathcal{X}^2(1)=3.46$, $p=.063$. For financial passwords, however, 57 out of these 74 participants reported "never" to "seldom", whereas only 17 reported "sometimes" to "often", $\mathcal{X}^2(1)=\ 21.62$, $p<.001$.

Finally, we asked these participants how the new shared password would be similar to their existing passwords. Their responses are summarized in Table 8. We can see that the semantic similarity is much more evident than the syntactic similarity in this case as well (participants follow options 1, 5 and 6 more frequently than option 2).

6.6. Password cracking

We did not only rely on what the participants said during the survey in Phase 2, we also analyzed on the basis of what they did in Phase 1. We exploited the passwords constructed by the participants in Phase 1 and tried to crack the financial and identity (higher-level) passwords of a participant by using that participant's content and sketchy (lower-level) passwords. For cracking purposes, we used the John The Ripper (JTR) password cracker.

Attack model: We first assume a scenario where an attacker compromises one lower-level password of each participant. We calculate the percentage of higher-level passwords that could be cracked by the attacker under this assumption. We also observe the effect when each additional lower-level password is compromised by the attacker. Specifically, we try to answer the following questions:

Research Question 1: What percentage of higher-level passwords could be cracked by an attacker by compromising one lower-level password of each participant?

Research Question 2: With compromise of each additional lower-level password, what additional percentage of higher-level passwords could be cracked by the attacker?

We compute the percentages of passwords cracked as follows. For each lower-level password, we employ it in cracking all of the user's upper-level passwords and report the overall percentage. For two to four lower-level passwords, we take the different combinations of multiple lower-level passwords when used individually to crack upper-level passwords and report the combined percentages.

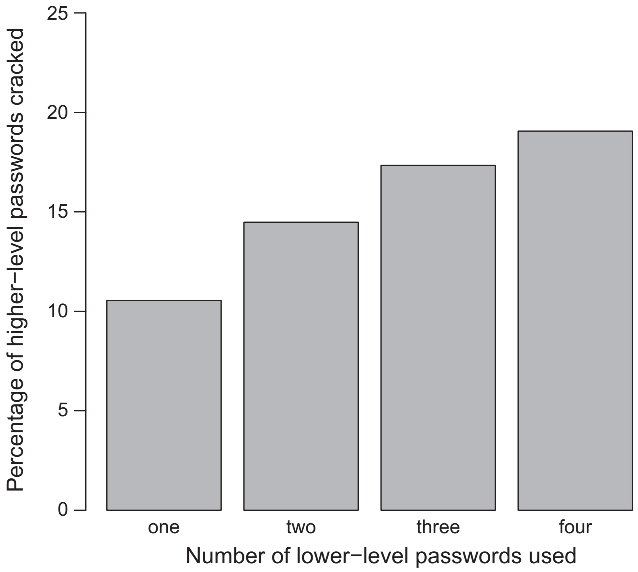


Fig. 7. Password cracking statistics (without wordlist).

Attack without wordlist: In this attack mode, we tried to crack the higher-level passwords of a participant by using the lower-level passwords only, without using any wordlist or dictionary. We performed cracking by using JTR in our modified wordlist mode (Section 5.4). We used the word mangling rules of JTR to mangle the lower-level passwords in order to guess the higher-level ones. These mangling rules include appending digits and replacing letters with similar symbols ($ instead of S, for example). For each participant, the wordlist consisted only the lower-level passwords of the same participant.

Fig. 7 demonstrates the password cracking statistics. By using one lower-level password of each participant, we could successfully crack 10.6% of higher-level passwords. The percentage increased to 19.1% (61 out of 320) when we used all the four lower-level passwords to crack the higher-level ones.

Less than half of the cracked higher-level passwords (29 out of 61) were the same as the lower-level ones. The rest were minor modifications of the lower-level passwords, such as appending digits, appending years, and capitalizing first/middle letter. Thus, even when a password is not directly reused, reuse with modification greatly increases the risks to users. Fig. 8 demonstrates the password reuse mechanism of the participants.

Attack with wordlist: In this attack mode, we impersonated a more sophisticated attacker who would use a wordlist along with the lower-level passwords for performing the cracking operations. For each participant, we combined the participant's lower-level passwords with the Cain & Abel wordlist and tried to crack the higher-level passwords of the same participant by using JTR in our modified wordlist mode.

Fig. 9 demonstrates the password cracking statistics. By using the Cain & Abel wordlist and one lower-level password of each participant, we could successfully crack 26.8% of higher-level passwords. The percentage increased to 33.1% (106 out of 320) when we used all the four lower-level passwords along with the Cain & Abel wordlist to crack the higher-level passwords.

We also tried to crack the higher-level passwords of the participants by using the Cain & Abel wordlist (in our modified wordlist mode) only, without using the lower-level passwords. Among the 320 higher-level passwords, we could successfully

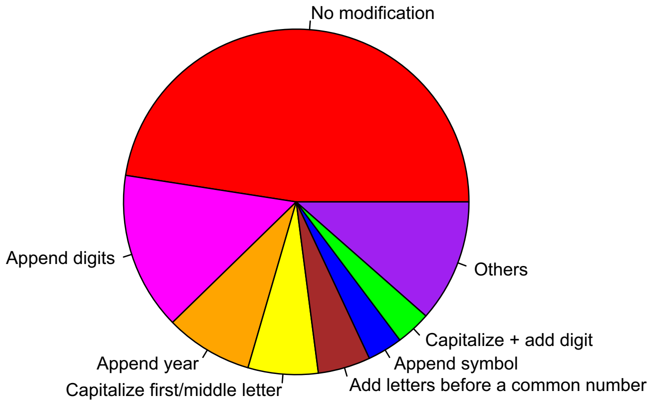


Fig. 8. Password reuse mechanism.

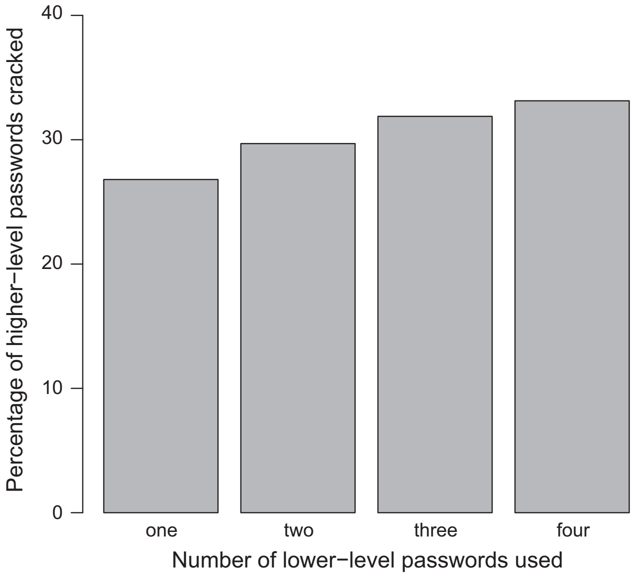


Fig. 9. Password cracking statistics (with wordlist).

crack 21.9% (70 out of 320) of passwords with this method. Thus, by combining the lower-level passwords with the Cain & Abel wordlist, we could successfully crack 51.4% more higher-level passwords (the number of cracked higher-level passwords increased from 70 to 106). A chi-square test showed that the difference is significant, $\mathcal{X}^2(1) = 9.6$, $p < .05$.

The average cracking time was 73.4 s (Intel Pentium 1.86 GHz dual-core processor, 1 GB RAM, Microsoft Windows Vista) when we used the wordlist. When we used the lower-level passwords, the cracking time was less than one second. For each participant, the JTR tests ran until all the entries of the dictionary were mangled and compared or all the passwords were cracked.

7. Discussion

Before discussing the various implications of our findings, it is important to highlight several limitations of our study.

7.1. Limitations

First, we do not dispute the fact that it is difficult to demonstrate ecological validity (Brewer, 2000) in any password study where participants are aware that they are creating passwords for experimental purpose, rather than for accounts they value in real life for regular use over a long period of time. However, in the context of this work, passwords for websites of different categories were created under conditions that should be affected equally by this issue. Furthermore, prior work suggested that involving a role-play scenario would motivate users to construct passwords more seriously than a survey scenario (Komanduri et al., 2011). We thus presented a role-play scenario for each website, and the sites were designed so that they would resemble real-world sites as much as possible (Section 5.2.1). We note, however, that our participants were not required to return on a second day to re-enter their passwords, and as such, some of them might have constructed less memorable random passwords.

It is also difficult to emulate the temporal, situational, environmental and other real-life contextual aspects of password construction in a laboratory study. For example, when constructing a password at home, a user with a pet on her lap might construct a password that has semantic relation to that pet. However, the results of a recent work of Fahl et al. (2013) on the ecological validity of password study reveal that passwords collected during user studies closely resemble users' actual passwords.

We also agree that the sample size of our study is not very large ($n = 80$) and the study is only of university students, who may vary significantly from other populations in their password behavior, and in particular their password sharing behavior. However, compared to the sample sizes of prior works (Wiedenbeck et al., 2005a; Notoatmodjo and Thomborson, 2009; Gaw and Felten, 2006), which were also laboratory experiments among students, our sample size can be considered reasonable. Finally, we note that the presence of an observer may, if anything, encourage users to create stronger passwords than they might otherwise. This notion was supported by the results of our pilot study.

In our survey, we have directly asked the users to rate the importance of their passwords for sites of different categories, which might have primed them to identify categories and gauge their relative importance. However, we have also analyzed the differences in lengths and the frequencies of capital letters, digits, and special characters for the passwords constructed during the first phase (passwords that corresponded to different categories in our hierarchy) in Section 6.2.1. These results supplement the findings of the survey.

Although there was a confounding effect for length (the minimum length requirement was different for different sites), there was no such effect for the frequency of capital letters. The policy for using capital letters was same for all four categories. The participants spontaneously used more capital letters in their higher-level passwords (Fig. 6). Moreover, the password policy was the same for all the identity, content, and sketchy sites regarding the use of digits and special characters. Yet, the participants used more digits and special characters in identity (higher-level) sites than content and sketchy (lower-level) sites (Fig. 6).

In this work, the categories of websites have been proposed by us based on prior work and our own observations; it did not originate from the users. During the first phase of the study, the participants constructed passwords for only eight websites that already conformed to our categorization. During the second phase, the categories of sites were described to participants and they subsequently answered questions about them. It would be unfair to claim based on our current study results that users do indeed classify accounts (and passwords) in this way independently in their real-life. Some non-primed feedback techniques such as card sorting might further help us to better understand the actual thought processes of the users. We plan to work on this in future.

Our password cracking results should be interpreted with caution, particularly by considering the fact that the participants were asked to construct eight passwords in a row in a short laboratory study with an artificial setting. It was a cognitively demanding task which might have prompted some of the participants to construct similar (if not exactly the same) passwords.

We acknowledge the fact that different password policies of the selected websites precluded us from making a fair comparison among the categories. The trade-off here was between realism and learning about user behavior. We chose the more realistic approach (enforced exactly the same password policies as they are enforced in the real sites). However, Section 6.2.1 contains the results of some fair comparisons and they are consistent with the basic claim of Hypothesis 1.

7.2. Implications

One important finding of our study is that users mentally classify their passwords into different levels according to the perceived importance level of the site. Our results suggested that users consider their financial passwords to be the most important. In our analysis, financial passwords were perceived to be significantly more important than identity passwords. Identity passwords, in turn, were perceived to be significantly more important than content and sketchy passwords. The perceived importance of passwords did not differ significantly between webmail accounts and social networking accounts, the two types of identity account that we studied.

Although the participants considered content passwords to be less important, their responses indicated that these passwords have a strong degree of similarity with their important financial and identity passwords. More than four-fifths of the participants reported that they use a fixed set of passwords for all kinds of content sites, which is a reasonable practice. More than half of these participants, however, further indicated that they reuse this fixed set of passwords in their important financial or identity accounts. Of those participants, about one-third reported that they "often" reuse them in identity sites, while one-fifth reported that they "often" reuse them in financial sites.

These findings suggest that a password used at a content site may be more valuable than the account that it protects. An account at a content site typically does not contain much sensitive information; users create it mainly for the purpose of customizing the site experience. The passwords used to protect these accounts, however, are valuable because they are reused frequently in identity or financial websites. If a content password is leaked, it can be used effectively to compromise important identity or financial accounts. This issue should be considered while formulating the authentication policies for content sites. Given how entrenched passwords are as an authentication mechanism, however, it may be more useful to help and encourage users at financial and identity sites to make stronger but memorable passwords that are clearly distinct from their content site passwords.

The passwords we collected from the first phase of the study showed that users construct stronger passwords (passwords of longer length with more capital letters, digits, and special characters) for higher-level sites. The rating of the eight websites also confirmed that users distinguish between higher-level sites and lower-level sites. However, the password cracking statistics and the responses of the survey in the second phase suggested that higher-level passwords have a good degree of similarity with lower-level passwords. Thus, it is apparent that while users do have a notion that sites have different levels of security and importance, expedience and simplicity of password management trump what they know are more secure behaviors.

Our work also advocates the need of more sophisticated password training for users to address this issue of password hierarchy. With the passage of time, users would gain more and more Web passwords and password reuse would become more widespread. Our work provides suggestions regarding the pattern of reuse users might maintain to keep their important accounts more secure. Assuming that reuse is necessary, users should reserve strong passwords for primary banking and email accounts and not reuse them elsewhere. For all kinds of content and sketchy accounts, a single weak password could be reserved.

Another important finding of our study is that the degree of semantic similarity is stronger than the syntactic similarity among passwords of different levels of a user. Our cracking methodology through JTR relied only on syntactic similarity. Through word mangling rules, it modified the lower-level passwords in various ways in order to guess the higher-level passwords. Semantic similarity was not examined or used. For example, multiple passwords of a user can be inspired from common source (e.g., music, film, and sports). If one of the passwords of a user is related to a personally meaningful word (e.g., the name of her cat), then it is likely that another of her passwords is also inspired by a similar thing (e.g., the name of her family's cat from when she was a child). In fact, the users' responses suggested that these practices are followed frequently (Table 8 and Table 4). Our cracking methodology did not leverage these kinds of semantic similarity. We believe that by exploiting semantic similarity, a larger percentage of higher-level passwords can be cracked. We leave this as a future work.

The issue of semantic similarity has a wider implication for shared passwords. Our survey responses showed that when creating a shared password, users frequently use a similar thought process as used when creating their identity (email/Facebook) passwords. This reveals a serious breach in the privacy of their identity accounts. As discussed before, users generally share accounts among close friends, family, or apartment mates. If the shared password has any kind of semantic similarity with her other passwords, then it becomes easier for these people to guess those other passwords. Passwords for shared accounts are also frequently created by reusing important identity or financial passwords. All of these findings highlight the indirect consequences of password sharing and suggest that password sharing perhaps should not be considered as a "nuanced practice engaged in with thought and care", as suggested by Kaye (2011).

8. Conclusion

In this work, we propose a hierarchy of users' Web passwords based on the perceived importance level of the sites and conduct a user survey to verify the hierarchy. The responses demonstrated that users consider their financial passwords to be significantly more important than their identity passwords, which, in turn, are considered to be significantly more important than their content and sketchy passwords.

We also conducted a laboratory experiment where we asked the participants to construct these four types of passwords. We exploited the content and sketchy (lower-level) passwords of a participant along with a password-cracking dictionary to crack that participant's identity and financial (higher-level) passwords. We could successfully crack almost one-third (106 out of 320) of the higher-level passwords in this method. This number is significantly higher than the number of passwords cracked by using the password-cracking dictionary only, without using the lower-level passwords.

This work also highlights the indirect consequences of password sharing. In our survey, we asked our participants regarding their password construction practices for shared accounts (accounts for paid subscription sites like Netflix or accounts for sharing a common service such as wireless Wi-fi). We found out that users use a thought process similar to the one used for creating an identity password when they create a shared password. Thus, passwords for shared accounts could be exploited to compromise important identity accounts.

An attacker could also exploit the content passwords to compromise users' identity accounts. Our survey results revealed that although most of the users are conscious and use a fixed set of passwords for content sites, a majority of these conscious users further reuse this fixed set of passwords for their important identity or financial accounts. These findings show that although users consider their identity passwords to be significantly more important than their lower security level passwords, they are not conscious enough to protect themselves from attacks that might leverage these lower security level passwords to guess their identity passwords. For financial passwords, users are relatively more conscious. However, the percentage of users who reuse their financial passwords to construct their lower-level passwords is not nominal.

We acknowledge the fact that our hierarchy of Web passwords did not originate from the users. We proposed it based on prior work and certain observations, and later verified it by conducting a comprehensive user study. There is more to do to completely understand how users mentally classify all of their password-protected accounts in real life. When cracking the higher-level passwords, we also did not consider the semantic similarity at all. We plan to work on these issues in future.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CNS-1117866.

References

Adams, A., Sasse, M.A., 1999. Users are not the enemy. Commun. ACM 42 (12), 40–46.

Adams, A., Sasse, M.A., Lunt, P., 1997. Making passwords secure and usable. In: BCS HCI.

Bonneau, J., Preibusch, S., 2010. The password thicket: technical and market failures in human authentication on the web. In: The Ninth Workshop on the Economics of Information Security.

Brewer, M.B., 2000. Research design and issues of validity. In: Reis, H.T., Judd, C.M. (Eds.), Handbook of Research Methods in Social and Personality Psychology. Cambridge University Press, New York.

Brostoff, S., Sasse, M.A., 2000. Are Passfaces more usable than passwords? A field trial investigation. In: Proceedings of Human Computer Interaction.

Chiasson, S., Forget, A., Stobert, E., van Oorschot, P.C., Biddle, R., 2009. Multiple password interference in text passwords and click-based graphical passwords. In: Proceedings of the 16th ACM Conference on Computer and Communications Security.

Clason, D.L., Dormody, T.J., 1994. Analyzing data measured by individual Likert-type items. J. Agric. Educ. 35 (4), 31–35.

Corner, M.D., Noble, B.D., 2002. Zero-interaction authentication. In: Proceedings of the 8th Annual International Conference on Mobile Computing and Networking.

Fahl, S., Harbach, M., Acar, Y., Smith, M., 2013. On the ecological validity of a password study. In: Proceedings of the Ninth Symposium on Usable Privacy and Security.

Florêncio, D., Herley, C., 2007. A large-scale study of web password habits. In: Proceedings of the 16th International Conference on World Wide Web.

Gaw, S., Felten, E.W., 2006. Password management strategies for online accounts. In: Proceedings of the Second Symposium on Usable Privacy and Security.

Goldberg, J., Hagman, J., Sazawal, V., 2002. Doodling our way to better authentication. In: CHI '02 Extended Abstracts on Human Factors in Computing Systems.

Haque, S.M.T., Wright, M., Scielzo, S., 2013. A study of user password strategy for multiple accounts. In: Proceedings of the Third ACM Conference on Data and Application Security and Privacy.

Hayashi, E., Hong, J.I., 2011. A diary study of password usage in daily life. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems.

Ives, B., Walsh, K.R., Schneider, H., 2004. The domino effect of password reuse. Commun. ACM 47 (4), 75–78.

Kaye, J.J., 2011. Self-reported password sharing strategies. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems.

Keith, M., Shao, B., Steinbart, P.J., 2007. The usability of passphrases for authentication: an empirical field study. Int. J. Hum.–Comput. Stud. 65 (1), 17–28.

Kim, H.-J., 1995. Biometrics, is it a viable proposition for identity authentication and access control? Comput. Secur. 14 (3), 205–214.

Komanduri, S., Shay, R., Kelley, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F., Egelman, S., 2011. Of passwords and people: measuring the effect of password-composition policies. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems.

Marino, C., Penedo, M.G., Penas, M., Carreira, M.J., Gonzalez, F., 2006. Personal authentication using digital retinal images. Pattern Anal. Appl. 9 (1), 21–33.

Notoatmodjo, G., Thomborson, C., 2009. Passwords and perceptions. In: Proceedings of the Seventh Australasian Conference on Information Security, vol. 98.

Orne, M.T., 1969. Demand characteristics and the concept of quasi-controls. In: Rosenthal, R., Rosnow, R.L. (Eds.), Artifact in Behavioral Research. Academic Press, New York.

Porter, S.N., 1982. A password extension for improved human factors. Comput. Secur. 1 (1), 54–56.

Preibusch, S., Bonneau, J., 2010. The password game: negative externalities from weak password practices. In: Proceedings of the First International Conference on Decision and Game Theory for Security.

Roddy, A.R., Stosz, J.D., 1997. Fingerprint features—statistical analysis and system performance estimates. Proc. IEEE 85 (9), 1390–1421.

Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F., 2010. Encountering stronger password requirements: user attitudes and behaviors. In: Proceedings of the Sixth Symposium on Usable Privacy and Security.

Sun, S.-T., Pospisil, E., Muslukhov, I., Dindar, N., Hawkey, K., Beznosov, K., 2011. What makes users refuse web single sign-on? An empirical investigation of OpenID. In: Proceedings of the Seventh Symposium on Usable Privacy and Security.

Syta, E., Kurkovsky, S., Casano, B., 2010. RFID-based authentication middleware for mobile devices. In: Proceedings of the 43rd Hawaii International Conference on System Sciences.

Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A., Memon, N., 2005a. Authentication using graphical passwords: effects of tolerance and image choice. In: Proceedings of the 2005 Symposium on Usable Privacy and Security.

Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A., Memon, N., 2005b. PassPoints: design and longitudinal evaluation of a graphical password system. Int. J. Hum.–Comput. Stud. 63 (1–2), 102–127.

Zhang, Y., Monrose, F., Reiter, M.K., 2010. The security of modern password expiration: an algorithmic framework and empirical analysis. In: Proceedings of the 17th ACM Conference on Computer and Communications Security.