Influencing User Password Choice Through Peer Pressure

by

Andreas Sotirakopoulos

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

The University Of British Columbia
(Vancouver)

December 2011

Abstract

Passwords are the main means of authenticating users in most systems today. However, they have been identified as a weak link to the overall security of many systems and much research has been done in order to enhance their security and usability. Although, many schemes have been proposed, users still find it challenging to keep up with password best practices. Our current work is based on recent research indicating that social navigation can be used to guide users to safer, more secure practices regarding computer security and privacy. Our goal is the evaluation of a novel concept for a proactive password checking mechanism that analyzes and presents to users, information about their peer's password strength. Our proposed proactive password feedback mechanism is an effort to guide users in creating better passwords by relating their password strength to that of other system users. We hypothesized that this would enable users to have a better understanding of their password's strength in regards to the system at hand and its users' expectations in terms of account security. We evaluated our mechanism with two between-subjects laboratory studies, embedding our proactive password checking scheme in the Campus Wide Login (CWL) mechanism for changing an account's password. In our study, we compared the password entropy of participants assigned to our proposed mechanism to this of participants assigned to the current CWL implementation (no feedback) as well as to the traditional horizontal bar, employed by many web sites, which provides feedback in the form of absolute password strength characterization. Our results revealed significant effect on improving password strength between our motivator and the control condition as well as between the group using the existing motivator and the control group. Although, we found a difference between the no feedback condition and the two feedback conditions, we

did not find any difference between feedback conditions (i.e., relative vs. absolute strength assessment). However, our results show that relating password strength to that of one's peers, while maintaining the standard visual cues, may yield certain advantages over lack of feedback or current practices.

Preface

A user study were conducted as part of this research. For this study (explained in Chapter 3), we obtained a human ethics approval (H11-00206) from the UBC Behavioural Research Ethics Board (BREB).

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to offer my gratitude to my supervisor, Dr. Konstantin Beznosov, for his support and mentorship these last two years.

Also, I would like to sincerely thank Dr. Cormac Herley, Dr. Serge Egelman and Ildar Muslukov for their invaluable contributions throughout the whole course of this project. Ildar Mulsukov developed the main components of the proxy web server and conducted part of the participant sessions in the first experiment. Furthermore, he provided feedback throughout the project contributing to design decisions and interpretation of the results. Dr. Cormac Herley and Dr. Serge Egelman provided feedback with the design of the study as well as with the interpretation of the results throughout the project. The initial idea for the Peer Pressure mechanism stemmed through discussions between Dr. Herley and Dr. Beznosov.

Thanks to Dr. Karon MacLean and Dr. Sidney Fels who kindly accepted to be in my committee.

I would like to thank my friends at the Laboratory for Education and Research in Secure Systems Engineering (LERSSE) for their constructive feedback on my research as well as for their friendship and support.

To Ioannis and Sofia, my beloved parents,

and

Kostas, my brother and best friend.

Chapter 1

Introduction

The evolution of networked computing and especially the Internet, with the many user centric/data sensitive capabilities readily available, have made user authentication a top priority in systems deployed today. The main authentication mechanism employed in millions of computer installations and web sites is passwords. Despite their predominance as a security mechanism as well as their ease of maintenance and deployment as means of authentication, passwords have been identified from the early years of their usage as the weak link in the security chain of many applications [18, 33, 38]. Password have maintained their predominance as form of authentication even in the face of new developments (e.g., biometric authentication devices) and this seems to remain the case for the foreseeable future. Moreover, it has been argued and shown that for systems with multiple users, the overall security of accounts and of the system is dependent upon the quality of individual account passwords. In an effort to secure their systems, administrators create mandatory password policies that users are required to follow when creating or altering their passwords. However, policies often require users to remember lengthy and/or complicated passwords or even randomly generated passwords. This might render the passwords ineffective [1] as users will resort to mechanisms (e.g., write the password on a post-it note and stick it on the PC) that might turn out to be more risky than having a slightly weaker but easier to remember password. To address this, one proposed solution has been to educate and supply guidance, during and prior to password creation, to the user [1, 43]. Research has shown that educated

users create better passwords than users that receive no guidance on how a good password is created [60].

A popular scheme that helps users choose strong passwords is to proactively check passwords. This mechanism is currently employed by many web sites serving millions of users. Most of the times, by using a number of criteria set by the developers, these proactive password checkers provide users with feedback labeling their password as weak or strong. In most cases, unless the chosen password violates a specific policy requirement (such as minimum password length) the user is usually allowed to use the chosen password even if it is indicated as weak or of medium strength. However, users are faced with many different implementations of proactive password checking mechanisms that yield different password strength assessments, based on a plethora of criteria administrators have chosen. Users receiving contradictory strength assessments, even for the same password, might become confused about what constitutes a good/strong password and/or lose confidence in the feedback they receive by such mechanisms thus rendering them useless or even counterproductive.

In this work we seek to investigate the possible advantages that a password strength meter, comparing the user's password strength to the one of his peers, would have over the traditional password strength meter and/or the lack of one. Our research focuses on answering whether, and to what extend, peer pressure motivators (PPM) stimulate users more effectively than not providing feedback or providing feedback by means of other types of existing motivators (EM) in creating better passwords. In addition, we seek to investigate whether PPM would affect the participants ability and willingness to maintain the chosen password, as well as, what trade-offs, in terms of labeling password strength, should be taken into consideration when implementing a PPM strength meter so as to be more effective in guiding/convincing users to create better passwords.

We consider our approach as an application of social navigation. Social navigation is utilized in creating user interactions with a system that are driven by other users of the system, not only the designer. In the realm of privacy and security, social navigation can be used to guide users towards safer, more secure decisions [6, 13, 23]. In the context of proactive password checking, providing the user with information about their peers' choices and giving feedback indicat-

ing the strength of the password, would be considered a type of social navigation. This might be more effective and better understood than an abstract decision about the password strength as indicated by a standard password strength meter. Also, users might end up creating better passwords in an effort to be better than a certain percentage of their peers in the system.

1.1 Study overview

In the present work, we conducted a between-subjects laboratory study having as participants UBC students, faculty and staff using the UBC's Campus Wide Login (CWL) system as a platform. During our study we manipulated the password interface to determine the effect of different types of password checking mechanisms. We did not reveal our study's purpose, but rather informed participants that they would assist in an evaluation of a current UBC portal (my.ubc.ca). Participants were told that they would participate in an evaluation of the current interface, performing a number of tasks, so as to identify points that a new portal design should take into account. While they tried to log into the web portal using their CWL account a proxy server we had installed to the computer uses redirected them to one of our prototypes and they were asked to change their password. This step was presented as an unrelated to the study, UBC IT, policy change. As participants changed their passwords using one of the prototypes, we gathered password strength data about their CWL password using our proxy server and prototypes. Participants were assigned to the following 3 conditions: no proactive password checking, proactive password checking following current industry practices (i.e., a horizontal bar indicating password strength in terms of weak, medium or strong), and a proactive password checking mechanism that employed peer pressure as a means of motivating users to choose better passwords. After about three weeks we contacted participants in order to conduct a follow-up study so as to judge whether the password they created, using one of the prototypes, was still in use or they changed it as a result of issues that stemmed by their effort to create too strong, hard to remember and manage passwords.

The main way in which we evaluated our prototypes' effect on password choice was bit-strength of the password. While participants chose their password we

recorded its bit-strength and we conducted statistical tests to evaluate the prototypes' effect on the strength of the chosen passwords among conditions. Furthermore, we recorded the time and number of trials participants needed to create their new password in each condition and examined whether certain prototypes had an effect on that. Finally, during the follow-up study, we examined whether passwords created using our prototypes were still in use.

Our results indicate that participants were motivated to create stronger passwords in the PPM and EM conditions compared to the Control and the ability of participant to maintain the passwords for the investigated period is not affected by the type of indicator presented to the user. Our data, since EM and PPM did not differ significantly on how they affected participants password choice, do not demonstrate whether peer pressure was the main reason for the improvement of password entropy in the passwords participants assigned to the PPM condition or it was the visual feedback that guided their choice.

Studying the effect of peers' choices on user password selection and demonstrating that by providing peer pressure feedback, users can be motivated to create equally, to the industry's standard method, strong passwords, is the main contribution of our work. Another contribution is the introduction of a paradigm that motivates password choice through feedback of a user's password strength in relation to the password strength of their peers.

1.2 Thesis outline

The remainder of this thesis is organized as follows.

1. Chapter 2 provides the related work and background information for this thesis. It includes the related work on user password practices and attitudes as well as social navigation and proactive password checking. Finally it includes an assessment of the proactive password mechanisms of various popular web sites

2. Chapter 3 presents our study's design which investigates the effect of PPM on password choice.

3. Chapter 4 presents the results of our study

4. Chapter 5 discusses points of interest as they have come up by our results' analysis as well as limitations of our approach.

5. Chapter 6 summarizes the contributions of this thesis, and introduces directions for the future research.

Chapter 2

Background and related work

2.1 Passwords and password policies

Passwords have been the prominent means for authentication almost since the need for user authentication and authorization emerged in multiuser environments. Along with passwords came the concerns about their security and usability. In the first UNIX systems different options for password creation and security were proposed and evaluated [38]. It was early understood that because of users' weak passwords practices and choices such as using the username as their password security risks came into being [7, 17]. The realization, in these early years, of the weakness a single ill-chosen password posed to the whole system led to large volumes of research in the creation of secure passwords and password policies and it has been proposed by many researchers that a good policy will help increase the security of user accounts in a given system [35, 48, 49, 51]. However, in practice, policies are not always easily understood or followed by users. This lack of understanding and inconvenience that strict policies place on users might lead to a drop in productivity and user frustration, as shown in [28, 31]. Also in [56] Vu et al. demonstrated that imposing password restrictions alone is not sufficient for creating stronger passwords and different techniques should be employed to ensure a stronger password creation strategy by users. But then, how users are going to decide how "strong" a password is?

Bruce Schneier in his article "The Psychology of Security" [46], describes se-

curity as both a reality and a feeling that is not always based on the actual security had or needed. In addition, he argues that security is a trade-off and that personalized risks are taken more seriously than generalized ones. Also, successful security systems must take into account these user perceptions in order to promote user relation to the security mechanisms in place and have a better response to them, as they will feel that they put into them the right amount of effort. This view is further reinforced by the work of Adams et al. in [2] where the authors have conducted a study that has shown that users conform to security mechanisms to the degree that their perception of security levels, information sensitivity and burden on their work practices matches their perception of the risk involved. Herley, in [28], argues that an overly restrictive password policy can be the cause for a bigger harm (particularly economic) than the harm the policy has meant to prevent. The research discussed above indicates that, depending on the system and on user expectations, password policies can have a severely negative impact on the security of the system instead of improving it. This leads to the conclusion that usability of passwords and password creation policies might be even more important than security measured in bit strength or time needed to crack a password for an account. Taking this considerations one step further Florencio and Herley in [19] demonstrated that web sites that care about competition, although they have huge assets to protect, seem to adapt more lax password policies than sites that don't have the need to compete - like universities for example. This implies that the password strength is not the only, even not the most important, defense against loss of valuable assets. In fact, an unusable, user-hostile policy might lead to loss of revenue and popularity instead protecting a company's assets.

Based on the above discussion it is evident that usable passwords and password policies are quite important and that the current practices for password creation guidance are not always fitting the systems and the users they are intended to protect.

2.2 Towards stronger passwords

Much research has been conducted on mechanisms and policies that will enable users to choose strong, memorable passwords. Various avenues for password cre-

ation have been explored. Among them, there are systems that employ graphical passwords ([52] for a survey), which utilize images instead of the traditional textual passwords to limit adversary's abilities to attempt brute-force attacks like the ones commonly used against systems with textual passwords as well as enhance memorability of the passwords. Graphical passwords have, however, their own drawbacks (e.g., susceptible to shoulder surface attacks - when an adversary can acquire a password by observing the owner while using it) and much research is still directed towards tackling them. Other avenues that might lead to stronger passwords have been sought, as well. In [20] Forget et al. present a system that aims at improving password strength by placing randomly-chosen characters at random positions into the password. This system was successful in increasing password security but at the same time users came up with strategies that would limit the mechanism's effectiveness when many random characters were placed into a password. In order to accommodate better password creation strategies Yan et al. in [60] suggested that mnemonic phrase-based passwords, memorable phrases condensed into passwords, could be employed and provide equal protection to this of random passwords. However, as it was demonstrated by Kuo et al. [36] even these passwords could be broken, especially as human mnemonic phrase dictionaries would become more available to attackers. Furthermore, a common, mechanism to help users in creating strong password has been proactive password checking.

2.2.1 Proactive password checking

It has being suggested that educating users, letting them understand the need for security and the rationale behind good password choices will lead to better overall security of a system as well as better attitude towards password policies on their part [43]. However there are cases that education and guidance are ineffective or the users might not be willing or savvy enough to read and understand the policies in place, let alone the reasoning behind them. In such cases alternative, automatic, mechanisms should be employed. Such a mechanism is the reactive password checker. The administrator periodically checks the system to find guessable passwords with password cracker programs. Accounts that are cracked are suspended until the passwords have been changed. The disadvantage of this mechanism is that

these checks consume resources and there is the possibility that between checks a vulnerable account is exploited. As a response to this disadvantage, proactive password checking has been proposed [4, 5, 7]. A proactive password checker is a mechanism that interacts with the user while they are creating or changing their account's password and informs them whether their password is one that could be easily guessed or not. Proactive password checkers operate as a form of user education at the time of creation of the password and can also be used to explain why the password chosen is inappropriate for the task (e.g., too short). Over the years proactive password checking has been extensively studied in various cases. A few systems that check passwords proactively based on different rule sets and try to discourage or disable users from using weak passwords can be found in [7, 34, 44, 61]. These systems may utilize the bit space (entropy) of passwords or the resemblance of a given password to commonly used passwords like for example "p@ssw0rd". In most cases, the password meter relies on designer choices about the rule-set employed (i.e., the policies that passwords must follow to be deemed fit for acceptance).

This is where the most important difference in our work lies. Instead of having the administrator of a system decide of the password strength needed for a given system we se the potential of letting the users of the system decide. Our notion is in agreement with prior research conducted by Brown et al. in [8] who, after presenting work that suggests that password requirements of easiness and obscurity are diametrically opposed [41], suggest that users should differentiate between items where security is important versus ones where a security breach would not lead to a compromise of critical data and create passwords of appropriate strength in each case. We believe that, for certain systems, this might be a good approach as the strength of the password will relate to the risk perception of the user and the value they place on their data thus no unneeded burden will be put on the users from overly strict (as perceived) password policies that do not necessary reflect the users' perception of data value.

2.3 Social navigation

2.3.1 An overview of social navigation

The main idea behind our research stems from the ideas of social navigation in computer interaction. In the general case, the term is used to describe the interaction of people with a place or a system which is based on the actions others have taken and the information trace they have left behind. As such, social navigation leads to a personalized, dynamically changing system. An example of a social navigation system, outside the realm of computing, could be as simple thing as a path in a forest, created by people that have passed through there, before its current user [37]. The path has been created dynamically by the its users and it is not part of the initial "design" as, for example, a city street might be. Systems with social navigation capabilities are utilized more and more in various areas of everyday life. For example, in the case of Ayers et al. [3], a system was developed that led to less energy consumption by consumers who employed it by utilizing feedback from the energy consumption levels of the consumer's peers. Such a system, could lead to huge energy conservation by guiding users to consume less without applying strict or hostile, as perceived by consumers, policies like price raising.

In computing, a social navigation system is a computing system that collects and aggregates behaviors, decisions, or opinions from users and provides this information to others, in order to guide their behavior and decision making [14]. This information can be either direct (e.g. in the form of reviews about a product) or indirect (e.g., in the form of popularity scoring based on views of a video on a web site). The notion of social navigation is by no means new in the realm of computer science. As early as 1945, Vannevar Bush [9] in his article "As We May Think" has discussed and explored the idea of people leaving trails in information space. These trails could be utilized by other users in various ways to interact with the system, depending on the system's design and their needs. Dieberger et al. have discussed social navigation as means that enable users to have an overview of how other users interact with the system instead of feeling isolated in their interaction with it. By introducing the term "social affordance", they discuss systems where interaction is created dynamically and in a way that users perceive it as one guided

by what their peers have done or are currently doing instead of what the designers want them to do [12]. Social affordance, therefore, might help users and designers to determine finer aspects, or even new ones, of a newly created system and the interaction of its users with it. In our case, by introducing such a social affordance, while users create passwords, we hope to explore different approaches in password selection and usage, dictated by the actual users of the system instead of a designer or administrator.

Furthermore, social navigation systems have undergone considerable investigate and research exists providing guidelines for their design, like the work done by Hook et al. [30]. Also, many systems have been, and still are, created following the principles of social navigation in various fields of computer applications. An early and interesting idea of a system that applies social navigation has been the one introduced by Hill et al. in [29]. In this work the authors have developed a system that creates indicators on the scroll bar of a document indicating positions in the document that have been edited or read and how often/much this has happened. This way users are able to quickly identify points in the document that are stable or are under revision by other users. Another early example of work on collaborative systems that employ social navigation is the Tapestry system developed by Terry et al in 1993 [54]. In that system the user's emails are assigned priority based on several filters the user has created. One way to filter messages is a collaborative filter which looks at recommendations from other Tapestry users and based on the preferences set it can assign priority to messages. To conclude this brief survey of systems in areas other than security, it is worth discussing the work of Svensson et al. on a system that uses social navigation for the presentation of food recipes to its users [53]. The recommendations are based on an algorithm that clusters recipes depending on how they are prepared (e.g., vegetarian) and lets users interact with direct (e.g., chatting capabilities of the system) or indirect (e.g., ordering of recipes within a recipe group) means. From the systems presented here, it is easily seen that social navigation has been successfully employed in many areas, including commercial ones, like Netflix, which use previous user's choices but also ones of their friends to recommend future movies.

2.3.2 Social navigation and computer security

More recently, security researchers have started to utilize social navigation in security and privacy. Research has demonstrated that users are unmotivated [15, 59] and not knowledgable enough [24, 45] to use and/or understand the complex security guidelines and practices they are needed to follow. For the average user, security is a secondary task that, some times, is an obstacle during the performance of a task. For this, users tend to find shortcuts and workarounds which might result in bad security practices (e.g., writing their ever changing account password on the proverbial post-it note and sticking it on the PC). Also the same research has shown that users prefer to delegate security to others. In particular, people prefer to delegate security duties to organizations (e.g., IT department) or trusted individuals that they consider knowledgable and have helped them in the past with security issues. However, since access to such individuals might not be always available and general guidelines set by IT experts might not fit every system or user interaction alternatives have been considered.

This is where social navigation comes into play. Direct approaches have been taken in order to utilize social navigation in computer security. An example of such an approach would be “PhishTank” [40] system where its users, rate various web sites as to whether they are phishing sites or legitimate ones. This is a classic user feedback/review system seen in many contemporary system designs, not necessary concerned with security (e.g., online bookstores having consumer reviews). Another, similar, approach concerned with the security of application installation on mobile phones is presented by Chia et al. in [10]. In that work, researchers seek to investigate how a closer circle of related users might guide one in making security decisions. They have utilized a users close social circle (or ”clique”), as compared to a larger community, to provide recommendations regarding the installation of an online application. It is demonstrated that friends’ negative advice about the installation of an application is regarded higher than community positive reviews. This is an interesting result, indicating that people regard the advice of users they feel they know better and are closer to, higher than the overall community of a system. This is a result that has been observed in other fields of research (i.e., economics). Peer pressure has been found to be more effective when it is enforced by

people one cares about and feel that his or hers action might affect them since guilt (internal pressure) seems to be more effective than shame (external pressure) in rising productivity [32]. Moreover, social navigation has been employed in security systems like Acumen [22] which is an internet explorer bar that supplies recommendations about actions regarding web site cookies. These recommendations are created by aggregating community choices on these web sties. The system uses colors to communicate recommendations to users as well as more detailed information should the user require it. Another system is a firewall, named Bonfire [23], which uses community feedback regarding allowing or not of internet access to various applications. In the Bonfire's case both colors (as visual cues) as well as tagging of application and choices, by other Bonfire users, are used.

DiGioia and Dourish have suggested to approach security mechanisms and security in general as a facet of interaction [13]. This is a step away from the classical implementation of social navigation mechanisms found in the web where reviews, comments and system suggestion based on user choices prevail. They attempt to bring social navigation to the periphery employing ideas drawn by Weiser's ubiquitous computing [57, 58] as well as the Tapestry system [54]. In their work they examine the security implications by determining patterns of conventional use and by disclosing the activities of others. Utilizing the Kazaa peer-to-peer application they examined how users can benefit from being presented with information about folder sharing choices others have made. They try to guide user choices on folder sharing utilizing subtle visual cues (e.g., folder icon) that depend on popular choices made by the user's peers. Also, by using a notion of piling and grouping different shared files into piles, they offer to the user an overview of the activities of other users. The security implications of such designs can be important in the sense that they serve the usable security concept of successfully incorporating the user into the determination of security instead of having designers taking all the security decisions for the user. Our work aims at taking this idea one step further, integrating social navigation, in the form of peer pressure, into a core security mechanism of most systems today. By utilizing visual and written cues we aim at subtly guiding users towards better password choices depending on the system and the system's community practices. We use the term peer pressure somewhat liberal in this context. By it we do not imply that others actively put pressure on an indi-

vidual to choose a good password rather we expect individual users by being aware of what their peers are doing to feel internal pressure in performing analogously.

2.4 Current practices of password strength measuring

Many websites today are enforcing rigid password polices (e.g. requirements of characters diversity, periodical password change, minimal password length etc.). It is also a common practice to show password's strength through visual or textual password meters, so that during password change or/and sign-up processes users are presented with feedback on their new password. In 2007 it was shown by Furnell [21] that users guidance in password selection varies from website to website. In order to check whenever results shown by Furnell are still relevant today, we partially repeated his tests on 5 well known websites.

2.4.1 Testing existing password strength meters

To understand the way most popular websites evaluate passwords, we used similarly to the work done by Furnell, the following criteria while trying to create an account with 5 popular web sites (GMail, YouTube, Facebook, MSN Live, Yahoo).

Password Entropy (PE) - is defined by Claude Shannon [47] and is used as a measure of the password's uncertainty (entropy).

Keyboard Layout (KL) - special algorithm which tests whenever or not the password is a set of sequential keys on the keyboard.

Black List Check (BKL) - tests whether the password is in the most common (popular) password list.

Dictionary Check (DC) - tests whether the password is a dictionary word. (Note: according to NIST [39] guidelines, the size of the dictionary has to be at least 50K words).

Advanced Dictionary Check (ADC) - test algorithm which uses the same dictionary as the DC test, but also checks whether a password is a result of dictionary words combinations.

Letters Substitution (LS) - it is an addition to BKL, DC and ADC algorithms tests, which reveals whether letters are replaced by their corresponding special characters, such as ”a”-¿ ”@”, ”s”-¿”$”, etc.

Profile Information Test (PIT) - tests whether passwords are checked against your public information, such as first name, last name, birthdate, username (or email address), etc.
Other heuristic (OH) - other heuristic algorithms used to check some specific aspects of the password. We use this type of heuristic to highlight that website uses some specific heuristic which is neither common nor significant.

For all passwords we used the same user identity where the selected name was John Smith, the selected date of birth was 01/01/1990 and the selected user was either the supplied email or testjohnsmith2010.

2.4.2 Results of testing

We had 2 main goals for our overview of current password meters:

1. Discover ways of showing password strength used by current websites;
2. Understand and discover different approaches used to generate password strength feedback.

In Appendix A we present password meter snapshots for the top 5 websites (GMail, YouTube, Facebook, Microsoft (MS) Live, Yahoo). The figures in the appendix show different system states of the password strength meters in different states from these 5 websites. The state of the meter depends on the user input. We also tested the algorithmic part the of existing strength meters which are used to convert password to verbal variable, such as 'Invalid','Weak','Strong', etc. Result of that survey are shown in Table 2.1. Also, in Table 2.2 the various verbal characterizations of passwords, for the sites surveyed are presented.

Results in Table 2.1 show that GMail uses most of the described techniques, although some of those are not fully implemented such as: LS is not recognizing $ sing as 's' letter, PIC is not checking for surname/forename. Also GMail account does not require birthdate, so we weren't able to check this aspect of PIC logic. Facebook uses all techniques described above although their implementation of LS is not perfect, e.g. 'p@ssw0rd' is a strong password for Facebook. Youtube failed in PIC because it accepted the "user's" "email" as a strong password. MSN does not allow user to use any type of his personal information by policy, that is much stricter than others, but it do not implement LS as part of the password checking.

Table 2.1: Password criteria used for password strength calculation by 5 popular websites.

Name	PE	KL	BKL	DC	ADC	LS	PIC	OH
GMail	+	-	+	+	+	-/+	-/+	
Facebook	+	+	+	+	+	-/+	-/+	+
MS Live	+	+	+	+	+	+	+	-
YouTube	+	-	+	-/+	-	-/+	–/+	
Yahoo	+	-	+	-	-	-	+	

Table 2.2: Password assessment levels

Website	Feedback given to the user
GMail	Too short, weak, fair, good, strong
Facebook	Too short, weak, medium, strong
MS Live	Weak, medium, strong
YouTube	Too short, weak, fair, good, strong
Yahoo	Too short, weak, strong

Another important aspect we tested was the minimum EM requirements for the web sites, Table 2.3. We see that they are quite different among them

We can see that among the web sites surveyed, there are huge differences on how password strength feedback is implemented and what strength assessments the users are presented with.

2.4.3 Differences in feedback among sites

A major motivator behind this work was the observation that users have to deal with a number of different and sometimes conflicting password strength assess-

Table 2.3: Minimum requirements across tested websites.

Website	Information sent back
GMail	8 characters minimum length
Facebook	6 characters minimum length.
MS Live	6 characters minimum length, cannot use username or email.
YouTube	(Same as GMail), 8 characters minimum length
Yahoo	6 characters minimum length, cannot use username or email.

ments and feedbacks at various sites. Many popular web sites seem to implement their own flavor of feedback indicators using different criteria in order to access the strength of the passwords provided by their users. As already presented, various sites use different heuristics and methods to calculate and communicate the feedback to their users.

It is easy to conclude and anticipate, from Tables 2.1 and 2.2, that passwords will be ranked quite differently among the 5 popular web sites we surveyed but we wanted to have concrete data for this hypothesis of ours. To that end, we used the Rockyou password dataset to calculate the password strength feedback a user would receive if they were to supply a password across three major web sites (Facebook, Microsoft Live and Google). These web sites have literally hundreds of millions of users and it is safe to assume that they are regarded as reputable and trustworthy by their client base. We used around 2.6 million randomly selected, unique passwords that met Google's minimum length requirements, from the Rockyou dataset, to calculate the feedback their users would receive among those web sites.In addition, we calculated the feedback users would receive in 13.6 million passwords that complied to MS Live and Facebook minimum requirements. The results are quite interesting, indicating huge discrepancies between the web sites not only based on the minimum requirement set for length (8 characters for Google and 6 for Facebook and MS Live) but also due to the way strength is calculated in each.

For our calculations we replicated the code found on the MS Live and Facebook web sites as it is publicly available in the form of Javascript whereas in the case of Google, where the code is not available, we opted to submit the passwords to their server via automatic queries and received the password strength assessment as a response. We did not try to reverse engineer the way they calculated the password strength as we could not be sure what kind of dictionaries they would be using for their checks.

Furthermore, due to difference in the implementations of the algorithms there are millions of passwords that would receive a high rating in MS Live but not Facebook (i.e., a password can receive a rating as "strong" with only 6 digits on Facebook but not on MS Live where at least it must have 7). Also MS Live uses an extensive dictionary where as Facebook does not seem to do any dictionary

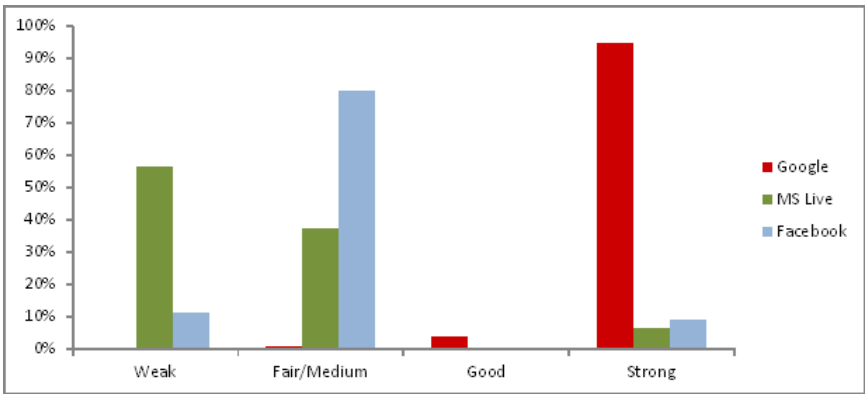

Figure 2.1: Assessment of passwords, compliant with Google's minimum requirements, across MS Live, Facebook and Google web sites.

checks. Google also controls for various dictionary and common passwords but not the same as MS Live.

In Figure 2.1, we present the percentage of passwords per password strength that would be assigned by MS Live, Facebook and Google and are compliant to Google's minimum requirement of 8 characters. In total we analyzed about 2.6 million passwords and it is evident, from the figure, that the differences, in the feedback users of these passwords would receive, would be great. Especially between the assessments of Google and MS Live/Facebook.

In Figure 2.2, we present the percentage of passwords per password strength that would be assigned by MS Live and Facebook to passwords that have, at least, a length of 6 characters or more. The passwords in this case are about 13.6 million and we can still see slight differences (several thousands) between MS Live and Facebook that derive from the fact that although close the way a strong password is defined defer slightly. Although, in this case the differences might seem very small we should keep in mind that similar percentages do not equal similar feedback on similar passwords. Rather, passwords that would be considered weak in Facebook's case are medium for MS Live and vice versa. When we looked into how the same password is assessed between MS Live and Facebook we found that almost 1 out of 4 passwords would receive a different assessment, in regards to strength, among those two web sites. This is a huge number of over 3 million pass-

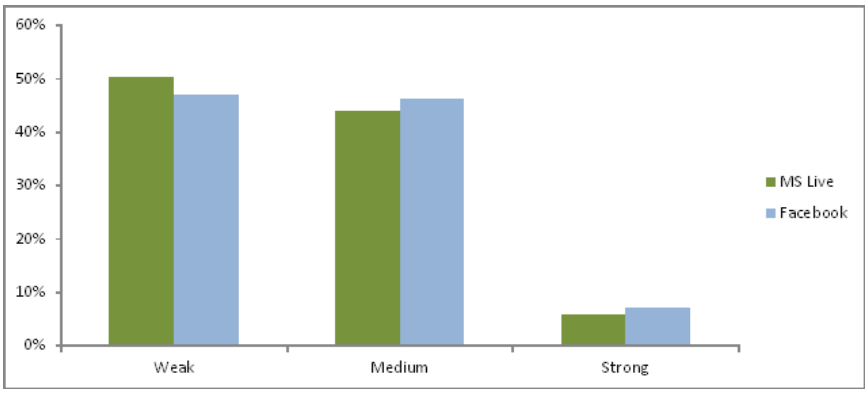

Figure 2.2: Assessment of passwords, compliant with Facebook's and MS Live's minimum requirements, across MS Live and Facebook.

words. Also, as seen from Figure 2.1 the more the minimum requirements become stricter, the more differences among web sites become evident.

From the figures above, it is evident that users that may try to use the same password, or similar, across these web sites, will receive non-uniform, even contradictory, feedback for no apparent, to them, reason. Even though the web sites give similar instructions/policies on what constitutes a strong password, users are known for not reading policies and even if they read them there is no explanation for penalties on the password's strength (e.g., due to the dictionaries used by MS Live or Google that are not readily available to the average user).

Chapter 3

Methodology

Our work aims at answering the following two research questions:

1. *RQ1*. To what extend do peer pressure motivators (PPM) stimulate users to create better passwords in comparison to other types of existing motivators (EM) and in the absence of any motivators?

2. *RQ2*. Does PPM have an impact on users choice to maintain their newly created passwords in the long run?

We have established three hypotheses based on our research questions:

1. *H1*. Participants exposed to our PPM condition, will create passwords with a higher entropy value, compared to participants that will be exposed to EM and to no proactive password checking (i.e., Control).

2. *H2*. The behavior of our participants towards our PPM and EM implementation will depend on computer expertise as well as password practices, such as using a password managers.

3. *H3*. Participants' choice of maintaining the new password will not be affected by the type of motivator a participant will be exposed to. In particular PPM will not lead participants in creating passwords that they will find difficult to handle and use every day.

To test our hypotheses, we decided to opt for a laboratory study that would utilize UBC's Campus Wide Login (CWL) service, which interface we altered to embed our password feedback mechanisms. Our study was implemented as a between subjects design with three conditions as described below.

1. Control Condition (CC). In this condition we replicated the current CWL change password web site which does not use any motivators to entice users to create stronger passwords. Figure 3.1.

2. Existing Motivator Condition (EM). A motivator that, following the common practice of most web motivators, is a horizontal bar that changes length and color (red, orange and green) and uses the words weak, medium and strong to indicate password strength. Figure 3.2.

3. Peer Pressure Motivator Condition (PPM). In this condition we implemented a vertical bar that used a green and a red sub-bar that informed the user whether the input password was stronger or weaker than a percentage of CWL users. Figure 3.3.

Our hypothesis was that participants would be motivated to choose better passwords than their peers in the system upon receiving feedback that would compare their password's strength to that of their peers. Since we did not have access to the actual CWL data we opted to use the password strength distribution of the RockYou database passwords, which complied with CWL's password policies, to seed the percentage feedback intervals of our meter. After designing and running our study, with 60 participants evenly spread among conditions, we did not find any statistically significant difference between the PPM and control conditions. We attributed that to the way we chose to display feedback to the participants in the PPM condition. It seemed that our intervals were making too easy for participant to reach above 50% of relative strength and thus the indicator failed in motivating them to create better passwords. We readjusted the intervals and re-ran the study, keeping all other parts exactly the same, using 47 participants. The exact intervals, for each experiment, are presented in 3.1.1.

3.1 Study design

In order to validate our hypotheses we needed a design that will have two main characteristics to ensure a certain degree of ecological validity in our study. We required a design that will ensure that participants will interact with a system and create a password they actually care about and rely on for their everyday work. Furthermore, we wanted to shift the focus from the actual password creation task to another primary task so as to maintain a realistic user case scenario. Most of the times, users change their password either on system demand, due to a password expiration policy or because they feel their accounts are in danger of being compromised.

We designed our study in order to satisfy those requirements. We chose our university's Campus Wide Login (CWL) account system as the platform on which we would implement our password feedback mechanisms. CWL is an account UBC students, faculty and staff use on a regular basis to access university services like E-Classes, grades, paying of fees, university email accounts etc. The CWL authentication and authorization platform is embedded into most major UBC websites that require users to log in and it is an important account for UBC members. As it was not feasible to implement our password strength meters on the actual CWL platform we needed to create an environment that would allow us to run a controlled study maintaining a realistic set up.

The UBC web site we chose in order to ask users to use their CWL account and test our password strength feedback mechanisms was the MyUBC web portal (http://my.ubc.ca). We felt that this web site was an appropriate choice for our study as its existence is well know among UBC members and it is used as a portal to access information about UBC, university email and interact with other members (e.g., posting sale ads). On the other hand, this web site is not one that most UBC members use very frequently. Because we intended to make changes in the login procedure we wanted a web site that participants wouldn't have used recently and frequently so as not to raise suspicions about our study goals.

To maintain an unbiased approach towards the password choice made by our participants we did not reveal our true study goals. Instead, we advertised our study as one aiming to redesign the current MyUBC portal. We claimed that participants

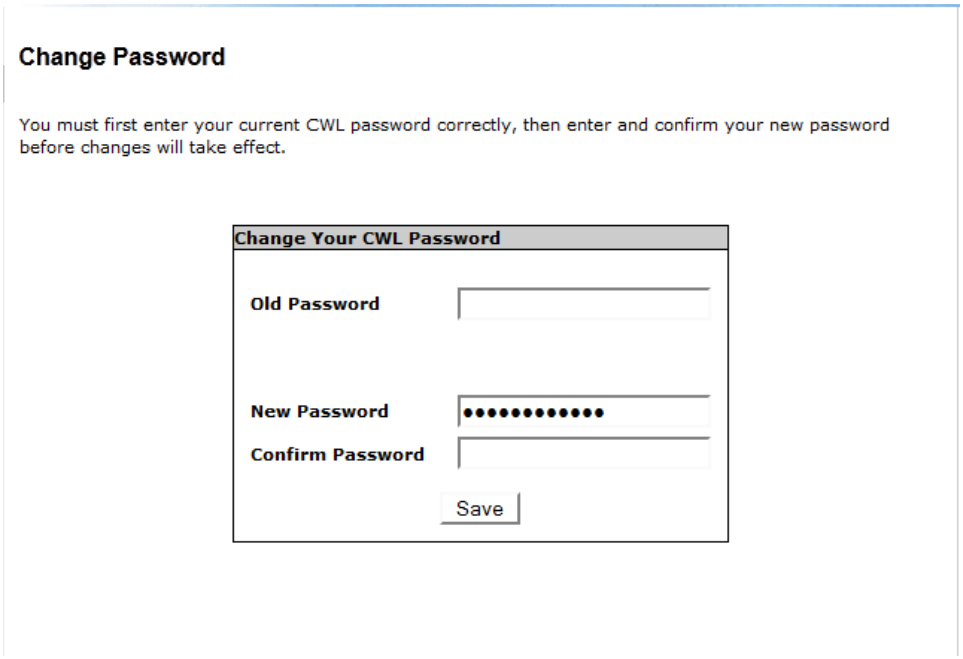

Figure 3.1: The control condition prototype.

will perform a number of tasks using the myUBC portal which were supposed to help the researchers assess the portal's usability and usefulness so as to give recommendations for a new interface for it.

The current design of the web site requires users to log in using their CWL account in order to access its services and we used that step in the user interaction in order to present our password feedback mechanisms.

We created a proxy server and installed it on the virtual machine users used to perform the tasks required. Each participant was randomly assigned to a condition by the proxy server. Upon inputing their account information they were redirected by the proxy to a web site that mimicked the actual CWL's web site password change layout, with the addition of the password feedback mechanism for the current condition. A pop-up window informed them that due to a new IT policy their password had expired and they need to create a new one as seen in Figure 3.4. The proxy server and the prototype interfaces are presented, in detail, in section 3.1.1.

Upon arrival, participants were greeted by a researcher and were shown to the

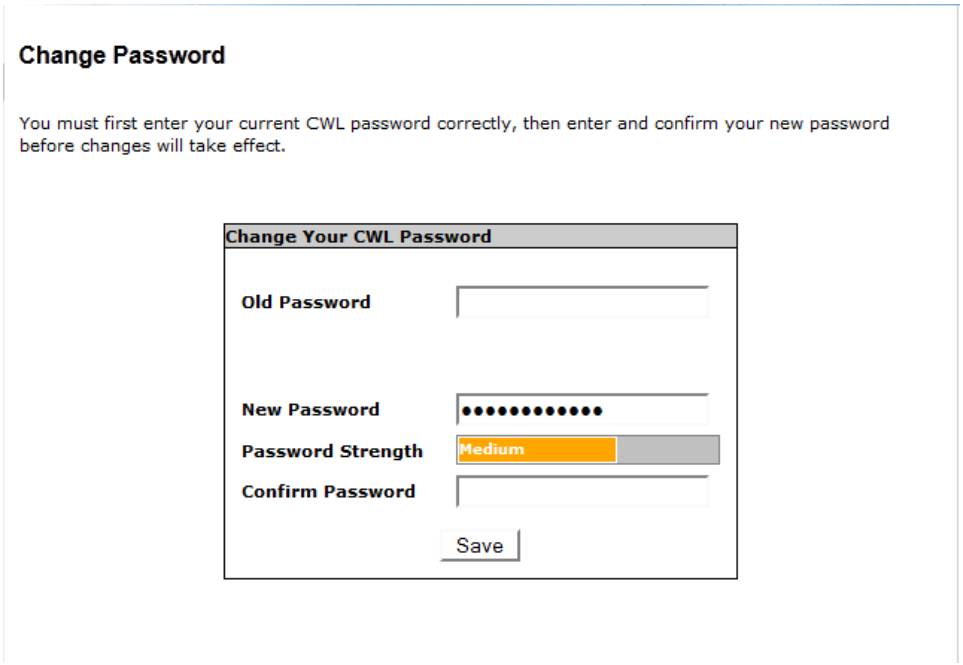

Figure 3.2: The EM condition prototype.

room where a computer was set up for this study. The researcher, having memorized the script, informed them about the supposed goals of our study and explained the experimental procedure. Each participant was first handed a consent form and a questionnaire used to gather demographic as well as computer expertise information. Also, the questionnaire included a series of dummy questions about the myUBC portal in order to reinforce the participants' belief in the study's advertised goals. Appendix B.1 presents the questionnaires.

After the participant had completed the questionnaire the researcher handed them the first task. After the completion of each task the next was handed to the participant. The three tasks were the following.

1. Add an ad in the classified section in the "other" section for a $50 coupon for the KEG restaurant in downtown Vancouver at 1499 Anderson Street.

2. Using the myUBC portal find the most popular question from the Vancouver - Ask Me.

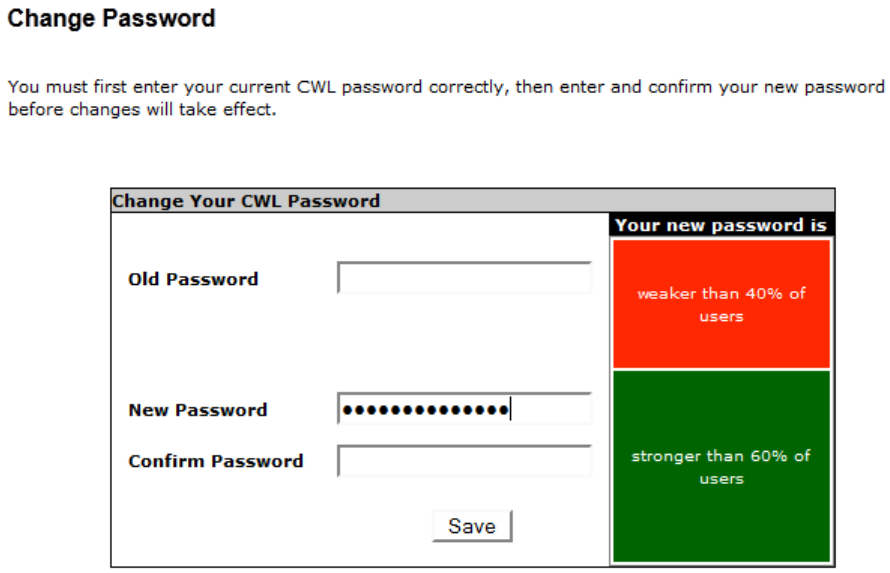

Figure 3.3: The PPM condition prototype.

3. Delete the ad created during the first task.

We felt that these three tasks required enough effort, on the part of the participants, in order to convince them for our goal to assess the various aspects of the portal's usability.

After the participant completed the third task the researcher asked them to complete a questionnaire, giving feedback on their experience using the web site while performing the tasks (see Appendix B.4). This step was part of our effort to maintain the deceit about the advertised purpose of our study (i.e., assessing the usability of the MyUBC portal) as we intended to have a follow-up session and we did not want participants to take any action regarding their CWL password on the account of finding out our study's true purpose. The follow-up session's purpose was to investigate whether they still used the password that they created or they ended up changing it because they found it too hard to remember. This choice was made because we wanted to investigate whether password motivators have an effect on the participants ability to manage their new password in the long run (i.e., lead users

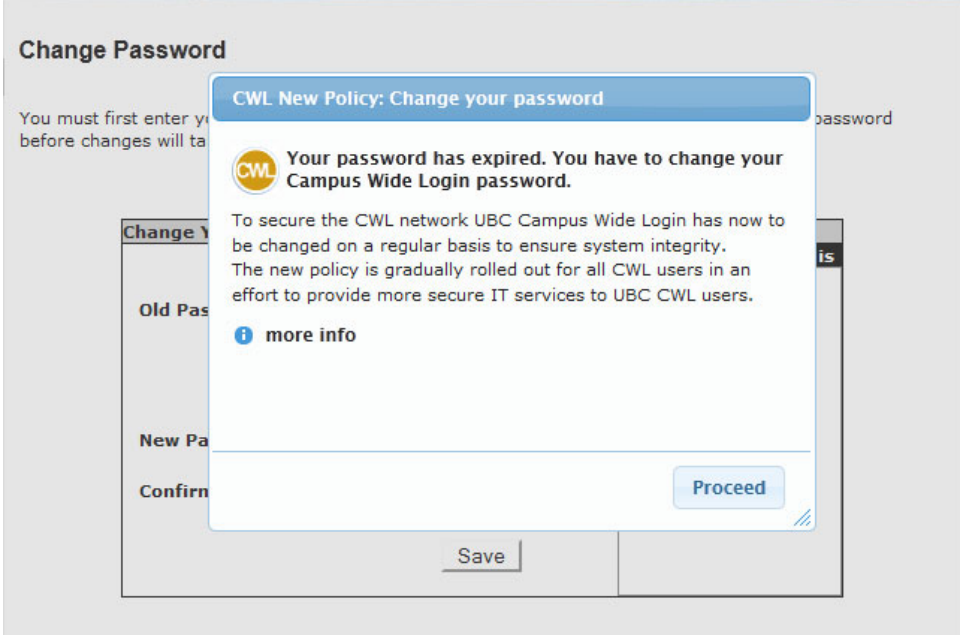

Figure 3.4: The pop-up window informing participants about the ”new UBC policy” for password expiration.

to create too complicated password that they found hard to remember after some time). When the second questionnaire was completed the researcher informed participants about when the second session would take place and the first session came to an end.

3.1.1 Proxy server and prototypes

The proxy server

We developed a proxy server (Figure 3.6), that was used to handle participants' condition assignment to the prototypes, their redirection once they attempted to log into the MyUBC portal using their CWL account information as well as the saving, in an sql database, of the password information for both the old and new password they supplied. The server was invisible to the participant once it was minimized. Participants could use their browser of choice between Firefox 5 and

Internet Explorer 9. The browsers had been configured to use the proxy server for their http and https requests. We created and installed an SSL certificate on the system so our participants would not see any SSL warnings while trying to change their password. Furthermore, our prototypes resided in an external to UBC server but the proxy server altered the URL so as to give the impression that the password change interface was the actual CWL interface (i.e., https://cwl.ubc.ca) and not alert knowledgable participants. The server intercepted the user's account information (i.e., username and password) but did not store the password in clear text.

When the participant tried to reach the my.ubc.ca website for the first time he was presented with the login page of the portal asking to use his CWL account as normal. The proxy server checked with the CWL system on whether the information provided by the participant was valid. In case the account was invalid the participant received the error message they would usually receive in such a case. Otherwise, the server redirected the participant to one of our prototypes. The prototype web page interface was loaded and a pop up was displayed. The pop up informed the participant that a new policy set by the UBC IT service called now for passwords to be changed in regular intervals. This new policy was presented as completely unrelated to our experiment. The pop up had a "more info" link that gave further information about the supposedly new change. No participants followed that link. All URLs were altered by our proxy server to ones that seemed to originate from the ubc.ca domain maintaining the impression that this requirement was truly one that UBC IT had set up. The proxy server did not allow navigation to the MyUBC portal unless the password was changed. If a participant tried to go back the MyUBC portal without changing their password, the server automatically redirected them to the password change webpage. After they had changed the password and their data were logged the proxy server became a transparent proxy and allowed all traffic to pass unchanged.

In Figure 3.5 we demonstrate the steps we took for accumulating our data in detail.

We managed to accumulate a rich dataset regarding participants' behavior while using the prototypes and choosing a password. The proxy server saved in a database the participant username, old password and new passwords (hashed as, out of eth-

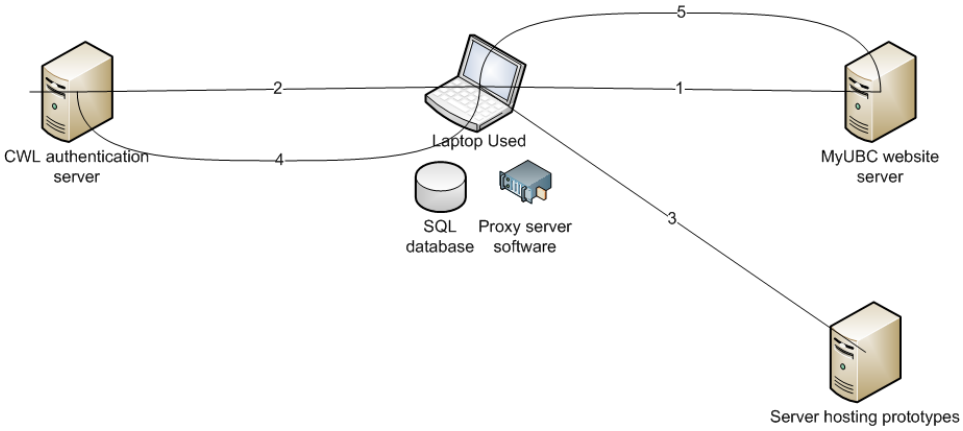

Figure 3.5:

1. The participant uses CWL username and password while trying to log into the MyUBC site.
2. The proxy server checks with the CWL authentication server whether the credentials are valid.
3. Upon validation of the credentials the proxy server redirects to the server hosting the prototypes. The participant chooses his new password and submits the form (along with the time spend on various components of the prototype.
4. The proxy server contacts the CWL authentication server and attempts to change the CWL password. Upon success saves in the local database the hashed values of the old and new password as well the time it took the participant to create the password.
5. The proxy server redirects to the MyUBC website and from that point on becomes transparent not affecting the participant's interaction with the MyUBC website.

ical considerations, we could not store them in clear text), the number of digits, lower and upper case letters, special characters and length of both old and new passwords as well as the strength of the password calculated using the Shannon entropy formula 3.1. In addition, the Levenshtein distance between the old and new password strings was calculated as a measure of how different the two password were. This was as we wanted to investigate whether participants would opt for small variation of their passwords or would choose a completely new one. Of course, there is also the potential that participants had a set of password and they

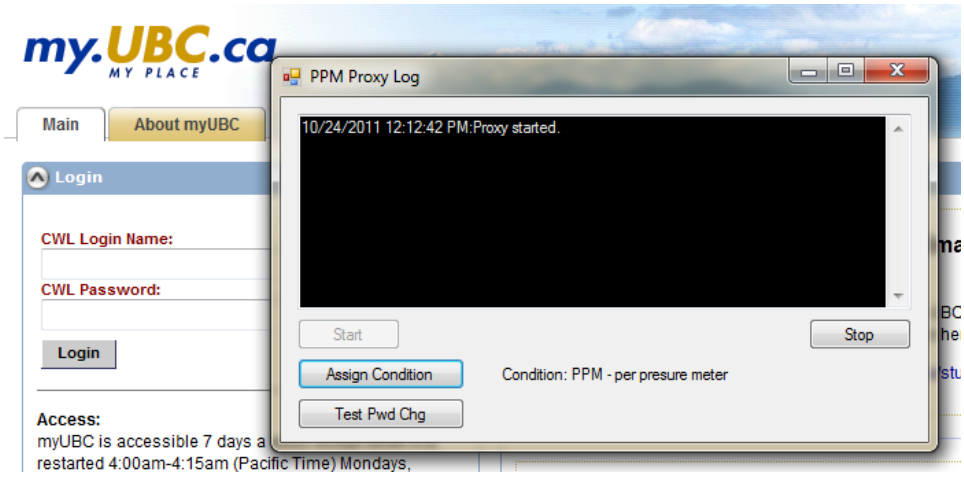

Figure 3.6: The proxy server's interface.

used a password from that particular set.

Prototypes

Our prototypes used javascript to enforce the CWL's native password restrictions for length: a minimum of 8 characters and maximum of 40 as well as the restriction that required a password to contain at least one digit and one letter. We did not add any further restrictions on password creation except for the fact that we did not allow participant to reuse their old password rather we logged their intention in doing so and required them to create a new one. Furthermore, the prototypes logged the time participants spent on the page, reading the pop up message, the number of times they pressed delete and backspace while creating their new password and also logged how many errors they did in password composition or length (not having at least on letter and digit and being between 8 and 40 characters long) and how many attempts to create a new password failed due to mismatches between the new password and the confirmation of this password. These data were submitted along with the password change HTML form when the submit button on the form was pressed and saved in our study's database by the proxy server. We were interested in the number of errors made and time needed by our participants while creating their new password so as to examine whether different feedback conditions yielded

any challenges for their users.

The prototypes, in two of the three conditions, displayed the password strength in the form of bars with different colors and wording according to each condition as seen in Figures 3.2 and 3.3. As we wanted to be consistent in our feedback to participants, password strength across conditions we had to come up with a way to decide on what constituted a strong, medium or weak password. For the PPM condition we needed data from the actual CWL user-base that would yield user percentages per password strength. We needed them because the PPM condition presented strength of password relatively to other users of the system indicating what percentage of users has a stronger and weaker password than the current password choice a participant was making. As we did not have access to these data we decided to take another approach in coming up with the percentages of users with different password strength. We decided to use passwords from the Rockyou dataset that complied to UBC's CWL password policy for 8 characters containing at least one letter and digit. This could introduce some uncertainty in the feedback percentages we created but we felt that it was a necessary risk we had to take. We used the simple Shannon algorithm to calculate the bit strength of these passwords and calculated what percentage of Rockyou accounts corresponded to different password strengths. We used those percentages to display the relative strength in our PPM condition. In Figure 3.7 we present the Shannon password entropy distribution as calculated for the Rockyou password dataset. We were interested in the percentages of compliant to CWL passwords per bit-strength. Also, as part of our investigation, we looked into the password entropy of different passwords (i.e., passwords of different composition). These results are presented in Figures 3.8 and 3.9.

In Figure 3.7 entropy is estimated as the log2 of possible password combinations. To calculate the password alphabet size we used Algorithm 1. A more detailed presentation of the Shannon entropy calculation and the reasoning behind our choice is presented in Section 3.1.1.

Furthermore, we had the EM condition bar's percentage coverage adjusted thus having an equal way of presenting password strength in each condition. In Table 3.1 and Table 3.2 we present the the choices of password strength feedback we made for experiments 1 and 2 respectively. These two tables show the bit strength

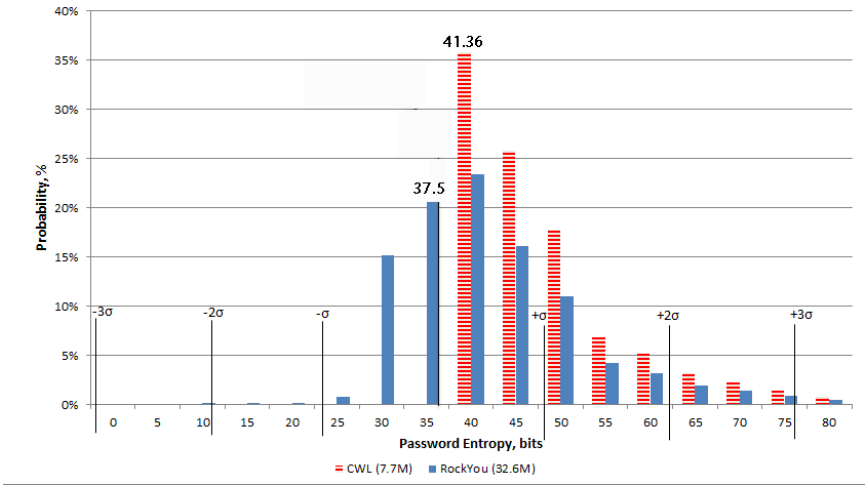

Figure 3.7: Distribution of the Shannon's entropy for the RockYou password database. Both for the general case and those who are CWL compliant.

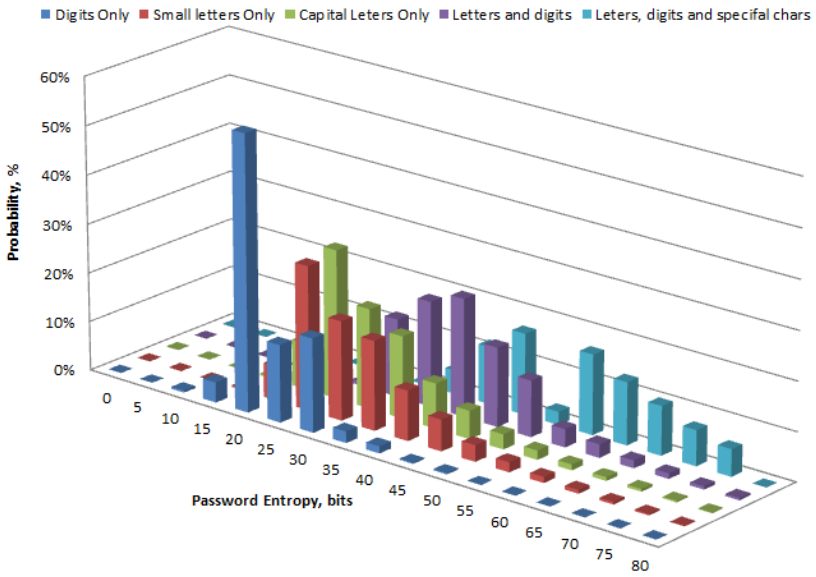

Figure 3.8: Distribution of the Shannon's entropy for the RockYou password database, separately for different types of passwords.

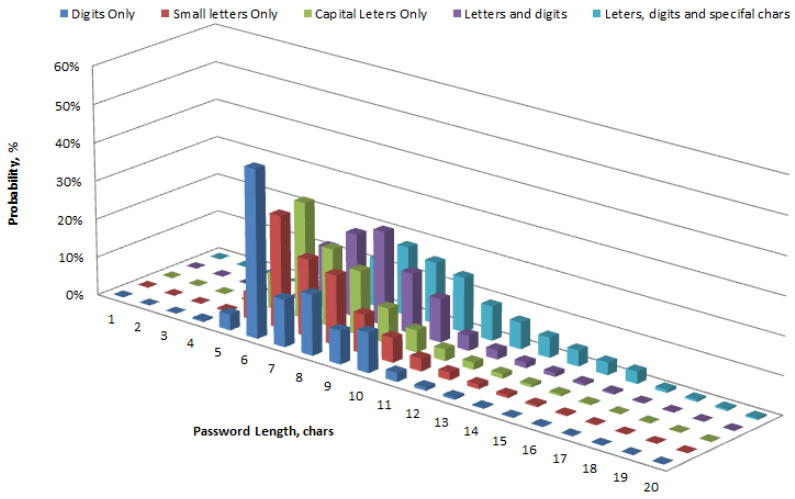

Figure 3.9: Distribution of the Password's length for the RockYou password database, separately for different types of passwords.

Table 3.1: Experiment 1: Password strength intervals used to provide feedback

Participants Bit Strength (x)	**PPM Feedback (stronger than)**	**EM Feedback**
x<=46	35%	Weak
46<x<=48	53%	Medium
48<x<=51	61%	Medium
51<x<=56.99	78%	Strong
56.99<x<=61.99	85%	Strong
61.99<x<=63.99	95%	Strong
63.99<x	100%	Strong

intervals our prototypes used to present password strength assessment to the user. For the second experiment we chose to crank up the intervals by 10 bits per interval so as to make more difficult for participants to achieve a higher strength (in terms of feedback). As shown in the results section, Chapter 4, this change successfully motivated participants in creating stronger passwords in the PPM condition, something that did not happen in experiment 1.

Table 3.2: Experiment 2: Password strength intervals used to provide feedback

Participants Bit Strength (x)	PPM Feedback (stronger than)	EM Feedback
x<=53.41	0%	Weak
53.41<x<=56.53	30%	Weak
56.53<x<=59.83	40%	Medium
59.83<x<=64.26	50%	Medium
64.26<x<=71.09	60%	Medium
71.09<x<=77.21	70%	Strong
77.21<x<=82.27	80%	Strong
82.27<x<=83.30	90%	Strong
83.30<x	100%	Strong

Algorithm for comparing passwords in terms of their strengths

in order to be able to show the strength of the password in EM and to compare different passwords between each other for the PPM conditions, we had to calculate a scalar value for password strength. The strength of the password can be estimated as an entropy (uncertainty of the password) according to Shannon's entropy estimation formula [47].

Another way to assess the password policy strength is according to NIST guidelines [16], which consists of six rules, enlisted below:

1. the entropy of the first character is taken to be 4 bits;
2. the entropy of the next 7 characters are 2 bits per character;
3. for the 9th through the 20th character the entropy is taken to be 1.5 bits per character;
4. for characters 21 and above the entropy is taken to be 1 bit per character;
5. A bonus of 6 bits of entropy is assigned for a composition rule that requires both upper case and non-alphabetic characters. This forces the use of these characters, but in many cases these characters will occur only at the beginning or the end of the password, and it reduces the total search space some-

what, so the benefit is probably modest and nearly independent of the length of the password;

6. A bonus of up to 6 bits of entropy is added for an extensive dictionary check. If the attacker knows the dictionary, he can avoid testing those passwords, and will in any event, be able to guess much of the dictionary, which will, however, be the most likely selected passwords in the absence of a dictionary rule. The assumption is that most of the guessing entropy benefits for a dictionary test accrue to relatively short passwords, because any long password that can be remembered must necessarily be a pass-phrase composed of dictionary words, so the bonus declines to zero at 20 characters.

Covert et. al showed in his work [11], that for printed English each character worth about 1.3 bits of entropy. This algorithm is relying solely on the password length; multiplying it by 1.3 we can estimate password entropy. But Covert's work only considers English dictionary words. We should note that the NIST guidelines do not refer to the password's strength in itself, rather to the bit-strength of the password policy and should not be used to calculate password bit-strength.

All entropy estimation algorithms, mentioned above, have pros and cons. They are easy to implement and don't require a lot of computational resources. But not all of them address such cases when users chose easy to guess password from a dictionary or most common password list. Florencio and Cormac in [18] showed that users often use lowercase or numerical passwords for websites, so that an attacker can use this knowledge and go through a reduced password space (numerical or lowercase letters) thus reducing the effective entropy and decreasing the strength of the users' selected passwords.

However, as the main research question in this work is whenever peer pressure influences users' password choice and the current state of the password policy on CWL does not check passwords against profile data and allows dictionary words to be used as part of the password, we decided to use the Shannon entropy estimation formula to rank passwords and do not use dictionary or a common password list checks. Although, such a formula can be considered naive and over optimistic, since we apply it consistently over all conditions in our study, both to calculate password strength and provide feedback, will enable us to have a way of compar-

```
nAlphabetSize = 0
if HasSmallLeter(sPassword) then
    nAlphabetSize+ = 26
end if
if HasCapitalLeter(sPassword) then
    nAlphabetSize+ = 26
end if
if HasDigit(sPassword) then
    nAlphabetSize+ = 10
end if
if HasSpecialSign(sPassword) then
    nAlphabetSize+ = 33
end if
```

Algorithm 1: Password's vocabulary size estimation algorithm.

ing passwords created by participants in different conditions. We recognize that participants have their own perceptions about password strength and what makes a password strong so feedback inconsistent to their view might make them lose confidence in the indicator providing. However, since CWL does not conduct any dictionary checks it would be quite hard to keep the conditions equal if we introduced dictionary checks. For example, participants assigned to the control condition would be punished (in bit strength) because of the use of a dictionary word without having any means of being aware of it like in the two other conditions.

The password strength is calculated as shown in Equation 3.1 with **l** being the password's length and **alphabetSize** calculated by Algorithm 1.

$$PasswordStrength = log_2(alphabetSize^l) \tag{3.1}$$

We decided not to use English dictionary in order not to introduce new password requirements to those CWL's website already has. We also found that there are only 50 out of 3157 most common passwords in the list of the crack tool John The Ripper [55] which comply with the current CWL password policies and they were used only 52690 times in RockYou passwords database (0.68% of all passwords complying with CWL policies).

Another obstacle with using dictionary words and the most common passwords

list for feedback purposes, in a way, that we will be able to compare EM and PPM conditions with the Control condition is violating equality among the conditions. In EM or PPM it is easy to employ dictionary criteria without confusing participants (i.e, if a password loses strength dramatically after the user typed the last character of the password, we can explain by showing a hint that the password was found in the dictionary or in the most common passwords list). However, it is hard to do so in the Control condition, where we are not supposed to provide any feedback on the users to guide them in password creating, as it is the case with the current CWL implementation.

Likewise with user's profile data, it is very desirable to test whenever user is creating his password on the basis of the publicly available information or not, but that would require change of the control condition (current state of the website), so that it will also check this aspect during password change/creation phases too, as well as how to provide a feedback to the user on that regard.

In general, our main goal was to investigate different ways to motivate users for good password choices not how restrictions policies work. That is why we decided to have our EM and PPM prototypes using the same password policy that the current state of CWL is using and not introduce any new requirements for the password.

3.1.2 Follow-up study

About two to three weeks later the researchers contacted, by email, the participants for the follow-up session. Participants were presented with a CWL login interface as shown in Figure 3.10. The login interface was similar to the CWL standard login interface but was created by the researchers. They were asked to log into the web site so as to complete a survey. When they input their account information, their password was hashed and along with the username a PHP scripted checked them against a database containing the usernames and hashed values of the password of the participants that had already taken part in the first session. The PHP script first checked whether the account provided by the participant was a valid CWL account by querying the actual CWL platform. If the account was a valid one, it checked to see whether the participant had actually taken part in the first session by querying

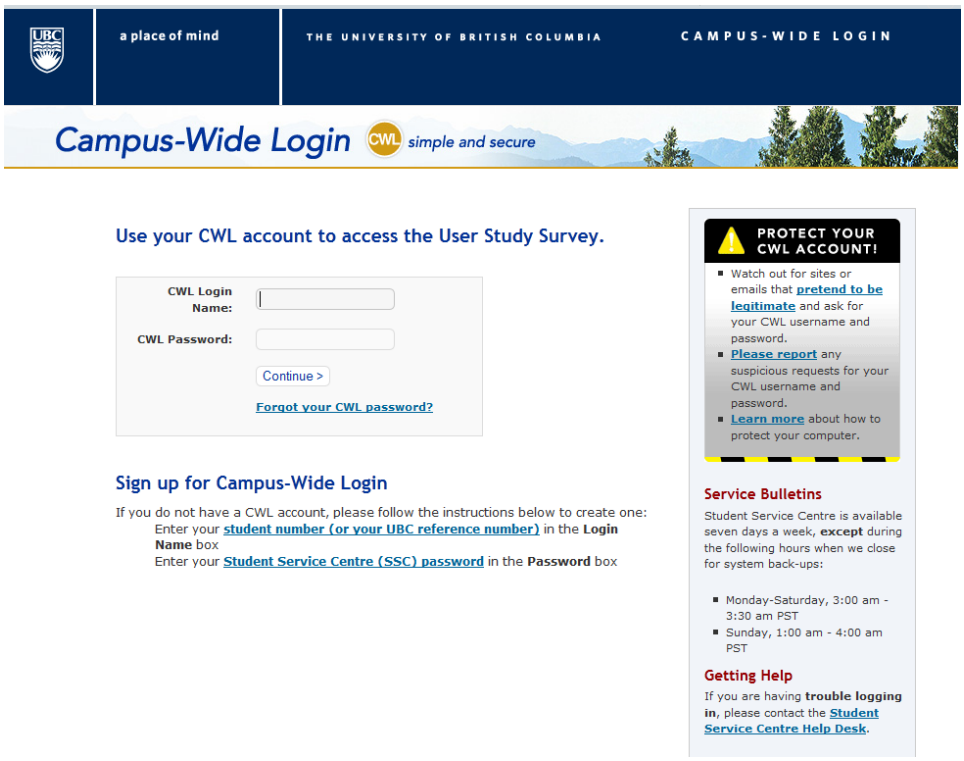

Figure 3.10: The proxy server's interface.

the username against our database. If the username was present in our database the password provided at that time was hashed and checked against the hash value of the one the participant had created during the first session.

In case the two hash values matched the participant was redirected to a survey asking questions about their password usage, choices and practices. They were also asked about how concerned they would be in case they had various account types (e.g., banks, email, CWL, Facebook) compromised. Finally participants were asked what, in their opinions, constitutes a good password.

In case their CWL password had changed since the last time they were in the lab with us, they were redirected to a similar, to the previous case, questionnaire with the addition of an open ended question asking why they chose to change their password. For both online surveys (changing and maintaining the new CWL password respectively), please see Appendix B.2 and B.3.

After completing the online survey, the participant was debriefed and the true purpose of the study was revealed.

3.1.3 Recruitment of participants

Participant recruitment was done via flyers hung around UBC campus boards as well as emails sent to department mailing list and Craigslist. They asked for students, faculty, and staff to participate in a forty minutes study during which they will evaluate and give feedback on the usability and design of the myUBC portal (see Appendix B.5). The only requirement for participation was to have a CWL account so as to have access to the myUBC portal. Potential participants were informed that they are going to perform a number of tasks before giving their feedback on the usability of the site as well as suggestions on how to improve the web site. Each participant was given $45 ($20 for the first session and $25 for the second session) as honorarium for their participation.

Chapter 4

Results

In this chapter, we present our results for experiments 1 and 2. In both cases we were particularly interested in how the two feedback conditions would fare in improving the mean password strength between the old and new passwords of the participants compared to the control condition and between each other. Statistical significance was achieved in the second experiment in the improvement of strength between old and new password vs. the control condition.

4.1 Second experiment

Upon analyzing our data from the initial study (see Section 4.2) we did not find a significant effect for PPM but we did for the EM condition compared to the Control. When we looked for the reasons behind this it occurred to us that it could be due to the fact that it was easy for our participants' chosen password to be rated as above average (i.e., stronger than the 50% of other users) even when choosing passwords of low, relatively, entropy. The main reason behind this is that we had designed the indicator in such a way that if, for example, a participant had chosen a password with calculated entropy equal to the one of a certain percentage of users the indicator would inform them that their password was stronger than this percentage. This placed passwords, rapidly, in high percentages demotivating users to put effort in creating a better password. To investigate whether our hypothesis was correct, we decided to re-adjust both the EM and PPM feedback shifting the feed-

back scale toward higher entropy values as shown in Table 3.2. Now it was more difficult for participant to achieve a "strong password" indication. As we see from the results presented in this section, our hypothesis was correct and PPM successfully guided participants in creating passwords with higher entropy values. This seems to imply that the notion of "competition" among users for better password could take actually place.

Additionally, this time the difference between old and new password entropy was statistical significant for both PPM and EM vs. the CC conditions. In this study we had 51 volunteers in the first session. Of them, one decided not to change their password when prompted as they felt that the environment was not a safe to do so. Another claimed that they could not remember their password and finally two participants' password was not logged into the database even though he went through with the password change when prompted by our prototype.

We present our results in the subsequent sections.

4.1.1 Old and new password strength

In Table 4.1 and Table 4.2 we present the mean values and standard deviations for the password entropy of the participants' old and new passwords respectively.

From the two tables we can see that our participants had a mean average password entropy that was quite high, even in their old passwords with a mean value, across conditions, of 49.68 bits of entropy. For comparison, we should mention that a password of 8 characters length with lower and upper case letters and digits would yield a bit entropy of 47.63 bits. As the entropy values for the new password did not follow a normal distribution a Kruskal-Wallis test was conducted on the new password entropy values among conditions to investigate differences in password entropy. The Kruskal-Wallis Test revealed a statistically significant difference in the new password entropy across conditions, $\chi^2(2,47) = 8.043$, p = 0.018. EM condition recorded a higher median score (Md = 60.11) with PPM having Md = 59.45 and Control Md = 51.70. To examine where the significance lied, we performed a Bonferonni adjustment to the alpha level. Instead of 0.05 the alpha level, for statistical significance, dropped to 0.017 and we conducted three Mann-Whitney U tests among our groups. The tests revealed a statistical signifi-

Table 4.1: Second study, old password entropy; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	49.98	15	9.90
EM	49.19	16	9.67
PPM	49.9	16	11.42
Total	49.68	47	10.14

Table 4.2: Second study, new password entropy; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	49.31	15	7.02
EM	60.75	16	16
PPM	64.91	16	21.35
Total	58.52	47	17.04

cant difference between Control and PPM $U = 54.50$, $z = -2.605$, $p = 0.009$, with an effect size of $r = 0.4$ (medium according to Cohen's criteria). The rest two comparisons between EM and Control and EM and PPM failed to reach significance with $p = 0.023$ and $p = 0.696$ respectively. In Figure 4.1, we present the comparison of the old and new password entropies as well as the differences between them (as seen in Table 4.5).

In Tables 4.3 and 4.4, we present the composition of the participants' passwords, old and new respectively. They were mostly comprised of lower case letters and digits having the average length above the minimum required, by CWL, password length of 8 characters containing at least on letter and one digit. Participants seem to rely on length, lower case and digits in order to create their passwords. Even when they were asked to change them and received feedback, still they did not increase, extensively, their use of special characters or capital letters. A one-way Analysis of Variance test was conducted in order to explore whether the individual components of the new passwords (i.e., length, number of upper and lower case letters etc) were statistically significantly different among conditions. No statistical significant differences were present among conditions ($p > 0.05$) in all cases.

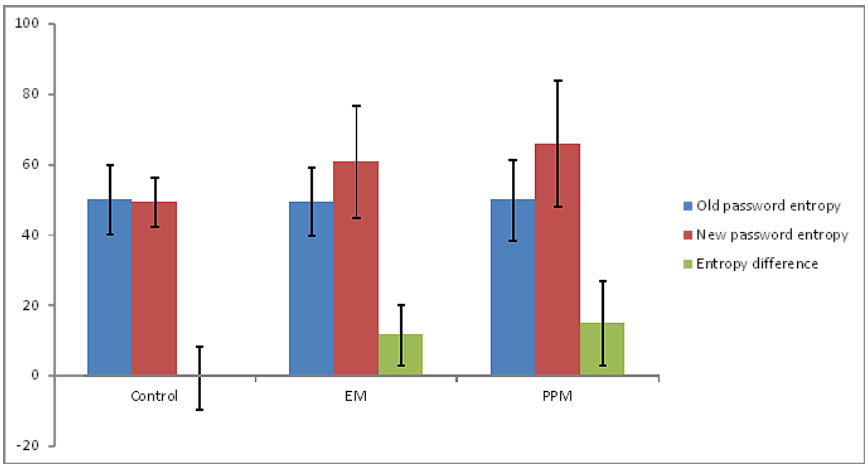

Figure 4.1: Second study, comparison of password entropies between old and new passwords as well as their differences.

Table 4.3: Second study, old password composition; Mean values

Condition	Length	Number of digits	Number of capital letters	Number of lower letters	Number of special characters
Control	9.67	2.67	0	7.00	0
EM	9.13	3.13	0.63	5.38	0
PPM	9.19	2.19	0.44	5.38	0
Total	9.32	2.66	0.36	6.30	0

Table 4.4: Second study, new password composition; Mean values

Condition	Length	Number of digits	Number of capital letters	Number of lower letters	Number of special characters
Control	9.27	2.33	0.53	6.40	0
EM	10.63	2.66	0.75	6.94	0.31
PPM	11.38	2.75	0.56	7.81	0.19
Total	10.45	2.57	0.62	7.06	0.17

Table 4.5: Second study, difference in password entropy between old and new passwords; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	-0.66	15	9.06
EM	11.57	16	8.73
PPM	15.01	16	19.53
Total	8.84	47	14.81

4.1.2 Improvement of password entropy between old and new passwords

Our main objective, during our study, was to investigate whether we could see an improvement in password entropy among conditions, particularly between PPM and EM vs. Control (CC) and vs. each other. In Table 4.5 the descriptive statistics for the difference in entropy between the old and the new passwords are presented. According to our *H1*, we hypothesized that PPM will lead participants, in an effort to do better than their peers, to create passwords of higher entropy value. Our hypothesis was partially confirmed, towards the Control condition. PPM led participants in creating passwords of higher entropy value compared to the control condition. We calculated the difference between the entropies of the new and old passwords. Since they adhered to a normal distribution we conducted an one-way ANOVA to investigate differences. The interaction effect between indicator and improvement in password entropy was significant F (2, 44) = 5.711, p = 0.006. The effect size, according to Cohen's criteria, is large (0.21). Post-hoc comparisons, using the Tukey HSD test, showed that the mean entropy difference was statistical significant between the EM and CC conditions p = 0.04 and between PPM and CC p = 0.007. These results confirm that there is significant effect of the PPM scheme on the improvement of password entropy compared to the Control but not compared to the EM condition p = 0.753.

4.1.3 Effect on computer expertise on password entropy among conditions

In order to evaluate our hypotheses (*H2*) that computer expertise would affect the password choice and password strength of our participants, we had a series of questions that helped judge the expertise of our participants as shown in Appendix B.1. We categorized participants as experts and non experts, using the questionnaire provided in Appendix B.1, and conducted a two-way between-groups analysis of variance to explore the impact of expertise and indicator type on the difference between old and new password entropy. As the Levene's Test of Equality of Error Variances yielded a significant result, $p = 0.013$ we had to adjust our alpha level to 0.01 instead of 0.05 for the two-way ANOVA test. With this alpha level the interaction effect between expertise and condition (i.e., indicator type) was not statistically significant, $F\ (2, 41) = 3.707$, $p = 0.033$.

4.1.4 Time and trials required to create the new password

We wanted to investigate whether the type of feedback, or its lack, had any effect on the time required by our participants to create their password. Therefore, we logged the time each participant spent on the password-change web page as a whole as well as in the "new password" textbox. Significantly more time spent in the case of one condition vs. another could indicate that the participant was taking feedback more "seriously" trying to achieve a higher strength score. A one-way between groups Analysis of Variance was conducted that revealed no statistically significantly differences among conditions for time spent on the web page $p = 0.157$. However, for the time spent in the "new password" textbox we had a statistically significant effect $F\ (2,\ 44) = 3.451$ $p = 0.041$, with a large effect size, 0.14. Post-Hoc comparisons with the Tukey HSD test revealed a significant difference between the CC and PPM conditions, $p = 0.037$ but not between any other conditions. This finding indicates that participant spent time trying to come up with a password that would yield a feedback indicating a "strong" password. This might arguably had an impact on the level of frustration a participant would develop while trying to succeed in his effort but we have no reliable way to measure this. In Table 4.6 we present the mean values of time spent in the new password

Table 4.6: Second study, Time spend in the new password textbox (in seconds); Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	7.13	15	5.85
EM	6.38	16	13.3
PPM	29.56	16	39.31
Total	17.11	47	25.78

text box, across conditions.

In addition to the time participants spent in creating the new password we kept track of their unsuccessful attempts to create a new password. We logged in the database five different types of possible errors a participant could make when submitting a new password. Errors in length (i.e., the password being less or more than 8 and 40 characters respectively), errors in password composition (i.e., the password did not contain at least one digit and one letter), failure to type the same password in both the "new password" and "confirm password" textboxes, failure to type correctly the old password in the "old password" textbox and finally, we logged attempts of our participants to use their old password as their new one. We conducted a series of chi-square tests to investigate a statistical significant relationship between our three conditions and the various errors but none was detected. It doesn't appear that an indicator type is related to a particular error-prone practice on behalf of our participants.

4.1.5 Levenshtein distance

When participants were asked to change their password there were many practices they could employ in order to come up with a new one. Especially, since we had decided to restrain the reuse of the old password as the new one. These practices could include coming up with a completely new passwords, adding a few characters to either end of their existing one or keeping components of the existing password and altering it slightly. Since, out of ethical considerations, we could not store the passwords in clear text we had to find a way to measure the relationship between the old and the new password. As a reliable way in doing so we chose Levenshtein

Table 4.7: Second study, levenshtein distance between old and new password; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	6.93	15	3.63
EM	6.38	16	4.40
PPM	7.50	16	4.21
Total	6.94	47	4.04

distance. Levenshtein distance is a metric which yields the difference between two strings (i.e., edit distance). The higher this is the greater the difference between the two strings is. Ideally, participants would choose a new password which would have little resemblance to the old one. Also we wanted to investigate whether conditions would have an effect on the password choice. In Table 4.7, we present the Levenshtein distance across conditions.

An one-way between conditions ANOVA was conducted to investigate differences in Levenshtein distance among conditions but no statistically significant differences where identified, $p = 0.742$.

4.1.6 Follow-up study

Changing of the new CWL password

Our study had a follow-up component as well. The aim of it was to investigate whether an indicator, especially PPM, would lead users to create password that, although of high entropy value, would be difficult to remember. It is a well documented fact that users can find overly complex/lengthy passwords hard to remember and end up either changing them to easier ones, if able, or writing them down. An ideal password is one that is hard for an adversary to guess but easy for the owner to recall from memory. Out of our initial 47 participants, in this study, 40 took part in its follow-up component. Of them 30 had kept the new password that they created in the first session and 10 had changed it. Of those 10 participants, 9 changed it back to their old one and only one came up with a completely new password. We asked then why they chose to change their password. 4 claimed that

Table 4.8: Second study, participants' passwords in follow-up

Condition	Did not change their new CWL password	Changed their new CWL password
Control	12	2
EM	9	3
PPM	9	5
Total	30	10

they reverted to their old one as they did not want to change it in all places that it was stored (e.g., browser cache) or they did not want to have to remember a new one; another 4 claimed that they had forgotten the newly created password and one claimed that they felt uncomfortable knowing that they changed their password using a "public computer". Finally one participant said that he thought of a better more secure password and changed his CWL password to that. When we investigated, using the Chi-Square test, for any condition impact on password change we did not find a significant effect, $\chi^2(2,40) = 1.714$, $p = 0.424$. In Table 4.8 we present the number of participants that changed, or not, the password they created during phase one of our study according to condition assigned.

Participants' assessment of the CWL account's importance and good password practices

Another point our follow-up study tried to investigate was participants's password recall practices and their perception of the importance of this particular account. Their perception of the CWL account was an important component of our study, as we wanted participants to care about their account and thus create passwords that they would feel comfortable using for a long time. When asked, in a form of a Likert scale questions, to rate how concerned they would be if one of their accounts was to be compromised, the majority rated CWL as very concerned or extremely concerned. Tables 4.9 and 4.10 present participant attitudes among those who did not change their password for the follow-up and those who did. We see that our assumption that participants would care about their CWL account was correct. Furthermore, when we investigated potential differences in account-value perception between the two groups (i.e., people who maintained their new CWL password and

those who did not), our independent T-Test reveal no statistical significant difference among the two groups, $p > 0.05$.

All of our participants claimed that recalling their password from memory was the method they used to remember their password. When, the participants that maintained their new password were asked to rate in a Likert scale the effort they had to put into remembering it they scored an average of 2.10 with 1 being very easy to remember and 5 very hard to remember. Moreover, we asked them to evaluate whether their new password is weaker, equally strong or stronger than their old one. Almost all of the 30 participants that had maintained their new password claimed that their new CWL password was equally strong (14) or stronger (15) than their old one, with only one claiming to be weaker.

Finally, in an open ended question, we asked participants to describe, in their opinion, what constitutes a good password. We received a variety of responses but a few basic concepts were prevalent in our participants opinions about what makes a password strong. Most of them believed that a combination of letters and numbers is sufficient but many also pointed out the need for lengthy, non-word based passwords and containing symbols. Also, participants stressed the importance of being able to easily remember their password although it should not be something that an adversary could easily guess. In total, it seems that many of our participants were aware of general good password practices when creating a password (e.g, adequate length, combination of various characters, not using dictionary words etc.). However, many of their answers were general with little detail (e.g., "a mixture of letters and numbers"). In addition, when asked whether they have used their CWL password in other accounts as well, 22 out of 40 claimed that they did. This indicates that users, even if they are aware of published good password practices, they prefer the well documented, convenience of maintaining a few password across many accounts instead of using a separate one for each.

4.1.7 Participant demographics

For the second study, as with the first, our participants were selected among the UBC students (current and alumni), faculty and staff. Our only requirement for participation was a valid CWL account and we had 51 volunteers in total for this

Table 4.9: Participants that maintained their new password. How concerned would you be if one of your following accounts/passwords had been stolen? (1: Not concerned at all, 5: Extremely concerned).

Account Type	Rating Average
CWL account	3.9
Bank account	4.87
Main email account	4.7
Facebook account	3.87
Forum I am subscribed to	2.79
Messenger account	3.23

Table 4.10: Participants that did not maintain their new password. How concerned would you be if one of your following accounts/passwords had been stolen? (1: Not concerned at all, 5: Extremely concerned).

Account Type	Rating Average
CWL account	3.30
Bank account	5
Main email account	4.4
Facebook account	4.0
Forum I am subscribed to	3.0
Messenger account	2.6

study with 47 of them going successfully through the password change and changing their passwords. Out of the 47 participants that successfully completed the password change study component, 32 were female and 15 male. 35 of our participants were students, 4 were staff, 2 faculty and 3 were prospective students. In Table 4.11 and Table 4.12 we present the age and education composition of our participants as reported by them. In case that a participant was a student we asked them about their level (i.e., graduate or undergraduate) and we present the results in Table 4.13.

Table 4.11: Second study; Age groups of participants

Age Group	N
18	4
19-24	25
25-30	12
31-35	2
36-45	1
46-55	2
56-65	1

Table 4.12: Second study; Completed education of participants

Highest Completed Education	N
Highschool	22
University	15
Graduate School	8
Professional School	1
Other	1

Table 4.13: Second study; If you are a student, you are a(n):

Student Level	N
Undergraduate	28
Masters	2
PhD	4
Other	1

Table 4.14: Second study; Student expertise:

Department type	N
EECE	12
Computer Science	3
Other	22

Table 4.15: First study, old password entropy; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	55.65	19	12.47
EM	51.19	20	10.41
PPM	49.73	20	9.12
Total	52.13	59	10.83

Table 4.16: First study, new password entropy; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	53.34	19	9.59
EM	59.91	20	13.06
PPM	52.33	20	10.1
Total	55.54	59	11.33

4.2 First experiment

In this section we present the data collected from our first experiment. In this experiment we had 60 participants in 3 conditions. However, in one occasion the participant's data were not recorded in the database by our proxy server, probably due to an exception in the proxy server.

4.2.1 Old and new password strength

In Table 4.15 and Table 4.16 we present the mean values and standard deviations for the password entropy of the participants' old and new passwords, from the first experiment, respectively.

From the two tables we can see that our participants had a mean average password entropy that was quite high, even in their old passwords. As the entropy values for the new password did not follow a normal distribution a Kruskal-Wallis test was conducted on the new password entropy values across conditions to investigate differences in password entropy. No statistical significant difference was observed among conditions, $p = 0.092$. In Figure 4.2, we present the comparison of the old and new password entropies as well as the differences between them (as

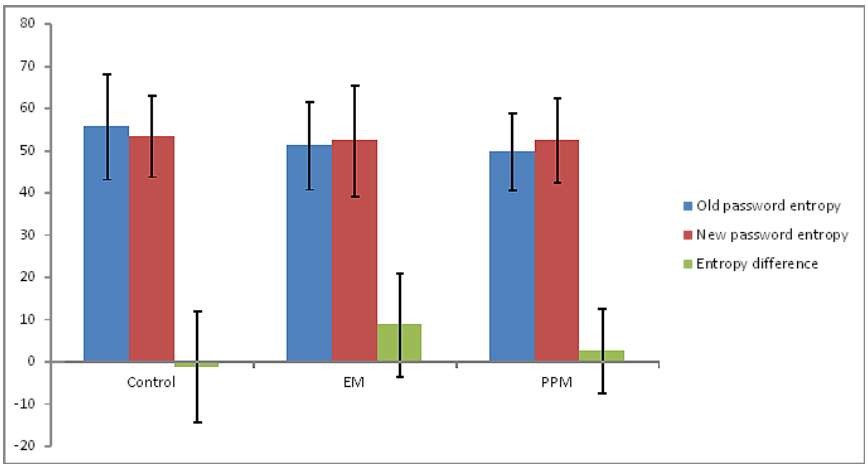

Figure 4.2: First study, comparison of password entropies between old and new passwords as well as their differences.

seen in Table 4.19).

Furthermore, as seen in Tables 4.17 and 4.18, where the composition of the passwords, old and new respectively, participants' passwords are presented, they were mostly comprised by lower case letters and digits having on average a length of 10.27 characters. This was above the minimum, required by CWL, password length of 8 characters containing at least on letter and one digit. Participants seem to rely on length, lower case and digits in order to create their passwords. Even when they were asked to change them and received feedback, still they did not increase, extensively, their use of special characters or capital letters. An one-way Analysis of Variance test was conducted in order to explore whether the individual components of the new passwords (i.e., length, number of upper and lower case letters etc) were statistically significantly different among conditions. No statistical significant differences were present among conditions ($p > 0.05$) in all cases.

4.2.2 Improvement of password entropy between old and new passwords

In order to investigate whether we could see an improvement in password entropy among conditions and especially between PPM and EM vs. Control and vs. each other. In Table 4.19 the descriptive statistics for the difference in entropy between

Table 4.17: First study, old password composition; Mean values

Condition	Length	Number of digits	Number of capital letters	Number of lower letters	Number of special characters
Control	10.05	4.32	0.21	5.21	0.32
EM	9.65	3.40	0.3	5.9	0.05
PPM	9.45	2.4	0.2	6.8	0
Total	9.71	3.36	0.24	5.98	0.12

Table 4.18: First study, new password composition; Mean values

Condition	Length	Number of digits	Number of capital letters	Number of lower letters	Number of special characters
Control	9.95	3.42	0.63	5.69	0.21
EM	11	3.9	0.45	6.55	0.1
PPM	9.85	2.7	0.15	6.9	0.05
Total	10.27	3.34	0.41	6.39	0.12

the old and the new passwords are presented. According to our *H1* we hypothesized that PPM will lead participant to create password of higher entropy value. Our hypothesis was not confirmed. We calculated the difference between the entropies of the new and old passwords. Since they adhered to a normal distribution we conducted an one-way ANOVA to investigate differences. The interaction effect between indicator and improvement in password entropy was significant F (2, 56) = 3.537, p = 0.036. The effect size is medium (0.11). Post-hoc comparisons, using the Tukey HSD, showed that the mean entropy difference was statistical significant between the EM and Control conditions only (p = 0.029). These results do not confirm that there is a significant effect of the PPM scheme on the improvement of password entropy compared to the Control and EM conditions.

4.2.3 Effect on computer expertise on password entropy among conditions

We had hypothesized (*H2*) that computer expertise would affect the password choice and password strength of our participants. We had a series of questions

Table 4.19: First study, difference in password entropy between old and new passwords; Descriptive statistics

Condition	Mean	N	Std. Deviation
Control	-1.29	19	13.18
EM	8.72	20	12.28
PPM	2.59	20	10
Total	3.42	59	12.38

that helped up judge the expertise of our participants as shown in Appendix B.1. We categorized participants as experts and non experts and conducted a two-way between-groups analysis of variance to explore the impact of expertise and indicator type on the difference between old and new password entropy. The interaction effect between expertise and condition (i.e., indicator type) was not statistically significant, F (2, 53) = 0.813, p = 0.449.

4.2.4 Time and trials required to create the new password

Significantly more time spent in the case of a condition vs. another could indicate that the participant was taking feedback more "seriously" trying to achieve a higher strength score. An one-way between groups Analysis of Variance was conducted that revealed no statistically significantly differences among conditions neither for time spent on the web page nor in the "new password" textbox, p = 0.359 and p = 0.310 respectively.

In addition to the time participants spent in creating the new password we kept track of their unsuccessful attempts to create a new password. We logged in the database five different types of possible errors a participant could make when submitting a new password. Errors in length (i.e., the password being less or more than 8 and 40 characters respectively), errors in password composition (i.e., the password did not contain at least one digit and one letter), failure to type the same password in both the "new password" and "confirm password" textboxes, failure to type correctly the old password in the "old password" textbox and finally, we logged attempts of our participants to use their old password as their new one. We conducted a series of chi-square tests to investigate a statistical significant relation-

ship between our three conditions and the various errors but none was detected. It does not appear that an indicator type is related to a particular error-prone practice on behalf of our participants.

4.2.5 Follow-up study

Changing of the new CWL password

Our study had a follow-up component as well. The aim of it was to investigate whether an indicator, especially PPM, would lead users to create password that, although strong, would be difficult to remember. Out of 55 participants that took part in the follow-up component of the study, 38 had kept the new password that they created in the first session and 17 changed it. When, those 17 participants were asked why they chose to change their password, 14 claimed that they reverted to their old one as they did not want to change it in all places that it was stored (e.g., browser cache) or they did not want to have to remember a new one; 2 claimed that they felt uncomfortable knowing that they changed their password using a "public computer". When we investigated, using the Chi-Square test, for any condition impact on password change we did not find a significant effect, $\chi^2(2,59) = 5.192$, $p = 0.075$.

Participants' assessment of the CWL account's importance and good password practices

Another reason for our follow-up study was that we wanted to probe participants about password recall practices and their perception of the importance of this particular account. Their perception of the CWL account was an important component of our study, as we wanted participants to care about their account and thus create passwords that they would feel comfortable using for a long time. When asked, in a form of a Likert scale question, to rate how concerned they would be if one of their accounts was to be compromised, the majority rated CWL as very concerned or extremely concerned. Tables 4.20 and 4.21 present participant attitudes among those who did not change their password for the follow-up and those who did. We see that our assumption that participants would care about their CWL account was

correct.

For the majority of our participants (38), recalling their password from memory was the method they used to remember their password. 1 used a password manager and another claimed that constantly uses 3 passwords so "it is bound to one of them". When, the participants that maintained their new password were asked to rate in a Likert scale the effort they had to put into remembering it they scored an average of 2.13 with 1 being very easy to remember and 5 very hard to remember. Moreover, we asked them to evaluate whether their new password is weaker, equally strong or stronger than their old one. Almost all of the participants that had maintained their new password (29 out of 30) claimed that their new CWL password was equally strong (14) or stronger (15) than their old one.

Finally, in an open ended question, we asked participants to describe, in their opinion, what constitutes a good password. We received a variety of responses but a few basic concepts were prevalent in our participants opinions about what makes a password strong. Most of them believed that a combination of letters and numbers is sufficient but many also pointed out the need for lengthy, non-word based passwords and containing symbols. Also, participants stressed the importance of being able to easily remember their password although it should not be something that an adversary could easily guess (e.g., it should not be one's address or phone number). In total, it seems that many of our participants were aware of general good password practices when creating a password (e.g, adequate length, combination of various characters, not using dictionary words etc.). However, many of their answers were general with little detail (e.g., "a mixture of letters and numbers"). In addition, when asked whether they have used their CWL password in other accounts as well, over 50%, 21 out of 40, claimed that they did. This indicates that users, even if they are aware of published good password practices, they prefer the, well documented, convenience of maintaining a few password across many accounts instead of using a separate one for each.

4.2.6 Participant demographics

Our participants were selected among the UBC students, faculty and staff. Our only requirement for participation was a valid CWL account and we had 59 volunteers in

Table 4.20: Participants that maintained their new password. How concerned would you be if one of your following accounts/passwords had been stolen? (1: Not concerned at all, 5: Extremely concerned).

Account Type	Rating Average
CWL account	3.87
Bank account	4.81
Main email account	4.39
Facebook account	3.36
Forum I am subscribed to	2.5
Messenger account	3.25

Table 4.21: Participants that did not maintain their new password. How concerned would you be if one of your following accounts/passwords had been stolen? (1: Not concerned at all, 5: Extremely concerned).

Account Type	Rating Average
CWL account	4.18
Bank account	5
Main email account	4.47
Facebook account	3.65
Forum I am subscribed to	2.19
Messenger account	3.35

total for this study with 29 of them being female and 30 male. 53 of our participants were students, 3 were staff, 1 faculty and 2 alumni. In Table 4.22 and Table 4.23 we present the age and education composition of our participants as reported by them. In case that a participant was a student we asked them about their level (i.e., graduate or undergraduate) and we present the results in Table 4.24.

Table 4.22: Age groups of participants

Age Group	N
18	5
19-24	29
25-30	21
31-36	4

Table 4.23: Completed education of participants

Highest Completed Education	N
Highschool	21
University	19
Graduate School	12
Professional School	1

Table 4.24: If you are a student, you are a(n):

Type of student	N
Undergraduate	30
Masters	15
PhD	8

Chapter 5

Discussion

In this chapter we discuss and evaluate our results as presented in Chapter 4.

5.1 Effect of PPM on password choice

Our main goal, in the present work, was to investigate whether peer pressure would affect participants' password choice. Our results do not seem sufficient to conclude that peer pressure is indeed a major factor in participants' password strength decisions when choosing for a new password. All old CWL passwords had been created without any password strength feedback. When we looked at the difference between the entropy values of the old and new passwords, across our three conditions, we saw that providing feedback steared users toward higher entropy values, indicating that they took feedback into account during their password choice (for both EM and PPM). Looking at the mean values of the entropy difference between old and new passwords across conditions, we see that passwords in the control condition had virtually zero improvement whereas passwords in the other two cases were significantly better.

Participants, in the PPM condition, scored a mean entropy of 64.9 bits. That would yield an average feedback of being stronger than 60% of their "peers" (i.e., than the percentage of RockYou users having smaller password entropy). Moreover, during the experiment participants were observed to try different passwords so as to achieve a better score on the PPM indicator and this is reflected in the

amount of time spent in the new password text box which had a mean value of 29.56 seconds which is large compared to the 7.13 seconds spent by participants in the control condition as well as 14 seconds in the EM condition. When we investigated for significance, we found that the time required, by participants, in PPM was significantly more compared to Control but there was no significant difference between PPM and EM or EM and Control. We believe, however, that the novelty, for our participants, of the design might had added to the time needed by them to understand and familiarize themselves with it.

We would like to stress the point that this research is not about evaluating whether PPM will yield strong passwords. What a strong password is, is considered, still, an open research question. Our work, rather, is exploring whether relating perceived (by the users) password strength to this of their peers' would provide sufficient motivation to try not be "left behind". This is a reason why we did not try to employ a very sophisticated method of calculating password strength (e.g.: [50]) but we used one that might be considered naive and overoptimistic, as it does not take into account password shortcomings like usage of dictionary words. For our purposes, however, it was adequate since all we needed was a scalar value for password strength (for comparison purposes) and we consistently applied the same bit-strength calculation and feedback intervals across all three conditions.

What our results show is that, provided feedback, participants created passwords of higher entropy value when compared to the Control condition. They also show that relating a password's to that of the participants' peers, maintaing the visual cues, is not different than giving an absolute password strength attribute (i.e., characterizing a password as weak, medium or strong) as in the case of an EM implementation. This is an interesting result as it shed a new light to aspect of users' strategies and practices when creating passwords and interpreting feedback. It also demonstrates that there might be alternatives to absolutely characterizing passwords' strength, with its potential drawbacks, as it happens in the industry today.

The current state of affairs in the industry is that each vendor is deciding on what a strong password, for their systems, is and adjusts their feedback to reflect this choice. This leads to great inconsistencies among feedback provided to the user, as shown in section 2.4.3 and it might prove a source of confusions for them.

Some users are trained with some basic principles about password strength (e.g., combination of letters and digits and no everyday words) but as feedback characterizes passwords differently according to the design choices made by vendors they see the same password being rated as weak in one occasion and strong in another. Also, it might lead users to lose confidence in the EM indicators and the feedback they provide, therefore defeat the very purpose that they are supposed to serve (i.e., help users in their decisions). We argue that PPM opens a new avenue for providing feedback to the users. Instead of having vendors deciding on how to label password strength, we propose that users will be in charge of deciding what is the appropriate password strength depending on the value they place on their data on a particular site. This can be perceived as analogous to "market forces". That is, the population's value of their data on a site is reflected on the absolute level of password strength on that particular site. This contrasts with how EM is implemented today. Of course, some minimum requirements should be set by the administrators, mainly to ensure a minimum of security but this requirements could be consistent across implementations. Minimum requirements at the least should enforce a lower bound for password length so as to avoid passwords that would be extremely easy to break (i.e., one characters long passwords).

5.2 Comparison between EM and PPM

Our results, do not demonstrate PPM performing better than the existing motivators (i.e, EM) as we had hypothesized. Such a difference would be a strong indication that password strength was affected by the peer pressure feedback component of the PPM indicator. As this was not observed we need to investigate the reasons for this.

5.2.1 High lower bound of old password entropy value

CWL's minimum password requirements called for quite strong passwords to begin with. At a minimum CWL passwords should be 8 characters long and have at least on digit and one letter, yielding a minimum entropy of 41.35 bits. Therefore, when a participant had to try considerably to come up with a password that would have a significantly higher entropy. Having the minimum set so high might

have affected the ability of PPM to outperform EM. Indeed, when we had firstly conceived the PPM idea we hypothesized that users would "compete" to be better than their peers. However, we also realized that in order for that to occur, adequate entropy space should be available to them. Very high entropy values are difficult to be achieved with an easily memorable password so having the lower bound high certainly narrows the space in which PPM can prove capable of leading to better entropy values. In our case, our participants were, in a sense, trapped between high, even unrealistic, demands of high password entropy and their own ability to come up with a password that they would feel comfortable remembering. We believe that PPM will perform even better in situations that the lower entropy bound is low and users do not need to come up with overly complex or lengthy passwords in order to see a feedback which will place them among higher percentages.

5.2.2 Design and risk communication

Another possible reason for PPM not being better than EM is that participants are conditioned to EM and are used to the horizontal bar design and its absolute characterization of their password's strength. Also, participants might do not understand the feedback provided, relating their passwords to the passwords of others, rather, due to habituation [25, 42], they relied exclusively on the visual cues (i.e., the bar) for their decision making. We should mention that we made a choice to keep PPM as close as possible to the EM condition (with the bar being vertical and wider to simply accommodate for the more information we had to convey) in order to compare those two conditions in terms of peer pressure and not design choices. It might be that a better PPM implementation, regarding design, should be sought and evaluated in the future.

Furthermore, the lack of a significant difference might lie deeper than the design of the indicator. It might lie to the way risk is communicated to the participant.The way EM communicates risk is quite different than PPM. EM tries to affect user choice by communicating risk directly (i.e., the password is weak -therefor insecure-). PPM has a subtler way to communicate risk. It informs the participant that his password is weaker than a percentage of the users of the system with the purpose that it will nudge them to strive for a high percentage. Although, this

seems to be adequate, under conditions, compared to no feedback at all, it doesn't seem to affect our participants more strongly than EM.

5.3 Setting feedback intervals in PPM

During our first study we analyzed the password strength of the passwords, that complied with the CWL standards, from the Rockyou database in order to come up with the distribution of password entropy and the percentage of users corresponding to each entropy value. This, in itself, could be problematic as there was no way to be sure that the two datasets (i.e., CWL and Rockyou) followed a similar entropy distribution; When we examined the old password results from the first experiment against the Rockyou entropy distribution we did not see great differences. In that study, if a participant had an equal entropy value with a percentage of Rockyou users, they would receive feedback indicating that their password was stronger than the sum of the percentages of passwords corresponding to lower entropy values plus the percentile of passwords with entropy values equal to their own. This proved to be a mistake on our part as it enable participants to reach over 50% quite easily (with just 46.01 bits of entropy when, in total, the average old password entropy value was 52.13). Participants had little motivation to carry on and thus PPM did not perform well, not being significantly better than the control condition.

These results, in conjunction to our results from the second study, proved, however, quite interesting. They demonstrated, for the implementors of PPM, the need to follow a "pessimistic" approach when choosing feedback intervals, as a liberal one might hinder PPM's ability to serve as an adequate motivator. In our case, this is even further demonstrated by how PPM fared when we readjusted the intervals not to include the percentages of "users" having a password with entropy value equal to the current entropy value of the participant and, in addition, we shifted the values 10 bits higher making it quite difficult for participants to achieve a high percentage, unless they created quite complex, lengthy passwords; (e.g., the password "iarkto@A1" having lower and upper case letters, symbols, digits and a length of 9 characters has an entropy value of 59.13 which would rank it at 50%).

Our approach might have been over-pessimistic but it served our goal of inves-

tigating the ability and extend of PPM's motivation capabilities. In a more realistic implementation, it might not be necessary to make it so difficult for users rather keep the percentages close to the actual distribution but always keeping in mind that a "pessimistic" approach to it is recommended.

5.4 Password composition

As presented in Chapter 3 we chose to use the Shannon algorithm to calculate the bit strength of the password and use this to provide feedback to our study's participants. Equation 3.1 shows that password bit strength depends on either the password composition (i.e., how many different items will a password include - capital letters, lower letters etc.-) or on the length of the password. Previous research has shown that users mostly choose passwords that are relatively simple without non letter symbols [26]. In our data analysis, as well, in both studies we show that participants almost did not use any special characters in their passwords with the second study not having a single participants having a special character in their old CWL password. When they were asked to create a new password, very few participants used special characters. It was surprising to see out of 47 participants, in our second study, where achieving a higher entropy was considerably more difficult compared to the first study due to the changes in the feedback intervals, only 7 using a special character. Instead, participants relied on password length to create a stronger password. When feedback was employed, in our second study, participants improved their password's length by 1 to 2 characters (adding mostly alphanumeric characters, Table 4.4), whereas in the Control condition the average length remained almost the same. This leads us to believe that even after many attempts to educate users so as to diversify their passwords' composition and not use solely numbers and letters they do not seem to incorporate this advice into their password choice practices. This should be taken into consideration, even in PPM deployments as it should be taken into account when designing an algorithm that will calculate password strength for the users and their peers.

5.4.1 Choice of new password and maintenance of it for a three weeks period

In an effort to judge whether participants came up with a completely new password rather than a slight modification of their old one, we calculated the Levenshtein distance between their old and new passwords. We saw that for across all passwords in the second study, which had an average length of 10.45 characters the Levenshtein distance had an average value of 6.94. Only a 20% of participants had a Levenshtein distance of 3 or less. This indicates that most participants chose a password that bore little resemblance to their old one. However, given the fact that password's Levenshtein distance was not affected by the condition they were into, leads us to believe that the presence of feedback does not affect new password's resemblance to an old password. Rather, it seems, that the act of merely asking participants to change their password (at least for accounts considered valuable as in the case of CWL) will be sufficient in ensuring a completely different new password.

With the term maintenance we defined mostly the choice to keep the newly created password. We expected high entropy passwords to be difficult to maintain in memory and difficult to use across various web sties therefore participant would change them more often than those with lower entropy. When we investigated whether participants would stick to their new passwords or they would find them too difficult to recall or manage and changed them we saw that in both studies over 3 out 4 participants did not changed their newly created CWL passwords two to three weeks later, see Table 4.8. Furthermore, of those who changed it, not all did it out of inability to remember the new one rather out of security considerations since they changed their password in a computer other than their own. We were surprised to see that participants did not find difficult to recall from memory their new passwords even in the second study where the average entropy was higher (especially for conditions receiving feedback). However, we also noticed that almost half of our participants claimed to have reused their password in other accounts. This is particularly interesting to us. It not only confirms a known practice but also raises questions about how exactly password feedback indicators affect user choice. If users have a set of predefined passwords that they rotate among accounts than an indicator might not be the factor that plays the major part in the

creation of a password. It may be just that depending on how difficult it is for them to achieve a "strong" feedback participants use passwords they have already created until they use one that will yield the desirable feedback. After all, previous research has shown that users have, on average, less accounts than passwords, having a number of passwords that they use among sites [18]. We believe it is important to take this knowledge into account when interpreting results like our own and in future work to try and control for such issues (e.g., by asking participant to create completely new passwords instead of just a password). Since in our case we opted for maintaining as much a realistic scenario as possible we could not instruct our participants on how to create their passwords. In the future, it would be useful to somehow, preferably indirectly, point participants towards avoiding reuse of passwords (e.g., using some roleplaying instructions). Alternatively, it could prove useful to investigate deeply, by participant inquiry, on how people employ password reuse strategies and what the effect of each indicator type might be in such a case.

5.5 Security considerations in a real-world PPM deployment

In a real-world PPM deployment one consideration could be that an adversary could utilize this extra information the system gives about the relative strength of other users' passwords in the system. Thus, an online brute force attack on the system could be made easier. However, online brute force attacks, if left unchecked, are a serious threat even to systems that do not reveal any information. Knowing a percentage of users with a password entropy less of the one an adversary has tried does not help them much if a lock out mechanism is employed on the web site, as security best practices dictate. Adversaries, with even low knowledge of security research results, know already that in any given high profile web site, among thousands of user accounts, weak links are bound to exist. The reason behind the fact that web sites like Gmail have not been compromise to the best of our knowledge, due to an online brute force attack is not the lack of weak passwords among their users rather the security mechanisms that are employed complementary to the password requirements the web site sets. PPM does not call for an utter abolishment of

password requirements. It simply proposes a way to inform users about the value others place on their account in a given system thus helping them make the appropriate choice themselves. This does not mean that minimum requirements should not exist rather that they could be uniform across web sites in a similar area (e.g.: email providers, forums, social networks) and adhere to a minimum that will not hinder the vendor's business model. What these requirements could be is beyond the scope of this research.

5.6 Additional dimensions of the PPM approach

Password strength motivation based on peer pressure is multidimensional and not all of its dimensions are investigated in this research. In our present work, we are trying to investigate whether password choice can be affected by feedback bearing information about password choices of other users of the system. In addition to this aspect the proposed peer pressure paradigm has a number of other interesting implications. At present administrators try to force or guide users towards strong password even in the cases that users do not understand or agree with this. This results, many times, to users resisting the policies enforced upon them leading to insecure practices (e.g., the proverbial post-it note on the PC). In the peer pressure paradigm users are free, through a indirect consensus, to set the password level appropriate for a particular web site. This does not necessarily leads to passwords that are harder to crack. It could very well lead to the exact opposite. Users of a system that is not perceived of high data value might end up with passwords that are relatively weak. By doing so, we expect to see various levels of average password strength and a different impact of the information provided through a PPM indicator to be observed in various implementations.

Furthermore, information about the choices of peers is expected to affect different users in various ways. Unwary populations might be more susceptible to manipulation through peer pressure whereas users with high technical knowledge and strong opinions might disregard this information.

From a research standpoint, in order to investigate such implications cannot easily investigate by a study design similar to our own. We believe that the most reliable way to investigate those dimensions of PPM is to employ longitude stud-

ies in the field. Install peer pressure indicators in various sites and observe the password level throughout long period of times. This way, we expect to not only be able to address the dimensions discussed above but also identify new ways of interactions with the indicator as well as weak points that we might have not taken into account in our present analysis and assessment of our approach.

5.7 Limitations

As with every study, our design has several limitations that might have affected our results.

5.7.1 Ecological validity

Our study does not claim to have as participants a representative sample of the general population. We had, as participants, volunteers drawn from a very specific population; That of the (mostly) students of the University of British Columbia. Such a student population has been characterized as WEIRD (Western, Educated, Industrialized, Rich and Democratic) by Henrich et al. [27] arguing that since most people do not fall under this characterization we should be cautious when making generalizing arguments based on studies having such a population as their sample.

However, we made the choice of using such a population sample purposefully. Having participants from a more general population would create a new, and in our opinion greater, problem. It would be hard to ensure that participants would value the accounts that they would be asked to change the passwords for. Also, it would be extremely difficult to recruit participants from the general population, that would be using a common service which we would have the mean to tamper with, as we did with CWL. Many study designers, that aim at using the general population, are faced with this problem when passwords is the subject under investigation. We believe that by using CWL as the platform where we deployed our prototypes, we ensured that participants valued their account, since CWL is central to their university life. This is evident from our results where participants rated a potential CWL account compromise as a cause for very serious or extreme concern.

5.7.2 Strict CWL password requirements

UBC, as many not-for-profit organizations [19], has a relatively strict set of password requirements in order for the CWL passwords to be eligible. Namely, a CWL password has to have a length of 8 characters including at least 1 digit and 1 letter. In comparison, Hotmail and Yahoo have, as minimum password requirements and without taking into account restrictions about using the username as part of the password, only 6 characters length with no restrictions on the composition of the password. Google on the other hand requires its users to set a password with at least 8 characters length but other than that sets no restrictions on the password's composition, see Table 2.3. The strict password requirements set by UBC regarding CWL accounts, might have affected our study's results and the displayed efficiency of our proposed motivator. We theorize that these strict requirements have limited our prototype's ability to demonstrate its full potential.

The lower bound set drove participants to have a high entropy password to begin with and their (dis)ability to be able to remember and overly lengthy and complex password set a natural upper bound for our prototype's effectiveness to motivate them. We believe that in lack of so strict password requirements participants might have demonstrated a different behavior regarding their password choice among conditions and this would enable us to evaluate better our prototype's capabilities. That is not to disregard the significant improvement we observed of our prototype over the Control condition but to stress that we believe that it has been tested under a certain (strict) environment that does not necessarily abides the industry standards. Further testing in sites with different (i.e., less strict) password requirements might yield even better results for PPM.

5.7.3 PPM prototype design

In our effort to investigate the effect on peer pressure we deployed our prototype without testing among various designs, using user input, to pick the one that would elicit the best response from our participants. The sole drive behind our prototype's design was to clearly convey the peer pressure motivation (i.e., present other users' password strength percentages in a clear manner) and keep the design close to the concept of the basic EM indicator (i.e., a bar having different percent colored and

a written indication depending on the password strength). We could not keep the bar horizontal, as with EM, because the information presented there was more and would make the prototype too difficult to read and understand. This design choice does not seem to have affected participants' perception or favoring it over EM as we can see from the results of both experiments. However, we are not sure that this is the most effective way to present PPM to participants. In contrast to the basic EM idea and design, which has been out for quite a long time and has surely undergone user evaluation, our design has not. We perceive this as a limitation of our study as we cannot be certain about the improvement in its efficiency a thoroughly evaluated PPM prototype might had. A user evaluation of various design ideas would surely strengthen a future PPM implementation.

Chapter 6

Conclusions

In this thesis, our goal was to investigate the effect of peer pressure on the password choice of users, to evaluate whether such an approach would be efficient in a real world system and to identify potential trade offs required by it. For this purpose we designed a large scale between-subjects laboratory study that attempted to evaluate the effect of peer pressure. We compared our mechanism to a traditional motivator and the lack of any feedback, using the Campus Wide Login (CWL) system of the University of British Columbia (UBC). During our study, participants were required to change their CWL passwords receiving, or not, feedback according to the condition they were assigned to. We had two conditions that provided feedback utilizing peer pressure (PPM) and absolute characterization of password strength (EM) as well as visual cues (i.e., a strength bar) and one condition that provided no feedback and served as the Control (CC). In the case of the PPM prototype, their password strength was related to the password strength of other users of the system in an attempt to motivate participants to choose stronger, in relation to their peers, passwords. As we did not have access to the actual CWL user data, we had to rely on the Rockyou password dataset to set the password entropy intervals that would provide the feedback presented to our participants. We used a naive password strength algorithm (i.e., Shannon's password entropy) to calculate password strength and judge our participants' password strength improvement across conditions.

In the following, we provide a summary of the findings and contributions of

our research.

- **Effect of feedback on password choice**: Our results indicate that strength feedback can affect password choice. When participants were faced with our PPM prototype they tended to create passwords of higher entropy compared to password created by participants assigned to the Control condition (lacking any kind of password strength feedback) but not higher to the EM condition. The main contribution of this results is that information on relative password strength, although does not seem to motivate participants to create passwords of higher entropy compared to EM, can, in the presence of visual cues, have the same effect as characterizing the password's strength in an absolute manner. This can help users by lifting possible confusion and uncertainty of users caused by the many different password strength feedbacks that they receive (even on the same password) on different sites and that depend on each site's motivator implementation. Rather, PPM enables users of a particular system to decide themselves about the what the appropriate strength of their password should be relatively to this of other system users.

- **Design considerations when deploying PPM**: When we analyzed the results of our first experiment, we saw that PPM did not seem to affect password entropy. Upon closer inspection of the results we noticed that participants in this condition reached an above average password strength quite easily. Although, bar coverage percentage (and subsequent analogous password strength verbal characterization) was the same for EM we identified that PPM's way of motivating users was based, in part, on the desire of participants to be above average. We confirmed this when we shifted our entropy intervals 10 bits up and repeated our study. PPM then performed better than the Control condition. This realization is our second contribution, which will help people wanting to employ such a motivator decide the interval set up of their motivator's feedback. It may seem that shifting the entropy values by 10 bits violates the idea behind letting users deciding on the appropriate password strength based on the value they place on their account and the choices of their peers. However, we should keep in mind that we

did not have the accurate CWL user data in order to seed our PPM indicator in the first place and we had evidence that the average password strength in CWL was considerable higher than this of the Rockyou password dataset. Also, in our first study we had been very optimistic in setting our intervals (indicating to the participant that they had stronger password even to those users that had equally strong password to them). Finally, in our study, we wanted to investigate if the very notion of peer pressure would affect participants' password choice rather than to precisely identify every aspect of how to best implement such a motivator.

6.1 Future work

This initial evaluation of the idea of peer pressure motivators could be enhanced by future research on the subject. We believe that, at this point, two main directions for future research should be considered.

- **Isolating the effect of the various indicator aspects**: The lack of difference between the passwords' bit-strength among EM and PPM conditions gives rise to various considerations about the aspects of a password strength feedback mechanism which influence user choice. Namely, in order to evaluate whether peer pressure is indeed effective, after taking into consideration the present results, we should investigate the effect of visual cues in both EM and PPM. On the internet there are numerous sites that have implemented EM without the use of a strength bar (i.e., Facebook) and rely solely on verbal instructions (font colors are applied to some times too). Such sites might prove good platforms for a future investigation of our mechanism. The best way to control for the effect the visual cues have on user choices would be to investigate among conditions with no visual cues where the only feedback comes in the form of information about absolute and relative password strength.

  In addition, we should consider motivator designs (for both EM and PPM) that deviate from the EM strength bar standard so as to control for habituation of users to this way of representing password strength. We believe that

by isolating the various aspects of the motivators, a future design will be able to identify where likely differences in participant response lie into.

- **Design evaluation of the PPM prototype**: Investigating, utilizing user input, various designs for PPM will enable us to come up with that design that will best convey the message of PPM both more aesthetically and efficiently. Different design options could and should be evaluated helping us identify different ways of wording the motivation to users as well as investigating design constraints (e.g., constraints with setting feedback intervals) that have not investigates thoroughly in this initial assessment of the PPM idea. Moreover, novel approaches on how to convey visually the meaning of peer pressure should be sought as it is possible that text might be insufficient for this purpose.

- **Large scale evaluation on several sites**: Deploying PPM prototypes on various, high traffic, web sites with different password requirements would enable us to evaluate our hypothesis that PPM's performance is affected by the minimum password requirements (i.e., password policies) that exist in online applications. Furthermore, by implementing the prototypes on various sites, which hold data of different value to their users, we will be able to evaluate the hypothesis that password choice, overall, will converge to particular levels that will reflect the majority of the users' assessment of the value of the data held at this web sites. For example, we hypothesize that implementing PPM on a web site like Facebook and on a general interest forum will yield different user password choice behavior.

Bibliography

[1] A. Adams and M. A. Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999. ISSN 0001-0782. http://dx.doi.org/http://doi.acm.org/10.1145/322796.322806doi:http://doi.acm.org/10.1145/322796.322806.

[2] A. Adams, M. A. Sasse, and P. Lunt. Making passwords secure and usable. In *Proceedings of HCI on People and Computers XII*, HCI 97, pages 1–19, London, UK, 1997. Springer-Verlag. ISBN 3-540-76172-1. URL http://dl.acm.org/citation.cfm?id=646684.702633.

[3] I. Ayres, S. Raseman, and A. Shih. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. Working Paper 15386, National Bureau of Economic Research, September 2009. URL http://www.nber.org/papers/w15386.

[4] F. Bergadano, B. Crispo, and G. Ruffo. Proactive password checking with decision trees. In *CCS '97: Proceedings of the 4th ACM conference on Computer and communications security*, pages 67 – 77, New York, NY, USA, 1997. ACM. ISBN 0-89791-912-2. http://dx.doi.org/http://doi.acm.org/10.1145/266420.266437doi:http://doi.acm.org/10.1145/266420.266437.

[5] F. Bergadano, B. Crispo, and G. Ruffo. High dictionary compression for proactive password checking. *ACM Trans. Inf. Syst. Secur.*, 1(1):3 – 25, 1998. ISSN 1094-9224. http://dx.doi.org/http://doi.acm.org/10.1145/290163.290164doi:http://doi.acm.org/10.1145/290163.290164.

[6] A. Besmer, J. Watson, and H. R. Lipford. The impact of social navigation on privacy policy configuration. In *SOUPS '10: Proceedings of the Sixth Symposium on Usable Privacy and Security*, pages 1–10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0264-7.

http://dx.doi.org/http://doi.acm.org/10.1145/1837110.1837120doi:http://doi.acm.org/10.1145/1837110.1837120.

[7] M. Bishop and D. V. Klein. Improving system security via proactive password checking. *Computers and Security*, 14(3):233 – 249, 1995. ISSN 0167-4048. http://dx.doi.org/DOI: 10.1016/0167-4048(95)00003-Qdoi:DOI:10.1016/0167-4048(95)00003-Q. URL http://www.sciencedirect.com/science/article/pii/016740489500003Q.

[8] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas. Generating and remembering passwords. *Applied Cognitive Psychology*, 18(6):641–651, 2004. ISSN 1099-0720. http://dx.doi.org/10.1002/acp.1014doi:10.1002/acp.1014. URL http://dx.doi.org/10.1002/acp.1014.

[9] V. Bush and J. Wang. As we may think. *Atlantic Monthly*, 176:101–108, 1945.

[10] A. N. Chia Pern Hui, Heiner Andreas. The wisdom of cliques: Use of personlized social rating for trustworthy application installation. Technical report, Nokia Research Center Helsinki, Finland, 2010.

[11] T. Cover and R. King. A convergent gambling estimate of the entropy of english. *Information Theory, IEEE Transactions on*, 24(4):413 – 421, July 1978. ISSN 0018-9448. http://dx.doi.org/10.1109/TIT.1978.1055912doi:10.1109/TIT.1978.1055912.

[12] A. Dieberger, P. Dourish, K. Höök, P. Resnick, and A. Wexelblat. Social navigation: techniques for building more usable systems. *interactions*, 7: 36–45, November 2000. ISSN 1072-5520. http://dx.doi.org/http://doi.acm.org/10.1145/352580.352587doi:http://doi.acm.org/10.1145/352580.352587. URL http://doi.acm.org/10.1145/352580.352587.

[13] P. DiGioia and P. Dourish. Social navigation as a model for usable security. In *SOUPS '05*, pages 101–108, Pittsburgh, Pennsylvania, 2005. ACM. ISBN 1-59593-178-3. http://dx.doi.org/http://doi.acm.org/10.1145/1073001.1073011doi:http://doi.acm.org/10.1145/1073001.1073011.

[14] P. Dourish and M. Chalmers. Running out of space: models of information navigation. In *Short paper presented at HCI*, volume 94, 1994.

[15] P. Dourish, R. E. Grinter, J. D. de la Flor, and M. Joseph. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004. ISSN 1617-4917.

[16] D. L. Evans, P. J. Bond, and A. L. Bement. Electronic authentication guideline. Technical report, National Institute of Standards and Technology, September 2004.

[17] D. C. Feldmeier and P. R. Karn. Unix password security - ten years later. In *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '89, pages 44–63, London, UK, UK, 1990. Springer-Verlag. ISBN 3-540-97317-6. URL http://dl.acm.org/citation.cfm?id=646754.704918.

[18] D. Florencio and C. Herley. A large-scale study of web password habits. In *WWW '07: Proceedings of the 16th International Conference on World Wide Web*, pages 657–666, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-654-7. http://dx.doi.org/http://doi.acm.org/10.1145/1242572.1242661doi:http://doi.acm.org/10.1145/1242572.1242661.

[19] D. Florêncio and C. Herley. Where do security policies come from? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 10:1–10:14, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0264-7. http://dx.doi.org/http://doi.acm.org/10.1145/1837110.1837124doi:http://doi.acm.org/10.1145/1837110.1837124. URL http://doi.acm.org/10.1145/1837110.1837124.

[20] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS '08, pages 1–12, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-276-4. http://dx.doi.org/http://doi.acm.org/10.1145/1408664.1408666doi:http://doi.acm.org/10.1145/1408664.1408666. URL http://doi.acm.org/10.1145/1408664.1408666.

[21] S. Furnell. An assessment of website password practices. *Computers and Security*, 26(7-8):445 – 451, 2007. ISSN 0167-4048. http://dx.doi.org/DOI:10.1016/j.cose.2007.09.001doi:DOI:10.1016/j.cose.2007.09.001. URL

http://www.sciencedirect.com/science/article/B6V8G-4PPF629-1/2/306f75b35d2dfabd6fa1e67a3e340c49.

[22] J. Goecks and E. Mynatt. Supporting privacy management via community experience and expertise. In P. Besselaar, G. Michelis, J. Preece, and C. Simone, editors, *Communities and Technologies 2005*, pages 397–417. Springer Netherlands, 2005.

[23] J. Goecks, W. K. Edwards, and E. D. Mynatt. Challenges in supporting end-user privacy and security management with social navigation. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 5:1–5:12, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-736-3. http://dx.doi.org/http://doi.acm.org/10.1145/1572532.1572539doi:http://doi.acm.org/10.1145/1572532.1572539. URL http://doi.acm.org/10.1145/1572532.1572539.

[24] J. B. Gross and M. B. Rosson. Looking for trouble: understanding end-user security management. In *Proceedings of the 2007 symposium on Computer human interaction for the management of information technology*, CHIMIT '07, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-635-6. http://dx.doi.org/http://doi.acm.org/10.1145/1234772.1234786doi:http://doi.acm.org/10.1145/1234772.1234786. URL http://doi.acm.org/10.1145/1234772.1234786.

[25] P. M. Groves and R. F. Thompson. Habituation: A dual-process theory. *Psychological Review*, 77(5):419 – 450, 1970. ISSN 0033-295X. http://dx.doi.org/10.1037/h0029810doi:10.1037/h0029810. URL http://www.sciencedirect.com/science/article/pii/S0033295X07622166.

[26] K. Harada, Y. Kuroki. A study on the attitude and behavior of computer network users regarding security administration. *REPORTS- NATIONAL RESEARCH INSTITUTE OF POLICE SCIENCE RESEARCH ON PREVENTION OF CRIME AND DELINQUENCY REPORTS*, 1996.

[27] J. Henrich, S. Heine, and A. Norenzayan. Most people are not weird. *Nature*, (466):29, 2010.

[28] C. Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *NSPW '09: Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, pages 133–144, New York, NY, USA, 2009. ACM.

[29] W. C. Hill, J. D. Hollan, D. Wroblewski, and T. McCandless. Edit wear and read wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92, pages 3–9, New York, NY, USA, 1992. ACM. ISBN 0-89791-513-5. http://dx.doi.org/http://doi.acm.org/10.1145/142750.142751doi:http://doi.acm.org/10.1145/142750.142751. URL http://doi.acm.org/10.1145/142750.142751.

[30] K. Höök. *Designing Information Spaces: The Social Navigation Approach*. Springer, Jan. 2003. ISBN 1852336617. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/1852336617.

[31] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 383–392, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. http://dx.doi.org/http://doi.acm.org/10.1145/1753326.1753384doi:http://doi.acm.org/10.1145/1753326.1753384. URL http://doi.acm.org/10.1145/1753326.1753384.

[32] E. Kandel and E. Lazear. Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–17, 1992. URL http://econpapers.repec.org/RePEc:ucp:jpolec:v:100:y:1992:i:4:p:801-17.

[33] D. Klein. 'foiling the cracker': a survey of, and improvements to, password security. In *USENIX Workshop Proceedings. UNIX Security II, 27-28 Aug. 1990*, USENIX Workshop Proceedings. UNIX Security II, pages 5–14, Portland, OR, USA, 1990. USENIX Assoc.

[34] D. V. Klein. "foiling the cracker": A survey of, and improvements to, password security, 1990.

[35] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2595–2604, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. http://dx.doi.org/http://doi.acm.org/10.1145/1978942.1979321doi:http://doi.acm.org/10.1145/1978942.1979321. URL http://doi.acm.org/10.1145/1978942.1979321.

[36] C. Kuo, S. Romanosky, and L. F. Cranor. Human selection of mnemonic phrase-based passwords. In *Proceedings of the second symposium on*

Usable privacy and security, SOUPS '06, pages 67–78, New York, NY, USA, 2006. ACM. ISBN 1-59593-448-0. http://dx.doi.org/http://doi.acm.org/10.1145/1143120.1143129doi:http://doi.acm.org/10.1145/1143120.1143129. URL http://doi.acm.org/10.1145/1143120.1143129.

[37] S. M. Difining and designing social navigation, 2000.

[38] R. Morris and K. Thompson. Password security: a case history. *Commun. ACM*, 22(11):594–597, 1979. ISSN 0001-0782. http://dx.doi.org/http://doi.acm.org/10.1145/359168.359172doi:http://doi.acm.org/10.1145/359168.359172.

[39] N. N. I. of Standards and Technology. http://www.nist.gov/.

[40] PhishTank. Phishtank: Phising on the internet. http://www.phishtank.com, July 2009.

[41] H. R. Doing a number on memory. APS Observer, 2001.

[42] C. H. Rankin, T. Abrams, R. J. Barry, S. Bhatnagar, D. F. Clayton, J. Colombo, G. Coppola, M. A. Geyer, D. L. Glanzman, S. Marsland, F. K. McSweeney, D. A. Wilson, C.-F. Wu, and R. F. Thompson. Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, 92 (2):135 – 138, 2009. ISSN 1074-7427. http://dx.doi.org/10.1016/j.nlm.2008.09.012doi:10.1016/j.nlm.2008.09.012. URL http://www.sciencedirect.com/science/article/pii/S1074742708001792. ¡ce:title¿Special Issue: Neurobiology of Habituation¡/ce:title¿.

[43] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the weakest link a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19:122–131, 2001. ISSN 1358-3948. URL http://dx.doi.org/10.1023/A:1011902718709. 10.1023/A:1011902718709.

[44] S. Schechter, C. Herley, and M. Mitzenmacher. Popularity is everything: a new approach to protecting passwords from statistical-guessing attacks. In *Proceedings of the 5th USENIX conference on Hot topics in security*, HotSec'10, pages 1–8, Berkeley, CA, USA, 2010. USENIX Association. URL http://dl.acm.org/citation.cfm?id=1924931.1924935.

[45] B. Schneier. *Secrets and lies*. Wiley, Indianapolis, Ind., 2000. ISBN 0471253111.

[46] B. Schneier. The psychology of security. In *Proceedings of the Cryptology in Africa 1st international conference on Progress in cryptology*, AFRICACRYPT'08, pages 50–79, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-68159-0, 978-3-540-68159-5. URL http://dl.acm.org/citation.cfm?id=1788634.1788642.

[47] C. Shannon. Prediction and entropy of printed english. *Bell Systems Technical Journal*, 1(30):50–64, 1951.

[48] R. Shay and E. Bertino. A comprehensive simulation tool for the analysis of password policies. *International Journal of Information Security*, 8: 275–289, 2009. ISSN 1615-5262. URL http://dx.doi.org/10.1007/s10207-009-0084-3. 10.1007/s10207-009-0084-3.

[49] R. Shay, A. Bhargav-Spantzel, and E. Bertino. Password policy simulation and analysis. In *Proceedings of the 2007 ACM workshop on Digital identity management*, DIM '07, pages 1–10, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-889-3. http://dx.doi.org/http://doi.acm.org/10.1145/1314403.1314405doi:http://doi.acm.org/10.1145/1314403.1314405. URL http://doi.acm.org/10.1145/1314403.1314405.

[50] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 2:1–2:20, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0264-7. http://dx.doi.org/http://doi.acm.org/10.1145/1837110.1837113doi:http://doi.acm.org/10.1145/1837110.1837113. URL http://doi.acm.org/10.1145/1837110.1837113.

[51] W. C. Summers and E. Bosworth. Password policy: the good, the bad, and the ugly. In *Proceedings of the winter international synposium on Information and communication technologies*, WISICT '04, pages 1–6. Trinity College Dublin, 2004. URL http://dl.acm.org/citation.cfm?id=984720.984724.

[52] X. Suo, Y. Zhu, and G. S. Owen. Graphical passwords: A survey. *Computer Security Applications Conference, Annual*, 0:463–472, 2005. ISSN 1063-9527. http://dx.doi.org/http://doi.ieeecomputersociety.org/10.1109/CSAC.2005.27doi:http://doi.ieeecomputersociety.org/10.1109/CSAC.2005.27.

[53] M. Svensson, K. Höök, J. Laaksolahti, and A. Waern. Social navigation of food recipes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, pages 341–348, New York, NY, USA, 2001. ACM. ISBN 1-58113-327-8. http://dx.doi.org/http://doi.acm.org/10.1145/365024.365130doi:http://doi.acm.org/10.1145/365024.365130. URL http://doi.acm.org/10.1145/365024.365130.

[54] D. B. Terry. A tour through tapestry. In *Proceedings of the conference on Organizational computing systems*, COCS '93, pages 21–30, New York, NY, USA, 1993. ACM. ISBN 0-89791-627-1. http://dx.doi.org/http://doi.acm.org/10.1145/168555.168558doi:http://doi.acm.org/10.1145/168555.168558. URL http://doi.acm.org/10.1145/168555.168558.

[55] J. the Ripper. A common passwords list by openwall project, http://www.openwall.com/passwords/wordlists/password.lst., 2011. Accessed on 10th Jan 2011.

[56] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. E. Schultz. Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, 65(8):744 – 757, 2007. ISSN 1071-5819. http://dx.doi.org/10.1016/j.ijhcs.2007.03.007doi:10.1016/j.ijhcs.2007.03.007. URL http://www.sciencedirect.com/science/article/pii/S1071581907000560.

[57] M. Weiser. The computer for the 21st century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3:3–11, July 1999. ISSN 1559-1662. http://dx.doi.org/http://doi.acm.org/10.1145/329124.329126doi:http://doi.acm.org/10.1145/329124.329126. URL http://doi.acm.org/10.1145/329124.329126.

[58] M. Weiser and J. S. Brown. Designing calm technology. *POWERGRID JOURNAL*, 1, 1996.

[59] A. Whitten and J. D. Tygar. Why johnny can't encrypt. In *In Proceedings of the 8th USENIX Security Symposium*, 1999.

[60] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: empirical results. *Security Privacy, IEEE*, 2(5):25 – 31, sep. 2004. ISSN 1540-7993. http://dx.doi.org/10.1109/MSP.2004.81doi:10.1109/MSP.2004.81.

[61] J. J. Yan. A note on proactive password checking. In *Proceedings of the 2001 workshop on New security paradigms*, NSPW '01, pages 127–135, New York, NY, USA, 2001. ACM. ISBN 1-58113-457-6. http://dx.doi.org/http://doi.acm.org/10.1145/508171.508194doi:http://doi.acm.org/10.1145/508171.508194. URL http://doi.acm.org/10.1145/508171.508194.

Appendix A

Existing password meters

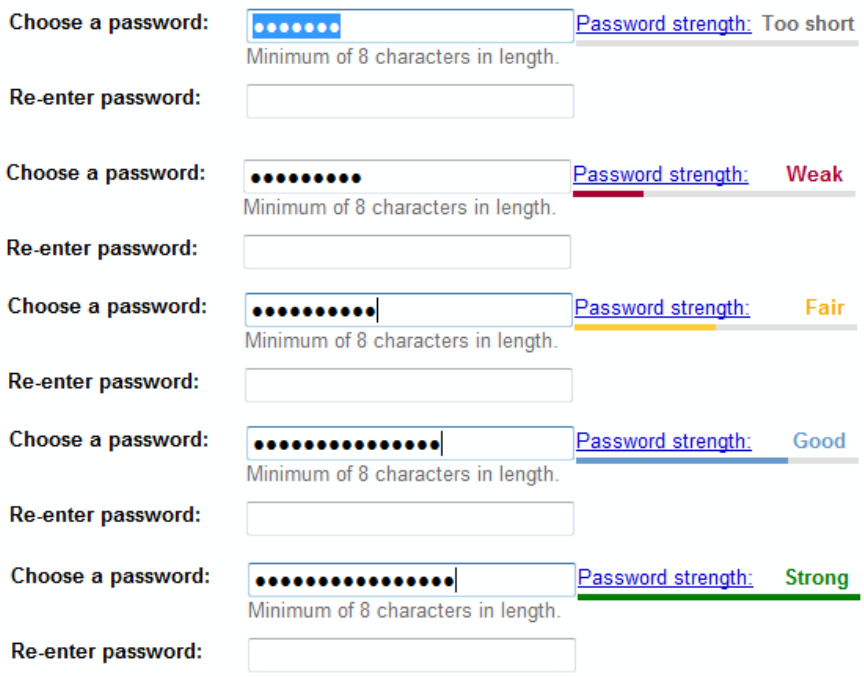

Figure A.1: GMail password meter

Figure A.2: Facebook password meter

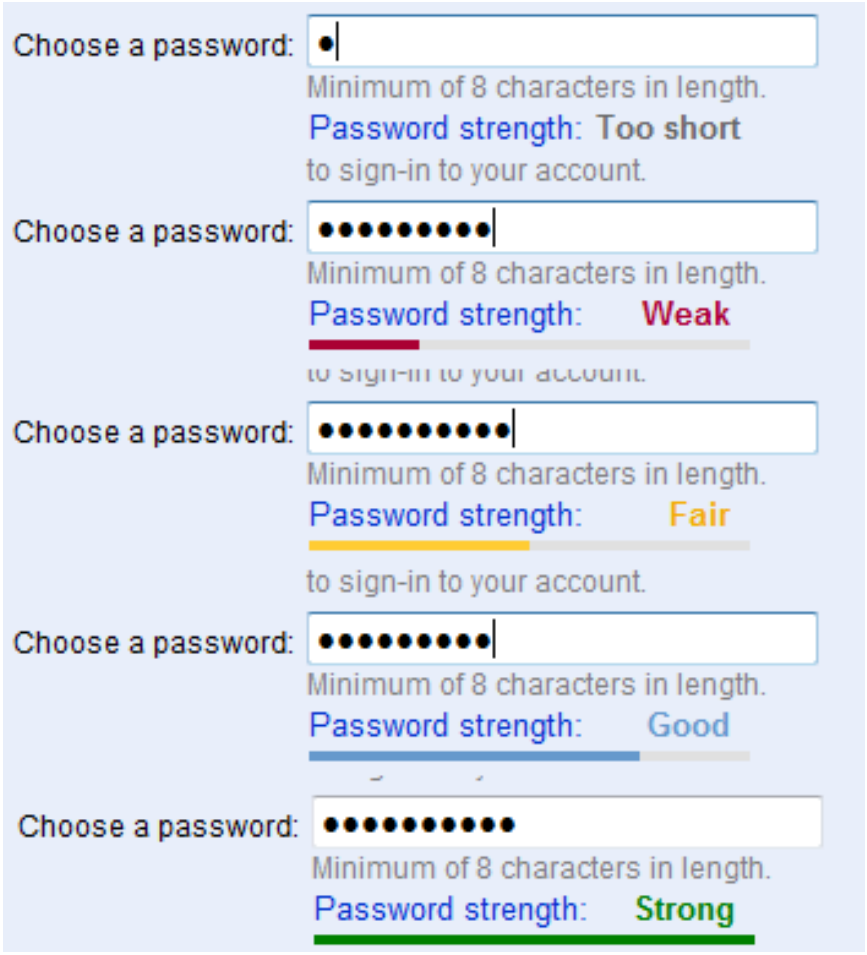

Figure A.3: YouTube password meter

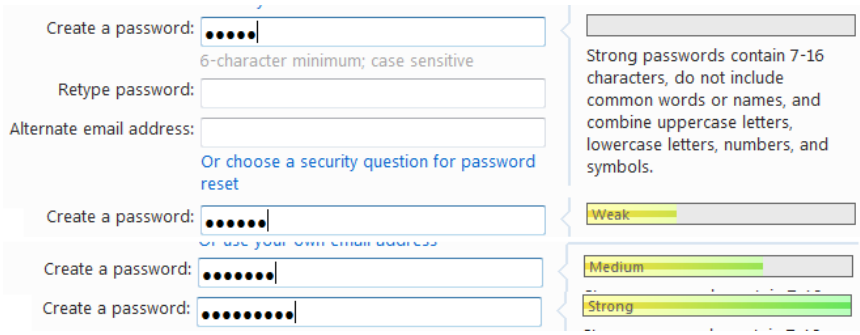

Figure A.4: MSN Live password meter

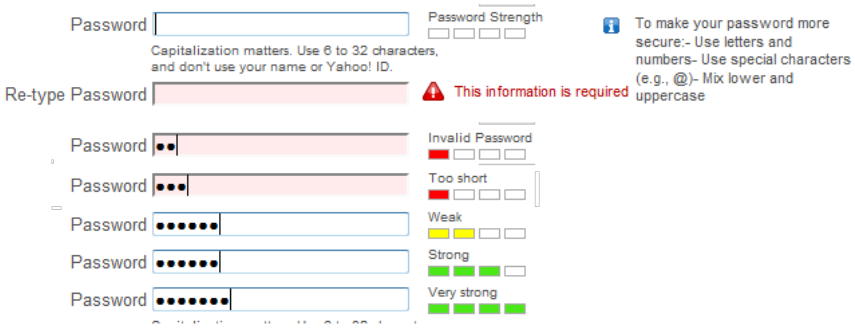

Figure A.5: Yahoo password meter

Appendix B

Study materials

B.1 Demographics and computer expertise survey

Demographics Questionnaire

1.

*1. Username

2. What is your gender?

- Male
- Female

3. What is your age?

- under 18
- 19-24
- 25-30
- 31-36
- 36-45
- 46-55
- 56-65
- over 65

4. What is your highest level of completed education?

- Highschool
- University
- Graduate School
- Proffessional School

Other (please specify)

5. What is your UBC affiliation?

- Prospective Student
- Student
- Faculty
- Staff

Other (please specify)

Demographics Questionnaire

6. If you are a student, are you

- ○ Undergraduate
- ○ Masters
- ○ PhD

Other (please specify)

7. What is your department?

2.

10. What do you use your computer for? (check all that apply)

- [ ] Multi Media (Music, Video, Photo)
- [ ] Games
- [ ] Word Processing
- [ ] Spread Sheets
- [ ] Presentation
- [ ] Web Surfing
- [ ] Email
- [ ] Instant Message
- [ ] Online Shopping
- [ ] Online Gaming
- [ ] Online Education
- [ ] Pay Bills
- [ ] Banking
- [ ] Research
- [ ] Programming
- [ ] Database Applications

Other (please specify)

11. How do you assess your technical knowledge of computers?

- ○ Advanced at operating system level
- ○ Basic Knowledge at operating system level
- ○ Advanced user of basic programs (web browsers, email, etc.)
- ○ Regular user of basic programs (web browsers, email, etc.)
- ○ Sporadic user of basic programs (web browsers, email, etc.)

12. Please indicate how difficult you find each of the following tasks to be.

	Very Easy	Somewhat Easy	Neutral	Somewhat Difficult	Very Difficult	Never Done
Copying and moving files between directories	○	○	○	○	○	○
Finding required information of services in the Internet	○	○	○	○	○	○
Loading (installing) new software onto a computer	○	○	○	○	○	○
Installing a device driver	○	○	○	○	○	○
Installing operating system	○	○	○	○	○	○
Computer programming	○	○	○	○	○	○
Administering a computer network server	○	○	○	○	○	○
Setting up a wireless network at home	○	○	○	○	○	○

13. How do you troubleshoot your computer?

	Never	Sometimes	Often	Always
I troubleshoot my computer myself.	○	○	○	○
I use system Help or online resources.	○	○	○	○
I get help from other people.	○	○	○	○

Demographics Questionnaire

3. myUBC portal usability

1. I use the myUBC portal

- More than 7 times a week
- About 4 times a week
- About 2 times a week
- Less than 2 times a week
- I never use the myUBC portal

2. I have used the myUBC portal to post an ad or to find an ad in the classified section.

- True
- False

3. I find the MyUBC portal a useful resource for retrieving information.

- True
- False

4. The myUBC portal is a valuable resource of information for my work in UBC.

	Strongly disagree	Disagree	Neutral	Agree	Strong agree	N/A
valuable resource of information						

5. The news section of the portal is an important source for news about UBC for me.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	N/A
News section importance						

6. The organization of the web site is clear and easy to follow.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	N/A
Clear organization of the site						

B.2 Follow-up survey; No password change

1.

***1. Email**

2. You have not made changes to your CWL account

It seems that you have not changed your CWL password since you were last in the lab with us.

*2. How often have you used your CWL in UBC sites since you were last in the lab with us.

	Have not used it	Less than once per week	Once a week	Several times a week	Once a day	Several times a day
Frequency of CWL usage in UBC sites	○	○	○	○	○	○

*3. How many times, besides the one time in the first session, within a year do you, normally, change your CWL password?

- ○ I never change my CWL password
- ○ 1
- ○ 2-3
- ○ More than 3

*4. When I chose a password for my CWL account I was concerned with my account's security

	Not Concerned at all				Extremely concerned
Concerned with account's security	○	○	○	○	○

*5. Do you find your current CWL password compared to your old CWL password

- ○ Weaker
- ○ Equally strong
- ○ Stronger

Other (please specify)

*6. Please rate the effort to remember your current CWL password

	Very easy to remember	Easy to remember	Neutral	Hard to remember	Very hard to remember
Effort to remember CWL password	○	○	○	○	○

*7. What is your practice for remembering your current CWL password?

- ☐ I recall it from memory
- ☐ I have written it down (on paper or on a computer file)
- ☐ I use a password manager

Other (please specify)

*8. How concerned would you be if one of your following accounts/password had been stolen?

	Not Concerned at all	A little concerned	Neutral	Very concerned	Extremely concerned	Not applicable
CWL account	○	○	○	○	○	○
Bank account	○	○	○	○	○	○
Main Email account	○	○	○	○	○	○
Facebook account	○	○	○	○	○	○
Forum I am subscribed to	○	○	○	○	○	○
Messenger account	○	○	○	○	○	○

9. Have you used the CWL password in other accounts?

- ○ Yes
- ○ No

Other (please specify)

*10. What, in your opinion, constitutes a good password?

3.

Dear Participant,

Thank you once for more taking part in our study. Our study's real purpose was to investigate users' practices in creating CWL passwords under different scenarios. We could not reveal this earlier as we did not want to bias your CWL password choice.

During the study you were asked to change your CWL password and, depending on the condition you were into, you were given some indicators about its strength. We then collected data that will help us assess the ability of those indicators to guide users in creating better passwords.

A good password is one that is hard to guess, in the event of an attack on the system, but will also be easy for the owner to retain in memory without having to write it down or put considerable effort into remembering it. Writing a password down is considered an insecure practice and having to put too much effort into remembering different passwords, has been shown to make users end up either writing it down or change the password for a less complex, easier to remember (and guess in an event of an attack) or more common password.

It is important, for your security and for the security of the systems you use, to choose strong and convenient to remember passwords.

Your password is safe. We have stored it in an one-way encrypted form and no one (including us) has access to the password's text.

Again, thank you for participating in our study. We would be happy to answer any questions you might have. If you know, personally, other participants that have not completed the second part of the study, please do not reveal its true purpose as this might affect their responses. Please send an email at [redacted]

Regards,
PPM Team.

B.3 Portal user experience survey

1.

***1. Email**

2. You have made changes to your CWL account

It seems that you have changed your CWL password since you were last in the lab with us.

*2. How often have you used your CWL in UBC sites since you were last in the lab with us.

	Have not used it	Less than once per week	Once a week	Several times a week	Once a day	Several times a day
Frequency of CWL usage in UBC sites	○	○	○	○	○	○

*3. How many times, besides the one time in the first session, within a year do you, normally, change your CWL password?

- ○ I never change my CWL password
- ○ 1
- ○ 2-3
- ○ More than 3

*4. Why did you decide to change your password since the last time you were in the lab with us?

5. Please rate the significance of the following statements in your decision to change your password

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Not Applicable
I found it hard to remember it	○	○	○	○	○	○
I change my password often as a practice	○	○	○	○	○	○
I felt that my password has been stolen	○	○	○	○	○	○
Other reason	○	○	○	○	○	○

Other (please specify)

6. When you choose a password for my CWL account I was concerned with my account's security

	Not concerned at all	A little concerned	neutral	Very concerned	Extremely concerned
Concerned with account's security	○	○	○	○	○

*7. How do you find your current CWL password compared to your old CWL password

- Weaker
- Equally strong
- Stronger

Other (please specify)

*8. Please rate the effort to remember your current CWL password

	Very easy to remember	Easy to remember	Neutral	Hard to remember	Very hard to remember
Effort to remember CWL password					

*9. What is your practice for remembering your current CWL password?

- I recall it from memory
- I have written it down (on paper or on a computer file)
- I use a password manager

Other (please specify)

*10. How concerned would you be if one of your following accounts/password had been stolen?

	Not Concerned at all	A little concerned	Neutral	Very concerned	Extremely concerned	N/A
CWL account						
Bank account						
Main Email account						
Facebook account						
Forum I am subscribed to						
Messenger account						

11. Have you used the CWL password in other accounts?

- Yes
- No

Other (please specify)

*12. What, in your opinion, constitutes a good password?

3.

Dear Participant,

Thank you once more for taking part in our study. Our study's real purpose was to investigate users' practices in creating CWL passwords under different scenarios. We could not reveal this earlier as we did not want to bias your CWL password choice.

During the study you were asked to change your CWL password and, depending on the condition you were into, you were given some indicators about its strength. We then collected data that will help us assess the ability of those indicators to guide users in creating better passwords.

A good password is one that is hard to guess, in the event of an attack on the system, but will also be easy for the owner to retain in memory without having to write it down or put considerable effort into remembering it. Writing a password down is considered an insecure practice and having to put too much effort into remembering different passwords, has been shown to make users end up either writing it down or change the password for a less complex, easier to remember (and guess in an event of an attack) or more common password.

It is important, for your security and for the security of the systems you use, to choose strong and convenient to remember passwords.

Your password is safe. We have stored it in an one-way encrypted form and no one (including us) has access to the password's text.

Again, thank you for participating in our study. We would be happy to answer any questions you might have. If you know, personally, other participants that have not completed the second part of the study, please do not reveal its true purpose as this might affect their responses. Please send an email at [redacted]

Regards,
PPM Team.

B.4 Follow-up survey; password change

MyUBC Portal User Experience

User Experience Evaluation

Please provide your feedback regarding your experience using the MyUBC portal.

1. Overall I found using the MyUBC portal

	Very easy to use				Very hard to use
Found using the portal	○	○	○	○	○

2. Overall I enjoyed the interaction with the portal's interface

	Not at all				Very much
Enjoyment	○	○	○	○	○

3. The tasks I was asked to perform were

	Very easy				Very hard
difficulty of tasks	○	○	○	○	○

4. Are there any additional services you would like the MyUBC portal to offer? What are they?

5. Would you use MyUBC portal more often if certain aspects of it were improved? If yes, what aspects would you like to see improved?

B.5 Recruitment of participants

The MyUBC Portal is Changing

UBC's **my.ubc.ca** portal is changing to better serve students, faculty and staff. We need users' feedback in order to identify usability issues with the current design as well as suggestions that will enhance the portal's usability.

All participants will receive **$45 for their participation**.

We require volunteers to participate in two sessions for **40 minutes** in total.

During this session you will be asked to complete a series of tasks and answer short online surveys that will help us draw valuable conclusions. **The only requirement to participate is to hold a valid CWL account in order to be able to use the MyUBC portal. No previous experience with the my.ubc.ca portal is required to participate in the site's evaluation.**

If you would like to participate in this study, please contact us at … **(**or call …).