Effects of Peer Feedback on Password Strength

Marc Dupuis
Computing & Software Systems
University of Washington Bothell
Bothell, Washington 98011 U.S.A.
marcjd@uw.edu

Faisal Khan
Computing & Software Systems
University of Washington Bothell
Bothell, Washington 98011 U.S.A.
faisalk@uw.edu

***Abstract*—Passwords have dominated the world of authentication. Their widespread use has made them a prized target for attackers. Various schemes have been employed to strengthen password security to resist such attacks. Numerous websites and applications use password meters to help users create a stronger password. The objective of having a password meter is to provide visual feedback to users on their choice of a password by labeling it as weak, medium, or strong, for example. In this paper, we incorporated social influence, which is the effect others have on an individual's attitude and behavior. This social influence, commonly known as peer feedback, was incorporated in the design of a peer feedback password meter. When participants were given explicit instructions to create a unique password, those that were provided with the peer feedback meter created stronger passwords when compared to those that had the traditional meter.**

Keywords—passwords, peer feedback, password feedback mechanisms, password strength meters

I. INTRODUCTION

According to some estimates, online users are expected to reach 4 billion by 2019, generating around 44 zettabytes of data by 2020 [1]. All of this data is valuable for the individual, the institution, and any malicious actor. Recent data breaches suffered by Yahoo!, Dropbox, and LinkedIn had exposed up to 732 million user details [2]. The exposure of personally identifiable information (PII) includes information such as email addresses, passwords, secret questions, and their answers. It has compromised the confidentiality and integrity of information and risked the privacy and security of the user. Institutions suffering these attacks face economic loss as well as loss of reputation. Some studies suggest that in 2015 cyber-crime victims spent around $126 billion to deal with the fallout of an attack [3]. Around 62% of the data breaches that occurred in 2016 were as a result of hacking, and out of those a staggering 81% leveraged the use of either prior stolen information or weak passwords [4].

Passwords have dominated the world of authentication and remain an important component of individuals protecting themselves from various threats [5]. Their widespread use has made them a prized target for attackers. If the attackers can get hold of a password, they can essentially compromise the security of an entire system. In one study, it was found that "123456" was the most popular password followed by "12345" [6]. Thus, it is not surprising that out of 500 compromises in 15 countries, 28% of those were as a result of weak passwords [7]. The compromised financial accounts (e.g., PayPal) went on to be sold in the black market for upwards of $60 per account [8]. Yet, other research revealed that around 15% of the passwords were either a partner's or child's name, followed by the names of football teams and pets [9]. In one survey it was found that 26.2% of the respondents have on average more than 50 accounts. Additionally, 50.9% of the respondents stated that they often reuse their passwords because they are easier to remember [10]. If one account gets compromised, this could potentially create a chain reaction of other associated accounts of that individual to be compromised as well. This includes accounts one may maintain with their employers, which increases the threat of a non-malicious insider [11].

A weak password is defined as phrases that are trivial to guess or those likely to be cracked by using a dictionary attack [12]. Researchers identified passwords as being the weakest link in the security chain at least as early as 1979 [13]. Many reasons have been speculated as to why a user may choose weak passwords to begin with. The reason comes down to human psychology. People prefer convenience, speed, and memorability of a weak password than the complexity of a strong password [14].

Kevin Mitnick, arguably the most famous hacker, pointed out that [8]:

The human side of computer security is easily exploited and constantly overlooked. Companies spend millions of dollars on firewalls, encryption, and secure access devices, and its money wasted because none of these measures address the weakest link in the security chain.

Mitnick's statement essentially means that no matter how much money and effort is spent building security around a system to prevent attacks, as long as humans – the weakest link in this security chain – continue to create hackable passwords, computer systems will remain vulnerable.

New and improved ways of authentication have been introduced in the industry, such as two-factor authentication and biometrics, but the use of passwords still dominates. In many cases, users are forced to choose stronger passwords in accordance with strict password policies. However, such policies tend to isolate a user and do not improve password quality and may, in turn, lead to the repetitive use of a single password across websites and services [15]. In such scenarios,

978-1-5386-4922-0/18/$31.00 ©2018 IEEE

a user might tend to use mechanisms (e.g. writing passwords down on paper) which might be even more detrimental to security than the use of weak passwords [16].

Various schemes have been employed to strengthen password security to resist such attacks. Numerous websites and applications use password meters to help users create a stronger password. As shown in Fig. 1, a password meter is a graphical representation that informs the user how secure the password is on some scale.

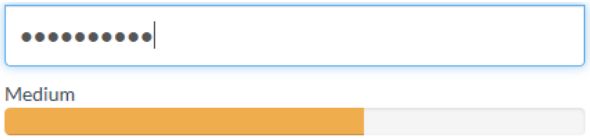

Fig. 1. Graphical password meter

The objective of having a password meter is to provide visual feedback to the user on their choice of a password by labeling it weak, medium, or strong, for example. The criteria for calculating the strength of a password is set by a developer or an institution, such as checking for minimum length, use of dictionary words, and use of special characters. One drawback of this approach is that there is a multitude of varying implementations of password meters, each having its unique set of password rules, which can yield a different result for the same password phrase. The differing results can cause confusion and even distrust by the user. When conflicting feedback for the same password occurs, it decreases a user's trust and willingness to comply with the system [17].

The evidence suggests the presence of password meters for password change and creation might lead to a more secure password [18]. While implementing such meters may encourage users to create longer and hence more secure passwords, users tend to dislike having to follow strict guidelines for choosing such passwords [19]. Another drawback is that it is hard to remember such passwords since password meters do not necessarily increase their memorability.

In this paper, we incorporated social influence, which is the effect others have on an individual's attitude and behavior [20]. This social influence, commonly known as peer feedback, was incorporated in the design of a password meter which we are calling a peer feedback meter. Our approach is an application of social navigation which states that users navigate through a system based on the inherent design of the system. This method also allows users to understand and interpret the actions of others as well [21]. The modified password meter was used to assess whether it resulted in a stronger password than a traditional password meter, and if so, to what extent.

The remainder of the paper is organized into the following sections. Section two provides background by examining related work on text-based passwords, password meters, and details on how a user navigates through social influences. Section three provides the design and methodology used in our study to investigate the effect of peer feedback on the strength of the password. Section four presents the results of our study. Finally, section five provides a summary of the results, the conclusion of our study, its limitations, and the implications of our work.

II. Background

Typically, user authentication falls into one of three categories: 1) "what you know" (e.g. text-based passwords or PINs); 2) "what you have" (e.g. a token or a smart card), and 3) "who you are" (e.g. biometrics) [22]. In this paper, we are going to focus on "what you know," specifically text-based passwords and how we can improve their effectiveness in security.

A. Passwords and Password Meters

More than 35 years ago, Morris and Thompson identified text-based passwords as the weak point in a system's security [13]. In the study, they found that approximately 86% of the users had weak passwords characterized either as short, all lower or uppercase, or could easily be found in the dictionary. Despite their efforts to raise awareness, contemporary password and password meter designs have not evolved at the same pace over the decades as they would have expected and still to this day remain one of the weakest links.

Researchers and industry experts understood that because of poor password choices by the user, the inherent security risks associated with it materialized [23]. Because of this, many researchers proposed that creating good password policies would help increase security [24]. However, later research established that these policies are often misunderstood and even become cumbersome for users, thereby defeating the purpose of having passwords. As a result, users default to insecure practices, such as writing their passwords down on post-it notes and attaching it to their workstations. These are the very practices that password policies hoped to mitigate [25].

Beyond password policies, password strength meters have been used as a visual representation of how secure a password is against the rules set by the developer. As shown in Fig. 1, they advise or force the user to choose a complex password if the password they chose is considered weak. This method helps to prevent the user from choosing easily guessable passwords and may even enforce the password policies. These password meters are also known as proactive password checkers and have been in use for many years [23]. Instead of relying on a user to create a robust and secure password, they continuously check the typed phrase against the set of rules and provide feedback on it (e.g. if the password is too short or is missing special characters). The evidence suggests that implementing password meters assists in creating stronger passwords [18].

1) Inner Workings of a Password Strength Meter

The goal of every password meter is to check if an attacker can easily crack a given password and provide this feedback to the end user. To achieve this, password meters employ the calculation of bit-space or entropy based on the length and the character set (e.g. lowercase, uppercase, numbers, symbols) used in the password. These password checkers often penalize a user for using a dictionary word since a dictionary attack can be used to crack such passwords easily. National Institute of Standards and Technology (NIST) recommend that passwords have a minimum of one lowercase, one uppercase, one number, and one special character with a minimum length of eight characters. Additionally, they should not be a permutation of the username and should not contain phrases from a language dictionary [26].

However, there is no formalized standard across the industry which defines how much entropy is considered weak, strong, or very strong when it comes to password meters. In fact, in a large-scale comparative analysis of various password meters used on popular websites (e.g. Microsoft, Google, Dropbox, Apple, Yahoo!, Skype, Twitter, PayPal), researchers found that these platforms do not explain the logic behind the strength assignment of the password and give varying strength scores for the same password [15].

This is where our work differentiates from previous work. In our study, administrators do not set the values of how strong a given password is; instead, it is left to the potential user of the system. Brown et al. support our view that users should be able to decide which application requires stringent password security (e.g. account with financial and banking information) as opposed to accounts like forums that may not require the same level of password strength [27].

B. Psychology and Social Aspects

Theoretically, using password meters can be used to block all possible weak passwords, but in reality, it is impractical as these passwords are often difficult to memorize [28]. Users tend to avert stringent rules [29], and as a result, indulge in the same insecure practices we mentioned earlier. There is always a trade-off to be made between security (i.e. using strong passwords) and usability (i.e. easy to remember passwords) [30]. This security tradeoff creates a problem for IT professionals. Users are usually aware of what a good password is, but despite this understanding, they are still inclined to take risks and are optimistically biased; they believe a negative event is unlikely to happen to them [31]. In a study of 20,907 people, researchers found that around 76% of the respondents knew that they should protect their information online yet they engage in sharing their passwords and other risky behaviors [3]. When their understanding is lacking, it may be due to them overestimating the benefits of a few predictable practices, such as adding digits or phrases, which inadvertently weakens their password [32].

In this study, we incorporate the recommendations of earlier research that password meters should be targeted towards users and be data driven. We provide feedback based on an individual's password proactively, which is unique to that specific individual. Through this study, we seek to develop a different type of password meter in hopes to achieve a better balance between the usability and security trade-off.

Incorporating social influences into cyber security is not new. In one study, researchers found that users shared their information depending on how their peers made decisions [33]. In another study, the researchers showed how the influence of framing impacts user tolerance on security delays [34]. In cognitive psychology, the notion of framing effect states that how a risk is framed can influence people's actions [35]. By manipulating the way information is presented, we can influence and alter decisions and judgments about that information. This theory suggests that people tend to avoid risks if a positive frame is offered and seek risks when a negative frame is presented. The idea of social influence has been adopted in other computing fields apart from security. For instance, YouTube and Netflix recommend content to users based on what other users have liked. Our design of the system reflects such social influence attributes. The feedback the password meter gives is based upon peer feedback and how the feedback is expressed is influenced by the framing effect.

Considering this past research, the following hypothesis was developed. In this study, we wanted to explore if users can be encouraged to generate stronger passwords if they are given feedback on how their passwords compare with their peers, which can be referred to as subjective norm. Subjective norm is defined as an individual's perception of what people in society think about their behavior [36]. This comparison with their peers includes the strength of passwords used by individuals in their university community. We are interested in finding out if the peer feedback password meter results in stronger passwords compared to a traditional password meter. Based on the literature, we believe it does.

H1: Using peer feedback password meters increases the password strength as compared to using the traditional password meter.

III. Research Methods and Design

For testing of the hypothesis, we designed a single blind experiment where we masked the true goals of this study so that the participants in this study would not consciously be biased when they saw both password meters. Thus, conducting a laboratory experiment was a logical choice as it gave us more control of the environment and any external influencing factors. We created a hypothetical website for University of Washington Bothell students called "Slack at UW." Slack is a cloud-based set of team collaboration tools and services which allow users to create groups, conduct real-time communication, and sharing of resources for projects.

We developed and tested a functioning prototype of an experimental peer feedback password meter, which displayed the password strength based on how it compared to previously registered users. However, to check the peer feedback meter

and its effect we only had to test how the visual representation of the peer feedback meter impacted the password strength. Hence, to be consistent, we changed how the score was calculated on the backend of the system for the experiment. Instead of calculating based on a comparison to previously registered users, we calculated based on the scoring algorithm discussed in section 3.1. Both the traditional and peer feedback meters used the same color schemes, scoring rules, and evaluation of feedback based on scores. The feedback evaluation is illustrated in Table 2. The value of the progress bar was calculated by multiplying the password strength value by two.

In our study, we informed each participant that this system is being developed specifically for University of Washington Bothell students to assist them in collaborating on group projects. Hence, we created a perception amongst the participants that this was a real prototype website soon to be launched for them and they had a real stake in it. This was the key in our study as previous research had shown that systems which contain sensitive or critical information (e.g. financial) about users are perceived to require stronger passwords as compared to accounts associated with blogs or forums [37]. In our view, our hypothetical website was around the middle to upper-middle level on the scale of sensitive information stored.

During the early part of our study and based in part on the survey they each had completed, we became concerned that participants were simply using an existing password rather than developing a new one. At approximately the midpoint of the study, we decided to investigate if explicitly stating in the oral instructions to users that they need to create a unique password would cause them to instead develop a new password rather than reusing an existing one. The use and subsequent effectiveness of cues have been shown in other research [33].

We created two variations of the same website. The first one was the control with the traditional meter as illustrated in Fig. 2.

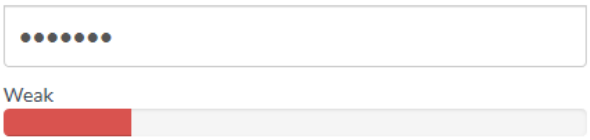

Fig. 2. Traditional password meter

The second one was our experimental variant with the peer feedback meter as shown in Fig. 3.

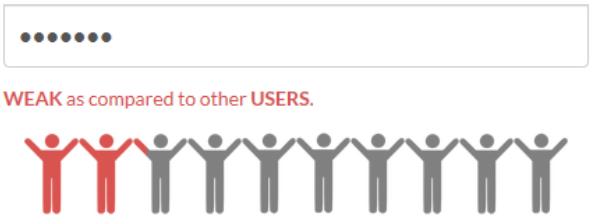

Fig. 3. Peer feedback meter

The difference between the two meters was how they were visualized and what feedback was given. The traditional meter simply stated the feedback and set the percentage value on the progress bar. The peer feedback meter, however, stated the feedback in comparison to other users and instead of displaying a traditional progress bar we created a 10-person series bar which reflected that it is a percentage from a population. Apart from this meter, both the variants had identical content and design.

A. *Password Scoring Algorithm*

An open source library for password strength calculation was used called “jQuery Password Strength Meter for Twitter Bootstrap” [38]. This library uses custom rules and settings to calculate the strength of a given password. The library allows a developer to use a scoring system which rewards or penalizes when rules match. Rules can be minimum length, word repetitions, characters used, etc. The default scoring rules are shown in Table 1.

TABLE I. ALGORITHM RULE BASED SCORES

Rule	Score
wordNotEmail	-100
wordLength	-50
wordSimilarToUsername	-100
wordSequences	-50
wordTwoCharacterClasses	2
wordRepetitions	-25
wordLowercase	1
wordUppercase	3
wordOneNumber	3
wordThreeNumbers	5
wordOneSpecialChar	3
wordTwoSpecialChar	5
wordUpperLowerCombo	2
wordLetterNumberCombo	2
wordLetterNumberCharCombo	2

The total score at the end is raised to the power of 1.4 based on the password length. This resulted in a possible range for the calculated password score of -56 to +75. For the traditional meter, the following score values, shown in Table 2, are used to determine the feedback verdict given back to the user.

TABLE II. TRADITIONAL METER SCORING

Score	Feedback verdict
Score < 0	Very Weak
0 > Score < 14	Weak
14 > Score < 26	Normal
26 > Score < 38	Medium
38 > Score < 50	Strong
Score > 50	Very Strong

We kept the same scoring rules for the peer feedback meter. However, instead of displaying the same feedback verdict, we calculated the score of the password and then calculated the percentage of users that score below the current user's score. Moreover, we used the score results to display the feedback. Again, this is what was done for the actual system created. For testing purposes, we used the same scoring algorithm for both variations.

Cognitive interviews were conducted to ensure that our understanding of the peer feedback provided to participants was consistent with our intent [39]–[41]. These participants interpreted the presentation of the peer feedback meter in a manner consistent with our expectations—a comparison with how strong their password is compared to others that had already registered on the system.

While a password could still be strong despite being weaker than a significant number of peers, the primary goal in this research was to determine how a user responds when presented with information on how the strength of his password compares with his peers. Thus, it is possible that other factors could be considered in such a comparison, such as the form or structure of the password apart from its strength, but that is not the focus of this research.

B. Design of the System

The website was coded in HTML, CSS, and JavaScript for the front end. The backend server functionality was coded in PHP. A MySQL database was used to store the data on the server. Apart from the website, there was another web page created which acted as an admin panel. It was password protected with credentials so only the research team would have access to the data.

The data stored for the study was the user's email address, degree program, their IP address and location (resolved through IP address lookup), the meter type (traditional or peer feedback) they were exposed to, their password strength calculated using the algorithm discussed in 3.1, the characteristics of the passwords (e.g. number of digits, lowercases, uppercase and repetitions), hashed password using "bcrypt" algorithm and salted with a random cryptographic salt, and the number of unique password tries along with their score. Each user was assigned a unique identifier or UUID which was based on Mersenne Twister Random Number Generator. The database itself was encrypted and each data column was encrypted with AES 256-bit algorithm for further protection.

Fig. 4 shows the workflow of how the system operated. The user would interact with the user front end–the "Slack at UW" website. Various web service calls can be made through the frontend for the required functionality. A password protected MySQL database was used to store the necessary data and information for the study. The researchers could interact using a password protected admin frontend to view data in the database.

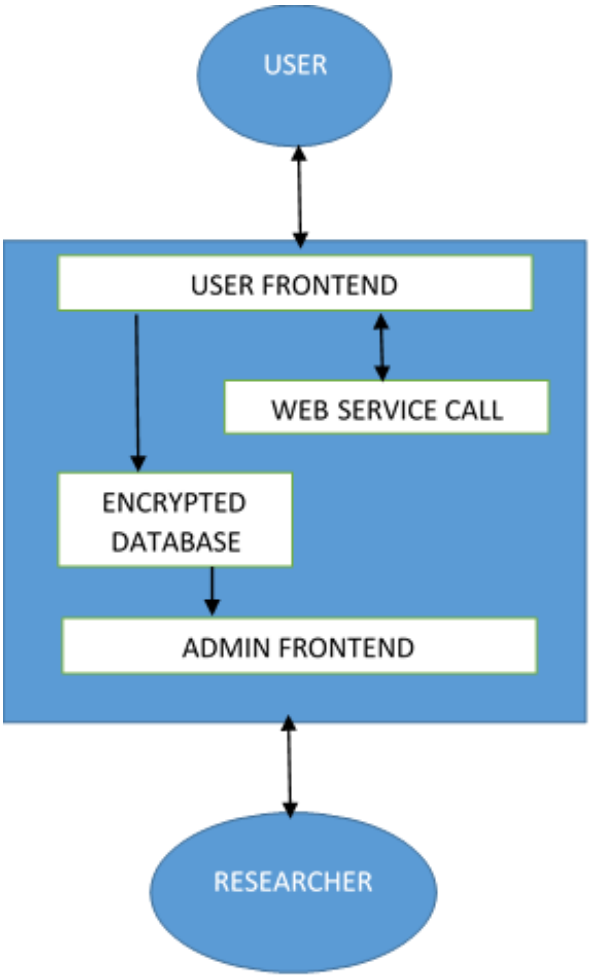

Fig. 4. System workflow diagram

We administered a survey to participants to collect demographic information along with Internet security behavior questions to gauge the sample population's expertise in online security and how they behave online. The survey also served as a distraction to determine if the participants would be able to recall the password they chose after approximately five minutes.

IRB approval was obtained before recruitment and subsequent participation in the study. For the recruitment of participants, flyers were posted around various campus buildings. The prospective participants were required to sign up by completing a qualifying survey where if selected they were invited to schedule an appointment to participate in the

study. The participants were provided with a $20 Amazon gift card as compensation for their time.

IV. Results

Forty-eight students were recruited as participants for this study. They ranged in age from 18 to 39 years old with 24 males and 24 females. Approximately 54% of participants identified as White/Caucasian with 42% females and 58% males, followed by 27% who identified as Asian/Pacific Islander with 31% males and 69% females.

Out of the 48 total participants, 14 were from a computing background and the rest were from a non-computing background.

Twenty-four of the participants were in the control group, while the other 24 were in the treatment group. Half of the participants in each group were given explicit instructions to create an account using a unique password that they had not used anywhere previously. The other half of the participants in each group were not given explicit instructions. The breakdown is presented in Table 3.

TABLE III. Participant Distribution Among the Groups

Non-Explicit Instructions		**Explicit Instructions**	
Male		**Male**	
Peer Feedback	6	Peer Feedback	6
Traditional	6	Traditional	6
Female		**Female**	
Peer Feedback	6	Peer Feedback	6
Traditional	6	Traditional	6
Total	**24**	**Total**	**24**

Both groups were told to explore the website and create an account on the sign-up page. After registering for an account, they were redirected to an online survey. Upon completion of the survey, they were asked to log in using the credentials they had created on the sign-up page.

Additionally, both groups had identical lab settings. To minimize any external influence, the participants and the researcher met individually in a meeting room. The participants were asked to read and sign a consent form. The researcher then presented the participant with either the control website with the traditional password meter or the experimental website with the peer feedback password meter. Participants were then given instructions. Care was taken to ensure each group had an equal number of both males and females. The participants were redirected to sign in with their credentials which they had created on the sign-up page after they had completed the online survey. The participants were not told about the sign-in beforehand as this approach ensured that the participants would not make a conscious effort to memorize or write down the password to find out if creating a stronger password had any effect on the memorability of the said password phrase.

For hypothesis testing, IBM SPSS version 19 was used to perform independent samples t-test on the results obtained. As mentioned in section three, we could not find a statistically significant result without the explicit instructions to create a unique password. However, when the cue was explicitly stated during both the traditional and peer feedback meters, a statistically significant difference between the control and treatment groups were found ($t=1.882$; $p < 0.05$). For those that used the traditional meter, the average strength of their passwords (39.81 ± 15.81) was lower than those that used the peer feedback meter (53.35 ± 19.27).

Out of all the participants, only 16.67% (eight participants) had difficulties remembering their entered password. Four of these participants could enter their correct password by the second attempt.

Further analysis of the data collected revealed that when the peer feedback meter was presented, it took an average of 6.42 tries until the user was able to finalize their password selection. Table 4 shows the average number of attempts it took to decide upon the final password for each treatment given.

TABLE IV. Number of Attempts to Submit Final Password

Condition	**Average # of attempts to submit final password**
Traditional meter with no explicit instruction	2.92
Peer-Feedback meter with no explicit instruction	2.83
Traditional meter with explicit instruction	5.50
Peer-Feedback meter with explicit instruction	6.42

Fig. 5 illustrates how the score of both password meter treatments within the explicit instruction groups influenced the resulting password strength of users. There was an increase in the average password strength of the peer feedback meter group as compared to the traditional meter group. This suggests that the peer feedback meter influenced the users to create a stronger password when every other condition was kept the same between the two groups.

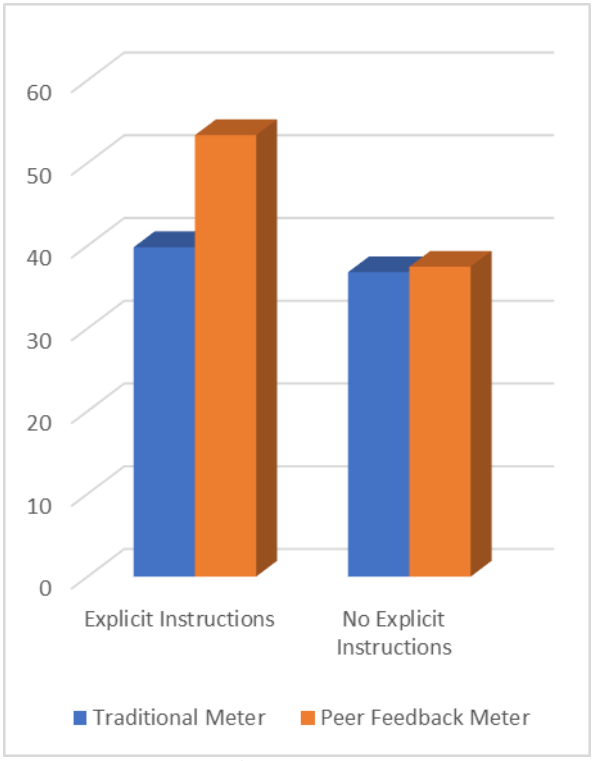

Fig. 5. Traditional vs. peer feedback meter

Taking a deeper look into the above results, Fig. 6 and Fig. 7 illustrate that males consistently created stronger passwords as compared to females; however, the difference found was not statistically significant. Fig. 6 illustrates the difference by gender when no explicit instructions were given.

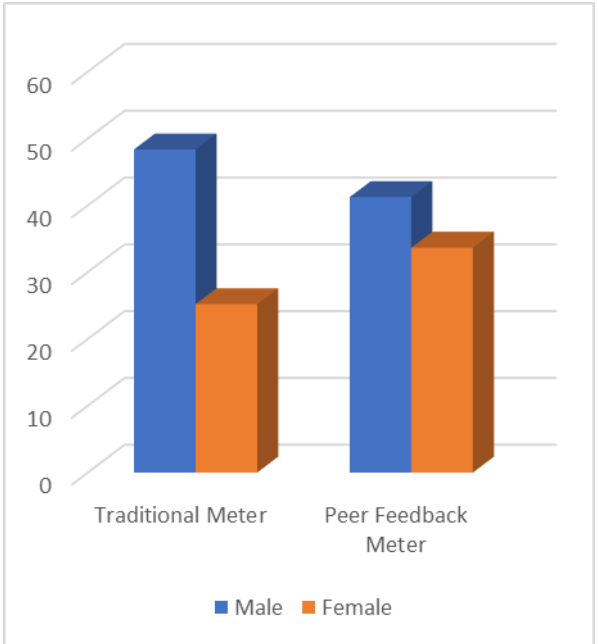

Fig. 6. Password strength by gender without explicit instructions given for a unique password

Fig. 7 also provides a breakdown by gender when explicit instructions were given to the participant to create a unique password.

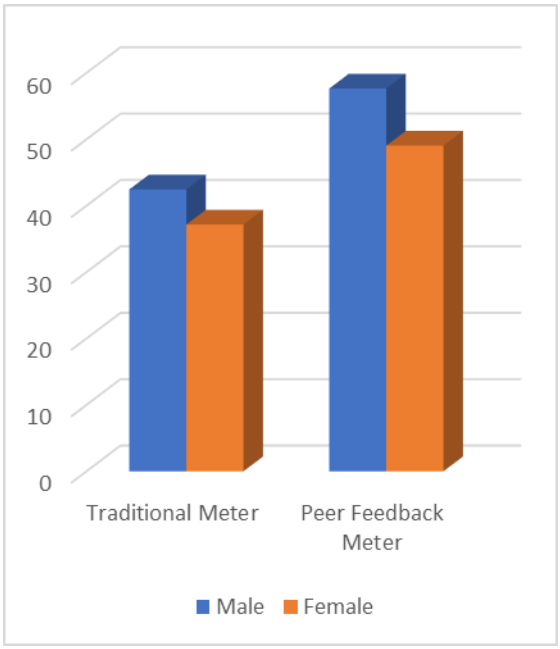

Fig. 7. Password strength by gender with explicit instructions given for a unique password

We analyzed the data collected from the surveys and present some of the results here. Approximately 59% of the participants only change their passwords if they have been notified of a security breach. This approach to managing passwords is problematic [42].

Another interesting insight we gathered from the survey results is that around 20% of the participants stated they create secure passwords for all their accounts, followed by emails and banking websites at 13%.

Approximately 52% of the users perceived that having a minimum of eight characters in a password constitutes as good, while around 10% of the users believed the password should be more than 10 characters.

V. Conclusion

The hypothesis under investigation in this study was whether or not peer influence had any effect on a user's password choice. To address this question, we conducted an experiment on a pool of 48 university students. In the design of this experiment, we had two forms of password meters, one was the traditional password meter and the other was a peer feedback password meter. Additionally, we administered this experiment with two different conditions—explicit instructions provided and no explicit instructions provided.

Our results suggest that the peer feedback meter, when administered alongside explicit instructions, would create relatively stronger passwords as compared to traditional

password meters. Thus, our hypothesis is supported. However, when no explicit instructions were given, we did not find a statistically significant difference. We believe that such prototype peer feedback meters could have the most benefit on platforms which already depend upon social connections between users, alongside a mechanism to prompt the user to create a unique password. Platforms like Facebook, LinkedIn, Google, and Outlook are prime examples. While the behavior of users on some of these platforms present their own security and privacy challenges [43], keeping unauthorized individuals from accessing accounts remains a top priority.

A. Limitations

Our experimental design had several limitations. First, our sample population was not a true representative sample of the general population. Our participants were university students whose mean age ranged from 20 to 24 years old. This age group is a part of "Net Generation" and is considered more intuitively aware of the digital and IT space around them [44]. Additionally, Western college-educated participants represent a unique cognitive makeup different from the population as a whole [45]. The laboratory experimental design of our study has an inherent limitation in that it has low generalizability and lacks ecological validity. Participants were out of their natural environment and we were able to exert some control over them. Thus, field experiments would need to be done to determine if the results obtained here would also be found in the natural environment [46].

Another limitation is that in our quest to find effects of peer influence in password meters, we did not test different variations of the GUI design. Although it is uncertain if a different design would or would not have any effect, different prototypes could be developed and tested in a pilot study to determine the effectiveness of each. Some of the prototype designs we initially considered include: 1) how would peer feedback influence users if they were told their password's strength compared to users from another geographical region, and 2) showing users a percentage value of how much stronger or weaker their password is when compared to others.

Additionally, our study involved compensation worth $20 and thus could influence some individuals to have participated in the study that otherwise would not have participated.

Finally, for new systems traditional password meters would need to be shown to gather password strengths data which would be used to compare against other users. This may introduce a vulnerability of the system. If attackers can breach the system and crack the database encryption, they might find accounts which have weaker passwords and prioritize by attacking them first. Long-term implementation of such a system can be complicated. In a scenario when most of the users are storing weak passwords by current scoring algorithms, if a new password is created which might be relatively better but not strong on its own, the peer feedback meter might show it as a much stronger password compared to others. Hence, this would lead to a general degradation of password strength quality over time. One way to avoid this is to use a combination of policies, such as enforcing complex character set and minimum password length to increase the entropy.

B. Future Work

Although user authentication mechanisms have improved over the years, more needs to be done. Our attempt was designed to investigate a direct relationship between peer influence and password strength through the use of peer feedback meters. However, we did not conclusively establish if peer feedback generated passwords led to more memorable passwords, which weighs heavily in the usability-security tradeoff. Future research should focus on investigating that relationship in detail by employing various other algorithms based on entropy to find a direct link between peer feedback and the increase in entropy.

We had only a single prototype design for the peer feedback password meter. A variety of such peer feedback meters should be tested to determine if they can improve the quality of generated passwords by delivering the feedback in a more user-friendly manner.

Additionally, our study had a small sample size. Conducting a large-scale study on systems which have a high volume of users would generate more reliable data. Moreover, conducting this study in such prototype peer feedback meters on functioning systems would help evaluate our hypothesis as the data of users can be compared to those of others. Also, it is worth mentioning that we did not check the password strength scores when no password meter is present at all. This would require a larger sample for the study and future research can design studies which compare all three scenarios.

This study was conducted in a lab setting whereas conducting it in realistic or field environment would produce results with higher external validity. Thus, future work should seek to collect data in a more naturalistic setting.

REFERENCES

[1] McAfee labs, "McAfee Threat Predictions 2016," 2016.

[2] Experian Data Breach Resolution, "DATA BREACH INDUSTRY FORECAST," no. 4, 2017.

[3] N. Symantec, "Norton Cyber Security Insights Report 2016," *Nort. Symantec*, pp. 3–9, 2016.

[4] Verizon, "2017 Data Breach Investigations Report," no. 10, p. 74, 2017.

[5] M. Dupuis, R. Crossler, and B. Endicott-Popovsky, "Measuring the Human Factor in Information Security and Privacy," in *The 49th Hawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, 2016.

[6] W. Han, Z. Li, L. Yuan, and W. Xu, "Regional Patterns and Vulnerability Analysis of Chinese Web Passwords," *IEEE Trans. Inf. Forensics Secur.*, vol. PP, no. 99, pp. 1–1, 2015.

[7] Trustwave, "2015 Trustwave Global Security Report," p. 90, 2015.

[8] Verizon, "2016 Data Breach Investigations Report," *Verizon Bus. J.*, no. 1, pp. 1–65, 2016.

[9] J. Godin, “Infosecurity Europe 2004: Office workers give away passwords for a chocolate bar,” *Univ. Wired*, no. April, p. 1, 2014.

[10] K. Security, “Crossing the password chasm,” no. January, 2017.

[11] M. Dupuis and S. Khadeer, “Curiosity Killed the Organization: A Psychological Comparison between Malicious and Non-Malicious Insiders and the Insider Threat,” in *Proceedings of the 5th Annual Conference on Research in Information Technology*, Boston, MA, USA, 2016, pp. 35–40.

[12] E. H. Spafford, “OPUS: Preventing weak password choices,” *Comput. Secur.*, vol. 11, no. 3, pp. 273–278, 1992.

[13] R. Morris and K. Thompson, “Password security: a case history,” *Commun. ACM*, vol. 22, no. 11, pp. 594–597, Nov. 1979.

[14] D. V Klein, “Foiling the cracker: A survey of, and improvements to, password security,” *Proc. 2nd USENIX Secur. Workshop*, pp. 5–14, 1990.

[15] X. D. C. De Carnavalet and M. Mannan, “A Large-Scale Evaluation of High-Impact Password Strength Meters,” *ACM Trans Inf Syst Secur*, vol. 18, no. 1, p. 1:1--1:32, 2015.

[16] A. Adams and M. A. Sasse, “Users are not the enemy,” *Commun. ACM*, vol. 42, no. 12, pp. 40–46, Dec. 1999.

[17] X. D. C. De Carnavalet and M. Mannan, “From Very Weak to Very Strong : Analyzing Password-Strength Meters,” *Ndss 2014*, no. February, pp. 23–26, 2014.

[18] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley, “Does My Password Go Up to Eleven? The Impact of Password Meters on Password Selection,” *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 2379–2388, 2013.

[19] B. Ur *et al.*, “How does your password measure up? The effect of strength meters on password creation,” 2012, pp. 65–80.

[20] L. F. Berkman, “Social support, social networks, social cohesion and health.,” *Soc. Work Health Care*, vol. 31, no. 2, pp. 3–14, 2000.

[21] P. DiGioia and P. Dourish, “Social navigation as a model for usable security,” in *Proceedings of the 2005 symposium on Usable privacy and security - SOUPS ’05*, 2005, pp. 101–108.

[22] L. O’Gorman, “Comparing passwords, tokens, and biometrics for user authentication,” *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, Dec. 2003.

[23] M. Bishop and D. V. Klein, “Improving system security via proactive password checking,” *Comput. Secur.*, vol. 14, no. 3, pp. 233–249, Jan. 1995.

[24] S. Komanduri *et al.*, “Of Passwords and People: Measuring the Effect of Password-Composition Policies,” *Proc. 2011 Annu. Conf. Hum. Factors Comput. Syst. - CHI 11*, p. 2595, 2011.

[25] P. G. Inglesant and M. A. Sasse, “The true cost of unusable password policies,” *Proc. 28th Int. Conf. Hum. Factors Comput. Syst. - CHI 10*, p. 383, 2010.

[26] W. E. Burr *et al.*, “Electronic Authentication Guideline,” Gaithersburg, MD, Nov. 2013.

[27] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas, “Generating and remembering passwords,” *Appl. Cogn. Psychol.*, vol. 18, no. 6, pp. 641–651, 2004.

[28] J. J. Yan, “A note on proactive password checking,” in *Proceedings of the 2001 workshop on New security paradigms - NSPW ’01*, 2001, p. 127.

[29] B. Ur *et al.*, “How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation Blase,” in *Security’12 Proceedings of the 21st USENIX conference on Security symposium*, 2012, pp. 5–16.

[30] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. (Belin) Tai, J. Cook, and E. Eugene Schultz, “Improving password security and memorability to protect personal and organizational information,” *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 8, pp. 744–757, Aug. 2007.

[31] M. Whitty, J. Doodson, S. Creese, and D. Hodges, “Individual differences in cyber security behaviors: An examination of who is sharing passwords,” *Cyberpsychology Behav. Soc. Netw.*, vol. 18, no. 1, pp. 3–7, Jan. 2015.

[32] B. Ur, J. Bees, S. M. Segreti, L. Bauer, N. Christin, and L. F. Cranor, “Do Users ’ Perceptions of Password Security Match Reality ?,” *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI 16*, pp. 3748–3760, 2016.

[33] A. Besmer, J. Watson, and H. R. Lipford, “The impact of social navigation on privacy policy configuration,” in *Proceedings of the Sixth Symposium on Usable Privacy and Security - SOUPS ’10*, 2010, p. 1.

[34] S. Egelman, D. Molnar, N. Christin, A. Acquisti, C. Herley, and S. Krishnamurthi, “Please Continue to Hold An empirical study on user tolerance of security delays,” 2010.

[35] D. Kahneman and A. Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econom. J. Econom. Soc.*, vol. 47, no. 3, pp. 263–291, 1979.

[36] I. Ajzen, “The theory of planned behavior,” *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 179–211, 1991.

[37] W. Khern-am-nuai, W. Yang, and N. Li, “Using Context-Based Password Strength Meter to Nudge Users ’ Password Generating Behavior : A Randomized Experiment,” *SSRN*, p. 27, 2016.

[38] T. Piper and A. Blanco, “jQuery Password Strength Meter for Twitter Bootstrap.” GitHub, 2016.

[39] P. C. Beatty and G. B. Willis, “Research Synthesis: The Practice of Cognitive Interviewing,” *Public Opin. Q.*, vol. 71, no. 2, pp. 287–311, 2007.

[40] P. Housen, “What the Resident Meant to Say: Use of Cognitive Interviewing Techniques to Develop Questionnaires for Nursing Home Residents,” *Gerontologist*, vol. 48, no. 2, pp. 158–169, 2008.

[41] M. Rosal, E. Carbone, and K. V. Goins, “Use of cognitive interviewing to adapt measurement instruments for low-literate Hispanics.,” *Diabetes Educ.*, vol. 29, no. 6, 2003.

[42] W. C. Summers and E. Bosworth, “Password policy : The good , the bad , and the ugly,” *Res. Gate*, no. February, 2015.

[43] M. Dupuis, S. Khadeer, and J. Huang, “‘I Got the Job!’: An Exploratory Study Examining the Psychological Factors Related to Status Updates on Facebook,” *Comput. Hum. Behav.*, vol. 73, pp. 132–140, 2017.

[44] D. Oblinger and J. Oblinger, “Is It Age or IT: First Steps Toward Understanding the Net Generation,” *Educ. Net Gener.*, vol. Chapter 2, no. 2, p. 2.1-2.20, 2005.

[45] J. Henrich, S. J. Heine, and A. Norenzayan, “Most people are not WEIRD,” *Nature*, vol. 466, no. 7302, pp. 29–29, Jul. 2010.

[46] J. B. Kessler and L. Vesterlund, “The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects,” in *Handbook of Experimental Economic Methodology*, no. 2006, Oxford University Press, 2015, pp. 391–406.