A Password Study to Explore Usability and Security of Text Passwords on Mobile and Traditional Devices

MUHAMMAD SAJIDUR RAHMAN, University of Florida, USA
LAKSHMAN RAMESH, University of Florida, USA

Text passwords remain the de-facto end-user authentication standard. Research has shown that there is an arms race between the password policy and the type of device used for password creation and reentry which affects security and usability [4]. This paper presents results from a 32 participant, between-group, simulated password study which shows that creating passwords on mobile takes longer (89.09s) than traditional devices (50.63s). In contrast, more attempts are required on traditional (2.40) than on mobile devices(0.80). Usage of special characters in password composition is also found to be lower in mobile condition(1.70). Overall, the study finds no significant difference in usability but a marginally significant difference in syntactic properties of text passwords between mobile and traditional devices in the context of 3class12 password policy.

1 INTRODUCTION

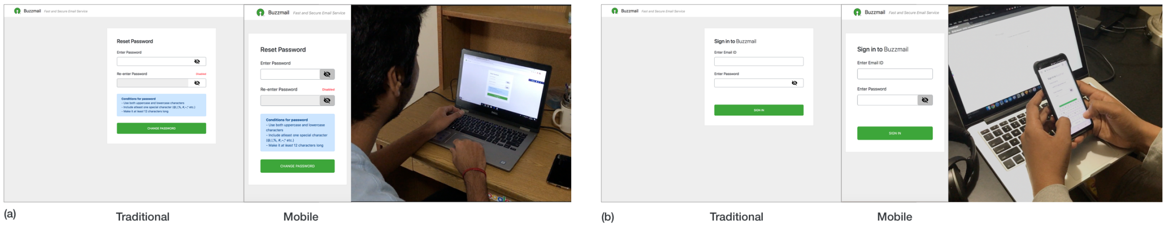
The last decade has seen a significant surge in the use of mobile (smartphone, tablet) and traditional devices (laptop, desktop), referred to as MD and TD from here on, for communication, banking, media consumption among other contexts. These are situations which often involve password creation and user-authentication, specially in mobile environment. Prior research speculated that passwords created on MD might be more vulnerable for offline attacks [7], as people tend to choose simpler passwords for ease of use and memorability [2]. Further, password usability and security in MD context is largely unexplored, which motivates us in our current study.

We hypothesize that password creation on MD are as usable (creation time, no. of attempts) and as secure (measured by password syntactic properties) as password creation on TD, in the context of 3class12 password policy. To this end, we conducted a 32-participant, two-part, between-subject simulated password study. Password usability was measured based on participants recorded behavior (e.g., creation time, no. of attempts) and security using syntactic properties of password, as proposed by previous research [3].

Our analysis shows that participants, on average, take longer time in creating passwords on mobile devices compared to traditional devices. Moreover, participants' self-reported user sentiment revealed that password entry on mobile condition is more difficult (77%) than on traditional condition (39%). Distribution of special characters was significantly different across MD and TD but we did not find any significant differences for other character cases (upper, lower, digit). Overall, the study finds no significant difference in usability of text passwords but a significant difference in password composition across devices in the context of 3class12 password policy.

2 BACKGROUND AND RELATED WORK

A lot of password research has been conducted in the past few decades. Shay et al. conducted large-scale online password studies, focusing 15 password-composition policies and characterized the resulting password by describing the number of character classes or complexity [6]. For instance, the 3class12 policy requires at least three different character classes in passwords of minimum length twelve. In the work most closely related to ours, Melicher et al. studied security and usability of passwords created on mobile and traditional devices and concluded that 3class12 policy would be suitable for usability whereas 2word16 would attain sufficient security against offline attacks [4]. Mazurek et al. studied passwords created under various policies for exploitable patterns, such as consisting of a dictionary word followed by a number and symbol [3]. Komanduri et al. examined end-user's password usage behavior under different policies and reported how stricter policies can overly burden and annoy users [2]. Research has also been conducted to assess effects of various input methods on password strength and usability. Haque et al. reported that entropy of passwords created on devices with physical keyboards were significantly different from those created with touchscreen keyboards [1]. In a related work, Zezschwitz



Conditions for password
- Use both uppercase and lowercase characters
- Include atleast one special character (@,!,%, #,~;* etc.)
- Make it *at least* 12 characters long

Conditions for password
- Use both uppercase and lowercase characters
- Include atleast one special character (@,!,%, #,~;* etc.)
- Make it *at least* 12 characters long

Fig. 1. (A) Part-1: password reset task in a traditional device and (B) part-2: password login task in mobile device. Same graphical user interface (A) was used on all devices, but adjusted to fit respective screen size.

et al. found that mobile users tend to create shorter passwords with fewer symbols and uppercase letters [7]. In our work, we examine 3class12 password policy across TD and MD to examine if it can conform with previous findings in terms of security and usability [4] [6]. For data analysis, we performed non-parametric Kruskal-Wallis test [4] due to non-normality found in dataset.

3 EXPERIMENT

The experiment followed a two-part, between-group, online design with the type of device used for password creation and re-entry as independent variable(3 levels). Participants used their own devices during study in their convenient time and place. They were assigned to one of the following condition groups: *Mobile to mobile (MM)*, *Traditional to mobile (TM)* and *Traditional to traditional (TT)* in a round robin manner. We had three dependent variables: *Time to create*, *Creation attempts* and *frequency of Special character*. Our overall methodology is based on techniques presented in [4]. We have released a replication package of this study (source code) in Github [5] under MIT open-source license.

In part one, participants received an email with a fictional email id ('userXXX@buzzmail.com') and a link to begin the study. After following the link, we asked participants to imagine that their email provider had suffered a data breach and required them to reset password, under 3class12 policy (part-a, Fig. 1). To proceed to password reset, participants were required to prove they were using correct device, such as MD or TD, which was verified by checking participant's browser HTTP User-Agent string. Until the correct device was verified, they could not proceed to the rest of the study. After resetting password, participants completed a brief demographic survey and reported their agreement on the statement, "I found it difficult to enter the password I created on this device" on a five-point Likert scale.

One day later after part-1, participants received another email with a link to continue the study. After visiting the link, they were asked to login using the password they created in part-1 (part-b, Fig. 1). Participants who made five incorrect attempts or clicked "Forgot Password" link, saw their password on screen. After password entry, participants completed a survey about password storage method and behavior.

4 DATA COLLECTION

Participants were recruited via convenience sampling. A total of 40 participants signed an online consent form, of which 32 completed part-1 and 29 returned to complete part-2 (age-range 21-34, mean 26, 69% male, 31% female).

The experimental apparatus were the participant's own devices(MD and/or TD) used during study (see Fig. 1). We measured and logged creation attempts (number of attempts taken to create a password), creation time (time elapsed in seconds between page load and submission), re-entry time (during part-2), no. of character deletions, no. of copy-pastes along with the text password created. All data was measured on the participant's machine to avoid network latency (via custom-build, client-side javascript data recorder), transmitted over HTTPS connection to a remote server(running PHP scripts for verification-validation-data storage) and then anonymized and stored in a JSON file per participant. At the end of study, all JSON files were downloaded and undergone data pre-processing

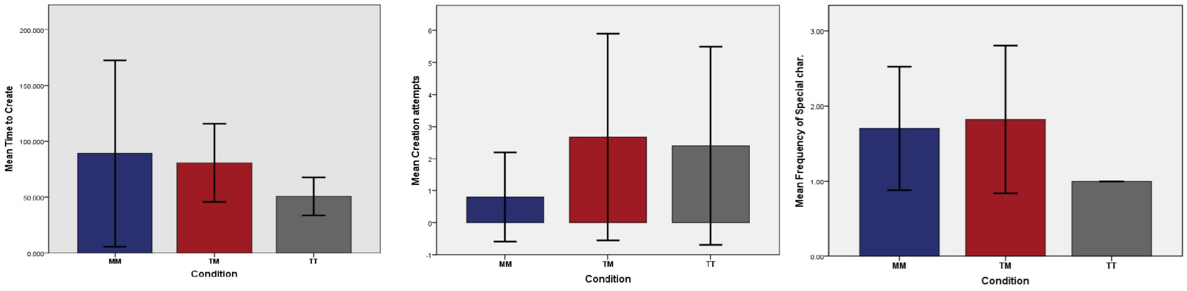


Fig. 2. Barcharts show mean values(from left) of *Time to create*, *Creation attempts* and *Special character frequency*. Error bars: +/- 1 Standard Deviation.

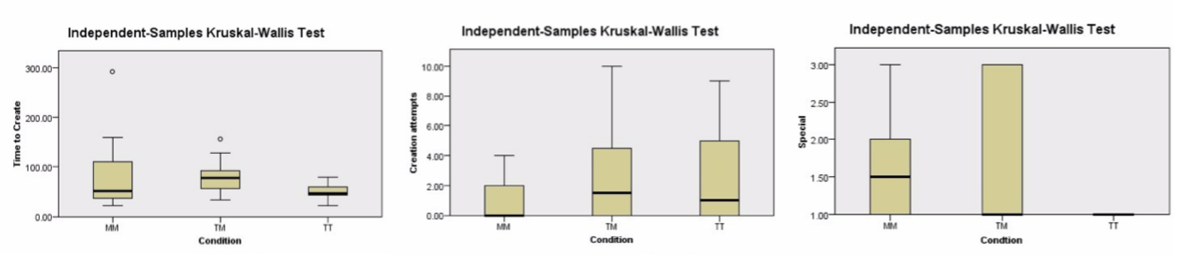


Fig. 3. Kruskal-Wallis Test, mean rank (from left): ‘creation time’ $16.50s$ for MM, $20.42s$ for TM and $11.80s$ for TT;‘creation attempts’ 12.70 for MM, 18.88 for TM and 17.45 for TT; ‘ ‘special character frequency’ 1.50 for MM and 1 for both TM and TT.

and conversion into a single wide-format csv file for further analysis. We used a custom-built Python script for this purpose.

5 DATA ANALYSIS AND RESULTS

Due to technical errors during second part of the study, we run all of our statistical analyses on data collected from first part. We used SPSS statistical software package for data analysis. Table 1 presents descriptive statistics of password usability metrics and syntactic properties. In particular, we tested the following hypotheses:

H1 Password creation on MD are as usable as password creation on TD.

H2 Syntactic properties of passwords created on MD are similar to the passwords created on TD.

5.1 H1: Password Usability for Mobile vs Traditional Devices

For pairwise omnibus comparison across three condition groups, we used Kruskal-Wallis One-way ANOVA tests for two variables *Time to create* and *Creation attempts*. Non-parametric statistical test was chosen as our quantitative data failed the tests of normality (Shapiro-Wilk, $p < .05$). The Kruskal-Wallis test showed that there was no significant difference of password creation time and creation attempts among the three condition groups (*Time to create*, $\chi^2(2) = 4.602, p > .05$; *Creation attempts*, $\chi^2(2) = 2.886, p > .05$). Thus, we failed to reject hypothesis *H1*, i.e. the usability of text password creation is found to be similar across types of devices, for 3class12 password policy. Figure 3 shows mean ranks for each of the variable across three condition groups.

5.2 H2: Syntactic Properties of Passwords for Mobile vs Traditional Devices

A Kruskal-Wallis One way ANOVA test was performed, using the three condition groups and password syntactic properties (no. of uppercase, lowercase, digits and special characters in a password). While the distribution of

Table 1. Descriptive statistics of password usability metrics and syntactic properties by condition

Condition		Time to Create (s)	Creation Attempts	Upper	Lower	Spcl.	Dig.
3class12MM(n=10)	Mean	89.09	0.80	1.50	6.50	1.70	3.80
	SD	83.31	1.40	0.85	3.57	0.82	2.53
3class12TM(n=12)	Mean	80.79	2.67	1.27	7.00	1.82	4.09
	SD	35.04	3.23	0.65	4.22	0.98	2.91
3class12TT(n=10)	Mean	50.63	2.40	1.60	6.60	–	4.30
	SD	17.00	8.99	3.10	1.27	–	2.58
Total(n=32)	Mean	73.96	2.00	1.45	6.71	1.52	4.06
	SD	52.93	2.78	0.93	3.73	0.81	2.61

SD = Standard deviation. Spcl. and Dig. refer to frequency of special characters and digit. Mean and SD is omitted for *(*Speical) as it was found to be constant under TT condition.

uppercase, lowercase and digits have been found non-significant across categories of condition, there was a significant difference of using special characters in MD vs TD ($\chi^2(2) = 6.567, p < .05$). A post-hoc pairwise comparisons showed the difference of special character usage to be marginally significant between group TT and MM ($p < .10$) and between group TT and TM ($p < .10$).

6 DISCUSSION AND CONCLUSION

Previous studies have shown that simulated password studies can be an alternative for real world data. Therefore, while not ideal, our experiment data tries to mimic real world password usage behavior, thus striving for external validity. Also, we tried to minimize any bias by not priming participants about passwords beforehand. Like any online study, our study had several limitations. We faced technical problems while running the study and asked few participants to retake part-2 of the study. This might affect their 'natural' behavior and responses. Also, participants may potentially choose a convenient time and place to complete the study, so our results may represent a best case scenario for password behavior, which may hurt internal validity. For study redesign, we want to add one more level in independent variable (Mobile to Traditional) to investigate the effect of device switching on password usability and security with a larger set of participants.

Results from our study confirm previous findings [4], that 3class12 password policy can be suitable for mobile context as well as in traditional device context, without compromising usability. Consistent with prior work [7], participants used less special character in MD than TD, but no significant difference was found in other character classes (upper and lower), which was in-contrast with previous findings [4]. Our work calls for further analysis of password strength under 3class12 policy across devices.

REFERENCES

[1] S. M. Haque, M. Wright, and S. Scielzo. 2013. Passwords and interfaces: towards creating stronger passwords by using mobile phone handsets. In *Workshop on Security and privacy in smartphones & mobile devices*. ACM, 105–110.

[2] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. 2011. Of Passwords and People: Measuring the Effect of Password-composition Policies. In *2011 International CHI Conference*. ACM, 2595–2604.

[3] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur. 2013. Measuring Password Guessability for an Entire University. In *2013 ACM Conference on Computer & Communications Security*. ACM, 173–186.

[4] W. Melicher, D. Kurilova, S. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek. 2016. Usability and security of text passwords on mobile devices. In *2016 CHI Conference on Human Factors in Computing Systems*. ACM, 527–539.

[5] M.S. Rahman and L. Ramesh. 2018. A Replication Package for Simulated Password Study. Retrieved April 23, 2018 from https://github.com/lr4994/HCC-Project

[6] R. Shay, S. Komanduri, A. L. Durity, P. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. 2016. Designing Password Policies for Strength and Usability. *ACM Trans. Inf. Syst. Secur.* 18, 4 (May 2016), 13:1–13:34.

[7] E. Von Zezschwitz, A. De Luca, and H. Hussmann. 2014. Honey, I shrunk the keys: influences of mobile devices on password composition and authentication performance. In *NordicCHI'14 conference on human-computer interaction*. ACM, 461–470.