Honey, I Shrunk the Keys: Influences of Mobile Devices on Password Composition and Authentication Performance

Emanuel von Zezschwitz, Alexander De Luca, Heinrich Hussmann
Media Informatics Group, University of Munich (LMU)
Amalienstr. 17, 80333 Munich, Germany
{emanuel.von.zezschwitz, alexander.de.luca, hussmann}@ifi.lmu.de

ABSTRACT
In this paper, we present the results of two studies on the influence of mobile devices on authentication performance and password composition. A pre-study in the lab ($n = 24$) showed a lower performance for password-entry on mobile devices, in particular on smartphones. The main study ($n = 450$) showed a trend that alphanumeric passwords are increasingly created on smartphones and tablets. Moreover, a negative effect on password security could be observed as users fall back to using passwords that are easier to enter on the respective devices.

This work contributes to the understanding of mobile password-entry and its effects on security in the following ways: (a) we tested different types of commonly used passwords (b) on all relevant devices, and (c) we present analytic and empirical evidence for the differences that (d) are likely to influence overall security or reduce secure behavior with respect to password-entry on mobile devices.

Author Keywords
Mobile Devices; Passwords; Usability; Performance

ACM Classification Keywords
H.4.6. Authentication: Human factors

INTRODUCTION
Alphanumeric passwords were introduced to computers in 1962 [7]. In the following decades, they were used by professionals for specific use cases and were never meant to be everyman's universal authentication mechanism. However, with personal computers in the 1980s and with the World Wide Web in the 1990s, alphanumeric passwords became omnipresent in users' daily life. Due to the intense growth of personalized web services that demand user authentication, people nowadays have to memorize a multitude of passwords [10]. To deal with this problem, people tend to choose weak passwords, which are easier to remember and often reuse passwords for multiple services [1].

NordiCHI'14, October 26 – 30 2014, Helsinki, Finland.

ACM 978-1-4503-2542-4/14/10 $15.00.
http://dx.doi.org/10.1145/2639189.2639218

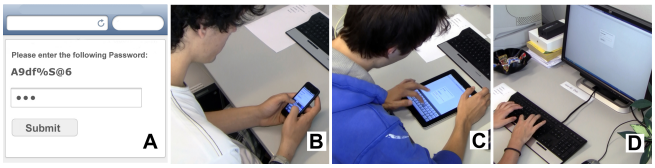

Figure 1. In the pre-study, input performance was assessed on an *Apple iPhone 5* (B), an *Apple iPad 4* (C) and a *Desktop PC* (D). The same graphical user interface (A) was used on all devices, but adjusted to fit the respective screen size.

In the 2000s, mobile devices with internet access became widely available. In June 2013, the Nielsen Company[1] announced that 62% of all U.S. mobile subscribers are using smartphones. In addition to this, the spread of tablet devices is constantly growing[2]. Such devices are used daily to access a diversity of web-based services [3, 6], although direct touchscreen interaction and text input on virtual keyboards is cumbersome and typing alphanumeric passwords is problematic [2]. More usable methods like graphical authentication and PINs are already used to unlock the device [22, 29], but are not yet supported by internet services. While prior research concerning usability factors on alphanumeric authentication focused mainly on memorability issues, we argue that input effort is an important factor on such devices as well and therefore the usage of mobile devices is likely to have an effect on password choice and user behavior.

We conducted two user studies to investigate the effects of smartphones, tablets and desktop computers on alphanumeric password-entry. First, we analyzed the effects of mobile devices on authentication performance in a controlled lab study. Then, we conducted a large-scale online study to gain insights into the impact of such devices on user behavior, password choice and security.

In this paper, we present the results of both studies and discuss the implications for the security and the usability of alphanumeric passwords. The results show that alphanumeric authentication on mobile devices is indeed cumbersome and that people opt for easier and thus weaker passwords for frequently used services. This negatively influences the effective password space.

[1] http://www.nielsen.com/us/en/newswire/2013/whos-winning-the-u-s-smartphone-market-.html, accessed: 03/06/2014
[2] http://www.idc.com/getdoc.jsp?containerId=prUS24253413, accessed: 03/06/2014

RELATED WORK

Researchers are focusing on the human factor in alphanumeric authentication since the 1990s. The first studies were based on self-reported data and revealed that knowledge-based authentication always comprises a trade-off between usability and security [1]. Adams et al. found out that user-selected passwords are often optimized for memorability. Therefore, people tend to use passwords, which are based on personal data (e.g. birthdays, names), which makes them easy to guess. More complex passwords are often written down to counter recall issues. In addition, password reuse is common and passwords are often shared with others.

In the following years, various experimental studies were conducted investigating the use of alphanumeric passwords in the World Wide Web. For example, Florencio et al. [10] conducted a long-term analysis of web-based authentication. They observed 500,000 users over a period of three months and confirmed that password reuse is very common and most users use weak passwords. Hayashi et al. [14] state that password use has become a daily task and web-based passwords are used in various locations on various devices. To counteract weak passwords, password guidelines [13, 26] and recommender systems [27] were proposed and evaluated. Even if those mechanisms can have positive effects [27, 28], adaptation to guidelines is often predictable [13], cumbersome [16] and leads to increased memorability issues [19]. As a consequence, particularly companies which depend on financial success are introducing usability-optimized password policies to avoid bothering customers [11]. In the recent years, large databases of user-selected passwords were disclosed and allowed the analysis of password space entropy. Bonneau et al. [4] analyzed 70 million passwords and showed that password composition is hardly influenced by the sensitivity of the protected data and that entropy in password space is low. The analysis of other password lists [20, 24] confirmed that users often chose the same weak passwords and a big part of the theoretical password space remains unused.

All described studies report on alphanumeric passwords in the context of desktop computers and therefore mainly focus on memorability issues. However, with mobile devices, input effort becomes a more important factor in authentication. While several studies focused on generic text input on small keyboards (e.g. [25]) and mobile devices (e.g. [15]), only a few publications focus on alphanumeric authentication on such devices. Bao et al. [2] analyzed the input effort of alphanumeric passwords on smartphones and desktop computers and found out that typing passwords is cumbersome and time consuming on both types of devices. The authors investigated general text input on mobile devices and the analysis of alphanumeric passwords has not been in focus. Therefore, deeper insights into password choice and password composition were not presented. Schloglhofer et al. [23] evaluated various authentication mechanisms considering the unlock of mobile devices. They conclude that alphanumeric passwords are by far the least usable solution. Furthermore, alternative solutions for fast alphanumeric password input have been proposed, but are still not widely supported (e.g. [17]). In addition, Schaub et al. [21] state that different software keyboards significantly influence authentication performance and might influence password composition as some characters are easier to enter than others. However, the analysis was restricted to smartphone keyboards and does not provide insights on the impact of the device itself.

We present the first large-scale analysis of the influences of tablets, smartphones and desktop computers on alphanumeric authentication. By gathering performance data in a laboratory experiment and collecting qualitative feedback via an online study, we are able to analyze the influence of mobile devices on password performance, password choice and security behavior. Thereby, we gathered novel insights into effects which are likely to influence the security of alphanumeric passwords in the long run.

PRE-STUDY: ASSESSING PASSWORD PERFORMANCE

To analyze the impact of mobile devices on authentication performance, we conducted a laboratory experiment evaluating different *Device*×*PasswordCategory* combinations.

Design

The study was conducted using a within-participants repeated measures design. The independent variables were *Device* with three levels (“smartphone”, “tablet”, “PC”) and *PasswordCategory* with three levels (“dictionary”, “internet”, “random”). That is, strings that resemble often discussed password complexities. *Device* was counterbalanced, *PasswordCategory* was randomized.

The dependent variables were *Authentication Speed* and *Error-Rate*. In addition, we collected qualitative data via questionnaire and video recordings.

Experimental Setup

The experiment was conducted in an isolated room at our premises. We used an Apple iPhone 5 (smartphone), an Apple iPad 4 10” (tablet) and a Windows PC with a 24” display and a Cherry JK-0100DE keyboard. We decided to use Apple products as mobile devices, because of their wide deployment and homogeneous keyboard layouts. Consequently, we were able to find consistently experienced users for all our devices.

All passwords consisted of eight characters, a common length for PCs as shown in [28]. Dictionary passwords (low complexity) were based on well-known dictionary words and comprised only lower case characters (e.g. “*casanova*”). “Internet” passwords (medium complexity) were designed with the objective to comply with commonly used password guidelines. They were not based on dictionary words but on pronounceable imaginary words. Such strings, which are built by alternating consonants and vowels can be chunked and are therefore assumed to be memorable [12]. Each “internet password” started with an upper case letter and ended with two digits (e.g. “*Yasana75*”). The third category was based on random strings (high complexity). Such passwords comprised two lower case letters, two upper case letters, two digits and two symbols in randomized order (e.g. “*A9df%S@6*”). Different passwords were used for each participant.

All devices displayed the same web-based user interface (see Figure 1) which was adjusted to fit the respective screen size.

It displayed a masked password field, a text field which was used for task descriptions (e.g. current password) and a submit button. User interaction was logged using JavaScript and a database.

Procedure

We started each session by explaining the task as well as the different levels of *Device* and *PasswordCategory*. The following procedure was used for each *Device*.

Training The user enters a short text (approx. 180 characters). No logging is done at this stage.

Typing Speed A second text (approx. 180 characters) is displayed. All participants enter the same text, but different texts are used on each device. To estimate the users' typing performance autocorrection is turned off and the input is logged.

Authentication The user enters four passwords of each category. That is, a total of 12 passwords are entered in randomized order. For each password, a maximum of three failures is allowed. After a correct authentication or three failed authentications, the next password is displayed.

The procedure was repeated until all three levels of *Device* had been tested. Users were allowed to take any hand posture but mobile devices had to be used edgewise (landscape format was not allowed). The texts of the training and the typing speed task were extracted from German newspapers. After the authentication task, the session ended with a short questionnaire collecting demographics and feedback on the used devices and passwords. The whole procedure took about 45 minutes, participants were rewarded with a 10 Euro shopping voucher.

Participants

We recruited 24 experienced users via various internet platforms and word-of-mouth advertising. All participants were required to use at least one of the examined mobile devices and a PC on a daily basis. All but one used a smartphone (17 iPhone users) on a daily basis and 15 stated to frequently use a tablet (9 iPads users). The group comprised 20 males and 4 females with an average age of 25 years (SD=7; 20-57 years).

Results

The training task was not analyzed. Seven authentication attempts were excluded as participants were interrupted or acted on the assumption of wrong passwords. First, we assess typing speed based on the natural language typing task and different keystroke models. After this, we focus on specific characteristics of the tested password categories. Our data was normally distributed and allowed for parametric tests, all post-hoc tests were Bonferroni corrected.

General Typing Speed

This analysis is based on the data of the initial typing speed trials. Each user entered approximately 180 characters of natural language text. We encouraged our participants to type as fast and correct as possible. Each user entered the same text but different texts were used on each device to minimize learning effects.

Device	Input Time	Errors (n)	Char Time
Smartphone	97.58 (23.40)	29	0.56 (0.13)
Tablet	82.78 (23.58)	22	0.47 (0.14)
PC	53.03 (18.35)	44	0.30 (0.11)

Table 1. Results of the typing speed task: overall input time, the total number of errors and character input time in seconds. Standard deviation is found in brackets.

Table 1 shows the measured performance data. A repeated measures ANOVA on the overall input time revealed a highly significant main effect for *Device* ($F(2, 46) = 47.05$, $p < .001$). Typing performance was significantly different on all devices with smartphone being the slowest and PC being the fastest ($p < .001$). Based on the keystroke-level model by Card et al. [5], an "average non-secretary typist" would need 53.2 seconds typing our trial text (190 key-strokes including shift) on a PC. In our study, participants needed 53.03 seconds on average which shows that our users can be considered trained, but not professionals. Even if Card's keystroke-model cannot be directly applied to mobile devices, the times indicate experienced users. The computed average character input time confirms this result. The error rate was not significantly influenced by *Device* ($p > .05$). Overall the number of errors when typing natural text was low with an average of 1.2 (SD: 2.2) errors on smartphones, 0.9 (SD: 1.3) errors on tablets and 1.8 (SD: 2.4) errors on the PC.

Keystroke Model

While our study design allows an in-depth analysis of influencing factors in a controlled environment, it does exclude important training aspects (e.g. motor memory) of daily password use. To assure that trained passwords would not significantly change the results of our experiment, we firstly assess performance differences on a keystroke level. While authentication is likely to become faster on all devices using trained passwords, we argue that the performance differences based on single keystrokes are likely to stay the same.

We defined two keystroke models analyzing the keyboard layouts of the mobile devices and the PC to map character transitions to the number of required keystrokes. We used these models to analyze the entered passwords. Entering a single character requires up to three keystrokes. For instance, typing a lower case "x" after a lower case "y" would require one keystroke on the mobile device, while a consecutively entered "+" takes three keystrokes. Entering characters on a PC can require up to three keystrokes as well (e.g. Shift + AltGr) though they can be performed in parallel.

Figure 2 shows the average input times for all tested combinations based on required keystrokes. An analysis of the mean character input times of dictionary passwords (1 keystroke only) shows similar typing speeds as found in the natural text trial. A repeated measures ANOVA reveals a highly significant main effect for *Device* ($F(2, 46) = 69.90, p < .001$). Character input of dictionary passwords using a smartphone (Mn=0.55 sec; SD=0.14) was significantly slower than using a tablet (Mn=0.44 sec; SD=0.09) or a PC (Mn=0.31 sec; SD=0.09), $p < .001$. Using dictionary passwords on the PC was significantly faster than on mobile devices ($p < .001$).

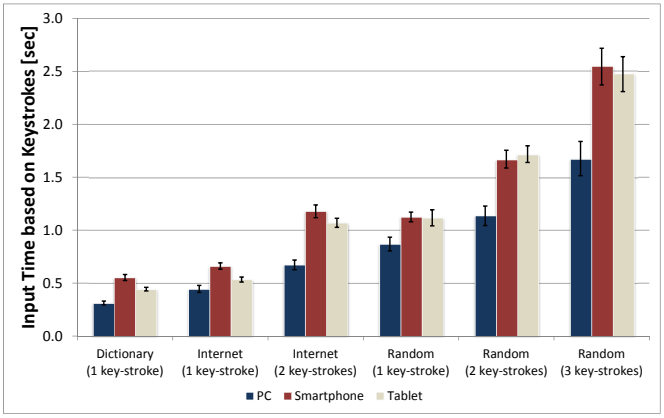

Figure 2. Input times for *Device* and *PasswordCategory* based on the respective keystroke models for iPhone, iPad and PC.

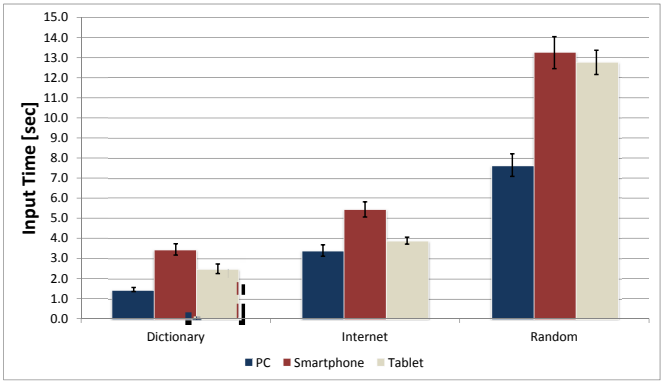

Figure 3. Password-entry times for *PasswordCategory* and *Device*. While both mobile devices generally perform worse than the PC, complex passwords seem to increase the effect.

Analyzing internet passwords (1-2 keystrokes) using a repeated measures 3×2 (*Device* $\times$ *Keystroke*) ANOVA showed highly significant main effects for *Device* ($F(2, 46) = 65.67$, $p < .001$) and *Keystroke* ($F(1.00, 23.00) = 236, 64, p < .001$; Greenhouse-Geisser corrected). In addition, a highly significant interaction effect for *Device* $\times$ *Keystroke* was found ($F(1.48, 33.97) = 30.42, p < .001$; Greenhouse-Geisser corrected). Post-hoc tests reveal that additional keystrokes significantly slow down input on mobile devices ($p < .001$). Input based on two keystrokes was 0.52 seconds slower on the smartphone, 0.54 seconds slower on the tablet and 0.23 seconds slower on the PC.

Next, we performed a 3×3 (*Device* $\times$ *Keystroke*) ANOVA based on the average input times of random passwords (1-3 keystrokes). The analysis revealed highly significant main effects for *Device* ($F(2, 42) = 29.16, p < .001$) and *Keystroke* ($F(1.30, 27.27) = 94, 90, p < .001$; Greenhouse-Geisser corrected). A significant interaction effect was found for *Device* $\times$ *Keystroke* ($F(3.04, 63.92) = 4.44, p < .05$; Greenhouse-Geisser corrected). Post-hoc tests show that the number of keystrokes significantly affects input times on all devices (all $p < .001$). However, mobile devices are significantly more affected by additional keystrokes than a PC ($p < .05$).

In summary, the keystroke analysis shows that input speed generally becomes slower, when string complexity is increased (see Figure 2). Input times based on one keystroke (dictionary passwords) are comparably fast to natural text. However, the analysis of internet passwords and random passwords shows that the performance of typing password-like strings is not comparable to natural language. The measured times of random passwords even exceed the estimated times by Card et al. [5], who proposed 0.50 seconds for typing random letters on a PC.

Authentication Speed

The results of this section are based on the average input speed of the last three authentications for each condition. Therefore, we analyzed 216 ($3 \times 3 \times 24$) distinct samples. Only correct authentication attempts were included into the analysis.

We distinguished the authentication time into three stages. The first stage, called orientation phase, is used for preparation and describes the time before the input starts. The second stage, called input phase, describes the time used for the actual password-entry. The last stage is called confirmation phase and is used to confirm the entered data. As our analysis showed that both the orientation phase and the confirmation phase are not significantly influenced by *Device* and *PasswordCategory*, we focus on the input phase.

A 3×3 (*Device* $\times$ *PasswordCategory*) ANOVA for input speed revealed highly significant main effects for *Device* ($F(2, 46) = 54.22, p < .001$) and *PasswordCategory* ($F(1.22, 28.14) = 336.94, p < .001$; Greenhouse-Geisser corrected) and a significant interaction effect for *Device* $\times$ *PasswordCategory* ($F(4, 92) = 19.81, p < .001$; Greenhouse-Geisser corrected). The average input times are shown in Figure 3. Post-hoc tests revealed that authenticating on mobile devices takes significantly more time than authentications using a PC ($p < .05$). In addition, using a smartphone takes significantly more time than using a tablet ($p < .05$). The post-hoc test of *PasswordCategory* reveals that all levels have a significant impact on the input time (all $p < .05$).

The post-hoc tests for the interaction effects showed that all levels of *PasswordCategory* perform better when entered on a PC. However, random passwords perform significantly worse, when entered on a tablet (Mn=12.8 sec; SE=0.60) or a smartphone (Mn=13.2 sec; SE=0.80) (all $p < .001$). No effect was found, when weaker passwords are used ($p > .05$). When focusing on mobile devices, the tablet outperforms the smartphone for all levels of *PasswordCategory*. The fastest combination is using dictionary passwords on a PC, (Mn=1.4 sec; SE=0.10). Entry times of tablet (Mn=3.9 sec; SE=0.17) and PC (Mn=3.4 sec; SE=0.29) do not significantly differ when internet passwords are used ($p > .05$).

Authentication Errors

Within 648 authentication sessions, 63 attempts failed (overall error rate: 9.7%). 37 authentications failed with a single error, meaning that the correct password was entered in the second try. The remaining 13 authentications were correctly finished within the third attempt. Consequently, all users were able to authenticate within three tries.

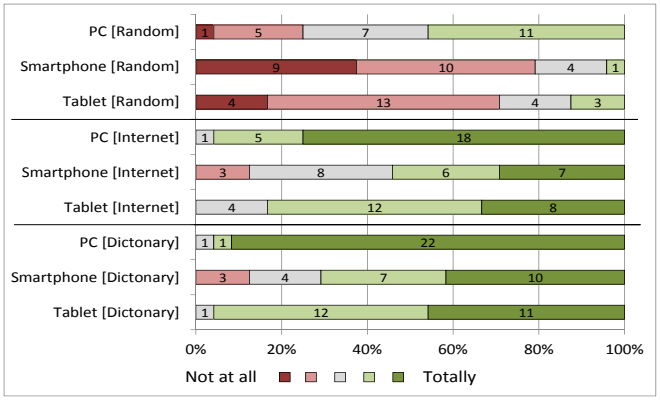

Figure 4. Answers for the statement: "[*Device*×*PasswordCategory*] is fast to use".

Although an ANOVA comparing the mean error rates of all combinations showed no significant main effects (all $p > 0.05$), the data indicates that authentication on smartphones is error prone. 47.6% of all errors were made with an iPhone, while performance on the tablet (23.8%) was comparable to the PC (28.6%). Focusing on *PasswordCategory*, dictionary strings seem to be the easiest passwords to enter (23.8%). Internet passwords led to 30.2% of all errors and random passwords were most difficult as 46.0% of all errors were based on such strings.

Looking at the combination of *Device* × *PasswordCategory* revealed that 71.1% of all errors with dictionary and internet passwords happened on mobile devices. Interestingly, authentications on the PC lead to 55.1% of all errors based on random passwords. A qualitative error analysis showed that a common error on the PC was mixing up symbols. Users entered for example "<" instead of ">". Since mobile devices have dedicated keys for such symbols, this error was not common on these devices. Though, authentication on mobile devices was prone to typing errors, where people selected keys neighbouring the target keys.

User Perception

In addition to the measured performance data, we asked the participants to compare the respective *Device* × *PasswordCategory* combinations according to their perceived ease-of-use and perceived speed. Concerning random passwords, 25.0% of our participants stated that authentication using a tablet is error prone; 41.7% stated the same for smartphones. However, only 4.2% agreed that random passwords are hard to enter using a PC. According to our participants, dictionary passwords and internet passwords are equally easy to use as only one participant disagrees on this statement.

Figure 4 gives an overview of the answers concerning perceived speed. Analog to the ease-of-use rating, people estimated random passwords to be the slowest and hardly made any difference between dictionary passwords and internet passwords. 79.2% stated that using random passwords on a smartphone was slow or very slow. According to the use of tablets, 70.8% stated the same. PC with random passwords was rated slow by 25%.

LARGE-SCALE STUDY: CHOICE AND PERCEPTION

In the pre-study, we showed a negative effect of mobile devices on password performance. Now, we are interested in whether this effect leads to a negative impact on password choice and user behavior when using mobile devices. Therefore, we conducted a large-scale user study and collected passwords on smartphones, tablets and desktop computers.

Design

The study was based on a mixed design. The survey was designed within-participants as the same questionnaire was handed-out to all users. An additional password creation task was based on a between-group design. Within the password creation task, we had the independent variable *Device* with three levels (PC, smartphone, tablet). Participants were randomly assigned to one of the three conditions with the prerequisite that they were used to the respective device (e.g. tablet for the tablet condition). The dependent variables of the password selection task were *Password* and *Error-Rate*.

Procedure

The online user study was distributed via Amazon Mechanical Turk[3]. We recruited 600 participants, that is 200 users per level of *Device* (pc, smartphone, tablet). Participants were required to use at least one mobile device (smartphone, tablet) and a PC on a daily basis. In addition, they needed to have the assigned device at hand as the password selection task had to be performed on the respective device.

After the task was accepted, the participants were redirected to an external URL hosting the questionnaire. The questionnaire asked for demographical data and investigated password experience and security behavior. Within the questionnaire, participants were asked to open a link on a specific *Device* depending on the assignment and to perform a password creation task. We used PHP Mobile Detect[4] to check if the required device was used for the task. The password creation page consisted of an introduction text and two password forms. Participants were asked to select a password for an imaginary service they would frequently use on the current device (PC, smartphone or tablet). As customary, participants had to type in their password twice. When the participants confirmed, the respective *Password* and *Error-rate* was stored in a database. For privacy reasons, passwords were separated from the survey data. Users were given two confirmation codes which were used to validate the completed tasks. The whole procedure took about 20 minutes, valid answers were rewarded with one USD.

Participants

Out of the 600 initially accepted workers, we had to obliterate 150 invalid submissions. All data was cleaned before evaluation following a strict coding. We had two levels of validation. First, we checked our two secret codes and the expenditure of time. We only accepted submissions with correct secrets and time spent over six minutes (avg. 16 minutes).

[3]https://www.mturk.com

[4]PHP Mobile Detect is an open-source script released under MIT License. (http://mobiledetect.net/)

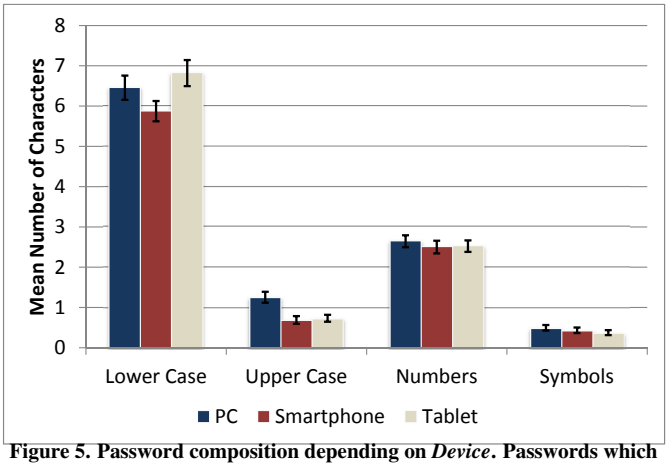

Figure 5. Password composition depending on *Device*. Passwords which were selected on smartphones are significantly shorter, PC-based passwords comprise significantly more upper case letters.

Device	Length	Lower Case	Upper Case	Numbers	Symbols
PC	10.85 (0.30)	6.46 (0.30)	1.26 (0.14)	2.64 (0.15)	0.50 (0.07)
Smartphone	9.50 (0.26)	5.88 (0.25)	0.69 (0.10)	2.50 (0.16)	0.43 (0.07)
Tablet	10.45 (0.18)	6.82 (0.32)	0.73 (0.09)	2.52 (0.14)	0.38 (0.06)

Table 2. The average number of chosen characters for each device group. Standard errors are reported in parentheses.

In the second step, we validated the given answers by checking (a) requirements and (b) contradictions. For example, we excluded participants who stated (a) not to use mobile devices and people who stated (b) to frequently use passwords on one question and to never use passwords on another question. The remaining 450 valid answers were based on 149 tablet users, 149 PC users and 152 smartphone users.

For the survey, we had 238 males and 212 females. The average age was 31 years (SD=9; Min=18; Max=67). All participants stated to be U.S. citizens, 27.3% had a technical background. The distinct groups of the password creation task had balanced demographical values. The PC group consisted of 91 male participants and 58 female participants. The average age was 30 years (SD=9; Min=18; Max=64). We had 82 male and 70 female smartphone users with an average age of 31 years (SD=9; Min=18; Max=63) and the tablet group comprised 65 male and 84 female participants with an average age of 31 years (SD=9; Min=18; Max=64).

Results

The results are based on 450 completed questionnaires including 450 password creation tasks. Password choice was analyzed distinguishing devices while the rest of the evaluation is based on all participants.

Password Choice

We report on the influences of mobile devices on the users' password choice. To ensure the users' privacy and security, password statistics were stored separately from all other data. Therefore, the analysis is restricted to statistical tests and not merged with qualitative answers. The results are based on 149 PC users, 152 smartphone users and 149 tablet users. Our data was normally distributed and allowed for parametric tests, all post-hoc tests were Bonferroni corrected.

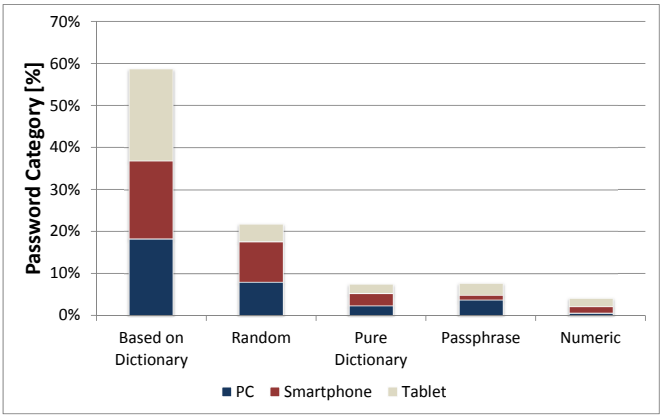

Figure 6. The distribution and number of the qualitative password categories depending on *Device*.

Category	Example	PC	Smartphone	Tablet
Dictionary	computer	11	13	10
Dictionary based	c0mputer123	82	84	98
Passphrase	ILoveComputers	17	5	13
Random	hjsd9847z	36	43	19
Numeric	1235213	3	7	9

Table 3. Qualitative password categories chosen by the participants. Most passwords were based on changed or extended dictionary words.

Table 2 reports the average number of used characters in each category, the data is visualized in Figure 5. A multivariate ANOVA comparing the mean password length revealed a significant main effect for *Device* ($F(2, 447) = 5.39, p < .05$). Post-hoc tests reveal that smartphone generated passwords (Mn=9.5; SE=0.26; Min=4; Max=25) are significantly shorter than PC-based passwords (Mn=10.6; SE=0.3; Min=4; Max=23) ($p < .05$). However, the length of passwords generated on tablets (Mn=10.5; SE=0.3; Min=3; Max=27) does not significantly differ from smartphone and PC passwords ($p > .05$). An ANOVA analyzing the means of used characters revealed a highly significant main effect for *Device* on password composition ($F(8, 890) = 3.12, p < .001$). While the post-hoc tests showed that *Device* did not have a significant impact on the use of lower case letters ($p > .05$), numbers ($p > .05$) and symbols ($p > .05$), both tablet and smartphone users used significantly fewer upper case letters ($p < .05$).

A detailed analysis of the distribution of those character groups shows that lower case letters are well established in all passwords. 96.6% of the PC passwords, 94.7% of the smartphone passwords and 93.3% in the tablet group comprise at least one lower case letter. Numbers are the second most important category as 89.9% of all PC passwords, 82.9% of all smartphone passwords and 86.6% of tablet passwords comprise at least one digit. Symbols are the least used group with 34.2% usage in PC passwords, 28.9% in smartphone passwords and 24.8% in usage tablet passwords. The significant difference becomes clear in the distribution of upper case letters. While 40.8% of the smartphone passwords and 46.3% of the tablet passwords used such characters, 63.1% of the PC passwords are composed using upper case letters.

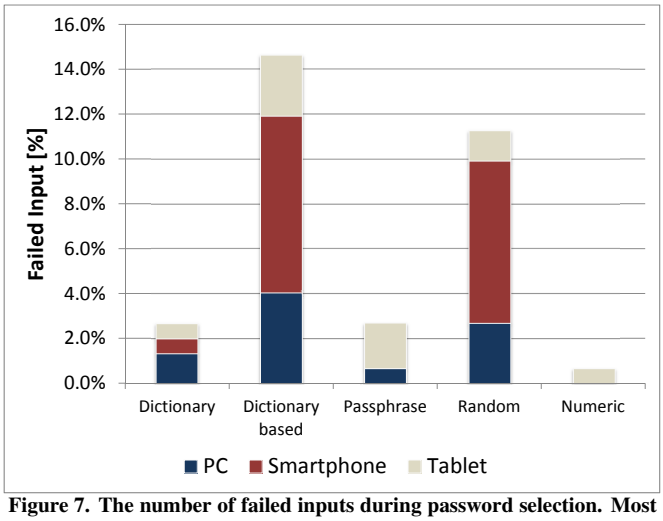

Figure 7. The number of failed inputs during password selection. Most errors happened using smartphones.

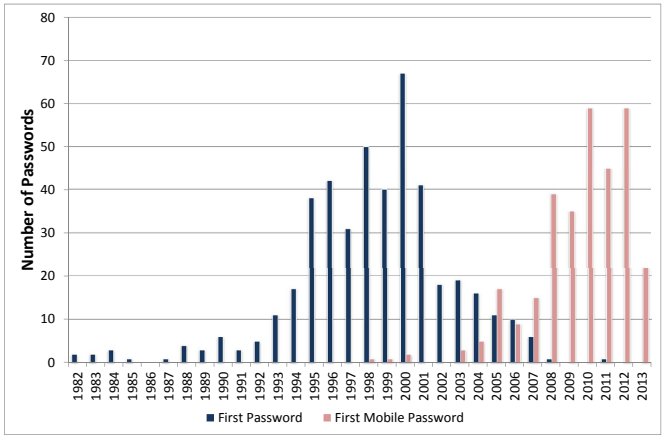

Figure 8. Years of the first password creation and the respective device. Passwords are created on PCs for several years, creation on mobile devices is a relatively new phenomenon.

We manually categorized the passwords by clustering the data. Table 3 and Figure 6 give an overview of the used password categories. All passwords fell into one of these categories. 58.7% of all passwords were based on changed dictionary words while only 7.6% of the users chose pure dictionary words (including names). A multivariate ANOVA reveals that *Device* had no significant effect on the password category ($p > .05$). Indeed, the used categories are nearly balanced between the devices. Interestingly, random passwords built the second biggest group. 24.2% of the PC users, 28.3% of the smartphone users and 12.8% of the tablet users relied on this password class.

Errors

This data is based on failed password confirmations. Within 450 password selection tasks, 48 (10.7%) errors occurred. Figure 7 gives an overview of the distribution of errors. An ANOVA showed no significant main effects for *Device* and *Category*. However, 50.0% of all errors occurred on smartphones. This indicates that such devices are error prone with an overall error rate of 15.8%. In comparison, error rates of tablets (7.4%) and PCs (8.7%) are lower.

At maximum, five consecutive errors were logged. The respective participant tried to select a password with the length of 20 using a smartphone. For tablets and PCs, a maximum of two failed attempts was logged.

Behavior & Experience

Our participants were experienced in both mobile device and alphanumeric password usage. 93.3% stated to use a smartphone on a daily basis, 62.0% use a tablet. Figure 8 shows the year of the very first password selection and the year of the first password selection using a mobile device. While most participants had used alphanumeric passwords for many years (Mn=1998; Min=1982; Max=2011; SD=4), selecting alphanumeric passwords on mobile devices became more common in the recent years. 69.3% of the participants stated that they already created alphanumeric passwords on a mobile device (Mn=2010; SD=3; Min=1998; Max=2013).

We used 10-point Likert scales ranking from one (*never*) to ten (*always*) to gather information about specific password behavior. The results revealed that most participants use password protected services on a daily basis. Most of the time, they are used on PCs or Laptops (median=8). However, usage of tablets (median=6) and smartphones (median=5) is also common. When authenticating, the participants most often type in their passwords manually. The reported median is eight for tablets and PCs and nine for smartphones. People rarely select new passwords, when they tend to use PCs (median=3) instead of mobile devices (median=2). 33.6% of our participants reported that mobile device use already influenced their password choice. Most stated to use more complex passwords on a PC than on mobile devices. When asked about their password creation behavior, 19.1% stated to use symbols on a PC, while only 13.1% stated the same for smartphones. 18.6% of our participants reported to use symbols, when creating passwords on a tablet. At the same time, 25.1% refuse from using symbols in passwords frequently used on a PC. This is true for 24.7% of tablet passwords and 43.8% of the smartphone passwords. 32.4% of our participants stated to generally use device-specific passwords. 20.0% additionally reported to use simpler versions of their desktop passwords on mobile devices.

Acceptance

Overall, the participants liked using passwords on mobile devices. 45.3% of our users' reported that passwords are their favorite way of authentication using mobile devices. 34.9% would rather use PIN and 16.4% would prefer patterns. The rest of the participants were in favor of biometric approaches (e.g. face recognition) to authenticate with external services.

However, most people are annoyed using complex passwords on their mobile device. For example, one participant stated:

> "*Passwords on mobile devices are usually easier to type in, therefore they are way more likely to get hacked. I hate entering complex passwords on mobile devices.*"

To evaluate the user perception concerning the input effort of strong passwords, we asked them to rate the following (ex-

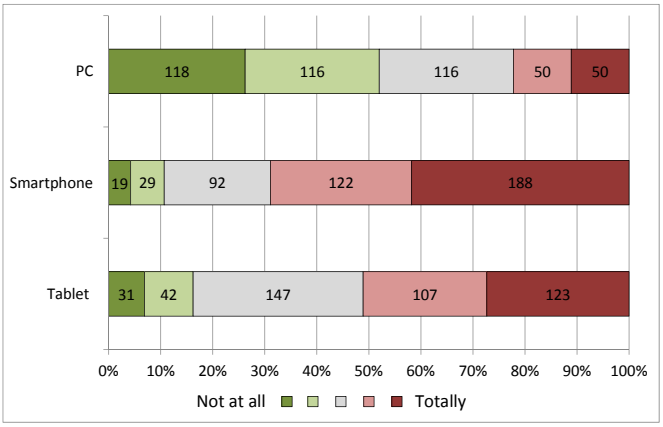

Figure 9. Participants' perception of strict password guidelines. Users are more willing to accept strict guidelines using desktop PCs.

emplary) password policy: (*a*) minimum length of 12, (*b*) no meaningful numbers or dictionary words, (*c*) minimum two digits, (*d*) minimum two symbols, (*e*) minimum one upper case letter.

The results are shown in Figure 9. People are more willing to deal with the additional effort of strict password policies, when using a PC. 52.0% state that complying with the respective guideline is not cumbersome, while only 10.7% stated the same for smartphones and 16.2% would be happy to comply with this guideline using a tablet. On the other hand, 68.9% of our participants see big usability problems in using this guideline on a smartphone.

DISCUSSION

The presented results indicate important effects of mobile device usage on alphanumeric authentication. In this section, we put our findings together and discuss their implications on the use of passwords and mobile devices.

Password Input Differs from Natural Language Input

An important question when designing the lab study was whether testing password-entry on mobile devices was different from entering "normal" text. Therefore, we analyzed both, natural language input as well as three different kinds of passwords. To no surprise, dictionary passwords performed similar to natural language as they constitute normal words. However, internet and random passwords behave completely different from natural language entry as they are significantly slower to input. This effect could be found for all devices and is consistently lower on the PC. Most interestingly, we could show that this effect is stronger for internet and random passwords. That is, the difference between entering dictionary passwords on the PC versus on the mobile devices is significantly smaller than when entering internet or random passwords.

This means that the negative effect of the mobile devices is higher when "better" or more complex passwords are used. It should be noted here that dictionary passwords should be avoided in any case and that internet and random passwords represent much more desirable passwords from a security point-of-view.

Password Creation on Mobile Devices Becomes Common

Figure 8 shows a summary of the years in which the 450 participants of the MTurk study firstly created alphanumeric passwords on desktop PCs and mobile devices respectively. The numbers nicely show that we are dealing with a rather new phenomenon but at the same time that more and more passwords are actually created on, for instance, smartphones. 69.3% of our participants stated to already have used mobile devices to create alphanumeric passwords. To the best of our knowledge, this work is the first to provide data to back up the claim that password creation on mobile devices is a new issue and worth pursuing as it comes with several new challenges.

Password Choice and Insecure Behavior

The results of the two studies show that password-entry on mobile devices can increase insecure behavior like using shorter passwords or passwords without upper case letters. For instance, in the pre-study, we could show that all three types of passwords were significantly slower to type in on both, tablets and smartphones. This effect is even stronger for random passwords as shown in Figure 3. In addition to high authentication times, authentication on mobile devices tended to be more error-prone.

The data from the online study backs up this claim. Passwords created on smartphones were significantly shorter than PC passwords and passwords on smartphones as well as tablets used significantly fewer upper case letters. Additionally, input errors were also more common on smartphones than on the other devices. It has to be noted here that we can also see that tablets are, to a certain extent, more robust to these effects than smartphones which might be partially due to their bigger size and thus, bigger keyboards.

Password Composition Strategies Depend on the Device

Our study results support the claim that smartphone use can have a negative effect on password security. In addition, 20.0% of our participants stated to use simpler versions of habitually used PC passwords. The findings further indicate that password composition strategies seem to depend on the device used to create the respective passwords. For instance, while 25.1% of participants refuse using symbols for passwords created on desktop PCs and 24.7% do the same for tablets, 43.8% of participants stated to completely leave out symbols from their password composition on smartphones. Over a third of our participants stated that mobile devices actively influence their password choice.

Possible Influence on Practical Security

We just discussed how the use of smartphones can have a negative effect on password strengths. This means that the theoretical security of the authentication is decreased. For instance, if more dictionary passwords are used, dictionary attacks are more likely to be successful again. Also, if the passwords are shorter, brute force attacks have a higher probability of success. However, the influence on security based on weak password choices can be even bigger. We argue that the results of our studies indicate that due to the effects of smartphones on password selection, the practical security can decrease as well. Slower input as well as shorter passwords

are more likely to be successfully shoulder surfed. Additionally, the increased error rates mean that password-entry has to be repeated which gives further possibilities to steal the password. That is, an attacker can more easily see the input and thus get access to the respective service.

Authentication for Mobile Devices

As mentioned before, alphanumeric passwords originated from computer environments (even before the first PCs were available). They were thus created for a very specific context. While alphanumeric passwords are seldom employed for unlocking mobile devices, many of the apps and services running on them still rely on alphanumeric passwords as their means of authentication. We assume that this is partially due to the fact that many of them come from desktop environments. Even apps that are only available for mobile devices use alphanumeric passwords.

When looking at mobile versions of websites and other desktop services, we can see that those are adapted to the mobile context, specifically attributes like screen size and input and output capabilities. This raises the question why the same does not hold for authentication. We argue that the results of our studies show that current input mechanisms provide a serious obstacle for using secure alphanumeric passwords in the mobile context, especially when it comes to smartphone use. Thus, there should be the goal to replace them with more appropriate authentication systems or to simplify the input of secure alphanumeric passwords. Simply storing passwords can open new security holes and is therefore not the perfect solution.

LIMITATIONS

Even if we are confident that both studies were thoroughly designed and conducted, there are inherent limitations concerning each of the approaches, which we would like to address in this section.

Participants of the laboratory study were asked to type in passwords which were displayed directly above the password field. Therefore, the performance analysis was not based on user-selected passwords. It is very likely that performance could improve on all devices when users type in self-selected passwords. However, our participants were highly familiar with all used devices and their respective keyboards. We argue that contrasts between the tested devices would not significantly change with self-selected passwords. In addition, since we restricted the password length to eight characters and tested three distinct classes, the results are not off-hand generalizable to all possible passwords.

We decided to utilize Amazon Mechanical Turk to collect qualitative data as this service eases acquiring large data sets. Recent work has indicated the ecological validity of online password studies [9] and MTurk was shown to be applicable to usable security studies [18]. On the downside, it makes it hard to influence the selection of participants. To ensure the quality and validity of the given answers [8], we added several control questions to the survey, asked for confirmation codes and monitored the expenditure of time. As a consequence, we were able to identify inaccurately answered surveys and excluded those from the analysis. We argue that, despite the limitations of such self-reported data, anecdotal evidence can greatly help to understand how users interact with computer systems.

The password selection task was contrived as our participants knew that they did not enroll for a real service. Consequently, users were aware of the fact that they would neither have to memorize the passwords nor would they have to use them frequently on their devices. As most users behave truthfully in such scenarios [9] and as we controlled the used devices, we assume that the data can nevertheless give valuable insights into the impact of mobile devices on password selection.

CONCLUSION & FUTURE WORK

In this work, we presented a large-scale analysis of the influences of mobile devices (tablets and smartphones) on alphanumeric passwords. By testing (a) typical password strings of various complexities and (b) directly comparing the impact of the three most relevant device classes, we were able to gain important insights into authentication performance, password creation and user behavior on mobile devices.

We showed that mobile devices have a significant impact on alphanumeric passwords. Our analysis revealed that passwords of the same complexity performed significantly slower on mobile devices and that this performance differs from natural text. As a consequence, users seem to opt for passwords which are easy and fast to enter on smartphones and tablets. For instance, user-selected passwords were significantly shorter on smartphones than the ones defined for desktop PCs. As we additionally showed that mobile devices are commonly used to select new passwords, this trend is likely to negatively affect overall password security. While memorability was one of the main limiting factors of password security for a long time (in the desktop context) [1], we argue that smartphone and tablet use has to be counted in that equation and input effort becomes more important.

Based on these results, we claim that secure alphanumeric passwords are unlikely to be used on mobile devices. As authentication on such devices becomes more common, this trend may further reduce the entropy of the user-selected password space. Therefore, we argue that web-based services should consider the requirements of mobile devices and provide adjusted authentication methods for the growing number of tablets and smartphones.

The impact of mobile device use on password authentication and password selection is a relatively new area of research. Therefore, there is a lot of room for further investigations. Future work should evaluate the effects of mobile devices under realistic conditions. Therefore, real authentication tasks with frequently used passwords should be analyzed. In addition, other effects on user behavior should be investigated. One interesting point to start with would be the analysis of password storage behavior in mobile apps in comparison to the same services on desktop PCs. If users are more likely to store passwords on smartphones due to the limited input modalities, this is likely to open new security holes.

ACKNOWLEDGMENTS

Special thanks go to Sarah Aragon Bartsch for her valuable help with the pre-study.

REFERENCES

1. Adams, A., and Sasse, M. A. Users are not the enemy. *Commun. ACM 42*, 12 (Dec. 1999), 40–46.
2. Bao, P., Pierce, J., Whittaker, S., and Zhai, S. Smart phone use by non-mobile business users. In *Proc. MobileHCI '11*, ACM (2011), 445–454.
3. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., and Bauer, G. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *Proc. MobileHCI '11*, ACM (2011), 47–56.
4. Bonneau, J. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Proc. SP '12*, IEEE (2012), 538–552.
5. Card, S. K., Moran, T. P., and Newell, A. The keystroke-level model for user performance time with interactive systems. *Commun. ACM 23*, 7 (July 1980), 396–410.
6. Chin, E., Felt, A. P., Sekar, V., and Wagner, D. Measuring user confidence in smartphone security and privacy. In *Proc. SOUPS '12*, ACM (New York, NY, USA, 2012), 1:1–1:16.
7. Corbató, F. J., Merwin-Daggett, M., and Daley, R. C. An experimental time-sharing system. In *Proc. spring joint computer conference '62*, ACM (1962), 335–344.
8. Downs, J. S., Holbrook, M. B., Sheng, S., and Cranor, L. F. Are your participants gaming the system?: Screening mechanical turk workers. In *Proc. CHI '10*, ACM (New York, NY, USA, 2010), 2399–2402.
9. Fahl, S., Harbach, M., Acar, Y., and Smith, M. On the ecological validity of a password study. In *Proc. SOUPS '13*, ACM (New York, NY, USA, 2013), 13:1–13:13.
10. Florencio, D., and Herley, C. A large-scale study of web password habits. In *Proc. WWW '07*, ACM (New York, NY, USA, 2007), 657–666.
11. Florêncio, D., and Herley, C. Where do security policies come from? In *Proc. SOUPS '10*, ACM (New York, NY, USA, 2010), 10:1–10:14.
12. Gasser, M. A random word generator for pronounceable passwords. Tech. rep., DTIC Document, 1975.
13. Grawemeyer, B., and Johnson, H. Using and managing multiple passwords: A week to a view. *Interacting with Computers 23*, 3 (2011), 256–267.
14. Hayashi, E., and Hong, J. A diary study of password usage in daily life. In *Proc. CHI '11*, ACM (2011), 2627–2630.
15. Hoggan, E., Brewster, S. A., and Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proc. CHI '08*, ACM (2008), 1573–1582.
16. Inglesant, P. G., and Sasse, M. A. The true cost of unusable password policies: password use in the wild. In *Proc. CHI '10*, ACM (2010), 383–392.
17. Jakobsson, M., and Akavipat, R. Rethinking passwords to adapt to constrained keyboards, 2011.
18. Kelley, P. G. Conducting Usable Privacy & Security Studies with Amazon's Mechanical Turk . In *Proc. SOUPS '10* (2010).
19. Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., Cranor, L. F., and Egelman, S. Of passwords and people: measuring the effect of password-composition policies. In *Proc. CHI '11*, ACM (2011), 2595–2604.
20. Malone, D., and Maher, K. Investigating the distribution of password choices. In *Proc. WWW '12*, ACM (2012), 301–310.
21. Schaub, F., Deyhle, R., and Weber, M. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proc. MUM '12*, ACM (New York, NY, USA, 2012), 13:1–13:10.
22. Schaub, F., Walch, M., Könings, B., and Weber, M. Exploring the design space of graphical passwords on smartphones. In *Proc. SOUPS '13*, ACM (New York, NY, USA, 2013), 11:1–11:14.
23. Schlöglhofer, R., and Sametinger, J. Secure and usable authentication on mobile devices. In *Proc. MoMM '12*, ACM (2012), 257–262.
24. Schneier, B. Real-world passwords. *Schneier on Security* (2006).
25. Sears, A., Revis, D., Swatski, J., Crittenden, R., and Shneiderman, B. Investigating touchscreen typing: the effect of keyboard size on typing speed. *Behaviour & Information Technology 12*, 1 (1993), 17–22.
26. Shay, R., Komanduri, S., Kelley, P. G., Leon, P. G., Mazurek, M. L., Bauer, L., Christin, N., and Cranor, L. F. Encountering stronger password requirements: user attitudes and behaviors. In *Proc. SOUPS '10*, ACM (New York, NY, USA, 2010), 2:1–2:20.
27. Ur, B., Kelley, P. G., Komanduri, S., Lee, J., Maass, M., Mazurek, M. L., Passaro, T., Shay, R., Vidas, T., Bauer, L., Christin, N., and Cranor, L. F. How does your password measure up? the effect of strength meters on password creation. In *Proc. Security '12*, USENIX Association (Berkeley, CA, USA, 2012), 5–5.
28. von Zezschwitz, E., De Luca, A., and Hussmann, H. Survival of the shortest: A retrospective analysis of influencing factors on password composition. In *Proc. INTERACT '13*. Springer Berlin Heidelberg, 2013, 460–467.
29. von Zezschwitz, E., Dunphy, P., and De Luca, A. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *Proc. MobileHCI '13*, ACM (New York, NY, USA, 2013), 261–270.