Revisiting Graphical Passwords for Augmenting, not Replacing, Text Passwords

Murat Akpulat
Gumushane University
Gumushane, Turkey
muratakpulat@
gumushane.edu.tr

Kemal Bicakci
TOBB University of Economics
and Technology
Ankara, Turkey
bicakci@etu.edu.tr

Ugur Cil
TOBB University of Economics
and Technology
Ankara, Turkey
ucil@etu.edu.tr

ABSTRACT

Users generally choose weak passwords which can be easily guessed. On the other hand, adoption of alternatives to text passwords has been slow due to cost and usability factors. We acknowledge that incumbent passwords remain difficult to beat and introduce in this study Type&Click (T&C), a hybrid scheme supporting text passwords with the graphical passwords. In T&C, users first type a text as usual and then make a single click on an image to complete the password entry. While largely preserving the login experience with the text passwords, the new scheme utilizes accumulated scientific knowledge in graphical password research (implicit feedback, persuasion during password creation, leveraging cued recall memory). The results of our user study suggest that T&C is promising for augmenting text passwords for improved security without degrading usability.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Authentication*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Graphical user interfaces (GUI)*

General Terms

Design, Experimentation, Security, Human Factors.

Keywords

Usable Security, Graphical Passwords, Passwords, Authentication.

1. INTRODUCTION

Coming in different flavors [5], graphical passwords are one among many proposed as alternative to traditional text passwords (see [8] for a large but still partial list of other proposals for replacement of passwords). Passwords prove themselves as a worthy opponent though [15] and the initial hype for replacing them seems to be dying down.

In this paper, we revisit graphical passwords with a motivation different than earlier work. Instead of aiming at replacing passwords, we explore whether we can improve security of text passwords without degrading their usability advantages and with a minimal change in users' habit of typing text as passwords. For this purpose, we augment text passwords with graphical passwords and introduce a hybrid scheme named as Type&Click (T&C in short) which incorporates both text and graphical elements.

In T&C, passwords consist of two parts; a text and a click on an image. Users first type the text part of their passwords as usual. As a second step, they make a single click on a given image. We note that these two steps are not independent. Adapting the idea from PCCP [12], the image displayed is a function of text input and changes while the user enters the text. A different text results in a different image, serving as an "implicit feedback" during login that the text part of the password is correct, which is useful information only for the legitimate user. T&C also utilizes other accumulated scientific knowledge in graphical password research including persuasion during password creation for stronger passwords and leveraging cued recall memory instead of pure recall for the second and graphical part of the password. The image serves as a cue to recall the location of the click point.

T&C is best suited for applications (e.g., access to an email account) where more secure alternatives such as OTP over SMS is not preferred due to usability or economic factors but which require more than the security achievable by today's common password practices (i.e., 6-8 characters user-chosen passwords [9]). T&C provides enhanced security against guessing attacks because the attacker needs to guess both text and graphical part of the password correctly for successful impersonation.

User study we conducted shows that compared to text passwords and PCCP [12], T&C performs better in terms of both password memorability and user satisfaction. The advantage over PCCP could be attributed to the fact that T&C asks for smaller change in users' experience with the text passwords. Our user studies also include the first - though not comprehensive - exploration of T&C regarding the important issue of recalling multiple passwords. We observed that users coped significantly better in recalling two different T&C passwords as compared to recalling two PCCP passwords.

The rest of the paper is organized as follows. We describe our hybrid scheme T&C in more detail in Section 2. We

ACSAC '13 Dec. 9-13, 2013, New Orleans, Louisiana USA
Copyright 2013 ACM 978-1-4503-2015-3/13/12 ...$15.00.
http://dx.doi.org/10.1145/2523649.2523672

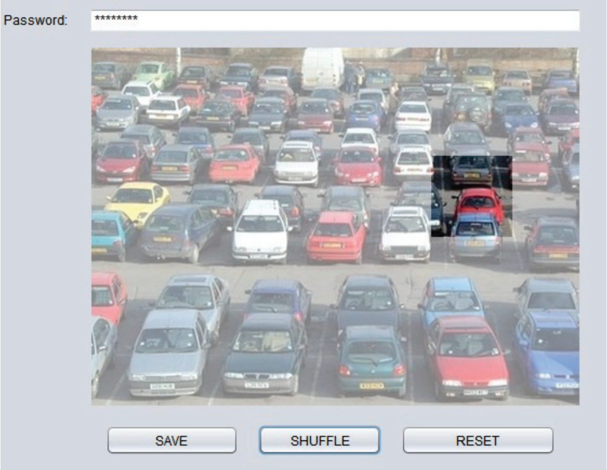


Figure 1: Password creation using T&C.

introduce the methodology and report the results of our usability evaluation work in Section 3. We discuss validation of hypothesis in Section 4. We present additional analysis and discussion in Section 5. Related work is presented in Section 6. We conclude in Section 7.

2. THE PROPOSED SYSTEM

In T&C, users first register as usual by choosing a username and entering other personal information. Then, they are directed to a second page as shown in Figure 1. On this page, initially, both password text field and the image panel underneath is empty. An image is displayed when the user starts entering the text part of the password. The displayed image is taken from an image pool (consists of 330 images in total) and is chosen by a deterministic function which takes the text password as input[1]. Hence, the image is dynamically changing as users type (and possibly erase) the text.

After user finishes typing the text part, he needs to make a click on the displayed image to finish creating the password. One of the findings in recent graphical password research [3] is that similar to the predictability problem seen in text passwords, there are popular regions called hotspots on images which users are more likely to make a click on. Hence to mitigate the problem of hotspots, PCCP [12] proposes the use of "viewport". Only the randomly located viewport (75×75 pixels) is not shaded in Figure 1 and the user could make a click only inside this viewport. In case the user does not want to make the click on the current viewport, he can change its location by clicking on the SHUFFLE button. The new location of the viewport is again determined randomly. The user is free to shuffle as much as he wants but the idea is that since shuffling requires effort, making a click on a location which is not hotspot becomes the "path of least resistance" [12]. In a sense, by use of a viewport, users are persuaded for avoiding hotspots and thus choosing a more secure password.

After making a click on the image, users could save their passwords. Then, they are directed to a password confirmation page. Before saving passwords, users can choose to start over at any time using the RESET button. Both password confirmation and login screens are similar to the screen given in Figure 1 except there is no viewport and SHUFFLE button. On the login page, above the password field there is a username field and user completes the login task on a single screen. During both for confirmation and login, users are not required to click on exactly the same location chosen during password creation. There is a tolerance region (19×19 pixels in size) centered on the click location.

In the design of T&C, other than making the password more secure against a guessing attack by contributing additional entropy into the users' passwords, there are three purposes of the graphical element. First, it provides an implicit feedback for the correctness of the text part of the password[2]. Second, the image triggers the memory and serves as a cue to recall the click point. Third, users which fall victims of a phishing attacks, could not reveal graphical part of their passwords due to not seeing the cueing image [14, 20].

Our current work is motivated by recent research work on graphical passwords together with the reluctance in practice on adopting new graphical password proposals. Passwords have this dilemma. On one hand, it has serious security problems. On the other, there is no solution on the horizon which improves their security while preserving benefits of passwords such as "Nothing-to-Carry" and "Easy-to-Learn" as described in [8]. We believe that T&C would be a viable solution to solve this dilemma since it requires a minimal change in users' password behaviors and it renders significantly more difficult guessing the password.

Having said that, we do not advocate that T&C is a better choice in all password use cases. For instance, it makes no sense for a site which asks passwords from its users before they download a white page to replace their long-standing login procedure because of their minimal (if any) security requirements. As mention previously, we introduce T&C for applications which require more security than the current password practices could provide but where more secure alternatives such as two factor authentication is costly, or otherwise not available. We also note that although not investigated in our user studies it seems reasonable to assume that on the down side T&C inherits the drawback of graphical passwords of being more vulnerable to shoulder-surfing attacks[3]. Other forms of password capture attacks [5] (malware, social engineering, etc.) are out of scope in our work.

3. USER STUDIES

3.1 Hypothesis

Our specific hypothesis with respect to usability of T&C were:

1. Participants will have higher recall success rates with T&C than with text passwords and with PCCP.

[1] The mapping is not one-to-one but many-to-one due to finite number of images available in our database. Nevertheless, the probability of an input different than the text part of the password leading to the same image is considered to be negligible (i.e., 1/330).

[2] In user studies, we observed that some users get benefit from not only the last image shown but even the intermediate images in the sequence as a feedback that they are in the correct path while typing their text passwords.

[3] In T&C, images corresponding to the prefixes of the password are also revealed to anyone who might be watching.

2. Participants will find T&C more secure and more usable than text passwords and PCCP.

3. Participants will have higher success rates with two T&C passwords than two text passwords and than two PCCP passwords.

To test these hypothesis, we first implement text passwords and PCCP together with T&C. The implementation of T&C was described in Section 2. Before going into other details, we summarize how PCCP [12] works, below:

PCCP is a successor of CCP and adds persuasion to it by the use of viewport which is explained in the description of T&C. In CCP, the passwords consists of one click point per each image (in the original implementation there are five images in total hence five click points form the password [12]). The image displayed is based on the click location on the previous image. A different click point results in a different image hence for the legitimate user the image sequence serves as an implicit feedback for the correctness of the password.

In our implementation, we choose the parameters same as in original implementation [12]. The only difference is the number of images and thus the number of clicks forming the password. We use 3 instead of 5 images for reasons described shortly. For the sake of consistency, other parameters are chosen as same in PCCP and T&C (image size: 451 × 331 pixels, tolerance region: 19 × 19 pixels, viewport: 75 × 75 pixels).

3.2 Equalizing Security of Passwords

In the design of user studies on new password proposals with the goal of comparing to text passwords, there is one important issue that should not be overlooked. As the trade-off between security and usability of passwords is well-known, a fair usability comparison of different password schemes requires that the passwords created and used in these schemes provide approximately the same level of security. For instance it is not fair to compare a text password consists of six-eight characters with a PCCP password which uses five images. The later provides a password entropy of around 43 bits whereas the former achieves much less [11]. This is because users tend to choose predictable passwords and the password space in practice falls short of what is achievable in theory.

On the other hand, it is not easy to establish straight rules to make equal the security of passwords chosen under different policies or created in different schemes. Unfortunately, researchers are still discussing the proper metric to evaluate the security of passwords. While the recent edition of NIST guideline builds its entire recommendation on the use of Shannon entropy as the security metric, researchers are questioning the suitability of this choice [6]. Bonneau has written that *"Shannon entropy has no direct correlation to guessing difficulty"* [6].

Although we acknowledge that Shannon entropy is not the most appropriate metric to evaluate the security of passwords, we opt to use NIST guideline[4] due to lack of any alternative for our purposes. We are unaware of any other guideline that sets a convenient framework to measure security of passwords created under different security policies. For instance, statistical metrics put forth by Bonneau require measuring the probability of individual passwords using a large password data set which cannot be obtained in any reasonable size user study [7].

To better grasp the basic idea underlying the design rationale of our user study, consider the following scenario.

Suppose you are a system administrator of a web site where you recently get a lot of complaints about identity thefts where you identify them to occur due to online guessing attacks [5]. Actually, you have already set seemingly a decent password policy. All passwords should have a minimum length of 8 characters. Based on the NIST guideline [11], however this provides roughly an effective password space of only 18 bits. To have stronger passwords, there are many alternatives including:

1. Set a higher minimum length restriction. According to NIST guideline, each additional character after the eighth character adds 1.5 bits of entropy to the password. So a password of length 14 has an estimated entropy of 27 bits [6].

2. Keep the minimum eight character requirement and ask users to click on an image in addition as part of his password. More precisely stated, use our new proposal, T&C. Here, if we assume that the clicks are uniformly distributed due to use viewport[7], then the math is simple for the calculation of additional password entropy. The click point adds approximately 9 bits of entropy as eq.(1) shows. So the password would have again an entropy of 27 bits in total.

$$\lceil \frac{451 \times 331}{19 \times 19} \rceil \approx 2^9 \tag{1}$$

3. Finally, we could give a try for a more radical change and use PCCP instead of text passwords. With the same assumption that hotspots are not exploitable, PCCP with 3 images would give us an entropy of approximately 27 bits.

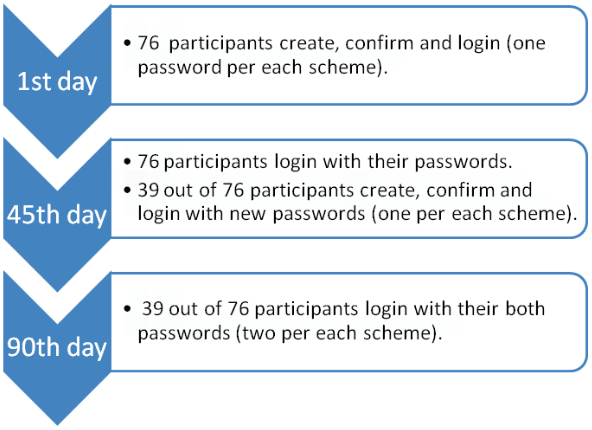
On the bottom-line, the explanation above serves as the preliminary security analysis required for a fair usability comparison. Other than the three alternatives given above, there are many other ways to establish more password entropy e.g., requirement of having special characters in the password, a dictionary check to ensure that password is not predictable with a dictionary attack, etc. For practical reasons, in our usability study we choose to content ourselves with the above three. The reason of choosing the last two alternatives is obvious. We decide to include the first alternative given above since minimum password length restriction is widely used and a well-known type of password policy among Internet users.

[4] According to a recent study [17], NIST guideline succeeds at its stated purpose of providing a "rough rule of thumb".

[5] There are many other ways to attack password-protected accounts (phishing, malware, etc.) but password guessing attacks are still prevalent and news still abounds for such attacks [2].

[6] Forcing users to remember passwords complying even stricter requirements can accommodate even weaker passwords. But on the overall, we assume NIST guideline is correct.

[7] Previous user studies on PCCP [12] shows that PCCP click-points have a flatter distribution than the other schemes. We initially assume T&C inherits the same property due to



- 76 participants create, confirm and login (one password per each scheme).
- 76 participants login with their passwords.
- 39 out of 76 participants create, confirm and login with new passwords (one per each scheme).
- 39 out of 76 participants login with their both passwords (two per each scheme).

Figure 2: Chronology of Lab Sessions.

3.3 Methodology

Our study included three lab sessions scheduled as presented in Figure 2. First session was completed by 76 participants. All participants (29 female, 47 male) were the students of Kelkit Aydın Doğan Vocational School, Department of Computer Technologies. At the time of the user study, one of the authors (first author) has been teaching courses registered by the participants. Participants were comfortable in using web sites requiring passwords. None of them had previously taken a security course or used graphical passwords. Our experimental work was performed with the approval of Kelkit Aydın Doğan Vocational School, Human Subjects Ethics Committee.

In the first session, participants were invited to the computer laboratory in groups of 6-7 students. The computer lab has desktop machines all having standard keyboard and mouse. For around 5 minutes, they were informed about the following issues:

1. The purpose of the user study (we did not mention T&C was our own design so as to not bias participants).
2. The importance of creating and using passwords not previously used in order to obtain ecologically valid results.
3. The request for pretending that the passwords they create and use would be the passwords of an email or an e-banking account.
4. The request for not writing down their passwords.
5. The demand that the text part of the T&C password should not be same or similar to the text password.

Then, for around 10 minutes, participants were free to try and get familiar with the three schemes.

We used a within-subjects design so that participants were asked to create three accounts, one per each scheme. Hence, all participants created and confirmed three passwords and login with them; one text password, one T&C password and one PCCP password in a counter-balanced fashion. Between password confirmation and login, participants answered part of the questionnaire. During login, they were free to try any number of passwords they want but we consider a login as successful if the user is able to login in first trial.

Table 1: Success rates in the first experiment.

	Success Rate First Login	Success Rate Second Login
PCCP	48/76	23/76
	63.16%	30.17%
T&C	62/76	49/76
	81.58%	64.47%
Text Password	58/76	33/76
	76.32%	46.05%

The first session was completed by telling participants that they would be re-invited 45 days later to login to each system again. The reason for such a long break between sessions was as follows. Users have many accounts for different purposes and they often do not access some of these for a long period of time. In fact, these infrequently visited accounts have the most serious password-related usability problem. Remembering a password is hardly an issue for a user who logins with it regularly. As a result, we consider this generous longitudinal factor quite reasonable.

In the second session which was conducted 45 days after the first one, participants were asked to login to three systems, again. With all three schemes, they completed successfully the login task (in first trial), had eventual success or decided to give up. Our first experiment was completed by a post-task questionnaire applied to all participants.

Participants were asked whether they want to create a second account for each system voluntarily. 39 of them responded positively. These participants were instructed not to use the passwords they have for the first accounts.

These 39 participants created and confirmed three new passwords and login with them. As a result, 39 users had two passwords per each of three schemes. Participants having two accounts per each scheme were re-invited 45 days later (90 days after the first session) to login to each system again.

In the third and last session which completes our second experiment, 39 participants attempted to login with their all six passwords (they login for the third time with three of these passwords and for the second time with the other three). Again, they were free to stop trying to login at any time they want.

3.4 Experimental Results

Here, we report the results of the experiments. We defer analysis and discussion of these results to the next section.

For each of three schemes, the following data were collected as performance measures for memorability and usability: login success rates, times for password creation, times for password confirmation, times for logins, number of shuffles.

Table 1 presents the results for first and second login success rates of the first experiment for three schemes. An attempt is considered as unsuccessful if one of the following conditions hold: (i) RESET button is used; (ii) part of the text password (or text part of T&C password) is erased; (iii) the password is incorrect when LOGIN button is pressed.

Table 2 presents timing information for the first exper-

use of viewport (this assumption is validated after the user study, see Security Analysis in Section 5.).

Table 2: Timing information for the first experiment (in seconds).

		Password Creation	Password Confirmation	First Login	Second Login (45 days later)
PCCP	Average	43.4/43.8	13.5/17.5	17.5/20.7	32.3/43.9
	Median	45.5/35.5	12.0/16.0	16.5/17.0	28.0/33.0
T&C	Average	33.7/35.0	10.8/12.9	18.8/19.6	31.5/46.0
	Median	25.5/26.5	10.0/11.0	18.0/18.5	25.5/30.0
Text Password	Average	12.9/16.0	9.7/9.9	16.3/17.0	21.6/26.1
	Median	9.0/9.5	9.0/9.0	14.0/15.0	18.0/18.0

iment. Average and median of the timing information of users who were successful is provided first. Second value in each cell of the table is for all users except those who unable to login and give up. The reported values are the times taken between the first typing (or first click) of the password and the click on the LOGIN button (or hitting the Enter key).

Table 3: Success rates in the second experiment.

	Success Rate Third Login First Passwords	Success Rate Second Login Second Passwords
PCCP	18/39	7/39
	46.15%	17.95%
T&C	25/39	28/39
	64.10%	71.79%
Text Password	22/39	24/39
	56.41%	61.53%

Table 3 presents the results for third login success rates with the first passwords and second login success rates with the second passwords. Success rates for first login with second passwords are not reported due to the similarity of the results for first login with first passwords.

Table 4: Timing information for the login tasks in the second experiment (in seconds).

		First Password	Second Password
PCCP	Average	24.0	29.4
	Median	24.0	26.0
T&C	Average	25.4	22.8
	Median	19.0	19.5
Text Password	Average	16.3	17.6
	Median	14.5	16.0

Table 4 presents timing information in the second experiment for the login tasks with the first and second passwords for users who were successful (in their first attempts).

Table 5 presents how many times the SHUFFLE button is clicked during password creation in the first and second experiment. Note that there are three pictures and a single picture in PCCP and T&C, respectively.

3.5 Questionnaire Responses

Table 6 and Table 7 summarizes the responses we collected via post-task questionnaire. Table 6 presents number of user preferences per each scheme (only one scheme is chosen in each question). Table 7 reports the results of Yes/No type questions.

Table 5: Number of shuffles in the first and second experiments.

	PCCP			T&C
	Picture 1	Picture 2	Picture 3	
Average	13.7/17.0	9.1/18.3	7.6/9.4	15.3/33.0
Median	5/13	5/8	4/5	9/12

Table 6: Questionnaire Responses

Question	T&C	PCCP	Text
1. Which method do you prefer for creating and using passwords?	51	15	10
2. In which method do you create your password the easiest?	37	13	26
3. In which method do you create your password the fastest?	23	15	38
4. Which method do you find the most secure?	41	25	10
5. Passwords in which method do you think are the easiest to remember?	43	11	22
6. For creating a password for a bank account which method would you use?	39	25	12

4. VALIDATION OF HYPOTHESIS

Below, we discuss the validation of hypothesis based on the results of the user study.

1. **Participants will have higher recall success rates with T&C than with text passwords and with PCCP.** *Hypothesis supported.* In the first experiment, as compared to text passwords and PCCP, there were less number of users who could not successfully login with their T&C password when they attempted to login 45 days after they create their passwords. To investigate whether there were any significant differences between schemes we have conducted the Friedman test which is the non-parametric alternative to the one-way ANOVA with repeated measure. The dependent and independent variables were login success rate and type of authentication schemes, respectively. 64.47% recall success rate of T&C was significantly higher than the 46.05% recall success rate of text passwords and 30.17% recall success rate of PCCP ($\chi^2(2) = 24.571$, $p = .000005$). Post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significant level set at *p=.025.* There were statistically significant differences between T&C and PCCP *(Z=-4.914, p=.000001)* as well as be-

Table 7: Questionnaire Responses (Yes/No type questions)

Question	Yes	No
1. Is the text part of your T&C password different than your text password?	58	18
2. Are the passwords created in the user study different than your earlier passwords?	74	2
3. In T&C method, does the picture make it easier for you to remember the text part?	60	16

tween T&C and text passwords *(Z=-2.921, p=.003)*.

Table 8: Test Statistics for Hypothesis 2

Question	Chi-Square	df	Asymp. Sig.
1	39.500	2	0.000
2	11.395	2	0.003
3	10.763	2	0.005
4	18.974	2	0.000
5	20.868	2	0.000
6	14.395	2	0.001

2. **Participants will find T&C more secure and more usable than text passwords and PCCP.** *Hypothesis supported.* Table 6 shows that users' perception regarding the usability and security of T&C is better than those of text passwords and PCCP[8]. We confirm these results are statistically significant using non-parametric chi square test. See Table 8.

3. **Participants will have higher success rates with two T&C passwords than two text passwords and two PCCP passwords.** *Hypothesis partially supported.* We perform the statistical experiments in two ways.

   First, in aggregate, 68% of T&C passwords were recalled successfully when each participant had two T&C passwords (together with two text passwords and two PCCP passwords). This success rate was higher than the 59% aggregate success rate of text passwords and 32% aggregate success rate of PCCP. The difference is statistically significant ($\chi^2(2) = 45.500, p < .05$).

   Second, if we compare third logins with first passwords and second logins with second passwords individually, then we see that there is a significant difference for the third logins ($\chi^2(2) = 10.571, p = .005$). After Bonferroni correction applied (p was set to .025), there is a significant difference between T&C and PCCP ($Z = -2.646, p = .008$) but not between T&C and text passwords ($Z = -1.732, p = .083$). For second login with second passwords, T&C vs. PCCP has a statistically significant difference ($Z = -4.583, p < .025$) but there is no significant difference between T&C and text passwords ($Z = -2.000, p = .046$).

5. ANALYSIS AND DISCUSSION

In the previous section, we present the results with respect to validation of our hypothesis. In this section we interpret and discuss the results of user study, itemized as follows:

1. Habituation of users to standard text based passwords leads to usability challenges for alternatives. Higher scores of text passwords over PCCP in our questionnaire (Questions-2,3,5 in Table 6) support this observation. On the other hand, as mentioned earlier, T&C performs better than text passwords with respect to perceived security and user satisfaction.

2. Users achieve higher recall success rates with T&C than with text passwords. We think that one major factor for this difference is the implicit feedback property of T&C. During the experiments, we observed that a significant portion of users verify the correctness of text part of their T&C passwords by checking whether the image displayed is the one they have seen.

3. We also attribute (at least partially) higher recall success rates of T&C over PCCP to the habituation effects. T&C asks less change in user habits as compared to PCCP. The results of our initial exploration of multiple password interference effects also support this argument.

4. Times for password creation, confirmation and login are comparable in T&C and PCCP. However, text passwords perform significantly better than both of these schemes. This is expected since typing through standard keyboard is faster than pointing and clicking with mouse[9]. In the experiments, we observed that users spend some time for verifying the text part of the password before clicking on the image, which also contributes to the relatively longer timings with T&C.

5. Previous work on PCCP applies a viewport size of 75 × 75 pixels. In our experiments, we adopt this size both for PCCP and T&C. We observed that some of the users assume that tolerance region and viewport sizes are equal and hence make mistakes during password confirmation and login.

6. Users make a single click hence issues such as geometric patterns [12] (e.g., clicks forming line segments) are not applicable in T&C. On the other hand, users click on the SHUFFLE button significantly more if we compare it to the number of shuffles per image in PCCP. We hypothesize that users have a shuffle budget. While the budget is spread out on multiple images in PCCP, they tend to spend all of it in the single image of T&C. In the subsection 5.2, we analyze security implications of this behavior.

5.1 Limitations and Future Work

Due to limited number of participants in our experiments, we preferred to use a within-subjects design methodology, which brings potential learning effects and interference between passwords harming ecological validity of our results.

[8] There is one exception though. Text passwords achieve a score higher than T&C in question-3 of Table 6 which is about user perception of password creation time. The timing information confirms that participants created text passwords the fastest.

[9] We speculate that timings may change in favor of T&C and PCCP when touch-screen devices are used because of their less friendly text input methods.

We acknowledge this limitation. In fact, 18 out of 76 participants have indicated that the text parts of their T&C passwords are similar to their text passwords. We did not analyze further the extent of using similar passwords.

As another limitation we should mention that users are recalling initially only one password per scheme in our study. Later, only a self-selected group have another set of passwords, which might make the memorability easier than it might be otherwise.

As a future work, we plan to conduct a between-subjects web-based study (e.g., using Mechanical Turk) which incorporates higher number of participants.

The results of our experiments also lead to the following research question which we plan to validate by a future user study: If minimum length requirement for the text part of the T&C passwords is exactly same as in the policy applied to the text passwords, which scheme performs better in terms of recall rates? In other words, how do users perform if implicit feedback property of T&C comes together with the additional burden of recalling the click location?

5.2 Security Analysis

The finding in previous work [12] was that click-points in PCCP is nearly indistinguishable from those of a randomly generated simulation dataset, thanks to the persuasion through the use of viewport. In T&C, users preferred to click on the SHUFFLE button more as compared to the number of its use per image in PCCP. In this section, we analyze whether this behavior lead to a situation where click-points are distributed in some recognizable manner.

We first look at how click points are distributed along the x- and y-axes in the first and second experiments and how they compare against randomly-generated datasets (Figure 3). Maximum and minimum median values of the simulated datasets (blue and red lines) are calculated using 100 simulated datasets. In T&C, click-points are quite uniformly distributed along x- and y-axes. The medians of click point distributions fall inside of the random range of the simulated datasets.

To analyze the security effects of user choices further, we investigate whether click-points within datasets are clustered around some coordinates or they are randomly dispersed. For this purpose similar to the previous analysis of PCCP [12], we get help from the J-statistics. For a given radius, J-function measures the clustering of points. J=0 and J=1 indicates clustering at the same coordinate and random dispersion (no clustering), respectively. Figure 4(a) and Figure 4(b) presents the J-function values for the datasets of experiment 1 and experiment 2, respectively (for four different types of J-function calculation methods, which have slightly different output values). Since a radius of 9 pixels is a good approximation for 19×19 tolerance regions, we should consult to the $J(9)$ values. We observe that $J(9)$ is very close to 1 in both figures, thus we conclude that there is no clustering effect in our experimental datasets.

Since the image clicked on is a function of the text part of the password, it is not straightforward to exploit image-specific hotspots in T&C. On the other hand, some regions (e.g., corners) could be clicked more independent of which background image is used [12]. We create heat-maps to depict the distribution of participants' click-points on the no-image background for both first and second studies (see Table 5). Color bands, from cyan to pink, represent varying intensities of click-point concentration (more pinkish areas mean more click-points). In addition to heat-maps, to explore hotspots further, we calculate the rough estimate values of password entropy for both observed and simulated datasets via the formula given in [4]. We found out that estimated password entropy value for observed dataset is completely between the maximum and minimum entropy values of the simulated dataset so this result gives an evidence that hotspot does not skew the click-point distribution.

As a summary of the security analysis, we report that no evidence is found regarding the negative security effects of user choices (i.e., use of SHUFFLE button) in T&C.

On the other hand, regarding shoulder surfing attacks, we note that if the images are observed and/or recorded, an attacker can step through the progression of images as he types to determine the exact password the user typed *i.e.*, at each character position, the attacker types until he gets to the image displayed at that point in time. He then does this for each successive key. This is a security weakness of T&C.

6. RELATED WORK

We overview related work under two headings; hybrid password schemes and (usability) comparisons to text passwords.

6.1 Hybrid Password Schemes

Earlier work on hybrid password schemes combining text and graphical passwords are briefly discussed as follows. Jermyn et al. [16] proposed a mechanism in which text passwords are augmented by some minimal graphical capabilities. The graphical assistance enables the decoupling of temporal order of input and the position in which characters are input. TwoStep [20] is a combination of text passwords and recognition based graphical passwords. Its implementation as a password manager was used by more than 4000 users. However no user study on TwoStep was reported.

Singh et al. [19] conducted a user study which compares the combination of text passwords and CCP [5] to text passwords (and PCCP [12]). They found that text passwords performed the best in terms of success rates and entry times. In their design of the hybrid scheme, text and graphical steps are not related. First, the text part of the password is entered and confirmed. Only if this first step succeeds, users are directed to the second graphical step. Another hybrid scheme which has independent steps for the entry of text and graphical parts is proposed by Khan et al. [18]. Phorcefield [14] could also be considered as a hybrid password scheme but its main motivation is protection against phishing attacks rather than increasing password strength.

6.2 Comparisons to Text Passwords

In recent years, many variants of graphical passwords were proposed [5]. Surprisingly, despite the fact that the rival is the traditional text-based passwords, only a few attempted to make a comparison with text passwords through user studies. Table 9 summarizes the results of user studies conducted to make a comparison between text passwords and different types of graphical password schemes. We infer from this comparison table that entry times for text passwords were consistently reported to be shorter than the times for graphical passwords, which is a result we confirm in our user study. On the other hand, mixed results were reported for

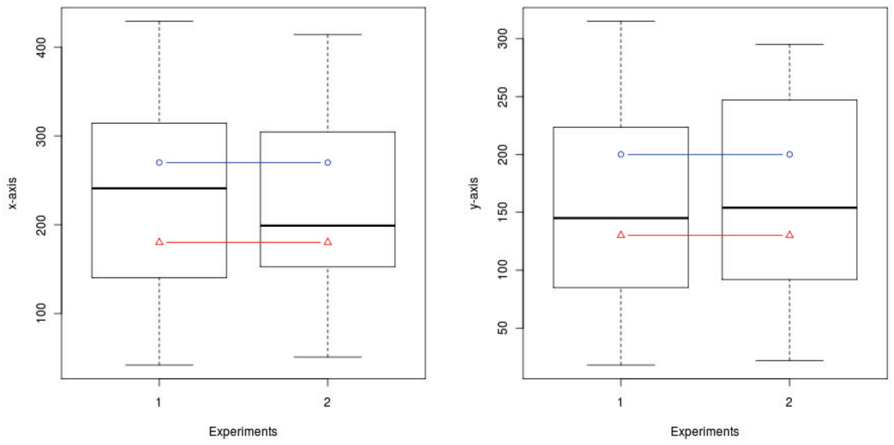
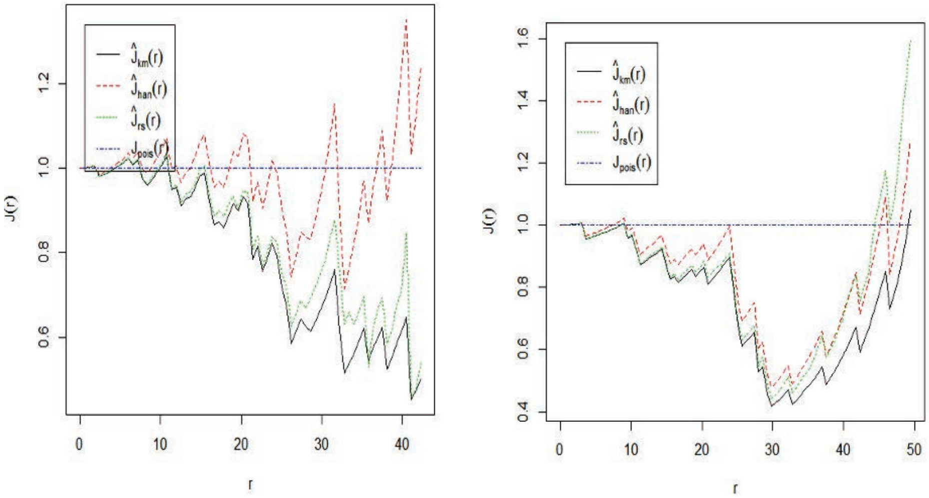


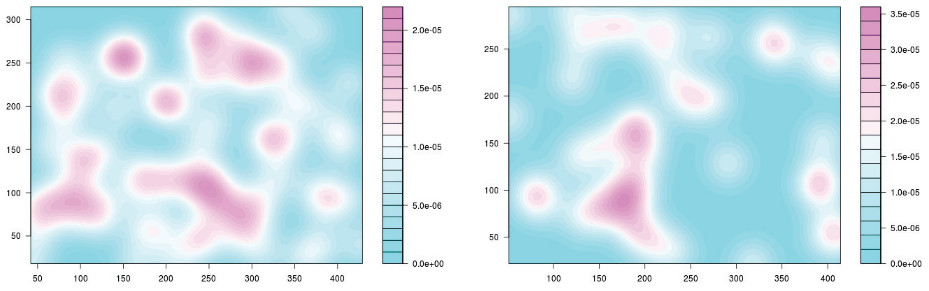
Figure 3: The box plots showing the distribution of click-points along the x-axis and y-axis of the image, respectively for the first and second experiments. The red line (with triangles) and the blue line (with circles) represent the max and min of median values for the simulation sets.



(a) The first experiment

(b) The second experiment

Figure 4: *J*-function values of the click-points in collected password datasets of T&C.



(a) Heat-map for the first experiment

(b) Heat-map for the second experiment

Figure 5: The hot-spot images for first and second experiments.

the login success rates. The mismatch between security levels of passwords compared is also worth mentioning.

7. CONCLUSION

In this paper, we applied the insight gained from graphical passwords to introduce T&C as a hybrid scheme combining text and graphical elements. During password creation, T&C adopts the idea of persuasion to influence users for stronger passwords. For login, users first enter their text passwords as usual. The image underneath changes as users type the password, providing an implicit feedback for the correctness of the text. The image also serves as a cue to recall the second part of the password. Users make a single click on the image and this completes the password entry.

To evaluate its usability, we compared T&C to PCCP and text passwords through a three-session lab study with 76 participants. In the user study, we set the parameters of the schemes so that we have passwords with approximately same level of security. We observed that recall success rates were significantly higher with T&C than with text passwords and with PCCP. Participants found T&C more secure and more usable than text passwords and PCCP. In T&C, users preferred to click on the SHUFFLE button more (as compared to number of its use per image in PCCP) but we found no evidence that this behavior led to a situation where click-points are distributed in some recognizable manner.

A common goal in password research is to increase password space while keeping usability impacts minimal. Our user study results suggest that the hybrid scheme T&C has a potential to achieve this goal.

8. REFERENCES

[1] The science behind passfaces. www.realuser.com/published/ScienceBehindPassfaces.pdf. Accessed: 03/03/2012.

[2] Weak password brings ’happiness’ to twitter hacker. www.wired.com/threatlevel/2009/01/professed-twitt. Accessed: 03/03/2013.

[3] K. Bicakci, N. Atalay, M. Yuceel, H. Gurbaslar, and B. Erdeniz. Towards usable solutions to graphical password hotspot problem. In *Computer Software and Applications Conference, 2009. COMPSAC’09. 33rd Annual IEEE International*, volume 2, pages 318–323. IEEE, 2009.

[4] K. Bicakci, N. B. Atalay, M. Yuceel, and P. C. van Oorschot. Exploration and field study of a browser-based password manager using icon-based passwords. In *Workshop on Real-Life Cryptographic Protocols and Standardization*, 2011.

[5] R. Biddle, S. Chiasson, and P. C. van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys 44(4)*, 2011.

[6] J. Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *IEEE Symposium on Security and Privacy*, pages 538–552, 2012.

[7] J. Bonneau. Statistical metrics for individual password strength. In *20th International Workshop on Security Protocols*, 2012.

[8] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *IEEE Symposium on Security and Privacy*, pages 553–567, 2012.

[9] J. Bonneau and S. Preibusch. The password thicket: technical and market failures in human authentication on the web. In *Proc. WEIS*, 2010.

[10] S. Brostoff and M. Sasse. Are passfaces more usable than passwords? a field trial investigation. *People and Computers*, pages 405–424, 2000.

[11] W. Burr. Electronic authentication guideline. *NIST special publication*, 800:63.

[12] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. C. van Oorschot. Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism. *IEEE Trans. Dependable Sec. Comput.*, 9(2):222–235, 2012.

[13] R. Dhamija and A. Perrig. Déjà vu: A user study using images for authentication. In *Proceedings of the 9th conference on USENIX Security Symposium - Volume 9*, SSYM’00, pages 4–4, Berkeley, CA, USA, 2000. USENIX Association.

[14] M. Hart, C. Castille, M. Harpalani, J. Toohill, and R. Johnson. Phorcefield: a phish-proof password ceremony. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 159–168. ACM, 2011.

[15] C. Herley and P. C. van Oorschot. A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy*, 10(1):28–36, 2012.

[16] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. The design and analysis of graphical passwords. In *Proceedings of the 8th USENIX Security Symposium*, pages 1–14. Washington DC, 1999.

[17] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 523–537. IEEE, 2012.

[18] W. Khan, Y. Xiang, M. Aalsalem, and Q. Arshad. A hybrid graphical password based system. In *ICA3PP 2011 Workshops, Part II, LNCS 7017*, pages 153–164, 2011.

[19] C. Singh, L. Singh, C. Singh, and L. Singh. Investigating the combination of text and graphical passwords for a more secure and usable experience. *International Journal of Network Security & Its Applications (IJNSA)*, 3(2), 2011.

[20] P. C. van Oorschot and T. Wan. Twostep: An authentication method combining text and graphical passwords. In *MCETECH*, pages 233–239, 2009.

Table 9: Results of user studies comparing text and graphical passwords. The entropy reported for graphical schemes is the theoretical maximum whereas the entropy of (text) passwords is calculated using NIST formula [11]. For Story [10], the mean of success rates for a ten-weeks period was reported. Successful login rate of Passfaces was reported as less than a third of the rate with text passwords [1].

Name of Method		Study Type	Design	Entropy (bits)	Password Assignment	# of Participants	# of Passwords Per User	Time for Create (sec)	Time (First Login) (sec)	Time (Second Login) (sec)	Success Rate (First Login)	Success Rate (Second Login)
Story [10]	Passwords	Field Study $n \times 10wk$	Within Subject	n.a.	User Choice	34	1 for each	n.a.	n.a.	n.a.	$\approx 3k\%$	
	Passfaces	Field Study $n \times 10wk$		12				n.a.	n.a.	n.a.	$k\%$	
Déjà Vu [13]	Passwords	Lab Study $2 \times 1wk$	Within Subject	14	User Choice	20	1 for each	25	18	24	95%	70%
	PIN	Lab Study $2 \times 1wk$		14				15	15	27	95%	65%
	Art	Lab Study $2 \times 1wk$		16				45	32	36	100%	90%
	Photo	Lab Study $2 \times 1wk$		16				60	27	31	100%	95%
PCCP [12]	Passwords	Lab Study $2 \times 2wk$	Between Subject	18	User Choice	34	6 for each	26	10	10	99%	31%
	PCCP	Lab Study $2 \times 2wk$		$43 - 73$	System Persuasion	83		91	18	27	99%	32%
PCCP (Web)	Passwords	Field Study $4 \times 1wk$	Between Subject	36	User Choice	21	3 for each	11	6	6	100%	56%
	PCCP	Field Study $4 \times 1wk$		43	System Persuasion	24		68	13	20	99%	67%