International Journal of Computing Academic Research (IJCAR)
ISSN 2305-9184 Volume 1, Number 2 (December 2012), pp. 66-78

http://www.meacse.org/ijcar

On the Deployment of Password Hints Using Pre-Attentive Visual Priming for One-Time Passwords

Kenneth Revett[1] and Ahmed Selim Bahaa[2]

[1] Faculty of Informatics and Computer Science
Loughborough University
[1 2]The British University in Egypt campus
El Sherouk City, Egypt

Abstract
Password based security is still the most prevalent form of controlling access to trusted resources accessed through computers. There are several difficulties associated with password based systems, the predominant one being password memorability. The average person has approximately 15 passwords to maintain, which engenders a significant cognitive burden if passwords are selected and utilised properly. One potential solution to the memorability issue is to provide users with password hints. For instance, literal hints in the form of a displaying a subset of the actual password characters *in situ* during have been deployed commercially. Although potentially effective, this approach compromises the password coverage space, effectively weakening the password. Further, this approach may render the password susceptible to shoulder surfing and other means of surveillance. In this work, a compromise was sought between enhancing password memorability while reducing the likelihood of successful surveillance based attacks. The scheme deployed in this work is based on a one-time password scheme (OTP). To enhance memorability, password hints are utilised, which are deployed in the form of pre-attentive hinting. The question explored in this work is whether pre-attentive hinting is sufficient to enhance memorability, without rendering the approach susceptible to a surveillance based attack**.**

Keywords: cognitive biometrics; false rejection rate; iconic memory; memory span; one-time passwords; pre-attentive visual priming; short term memory.

Introduction

Passwords are the most prevalent form of user authentication for access to the vast majority of trusted computing facilities. Passwords which are user selected tend to be at best moderately strong – in that they typically contain semantic information that weaken the password in that they may be subject to successful dictionary attacks [1]. Furthermore, users tend to have on average 15 accounts that require password based authentication. Users will typically use very similar passwords – ones with a common root and append digits such as '123' to the end of the root. The typical scenario indicates that users will generate a small subset of passwords that they will utilise for multiple authentication portals. Therefore, if one can successfully crack one password – they will more than likely gain access to several restricted access portals. One countermeasure to this problem is password hardening, where passwords are fortified with certain features that render them less prone to a brute force off-line attack. This feature should be enforced when a user selects a password – the system will generate a hardness score, and will only accept a password with a

suprathreshold score. This process in itself renders passwords less memorable as the user will typically be asked to make at least 2 changes to their password in order to reach the desired hardness threshold. For example, when asked to change their password to meet the hardness criteria, the users may provide new passwords which are typically minor variations to a root password. This approach may in turn lead to a reduction in the degree of password solidification. Furthermore, this scheme generally does not obviate the issue of duplicate passwords across multiple authentication portals. Notwithstanding these issues, a very intriguing question remains: given a hardened password of say seven characters, why do humans find it so difficult to remember them?

To address this important question, one can begin by investigating the experimental/cognitive psychology literature. In a classic paper by George Miller, the notion of short term and long term memory storage was discussed in terms of limits and capacities [2]. Miller has indicated that there is a magical number: 7 ± 2, which represents the number of items that can be stored and processed without significant rehearsal and training. This limit occurs across a variety of modalities. For instance, a series of pure tones may be presented to the subject, and the task for the subject is to be able to rate the pure tone on some pre-determined nominal scale. In a study by Pollack and colleagues, a series of pure tones (100-8,000 Hz) were presented to subjects whom were asked to identify the tones by assigning numerical values to them [3]. The result of this study indicated that subjects were only able to reliably identify approximately 2.5 bits (@ 6 different tones). In another study, subjects were asked to rate the saltiness of a solution containing ordinary table salt, again by assigning a nominal scalar value to each salt solution. The results from this study indicated that subjects could identify 1,9 bit (approximately 4 different salt concentrations) reliably [4]. These and other important results indicate that humans have on average an upper bound for making absolute discriminatory judgments of approximately 2.6 bits (about 7 distinct judgments). This appears to be independent of sensory modality – and forms the basis of the channel capacity of the human brain. This important observation can be summarised by the channel capacity rule, which states that humans can discriminate 7 +/- 2 distinct aspects of a stimulus. It should be noted these results were obtained from a uni-modal stimulus. The obvious question is whether these results extrapolate to more cognitively demanding tasks.

Klemmer and Flick investigated the discriminatory capacity of identifying the x,y coordinate of a dot within a square [5]. Data was already available with respect to the uni-dimensional task – that is locating a position of a point along an interval was approximately 3.25 bits. Therefore, one could imagine that identifying the position of a dot in 2D would require two such operations, yielding a bit rate of 6.5 bits. The authors measured a bit rate of 4.6 bits, considerably less than the expected from a linear summation of two 1D position location tasks. This is a general observation: multi-dimensional sensory modality does not yield a linear superposition of the individual tasks [6],[7]. The yield is always much less than expected – typically on the order of 60% - depending on the modalities involved (though never achieving 100%).

Other researchers have re-examined Miller's magical number and have provided additional information on the subject of memory span generally. A memory span can be loosely defined as the number of elements that can be recalled correctly in 50% of the trials [8]. Cowan has examined the concept of memory span extensively, and his and related work indicate that '7' may actually be too generous. A more realistic value is approximately 2-bits – or 4 chunks of information which can be stored in working (short term memory) [9,10]. These results were modality and dimensionally independent, and apply in the case of password recall, a classic example of a memory span task. These observations have been challenged in the cognitive science literature primarily by Baddeley and his followers [11].

The principle ground for argument is whether or not there is any word length effect on free recall. Baddeley's argument is based on the concept of 'chunks,' as defined by Miller – which he defines as a subjectively meaningful unit. Within this context, the number of chunks recalled is typically on the order of 2-4 items, a constant, independent of the content of the chunks. Baddeley has provided evidence that this concept doesn't hold, and has refined the notion in his word length argument [11]. This argument basically states that we are more likely to recall five monosyllabic words than five polysyllabic words. The word length in this scenario refers to the number of syllables in each word. In the context of a password, this would refer to the length of the characters in the password. There is evidence for both arguments, and the matter will not be addressed in this paper as password lengths were kept constant). Furthermore, Baddeley argues that the maximal amount of information that we can recall is approximately that which can be perceived within 2 seconds. We used this upper limit as maximal stimulus presentation time, as will be discussed more fully in the methods section. Notwithstanding these fundamental arguments, there are other issues that may affect a person's recall capacities.

The limits to working memory are influenced by factors such as level of distraction, age, educational level, and gender [12]. Of relevance to the current study, the work by Hester demonstrates that digit span decreases by approximately 10% between the ages of 20 and 45 [13]. In the current study, these factors were addressed as much as possible. Distractions were kept to a minimum by utilising a soundproof and secure laboratory space. All subjects were male – so in effect, this issue was not addressed in this study. The subjects were all selected from a university setting – so education was similar in many respects. Age was controlled by utilising 2 groups of subjects, each at approximately the same age range – see methods section for details). All the evidence suggest that working memory for memory span tasks, regardless of the context (images, words, smells, tones, text) is about 4 chunks. It is interesting to note that a typical PIN used to access ATMs are four digits (which provide direct access to secure resources: i.e. our money!!!) and passwords, which may only house our Facebook and music accounts, our protected with 6+ digit passwords! The trend in secured access sites is to enhance the level of security by placing the burden of stringency of passwords on the users. So, is there anything users can do to enhance their skills in this domain?

The performance on memory span tasks can be improved *in theory* by increasing the length of the span, given a fixed recall rate, increasing the accuracy of recall given a fixed length span, or both. To date, many strategies have been proposed for specific types of spans, but as of yet, the only general purpose approach is rehearsal – based on short term storage and repetition (conscious and possibly subconscious?) strategies. Even with this approach, the effect is generally to increase the accuracy, without any significant increase in the span length. Hence we would expect that one would be able to remember 4 digits quite accurately (the Cowan limit). But what about passwords, which are significantly longer than this limit – what can be done to enhance recall accuracy in this context?

One approach is to provide a cue or a hint, designed to jog the person's memory and in turn, enhance recall. The deployment of hinting has been suggested as a mechanism to enhance the memorability of passwords [14]-[17]. These authors suggest that providing partial clues (hints) to the user would enhance the memorability of their password, and further, this would allow users to select stronger passwords. In Hertzum's study [18], the effect of providing subsets of a user's password was investigated when the users were provided with a set of five passwords to remember at 1 and 4 weeks after learning their passwords. On average, users remembered 4 out of 5 passwords with hints, compared with approximately 2 of 5 passwords without hints after 1 week. After 4 weeks, the hinted password group's performance dropped by approximately 20%, while the non-hinted group dropped by approximately 5%. These results simply indicate that hinting can increase the likelihood that a user will remember a set of passwords more effectively than without

hints. In the study published by Lu and Twidale, a similar hinting scheme was deployed, with the hints being selected by the user during the creation phase [16]. These hints would be stored on the server and the local host, but slightly altered visually, rendering them difficult to decipher (semi-public) without prior knowledge of the password. The principle results from this study indicated that 67% of the passwords (8/12) were remembered 10 days after creation when hints were provided. In contrast, 0% (0/9) passwords were correctly remembered without hints. Both of these studies indicate that hinting enhances memory recall of passwords when measured within a relatively short period of time (7-28 days). These results are promising, though one would like the recall percentages to reach 100% in most cases. One plausible explanation for falling short of the mark in these studies is that deploying partial hinting may actually be counter-productive. Partial hinting may engender a conflict between two memory based systems – recall versus recognition. When a subset of the password characters is presented *in situ* – the subject may engage in *both* recall and recognition strategies. As a result, neither cognitive system is fully engaged, yielding a compromised response. It may be that displaying a partial hint can not adequately engage memory based cognitive strategies that will yield maximal performance in recall tasks. If one were to display the entire password, then the system is compromised if someone is able to shoulder surf or otherwise record the session. A solution to this problem is to present the password hint pre-attentively. This work explores experimentally whether pre-attentive priming can provide a hint that significantly enhances recall of a user's password in such a way as to minimize shoulder surfing and video surveillance as a means of acquiring someone's password.

In this work, the deployment of pre-attentive information was utilised as an alternative to coping with adaptive learning strategies. Pre-attentive (also called subliminal) refers to providing stimuli within a time frame that is pre-conscious. For visual stimuli, the threshold for conscious perception of a stimulus is approximately 20-30ms, depending on the individual [19]. Subliminal response priming has been described since the late 1890s by Sidis [20]. Subjects were able to guess numbers printed on distant cards with above chance accuracy, despite indicating they were not able to make out what was printed. In typical priming scenarios, subjects are required to make a choice reaction and the amount of time required to react is the dependent variable. In a subliminal version, pre-attentive information specific to one of the choices is presented prior to the choice, and the reaction time is recorded with and without the subliminal stimulus (primer). Typically, the reaction times (RT) are shorter when the subliminal primer has been presented to the subject [21]. More generally, and in the current context, priming refers to providing a copy (or at least a significant hint) of the stimulus that is to be referenced in a particular task [22]. In this case, priming simply refers to displaying the target stimulus, which is the user's password. A further constraint in the current experiment is to provide priming in such a way as to benefit the recipient without compromising the information (i.e. revealing the password) to outsiders.

In this study, pre-attentive priming was deployed in order to determine whether it would alter memory span for 7-alphanumeric based passwords. In one study, subjects were asked to memorise a set of 10 hardened passwords during a single session. They were then asked to enter the passwords 7 days later, either with or without pre-attentive priming. Further, the same set of subjects were asked to enter naïve passwords (created on the spot using a One-Time Password scenario), with and without pre-attentive priming. The false rejection rate (FRR) and indirectly, the false acceptance rates (FAR) were calculated in order to place the results within the biometrics domain. The question addressed here is whether the priming can be performed in such a way as to reduce the FRR, without enhancing the likelihood of successful shoulder surfing and increasing the false acceptance rate (FAR). The next section describes the experimental methodology deployed in this study in detail.

Methodology

A set of 4 university college students (aged 20-24, median 21) and 3 'mature' subjects (faculty members, aged 35-54, median 43) were utilised in this study. All subjects (from the same computer science department) were normal with respect to visual acuity, familiar with password based security, and all were briefed on the nature of the experiment. The experiment was performed in a university laboratory setting, with ambient lighting, a minimal of background noise (less than 15 dB) and without local distractions. The text was displayed either in the middle center of the screen or at the bottom center (for hints) on a high resolution 23" HD Samsung Synchmaster SA950 operating a high definition resolution – 1920x1080 pixels, 120 Hz, in 16 point New Times Roman font, white on a black background. All passwords deployed in this study were developed according to the following strategy: a digit or a character is selected at random by concatenating the 62 possible upper/lower case characters and digits 0-9 in a vector and randomly selecting an index. The characters are placed in at the start of the sequence, followed by the upper case, and finally the lowercase characters 'A'..'Z' and 'a'..'z' respectively. Then 7 random numbers were generated, without replacement, to form an alphanumeric password of length 7. The passwords were generated on the local host, to avoid any issues with requiring the password to be transmitted and hence potentially intercepted by a variety of man-in-the-middle attacks [11]. This protocol generally produces moderately hard passwords, according to our university's IT department definition. No two passwords overlapped by more than 2 characters, checked programmatically before starting the experiment.

In the first set of experiments, users were asked to memorise and enter a series of ten 7-character passwords generated randomly as previously described. Each password was mapped to a simulated application that had a corresponding password authentication text box associated with it. The users were asked to memorise and enter the passwords successfully 5 times in a given session (typically took about 2-3 minutes for each password). The subjects then were asked to return to the university lab one week later for testing purposes, and instructed not to study their password set before returning. The subjects were then placed in the same lab that was used for training, and were asked to enter their passwords. Note that the user entered the password 3 times, regardless of whether they entered it correctly on the first or second attempts. The FRR was calculated based on the failure rate while entering each subjects set of 10 memorised passwords. These subjects were asked to enter the same set of 10 passwords 1 week later, but with the addition of hints. The purpose of this experiment was two-fold: first, to define a standard FRR baseline, and second, to quantify the effects of hinting through any change in the FRR values. The effect of pre-attention was examined in this experiment – as hints were provided pre-attentively and attentively. Note that during all authentication trials with preattentive hints, the second and third password entry attempts were preceded by a visual primer which was placed at the bottom centre of the login screen in a clear moderate-sized font (16 point, New Times Roman, white on a black background).

In order to reduce the cognitive load of recall, OTP passwords were generated programmatically on the local host, using the same protocol described above. The users were not given any time to examine the passwords – all subjects were naïve to all passwords in this set of experiments. The purpose of these experiments was to determine if users could recall these passwords without any prior exposure. Furthermore, the constraint of minimizing surveillance techniques placed severe restrictions on how they could be presented to the user. The basic scenario is as follows: the OTP password was placed at the center of the screen, 16 point, Times new Roman, white on an all-black background for a fixed time interval: {0.5, 1.0, 1.5, and 2.0 seconds}, which was selected randomly, but fixed for a set of 10 passwords. The OTP was displayed for a given time frame, which was overwritten by a login entry box. The subjects would enter the password 3 times before a new password was presented (separated by a 1 second focusing

(blank black screen) delay. On the 2nd and 3rd attempts to enter the passwords, a hint would appear at the bottom middle of the screen. The hint was the password to be entered, in the same fonts etc. as the password previously displayed. The hint would remain on the screen for a brief time interval: {10ms, 20ms, 30ms, or 50ms}. Subjects were made aware that hints would appear, and were instructed to look for them at the 2nd and 3rd attempt. The hint time interval was evaluated in random order, but again, the duration was fixed for a set of 10 login entries (a block). So, all subjects would enter a total of 160 password (16 blocks of 10 passwords), and for each password, three attempts, for a total of 480 password entries. Each entry took approximately 3-5 seconds, so subjects spent a total of 25-40 minutes of actual password entry time to complete the study. Note in order to reduce fatigue and boredom effects, users were given a 2-3 minute rest after 4 blocks (every 6-10 minutes). Note that users were not allowed to enter their passwords while it was displayed the screen. The password was displayed, then it could be entered once the login entry textbox was displayed (which overwrote the password prompt). In contrast, the pre-attentive/attentive hints were displayed concurrently with the login box, so users could utilise that information during entry of their password.

A separate experiment was carried out in order to examine the likelihood of successful acquisition of the password by a shoulder surfer or video surveillance attack. In this experiment, would-be attackers were positioned around a user entering their password using the OTP and hinting paradigm. The attackers were instructed to view from any suitable angle and record the characters they viewed while the user entered their password. The number of contiguous characters collected served as the basis for measuring the false acceptance rate (FAR) of the system generally, and served to indicate how effective the OTP strategy would be against such an attack. A video surveillance camera was also deployed in this experiment, recording the activities at 30 fps, using a wide angle 2MP camera. This data was analysed off-line to determine how clearly the password appeared on the streaming video, for surveillance analysis purposes.

The data that was collected from this experiment consisted of the actual characters that the subject entered for all passwords across all three trials/password. The entries were scored in various ways in order to evaluate the effect of pre-attentive priming (hinting) and the effect of age on immediate memory of the hints. For statistical significance, a student's 2-tailed sample t-test was performed to evaluate the effects of the various experimental protocols utilised in these experiments. A summary of the results is presented in the next section.

Results

The first experiment consisted of determining the FRR when subjects were asked to enter their passwords one week after a short memorization phase. These results are summarised in Table 1. Note that for each subject, a total of 10 unique hardened passwords were utilised, and the FRR was counted as total failed entries (3 possible entries/password).

Table 1. False Rejection rate (FRR) for the subjects deployed in this study. Note each subject was asked to enter 10 separate passwords three times each for a total of 30 password entries. Note that the last 3 entries were from the 'mature' group of subjects. Note that the last 3 entries were acquired from the 'mature' subjects (in italics). The bottom entries are reported as the mean ± the s.d for all subjects.

	Incorrect entries for each trial (3trials/password)	Percentage failed attempts	Number of successful logins
Subj1	**6/30 (incorrect)**	**20.0%**	**7/10**
Subj2	**10/30 (incorrect)**	**33.3%**	**6/10**
Subj3	**9/30 (incorrect)**	**30.0%**	**7/10**
Subj4	**10/30 (incorrect)**	**33.3%**	**5/10**
Subj5	***16/30 (incorrect)***	***53.3%***	***3/10***
Subj6	***15/30 (incorrect)***	***50.0%***	***4/10***
Subj7	***20/30 (incorrect)***	***66.7%***	***2/10***
		41.0 ± 16.2	**4.9 ± 2.0**

This experiment establishes the baseline FRR within the context of long term memory, as the subjects were asked to enter their passwords after 7 days of memorizing it. Note that there were no hints provided during this experiment, and the password was not displayed on the screen prior to entering it. Instead, each subject was asked to select an application via a typical windows based icon, which would simply launch a password entry box. The subjects were already exposed to this exact same protocol during the training phase. This experiment was revisited, with the addition of cuing, in order to examine the effects it has on long term memory. In order to reduce practise effects, this experiment was deferred.

In the next experiment, the same set of subjects were utilised in a one-time password (OTP) protocol. A block consisted of the presentation of passwords at a fixed presentation time of: {0.5, 1.0, 1.5, and 2.0 seconds}. The order of the blocks was selected randomly, but applied consistently across all subjects (the order for the data presented in this paper was: 1.5s, 1.0 s, 0.5 s, and 2.0 seconds respectively). Each block in turn contains 40 unique passwords generated via the OTP approach discussed above. The passwords were further allocated into 4 sub-blocks, each containing 10 passwords. The sub-blocks consisted of 4 randomly selected pre-attentive/attentive hinting times: {10, 20, 30, and 50 ms}. The hints were the actual password, displayed at the bottom center of the screen for the time periods indicated above. The hints would start on the 2^{nd} and 3^{rd} trials for each password, providing a no hint/hint scenario of 1:3. The order of the hint presentation times was selected randomly, but fixed across all subjects. Each subject would then complete 10 passwords with a particular hint time, then repeat for the next hint time, until they complete all 4 hint times, completing 1 block. The subjects would then take a short break and repeat this process across all password presentation times (0.5, 1.0, 1.5, 2.0, in units of seconds).

The protocol is summarised as follows:

- Generate 40 unique passwords per presentation time (block) via the OTP protocol
- for each block, assign one of the hinting times to each such that there are 10 passwords associated with each of the 4 pre-attentive/attentive hinting times (each is a trial)
- For each trial, allow the user to enter the password 3 times, for a total of 120 entries, taking approximately 6-10 minutes/block of clock time.

The characters entered by the subjects was collected and stored for off-line analysis. In particular, the raw data was analysed in terms of the number of characters entered for each of the 120 password entries, across all 4 blocks. Table 2a presents data from a single subject in this experiment (selected randomly from the set of 7 subjects) in the pre-attentive hinting experiment and Table 2b presents the grand average data across all 7 subjects. The data presented in Table 2a,b

are summarised statistically in Table 3, which examined the data using a 2-tailed Student's t-test for the subject selected in Table 2a.

Table 2. Effects of pre-attentive hinting on number of correctly entered password trials. The data was selected from one of the subjects randomly. In a) the averages over each of the 10 trials per pre-attentive times indicated in the rows, at each of the presentation times (the columns). The values reported are the number of successful login attempts (if 1 or more of the 3 attempts was successful). In b), the grand average across all 7 subjects whom participated in this study.

a.

	0.5 seconds	1.0 seconds	1.5 seconds	2.0 seconds
10 ms	**4/10**	**8/10**	**9/10**	**9/10**
20 ms	**6/10**	**9/10**	**10/10**	**10/10**
30 ms	**7/10**	**10/10**	**10/10**	**9/10**
50 ms	**7/10**	**9/10**	**10/10**	**10/10**

b.

	0.5 seconds	1.0 seconds	1.5 seconds	2.0 seconds
10 ms	**5.2**	**8.7**	**9.4**	**9.2**
20 ms	**6.5**	**9.6**	**9.8**	**9.5**
30 ms	**8.3**	**10.0**	**10.0**	**9.6**
50 ms	**9.1**	**9.5**	**10.0**	**10.0**

Table 3. Results of the student t-test on the effect of pre-attentive priming on the number of correct contiguous characters entered during the three trials for each password. The data in the second row are the number of contiguous characters entered for the hinted and non-hinted cases respectively. The data was selected from a subject selected at random with a presentation time of 1.5 seconds. All p values marked with an '*' are statistically significant using a 2-tailed matched student-t test statistic.

10 ms	20 ms	30 ms	50 ms
6.3/4.1	6.7/3.7	7.0/4.6	7.0/5.1
P< 0.026*	**P< 0.010***	**P< 0.0002***	**P< 0.0001***

Note that in the experiment with pre-attentive hinting, the first attempt to enter the password was not accompanied by a hint. This data was analysed in order to obtain a baseline value for OTP, based on short term recall, analogous to the LTM FRR data presented in Table 1. A summary of this data is presented in Table 4. In this experiment, one-time passwords were presented to subjects (generated as indicated above), for each of the four password presentation times (0.5, 1.0, 1.5, and 2.0 seconds). No hinting was provided, and each subject had 1 chance to login across 40 unique passwords. The results are summarised in Table 4.

Table 4. Summary of successful login entries as a function of password presentation time. Note that the entries recorded are success/trials. Note that since there was only 1 attempt/trial in this scenario, then each constitutes a successful login attempt. The bottom row lists the maximum for trials/30 attempts, and the number of successful logins.

	0.5 seconds	**1.0 seconds**	**1.5 seconds**	**2.0 seconds**
Subject 1	0/40	0/40	0/40	1/40
Subject 2	0/40	1/40	2/40	3/40
Subject 3	0/40	0/40	0/40	1/40
Subject 4	0/40	0/40	1/40	0/40
Subject 5	0/40	1/40	0/40	2/40
Subject 6	0/40	0/40	0/40	0/40
Subject 7	0/40	0/40	0/40	0/40
	0/280	**2/280**	**3/280**	**7/280**

In the next experiment reported in this paper, we re-visited the first experiment (data summarised in Table 1) – which examined the long term memorability of the OTP based passwords (see Table 5). The same subjects and the same passwords were utilised in this experiment, with the addition of pre-attentive /attentive hinting at the 2nd and 3rd login attempts. Note that there was no password presentation in this experiment – subjects were cued by the application associated with each password. They then were prompted with a password entry box, where they entered the corresponding password. The subjects entered their password 2 more times, with pre-attentive hinting placed at the bottom center of the screen. The password entries were collected and analysed as per the data reported in Table 1 for direct comparison in terms of the effect of pre-attentive/hinting hinting on long term memory based recall.

The last experiment involves estimating the likelihood that a shoulder surfer would be able to extract the password while an authentic user was entering it using the current system (OTP + hinting). Note that the hint was placed at the bottom center of the screen to reduce shoulder surfing success. The imposters were instructed to sit behind the subjects any way that afforded them any view of the hint and OTP password, provided they maintained a 6 feet radius away from the subject. They were instructed to record the characters as they became visible to them on a sheet of paper while silently observing the authentic users entering their passwords (results are summarised in Table 6).

Table 5. Summary of the subject data generated from subjects where long term memory based passwords were utilised, with the addition of pre-attentive/attentive hinting at the times indicated by the columns. Note the passwords and subjects are the same as those that generated the data in Table 1. The values reported are number of correctly entered attempts over all 30 trials. The parenthetical values are number of successful logins (1 or more of 3 attempts were successful). The bottom row is the mean ± standard deviation (s.d.). Note the averages were across all 7 subjects: the 'mature' subjects are in bold.

Subject / Time	**10 ms**	**20 ms**	**30 ms**	**50 ms**	**Average**
Subj1	3/30(9/10)	4/30(9/10)	2/30 (10/10)	1/30(10/10)	2/30 (9.75)
Subj2	4/30(9/10)	2/30(10/10)	1/30(10/10)	2/30(10/10)	2.3/30(9.75)
Subj3	1/30(10/10)	2/30(10/10)	3/30(9/10)	2/30(10/10)	2.0/30(9.75)
Subj4	3/30(9/10)	1/30(10/10)	2/30(10/10)	1/30(10/10)	1.75/30(9.75)
Subj5	**4/30 (9/10)**	**5/30 (9/10)**	**4/30 (9/10)**	**6/30 (8/10)**	**4.75/30 (8.75)**
Subj6	**5/30 (8/10)**	**8/30 (8/10)**	**5/30 (8/10)**	**4/30 (9/10)**	**5.5/30 (8.5)**
Subj7	**7/30 (8/10)**	**9/30 (7/10)**	**6/30 (8/10)**	**5/30 (9/10)**	**6.76/30 (8.0)**
	3.9 ± 1.9 / 8.9±0.7	**4.4 ± 3.1/8.9±0.7**	**3.3 ± 1.8/9.1±0.9**	**3.0 ± 2.0/9.4±0.8**	

Table 6. Summary table for seven subjects reported as the maximum number of contiguous characters correctly recorded from all of the three trials/password (30), for each password presentation times (120 entries in total). The last row is the mean and s.d. for each corresponding column. The '*' indicate statistical significance ($p < 0.05$). Note that 'NS' indicates not-statistically different.

	0.5 Seconds	**1.0 Seconds**	**1.5 Seconds**	**2.0 Seconds**	
10 ms	2.1/7	2.3/7	2.8/7	3.9/7	NS
20 ms	2.2/7	2.6/7	3.1/5	3.6/7	NS
30 ms	2.8/7	3.0/7	3.0/7	3.3/7	NS
50 ms	2.4/7	3.4/7	3.3/7	4.1/7	NS
	2.38 ± 0.31	**2.83 ± 0.48**	***3.05 ± 0.21****	**3.73 ± 0.35**	

Conclusions

This pilot study was designed to investigate the feasibility of deploying password hints in order to assist users if and when they fail to remember their passwords. If a user fails to remember their password, they can request a password hint, which is implemented by displaying the password briefly on the screen (or possibly via a mobile). This is a purely implementation issue which has little bearing on the purpose of this study – the assumption is that this option is available.

This study demonstrated that the ability of a user to remember relatively hardened password with a minimal delay of 1 week is less than perfect in many cases (100% true in this set of subjects). In fact, in this study, a group of seven users were less than 50% successful in recalling passwords that they spent time memorizing, with only a minimal delay (7 days). These results were somewhat higher than those reported by Hertzum in a similar study [18]. These results were significantly enhanced with the addition of hinting, both pre-attentive and attentive forms. As can be seen by the results presented in Table 5, the hinted long term memory based results were significantly improved, with the best results obtained with 2.0 second presentation time (94% success, or an FRR of 6%). These results are comparable to physiological biometrics generally (e.g. keystroke dynamics based approaches) [23].

Hints were provided in these experiments in a non-standard, subliminal (or pre-attentive) fashion, an approach not currently reported within this context. Other authors have proposed hints – typically in the form of placing *in situ* a few elements of the password. These studies

demonstrated that hinting significantly enhanced the number of passwords that were correctly remembered [18]. The idea was that the hints would jog one's memory. The results of this study focused on the generation of passwords – that hinting allowed users to create more memorisable passwords, without necessarily paying attention to the hardness of the password. Another system, referred to a MiFa (MInimal Feedback hints for remote Authentication), also uses hints to enhance password recall [16]. The authors conclude that "a few carefully revealed hints will jog an authorized user's memory, but will be of insufficient help to an unauthorized user who does not know the password in the first place." Their results indicate that hints are effective, but care must be exerted, otherwise this process may reduce the password space, weakening passwords generally.

The deployment of pre-attentive hints in this study provided a unique approach to hinting. Firstly, it was used in conjunction with a one-time password (OTP) based approach, which obviates the need to remember passwords. Other authors have deployed this approach (indeed there are several commercial versions available) with variable results. Rubin has proposed the use of such an approach, focusing on the creation of passwords (based on a one-way hash function) [17]. The current approach could certainly implement passwords in this fashion – the mechanism for creating passwords is not the critical issue in this work. In this work, a simple local host application generated the passwords, with many features of typical hardened passwords. The passwords were randomly generated, without repetition, and even though there was no requirement for capital letters and digits, most of the passwords contained these elements.

The principle result of this study was that OTPs can be effective provided support is provided – in this, in the form of hints. When OTP passwords were entered without any hints, the accuracy dropped precipitously – to virtually 0%. These values were obtained when the password was entered the first time during the OTP experiments, which does not utilise pre-attentive hinting. When hinting was provided, the best combination of presentation time and hinting was 1.5 seconds and 30ms respectively. The results were only slightly higher (non-statistically different) for 1.5 seconds and 20ms. These values are quite reasonable to reproduce on most modern day computer hardware. The 30ms pre-attentive hint is admittedly at the borderline between pre and conscious awareness, as it is generally held that 20ms is the typical border for 50% of the population in terms of visual awareness of an object [19]. The 20ms results are quite acceptable in terms of the FRR of approximately 10%. The final question to be addressed is how safe is this approach to should surfing and video surveillance? In order to determine how much information was conveyed to a potential intruder, a small experiment on shoulder surfing was investigated.

When a subject was entering their password, the other subjects (1 at a time) were asked to write down what they could read during the pre-attentive hinting phase of the experiments (the results are summarised in Table 6.). These subjects knew where to look, and recorded on paper the password characters they could make out. The ability to shoulder surf was not very successful, even when the pre-attentive hint was maximal (50ms), and a password presentation time of 1.5 seconds. No user was able to acquire the entire password during all of the trials. At most 2-4 contiguous characters were remembered. These data provide a value for FAR – the false acceptance rate. Of the total of 280 attempts, 12 were successful (yielding an FAR of approximately 4.3% - quite a reasonable value for a one-time password based approach. When queried, the subjects indicated that the password was not directly visible due to the location at the bottom center of the screen, which was generally obscured by the subject unless viewed at an extremely acute angle. Although not tested in this experiment, it would seem the same issue would preclude the viewing angle of a suitable camera from acquiring the password characters.

The results of this experiment demonstrate that OTP supplemented with hinting may provide a viable alternative to having to remember a number of unique and hardened passwords. Pre-attentive hinting (in this study, any stimulus with less than or equal to 30ms in duration) greatly improved the recall rate, placing the FRR (@ 10%) in line with typical keystroke dynamics and

related biometric approaches [23]. When queried, all subjects indicated that they were not always aware of the 20 and 30ms hints. The 50ms hints were visible by all subjects, though at times they claimed not to recall the majority of the context. None of the subjects claimed to be able to see the 10ms hint to any appreciable degree. It is interesting to note that such pre-attentive cues provide sufficient cognitive information that subjects utilised for task completion. These results are consistent with other published works on this topic [6],[7].

This work is clearly a starting point for a much larger investigation of the utility of pre-attentive hinting. For one, the password length was fixed at 7 characters – it would be interesting to see the relationship between the length of the passphrase and the resulting impact of this form of OTP. Further, from a methodological perspective, acquiring information regarding the non-preattentive trials as the first of 3 consecutive trials may be somewhat biased. The first trial does not benefit from repeated entries as the other 2 may, and combined with the pre-attentive cue may bias the results unfavorably. This effect will be assessed by randomizing the order of the three trials per password exhaustively. Certainly, a much larger cohort is required in order to provide statistical viability to this study. The cohort consisted of 2 small groups (4 and 3 subjects) respectively representing different age groups. Although not directly indicated, the results of all age related experiments clearly demonstrate a statistically significant reduction in the recall rate (hence and increase in FRR) for the 'mature' group. Gender was not addressed in this study as all subjects were male. Lastly, the education level of the subjects was constant between both groups, with the later possessed greater than 16 years of education (the former between 12-16 years). These issues will be addressed as much as possible in a follow up study, which is currently under way.

References

[1] Monrose F, & Rubin A., Authentication via keystroke dynamics. In: Proceedings of the Fourth ACM Conference on Computer and Communications Security, Zurich, Switzerland, 2–4 April, pp. 48–56, 1997.

[2] Miller, G., A., The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological Review, 63: pp. 81-97, 1956.

[3] Pollack, I., The assimilation of sequentially encoded information, American J. Psychology, 66: pp. 421-435, 1953.

[4] Marois, R. & Ivanoff, J., Capacity limits of information processing in the brain, TRENDS in Cognitive Science, 9(6), pp. 296-305, 2005.

[5] Klemmer, E.T. & Frick, F.C., Assimilation of information from dot and matrix patterns, J. Experimental Psychology, 45: pp. 15-19, 1953.

[6] Cheesman, J., & Merikle, P.M., Priming with and without awareness, Perception and Psychophysics, 36, pp. 387-395, 1985.

[7] Moore, C.M. & Egeth, H., Perception without attention: evidence of grouping under conditions of inattention, Journal of experimental psychology 23(2), pp. 339-352, 1997.

[8] Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4, pp. 417–423.

[9] Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. Journal of Vision, 11(5):4, 1–34

[10] Cowan, N., The magical number 4 in short-term memory: a reconsideration of mental storage capacity, behavioral Brain Sciences, 24, pp. 87-185, 2001.

[11] Baddeley, A.D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behaviour 14, pp. 575–589.

[12] Schacter, D. L., 1992. Priming and multiple memory systems: Perceptual mechanisms of implicit memory. Journal of Cognitive Neuroscience 4(3), pp. 244-256.

[13] Hester, R., Kinsella, G., & Ong, B. (2004), Effect of age on forward and backward span tasks, Journal of the International Neuropsychological Society, 10(4), pp. 475-481.
[14] Haller, N., The S/KEY One-Time Password System, Proceedings of the ISOC Symposium on Network and Distributed System Security, pp. 151-157, February 1994, San Diego, CA
[15] Rathgeb, C. & Uhl, A., A survey on biometric cryptosystems and cancelable biometrics, EURASIP Journal on Information Security 2011, pp. 3-28.
[16] Lu, B., and Twidale, M.B. Managing multiple passwords and multiple logins: MiFA – minimal-feedback hints for remote authentication. In Proceedings of the IFIP INTERACT, 2003 Conference. IOS Press, Amsterdam, pp. 821-824, 2003.
[17] Rubin, A.D., Pseudo-Random Functions for One-Time Passwords, Proc. 5th USENIX UNIX Security Symposium (June, 1995).
[18] Hertzum, M., Minimal-Feedback Hints for Remembering Passwords, Proceedings of the ACM Interactions, 13(3), pp. 38-40, 2006.
[19] Kiesal, A., Kunde, W., & Hoffmann, J., Mechanism of subliminal response priming, Advances in Cognitive Psychology, 3(1-2), pp. 307-315, 2007.
[20] Sidis, B., The psychology of suggestion, New York, Appleton, 1992.
[21] Ansorge, U., Heumann, M., & , Scharlau, I., Influences of visibility, intentions, and probability in a peripheral cuing task, Consciousness and Cognition, 11(4), pp. 528-545, 2002.
[22] Kiefer, M., Executive control over unconscious cognition: attentional sensitization of unconscious information processing, Frontiers in Human Neuroscience, 6(61) (September, 2012), pp. 65-76.
[23] Revett, K. Behavioral biometrics: a remote access approach, ch 8., Wiley & Sons, 2008