Shoulder-Surfing Resistance with Eye-Gaze Entry in Cued-Recall Graphical Passwords

Alain Forget, Sonia Chiasson, & Robert Biddle
School of Computer Science, Carleton University, Ottawa, Canada
aforget@scs.carleton.ca, chiasson@scs.carleton.ca, robert_biddle@carleton.ca

ABSTRACT
We present *Cued Gaze-Points* (CGP) as a shoulder-surfing resistant cued-recall graphical password scheme where users *gaze* instead of mouse-click. This approach has several advantages over similar eye-gaze systems, including a larger password space and its cued-recall nature that can help users remember multiple distinct passwords. Our 45-participant lab study is the first evaluation of gaze-based password entry via user-selected points on images. CGP's usability is potentially acceptable, warranting further refinement and study.

Author Keywords
eye tracking, graphical passwords, usable security

ACM Classification Keywords
H.5.2 Interfaces and Representation: User Interfaces – Input devices and strategies; K.6.5 Computing Milieux: Security and Protection – Authentication

General Terms
Experimentation, Human Factors, Security

INTRODUCTION & BACKGROUND
Graphical passwords are proposed as more memorable and secure authentication methods that leverage the human ability to more easily recognise and recall images over text [14]. One disadvantage to most graphical password schemes is their susceptibility to *shoulder-surfing*: attackers may observe or record users as they enter passwords and subsequently log in with the observed credentials. Text passwords and PINs may also be vulnerable to shoulder-surfing [16,18]. Some shoulder-surfing resistant graphical password schemes rely on obfuscation or challenge-response [12, 18].

Recent proposals use eye-gaze input, which should become affordable in the near future. Kumar et al. [13] first implemented a gaze-based authentication system. Their EyePassword scheme displays an on-screen keyboard whereupon users gaze at the letters of their password. However, EyePassword remains vulnerable to guessing attacks due to the predictability of text passwords. De Luca et al. [6] have proposed eye-gesture methods for shoulder-surfing resistant authentication. Dunphy et al. [9] tested gaze control with PassFaces, a recognition-based graphical password system.

We present *Cued Gaze-Points* (CGP): a cued-recall graphical password scheme using eye-gaze as an input mechanism. Cued-recall (or locimetric) password systems show a graphical cue that triggers the user's memory of their password, and therefore facilitates memory of multiple distinct passwords. Unlike similar click-based schemes, CGP is shoulder-surfing resistant since there is no on-screen indicator revealing users' gaze-point locations. Our 45-participant user study is the first evaluation of gaze-based password entry with user-selected points on images, rather than pre-defined regions.

Cued-recall click-based graphical passwords consist of clicks on specific points of one or more images. Such systems offer fast login times and larger theoretical password spaces than other graphical password systems [1, 17]. PassPoints [19], where passwords consist of 5 clicks on one image, has been extensively evaluated. Although usable, PassPoints users often select predictable passwords [3,7,15], raising significant security concerns. In response, *Cued Click-Points* [4] (CCP) users sequentially choose one click-point on each of 5 distinct images. Each subsequent image is determined by the user's previous click-point location. Attacking CCP requires more effort since it uses a large number of images, rather than only one. Chiasson et al. found CCP users less likely to select passwords in predictable patterns [3].

CUED GAZE-POINTS
Cued Gaze-Points is an eye-gaze version of Cued Click-Points, where users select points on a sequence of images with their eye-gaze instead of mouse-clicks. For each of a password's 5 gaze-points, users look at their desired point and hold the space bar for a few seconds to record their gaze. The system determines the user's intended gaze-point as the centre of the tolerance-sized area of highest gaze density. This emphasises points where the user has primarily gazed and eliminates outliers. We chose the space bar to trigger gaze input because users have difficulty performing eye-gestures to indicate exact input [11]. Kumar et al. [13] and De Luca et al. [6] similarly use the space bar. However, they respectively use on-screen character selection and eye-gestures as passwords, while we use gaze-points.

Eye-trackers typically require a multi-step calibration process for each new session or user. We use an alternative process that is quicker and tailored for the short duration of authentication. Before creating or entering a CGP password,

CHI 2010, April 10 – 15, 2010, Atlanta, Georgia, USA

users perform a 1-point calibration by gazing at a displayed point in the centre of the image area while holding the space bar. When the space bar is released, CGP uses the gaze density method to calculate the user's gaze-point (as above). The distance between the gaze-point and the on-screen point is used as an offset for the subsequent password entry.

EXPERIMENT

What is the usability and security cost of preventing shoulder-surfing attacks with CGP? To answer this, we conducted a 2-condition between-subjects lab study following the published study methodology for CCP [4], including the same image set. Our goal was to evaluate eye-gaze as a cued-recall authentication method; more ecologically valid field studies may only be possible when eye tracking becomes ubiquitous. Eye-gaze accuracy is limited by human eye physiology, so we adjusted CGP's configuration as follows.

We used a 17" Tobii 1750 eye-tracker with resolution of 800×600, instead of the CCP study's 1024×768. This made the physical size of the images in our study $\frac{1024}{800} = 1.28$ times larger linearly than in the CCP study. People have full vision acuity within $\sim 1^\circ$ of their gaze's centre [8]. Thus, a 1° radial target on a 17" monitor with a 800×600 resolution that is 25 inches away from the user forms a circular target with a diameter of *51 pixels*. Kumar et al.'s [13] on-screen keyboard keys were of similar size. Although the area of full vision acuity forms a circular target on the screen, we used square tolerance regions because a grid system is necessary to store passwords securely with Centered Discretization [5]. Our first condition, *T-51*, denotes a 51×51-pixel tolerance square. Since smaller tolerance squares equate to more total squares (and therefore stronger passwords), we tested a second condition, *T-31*, to see how shrinking the tolerance square to 31×31 pixels affected usability.

Forty-five (45) participants were randomly assigned to a condition and completed individual 1-hour sessions. We told participants to choose points that were easy to remember and repeat, but difficult for others to guess. We then instructed them on proper eye tracker use. Perception and demographics questionnaires were given at the beginning, middle, and end of the session. Participants familiarised themselves with the system during two practice trials, followed by 6 to 10 regular trials, as time permitted.

Each individual trial corresponded to one password. Participants first created and confirmed a 5-gaze-point password. If they made errors, they could re-try until successful or skip the trial at any point. Next, participants answered two 10-point Likert-scale questions about the usability of their current password, and completed a 30-second distraction task intended to clear visual working memory and simulate a longer passage of time. Finally, users logged in with their password, re-trying until successful or skipping if they could not remember it. This entire process was repeated every trial.

Table 1 compares the number of participants, trials, and general results from our CGP study conditions, T-51 and T-31. In the right-hand column, we show the published results from the CCP study for comparison.

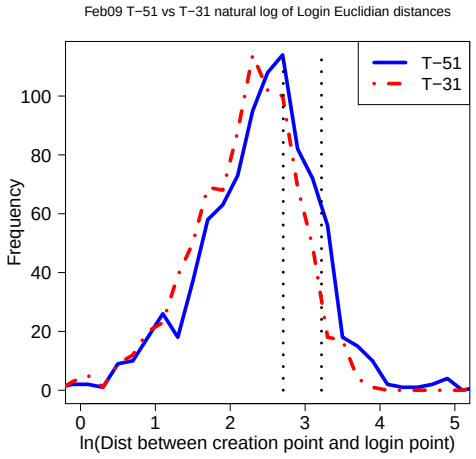


Figure 1. Frequencies of Euclidian distances between the creation and login points for passwords created, scaled by natural logarithm (ln). The vertical lines denote the conditions' tolerance square boundaries.

Successes. The success percentages and chi-square significance tests in Table 1 show that larger tolerance squares (T-51) made the system significantly easier for participants to confirm and log in than with smaller tolerance squares (T-31). In comparison with CCP, T-51 participants logged in 73% without error, while CCP users did so 96% of all logins. Note T-51's 3-try login rate (93%) and CCP's 1-try login rate (96%) are comparable, as are the confirm success rates of T-51 3-try (82%) and CCP 1-try (83%). Thus, CGP users can successfully re-enter passwords, but may require more than one attempt.

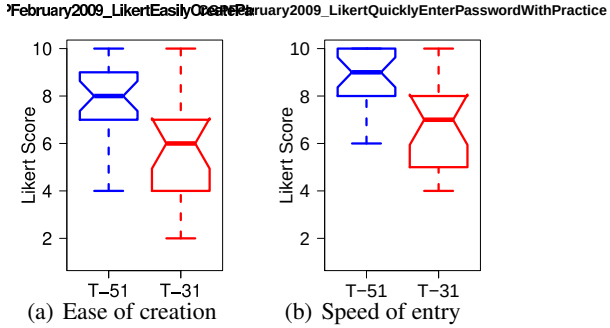
Errors. CGP participants using larger tolerance squares (T-51) were significantly less prone to login errors. Despite larger tolerance squares leading to higher confirm success rates, no difference in confirm errors was observed. CGP users generally committed more errors per trial than CCP users. This is not surprising, since users are probably more practiced at pointing with a mouse than with their gaze.

Times. CGP passwords were created more quickly with larger tolerance squares (T-51). However, for confirm and login, there was high variability (standard deviations) in the time taken, resulting in non-significant differences between conditions. Participants were slower on average to create, confirm, and log in with T-51 than with CCP. This is not surprising because CGP times include typing their username, calibrating, and recording a few seconds of gaze for each point, while mouse-clicking is rapid and CCP times begin at the first click-point [4]. CGP users also committed far more errors than CCP users, and time elapsed during errors and re-tries is included.

Accuracy. Figure 1 plots the frequencies of the Euclidian distances between the creation and login points for passwords created with T-51 and T-31 respectively, scaled by natural logarithm for greater detail. We use a line graph instead of overlaid histograms to make comparisons between the conditions easier. Figure 1 shows that T-51 login gaze-points were farther from the corresponding creation gaze-point than were those of T-31 ($t(1756.60) = 5.65, p < .00001$). This is also true for confirm ($t(2225.13) = 6.53, p < .00001$). This

Table 1. Main results from our CGP Tolerance-51 and -31 (T-51 & T-31) study conditions and Chiasson et al.'s CCP study [4].

System	CGP T-51	CGP T-31	T-51 vs T-31 Sig Tests	CCP T-19
# of Participants	25	20	n/a	24
# of Trials	169	141	n/a	257
Successful Confirms on 1st try	67%	50%	$\chi^2(1,310) = 9.34, p < .005$	83%
Successful Confirms $\leq$ 3 tries	82%	68%	$\chi^2(1,310) = 7.66, p < .01$	n/a
Successful Logins on 1st try	73%	54%	$\chi^2(1,254) = 10.46, p < .005$	96%
Successful Logins $\leq$ 3 tries	93%	79%	$\chi^2(1,254) = 11.46, p < .001$	n/a
Mean Confirm Errors (per trial)	1.21	1.68	$t(308) = -1.40, p = .16$	0.39
Mean Login Errors (per trial)	0.51	1.11	$t(152.73) = -3.20, p < .005$	0.05
Mean (SD) Create Time (s)	44.2 (22.0)	50.1 (23.2)	$t(308) = -2.26, p < .05$	24.7 (16.4)
Mean (SD) Confirm Time (s)	47.1 (78.5)	64.3 (85.3)	$t(308) = -1.84, p = .066$	10.9 (13.1)
Mean (SD) Login Time (s)	36.7 (35.9)	53.5 (45.9)	$t(260.53) = -1.94, p = .053$	7.4 (5.5)



(a) Ease of creation (b) Speed of entry

Figure 2. Box plots of Likert scale responses. High scores favour CGP.

shows that CGP participants gazed more precisely when using a smaller tolerance square. We suspect this occurred because T-31 users required greater precision to re-enter their passwords. Although eye-tracking is limited by eye physiology, somewhat better precision may be achieved with additional care. We also examined the x- and y-coordinates of incorrect gaze-points to see if CGP participants committed more errors at the edges of the image, due to our 1-point calibration. We found no such evidence, since the errors were evenly distributed across the image.

User perception. Figures 2(a) and 2(b) show notched box plots of post-test responses to two 10-point Likert scale questions. Higher scores favour CGP. Figure 2(a) shows that T-51 users felt they could easily create CGP passwords, while T-31 users were neutral. Figure 2(b) demonstrates that most CGP users felt they could quickly enter passwords with practice. This suggests that participants felt that any difficulties were because they lacked familiarity with eye-tracking.

DISCUSSION & CONCLUSION

CGP offers a distinct advantage over conventional graphical password systems: shoulder-surfing is very difficult without visible indicators of user selection. Even with a recording of users' eyes during login, it would be very difficult to align and synchronise this data with the screen with sufficient accuracy. This lab study was our users' first interaction with an eye tracker. It is encouraging to see that 93% of login attempts in the T-51 condition were eventually successful, indicating that users are capable of using the system with additional practice. Participants also indicated confidence in their ability to improve with practice.

Results of this initial study show a clear trade-off between usability and security. We found the smaller tolerance size too difficult to use with eye-tracking technology. The larger tolerance size proved considerably more usable. However, this configuration's smaller theoretical password space (TPS, the total number of all possible distinct passwords in a system) makes it more vulnerable to password guessing attacks. This would be an acceptable trade-off in certain environments. For example, CGP T-51 is much more secure than ATM PINs, because of a larger password space. TPSs grow exponentially, and are typically compared in log_2. An image size of 451×331 gives a grid of approximately 63 51×51 squares. With 5 gaze-points per password, the password space of CGP T-51 is $log_2(63^5) \approx 29.9$, while 4-digit PINs only offer $log_2(10^4) \approx 13.3$. CGP can also offer an even larger TPS by using larger images and/or adding gaze-points. For example, using an 800×600 image and 7 points, CGP T-51 would have a password space equivalent to an 8-character password using a full 95-character US keyboard ($log_2((16 \times 12)^7) \approx 53.1$ vs $log_2(95^8) \approx 52.6$). TPS is only an estimate of security against guessing attacks, since not all points on an image or text character combinations are equally likely to be chosen by users. We address this issue in other work [2, 10].

We next compare our CGP T-51 results (Table 1) to other gaze-based authentication systems. Precise comparisons are not possible due to differences between study methodologies, tasks, and assessment. For example, CGP success rates and times include username entry, calibration time, failed login attempts, and password recall time, in order to present more ecologically valid results. Conversely, Kumar et al. [13] and De Luca et al. [6] only reported the mean password entry time from the first to last gaze, and their participants received more training than CGP participants. We also count a login failure when users either enter an incorrect password or choose to re-enter their password if they see an unfamiliar image (by having previously gazed at the wrong point). This latter re-entry is analogous to erasing and re-typing a text password, which Kumar and De Luca do not count as failed logins. CGP users chose their own passwords, while Kumar and De Luca assigned passwords to users and showed them their passwords before entry, thus not testing password memorability in these conditions. CGPs success rates are lower as a result of using a stricter definition of "success", which is more representative of real-world usage.

Kumar et al. [13] tested their EyePassword system with two conditions: Dwell (when the users stare at their desired target) and Trigger (where users press the space bar to record their gaze-point). EyePassword participants took 10.7 and 9.2 mean seconds to enter Trigger and Dwell passwords respectively. The 1-try success rates for Trigger and Dwell were 85% and 97% respectively. EyePassword performance appears better than CGP T-51, but EyePassword users saw their password immediately before entering it, while CGP users relied solely on memory. Each EyePassword on-screen keyboard key had a focus point in the centre to help users focus their gaze. Obviously, CGP cannot show users their gaze-point, as this would reveal the password to an attacker. EyePassword and text passwords also share the same password space weaknesses: weak user-chosen passwords.

De Luca et al. [6] reported that EyePassShapes users took an average 12.5 seconds to log in, and achieved a 1-try login success rate of 86%. This suggests that EyePassShapes is quicker and easier to use than CGP T-51. However, we believe EyePassShapes may be more vulnerable to shoulder-surfing, since attackers could simply watch users' eye movements to capture passwords. Furthermore, EyePassShapes passwords included only 7 gestures of 8 possible directions each, giving a password space of $log_2(8^7) = 21$, which is smaller than CGP T-51 (29.9, see above).

Login times were not reported in Dunphy et al.'s [9] gaze-based PassFaces study. Their 1-try and 3-tries login success rates were 40% and 65% respectively, which are lower than either CGP condition. CGP also has a larger password space (29.9, see above) than PassFaces ($log_2(9^5) \approx 15.8$).

CGP is the first implementation of a shoulder-surfing resistant cued-recall graphical password system using eye-gaze. This approach has a number of advantages over similar gaze-based schemes and with sufficiently large tolerance squares, the system's usability is potentially acceptable. CGP's password space is larger than similar gaze schemes, and hence more secure against password guessing attacks. Moreover, CGP's cued-recall nature can help users remember multiple distinct passwords, as distinct images on different systems will help users remember their different gaze-points. While CGP's usability should be improved before real-world deployment, we believe CGP is a valuable contribution to eye tracking and graphical password research.

ACKNOWLEDGEMENTS

We thank the anonymous referees whose comments improved this paper's clarity. This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC). Partial funding from the NSERC Internetworked Systems Security Network (ISSNet) is also acknowledged.

REFERENCES

1. R. Biddle, S. Chiasson, and P.C. van Oorschot. Graphical passwords: Learning from the first generation. Technical Report TR-09-09, School of Computer Science, Carleton University, 2009.
2. S. Chiasson, A. Forget, R. Biddle, and P.C. van Oorschot. Influencing users towards better passwords: Persuasive Cued Click-Points. In *BCS-HCI '08: People and Computers XXII*. British Computer Society, 2008.
3. S. Chiasson, A. Forget, R. Biddle, and P.C. van Oorschot. User interface design affects security: Patterns in click-based graphical passwords. *Int. J. Information Security*, 8(6), 2009.
4. S. Chiasson, P.C. van Oorschot, and R. Biddle. Graphical password authentication using Cued Click Points. In *ESORICS, LNCS 4734*, 2007.
5. S. Chiasson, J. Srinivasan, R. Biddle, and P.C. van Oorschot. Centered discretization with application to graphical passwords. In *UPSEC*, 2008.
6. A. De Luca, M. Denzel, and H. Hussmann. Look into my eyes! Can you guess my password? In *SOUPS*. ACM, 2009.
7. A. Dirik, N. Menon, and J. Birget. Modeling user choice in the PassPoints graphical password scheme. In *SOUPS*. ACM, 2007.
8. A. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, 2nd edition, 2007.
9. P. Dunphy, A. Fitch, and P. Olivier. Gaze-contingent passwords at the ATM. In *COGAIN*, 2008.
10. A. Forget, S. Chiasson, P.C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *SOUPS*. ACM, 2008.
11. R. Jacob and K. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, and H. Deubel, eds., *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, chapter 4 commentary. Elsevier Science, 2003.
12. S. Komanduri and D. Hutchings. Order and entropy in picture passwords. In *GI*. ACM, 2008.
13. M. Kumar, T. Garfinkel, D. Boneh, and T. Winograd. Reducing shoulder-surfing by using gaze-based password entry. In *SOUPS*. ACM, 2007.
14. D. Nelson, V. Reed, and J. Walling. Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 1976.
15. P.C. van Oorschot and J. Thorpe. On predicting and exploiting hot-spots in click-based graphical passwords. Technical Report TR-08-21, School of Computer Science, Carleton University, 2008.
16. V. Roth, K. Richter, and R. Freidinger. A PIN-entry method resiliant against shoulder surfing. In *CCS*. ACM, 2004.
17. X. Suo, Y. Zhu, and G. Owen. Graphical passwords: A survey. In *ACSAC*. IEEE, 2005.
18. F. Tari, A. Ozok, and S. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *SOUPS*. ACM, 2006.
19. S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *Int. J. Human-Computer Studies*, 63(1-2), 2005.