Multiplexed Input to Protect Against Casual Observers

Andrea Bianchi
Department of Computer Science & Engineering
Sungkyunkwan University
Suwon, Korea
andrea.whites@gmail.com

Ian Oakley
Department of Human and Systems Engineering
Ulsan National Institute of Science and Technology
Ulsan, Korea
ian.r.oakley@gmail.com

ABSTRACT
It has become common to authenticate to private content on smartphones or tablets in the presence of trusted individuals such as friends and coworkers - casual, inadvertent and non-malicious observers to whom users do not want to needlessly expose their passwords. Existing shoulder-surfing resistant authentication schemes provide security but are overly burdensome for users in these relaxed scenarios. This paper explores the idea of using heterogeneous multiplexed input codes as a simple technique for creating good-enough solutions to protect against casual observation in non-malicious and relatively secure settings - trading off security for usability.

Author Keywords
Authentication, multiplexed PIN, casual observation.

ACM Classification Keywords
H5.2. [User Interfaces]: Input devices and strategies.

INTRODUCTION
The rapidly developing power and ubiquity of mobile devices such as tablets and mobile phones is increasing their use as platforms for shared activities such as entertainment or collaborative work. In such scenarios, friends and coworkers use, share and view the same device easily and seamlessly [6]. However, whenever access to secure data or services is required, the observable nature of common authentication schemes such as PINs mean that users need to rely on social etiquette (e.g., looking away from the screen while another user types a password, or hiding data entry with a hand [4]) to avoid disclosing their access codes. Indeed, due to the advent of easily observable password schemes such as the Android Pattern Lock (crackable even from the smudges left over time [1]), we argue that politely glancing away from a password entry process or accidentally learning a friend or colleagues PIN has become both embarrassing and commonplace. One solution to this problem is to use authentication methods that provide resistance to shoulder-surfing [3, 7]. However, such systems target security against malicious attackers and typically place high demand on users - authentication becomes a difficult, challenging task.

In this paper we argue for interfaces designed to protect against casual shoulder-surfing. We define this as a social or collaborative situation in which users perform authentication in front of a trusted party - not a malicious adversary - and wish to keep their access codes private without the burdens of entering a fully observation-resistant PIN. In the remainder of this paper we describe the state of the art technology in this space, present three systems which rely on multiplexing orthogonal cues to conceal input during authentication and perform user and security studies on these systems to explore their viability.

RELATED WORK
Issues of privacy in single display groupware [6], situations in which multiple users share a single display or device, have long been studied in HCI. To authenticate in such situations, Tan et al. [7] proposed a spy-resistant keyboard to mitigate the effects of shoulder surfing on large shared screens. More recently, Kim et al. [4] presented a series of prototypes for a shared tabletop that enforce privacy- oriented behaviors when authenticating in front of coworkers. Other approaches include the work of Watanabe et al. [9], who built a system that obfuscates a PIN by crowding the device screen with dummy cursors and Morris et al. [5], who relied on individual audio channels to privately convey information to users. Hayashi et al. [3] present an approach that relies on specific image distortions known only to the user. Finally, De Luca et al. [2] presented ColorPIN, a system that tackles shoulder-surfing by multiplexing multiple information channels (numbers and colors) into compound on-screen items that can be indirectly selected from a keyboard. In this system, the compound nature of each item means that a single observed input is insufficient to determine a PIN. A similar idea is proposed by van Eekelen et al [8] and this paper directly

builds on this prior work by designing and formally evaluating a novel multiplexed authentication system.

THREAT MODEL

This paper considers casual shoulder surfing threats. These involve "attackers" who are non-malicious colleagues, friends or acquaintances of the user in situations where they share the same view of an authentication process on a physical display by, for example, sitting or standing adjacently. Users are assumed to be comfortable authenticating in front of the attacker, so do not purposely attempt to conceal their passwords by, for example, obstructing the attackers view (e.g. hiding entry with their hand, as in Kim et al.'s ShieldPIN [4]). However, they also do not intend to openly disclose their passwords. We argue this practical scenario is common in many social and working environments - users authenticate comfortably in public yet do not wish to explicitly reveal their passwords.

SYSTEM DESCRIPTION

We developed three different applications in Java for Android running on a 10.1 inch touchscreen Samsung Galaxy tablet: NumberPIN, ColorPIN and ShaPIN. There were two versions of each: one-time randomization and per-input randomization. In the first case the input interface (e.g. the keypad) is randomized only once per full PIN entry. In the second version, the input interface is randomized immediately after every PIN item entry.

NumberPIN

The NumberPIN system (Figure 1) is a numerical keypad, where the numbers from one to nine (zero is excluded in line with ColorPIN [2]) are displayed in random arrangement on a 3x3 grid. In this system a PIN is composed of a sequence of four digits and users enter items by tapping the on-screen numbers. The interface also features an OK button to authenticate using the entered PIN, and a clear button to reset the input. At the top of the screen a strip of four colored squares represents the current status of the PIN input (empty, entered, authenticated, authentication failed). In the one-time randomization version of this application, the location of the on-screen numbers is randomized once per PIN, while in the per-input randomization, the location of numbers is shuffled at every input. This interface was used to establish baseline performance and has a theoretical entropy, or possible password space, of 6,561 (9^4).

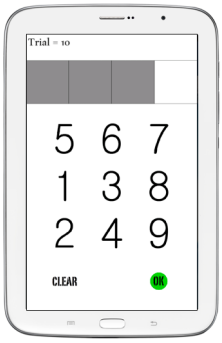

Figure 1. The NumberPIN interface.

ColorPIN

The ColorPIN system (Figure 2) is a replica of the application described in De Luca et al. [2] except that it is ported from a PC to a mobile device. As in the original ColorPIN, nine digits (zero excluded) are shown in ascending order from one to nine in a 3x3 grid. Each location in the grid is associated with three random characters, one displayed in black, one in white and one in red. In total 27 characters are arranged such that there are nine randomly selected letters from the English alphabet, each repeated in three different locations. In the one-time version of this system, each of the letters associated with the numbers are randomized only once per PIN while in the per-input version they are randomized after every input. The per-input scheme was used in the original paper and is the main reason for including this condition in the current work. The system also features a graphical representation of the PIN entry state, as with NumberPIN. Input was made via a virtual keyboard at the bottom of the screen. Pressing enter confirmed the final PIN and spacebar signified a reset.

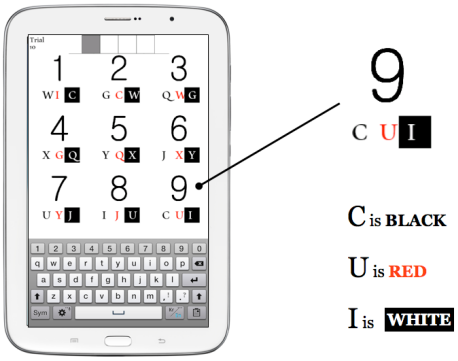

Figure 2. The ColorPIN interface, as explained in previous work, and ported on a mobile device.

In this system a PIN is composed of a sequence of four numbers associated with the colors black, red and white (e.g., 1-red, 2-white, 3-red, 4-black). Users input a PIN by tapping the letter on the keyboard that corresponds to the correct number-color combination. Invalid key entries are not possible; the system ignores entry of such characters. As the same letter occurs three times in each layout, observers cannot identify selected items from a single observation. The theoretical maximum entropy of ColorPIN is 531,441 [2].

ShaPIN

This paper introduces ShaPIN (Figure 3), a system that extends the idea of combining multiple selectable items into a single input element. In ShaPIN, each of the nine buttons in the 3x3 grid layout takes the form of a visually unified object featuring a random combination of heterogeneous PIN items: colors, shapes, numbers and letters. There are three possible colors (Red, Green, Blue), three possible shapes (Square, Circle, Triangle), nine digits (zero excluded) and nine letter groups taken from a typical telephone keypad mapping (ABC, DEF, GHI, etc.). The number of colors, shapes, numbers and letters are divisors of nine so that these features can be evenly distributed among the keypad buttons. As in the previously described systems a graphical representation of the current PIN sits at the top of the screen and confirmation and reset buttons lie at the bottom.

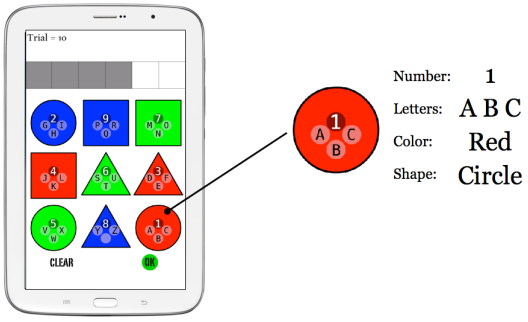

Figure 3. The ShaPIN interface: as shown on the right, interface buttons are multiplexing cues of different types, such as numbers, letters, colors and shapes.

In ShaPIN a user's PIN is composed of any sequence of six heterogeneous elements, such as a color, a letter, a color, two numbers and a shape (e.g., Blue, ABC, Red, 6, 3, Square). To enter an item, users simply touch any on-screen button showing the item they need to select. As the features appear in multiple buttons and in combination with other features, we argue that a casual observer will be unable to determine the true features a user selects during authentication. In work described in this paper, we enforced the use of PINs composed of three high entropy items (numbers or letters) and three low entropy items (colors or shapes) leading to an overall entropy of 19,683.

THEORETICAL SYSTEM COMPARISON

Firstly, it is worth noting that only NumberPIN is composed of a regular digit-based PIN; ColorPin and ShaPIN rely on more complex codes and the use of the term PIN in this paper is solely for consistency with prior work [2]. Indeed, ColorPIN and ShaPIN share considerable similarities in how their multiplexed PINs function. Basically, although user selections in both systems target multiple items (and a full PIN entry contains a large set of possible codes), in a full and correct authentication, only one of these candidates represents the user's PIN. The advantages of this multiplexed approach lies in the resilience to observation it provides - it is impossible to infer a PIN from the many possible candidates after a single observation. Furthermore, the ability to determine a PIN after repeated observations depends on how much additional information is disclosed by each new input process. For example, if a second entry takes place on a group of multiplexed items that is identical to that of an initial observation, no additional information is conveyed and an observer remains stymied. However, in contrast, if a target item is selected when it is multiplexed with an entirely new set of items, an attacker can compare the two separate observations to easily deduce the PIN item.

Extending this logic, it is clear that the smaller a cue set, the more overlap there will be between repeated trials. For example, in ShaPIN, the three colors and shapes will appear in the same compound item 33% of the time, making them relatively observation resistant. In contrast, items from the larger number and letter sets have a much lower probability to reappear with similar items (11%) and are thus easier to spot via visual snooping. This points to a novel tradeoff for multiplexed authentication in which high entropy items that are resistant to brute force need be balanced against low entropy items that are more secure against observation. We argue that ShaPIN offers better security against multiple observations than ColorPIN because it uses a larger space of possible features and feature types and more effectively balances entropy against susceptibility to observation.

Beyond this similarity in the use of multiplexed items ColorPIN and ShaPIN differ in the way PIN items are recognized and input. ColorPIN uses an indirect mapping and selection task - selecting a character signified by a spatially distant number and color. In contrast, ShaPIN, like NumberPIN, relies on direct on-screen selection of items, a potentially simpler task. The study in this paper explores differences in performance and security of the systems described above in order to draw conclusions about the advantages and disadvantages of authentication systems using multiplexed items with direct and indirect input.

EVALUATION

The study was completed by 12 participants (5 female, mean age 26), students at one of the institutions involved in this work. On a five-point scale they reported a high level of familiarity with computers (4.8), smartphones (4.2) and touchscreen devices (3.8). They were not compensated. In terms of design, two variables were manipulated in this experiment: interface type (three levels) and randomization type (two levels). The interface type variable encompassed use of the NumberPIN, ColorPIN and ShaPIN systems. The randomization variable affected when interface elements

were reconfigured on the screen – either one-time or per-item. The experiment was repeated measures – all participants experienced all conditions. The three levels of the interface variable were fully balanced, resulting in six order conditions, each containing two participants. Within each of these six order conditions one of the participants completed the one-time trials before the per-item trials while the other completed the opposite arrangement, also effectively balancing presentation of this variable.

Each condition was structured identically and composed of 20 PIN entries, the first five of which were treated as practice and discarded. During completion of these initial trials, participants were able to view their randomly generated PINs and an experimenter was present in order to answer questions. After this time participants completed the remaining 15 trials solo and in a quiet room. Between

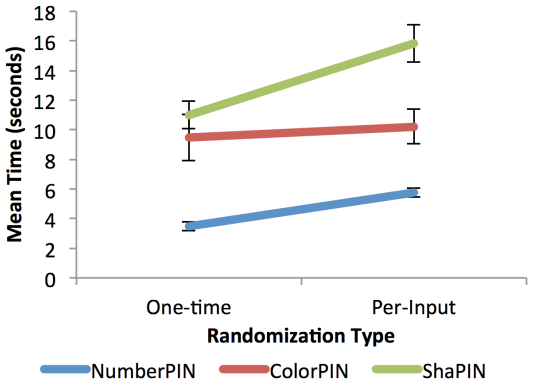

Figure 4. Mean authentication time per PIN.

conditions participants were encouraged to rest. Each PIN entry trial was also structured similarly. First participants pressed the reset button to start, entered their PIN and then pressed the confirmation button to end the trial and move to the next one. They were also able to cancel their current PIN

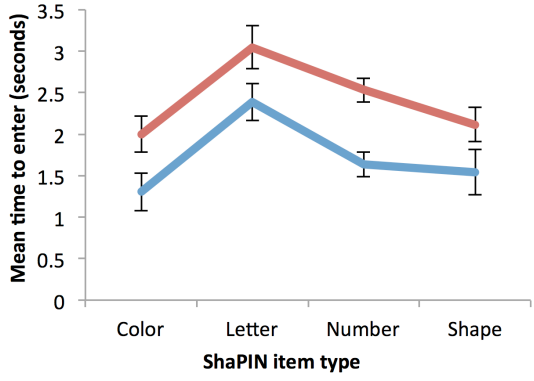

Figure 5. Resets and errors.

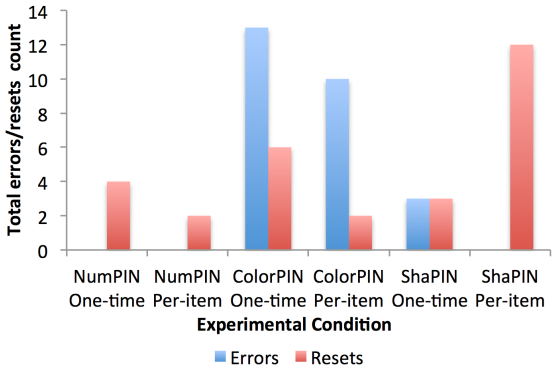

Figure 6. Mean authentication time per-item type in ShaPIN.

entry (subsequently referred to as a reset) by pressing the reset button at any time. The key measures in this study were mean task completion time, error rate (indicated by entry of an incorrect PIN) and reset rate. Additionally time data was stored for the entry of each individual PIN item. Finally, video was also recorded of participant's input – the tablet was typically used flat down on a tabletop and a camera was positioned looking down at the screen and users' hands as they entered data.

Results

Mean PIN entry times are in Figure 4. This data was analyzed with a two-way repeated measures ANOVA and effect sizes are reported at partial eta squared (η_p^2). The interaction between variables ($F(2, 22) = 12.23$, $p<0.01$, $\eta_p^2 = 0.528$) as well as both main effects were significant: interface ($F(2, 22) = 42.33$, $p<0.01$, $\eta_p^2 = 0.794$) and randomization ($F(1, 11) = 19.74$, $p<0.01$, $\eta_p^2 = 0.642$). Post-hoc t-tests with Bonferroni CI adjustments showed that all interface types differed from one another (all at $p<0.01$). Total error and reset data (e.g. summed counts) are shown in Figure 5. These data are sparse (for example, no errors are recorded for three conditions) and unsuitable for parametric analysis. Accordingly, Freidman tests were used to examine this data. An effect of interface type was found to be significant for errors ($p<0.01$) but not resets ($p=0.095$). Pairwise comparisons showed ColorPIN led to more errors than both the other interfaces ($p<0.05$). Finally, we also analyzed individual item entry times for each ShaPIN item type and randomization type (Figure 6). A two-way ANOVA showed no significant interaction ($F(3, 136) = 0.198$, $p = 0.9$) and significant, but fairly weak, main effects: item type ($F(3, 136) = 9.05$, $p<0.01$ $\eta_p2 = 0.166$) and randomization ($F(1, 136) = 20.96$, $p<0.01$ $\eta_p2 = 0.134$). Post-hoc t-tests with Bonferroni corrections on the item types showed that colors and shapes resulted in faster performance than numbers and letters and that numbers were faster than letters (all at $p<0.01$).

Security Study

We conducted a security study on the videos from two randomly selected experimental participants with the goal of determining the resilience to casual observations of the three interfaces. Two attackers completed this study. Both were knowledgable about security and one served on the project development team and thus had intimate knowledge of the system operation. The other received detailed instructions. Each attacker received 12 videos (six per participant), each containing the 15 experimental PIN inputs from one of the study conditions. Attackers were not limited in their exposure to the videos and were encouraged to take written notes to help them crack the PINs. Each was granted three attempts to crack each PIN and they were asked to report how many trials they observed in order to do so.

All systems were successfully cracked by both attackers, as follows: NumberPIN after watching one input; ColorPIN after 2.75 inputs; and ShaPIN after 3.875 inputs. There was no variation between one-time or per-item randomization. These results match those reported from literature [2]. Although these numbers are small, ShaPIN did achieve the greatest security against observation (30% more trials were required compared to ColorPIN), a fact that attackers reported is partially due to the longer six-item PIN used.

DISCUSSION AND CONCLUSION

A number of conclusions can be drawn from the studies. First, ColorPIN led to the largest number of errors: more than seven times greater than ShaPIN and in contrast to flawless performance in the NumberPIN baseline. While training may be able to mitigate this problem [2], we argue that the indirect mapping between the keyboard used for input and the interface elements shown on the screen is fundamentally complex and error prone. Reinforcing this point is the fact that our ColorPIN implementation resulted in performance that is 23% faster than the original [2]. We attribute this to a more direct mapping - essentially the closer physical proximity of keyboard and display elements in our tablet interface led to this improvement over the previous PC based interface. This point is further supported by the fact that ShaPIN, in the one-time configuration, is no slower than ColorPIN, despite requiring the entry of six, rather than four, PIN items. We again argue this more rapid entry of individual PIN items is partly due to ShaPIN's direct input: users simply click on their desired targets. Results contrasting the one-time and per-item conditions also suggest a valuable lesson. Basically while per-item randomization did not increase the security of the system, it did show the potential to reduce overall performance by increasing reset rate and input time in the ShaPIN interface. We therefore suggest that multiplexed PINs operate best with a one-time randomization before input starts.

In sum, this paper evaluates two different systems that explore multiplexed PINs. Our data show that multiplexed passwords improve resistance to observation by casual observers compared to baseline numerical PINs, at the cost of slower input times. More specifically, we found that direct mapping is the best design strategy and that randomization should be performed only once per PIN.

Future research will further explore the design space of the ShaPIN system by adapting the technique to leverage the cue types, such as colors or shapes, that support faster input.

ACKNOWLEDGMENTS

This work was partially supported by the National Research Foundation of Korea (NRF) Basic Science Research Program grant funded by MOE NRF-2010-0020210 and NRF 2014R1A1A1002223.

REFERENCES

1. Aviv, A., Gibson, K., Mossop, E., Blaze, M., Smudge attacks on smartphone touch screens, Proc. of USENIX, (WOOT'10), pp. 1-7.
2. De Luca, A., Hertzschuch, K., Hussmann, H., ColorPIN: securing PIN entry through indirect input, Proc.of CHI '10, pp.1103-1106.
3. Hayashi, E., Dhamija, R., Christin, N., Perrig, A., Use Your Illusion: secure authentication usable anywhere, Proceedings of SOUPS '08, pp. 35-45.
4. Kim, D., Dunphy, P., Briggs, P., Hook, J., Nicholson, J.W., Nicholson, J., Olivier, P., Multi-touch authentication on tabletops, Proc.of CHI'10.
5. Morris, M.R., Morris, D., Winograd, T., Individual audio channels with single display groupware: effects on communication and task strategy, Proc.of CSCW'04.
6. Stewart, J., Bederson, B.B., Druin, A., Single display groupware: a model for co-present collaboration, Proc.of CHI '99, pp. 286-293.
7. Tan, D.S., Keyani, P., Czerwinski, M., Spy-resistant keyboard: more secure password entry on public touch screen displays, Proc.of OZCHI'05, pp. 1-10.
8. van Eekelen, W. A. J., van den Elst, J., Khan, V.J., Picassopass: a password scheme using a dynamically layered combination of graphical elements,Extended Abstracts of CHI'13, pp. 1857-1862.
9. Watanabe, K., Higuchi, F., Inami, M., Igarashi, T., CursorCamouflage: multiple dummy cursors as a defense against shoulder surfing, SIGGRAPH Asia 2012 Emerging Technologies.