Pitfalls of Shoulder Surfing Studies

Oliver Wiese and Volker Roth
Secure Identity Research Group
Freie Universität Berlin
Berlin, Germany
Email:{oliver.wiese,volker.roth}@fu-berlin.de

***Abstract*—We review empirical studies that evaluate the resilience of various PIN entry methods against human shoulder surfers. Conducting such studies is challenging because adversaries are not available for study and must be simulated in one way or another. We were interested to find out whether there is a common standard how these experiments are designed and reported. In the course of our research we noticed that subtle design decisions might have a crucial effect on the validity and the interpretation of the outcomes. Getting these details right is particularly important if the number of participants or trials is relatively low. One example is the decision to let simulated adversaries enter their guesses using the method under study. If the method produces input errors then correct guesses may not be counted as such, which leads to an underestimation of risk. We noticed several issues of this kind and distilled a set of recommendations that we believe should be followed to assure that studies of this kind are comparable and that their results can be interpreted well.**

I. Introduction

Hardly any security problem has attracted as much attention in the human-computer interaction community as the problem of how to design password input methods that resist human shoulder surfing. What makes this problem interesting is the inherent challenge between usability and security. Increased security cannot be expected to come without a usability cost, but how usable can password input be while still offering some security improvement over traditional password entry? A multitude of proposals towards this end have been published in the scientific literature. Unfortunately, few authors have conducted comparative studies and comparing different proposals found in the literature is difficult. While most proposals come with an empirical shoulder surfing study, these studies used a variety of different setups and assumptions. The conclusions from these studies are rather coarse-grained. Typically, the respective authors conclude that their input method is "secure," where security is defined simply as the inability of study participants to guess a user's secret within a given number of observation attempts. If all proposed input methods are secure along this dimension then this leaves only perceived usability and measurable performance metrics such as entry time and input error rate as criteria for comparison. However, it is not plausible that all the proposals in the literature are equally secure. Towards understanding better how this problem should be studied and how proposals can be compared we reviewed a number of shoulder surfing studies. During that activity we noticed a variety of study design issues than can have a subtle or even a considerable influence on the validity and interpretation of the outcomes of such studies. In this short paper, we collect and discuss these issues so that they can be taken into consideration when interpreting the existing literature, and when designing similar shoulder surfing studies in the future. Another outcome of our research is the proposal of an improved methodology to study shoulder surfing security.

USEC '15, 8 February 2015, San Diego, CA, USA
 ISBN 1-891562-40-1
http://dx.doi.org/10.14722/usec.2015.23007

II. Contemporary Modeling

In this section, we review the contemporary process for studying authentication schemes in our threat model and make preliminary conclusions. We use a publication by Kim et al. [1] as a frequent example throughout our text. The paper presents several input schemes, it has been published at CHI 2010 and it has been cited 23 times at the time of writing, which makes this paper an interesting case study because the paper probably influenced a number of researchers. Our discourse is not specific to their paper, though. Related literature published at reputed HCI and SEC conferences use comparable approaches. We refer to additional literature throughout the text.

A. Input Methods and Input Errors

Kim et al. [1] evaluate their schemes using a custom FTIR-based multi-touch display for input and derive touch pressure information from it. Multi-touch displays based on the FTIR phenomenon were popularized by Jeff Han in 2005 [2]. Put simply, infrared light is injected and trapped in a plexiglass waveguide, which acts as the touchable surface of a rear-projection or a flat panel display. Touching the waveguide frustrates the internal reflection and allows the infrared light to escape at the position of the touch to where it can be recorded with an infrared camera. Using simple image processing techniques, touch positions can be extracted from the resulting video. For our discussion, it is important to note that the FTIR effect does not depend on the pressure exercised but only on the surface of touch. Although pressure can be estimated by measuring the area of the recorded blobs. The more pressure a touch exercises, the larger is the area the touch covers on the surface, up to a point. The precision with which the pressure can be estimated depends on the resolution of the camera that is used and its distance to the surface where the touch occurs. If the camera resolution is comparatively low then this affects the reliability of the estimates and may cause input errors.

Unfortunately, Kim et al. did not report the error input rate. Why is this information important?

Studies often measure the success of the adversary as follows. The investigator (i) asked study participants who posed as adversaries to enter their PIN guesses *using the same experimental input mechanism* as the user whom they attacked, and (ii) counted how often the adversary entered the PIN correctly, followed perhaps by a more detailed analysis of how many digits were wrong [1], [3]–[5]. The consequence is that an adversary may make errors when entering her PIN guesses, for technical reasons, for lack of experience with the experimental scheme, or because the scheme is difficult to use. This leads to an under-estimation of risk. Furthermore, a novel scheme causes greater cognitive load during experiments than would be the case once the scheme has become accepted and users have become proficient at it. This leads us to our first conclusion.

Conclusion 1: Investigators should measure the success of adversaries in a fashion that rules out input errors. Investigators should verify that adversaries reported what they intended to report.

As a general principle, we must be careful to design usability studies and security studies so that these two factors do not interfere. For example, Zakaria et al. [6] asked the simulated adversary to draw the secret on paper. This is a simple solution to avoid input errors.

B. Side-Channels due to User Behavior

SlotPIN is a human-computer authentication scheme for tabletops where the tabletop displays multiple reels, each with a permutation of the numbers 0-9. The user must align the digits of her PIN number in one row using soft input reels below the digit reels. One complete observation leaves the observer with only 10 possible PIN candidates, and two such observations yield a unique solution with high probability. Towards an improvement, Kim et al. [1] propose *CuePIN*. This variant requires that the user performs a shielding gesture with her hand on the touch-sensitive tabletop, which displays a challenge letter within the shielded area. This is an elegant idea because the interaction design *constrains* the user's actions so that the only perceivable and feasible action *encodes a security best-practice.* The reel positions are labeled with letters and the user authenticates herself by aligning the digits of her PIN on the reels with reel positions the secret challenge letters indicate.

Kim et al. did not study these two schemes but they are nevertheless useful to highlight the need to perform a usability study before studying a scheme's security. It is conceivable that users do not rotate the reels arbitrarily but so that their secret digits move towards the challenge letter. If this were found to be the case then an observer can exclude all digits as candidates that wrap around on the virtual reel.

Some authors have noticed already that user behavior yielded information on the user's secret. Examples are fake cursor schemes [7], [8]. Briefly, such schemes show n identical cursors instead of one and only one of them, the "real" cursor, can be used to input the secret. In order to input their secrets, users must distinguish real cursors from fake ones, which they typically do by eliciting a movement pattern they can recognize. This undermines security because the adversary has the same goal and the cursor signal that users elicit is not assumed to be secret. For example, De Luca et al. [8] noticed that users moved the cursor to the border in order to recognize it, which worked because the fake cursors did not move to the border. This leads to another conclusion.

Conclusion 2: Researchers should study how users interact with a scheme before studying the scheme's security. The way in which users interact with a scheme may reveal behavioral side-channels.

Sasamoto et al. [9] also noticed that user behavior can leak information. There is a subtle difference, though, in what they noticed. In their study, leakage occurred because users failed to meet the scheme's interaction requirements, for example, by not fully covering a secret output with their hand, or because users gave information away independently of their interaction with the scheme, for example, by means of utterances. A similar example is due to De Luca et al. [10] who reported that two users "pointed on the number they wanted to input" which informed the simulated adversary (2 out of 48 authentication sessions were successful). What we have described before are cases in which information leaks occur even if all stated interaction requirements of an input method are met.

Conclusion 3: A design should seek to minimize the opportunity for behavioral side-channels by eliminating user choice from users' input (because humans are not good sources of randomness).

In other words, the input scheme itself should fix all random choices that are relevant for security. Users' behavior should ideally be deterministic in relation to these random choices, that is, user responses are fixed along with the random choices of challenges. This can be difficult to achieve in practice because even timing user behavior can leak information about a secret [11], [12]. In summary, we have at least three types of side-channels to consider: interaction designs that allow users choices that encode secret information, violation of proper input procedures, and interaction that has a timing dependence on secrets.

C. Isolating the Sources of Security

PressureGrid is a 3-by-3 grid layout of digits one to nine (see Figure 1). The grid is rotated by 45 degrees and touch-sensitive colored regions protrude from the lower left and right sides. The user places three fingers of her left hand on the colored regions left of center, and three fingers on the colored regions right of center. The software detects where the fingers are and deforms the regions so that each finger lies in exactly one region. The colors of the regions flicker, in the assumption that this makes it more difficult to detect when fingers exert pressure, for example, by detecting a change of color under the fingernails. In order to enter a digit, the user increases the pressure of the fingers that indicate the row and column of the desired digit. One hypothesis on which the security of PressureGrid rests is that exercising pressure without lifting any fingers is difficult to observe. If the input is not observable then PressureGrid does not leak information on the entered PIN digits. Kim et al. [1] extended the general idea in two dimensions. One dimension is that the contents of the grid

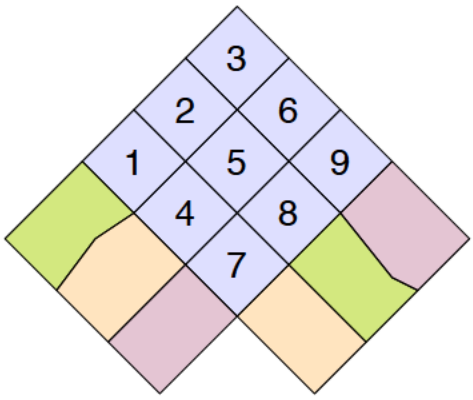

Fig. 1. The PressureGrid display with digits in a fixed arrangement. Other design alternatives are photos instead of digits and a randomized arrangement instead of a fixed one. A '2' is input by pressing on the two green input fields.

can be fixed or randomized. The second dimension is that the grid contents can be digits or face photographs, mimicking the well-known *Passfaces* system.

The PressureGrid scheme combines a number of elements that all may contribute to its security, for example:

- The flickering lights
- Minimizing finger movement by detecting pressure
- The distance of hands relative to the grid
- The two dimensions we mentioned before

Unfortunately, Kim et al. studied these elements jointly and not separately. This makes it difficult to know which elements contribute substantially to the scheme's security and which elements contribute only marginally. A breakdown of the elements may also provide new insights. For example, a key ingredient of PressureGrid is its pressure-sensing aspect, the goal of which is to make input observation difficult. However, in randomized grids, observations are only useful if some input and some output can be observed simultaneously. Since humans have a limited field of view, the distance of the hands relative to each other and to the grid thus play a similar role as the pressure sensing, and may even replace it in some use cases. This leads us to another conclusion.

Conclusion 4: An empirical security study of a scheme should isolate the core contributors to any improvements in a scheme's security.

Another motivation example for this recommendation is the study by Schaub et al. [5], who found that virtual keyboards on different smartphones yield differences in input error rates and security against observation attacks. Interestingly, they reported that virtual keyboard variants with lower usability proved more resistant against shoulder surfing. Among the reasons they offer are small button size and switching through characters. Another explanation might be that users make input mistakes more frequently and need to correct them and this increases the difficulty of observing the correct input. However, the question what determines security is not scrutinized further. The problem is that if a mechanism turns out to be secure only because it is not very usable then the security is a questionable benefit. In particular, a mechanism may be believed to be secure even after the reason of its security, that is, lack of usability, has been fixed in subsequent usability improvements.

D. Distractor Tasks

Kim et al. [1] performed a comparative study of four authentication schemes: regular PIN entry, Passfaces, PressureGrid with fixed digits *(PressurePIN)*, and PressureGrid with randomized faces *(PressureFace)*. They recruited 21 participants whom they randomly split into seven groups of three. One of each group was randomly assigned the user role and the other two posed as adversaries. All participants were briefed about the study and the schemes. All groups performed the experimental task for all schemes in random order. At the beginning of each task, the user practiced the scheme, without being observed, until he could authenticate himself successfully three times in a row. Following that, he authenticated himself three times while the adversaries observed his interaction. Afterwards, the adversaries had to perform a distractor task, they had to read a text for 30 seconds. Kim et al. do not mention that adversaries were asked questions about the text and hence it is possible that some adversaries performed the distractor task faithfully while others did not. Next, the adversaries were asked in random order to authenticate themselves using the experimental scheme. They had three tries each. The experiment concluded with an exit questionnaire.

Studies other than Kim et al.'s did not subject adversaries to a distractor task between the observation phase and the guessing phase. On the contrary, several studies allowed adversaries to make notes using pen and paper [4]–[6], [13]. Kim et al. argued that adversaries will not likely have the opportunity to authenticate themselves right after having observed user input. While this may often be true, adversaries will attempt to rehearse the observed information instead and will attempt to persist it, for example, by making notes. The distractor task impedes rehearsal and thus impedes transfer of the observed information into long-term memory. In Kim et al.'s study, only 10 of 14 observers were able to succeed in the regular PIN condition, whereas in other studies the success rate is typically 100% in this condition. This is indicative of the negative impact the distractor task has on the performance of adversaries. Indeed, some who failed to reproduce the correct PIN mentioned that they had forgotten it. From a security view point, it is preferable to give adversaries optimal conditions for their attack because in this fashion, the conclusions of the study will hold for a more significant variety of environments. It is also important to keep in mind that adversaries will attempt to shape the environment to their advantage.

Conclusion 5: Investigators should not require adversaries to perform distractor tasks between the observation phase and the guessing phase of an experiment. Instead, adversaries should be given the opportunity to rehearse what they have perceived.

E. Number of Observations

Table II lists the number of observations the adversaries were allowed to make in different studies. Many studies let adversaries observe three or fewer user authentications [1], [3]–[6], although this choice is quite arbitrary. In a realistic environment, the observation attack is either opportunistic or

Scheme	Input of adversary	Distractor task	Input error rate	Success rate of adversary
Back-of-Device [8]	missing	not mention	3.5% to 26.4%	38% to 100%
ColorPin [10]	missing	not mention	0%	4.17%
(extended) DAS [6]	paper	no	(see caption) †	40% to 77%
Pressure PIN/FACE [1]	scheme	yes	missing	5%/0%
Cognitive Trapdoor Game [13]	scheme	no	9% to 20%	0%
Undercover [9]	missing	not mention	26.32/52.63%	22.37%
WYSWYE [14]	missing	not mention	29%/ 25%	0%
XSide [15]	missing	not mention	12%	9% to 53%

TABLE I. SUMMARY OF DIFFERENT SECURITY STUDIES. † THE MEAN NUMBER OF ATTEMPS FOR A SUCCESSFUL LOGIN WAS BETWEEN 1.0 AND 1.4.

planned. In the former case, the adversary might be able to observe only one authentication of a user. An example would be a thief who observes someone unlocking her smartphone in a cafe and subsequently steals the smartphone. In the latter case, the adversary may observe arbitrarily many authentications of a user. For example, a worker may observe a co-worker logging into a collaborative system regularly.

Conclusion 6: Investigators should allow adversaries a number of observations that matches their assumptions about the scenario and the environment where the scheme will be deployed. Investigators should clearly document their assumptions.

Weaker assumptions are preferable because it means that the scheme is applicable more widely. Clearly, a scheme that resists observation attacks in a *planned* scenario also resists these attacks in an *opportunistic* scenario. Dunphy et al. [16] already went into that direction in an earlier study. They modeled a *friend attack* and allowed adversaries to make up to 10 observations at their own discretion before guessing the secret. Perhaps we need a well-defined catalogue of attack scenarios to which authors can refer explicitly.

Kim et al. [17] took yet another approach. They assumed that subsequent shoulder surfing attempts would be 10 times harder than previous ones and weighted the combinatorial success probability of adversaries after k observations by 0.1^k. Such an assumption is quite arbitrary and we cannot recommend that approach. It is worth noting that their weighting introduced an error that leads to counter-intuitive results. The error is apparent in Fig. 8 of their paper, which indicates cases for which, all other things being equal, more observations yield a smaller success probability than fewer observations.

F. Live versus Video Observations

In several studies, simulated adversaries were shown videos of user input and had to guess the secrets of users, instead of observing actual users. An argument in favor of video recordings would be that the input that adversaries receive is consistent and repeatable. However, Schaub et al. [3] conducted a study of multiple input schemes in which the adversary had to observe live input and video. In most schemes, the success rate of adversaries was lower for video observations than for live observations. For example, the success rate of adversaries against Pass-Go and UYI was about 0.7 for live input and 0.2 for video. This prompts us to add the following recommendation.

Conclusion 7: We recommend prefering live observations to study human shoulder surfing unless good reasons speak in favor of video.

However, since this recommendation is based on a single data point it should not be regarded as cast in stone. Rather, we add it to emphasize that the choice of video versus live observations is a decision that should be made deliberately, based on the adversarial model under consideration.

G. Simulation of the Adversary

Multiple options exist when it comes to modeling the adversary in relation to the user. We encountered three dominant approaches:

1) Participants are cast into the roles of adversaries and users. The adversaries observe authentication sessions of users.
2) One expert adversary observes the authentication sessions of all participants.
3) Participants are cast into the role of adversaries and observe authentication sessions of an expert user.

Table II provides a breakdown of these approaches by scheme. The different approaches are rooted in the different reasons that may cause a simulated adversary to succeed or fail in his task to guess secrets. A typical approach would be Approach 1, which bases one's estimates on a population average. However, it is difficult to account for factors such as learning, motivation and aptitude in such a setup. If one is concerned about over-estimating security in this fashion, an attempt to remedy that is to find an "expert" adversary and to re-use him in experiments with a group of recruited users, that is, Approach 2. However, over-estimation may occur also because users have little practice with the scheme under study. Hence they may be slower than they would be once the scheme is adopted and users become increasingly proficient with it. This speaks in favor of pitting recruited adversaries against an "expert" user, that is, Approach 3.

Conclusion 8: Since learning should be easy in the case of users but difficult in the case of adversaries it is better to use expert adversaries than non-experts if the goal is to avoid over-estimating security.

However, this still leaves the question open what constitutes an "expert" adversary. Studies that relied on experts often assumed that authors or experimenters would be good expert adversaries or expert users perhaps because they have the best insight into their schemes or the most practice with it. This, however, still has limitations as we are going to argue in the next section.

H. The Role of Strategy

Observation strategies and information gleaned from partial observations are often discounted in the interpretation of

Scheme	Adversary	User	Observations	
			Type	Count
Back-of-Device [8]	expert	participants	video	1
CCP,MIB, Pass-Go, TAPI,UYI [3]	participants	experimenter	live/ video	1
ColorPin [10]	expert	participants	video	1
(extended) DAS [6]	participants	experimenter	live	1
Passfaces^{TM} [4]	participants	experimenter	live	1
Pressure PIN/FACE [1]	participants	participants	live	3
Cognitive Trapdoor Game [13]	participants	participants	video	1
Undercover [9]	authors	participants	video	missing
WYSWYE [14]	participants	participants	screenshot	1
XSide [15]	experts	participants	video	1

TABLE II. SUMMARY OF DIFFERENT SECURITY STUDIES. TYPE DENOTES HOW THE ADVERSARY OBSERVED USER INPUT (LIVE VS. VIDEO).

exprimental results. For example, two of 14 adversaries in Kim et al.'s study [1] succeeded in the PressurePIN experiment. They reported afterwards that they watched one hand when the user authenticated himself for the first time, and watched the other hand when the user authenticated himself for the second time. By combining this information, they were able to infer the entered PIN number. In other words, these adversaries developed a detailed strategy that guided what observations they wanted to make. If a strategy is easy to use yet not immediately obvious then this threatens the validity of a study's results. Once the strategy becomes known, many more adversaries might be able to break a scheme than were able to do so in the experiments. In the relatively short timeframe of a study, investigators cannot expect that participants develop suitable strategies. Most studies do not even provide such an opportunity to participants before the experiment. One exception is the study of Roth et al. [13], who mention that they actively encouraged participants to think of strategies to attack their proposed scheme beforehand. Other investigators were aware that strategy plays a role, but limited themselves to asking participants about their use of strategy in exit questionnaires [4], [8].

Conclusion 9: When studying authentication schemes empirically, investigators must take observation strategies into account.

Of the conclusions we offered thus far we find the last one the most interesting because it is not immediately clear how the stated requirement can be met. Interviewing participants does not reliably uncover feasible strategies, and even if investigators instruct adversaries on feasible strategies, participants cannot be relied upon to master such strategies within the relatively short training period that is realistic in an experimental setting. Hence, a more rigorous and controllable approach is needed.

III. MINI CASE STUDIES

Kim et al. [1] proposed several schemes for tabletop devices with pressure sensitive input and asked study participants to pose as adversaries in an observation attack. Schaub et al. [3] explored the design space of graphical passwords on smartphones. They implemented six published schemes and studied their resilience against observation attacks, asking study participants to pose as adversaries. Adversaries had to enter their guesses using a phone and the experimental scheme. Tari et al. [4] studied the security of *Passfaces* and regular PIN entry against observation attacks. They also asked participants to pose as adversaries and let them enter guesses using the scheme under study. Tari et al. did not study input error rates but these errors were probably fairly low for the schemes they studied. Schaub et al. [5] conducted a comparative study of virtual keyboards on different smartphones, including their resistance to observation attacks during password-entry. Again, participants posing as adversaries had to enter their guesses directly on the device. Schaub et al. asked adversaries about preconceived strategies in an exit interview (i.e., whether adversaries focused on a virtual keyboard, finger movement or entry field) and correlated their strategies with a measure of their guessing success.

Zakaria et al. [6] studied variants of schemes that Jermyn et al. [18] proposed and how resilient they were are against observation attacks. In their study, they asked participants to pose as adversaries. In contrast to the studies we mentioned before, adversaries entered their guesses on paper, using printouts that resembled the principal interface. De Luca and various co-authors proposed and analyzed several mechanisms [8], [10] and analyzed their resilience to observation attacks. In one study, participants pointed to what they wished to enter as their guess instead of using the experimental scheme [10]. The other study's report [8] mentions that user strategy played a role in both performing the experimental scheme successfully and at the same time provided observation hints to the adversary. The report also noted that some users were aware of the security implications of their interaction strategies. It is not clear from the descriptions whether adversaries used the experimental scheme to input their guesses. Tan et al. [19] studied the design of a virtual keyboard meant to be resilient against observation attacks. They recruited participants as adversaries as well. They did not consider feasible strategies or input errors by adversaries.

Dunphy et al. [16] studied a graphical password scheme on smartphones, including the scheme's resilience to observation attacks, in a fashion comparable to what we have repeatedly described. Renaud and Maguire [20] asked study participants to enter their guesses on paper, as did Roth et al. [13], and assessed participants' confidence that they entered a guess correctly, using a questionnaire. Biddle et al. [21] surveyed twelve years worth of publications on graphical password entry systems, some of which were studied in the publications we just cited, and reported password entry success rates ranging from the sixties and lower to the nineties, in percent. In various short papers, notes and extended abstracts, researcher present additional password schemes and report that they conducted a security study. However, the descriptions of their studies are often quite brief [22]–[24]. Of these studies, those that used the experimental scheme for measuring the success of observation attacks were susceptible to misinterpretation of results due to

input errors.

A number of authors proposed human-computer authentication schemes without studying their schemes' resilience to observation attacks. For example, Dhamija and Perrig [25] proposed a scheme based on computer-generated random art portfolios. Jermyn et al. [18] proposed drawing passwords on a grid, and Thorpe at al. [26] presented a scheme whereby users authenticate themselves by identifying a secret location on a geographic map. Watanabe et al. [7] introduce the fake cursor scheme *CursorCamouflage.* Wiedenbeck et al. [27] proposed a widely cited shoulder-surfing resistant graphical password scheme but they did not present a security study or a formal security analysis. Four years later Asghar et al. [28] analyzed the scheme and presented two possible attacks.

IV. Conclusion

We have reviewed and discussed a variety of research proposals in the literature that present password input methods meant to protect against human shoulder surfers. Many of these proposals come with a shoulder surfing study. We found that the ways shoulder surfing security is studied is varied and the outcomes can be difficult to compare and interpret (see, e.g., Tables I and II). Moreover, a number of study design choices may have a subtle or even significant impact on the validity and interpretation of the outcomes. A significant challenge is the opportunity for observation strategies that, once they become known, can cause significant changes in the actual security of published schemes. The current approach of measuring all-or-nothing success of participants cast in the role of adversaries is unsatisfying. If we wish to make progress towards a better assessment of shouder surfing security we need to develop instruments that take potential strategies into account or eliminate strategies as an uncertain element.

References

[1] D. Kim, P. Dunphy, P. Briggs, J. Hook, J. W. Nicholson, J. Nicholson, and P. Olivier, "Multi-touch authentication on tabletops," in *Proc. CHI*, 2010, pp. 1093–1102.

[2] J. Y. Han, "Low-cost Multi-touch Sensing Through Frustrated Total Internal Reflection," in *Proc. UIST*, 2005, pp. 115–118.

[3] F. Schaub, M. Walch, B. Könings, and M. Weber, "Exploring the Design Space of Graphical Passwords on Smartphones," in *Proc. SOUPS*, 2013, pp. 11:1–11:14.

[4] F. Tari, A. A. Ozok, and S. H. Holden, "A Comparison of Perceived and Real Shoulder-surfing Risks Between Alphanumeric and Graphical Passwords," in *Proc. SOUPS*, 2006, pp. 56–66.

[5] F. Schaub, R. Deyhle, and M. Weber, "Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms," in *Proc. International Conference on Mobile and Ubiquitous Multimedia*, 2012, pp. 13:1–13:10.

[6] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan, "Shoulder Surfing Defence for Recall-based Graphical Passwords," in *Proc. Symposium on Usable Privacy and Security*, ser. SOUPS, 2011, pp. 6:1–6:12.

[7] K. Watanabe, F. Higuchi, M. Inami, and T. Igarashi, "CursorCamouflage: Multiple Dummy Cursors As a Defense Against Shoulder Surfing," in *SIGGRAPH Asia 2012 Emerging Technologies*, ser. SA, 2012, pp. 6:1–6:2.

[8] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann, "Using Fake Cursors to Secure On-screen Password Entry," in *Proc. CHI*, 2013, pp. 2399–2402.

[9] H. Sasamoto, N. Christin, and E. Hayashi, "Undercover: Authentication Usable in Front of Prying Eyes," in *Proc. CHI*, ser. CHI, 2008, pp. 183–192.

[10] A. De Luca, K. Hertzschuch, and H. Hussmann, "ColorPIN: Securing PIN Entry Through Indirect Input," in *Proc. CHI*, 2010, pp. 1103–1106.

[11] T. Perković, M. Čagalj, and N. Saxena, "Shoulder-Surfing Safe Login in a Partially Observable Attacker Model," in *Proceedings of the 14th International Conference on Financial Cryptography and Data Security*, ser. FC'10, 2010, pp. 351–358.

[12] T. Perković, S. Li, A. Mumtaz, S. A. Khayam, Y. Javed, and M. Čagalj, "Breaking Undercover: Exploiting Design Flaws and Nonuniform Human Behavior," in *Proc. Symposium on Usable Privacy and Security*, ser. SOUPS '11, 2011, pp. 5:1–5:15.

[13] V. Roth, K. Richter, and R. Freidinger, "A PIN-entry Method Resilient Against Shoulder Surfing," in *Proc. CCS*, 2004, pp. 236–245.

[14] R. A. Khot, P. Kumaraguru, and K. Srinathan, "WYSWYE: Shoulder Surfing Defense for Recognition Based Graphical Passwords," in *Proc. OzCHI*, 2012, pp. 285–294.

[15] A. De Luca, M. Harbach, E. von Zezschwitz, M.-E. Maurer, B. E. Slawik, H. Hussmann, and M. Smith, "Now You See Me, Now You Don't: Protecting Smartphone Authentication from Shoulder Surfers," in *Proc. CHI*, 2014, pp. 2937–2946.

[16] P. Dunphy, A. P. Heiner, and N. Asokan, "A Closer Look at Recognition-based Graphical Passwords on Mobile Devices," in *Proc. SOUPS*, 2010, pp. 3:1–3:12.

[17] S.-H. Kim, J.-W. Kim, S.-Y. Kim, and H.-G. Cho, "A New Shoulder-surfing Resistant Password for Mobile Environments," in *Proc. 5th Int'l Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC, 2011, pp. 27:1–27:8.

[18] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin, "The design and analysis of graphical passwords," in *Proc. USENIX Security Symposium*, 1999.

[19] D. S. Tan, P. Keyani, and M. Czerwinski, "Spy-resistant Keyboard: More Secure Password Entry on Public Touch Screen Displays," in *Proc. OzCHI*, 2005, pp. 1–10.

[20] K. Renaud and J. Maguire, "Armchair Authentication," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, ser. BCS-HCI '09, 2009, pp. 388–397.

[21] R. Biddle, S. Chiasson, and P. Van Oorschot, "Graphical Passwords: Learning from the First Twelve Years," *ACM Comput. Surv.*, vol. 44, no. 4, pp. 19:1–19:41, Sep. 2012.

[22] J. Nicholson, P. Dunphy, L. Coventry, P. Briggs, and P. Olivier, "A Security Assessment of Tiles: A New Portfolio-based Graphical Authentication System," in *CHI Extended Abstracts*, ser. CHI EA, 2012, pp. 1967–1972.

[23] A. Bianchi, I. Oakley, V. Kostakos, and D. S. Kwon, "The Phone Lock: Audio and Haptic Shoulder-surfing Resistant PIN Entry Methods for Mobile Devices," in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI, 2011, pp. 197–200.

[24] T. Kuribara, B. Shizuki, and J. Tanaka, "Vibrainput: Two-step PIN Entry System Based on Vibration and Visual Information," in *CHI Extended Abstracts*, ser. CHI EA, 2014, pp. 2473–2478.

[25] R. Dhamija and A. Perrig, "Deja Vu: A User Study Using Images for Authentication," in *Proc. 9th USENIX Security Symposium*, 2000, pp. 4–4.

[26] J. Thorpe, B. MacRae, and A. Salehi-Abari, "Usability and Security Evaluation of GeoPass: A Geographic Location-password Scheme," in *Proc. SOUPS*, ser. SOUPS, 2013, pp. 14:1–14:14.

[27] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget, "Design and Evaluation of a Shoulder-surfing Resistant Graphical Password Scheme," in *Proc. Working Conference on Advanced Visual Interfaces*, ser. AVI, 2006, pp. 177–184.

[28] H. J. Asghar, S. Li, J. Pieprzyk, and H. Wang, "Cryptanalysis of the Convex Hull Click Human Identification Protocol," in *Proceedings of the 13th International Conference on Information Security*, ser. ISC'10, 2011, pp. 24–30.