Behaviour & Information Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

User-centred multimodal authentication: securing handheld mobile devices using gaze and touch input

Mohamed Khamis, Karola Marky, Andreas Bulling & Florian Alt

To cite this article: Mohamed Khamis, Karola Marky, Andreas Bulling & Florian Alt (2022) User-centred multimodal authentication: securing handheld mobile devices using gaze and touch input, Behaviour & Information Technology, 41:10, 2061-2083, DOI: 10.1080/0144929X.2022.2069597

To link to this article: https://doi.org/10.1080/0144929X.2022.2069597

Published online: 06 May 2022.

BEHAVIOUR & INFORMATION TECHNOLOGY
2022, VOL. 41, NO. 10, 2061–2083
https://doi.org/10.1080/0144929X.2022.2069597

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

User-centred multimodal authentication: securing handheld mobile devices using gaze and touch input

Mohamed Khamis [a], Karola Marky [a], Andreas Bulling [b] and Florian Alt [c]

[a]School of Computing Science, University of Glasgow, Glasgow, UK; [b]Institute for Visualization and Interactive Systems, University of Stuttgart, Stuttgart, Germany; [c]CODE Research Institute for Cyber Defence, Bundeswehr University Munich, Munich, Germany

ABSTRACT
Handheld mobile devices store a plethora of sensitive data, such as private emails, personal messages, photos, and location data. Authentication is essential to protect access to sensitive data. However, the majority of mobile devices are currently secured by singlemodal authentication schemes which are vulnerable to shoulder surfing, smudge attacks, and thermal attacks. While some authentication schemes protect against one of these attacks, only few schemes address all three of them. We propose multimodal authentication where touch and gaze input are combined to resist shoulder surfing, as well as smudge and thermal attacks. Based on a series of previously published works where we studied the usability of several user-centred multimodal authentication designs and their security against multiple threat models, we provide a comprehensive overview of multimodal authentication on handheld mobile devices. We further present guidelines on how to leverage multiple input modalities for enhancing the usability and security of user authentication on mobile devices.

ARTICLE HISTORY
Received 15 October 2021
Accepted 22 March 2022

KEYWORDS
Privacy; Security; Smartphones; Eye tracking; Multimodal; Usable Privacy and Security

1. Introduction

Today's mobile devices allow users to access private data and perform sensitive actions, such as viewing personal photos or messages as well as making online payments. To protect access to said data and actions, users employ authentication mechanisms to lock their phones. These authentication mechanisms include knowledge-based schemes – like PINs and unlock patterns – and biometric schemes – such as fingerprint authentication and facial recognition. Knowledge-based and biometric schemes suffer from several vulnerabilities: The susceptibility of knowledge-based authentication schemes to shoulder surfing was demonstrated repeatedly (De Luca et al., 2013; Eiband et al., 2017; Khamis et al., 2016; von Zezschwitz et al., 2015; Khamis, Trotter, et al., 2018). These schemes are also vulnerable to thermal attacks (Abdelrahman et al., 2017; Abdrabou et al., 2021, 2020) and smudge attacks (Aviv et al., 2010; Schneegass et al., 2014; von Zezschwitz et al., 2013). While there is no evidence that biometric authentication is vulnerable to these side-channel attacks at the time of publishing this paper, biometric data can be stolen remotely (Stokkenes, Ramachandra, and Busch,2016; Zhang et al., 2015), and once leaked they cannot be changed by users. These are among the reasons Android and iOS require users to set a backup PIN, pattern or password as a fallback method, citing the insecurity of biometric authentication (Google,2016). Requiring a fallback method opens the door for 'bypass attacks' (Tiefenau et al., 2019) where, for example, an attacker may intentionally push their finger against the fingerprint sensor until the system prompts them to use the fallback method, which is vulnerable to the aforementioned side-channel attacks that impact knowledge-based schemes.

This means that we need more secure and usable authentication methods for mobile devices that are resilient to shoulder surfing, thermal and smudge attacks. To combat these threats, this work proposes the usage of multimodal user authentication on mobile devices by combining gaze and touch input to enter passwords. To realise this, we propose two multimodal authentication schemes: GazeTouchPass and GazeTouchPIN. The key differences between these two schemes are as follows: **GazeTouchPass** requires passwords that are composed of both gaze input and touch input. For example, a GazeTouchPass password can be 'Gaze left', 'Touch 1', 'Gaze right', 'Touch 2'. The second system **GazeTouchPIN** uses numeric PINs but allows users to enter them

CONTACT Mohamed Khamis mohamed.khamis@glasgow.ac.uk School of Computing Science, 18 Lilybank Gardens, Glasgow, G12 8RZ, UKniversity of Glasgow, Glasgow, UK

using gaze and touch. For example, to enter '1', the user needs to touch a pair of digits (either '1 and 2' or '0 and 1' depending on the currently shown layout) and then gaze to the left in case if the layout shows '1 and 2' or to the right in case of '0 and 1'. GazeTouchPass and GazeTouchPINare knowledge-based schemes that are resilient to smudge and thermal attacks by design because of relying on gaze input (Katsini et al., 2020).

After proposing the two schemes, we report five usability and security lab studies, with a total of 76 participants. First, we evaluate the schemes' usability in two usability studies – one for each scheme – shedding light on efficiency, error rates, and memorability. Second, we evaluate the resistance of each scheme against advanced shoulder-surfing attacks through three security studies. For this, we considered three realistic threat models: (1) *iterative observation attacks* where the attacker first observes the user's gaze input in one occasion, then observes their touch input in another occasion, and finally combines the observations to infer the password; (2) *side observation attacks* where the attacker finds the ideal angle from which they can see the user's gaze and touch input at the same time; and (3) *multiple shoulder surfers* where a pair of attackers simultaneously observes the user during authentication, each focusing on either gaze or touch input. The usability studies reveal that entering a 4-symbol multimodal password using GazeTouchPass takes 3.14 seconds on average, while a 4-digit PIN entered using GazeTouchPINrequires 10.82 seconds. The results of our security studies show that multimodal authentication using gaze and touch significantly improves resilience to observation attacks in all investigated threat models compared to a unimodal authentication baseline that uses touch to enter 4-digit PINs. However, GazeTouchPass is particularly more secure against side observation attacks, whereas GazeTouchPIN is more secure iterative observation attacks. Based on our investigations, we conclude with guidelines for designing user-centred multimodal authentication.

Research Contribution: In summary, this article makes three main contributions: (1) we introduce the concept of multimodal authentication using a combination of touch and gaze on mobile devices, (2) we present the implementation of two schemes, GazeTouchPass and GazeTouchPIN, and an evaluation of their usability and security considering three advanced yet realistic threat models, and (3) we outline guidelines for designing usable and secure multimodal authentication.

The rest of the paper is structured as follows: Section 2 discusses related work and highlights key differences to previous research. Section 3 presents the concept and implementations of GazeTouchPass and GazeTouchPIN, as well as the three threat models considered in this work. Section 4 reports on two usability studies evaluating GazeTouchPass and GazeTouchPIN respectively. Section 5 presents three security studies: The first two studies focus on one system each, and assess their observation resistance against two threat models, whereas the third study evaluates both systems against the third threat model. Section 6 discusses the results and outlines our guidelines for usable and secure multimodal authentication.

2. Related work

We build on several strands of prior work, most notably shoulder-surfing resistant authentication, gaze for authentication, and multimodal authentication on mobile devices.

2.1. Shoulder-surfing resistant authentication

State-of-the-art approaches to counter shoulder-surfing aim to make eavesdropping of password entries difficult for attackers. Multiple previous works rely on presenting users with *cues* that impact the way they enter their passwords. Examples of schemes that incorporated visual cues include SwiPIN (von Zezschwitz et al., 2015) and CueAuth (Khamis, Trotter, et al., 2018) which display arrows on each digit on a 10-digit PIN pad. Users then indicate their input by performing a gesture that matches the arrow on the digit they wish to enter. Other approaches employed haptic cues, such as VibraPass (De Luca, von Zezschwitz, and Hußmann,2009) that uses haptic cues to communicate to users whether they should enter correct or incorrect PIN digits to confuse shoulder surfers. Bianchi et al. proposed a number of authentication schemes that use haptic and audio cues: PhoneLock (Bianchi et al., 2011), SpinLock (Bianchi, Oakley, and Kwon,2011), TimeLock (Bianchi, Oakley, and Kwon,2012) and Colorlock (Bianchi, Oakley, and Kwon,2012). While those schemes are promising for resisting shoulder-surfing attacks, a common issue in cue-based authentication is a long authentication duration due to the time required to observe the cue before providing input. For instance, SwiPIN requires 3.7 seconds to authenticate (von Zezschwitz et al., 2015), whereas PhoneLock requires up to 28 seconds (Bianchi et al., 2011).

The aforementioned schemes inspire our multimodal schemes, in particular GazeTouchPIN . We learned from previous work that cue-based authentication is secure against observation but can be significantly slower when displaying too many cues. Further, using cues that require time to perceive (e.g.vibration patterns), or when users need to perform a linear search (e.g.find a digit in a completely randomised

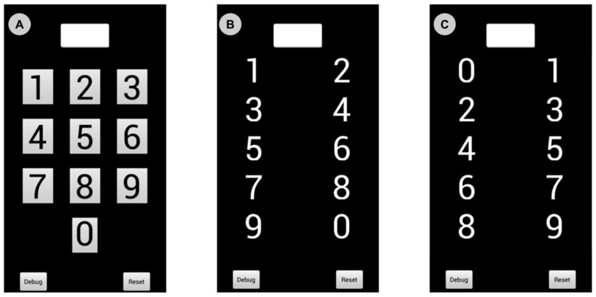


Figure 1. Layout (a) was used for GazeTouchPass and touch-only (GazeTouchPIN 's baseline). Layouts (b) and (c) are the two possible layouts for the touch+random as well as for the GazeTouchPIN system.

arrangement of digits). Thus, in GazeTouchPIN , users are shown one of only two random layouts (see Figure 1(b,c)). The choice of layout to display is determined randomly at the entry of each of the 4-digit PIN.

2.2. Gaze for authentication on mobile devices

There has been significant progress recently in gaze estimation, allowing eye tracking (Hohlfeld et al., 2015; Wood and Bulling.,2014; Ishimaru et al., 2013; Krafka et al., 2016; Khamis, Baier, et al., 2018) and the detection of gaze gestures (Khamis et al., 2016; Vaitukaitis and Bulling,2012; Zhang, Kulkarni, and Morris,2017) using front-facing cameras that are readily integrated in mobile devices. For an overview of eye tracking on mobile devices, we refer the reader to the survey by Khamis, Alt, and Bulling (2018).

Gaze was shown to be a promising modality for password entry in desktop settings (Best and Duchowski,2016; Cymek et al., 2014; De Luca, Denzel, and Hussmann,2009; De Luca, Weiss, and Drewes,2007; Forget, Chiasson, and Biddle,2010; Kumar et al., 2007; Sakai et al., 2016; Sluganovic et al., 2016; Khamis, Trotter, et al., 2018; Abdrabou et al., 2019). Gaze is also a popular choice for biometric authentication (Kinnunen, Sedlak, and Bednarik,2010; Song et al., 2016; Rigas, Abdulin, and Komogortsev,2016). Researchers have also utilised gaze for improving password selection (Alt et al., 2016; Bulling, Alt, and Schmidt,2012), password recall (Sridharan et al., 2016) and understanding user's password choice strategies (Katsini et al., 2019). For a review of the use of gaze for both knowledge-based and biometric authentication, we refer the reader to the work of Katsini et al. (2020).

Prior work shows that gaze is hard to observe (Almoctar et al., 2018), however by observing the user's eyes (instead of the screen), attackers may still eavesdrop password (De Luca, Denzel, and Hussmann,2009). To offset such an attack, schemes based on Electrooculography (EOG) have been demonstrated to work even with closed eyes when users where EOG glasses (Dieter Findling, Quddus, and Sigg,2019). Compared to existing schemes, the novelty of our schemes lies in the combination of gaze and touch input on unmodified mobile devices. Consequently, attackers would need to (a) observe the user's gaze input, (b) observe the user's touch input, and (c) combine both observations. For these reasons, we opted for evaluating our schemes under threat models that go beyond simple one-time observations.

2.3. Multimodal authentication on mobile devices

Although GazeTouchPass and GazeTouchPIN are the first authentication schemes that combine gaze and touch on mobile devices, there have been other schemes that employ multiple modalities. For example, PhoneLock (Bianchi et al., 2011), SpinLock (Bianchi, Oakley, and Kwon,2011), TimeLock (Bianchi, Oakley, and Kwon,2012), and Colorlock (Bianchi, Oakley, and Kwon,2012) resist shoulder surfing by using combinations of audio and haptic feedback as cues for password entry. The idea behind these systems is using a hidden output channel for cues that only users can perceive. Using cues has a positive influence on shoulder surfing resistance which inspired our implementation of GazeTouchPIN , where we use a randomised visual cue that is difficult to observe simultaneously while observing the user's eyes.

Another feature of GazeTouchPass and GazeTouchPIN is that they split the attacker's attention because gaze input and touch input need to be observed simultaneously. In terms of input-splitting, XSide by De Luca et al. (2014) is most similar to our work. XSide exploits the back of the device interaction to make observations more difficult. It was found that splitting the input strongly influences the observation resistance of a system as it requires splitting the attackers' attention. This conclusion influenced the design of our systems as we demonstrate in the following sections.

Unlike the aforementioned multimodal schemes, users of GazeTouchPass and GazeTouchPIN do not need any additional hardware (e.g. motors, earplugs or double-sided touch screens). The users only need one hand for interaction, which is preferred by users over two-handed interaction (Karlson, Bederson, and SanGiovanni,2005).

While preliminary evaluations of GazeTouchPass and GazeTouchPIN were reported in Khamis et al.

(2016), Khamis, Hassib, et al. (2017) and Khamis, Bandelow, et al. (2017), we significantly extend that work by (a) directly comparing the GazeTouchPass and GazeTouchPIN , (b) presenting guidelines for usable and secure multimodal authentication, (c) reflecting on previous work in that topic in more depth, (d) including an in-depth discussion that reflects on the results, ethical considerations, contributions in practice, and comparison to related work, and (e) reporting results on the memorability of GazeTouchPass.

3. Multimodal authentication using gaze and touch

In this section, we present the concept and implementations of each of GazeTouchPass and GazeTouchPIN . Both schemes are implemented as Android apps and do not require any additional hardware, because the gaze gestures are detected using the front-facing camera that is readily integrated into off-the-shelf mobile devices. Even though there is a recent uptake of front-facing depth cameras, which typically improve eye tracking accuracy (Khamis, Alt, and Bulling,2018), we used standard video (RGB) front-facing cameras to ensure compatibility with the majority of smartphones. The user's face and eyes are first detected using a Viola–Jones detector (Viola and Jones,2004). We then adapted a method proposed by Zhang, Bulling, and Gellersen (2014) for detecting gestures to the left and to the right without the need for eye tracking calibration. We were careful to avoid requiring calibration because calibration is known to be perceived as a tedious and a time-consuming task (Majaranta and Bulling,2014). We further followed the recommendation by Katsini et al. that gaze-based authentication should not require calibration due to its negative impact on usability (Katsini et al., 2020). While the method by Zhang, Bulling, and Gellersen (2014) measures the distance between the user's pupil centre and the eye corner in each eye, our method measures the distance between the face's centre and the pupil for each eye. We opted for relying on the face's centre rather than eye corner as low-resolution cameras are more likely to accurately detect the face rather than the eye corner. Gaze directions are then estimated based on the ratio between both distances.

3.1. GazeTouchPass

GazeTouchPass combines touch-based PINs (0–9) and gaze gestures (left and right) for authentication. The system uses a theoretical password space of (12^n), where n denotes the length of the password. Our prototype uses a length of n=4 to allow comparing GazeTouchPass to prior work. However, a deployed version of the system would allow longer inputs and would require a minimum length of inputs using each modality to ensure higher security. The user interface consists of a 10-digit keypad as shown in Figure 1(a). Users log in by touching digits and moving their eyes to the left or right.

Table 1. Sample GazeTouchPass passwords.

Condition	Example 1	Example 2
0-switches (*baseline*)	1-2-3-4	left-right-left-left
1-switch	left-1-2-3	1-2-left-right
2-switches	left-1-left-right	left-1-2-right
3-switches	1-left-2-right	left-1-right-2

Notes: We studied the effect of the number of switches between gaze and touch input (*modality-switch-count*). We expect that the more switches between modalities a password has, the more resistant it is to shoulder surfing. 0-switches is the baseline condition used when evaluating GazeTouchPass , as it represents a unimodal password consisting of touch input only or gaze input only.

Examples of GazeTouchPass passwords are shown in Table 1. Because GazeTouchPass passwords consist of two types of input – gaze input and touch input – they introduce a new feature to passwords which we refer to as *modality-switch-count* that denote a change from one input method to another. We expect that the higher number of switches from gaze to touch input or vice versa, the more difficult it will be to observe it. Namely, we expect a password, such as '1-left-2-right' (3-switches), to be more secure than '1-2-left-right' (1-switch). The reason for this is that from the perspective of an attacker, each *modality-switch* is equivalent to a switch of the attacker's focus between the touchscreen and the eyes (see Figure 2 Camera C).

3.2. GazeTouchPIN

GazeTouchPIN differs from GazeTouchPass in a number of ways. While GazeTouchPass combines gaze and touch into multimodal passwords (e.g. left-3-right-4), GazeTouchPIN uses classical 4-digit PINs that are entered using gaze and touch input (e.g. 1234). In GazeTouchPIN , users select the digit they wish to enter in two steps: in Step (1), they select a pair of digits from one of two layouts shown in Figures 1(b,c), before Step (2) gaze left or right to indicate the desired digit. For example, if a user is shown Layout B in Figure 1, touches the pair (1, 2), and then gazes to the right, then they have selected '2'. The choice of layout is determined randomly at every entry (e.g. four times for a 4-digit PIN). The reason we opted for showing a random layout is as follows: We expect this approach to be resistant to typical shoulder surfing attacks; at every entry of a 4-digits PIN, observing the touchscreen

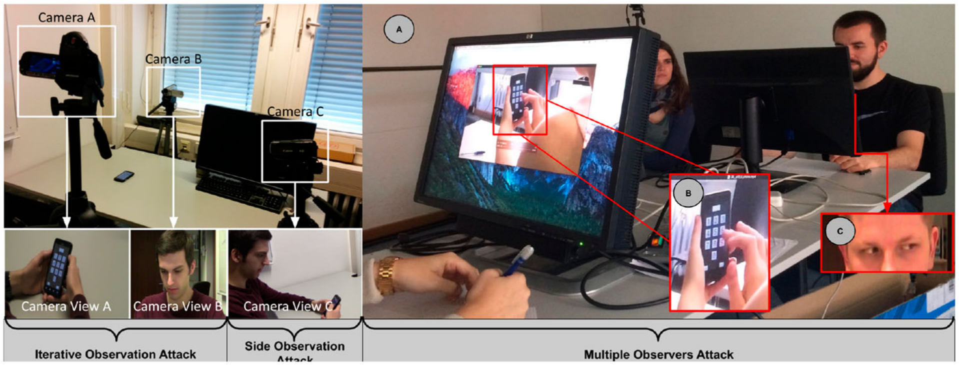


Figure 2. The figure shows the camera setup used for both usability studies. To prepare videos for the subsequent security studies, we recorded users using three cameras. Camera A recorded the phone screen (phone-view) to observe the touch input. Camera B recorded the participant's face (eyes-view) to observe the eye movements. Camera C simultaneously recorded the screen and the user's eyes (side-view). The views from Camera A and B were used to evaluate the schemes' resistance to iterative observation attacks, whereas the view from Camera C was used to evaluate resistance to side observation attacks.

would result in a pair of digits. An attacker who observes all touch inputs would still have to try 2^n possibilities to determine the correct PIN combination (where n denotes the number of digits in the PIN). Moreover, if an attacker observes one modality input after another (e.g. observing the eyes after observing the touchscreen), the attacker would not know which layout the user is responding to. There is only a $\frac{1}{2^n}$ chance that the attacker observes matching touch and gaze input. This makes the approach resistant to iterative attacks. In contrast to GazeTouchPass , attackers of GazeTouchPIN can predict which modality will be used next; users of GazeTouchPIN perform a touch input followed by a gaze gesture. Nevertheless, even when observing from an optimal side view that shows the user's eyes and touchscreen clearly, attackers would have to quickly switch focus between the eyes and the screen. GazeTouchPIN uses 4-digit PINs, thus maintaining the memorability and the password space of classical PIN-based systems, which has been studied extensively in prior work (von Zezschwitz, Dunphy, and De Luca,2013).

At the same time, having only two layouts supports learning effects and avoids any cognitive load caused by selecting from a completely randomised arrangement of digits.

3.3. Threat models

In this section we describe the threat models we evaluate GazeTouchPass and GazeTouchPIN against. The security evaluations are reported in Section 5.

The traditional threat model for shoulder-surfing attacks where an attacker observes the user during input would be trivial and of low effectiveness against our proposed schemes. Thus, we cover three advanced shoulder-surfing attacks. In each threat model, the user is in a public space and the attacker(s) know how the authentication schemes work. After observing the password, the attacker(s) get hold of the device (e.g. by stealing it or as the user leaves it unattended), and try to log in using the observed password.

3.3.1. Threat model 1: side observation attacks

In this threat model, the user is observed from a viewpoint that allows the user's gaze input as well as touch input to be eavesdropped (e.g. in a train). The distance between the attacker and the user is close enough to see the touchscreen, but far enough to reduce the effort of switching focus back and forth between the user's eyes and the device's touchscreen (see Camera C in Figure 2).

3.3.2. Threat model 2: iterative observation attacks

In this model, the attacker has the chance to observe the user twice: (1) the attacker exclusively focuses on one modality per observation – for example, first on the input on the screen (Camera A in Figure 2) and (2) on the users' eyes (Camera B in Figure 2), or vice versa. The challenge of this attack is to correctly observe both sequences and to correctly combine them later.

3.3.3. Threat model 3: multiple observers attacks

In this threat model, two adversaries are simultaneously observing the user. The pair decides upfront on an observation strategy. Each of the two has a chance to observe part of the authentication process from an optimal angle (see Figure 2). The attackers then discuss how their observations can be combined. This threat model is motivated by previous work that showed that multiple people sometimes simultaneously shoulder surf user (Eiband et al., 2017), and by real-world theft, pick-pocketing and burglary situations, where there are often multiple adversaries.

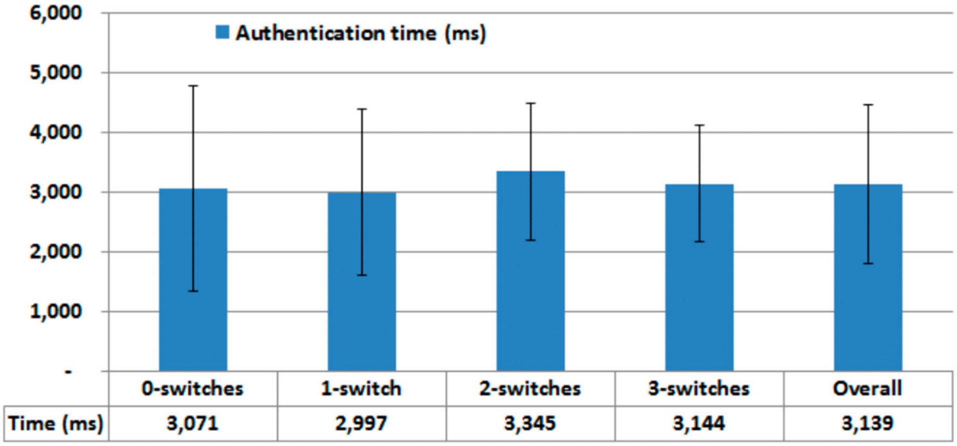


Figure 3. Mean authentication times for passwords with different numbers of modality switches. Error bars represent the standard deviation. Authentication times do not vary significantly among different number of modality switches. Overall mean authentication time is 3.1 seconds ($SD = 1.3$).

4. Usability evaluations

We evaluated the usability of GazeTouchPass and GazeTouchPIN in two separate user studies. Both user studies used a within-subjects design. We detail the independent variables of each study in the sections below.

4.1. Usability study 1: usability of GazeTouchPass

The aim of this study was to analyse the usability of GazeTouchPass and to collect video recordings of gaze and touch input for the subsequent security studies. The study had one independent variable: *modality-switch-count* , which had four conditions: 0-switches (baseline), 1-switch, 2-switches, and 3-switches (see Table 1). As this a repeated measures experiment, each participant went through all conditions by performing 16 authentications (4 passwords × 4 conditions) using randomly generated passwords. Recall that GazeTouchPass passwords consist of digits (0–9) and gaze directions (left and right) as detailed in Section 3.1.

4.1.1. Usability study 1 participants

We recruited 13 participants (4 males and 9 females), aged between 21 and 35 years ($M = 25.23$, $SD = 3.8$) through mailing lists. All participants had normal or corrected-to-normal vision. Five reported to use PINs as authentication mechanism. Others used lock patterns, graphical passwords, and TouchID. Participants were compensated with an online shopping voucher.

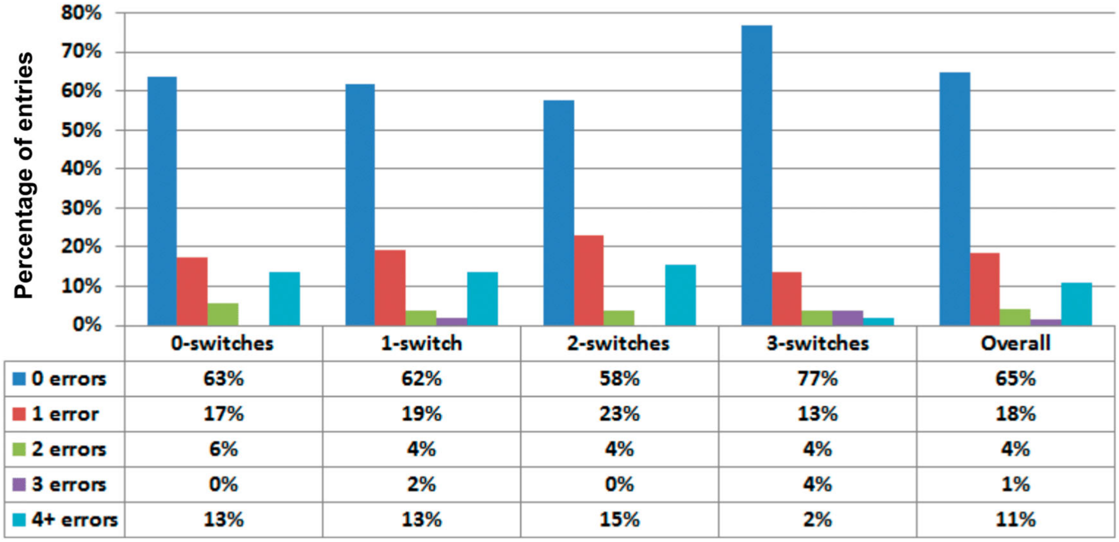
4.1.2. Usability study 1 procedure

Upon arrival participants were asked to sit at a table in a meeting room. The experimenter then explained the study and asked the participant to sign a consent form. Afterwards, the experimenter started the app on the smartphone, described how it worked and handed it to the participant. Each participant was then allowed to perform four training runs, one per condition, to get acquainted with the system. Those authentication attempts were excluded from further analyses. At each authentication attempt, the experimenter read out the password to be entered according to a previously generated, randomised list. We logged all authentication attempts and recorded the participants using three HD video cameras (see Figure 2). Participants repeated entry in case of an unsuccessful login. After entering all 16 passwords, we then asked the participant to freely define a GazeTouchPass password of their own choice. We did not set any requirements for that password. This step was done to evaluate memorability at a later stage and to understand user choices of passwords. We concluded the study with a semi-structured interview.

4.1.3. Usability study 1 results

Each participant entered 16 passwords, each four representing one condition, resulting in a total of $13 \times 16 = 208$ GazeTouchPass password entries. Three videos were recorded per password entry for each camera view (624 videos). We evaluated the system's usability by operationalising efficiency as input speed and effectiveness as error rates.

Input Speed. We measured the time taken to input the passwords starting from the moment the user touches the screen for the first time till the moment the fourth entry is detected by the system. Figure 3 suggests that mean authentication times do not vary greatly among different number of modality switches. Overall mean authentication time is 3.1 seconds (SD=1.3). For our analysis, we first excluded 3 out of 72 input time measurements as outliers ($> \mu + 3\times$



	0-switches	1-switch	2-switches	3-switches	Overall
🟦 0 errors	63%	62%	58%	77%	65%
🟥 1 error	17%	19%	23%	13%	18%
🟩 2 errors	6%	4%	4%	4%	4%
🟪 3 errors	0%	2%	0%	4%	1%
■ 4+ errors	13%	13%	15%	2%	11%

Figure 4. Number of attempts before a successful entry. Errors are less for passwords with 3-switches; consecutive gaze gestures can be error prone, while 3-switches in an $n = 4$ password can be only achieved by alternating gaze and touch input.

SD). No significant main effects were found for *modality-switch-count* on authentication time ($p > .05$).

Error Rates. We also logged the number of failed login attempts, which were false detection by the system. Figure 4 shows that there were fewer errors in the case of passwords with 3-switches. While providing multiple consecutive gaze gestures can be error prone, having 3 switches in a 4-digit password can be achieved only by alternating gaze gestures and digits.

Qualitative Feedback. After interaction, we gathered qualitative feedback from the participants through a short interview. Six out of 13 participants reported they would use GazeTouchPass as a primary authentication scheme. Nine reported that they would not use it for daily unlocking, but rather for insecure situations (e.g. surrounded by others) or to protect sensitive data, such as online banking apps. One participant indicated that they would be willing to use GazeTouchPass for a one-time unlock (e.g.when switching the phone on). Four participants said they would not be willing to do anything extra for higher security; two of them added that they do not use any lock mechanism on their phones.

Memorability. We informed the participants that they would be asked for the passwords they selected for the memorability test in the future, without specifying a date. We emailed the participants five days after the study asking them for the passwords they selected. Participants had up to three guesses to provide their password. Eleven out of 13 participants remembered their passwords – 10 were correct on the first guess, one was correct on the second guess, and two could not correctly recall their password after three guesses.

4.2. Usability study 2: usability of GazeTouchPIN

Similar to GazeTouchPass 's usability evaluation, the aim of this study is to evaluate the usability of GazeTouchPIN and to collect realistic password entries for the subsequent security study.

GazeTouchPIN uses 4-digit PINs, thus maintaining the memorability and the password space of classical PIN-based systems, which has been studied extensively in prior work (von Zezschwitz, Dunphy, and De Luca,2013). Thus, this usability study focuses only on efficiency and effectiveness (i.e. input speed and error rate). To understand the impact of using gaze and touch to enter 4-digit PINs, and to distinguish the impact of the randomised layout from that of gaze and touch input, we compared GazeTouchPIN to two baselines. To understand how GazeTouchPIN performs compared to standard unimodal 4-digit PINs, and to differentiate the impact of the random layout from the impact of input using touch and gaze, we include one independent variable (input method) with the following three conditions:

(1) The *touch-only* (Figure 1(a)) method uses the traditional PIN keypad (baseline). This served as a baseline that uses touch input only.
(2) The *touch+random* (Figure 1(b,c)) method uses touch to select the desired digit from one of two randomly shuffling layouts. This will provide insights about the shuffling idea and help distinguish the impact of the multimodal factor.
(3) GazeTouchPIN (Figure 1(b,c)) uses touch input to select a pair of horizontally aligned PIN digits and

then a gaze gesture to the left/right to select the desired PIN.

4.2.1. Usability study 2 participants

We recruited 12 participants (2 females, 10 males), aged between 19 and 31 years ($M = 24.8$, $SD = 3.6$), through mailing lists. Asked about whether they use authentication on their phones, participants reported using TouchID, lock pattern and PINs. All participants had normal or corrected-to-normal vision.

4.2.2. Usability study 2 procedure

We followed a procedure similar to the one used in the usability study of GazeTouchPass . Participants were allowed to perform three training runs, one with each method, to get acquainted with the different methods. Furthermore, in this usability study the experimenter read out the input method to be used in addition to the PIN at each authentication attempt according to a previously generated randomised list. Participants entered 6 pre-defined PINs using all three authentication methods, resulting in 6 PINs × 3 methods = 18 authentications in random order. We logged all authentication attempts and showed the home screen after every successful login. We recorded the participants using three HD video cameras in a similar setup (Figure 2). We concluded the study with a semi-structured interview.

4.2.3. Usability study 2 results

In total we recorded 54 videos per participant (6 passwords × 3 methods × 3 camera views). Apart from the videos, we analysed the data with regard to input speed and error rate.

Input Speed. Figure 5 summarises the time needed to authenticate for each method. Prior to analysis, we excluded 2 out of 216 input time measurements as outliers ($> \mu + 2.5\times$ SD). A repeated measures ANOVA showed significant effects for input method on input speed ($F_{1.021,9.192} = 156.106$, $p < .001$). Post-hoc analyses using Bonferroni correction revealed that there was a significant difference ($p < .001$) in input speed between touch-only input ($M = 1677$, $SD = 120$) and GazeTouchPIN input ($M = 10, 817$, $SD = 712$). There was also a significant difference ($p < .001$) between touch+random input ($M = 3210$, $SD = 124$) and GazeTouchPIN input ($M = 10, 817$, $SD = 712$). The third pair (touch-only vs touch+random) was also significantly different ($p < .001$).

Error Rates. The results show that the error rate of three participants decreased using GazeTouchPIN input as they entered more PINs. Figure 6 shows that the more PINs participants enter using GazeTouchPIN , the less errors occur, which suggests that there is a learning effect. For example, 10 out of 12 participants entered their fifth and sixth PIN correctly on their first attempt. Participants 2 and 6 never failed, while participants 1, 7 and 11 failed once each. Finally, participant 4 improved steadily from 4 failures at the first GazeTouchPIN input to no failures when entering the last PIN.

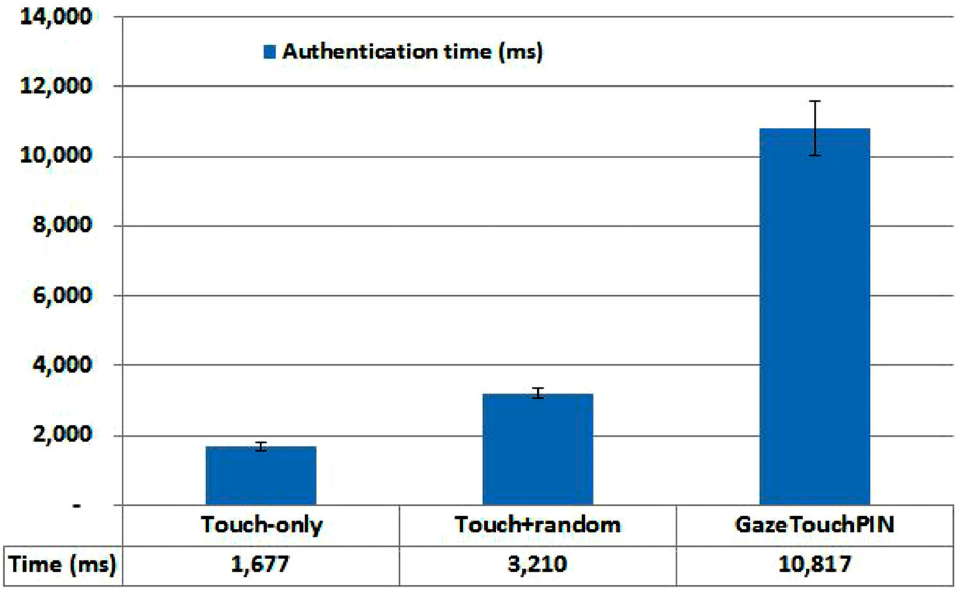
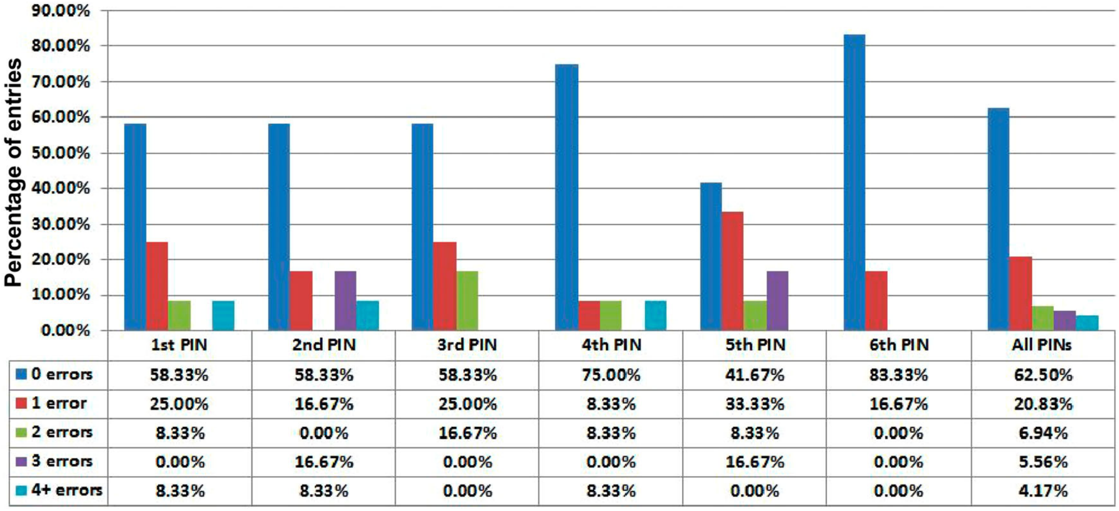


Figure 5. GazeTouchPIN requires on average significantly more time compared to touch+random and touch-only . Participants performed faster over time, with a mean input time decreasing from 10.8 at the first GazeTouchPIN entry to 9.5 seconds at the sixth entry.



	1st PIN	2nd PIN	3rd PIN	4th PIN	5th PIN	6th PIN	All PINs
■ 0 errors	58.33%	58.33%	58.33%	75.00%	41.67%	83.33%	62.50%
■ 1 error	25.00%	16.67%	25.00%	8.33%	33.33%	16.67%	20.83%
■ 2 errors	8.33%	0.00%	16.67%	8.33%	8.33%	0.00%	6.94%
■ 3 errors	0.00%	16.67%	0.00%	0.00%	16.67%	0.00%	5.56%
■ 4+ errors	8.33%	8.33%	0.00%	8.33%	0.00%	0.00%	4.17%

Figure 6. Number of attempts before a successful entry using GazeTouchPIN across all participants. Each participant entered 6 PINs using GazeTouchPIN , the graph shows that users tend to enter their PIN correctly at the first attempt as they enter more PINs.

Qualitative Feedback. Participants noted that the touch+random and GazeTouchPIN were more secure than the regular touch-only method. Despite longer login times, all participants agreed that with some training they would be able to enter PINs even faster. This aligns with the quantitative data, which showed that the mean input time of the participants' first entry using GazeTouchPIN is 10.8 seconds, which decreased to 9.5 seconds at their sixth entry using GazeTouchPIN . This is a decrease of 12%, which is promising especially because the participants were using GazeTouchPIN for the first time. Asked for application areas, participants voiced that they find GazeTouchPIN particularly useful in situations where they are more exposed, such as in public transport. Also using the approach as a second layer of authentication for particular cases (e.g.online banking applications, or for opening messages from a specific person) was mentioned as an application area. Overall while one participant reported that he would use GazeTouchPIN for frequent phone unlocking, 10 participants reported they would use it to protect sensitive data or in situations where they feel observed. One participant explicitly mentioned that he was not too much concerned about the security of his phone ('My phone isn't that important to me'). He stated to be too impatient for permanently using GazeTouchPIN . However, he would like to use it at ATMs to achieve a higher level of security.

The feedback received in this study matches the input by participants of GazeTouchPass 's security study, suggesting that GazeTouchPIN is attractive for security-aware users, while less concerned users would use it in sensitive contexts only.

5. Security evaluations

We evaluated the security of GazeTouchPass and GazeTouchPIN in terms of observation resistance in three user studies. The first two studies focus on GazeTouchPass and GazeTouchPIN respectively, and both studies cover side observation attacks and iterative observation attacks (i.e. threat models 1 and 2 described in Section 3.3). The third security study evaluates both GazeTouchPass and GazeTouchPIN against multiple observers attacks (i.e.threat model 3 described in Section 3.3). Both user studies used a within-subjects design. We detail the independent variables of each study in the sections below.

5.1. Security study 1: security of GazeTouchPass

The aim of this study was to analyse the security of GazeTouchPass in terms of observation resistance against iterative observation attacks and side observation attacks (i.e. threat models 1 and 2 as described in Section 3.3). To evaluate the security, we used the recordings from the preceding usability study to create consistent conditions. Because the recordings showed the participants of the usability study, we obtained their consent for using these videos and screenshots from them for further investigations and publications. The videos were played to the security participants on a computer screen. The security study participants were specifically instructed to try recovering digits and eye moves from the video to mimic an attack. While and after observing the videos, participants were asked to take notes of the observed digits and eye movements.

When performing iterative observation attacks against GazeTouchPass , participants noted the pauses between gaze gestures and then tried to fill the gaps with digits observed through the phone-view . Following a repeated-measures design, participants took the role of an attacker and watched videos of users authenticating using GazeTouchPass . There were two independent variables: 1) *modality-switch-count* (0-switches, 1-switch, 2-switches, 3-switches) and 2) threat model (side observation attacks, iterative observation attacks). This means participants observed successful authentication attempts using all four possible *modality-switch-count* and observed from three angles to cover both threat models (see Camera Views angles in Figure 2). Each participant independently attacked eight passwords of each condition of *n*-switches – half of which were side observation attacks (i.e. using the side-view as shown in Figure 2 Camera View C), while the others were iterative observation attacks (i.e. using the eyes-view and the phone-view as shown in Figure 2 Camera View A and B respectively). In iterative observation attacks, the experimenter alternated the order of the observed view. This results in a total of 32 attacked passwords. The order of videos was randomised per participant. To avoid learning effects, no participant attacked the same password from different views.

5.1.1. Security study 1 participants

We recruited 13 participants (6 females, 7 males), aged between 21 and 33 years ($M = 24.2$, $SD = 3.4$), through mailing lists. None of them had participated in the usability study of GazeTouchPass (Usability Study 1). Participants were compensated with an online shopping voucher. In addition, all participants took part in a draw for an additional 20 Euro voucher, where chances of winning increased with the number of successfully attacked passwords. This was done to motivate participants to put an effort in their observation attacks.

5.1.2. Security study 1 procedure

The experimenter introduced the study procedure, the task, and the reward mechanism. After explaining how GazeTouchPass works, participants had the chance to try and get acquainted with the app themselves. They were then given draft papers and the experimenter started playing the videos. They were given blank papers to take notes during the observation attacks if they wish, then the experimenter started playing the videos. Based on their observations, participants provided up to three guesses for each authentication attempt they observed. Each participant was allowed to watch the video sequences relevant to the current password once on a 24 ″ monitor. The study was concluded with a final questionnaire and a short semi-structured interview. In total, participants performed $13 \times 32 = 416$ attacks with up to three guesses each.

5.1.3. Security study 1 results

In the following we report on the successful attacks against GazeTouchPIN as well as on results of the questionnaire and semi-structured interviews.

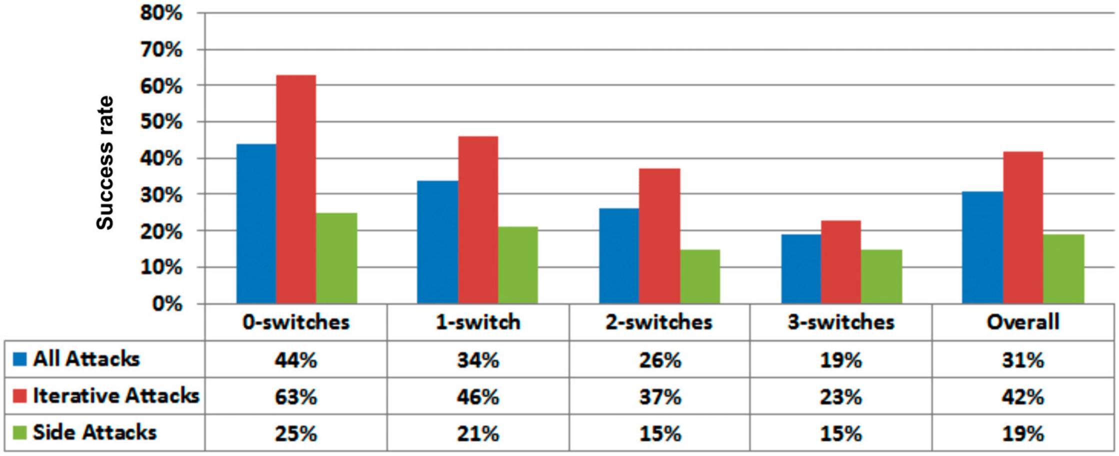
Successful Attacks. For each attack, we calculated the Levenshtein distance between the guesses and the correct password. The use of Levenshtein distance to measure closeness of observation attacks is the standard in previous work (Katsini et al., 2020; von Zezschwitz et al., 2015). Only the guess closest to the correct password was considered for further analysis. Moreover, we calculated the overall success rate in attacking passwords for each number of modality switches and for each attack type (iterative observation attack vs side observation attack). An attack is considered successful if at least one guess matched the correct password. Figure 7 summarises the successful attack rate against passwords with different *modality-switch-count*, observed through the side-view or through the phone-view and the eyes-view.

A two-way repeated-measures ANOVA showed significant main effects for *modality-switch-count* on attack success ($F_{3,36} = 3.86, p < .05$) and for the view angle on attack success ($F_{1,12} = 51.05, p < .0001$). There were no interaction effects between *modality-switch-count* and view angle ($p > .05$).

This suggests that distance between the guesses and the correct password depends on the *modality-switch-count*. Post-hoc analysis with Bonferroni correction showed a significant difference ($p < .05$) in attack success for passwords with 0-switches ($M = 1.25, SD = 0.14$) compared to those with 3-switches ($M = 1.9$, $SD = 0.1$). This means guesses against passwords with 0-switches in modality (baseline) are closer to the correct pattern than those with 3-switches. The other pairs did not show any significant differences ($p > .05$).

Post-hoc analysis with Bonferroni correction revealed that there was a significant difference ($p < .0001$) in attack success for passwords attacked iteratively ($M = 1.38, SD = 0.138$) compared to passwords attacked from the side ($M = 1.913$, $SD = 0.123$). This suggests that guesses against passwords observed iteratively (threat model 2) are closer to the correct password compared to those observed from the side (threat model 1).

Qualitative Feedback. When asked in the questionnaire how easy it was to attack passwords for each view (5-point scale; 1=Very easy; 5=Very difficult), participants found side attacks to be very difficult ($Med = 5$, $SD = 0.66$), while iterative attacks were perceived to be



	0-switches	1-switch	2-switches	3-switches	Overall
🟦 All Attacks	44%	34%	26%	19%	31%
🟥 Iterative Attacks	63%	46%	37%	23%	42%
🟩 Side Attacks	25%	21%	15%	15%	19%

Figure 7. Success rate when attacking passwords entered using GazeTouchPass in Security Study 1. Passwords with higher *modality-switch-count* are significantly more secure against observations compared to those with fewer *modality-switch-count* . Side attacks are always less successful than iterative attacks due to the difficulty of continuously switching focus back and forth from the eyes to the touchscreen.

easier ($Med = 3$, $SD = 0.96$). In the interviews, eight participants expressed that attacking touch-only and gaze-only passwords was easiest. One participant reported it was easier to break passwords with consecutive inputs of the same modality. There was a disagreement among participants regarding which modality was more difficult to observe. While some found gaze input to be more difficult to observe than touch input, others found gaze input easier. Participants reported side observation attacks to be harder as it was difficult to concentrate on the eyes and the display at the same time. Three participants said that they had trouble finding the right order during iterative observation attacks. They also reported that it is harder to attack passwords entered quickly. It is expected that users will authenticate faster as they use the system more often, making the system even more secure.

5.2. Security study 2: security of GazeTouchPIN

This study also followed a repeated-measures design with the aim to analyse the security of GazeTouchPass in terms of observation resistance against iterative observation attacks and side observation attacks (i.e.threat models 1 and 2 as described in Section 3.3). As done in Security Study 1, we used the videos that were collected from the participants of the usability study of GazeTouchPIN . The participants had consented to using the videos with their faces in publications and further user studies. In total, each participant of Security Study 2 attacked 24 PIN entries – 8 for each input method: touch-only , touch+random, and GazeTouchPIN . Participants performed half of the 24 attacks using the side-view and the other half using the phone-view . For iterative attacks against the GazeTouchPIN method, participants were provided both the eyes-view as well as the phone-view . Half of these started by the eyes-view , while the other half started with the phone-view . When observing GazeTouchPIN , participants noted down the gaze gestures and the pairs of digits selected every time.

For any two observations against GazeTouchPIN , there is a $\frac{1}{2^n}$ chance (where n is the number of PIN digits) that the phone-view and the eyes-view match. Hence, we randomly assigned the views such that there was a $\frac{1}{16}$ chance for a match (given that we used 4-digit PINs). The order of PINs and methods was randomised per participant. To avoid learning effects, no participant attacked the same password from different views.

5.2.1. Security study 2 participants

We recruited 18 participants (5 females) aged between 18 and 36 ($M = 24.6$, $SD = 4.54$) through mailing lists. We employed a reward system identical to the one used in GazeTouchPass 's security study (Security Study 1 described in Section 5.1). None of the participants of the security study had participated in the usability study of GazeTouchPass.

5.2.2. Security study 2 procedure

We followed the same procedure and reward mechanism used for GazeTouchPass 's security study (see Section 5.1.2). We additionally allowed participants to examine the layouts at any time during the study (see Figure 1).

5.2.3. Security study 2 results

In the following we report on the successful attacks as well as on interview results and a questionnaire.

Successful Attacks. In total, participants performed $18 \times 24 = 432$ attacks, providing three guesses for each. We calculated the Levenshtein distance in the same manner as in GazeTouchPass 's security study. Figure 8 shows the rate of successful attacks against PINs entered using each of the methods, observed either through the side-view or through the phone-view and the eyes-view. All three graphs show that the success rate is lower for GazeTouchPIN.

A repeated-measures ANOVA showed significant main effects for input method ($F_{2,34} = 42.36$, $p < 0.001$) on attack success. This suggests that the distance between the guesses and the correct PIN depends on the input method. No significant main effects were found for the number of PINs attacked so far.

Post-hoc analysis using Bonferroni correction revealed that there was a significant difference ($p < .001$) in the distances for PINs entered using GazeTouchPIN ($M = 1.88$, $SD = 0.11$) compared to touch-only PINs ($M = 0.65, SD = 0.1$). There was also a significant difference ($p < .005$) in the distances for PINs entered using GazeTouchPIN ($M = 1.88$, $SD = 0.11$) compared to touch+random PINs ($M = 1.37$, $SD = 0.13$). The final pair was also significantly different ($p < .001$). This means that guesses against PINs were statistically closer to the correct PIN in case of touch-only PINs, followed by touch+random PINs. However guesses against GazeTouchPIN PINs were the least similar to the correct one.

Questionnaire and Interviews. All participants reported that attacking multimodal PINs (GazeTouchPIN) through the side-view was the most difficult task. Some attributed this to the difficulty of focusing on the eyes and phone in parallel, particularly if the users were fast in entering their password. 'It is just very hard to concentrate on two numbers, look at his eyes, then again at the screen', said P0. One participant noted that she had to keep track of: (1) the user's finger, (2) which layout is displayed and (3) the eye movements. Another participant seconded her, adding that he found it particularly difficult when the user used multiple fingers when entering the password. 'It is only possible when there is a long gap between row selection and eye movements', said P2, implying that GazeTouchPIN 's entry speed is also influential. Multiple participants indicated that shuffling the layout confused them. After the study, participants were asked how easy it was to attack the passwords for each method and view (5-point scale; 1=Very easy; 5=Very difficult). Note that they were not aware of how many of their attempts were successful during the study. Table 2 summarises the median scores of the perceived difficulty. It can be seen that side attacks are the hardest, with a median score of *Very Difficult*.

5.3. Security study 3: security of GazeTouchPass and GazeTouchPIN against multiple observers

The main goal of this study is to investigate how the multiple observers threat model (threat model 3) influences the security of GazeTouchPass and GazeTouchPIN. We used the videos recorded during usability studies 1 and 2 (Section 4) and used in security studies 1 and 2.

5.3.1. Security study 3 design

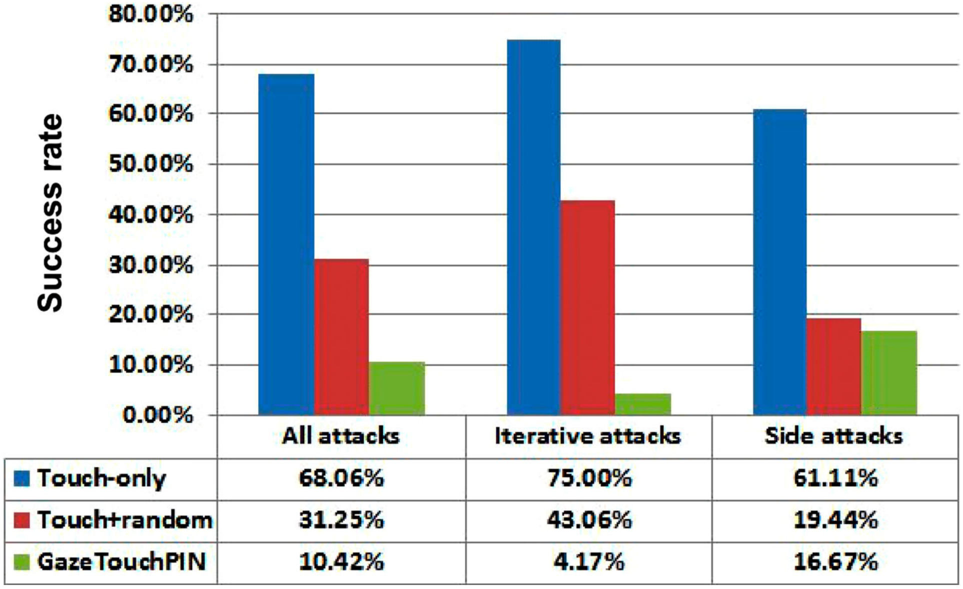
The study was designed as a repeated-measures experiment with a single independent variable: the password type. As explained above, GazeTouchPass passwords can consist of multiple switches in input modality. Hence, we included four conditions: 3-switches, 2-switches, 1-switch, and 0-switches. The last condition refers to having no switches in modalities when entering the password i.e. a unimodal password. This means that when two observers attack GazeTouchPass with 0-switches, they will be both observing the same modality. This was considered a baseline in our experiment. The fifth and last condition is GazeTouchPIN. Each team of attackers observed 3 passwords of each type. This means that each team attacked 15 passwords in total (3 passwords × 5 password types). The conditions were counter-balanced using a latin square.

5.3.2. Security study 3 participants

We invited 20 participants (9 females, 11 males) in pairs of two to take the role of an attacker team. The study was advertised through mailing lists.

5.3.3. Security study 3 procedure

We invited participants in teams of two. The experimenter explained the study and asked the participants to sign a consent form. Participants were then explained how GazeTouchPass and GazeTouchPIN worked and had the chance to try them out, and watch videos showing how they work. Each team then watched two video clips on two different 17 ″ displays (see Figure 9). Both videos started at the same time. The participants were free to examine the layouts (Figure 1) and to take notes at any time during the study. The pair were allowed to communicate at any time, for example, to discuss strategies. The pair were positioned at opposite sides of the table, to simulate an attacker observing the user's face, and another one observing the user's touchscreen. After each video, the participants had time to discuss their solution and could state up to three guesses for the password. We concluded with a short semi-structured interview.



	All attacks	Iterative attacks	Side attacks
🟦 Touch-only	68.06%	75.00%	61.11%
🟥 Touch+random	31.25%	43.06%	19.44%
🟩 GazeTouchPIN	10.42%	4.17%	16.67%

Figure 8. Success rate of attacking PINs entered using the three methods in Security Study 2. GazeTouchPIN provided the highest level of security among the tested methods, in particular against iterative attacks.

5.3.4. Security study 3 results

We report on the successful attacks and the results of the semi-structured interviews. *Successful Attacks* A repeated-measures ANOVA with Greenhouse–Geisser correction showed a significant main effect for the password type ($F_{1.87,16.82} = 4.32, p < .05$). Post-hoc analysis with Bonferroni correction revealed a significant difference between GazeTouchPass with 0-switches and GazeTouchPass with 2-switches ($p < .05$). Although the other pairs were not significantly different ($p > .05$), we found a tendency for more successful guesses against GazeTouchPass with no switches, compared to GazeTouchPass with 1-, 2-, and 3-switches (see Figure 10). This result matches previous work (Khamis et al., 2016), which reported that the more switches in a GazeTouchPass password exist, the harder it is to observe. Furthermore, we found that GazeTouchPIN is less secure than many configurations of GazeTouchPass. This is expected since the random layout is no longer as effective when two attackers are observing the user at the same time.

Table 2. Perceived difficulty (1=Very easy, 5=Very difficult) of attacking the three methods in each of the views.

	Touch-only		Touch +random		GazeTouchPIN	
	Phone	Sideview	Phone	Side	Phone+Eyes	Side
Median	2.00	3.00	3.00	4.00	3.00	5.00
StDev	1.00	0.90	0.80	0.65	1.35	0.98

Notes: Participants found it most difficult to attack touch+random and GazeTouchPIN from the side. Their perception of the difficulty of iterative observation attacks was misplaced because there was a $\frac{1}{2^n}$ chance of seeing a matching phone-view and eyes-view .

Qualitative Feedback In the short interviews, the participants indicated their relationship to the other attacker in their team. In six teams, the attackers were friends, in three of them they were acquaintances, and the remaining pair were strangers. We did not find any effect of the relationship between the attackers on successful guesses. Participants reported that they devised strategies with their partners. For example, they would count in their heads to try to estimate the positions of the inputs from the other modality. The attacker who observed the touch input was able to see whether the successful login screen was shown after the last touch input, or if the last touch input was followed by a pause. This gave the attackers hints about the positions of the observed inputs.

6. Discussion

6.1. GazeTouchPass is secure against side observation attacks

GazeTouchPass passwords that use 2- and 3-switches are particularly secure against side observation attacks (only 15% success rate), even when compared to GazeTouchPIN due to the fact that attackers cannot predict whether the user's next input is gaze-based or touch-based in case of GazeTouchPass . When using GazeTouchPIN , however, side observation attacks performed slightly better (17% success rate) as the adversary expects gaze input right after each touch input.

In case of multiple observers, GazeTouchPass is less secure but still better than the baseline and than

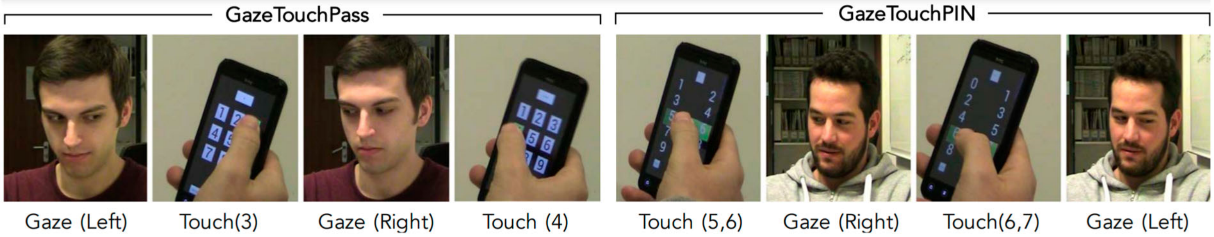


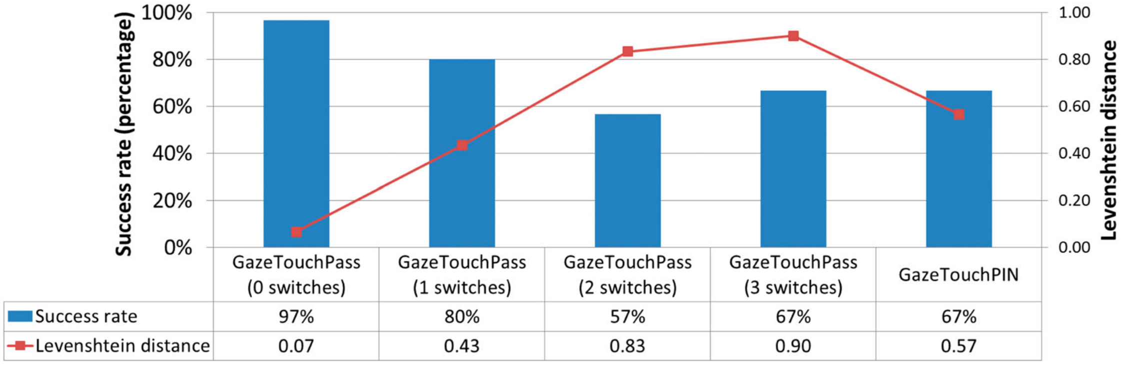
Figure 9. This work proposes and evaluates the use of gaze and touch for user-centred authentication on smartphones. This multimodal approach increases resilience to shoulder-surfing attacks as attackers need to observe the user's eyes and the touchscreen simultaneously to find the password. GazeTouchPass (left) enables passwords with multiple switches between input modalities during authentication. In the example, the user authenticates using: Left-3-Right-4. GazeTouchPIN (right) uses multiple modalities and complicates attacks by using one of two random layouts during PIN entry. In the example the user enters the digit 6 two times in a row.

GazeTouchPIN when using 3-switches. This can be seen in Table 3, which shows a comparison between the success rates in security study 3 compared to security studies 1 and 2. The reason behind the higher success rate against the baseline condition (GazeTouchPass with 0-switches) is that both attackers saw the same video. Attackers were able to discuss their guesses afterwards, and this allowed them to fine-tune the three submitted guesses based on two observations instead of only one. Attackers performed better against the other conditions of GazeTouchPass as well due to the same reason: overall, the team had higher exposure to the password and was able to better identify the pauses between the inputs from different modalities. For example, observing Touch(1), Pause, Touch (2) in the phone view (Figure 2 Camera B) suggests that there is one or more gaze inputs in between those two touch inputs. These pauses in turn help the attackers identify how to order their observations. At the same time, the main reason behind incorrect guesses against GazeTouchPass is the ordering of inputs; in the vast majority of cases, the correct inputs were observed by the attackers, but the guessed order was incorrect (e.g. guessing Touch(1), Gaze(Left), Touch(2), Gaze(Right) instead of Gaze(Left), Touch(1), Touch(2), Gaze (Right)).

6.2. GazeTouchPIN is secure against iterative observation attacks

GazeTouchPIN is superior over GazeTouchPass in protecting against iterative observation attacks (only 4.2% success rate) because of the randomness of the layout. Iterative observation attacks against GazeTouchPass are complicated but still possible (23%–46%), given that the adversary paid attention to all inputs and noted the gaps in-between.

Attacks by multiple observers are effective against GazeTouchPIN due to the parallel observations. GazeTouchPIN was very secure against iterative observation attacks because each time the user enters a digit, the layout could have been different. This made it unfeasible for attackers to identify which layout the user is



	GazeTouchPass (0 switches)	GazeTouchPass (1 switches)	GazeTouchPass (2 switches)	GazeTouchPass (3 switches)	GazeTouchPIN
Success rate	97%	80%	57%	67%	67%
Levenshtein distance	0.07	0.43	0.83	0.90	0.57

Figure 10. The figure shows that, similar to previous work, the Levenshtein distance is larger in case of 2- and 3-switches. This means that GazeTouchPass is more secure when more switches exist in the password. GazeTouchPIN is far less secure against our threat model compared to previously studied ones, since the random layout is no longer effective when two attackers observe simultaneously. Overall, while success rates are much higher in the multiple observers threat model compared to models studied in the past, both schemes still outperform the baseline.

Table 3. Compared to previous evaluations of GazeTouchPass and GazeTouchPIN (Sections 5.1.3 and 5.2.3), the multiple attackers threat model results in more successful attacks against the said schemes.

	GazeTouchPass				
Number of attackers	0-switches (baseline)	1-switch	2-switches	3-switches	GazeTouchPIN
One attacker	63%	46%	37%	23%	4%
Two attackers	97%	80%	57%	67%	67%

responding to when observing their eye movements. This security advantage is no longer present in case of parallel multiple observers attacks; the attacker observing the screen could note down the touch input and the shown layout, while the other one observes the gaze input. Combining the observations in this case would be trivial.

6.3. Usability and security trade-off

GazeTouchPass and GazeTouchPIN are significantly more secure than the baselines. It should be noted that all previous conclusions are based on the assumption that the attacker knows how the observed system works. The threat models we propose are realistic but also ensure optimal attacking conditions. Additionally, participants of the security studies were highly motivated and trained. This is evidenced from their performance against the baselines which was as high as 75% (see Figures 7–10). This is comparable to results from state-of-the-art schemes; attackers of ColorSnakes (Gugenheimer et al., 2015) and XSide (De Luca et al., 2014) achieved 75% and 53% success rate against the respective baselines.

On the downside, both GazeTouchPass and GazeTouchPIN suffer from lower usability compared to the less secure baselines. Mean authentication time using GazeTouchPass is approximately 3.1 seconds, and ≈200 more milliseconds for passwords with 2 switches. While this is slightly slower compared to the baseline and common schemes such as PINs. For example, von Zezschwitz, Dunphy, and De Luca (2013) report 1.5 seconds for PINs and 3.13 seconds for lock patterns. GazeTouchPass is faster than some of the security-optimised state-of-the-art and multimodal authentication systems (see Table 4). In terms of usability, input time is faster using GazeTouchPass compared to GazeTouchPIN. We expect that participants will authenticate faster as they use the systems more frequently due to training effects. We already observed preliminary evidence of this; mean authentication time using GazeTouchPIN decreased from 10.8 to 9.5 seconds as participants used it more often. Since users unlock their phones almost 50 times a day (Harbach, Luca, and Egelman,2016), we recommend the use of GazeTouchPIN in sensitive contexts rather than on regular basis. Overall, and as several participants indicated, multimodal authentication

Table 4. Comparison of GazeTouchPass and GazeTouchPIN with state-of-the-art schemes using gaze-based authentication (De Luca, Denzel, and Hussmann,2009; De Luca, Weiss, and Drewes,2007; Forget, Chiasson, and Biddle,2010; Kumar et al., 2007; Liu et al., 2015; Sluganovic et al., 2016), input-splitting (De Luca et al., 2014) and multiple modalities (Bianchi et al., 2011; Bianchi, Oakley, and Kwon,2011, 2012).

System	Input time	Successful attacks
GazeTouchPass		
3-switches (Side)	3.1 s	15%
3-switches (Iterative)	3.1 s	23%
2-switches (Side)	3.3 s	15%
2-switches (Iterative)	3.3 s	37%
1-switches (Side)	3.0 s	21%
1-switches (Iterative)	3.0 s	46%
0-switches (Side)	3.0 s	25%
0-switches (Iterative)	3.0 s	63%
GazeTouchPIN (Side)	10.82 s	17%
GazeTouchPIN (Iterative)	10.82 s	4%
Authentication schemes that use gaze		
EyePassShapes (De Luca, Denzel, and Hussmann,2009)	12.5 s	42%
EyePIN (De Luca, Weiss, and Drewes,2007)	48.5 s	55%
CGP (Forget, Chiasson, and Biddle,2010)	36.7 s	
EyePassword (Kumar et al., 2007)	9.2 s–12.1 s	
Liu et al. (2015)	9.6 s	
EyeVeri (Song et al., 2016)	5 s–10 s	
Sluganovic et al. (2016)	5 s	
Multimodal authentication schemes		
PhoneLock (Bianchi et al., 2011)	12.2 s–28.2 s	
SpinLock (Bianchi, Oakley, and Kwon,2011)	10.8 s–20.1 s	
TimeLock (Bianchi, Oakley, and Kwon,2012)	10 s	
ColorLock (Bianchi, Oakley, and Kwon,2012)	10 s	
GazeGestureAuth (Abdrabou et al., 2019)	19.34 s–20.63 s	
Authentication Schemes that split input		
CueAuth (gaze) (Khamis, Trotter, et al., 2018)	26.46 s	0.03%
XSide (De Luca et al., 2014)		
front 1-switch start	3.9 s	38%
front 1-switch end	3.7 s	13%
front 2-switches	3.8 s	28%
back 1-switch start	3.8 s	19%
back 1-switch end	4.1 s	16%
back 2-switches	4.0 s	9%

Note: GazeTouchPass shows a balance between security and usability, with lower authentication times and less successful attack rates, while GazeTouchPIN shows superior resistance to iterative attacks while maintaining good usability. This suggests that multimodal schemes are promising for secondary authentication, where users feel observed or want to protect sensitive data.

can be particularly useful as a secondary authentication mechanism that users can choose to opt to when feeling observed (e.g. public setting), or when accessing critical data (e.g. online banking).

6.4. Comparison to state-of-the-art

GazeTouchPass demonstrates a balance between security and usability, with lower authentication times and less successful attack rates compared to related authentication systems, while GazeTouchPIN shows superior resistance to iterative observation attacks while maintaining reasonable usability (Table 4).

We compared our systems against gaze-based authentication schemes that, like GazeTouchPass , transform the password space (De Luca, Denzel, and Hussmann,2009; De Luca, Weiss, and Drewes,2007; Forget, Chiasson, and Biddle,2010; Kumar et al., 2007; Khamis, Oechsner, et al., 2018), as well as with systems that, like GazeTouchPIN , obscure numerical passwords using gaze input (Liu et al., 2015; De Luca, Weiss, and Drewes,2007; Khamis, Trotter, et al., 2018; Abdrabou et al., 2019). We found that GazeTouchPass and GazeTouchPIN are faster and more secure than desktop-based systems (Table 4). GazeTouchPass is faster than the system proposed by Liu et al. (2015). Its security was not formally evaluated, however, it uses a password space of 4^n only, while our systems use 12^n and 10^n respectively. Although the security of PhoneLock (Bianchi et al., 2011), SpinLock (Bianchi, Oakley, and Kwon,2011), TimeLock (Bianchi, Oakley, and Kwon,2012) and ColorLock (Bianchi, Oakley, and Kwon,2012) was not evaluated in a way comparable to our studies, our systems require a shorter authentication time (Table 4).

XSide is based on input-splitting (De Luca et al., 2014), where a double-sided touchscreen is used for password entry. Our systems, on the other hand, can work on off-the-shelf mobile devices without any additional hardware. XSide is faster than GazeTouchPIN , but slower than GazeTouchPass . Similar to our systems, the number of switches in a password entered using XSide influences its security; in most cases GazeTouchPass and GazeTouchPIN are more resistant to observations (Table 4).

A further distinction of our work is that we consider advanced shoulder-surfing tactics, which allowed studying the security of our systems in worst-case scenarios that are nevertheless realistic. For example, the security study of XSide considered side observation attacks in case of split input, while iterative observation attacks could be more successful.

6.5. Splitting the attacker’s attention is key to resisting observation attacks

Although multiple observers perform better than single ones when attacking GazeTouchPass and GazeTouchPIN compared to single observers, their success is significantly worse than when attacking the baseline (see Table 3). This means that while these schemes are not as effective against multiple attackers as they are against single observers, they are still more secure than the baseline.

6.6. Password selection strategies

Multimodal passwords entered using GazeTouchPass were remembered by the vast majority of participants after 5 days. It is also expected that with frequent use, users would find it easier to recall passwords. By examining the practical password space of GazeTouchPass we find that users exploit different features of GazeTouchPass that make it more secure. There was a focus on selecting passwords with multiple switches in modality, and also on ones starting with gaze input; the security studies participants reported these were the most difficult to break. GazeTouchPIN is based on the widely used PINs, hence its users will not have to memorise new passwords.

6.7. Attacking strategies

When performing iterative observation attacks, participants in GazeTouchPass ’s security study employed a gap-filling strategy when combining observations. In addition to noting the gestures and the digits, participants also noted the pauses when observing either the eyes-view or the phone-view . This approach, however, does not always serve its purpose. We logged multiple cases where participants observed all inputs but guessed an incorrect order (e.g. 9-1-right-left instead of 9-right-1-left). Security-aware users can in fact exploit the gaps to confuse observers; a user could intentionally introduce an unneeded gap before providing the next entry. This strategy was far easier to implement in case of multiple attackers. Our participants reported that they split the tasks. The attacker who observed the touch input was usually responsible for determining the last input and allowed the team to gather insights about the number of modality switches.

Attackers of GazeTouchPIN noted inputs in a similar manner. Rather than observing the two selected digits, four participants wrote down one of them and noted whether it was on the right or on the left. They then checked the layouts (Figure 1) to determine which row was selected in which layout. However these strategies were less effective for iterative observation attacks against GazeTouchPIN , where 69 out of 72 attacks failed because of the low probability ($\frac{1}{2^n}$) of seeing a matching phone-view and eyes-view at different authentication attempts. Iterative observation attacks were highly successful however, in the presence of multiple attackers. In case of side observation attacks, attackers reported they had to switch focus back and forth between the eyes and the touchscreen, which was particularly harder as users authenticate faster. Side observation attacks performed worse on GazeTouchPass due to the unpredictability of the switches in input modality.

In all cases, participants of the security studies reported that as users authenticate faster, the harder it is to attack the passwords. This is another positive aspect of multimodal authentication, since results indicate that users are expected to input passwords faster as they use the system more often.

6.8. Other threat models

Although we considered advanced threat models that assume a better-than-naive attacker, there are various other threat models that our system can be compared against.

Similar to iterative observation attacks, insider attacks (Wiese and Roth,2016) combine multiple partial observations. But instead of observing the entire password, the insider model relies on using partially collected observations to reduce the entropy of the currently observed password when performing brute force attacks. In input-splitting schemes, such as XSide (De Luca et al., 2014) and our systems, an insider could focus on inputs observed from one view and guess the other inputs. For example, observing one input in a 4-digit PIN reduces the space from 10^4 to 10^3. While a single observation on any system reduces the password space dramatically, our systems still have the advantage of not leaking the order of observed input from any of the views. For example, by observing a gaze gesture from the eyes-view on GazeTouchPass , the attacker would not know where in the password the gesture is with respect to the other inputs.

Another interesting direction for future work is to investigate combined threat models. For example, an attacker could observe a user's gaze input while authenticating using GazeTouchPIN or GazeTouchPass, and then perform a thermal attack (Abdelrahman et al., 2017) or a smudge attack (Aviv et al., 2010) to infer touch input.

While it was infeasible to address all possible threat models in our studies, we intend to study other models in future work. Implementations of our schemes will lock users out after multiple failures to counter guessing attacks, and a minimum number of switches will be required by GazeTouchPass.

6.9. Limitations and future work

Video-based gaze estimation has its known limitations; varying light conditions, reflections of eye glasses and heavy makeup can affect the quality of eye tracking (Majaranta and Bulling,2014), and some eye tracking algorithms rely on the presence of a full face in the camera's view, which is not always the case in day-to-day smartphone use (Khamis, Baier, et al., 2018). For this reason we opted for simple eye gestures that can be robustly detected by front-facing cameras. However, we acknowledge that the use of better eye tracking equipment (e.g. infrared light sources and sensors) can enable a wider range of eye movements to be detected robustly. A direction for future work is to run our systems on infrared-supported mobile devices. Moreover, as processing power of mobile devices improve, mobile-based gaze tracking approaches that have been used for offline processing (e.g.Huang et al., 2017; Wood and Bulling.,2014) can be employed in real-time.

Although mean authentication times using both systems are comparable to state-of-the-art systems, they are generally longer compared to the more popular and insecure PINs and patterns. The trade-off between usability and security has been discussed in previous works. Therefore we believe that multimodal authentication, being significantly more secure, offers users a tangible benefit in protecting their sensitive data, and would recommend it for secondary authentication which security-aware users can opt for in sensitive contexts. A future long-term study where participants use the systems over a number of weeks will reveal how learning effects will impact input time.

6.10. Guidelines for multimodal authentication on mobile devices

Based on our experimentation with two concepts for multimodal authentication, and based on the results of the five user studies, we developed the following recommendations:

- *Exploit data from new and improved sensors on mobile devices to improve the usability and the security of authentication.* By using gaze as an additional modality for authentication, we improved the security of authentication significantly as shown in Table 4. The usability of our systems is expected to improve using the newly available depth sensors in front-facing cameras. As newer sensors are integrated into smartphones, researchers and practitioners are encouraged to exploit them to improve authentication albeit by processing the data locally on the smartphone (see a discussion of ethical considerations in Section 6.13).

- *Minimize the number of layouts when introducing random elements into the authentication procedure.* While prototyping GazeTouchPIN, we wanted to add a random element to ensure that observing the eyes-view and the phone-view on two different occasions does not leak the PIN. Had we displayed a completely randomised layout, participants would have had to spend more time to find the digits they wish to enter. Instead, randomly showing one of two layouts supports learning effects and hence usability and potentially memorability.
- *Offer multimodal authentication as security add-on for specific tasks or accounts.* Feedback from our participants shows that while users may not appreciate the increased security at the expense of usability for their daily smartphone unlocking, they are willing to use our systems for sensitive tasks that they do not perform as often, such as booting the phone, or accessing online banking.
- *Tailor multimodal authentication to the user's location and environment.* Many of our participants reported they would use our systems in situations where they feel vulnerable. For example, users are more likely to be observed while using public transportation, hence we need to consider the three threat models detailed above. In other locations, such as at home, weaker threat models might be more appropriate.
- *Increasing the number of switches from one modality to another improves security.* Switching from gaze to touch input, or from touch to gaze input in GazeTouchPass improves observation resistance significantly. This is because each switch requires the attack to switch attention, thereby complicating the attack. In fact, similar results were observed when introducing elements to authentication that require attackers to switch attention (e.g. see RubikAuth Mathis et al., 2021 and XSide De Luca et al., 2014).

6.11. Contributions in practice

The guidelines we presented are useful to practitioners who wish to employ multimodal authentication on their systems. We implemented and evaluated our systems for mobile devices. However, some recommendations apply to other platforms as well e.g. ATMs, public displays or mixed reality headsets.

6.12. Open challenges for practical application

We evaluated the usability of our systems in the lab. A long-term user study in the wild may reveal interesting insights into daily usage of multimodal authentication. Some challenges of gaze interaction on mobile devices may be present when authenticating using gaze in daily scenarios. These challenges include: the visibility of the user's eyes, accuracy in shaky environments, lighting conditions, and the privacy implications of collecting gaze data (Khamis, Alt, and Bulling,2018).

As discussed in Section 6.9, the user's eyes may not always be fully visible to the front-facing camera due to clothing, reflections, or the user's holding posture (Khamis, Baier, et al., 2018; Huang, Veeraraghavan, and Sabharwal,2017). This is amplified by the fact that many gaze estimation algorithms rely on first detecting the user's full face. More work is needed to maximise gaze estimation accuracy even if only part of the face (or one eye) is visible.

Users are often on the move while interacting or unlocking their smartphones. This means that a lot of the gaze data will be inaccurate due to shaky environments. More research is needed to study how well gaze-based authentication works in these scenarios, and develop methods to guide users into a setting (e.g. an ideal holding posture) to allow accurate gaze estimation. This is also needed to overcome the problem of lighting conditions; sunlight may make depth data less reliable, whereas dark environments complicate gaze estimation in RGB videos.

Finally, another challenge is that even though the aim of this work is to improve security, the collection of gaze data has privacy and ethical implications. We discuss those in the next section.

6.13. Ethical considerations

In our work, we estimated the gaze gestures on the mobile devices directly. Another approach is to outsource the gaze estimation process to a remote server (Khamis, Alt, and Bulling,2018). Practitioners may be tempted to do this as gaze estimation is a CPU heavy task that may cause the smartphone to heat up and drain its battery. However, doing so may have significant implications on privacy. Gaze data can reveal very sensitive information about the user (Katsini et al., 2020). For example, the users personality traits, mental state, emotions, and visual interests can be determined from their eye movements (Katsini et al., 2020). Thus, we strongly recommend that real-world implementations of multimodal authentication process gaze data locally on the smartphone without sending the gaze data elsewhere to ensure privacy and avoid unethical exploitation of the user's sensitive data. Another ethical issue is that as smartphones' capability of accurate gaze estimation improves, the users' smartphones will start to pose a privacy risk on bystanders. This is because gaze behaviour may be captured by the smartphone cameras, which are increasingly improving in terms of lens angle and by incorporating depth sensors. Our studies met the ethics regulations of Ludwig Maximilian University of Munich, where the studies took place.

7. Conclusion

We proposed to combine gaze and touch for multimodal user authentication on mobile devices by exploiting the front-facing cameras readily available in these devices for estimating users' eye movement. We presented two novel authentication schemes that enhance security by requiring attackers to observe both input modalities. While GazeTouchPass (multimodal passwords) is more resilient to side observation attacks because of having to quickly switch focus between phone and eyes, GazeTouchPIN (multimodal selection of PINs) is more superior against iterative observation attacks due to the random choice of layout. We demonstrated that both schemes are significantly more secure than current singlemodal schemes, including attacks that involve multiple observers. These findings underline the potential of using gaze input to increase security against basic and advanced shoulder-surfing attacks. We expect these advantages to multiply with further advances in remote gaze estimation on mobile devices.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work has been funded, in part, by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC [grant number EP/S035362/1], an EPSRC New Investigator award [EP/V008870/1], and the Royal Society of Edinburgh [award number #65040].

ORCID

Mohamed Khamis http://orcid.org/0000-0001-7051-5200
Karola Marky http://orcid.org/0000-0001-7129-9642
Andreas Bulling http://orcid.org/0000-0001-6317-7303
Florian Alt http://orcid.org/0000-0001-8354-2195

References

Abdelrahman, Yomna, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. "Stay Cool! Understanding Thermal Attacks on Mobile-Based User Authentication." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 12. New York, NY: ACM.

Abdrabou, Yasmeen, Yomna Abdelrahman, Ahmed Ayman, Amr Elmougy, and Mohamed Khamis. 2020. "Are Thermal Attacks Ubiquitous? When Non-Expert Attackers Use Off the Shelf Thermal Cameras." In *Proceedings of the International Conference on Advanced Visual Interfaces*, Arcticle 47, 5. New York, NY: Association for Computing Machinery. doi:10.1145/3399715.3399819.

Abdrabou, Yasmeen, Reem Hatem, Yomna Abdelrahman, Amr Elmougy, and Mohamed Khamis. 2021. "Passphrases Beat Thermal Attacks: Evaluating Text Input Characteristics Against Thermal Attacks on Laptops and Smartphones." In *Human-Computer Interaction – INTERACT 2021*, edited by Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen, 712–721. Cham: Springer.

Abdrabou, Yasmeen, Mohamed Khamis, Rana Mohamed Eisa, Sherif Ismail, and Amrl Elmougy. 2019. "Just Gaze and Wave: Exploring the Use of Gaze and Gestures for Shoulder-Surfing Resilient Authentication." In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*, Article 29, 10. New York, NY: ACM. doi:10.1145/3314111.3319837.

Almoctar, Hassoumi, Pourang Irani, Vsevolod Peysakhovich, and Christophe Hurter. 2018. "Path Word: A Multimodal Password Entry Method for Ad-Hoc Authentication Based on Digits' Shape and Smooth Pursuit Eye Movements." In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, 268–277. New York, NY: Association for Computing Machinery. doi:10.1145/3242969.3243008.

Alt, Florian, Mateusz Mikusz, Stefan Schneegass, and Andreas Bulling. 2016. "Long-Term Memorability of Cued-Recall Graphical Passwords with Saliency Masks." In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. New York, NY: ACM. doi:10.1145/3012709.3012727.

Aviv, Adam J., Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. 2010. "Smudge Attacks on Smartphone Touch Screens." In *Proceedings of the 4th USENIX Conference on Offensive Technologies (WOOT'10)*, 1–7. Berkeley, CA: USENIX Association. http://dl.acm.org/citation.cfm?id=1925004.1925009.

Best, Darrell S., and Andrew T. Duchowski. 2016. "A Rotary Dial for Gaze-based PIN Entry." In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*, 69–76. New York, NY: ACM. doi:10.1145/2857491.2857527.

Bianchi, Andrea, Ian Oakley, Vassilis Kostakos, and Dong Soo Kwon. 2011. "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 (TEI '11)*, 197–200. New York, NY: ACM. doi:10.1145/1935701.1935740.

Bianchi, Andrea, Ian Oakley, and DongSoo Kwon. 2011. “Spinlock: A Single-Cue Haptic and Audio PIN Input Technique for Authentication.” In *Haptic and Audio Interaction Design*, edited by Eric W. Cooper, Victor V. Kryssanov, Hitoshi Ogawa, and Stephen Brewster. Lecture Notes in Computer Science, Vol. 6851, 81–90. Berlin: Springer. doi:10.1007/978-3-642-22950-3_9.

Bianchi, Andrea, Ian Oakley, and Dong Soo Kwon. 2012. “Counting Clicks and Beeps: Exploring Numerosity Based Haptic and Audio {PIN} Entry.” *Interacting with Computers* 24 (5): 409–422. doi:10.1016/j.intcom.2012.06.005 .

Bulling, Andreas, Florian Alt, and Albrecht Schmidt. 2012. “Increasing the Security of Gaze-based Cued-recall Graphical Passwords Using Saliency Masks.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*, 3011–3020. New York, NY: ACM. doi:10.1145/2207676.2208712.

Cymek, Dietlind Helene, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. “Entering PIN Codes by Smooth Pursuit Eye Movements.” *Journal of Eye Movement Research* 7 (4): 1–11.

De Luca, Alexander, Martin Denzel, and Heinrich Hussmann. 2009. “Look into My Eyes!: Can You Guess My Password?” In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS ’09)*, Article 7, 12. New York, NY: ACM. doi:10.1145/1572532.1572542.

De Luca, Alexander, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. 2014. “Now You See Me, Now You Don’t: Protecting Smartphone Authentication from Shoulder Surfers.” In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI ’14)*, 2937–2946. New York, NY: ACM. doi:10.1145/2556288.2557097.

De Luca, Alexander, Emanuel von Zezschwitz, and Heinrich Hußmann. 2009. “Vibrapass: Secure Authentication Based on Shared Lies.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’09)*, 913–916. New York, NY: Association for Computing Machinery. doi:10.1145/1518701.1518840.

De Luca, Alexander, Emanuel von Zezschwitz, Ngo Dieu Huong Nguyen, Max-Emanuel Maurer, Elisa Rubegni, Marcello Paolo Scipioni, and Marc Langheinrich. 2013. “Back-of-Device Authentication on Smartphones.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13)*, 2389–2398. New York, NY: ACM. doi:10.1145/2470654.2481330.

De Luca, Alexander, Roman Weiss, and Heiko Drewes. 2007. “Evaluation of Eye-gaze Interaction Methods for Security Enhanced PIN-entry.” In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI ’07)*, 199–202. New York, NY: ACM. doi:10.1145/1324892.1324932.

Eiband, Malin, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. “Understanding Shoulder Surfing in the Wild: Stories from Users and Observers.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*, 11. New York, NY: ACM.

Findling, Rainhard Dieter, Tahmid Quddus, and Stephan Sigg. 2019. “Hide My Gaze with EOG! Towards Closed-Eye Gaze Gesture Passwords That Resist Observation-Attacks with Electrooculography in Smart Glasses.” In *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia (MoMM2019)*, 107–116. New York, NY: Association for Computing Machinery. doi:10.1145/3365921.3365922.

Forget, Alain, Sonia Chiasson, and Robert Biddle. 2010. “Shoulder-Surfing Resistance with Eye-Gaze Entry in Cued-Recall Graphical Passwords.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’10)*, 1107–1110. New York, NY: ACM. doi:10.1145/1753326.1753491.

Google. 2016. “Unlock With Your Fingerprint.” Webpage. Accessed January 9, 2017. https://support.google.com/nexus/answer/6285273.

Gugenheimer, Jan, Alexander De Luca, Hayato Hess, Stefan Karg, Dennis Wolf, and Enrico Rukzio. 2015. “ColorSnakes: Using Colored Decoys to Secure Authentication in Sensitive Contexts.” In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’15)*, 274–283. New York, NY: ACM. doi:10.1145/2785830.2785834.

Harbach, Marian, Alexander De Luca, and Serge Egelman. 2016. “The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*, 4806–4817. New York, NY: ACM. doi:10.1145/2858036.2858267.

Hohlfeld, Oliver, André Pomp, Jó Ágila Bitsch Link, and Dennis Guse. 2015. “On the Applicability of Computer Vision Based Gaze Tracking in Mobile Scenarios.” In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’15)*, 427–434. New York, NY: ACM. doi:10.1145/2785830.2785869.

Huang, Michael Xuelin, Jiajia Li, Grace Ngai, and Hong Va Leong. 2017. “ScreenGlint: Practical, In-Situ Gaze Estimation on Smartphones.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*, 2546–2557. New York, NY: Association for Computing Machinery. doi:10.1145/3025453.3025794.

Huang, Qiong, Ashok Veeraraghavan, and Ashutosh Sabharwal. August 1, 2017. “TabletGaze: Dataset and Analysis for Unconstrained Appearance-Based Gaze Estimation in Mobile Tablets.” *Machine Vision and Applications* 28 (5): 445–461. doi:10.1007/s00138-017-0852-4 .

Ishimaru, Shoya, Kai Kunze, Yuzuko Utsumi, Masakazu Iwamura, and Koichi Kise. 2013. “Where Are You Looking at? -- Feature-Based Eye Tracking on Unmodified Tablets.” In *Proceedings of the 2nd IAPR Asian Conference on Pattern Recognition*, 738–739. Piscataway, NJ: IEEE. doi:10.1109/ACPR.2013.190.

Karlson, Amy K., Benjamin B. Bederson, and John SanGiovanni. 2005. “AppLens and launchTile: Two Designs for One-Handed Thumb Use on Small Devices.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’05)*, 201–210. New York, NY: ACM. doi:10.1145/1054972.1055001.

Katsini, Christina, Yasmeen Abdrabou, George Raptis, Mohamed Khamis, and Florian Alt. 2020. "The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions." In *Proceedings of the 38th Annual ACM Conference on Human Factors in Computing Systems (CHI '20)*, 21. New York, NY: ACM. doi:10.1145/3313831.3376840.

Katsini, Christina, Christos Fidas, Marios Belk, George Samaras, and Nikolaos Avouris. 2019. "A Human-Cognitive Perspective of Users' Password Choices in Recognition-Based Graphical Authentication." *International Journal of Human–Computer Interaction* 35 (19): 1800–1812. doi:10.1080/10447318.2019.1574057 .

Khamis, Mohamed, Florian Alt, and Andreas Bulling. 2018. "The Past, Present, and Future of Gaze-Enabled Handheld Mobile Devices: Survey and Lessons Learned." In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*, Article 38, 17. New York, NY: Association for Computing Machinery. doi:10.1145/3229434.3229452.

Khamis, Mohamed, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. "GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices." In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*, 2156–2164. New York, NY: ACM. doi:10.1145/2851581.2892314.

Khamis, Mohamed, Anita Baier, Niels Henze, Florian Alt, and Andreas Bulling. 2018. "Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones Used in the Wild." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 1–12. New York, NY: Association for Computing Machinery. doi:10.1145/3173574.3173854.

Khamis, Mohamed, Linda Bandelow, Stina Schick, Dario Casadevall, Andreas Bulling, and Florian Alt. 2017. "They are all After You: Investigating the Viability of a Threat Model That Involves Multiple Shoulder Surfers." In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*, 5. New York, NY: ACM. doi:10.1145/3152832.3152851.

Khamis, Mohamed, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, and Florian Alt. 2017. "GazeTouchPIN: Protecting Sensitive Data on Mobile Devices using Secure Multimodal Authentication." In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*, 5. New York, NY: ACM. doi:10.1145/3136755.3136809.

Khamis, Mohamed, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. "VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements." In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI '18)*, 7. New York, NY: ACM. doi:10.1145/3206505.3206522.

Khamis, Mohamed, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018, December. "CueAuth: Comparing Touch, Mid-Air Gestures, and Gaze for Cue-Based Authentication on Situated Displays." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (4): 21. Article 174 . doi:10.1145/3287052.

Kinnunen, Tomi, Filip Sedlak, and Roman Bednarik. 2010. "Towards Task-Independent Person Authentication Using Eye Movement Signals." In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*, 187–190. New York, NY: ACM. doi:10.1145/1743666.1743712.

Krafka, Kyle, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. "Eye Tracking for Everyone." In *Proceddings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2176–2184. Piscataway, NJ: IEEE. doi:10.1109/CVPR.2016.239.

Kumar, Manu, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. "Reducing Shoulder-Surfing by Using Gaze-Based Password Entry." In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*, 13–19. New York, NY: ACM. doi:10.1145/1280680.1280683.

Liu, Dachuan, Bo Dong, Xing Gao, and Haining Wang. 2015. "Exploiting Eye Tracking for Smartphone Authentication." In *Proceedings of the 13th International Conference on Applied Cryptography and Network Security (ACNS '15)*, 20. Springer. doi:10.1007/978-3-319-28166-7_22.

Majaranta, Päivi, and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human–Computer Interaction*, 39–65. London: Springer. doi:10.1007/978-1-4471-6392-3_3.

Mathis, Florian, John Williamson, Kami Vaniea, and Mohamed Khamis. 2021. "Fast and Secure Authentication in Virtual Reality Using Coordinated 3D Manipulation and Pointing." *ACM Transactions on Computer-Human Interaction (ToCHI)* 1 (28): 44. Article 6. doi:10.1145/3428121.

Rigas, Ioannis, Evgeniy Abdulin, and Oleg V. Komogortsev. 2016. "Towards a Multi-Source Fusion Approach for Eye Movement-Driven Recognition." *Information Fusion* 32: 13–25. doi:10.1016/j.inffus.2015.08.003. SI: Information Fusion in Biometrics.

Sakai, Daiki, Michiya Yamamoto, Takashi Nagamatsu, and Satoshi Fukumori. 2016. "Enter Your PIN Code Securely!: Utilization of Personal Difference of Angle Kappa." In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*, 317–318. New York, NY: ACM. doi:10.1145/2857491.2884059.

Schneegass, Stefan, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. "SmudgeSafe: Geometric Image Transformations for Smudge-resistant User Authentication." In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*, 775–786. New York, NY: ACM. doi:10.1145/2632048.2636090.

Sluganovic, Ivo, Marc Roeschlin, Kasper B. Rasmussen, and Ivan Martinovic. 2016. "Using Reflexive Eye Movements for Fast Challenge-Response Authentication." In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*, 1056–1067. New York, NY: ACM. doi:10.1145/2976749.2978311.

Song, Chen, Aosen Wang, Kui Ren, and Wenyao Xu. 2016. "EyeVeri: A Secure and Usable Approach for Smartphone User Authentication." In *Proceddings of the IEEE International Conference on Computer Communication (INFOCOM'16)*, 1–9. San Francisco, CA: IEEE.

Sridharan, Srinivas, Brendan John, Darrel Pollard, and Reynold Bailey. 2016. "Gaze Guidance for Improved Password Recollection." In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*, 237–240. New York, NY: ACM. doi:10.1145/2857491.2857537.

Stokkenes, Martin, Raghavendra Ramachandra, and Christoph Busch. 2016. "Biometric Authentication Protocols on Smartphones: An Overview." In *Proceedings of the 9th International Conference on Security of Information and Networks (SIN '16)*, 136–140. New York, NY: ACM. doi:10.1145/2947626.2951962.

Tiefenau, Christian, Maximilian Häring, Mohamed Khamis, and Emanuel von Zezschwitz. 2019. "'Please Enter Your PIN' – On the Risk of Bypass Attacks on Biometric Authentication on Mobile Devices." arXiv:1911.07692 [cs.HC].

Vaitukaitis, Vytautas, and Andreas Bulling. 2012. "Eye Gesture Recognition on Portable Devices." In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*, 711–714. New York, NY: ACM. doi:10.1145/2370216.2370370.

Viola, Paul, and Michael J. Jones. 2004. "Robust Real-Time Face Detection." *International Journal of Computer Vision* 57 (2): 137–154. doi:10.1023/B:VISI.0000013087.49260.fb.

von Zezschwitz, Emanuel, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. 2015. "SwiPIN: Fast and Secure PIN-Entry on Smartphones." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 1403–1406. New York, NY: ACM. doi:10.1145/2702123.2702212.

von Zezschwitz, Emanuel, Paul Dunphy, and Alexander De Luca. 2013. "Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-Based Authentication on Mobile Devices." In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)*, 261–270. New York, NY: ACM. doi:10.1145/2493190.2493231.

von Zezschwitz, Emanuel, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. 2013. "Making Graphic-based Authentication Secure Against Smudge Attacks." In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*, 277–286. New York, NY: ACM. doi:10.1145/2449396.2449432.

Wiese, Oliver, and Volker Roth. 2016. "See You Next Time: A Model for Modern Shoulder Surfers." In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*, 453–464. New York, NY: Association for Computing Machinery. doi:10.1145/2935334.2935388.

Wood, Erroll, and Andreas Bulling. 2014. "EyeTab: Model-based Gaze Estimation on Unmodified Tablet Computers." In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*, 207–210. New York, NY: ACM. doi:10.1145/2578153.2578185.

Zhang, Yanxia, Andreas Bulling, and Hans Gellersen. 2014. "Pupil-Canthi-Ratio: A Calibration-Free Method for Tracking Horizontal Gaze Direction." In *Proc. of the 2014 International Working Conference on Advanced Visual Interfaces (AVI 14) (2014-05-27)*, 129–132. New York, NY: ACM. doi:10.1145/2598153.2598186.

Zhang, Yulong, Zhaonfeng Chen, Hui Xue, and Tao Wei. 2015. "Fingerprints On Mobile Devices: Abusing and leaking." In *Black Hat Conference*.

Zhang, Xiaoyi, Harish Kulkarni, and Meredith Ringel Morris. 2017. "Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 2878–2889. New York, NY: Association for Computing Machinery. doi:10.1145/3025453.3025790.

Appendix. Participant demographics

Table A1. Details of the demographics of our participants.

Usability Study 1			Usability Study 2			Security Study 1			Security Study 2			Security Study 3		
P1	female	22	P1	male	23	P1	male	33	P1	male	18	P1	female	32
P2	female	22	P2	male	26	P2	female	22	P2	male	26	P2	male	36
P3	male	27	P3	female	27	P3	male	25	P3	male	21	P3	female	23
P4	male	35	P4	male	27	P4	female	21	P4	male	25	P4	male	25
P5	female	27	P5	male	31	P5	male	23	P5	male	28	P5	female	21
P6	female	24	P6	male	24	P6	male	25	P6	male	22	P6	female	24
P7	female	30	P7	female	29	P7	male	29	P7	male	19	P7	male	24
P8	female	23	P8	male	21	P8	female	25	P8	male	23	P8	male	28
P9	male	27	P9	male	21	P9	female	21	P9	female	23	P9	male	19
P10	female	21	P10	male	22	P10	male	22	P10	female	21	P10	female	19
P11	female	23	P11	male	19	P11	female	23	P11	female	27	P11	male	26
P12	male	25	P12	male	25	P12	female	23	P12	male	28	P12	female	22
P13	female	22				P13	male	22	P13	female	32	P13	male	27
									P14	male	36	P14	female	22
									P15	female	18	P15	female	39
									P16	male	23	P16	male	24
									P17	male	26	P17	male	22
									P18	male	23	P18	female	23
												P19	male	25
												P20	male	24