Communicating Device Confidence Level and Upcoming Re-Authentications in Continuous Authentication Systems on Mobile Devices

Lukas Mecke[1,2†], Sarah Delgado Rodriguez[2‡], Daniel Buschek[2†], Sarah Prange[1,3,2†], Florian Alt[3]
[1]*University of Applied Sciences Munich, Munich, Germany, {firstname.lastname}@hm.edu*
[2]*LMU Munich, Munich, Germany,† {firstname.lastname}@ifi.lmu.de, ‡S.Delgado@campus.lmu.de*
[3]*Bundeswehr University Munich, Munich, Germany, {firstname.lastname}@unibw.de*

Abstract

Continuous implicit authentication mechanisms verify users over time. In case the device's confidence level (DCL) is too low, the user is prompted with a re-authentication request, which has been shown to annoy many users due to its unpredictable nature. We address this with a novel approach to enable users to anticipate the need for re-authentication with two indicators: (1) a *long term indicator* shows the current DCL and its development over time, and (2) a *short term indicator* announces that re-authentication is imminent. In both cases voluntary re-authentication allows the DCL to be raised and a device lock to be avoided. We tested the indicators in a four week field study (N=32). Our results show that both indicators were preferred over giving no indication and that importance and sensitivity of the interrupted task have a strong impact on user annoyance. Voluntary re-authentications were perceived as positive.

1 Introduction

Smart phones enable access to sensitive information, both on the device itself and in the cloud, that need to be protected. At the same time, traditional smart phone authentication is based on explicit authentication mechanisms, such as PINs, lock patterns, TouchID, and FaceUnlock. The use of such explicit mechanisms creates a considerable authentication overhead. Harbach et al. showed that smartphone users authenticate on average 47.8 times per day [16], spending 2.9% of their time on authentication.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

Researchers have proposed several methods to reduce authentication overhead, including time- or app-based approaches [7,18] as well as implicit authentication mechanisms that authenticate users based on their context [17,23] or their behaviour [6,11,12,24,27,28].

One caveat of such implicit authentication systems is that they can trigger explicit re-authentication; that is: asking users to confirm their identity via a second factor, in case the mechanism is unable to confirm the current user's identity [13,19,21]. Such re-authentication events are likely to interrupt other tasks and, hence, annoy users [20].

Reasons for this annoyance include the unpredictability of interruptions and the sensation of not being correctly informed about the current state of the implicit authentication system [2,9,20]. Moreover, users wish to influence the timing of the interruption in some way [2,22].

To address this, we propose (1) a long term indicator (*LT*), informing users about the current device confidence level (*DCL*) and thus enabling upcoming re-authentication to be anticipated, and (2) a short term indicator (*ST*), enabling users to finish their task. To avoid system-side locking of the device we (3) provide *voluntary* re-authentication (cf. Figure 1).

We investigated these indicators in a field study (N=32) where participants used them in everyday life. We found that people preferred our indicators to a system that interrupts them in an unpredictable way. Their perception strongly depended on the importance of the interrupted task. Voluntary re-authentication was perceived less annoying. Our research is complemented by deriving implications for future implicit authentication systems.

We contribute (1) novel designs to announce upcoming re-authentications and allow for voluntary re-authentication; (2) findings from a 4-week field study, testing the two indicators and their combinations; and (3) a set of implications for future implicit authentication mechanisms based on our findings.

Figure 1: We propose to use indicators to communicate both the current device confidence level (*DCL*) and the need for re-authentication for continuous implicit authentication systems on mobile devices: (1) a *long term* indicator illustrates the current *DCL* and its development over time via a task bar icon, and (2) a *short term* indicator announces an upcoming re-authentication via darkening the screen. Our system also allows for (3) *voluntary* re-authentication to avoid system-side locking of the device.

2 Underlying Use Cases

Implicit authentication has two major use cases: a) as an effortless, independent main authentication mechanism [19]; or b) as a second line of defence against unauthorised access to the private smartphone [21]. The first use case is particularly suitable for smartphone users that currently do not use any kind of authentication on their devices due to the required effort of explicit mechanisms. Hence, users would need to authenticate less frequently than with traditional explicit authentication approaches [16, 19]. The second use case provides an additional security barrier for devices which were already unlocked using an explicit mechanism [21].

In both cases, the reaction of the system to an unsuccessful authentication determines the provided security. An imminently triggered re-authentication prompt, as suggested by Khan et al. [19], promises to be one of the most secure approaches. But such interruptions could also be triggered by false rejects during an authorized usage and can therefore cause usability issues [20]. Some commercial products (e.g., Smart Lock[1]) instead keep the device unlocked and require re-authentication only after the session has ended. While this avoids interruption it also imposes a security risk, in case an attacker gets hold of the device within this time frame.

In this work we address systems that use interruptions to immediately lock the device as proposed by related work to minimise security risks. As previously shown, this can induce annoyance among users, which we aim to mitigate with appropriate indications to prepare users for upcoming re-authentications. Next, we discuss related work in this direction.

3 Related Work

3.1 Implicit Authentication

Many current authentication mechanisms rely on explicit authentication (i.e., recalling a secret or presenting a token or biometric feature [25]). The term *implicit authentication*[2], in contrast, describes the process by which a user is authenticated without requiring explicit interaction. In implicit authentication systems, the initial explicit authentication step to gain access to the device is replaced or complemented by a continuous evaluation of the users' identity that is reflected in a *device confidence level (DCL)*. Similar to a fallback in explicit authentication systems, an explicit so called *re-authentication* is required in case the device can not verify the user's identity.

Methods suggested for implicit authentication rely on the user's context [5, 17, 23, 26] and behavioural features. Examples include mechanisms that authenticate users based on gait recognition [12], continuous eye-tracking [24], or the users' tap or app-execution behaviour [6, 11, 27, 28].

There are several works pointing out the positive effects of implicit authentication. Hayashi et al. [17] found that implicit authentication could reduce explicit authentication by 68%. Riva et al. [26] report a decrease of 42%.

Several studies report on implicit authentication being perceived convenient and easier to use than traditional methods [8, 15, 20]. Finally, in a study by Crawford and Renaud [9] 90% of the participants indicated they would consider using implicit authentication and 73% felt it was more secure than authenticating explicitly.

3.2 Research on Re-Authentication

While implicit authentication is generally perceived positive and can indeed reduce authentication overhead, previous work found that the need for re-authentications can strongly disrupt those positive effects. Khan et al. [20] found that re-authentications, due to *false rejects (FR)* (i.e., cases in which the system rejected the legitimate user), were perceived annoying by 35% of their participants. This was due to both the unpredictable nature of the interruption and the need to switch the context for re-authentication. Another finding, also supported by the study of Crawford and Renaud [9], was that security barriers – like re-authentication – helped users to build a mental model of the system's security and thus led to a stronger perception of security.

[1] Smart Lock: https://support.google.com/android/answer/9075927?hl=en, last accessed June 25, 2019

[2] also called transparent or continuous authentication (e.g., [10]).

3.3 Interruptions

Work by Bailey et al. [4] found that interrupting users is perceived as rude and decreases task performance. They also found timing of an interruption to be highly important, as interrupted tasks were perceived as more difficult. Thus, they suggest using *attention manager* systems to detect phases of low memory load and schedule interruptions during these.

Adamczyk and Bailey [1] further investigated the impact of triggering interruptions at opportune moments. They were able to show that better timed interruptions are perceived as less annoying, less frustrating and more respectful. They also require less mental effort. Fischer et al. [14] aimed at identifying such opportune moments for interruptions with smartphones with the goal of identifying the best timing for delivering notifications. Although their participants did not clearly prefer the suggested interruptions after finishing a task compared to random interruptions, they found people attending faster to notifications in the task-dependent condition.

McFarlane [22] studied interruptions in general and found that making interruptions more predictable made them less annoying and had a positive effect on user performance in the interrupted task. He also found that letting users determine the moment of interruption made interruptions less annoying. Agarwal et al. [2] found similar results in their study. They tested different mechanisms to delay the re-authentication interrupt, using gradual dimming of the screen and transparent overlays to reduce context switch overhead and unpredictability of the interrupt. They found indications that participants were less annoyed when they could predict the interruption. Participants liked the introduced *grace period* (i.e., the delay of the re-authentication) and performance was increased as users tried to finish their tasks before the device was locked.

3.4 Implications of Related Work

From the insights in prior work we derive three opportunities for handling re-authentication interrupts in continuous authentication systems:

1. *Show current state*: Crawford and Renauds [9] found that users disliked the idea of a totally invisible authentication mechanism. Khan et al. [20] suggested indicating the current system status to address similar concerns voiced by participants of their study. This suggests that users' general desire for system feedback is particularly true for authentication as well.

2. *Announce interrupts*: Agarwal et al. [2] and McFarlanes et al. [22] found that predictable interruptions make users feel less annoyed.

3. *Delay interrupts*: Instantly locking the device when re-authentication is required can heavily disrupt the interaction flow [4]. Prior work showed that users liked having a *grace period* to finish their tasks in these situations [2].

4 Concept Development

In this section we report on the development process for our re-authentication concepts: We introduce design considerations revolving around *presentation strategy* and *integration with the smartphone*. These considerations provide the framing for a subsequent focus group in which participants brainstormed about specific designs. In the next section we describe our final concept for indicating upcoming re-authentications based on related work, our design considerations and our findings from the focus group.

4.1 Design Considerations

4.1.1 Presentation Strategy

From related work we derive two approaches for presenting a re-authentication indicator: *long-term* and *short-term*. We consider and investigate both.

Long Term Indicator To show the current state of the system, we consider a permanent indicator displaying the device confidence level (*DCL*) to show that the system is active. This also serves as a means to anticipate upcoming re-authentication.

Short Term Indicator To inform users about the imminent need for a re-authentication, we propose a short term indicator, granting a grace period.

4.1.2 Integration with the smartphone

The re-authentication indicator can be integrated with the smartphone in different ways: by means of static elements with the main purpose of permanently showing the current system status; by using dynamic elements, announcing an upcoming re-authentication request; or a combination of both approaches (hybrids).

Static Elements A well-suited static element on mobile devices is the task bar, as it is (with few exceptions) always shown. Possible elements are icons, percentages, progress bars or changes to the bar itself (e.g., changing colour) to indicate the current *DCL*.

Dynamic Elements On-screen dynamic elements include distortions of the screen content (e.g., darkening, desaturation, pixelation, etc. [2,3]) or a notification. Off-screen elements include vibration, sound, the use of the flashlight, or the notification light.

Hybrids An element that can be used both statically and dynamically is a floating action button, overlaying screen content. Such buttons can show both *DCL* and upcoming re-authentication requests, either colour coded or in the form of e.g., a counter. In particular, a floating action button could also remain invisible and only (gradually) appear to announce a re-authentication.

4.1.3 Freedom of Authentication

To address annoyance due to having to wait for the grace period to finish [2], we propose allowing explicit re-authentication at any time and in particular during the grace period.

4.2 Focus Group

The focus group served two purposes: (1) To collect novel design ideas for re-authentication concepts, focus group participants engaged in an open brainstorming session. (2) To understand users' preferences regarding the design opportunities, participants discussed several designs, covering different aspects of our considerations. We recruited five HCI students from our university (4 female, 1 male) for their expertise in interface design.

4.2.1 Procedure

We first introduced participants to the concept of continuous implicit authentication and explained the terms 'device confidence level' (*DCL*) and 're-authentication'. Afterwards, we asked them to sketch ideas of how the current *DCL* and the need for re-authentication could be communicated to users. We provided print-outs of smartphone home-screens. Furthermore, we nudged them to think beyond visual cues. Following the sketching phase we asked them to present their ideas and discussed them. We then presented a set of our own indicator designs and asked participants to discuss those. Finally we asked participants to rank all designs (their own and our presented ones) and comment on why they chose a ranking.

4.2.2 Focus Group Results

Results covered integration with the smart phone, visual design, modalities, and re-authentication mechanism.

Participants favoured approaches that subtly *integrate the indicator with the smartphone*. In particular, they felt that the indicator would optimally be placed in the task bar. Floating action buttons were perceived as too intrusive. Notifications received mixed opinions: While some participants argued that they were intrusive, others described them as the natural way the device would communicate announcements.

Regarding the *visual design*, participants suggested indicators gradually changing appearance (such as colour) to make users aware of diminishing *DCL*. Abrupt colour changes were considered too intrusive. A positively perceived idea was dimming the screen (similar to the method used in [2]).

Regarding *modality*, participants mentioned notifications and vibration to announce upcoming re-authentication.

As *re-authentication mechanism*, most participants mentioned biometric methods (fingerprint or face recognition) to make the process as smooth as possible. This is in line with feedback from participants in the study by Khan et al. [20].

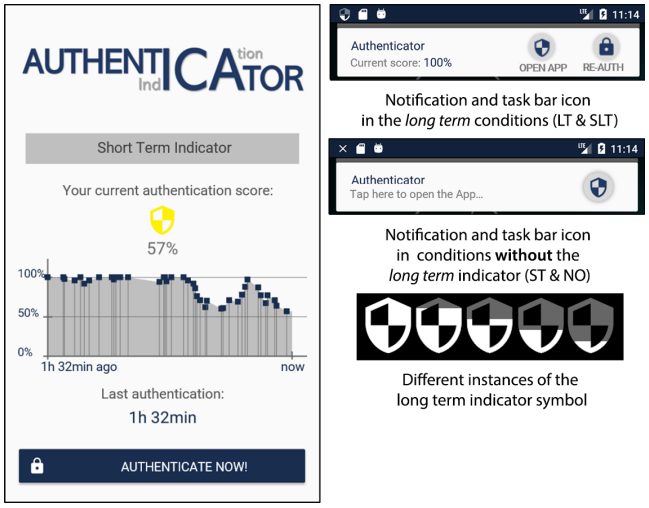

Figure 2: Different elements of the Authenticator app. Left: the main application with the device confidence level (*DCL*) visualised as a graph. Right: The notification and icon shown in the *long term* conditions (top), in the conditions without a *long term* indicator (middle) and the instances of the indicator symbol showing the current *DCL* in the task bar.

5 Authenticator

Based on the recommendations and suggestions both from related work and the focus group we built an android app, called *Authenticator*. The app simulates an implicit authentication system. It provides two different types of indicators that can be combined but also work independently.

5.1 Indicator Designs

Our prototype supports two indicators, namely a *short term* and a *long term* indicator.

5.1.1 Long Term Indicator (LT)

To realise the long term indicator, our application places a permanent (non dismissable) notification in the task bar (cf. Figure 2 right top). As an icon we used a shield that gradually darkens in five steps, according to the *DCL* (cf. Figure 2 right bottom). In the notification, we displayed the current *DCL* value together with a button to open the control application and *re-authenticate voluntarily*. While we decided to permanently display the indicator in our study, it could also be implemented as an on-demand information source (comparable to e.g., battery level) to free up space in the task bar.

5.1.2 Short Term Indicator (ST)

The short term indicator gradually darkens the screen once the *DCL* falls below 20% (Figure 1 centre). It is therefore only visible, when a re-authentication is imminent. To avoid

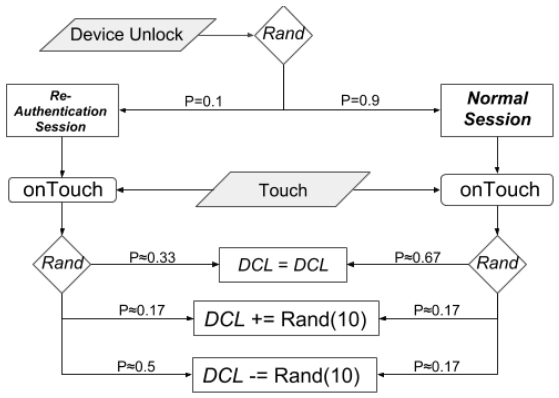

Figure 3: Schematic presentation of our simulated implicit authentication mechanism: Upon unlock of the device we determined (based on the desired false acceptance rate of 10%) whether a re-authentication should be triggered in this session (*re-authentication session*). The probabilities of user touches influencing the device confidence level (*DCL*) are altered accordingly; leading to decreases being more likely in *re-authentication sessions*. In *normal sessions* the *DCL* is more likely to remain stable.

annoyance through waiting for the grace period to end (cf. [2]), we display a notification as the dimming period begins. It shows a button to allow the user to *voluntarily re-authenticate* at any point within the grace period (Figure 2 right top).

In the study by Agarwal et al. [2] a duration of 4 seconds was chosen as shorter amounts did not allow for anticipation of the re-authentication and for longer duration testers had to wait too long for the re-authentication to appear. Due to the introduction of voluntary re-authentications the latter finding does not hold in our setting so we also explored longer grace periods. Through testing with five participants we determined a grace period duration of 8 seconds to be suitable. To address the remaining uncertainty we included a question about the desired length of the grace period in the final questionnaire.

5.2 Simulated Implicit Authentication

We followed related work and used a simulated system: Khan et al. [20] interrupted sessions after a random time period of between 5 and 30 seconds. Using a simulated system provides more control for our evaluation of the indicator concepts and helps to avoid differing false reject rates (e.g., due to hand posture) that might have an influence on the results [7, 9, 20]. We thus favoured a simulated system based on the number of touch interactions over a real implicit authentication system to keep conditions comparable. Following the medium-level false reject rate of 10% used in related work [20], our system triggers re-authentication in approximately one out of ten sessions[3]. To achieve this, we simulated DCL fluctuations as follows (cf. Figure 3):

[3] A session refers to the time between two unlocks.

5.2.1 Selection of Re-authentication Sessions

We flagged a session as a *re-authentication session* with a probability of 0.1 (to achieve 10% false rejects) upon unlocking the device. This flag influenced the random DCL fluctuations (see Figure 3) such that a re-authentication would likely appear in this session. For cases where sessions were too short for a re-authentication request to appear (i.e., the *DCL* did not fall below the threshold before the session ended), the flag would persist until a re-authentication was triggered. Depending on the flag being set or not, changes to the *DCL* were simulated differently, as explained next.

5.2.2 Alterations to the DCL

Depending on the chosen type of session (*re-authentication* or *normal*) the goal was to either decrease *DCL* or keep it stable while adding some fluctuation to make the results more believable. Each touch by the user had a chance to either trigger a change to the *DCL* (0.67 if it was a *re-authentication session*, 0.33 in a *normal session*) or leave it unchanged (with inverse probability accordingly). For *re-authentication sessions*, a decrease of the *DCL* was more likely (0.5) in comparison to increases (0.17). In *normal sessions* the probability for decreases and increases was equal at 0.17 (compare Figure 3 for an overview of the whole process). Both decreases and increases to the *DCL* could trigger a random change between 1% and 10%. Decreases resulting in a *DCL* below 20% were only executed in *re-authentication sessions*.

All probabilities were determined through a pre-study with five testers so as to create fluctuation of the *DCL* that seemed natural. A re-authentication was triggered as the *DCL* fell below 20% and completing a re-authentication reset the *DCL* to 100%. Re-authentication was suspended during calls.

5.2.3 Usage

Using this method we achieved an actual false reject rate of 7.65% in our 4-week field study. The deviation from the goal (10%) is a result of sessions that were too short to trigger a re-authentication. While we forced the next session to be a re-authentication session in those cases as described above, we did not adjust probabilities afterwards to mitigate effects on the overall false reject rate.

5.3 Re-Authentication

Voluntary re-authentication was possible using the control application (Figure 2 left) or one of the notifications tied to the indicators (Figure 2 right), i.e., the permanent notification or the notification displayed during the grace period. Information about the current *DCL* was provided by the permanent notification icon (discretised), the permanent notification, and the control application. The latter additionally featured a graph, displaying the history of the *DCL* over time (Figure 2 left).

The *re-authentication process* itself was implemented by locking the device and, hence, forcing the user to authenticate by using their default unlock mechanism. Due to technical restrictions it was not possible to offer biometric methods for re-authentication as Android requires using the backup authentication scheme in cases where the device is locked by an app. Using those methods was still possible for normal locks, i.e., locks that were not triggered by our app.

6 Evaluation

Our evaluation was guided by the these research questions:

Q1 – *Can indicators reduce annoyance caused by unpredictable re-authentication requests?* We hypothesise this to hold true due to results from related work [2, 22].

Q2 – *Are there other factors influencing annoyance caused by re-authentication requests?* We propose location, task and importance and sensitivity of the interrupted task as possible factors.

Q3 – *Do indicators nudge users to voluntarily re-authenticate?* We expected an increasing number of voluntary re-authentications for short term (due to the option to re-authenticate during the grace period) and long term indication (due to the added feedback from the task bar symbol and the graph visualisation of the *DCL*).

Q4 – *How do users perceive and respond to the introduction of voluntary re-authentication?* We expected users to like this feature, as prior work showed that letting users determine the interruption time reduced annoyance [22].

6.1 Study Design

To answer our research questions we conducted a field study (N=32). The study employed a within-subject design. Participants tested a set of four conditions for one week each, resulting in a total study length of four weeks. The order of conditions was counterbalanced.

1. (**NO**) ***No** Indication*: Our (simulated) implicit authentication scheme runs transparently in the background. Re-authentication is requested without prior indication, which resembles the current practical standard. Voluntary re-authentication is only possible from the control app, but not from notifications.

2. (**ST**) ***S**hort **T**erm*: Only the *short term* indicator is shown. Voluntary re-authentication is possible from the control app and the notification triggered with the grace period.

3. (**LT**) ***L**ong **T**erm*: Only the *long term* indicator is shown. Voluntary re-authentication is possible from the control app and the permanent notification.

4. (**SLT**) ***S**hort & **L**ong **T**erm*: Both indicators are present. All options for voluntary re-authentication are possible.

Note how both *NO* and *ST* can serve as baselines here. The *NO* condition, i.e., locking the device without giving indication, is the current *practical* state of the art and thus a natural baseline. Furthermore our *ST* condition is based on the best performing method from the study by Agarwal et al. [2] (including their recommended change of allowing for re-authentication during the grace period). As such, *ST* serves as a baseline for the best currently known scheme for indicating re-authentications.

6.2 Procedure

We recruited participants through a University mailing list and via social media. They were asked to sign a consent form and install our app from the Google Play Store, using an installation guide we provided on a dedicated website. This website also provided additional information about all study conditions and answers to frequently asked questions.

Participants had to *use the application* for four weeks with conditions automatically switching each week. They used their phones as usual with occasional interruptions by our system and a maximum of three (dismissible) *experience sampling questionnaires* per day after successful re-authentication. After each condition switch, we asked participants to fill a *weekly questionnaire* about their experience. After all conditions we concluded with a *final questionnaire*.

After four weeks, participants could uninstall the app and we invited them to participate for a *final semi-structured interview* to collect qualitative feedback (in person or via telephone). Participants received €20, plus €5 if they participated in the interview.

6.3 Collected Data

We collected *usage data* on participants' devices, including executed apps, and aggregated touch interactions, unlocks, and re-authentications. Collected data was stored on the device and transferred to our server once per day.

The *experience sampling questionnaires* asked for current location and interrupted task. We also asked if the interrupted task was perceived as sensitive and important and if the interruption was perceived as annoying.

In our *weekly questionnaires*, participants rated on a 5-point Likert scale if they felt rewarded by an increasing *DCL*, if they felt motivated to re-authenticate voluntarily, and if they perceived the system as obstructive, annoying, and easy to use. We also asked for free feedback on what they liked and disliked about the current indicator and the system in general.

In the *final questionnaire* we asked participants to rank the four conditions and explain their decision. In particular, we asked which features of the first and last choices contributed

Gender	14 (44%)	Female
	18 (56%)	Male
Mean Age	28.3	
Occupation	2 (6%)	Homemaker or retiree
	8 (25%)	Working
	22 (69%)	Student
Primary Unlock Mechanism	1 (3%)	Password
	2 (6%)	PIN
	2 (6%)	Face Recognition
	6 (19%)	Pattern
	21 (66%)	Fingerprint
Secondary Unlock Mechanism	3 (9%)	Password
	8 (25%)	PIN
	10 (31%)	Pattern
	11 (34%)	None
smart phone usage (mean)	52.7	Estimated daily unlocks
	3.6	Estimated daily usage (h)

Table 1: Demographics of the participants of our four week field study (N=32).

to their decision. For the specific indicators, we asked participants whether they would modify the duration of the grace period, if they were stressed due to the grace period, and if the long term indicator helped predicting re-authentications.

Furthermore, participants rated several statements on a 5-point Likert scale: Did they like the system, were they annoyed by the vibration or notification (*ST*), did they feel that the system influenced their behaviour, and did any bugs influence the system performance? Similarly, we asked participants if the experience sampling was annoying, and if it influenced their behaviour or the perception of the system.

Moreover, we asked if participants had read the introduction on the website and watched the introductory video we provided, if they had previous knowledge about implicit authentication, and if they had looked up app functionality or how implicit authentication worked in general on our website or other sources. Finally, we asked if they always locked their phone after use, if they thought re-authentication interrupts were more annoying than traditional authentication, and if they would consider using implicit authentication.

In the *final interview*, we asked participants to share their experiences with the systems guided by a few questions.

6.4 Participants

We recruited 36 participants. Four were excluded since their data was not properly transferred to our server. The remaining 32 people had a mean age of 28 years (18 male and 14 female; Table 1). Three participants did not submit a final questionnaire, resulting in a reduced set of 29 answers for these questions. For practical reasons we conducted the study in two runs (i.e., not all participated in parallel).

All but two participants partially agreed (n=7) or agreed (n=23) that the restriction of access to their smartphone (authentication) was important (5-point Likert scale). Participants self reported their technical knowledge as high (median=4).

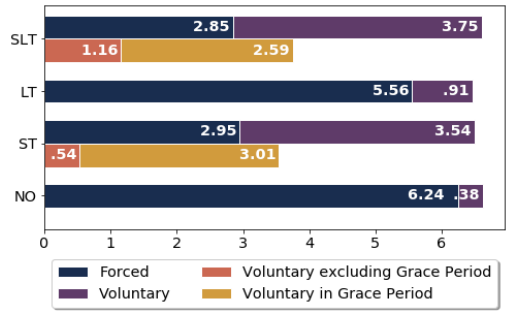

Figure 4: Average daily re-authentications by condition. Re-authentications are divided in voluntary and forced re-authentications and voluntary re-authentications are again subdivided in re-authentications during and excluding the grace period (where applicable).

6.5 Study Limitations

As participation were self-selected, our sample may not represent the general population. Our simulation might differ from the dynamics when using real implicit authentication systems. Moreover, our prototype added re-authentication on top, whereas a real system could in turn remove the initial device unlock authentication. This might have negatively affected participants' perception of our system. However, the goal was not to evaluate the general concept of implicit authentication itself but indicators for re-authentication.

7 Results

In the following report, quantitative results were tested for significance using repeated measures ANOVA with Greenhouse-Geisser correction and Bonferoni post-hoc tests. Ordinal results were tested using a Friedman test with Conover's post-hoc tests. We report significance at the level of $p < 0.05$. No effects of ordering were observed.

7.1 Usage Data

Over the course of the four week field study we observed a total of about 3.6 million touches and about 74.200 unlocks (average 84.7 unlocks per day and user) of which 5679 (7.65%) were re-authentications (1910 were voluntary, of which 646 were outside of the grace period).

The *average number of daily re-authentications* per condition is shown in Figure 4. We found no effect of the indicators on the average number of daily re-authentications. However, we found a significant difference for the average number of daily *voluntary* re-authentications ($F(1.95, 60.44)=14.75$, $p<.001$, $\eta^2 =0.322$). Post-hoc tests revealed significantly more voluntary re-authentications for all indicators ($p<.04$) compared to none (*NO*); and also significantly more for *ST* ($p=.001$) and *SLT* ($p=.003$) compared to *LT*.

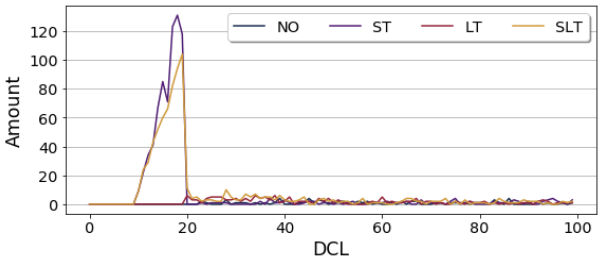

Figure 5: Distribution of *DCL* at voluntary re-authentication. There are no re-authentications below 20% for *NO* and *LT* as they had no grace period but instantly locked the device.

We also analysed re-authentications *excluding* those in the grace period, since these are arguably not strictly voluntary: We found a significant difference for relative daily voluntary use, that is, the ratio of voluntary to all re-authentications ($F(2.82, 84.53)=59.09$, $p<.001$, $\eta^2=0.165$). Post-hoc tests revealed significantly higher relative voluntary re-authentication for both *LT* ($p=.014$, Mean=14.56%) and *SLT* ($p=.008$, Mean=17.63%), compared to *NO* (Mean=5.67%). Relative voluntary re-authentications *during* the grace period were significantly higher ($F(1.0, 30.0)=5.01$, $p=.032$, $\eta^2=0.144$) for *ST* (Mean=47.49%) than for *SLT* (Mean=38.93%).

In 49.6% of cases, participants re-authenticated *before* the grace period was over, that is, they did not wait for system-triggered re-authentication (Mn=3.29s, SD=1.46). Outside of the grace period, there was no particular *DCL* at which people preferred to voluntarily re-authenticate (Figure 5).

In summary, we did not observe an effect of the indicators on the *total* average daily re-authentications. However, *voluntary* re-authentications were more common when using indicators. This can be mainly attributed to re-authentications *outside* the grace period for conditions including the long term indicator and re-authentications *during* the grace period for conditions using the short term indicator.

7.2 Experience Sampling

7.2.1 General Results

We collected 1557 answers for the experience sampling questionnaires. On a 5-point Likert scale, annoyance was rated neutral *over all conditions* (Median=3). The statements that the interrupted task was sensitive and that the interrupted task was important were also rated neutral (both Median=3). We could not find a significant impact of indicators on any rating.

Regarding the *authentication context*, participants most frequently reported "at home" for the *place* where they were interrupted, followed by transit and work. The most frequent *tasks* that were interrupted were chatting, reading, searching for information, "nothing"[4] and writing. This aligns with our logged data about the interrupted apps.

[4]This includes both cases where participants actually did nothing in particular or were not interrupted, as the re-authentication was voluntary.

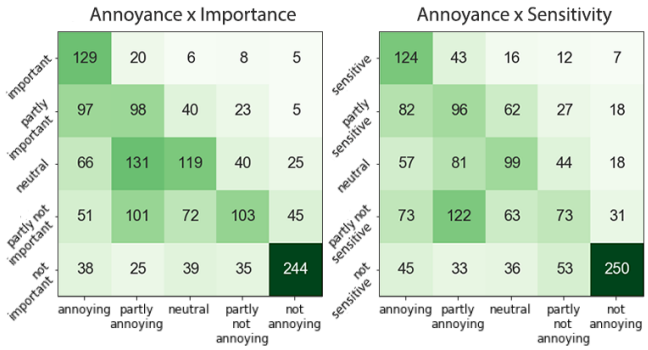

Figure 6: Frequencies of reported annoyance by importance of the interrupted task (left) and by sensitivity of the interrupted task (right). Colour encodes the shown counts.

7.2.2 Annoyance

We found significant positive (Spearman) correlations between perceived annoyance and importance of the interrupted task ($r_s=0.569$, $p<.001$) and between perceived annoyance and sensitivity of the interrupted task ($r_s=0.489$, $p<.001$), see Figure 6. We could not find effects of the day of the week or the day since the specific condition started.

The annoyance of voluntary re-authentication was perceived neutral (n=273, Median=3), similar to forced re-authentication (n=1277, Median=3). The degree to which people were annoyed by voluntary re-authentication did not significantly differ based on whether it happened during (n=76, Median=3.5) or outside of the grace period (n=136, Median=3). Voluntary re-authentication was labelled as such in the experience sampling in only 18.3% of the cases.

When comparing annoyance for the most frequently reported tasks in the experience sampling, a Friedman test revealed a significant effect of task on annoyance through re-authentication ($\chi^2(5)=36.16$, $p<.001$, $W=0.604$). Conover's post-hoc tests found that the interruption of the task "voluntary/nothing" was perceived as less annoying (Median=1) when compared to chatting ($p<.001$, Median=4), reading ($p=.002$, Median=3), searching for information ($p<.001$, Median=4), writing ($p<.001$, Median=4) and all other tasks ($p<.001$, Median=4).

In summary, we found that the annoyance caused by an interruption was influenced by a) the sensitivity of the data accessed during the interrupted task, b) the importance of the interrupted task, and c) by the task itself, as the reported task "voluntary/nothing" was perceived as less annoying.

7.3 Weekly Questionnaires

7.3.1 Voluntary re-authentications

For the weekly questionnaires we found significant differences for the motivation to voluntarily re-authenticate

($\chi^2(3)=10.05$, p=.018, W=0.498) and the feeling of reward by an increased *DCL* after re-authentication ($\chi^2(3)=21.74$, p<.001, W=0.618) with regards to the different indicators. Post-hoc analysis revealed that for *SLT* (Median=3) participants felt significantly more motivated to voluntarily re-authenticate than for *NO* (Median=1, p=.009). For all conditions using an indicator participants felt significantly more rewarded (Median-ST=2, Median-LT=2, Median-SLT=3) than in the *NO* condition (Median=1, p<.02). We found no significant differences on *perceived annoyance* of the system.

Thus, while we cannot provide evidence for a general effect of our indicators on the annoyance, we did find a positive influence of the long term indicator on the motivation to voluntarily re-authenticate. The feeling of being rewarded for re-authentication by the increased *DCL* was also significantly higher for the conditions including the long term indicator.

7.3.2 Perception of Indicators

Participants liked about the indicators that interruptions were less sudden compared to no indication (mentioned by 22 people) and that the *DCL* was visible at any time for the conditions with a long term indicator. In the *NO* condition, participants liked that re-authentication was fast (9 mentions). The gradual darkening was positively mentioned by ten participants for *ST* and by eight for *SLT*.

Interrupts were perceived as sudden by fifteen participants in the *NO* condition and by ten, four and three participants in the *LT*, *ST* and *SLT* conditions, respectively. Seven participants reported they overlooked the *DCL* visualization in the *LT* condition. Interrupts were in general perceived as annoying in all conditions (mentioned by 10, 9, 7 and 8 participants for the *NO, ST, LT* and *SLT* conditions, respectively).

7.4 Final Questionnaire

7.4.1 Ranking

In the final questionnaire, participants were asked to rate their experience with the system in general. The *overall ranking* of the different conditions (Figure 7) reveals that the combination of both *long term* and *short term* was preferred. No indication (*NO*) was ranked last. Long term (*LT*) and short term (*ST*) ranked second and third. Based on the open questions, the following reasons contributed to their choice: Sixteen participants stated to not like the sudden interruptions without indication. The combination of both short and long term (*SLT*) was particularly liked for the best overall overview and control and the continuous visualization of the *DCL* (10 and 9 mentions).

7.4.2 General Perception

As a response to our Likert scale questions, participants did not find vibration and notifications particularly annoying

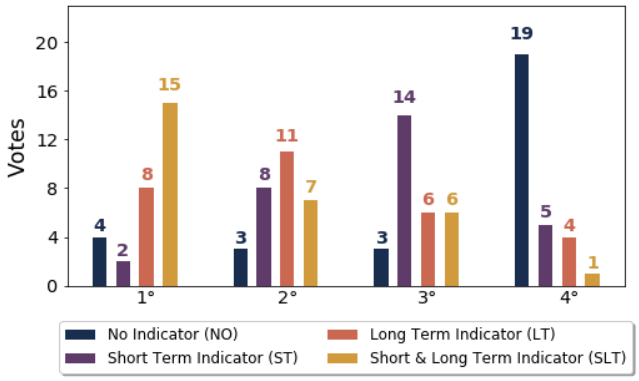

Figure 7: Participants' ranking of the different indicators. The combination of long- and short term indicator was the most preferred method while no indication was least preferred.

(Median=2). They felt neutral towards being stressed by the dimming during the grace period (Median=3). The long term taskbar symbol was considered to be helpful (Median=4) to predict re-authentications.

Participants remained neutral (Median=3) towards a possible influence of the system on their behaviour. They partly liked the design (Median=4) and partly disagreed to being negatively influenced by bugs (Median=2). They felt neutral (Median=3) about the experience sampling being annoying or influencing their behaviour or perception.

No one had profound knowledge about implicit authentication before the study nor did they review implicit authentication from other sources than the material provided by us (Median=1). There was general agreement on having read the introduction on the website and having watched the whole introductory video (Median=5).

In general, participants agreed to always locking their device (Median=5) and to authentication interrupts being more annoying than traditional authentication up front (Median=5). Regarding whether they would use the concept of implicit authentication in general, participants remained neutral (Median=3; 10 agreed or partly agreed, 5 neutral, and 14 disagreed or partly disagreed).

Finally, people would have liked a slightly longer grace period. On average they suggested 10.14 s (range 2 s–60 s).

8 Discussion & Implications

8.1 Importance & Sensitivity

While we did not find a significant effect of indicators on perceived annoyance via experience sampling, we gained related evidence and insights: We found a significant impact of *sensitivity* and *importance* of an interrupted task on the perceived annoyance. This was also pointed out in the final in-

terviews where five of the eight participants found the system interrupting an important or stressful task to be a particularly negative event:

> *I remember when I had to make a really important call and my screen was locked before I could do it. I had to answer the feedback, too, before I could finally call. Then, it was really annoying, but usually the interrupts were no problem.*

As a key insight, the situations in which participants perceived interrupts as annoying were also those that they rated as sensitive, hence, those that would require increased protection when relying on a real implicit authentication system. It might be possible that users were biased as they knew their phone was protected by their primary locking mechanism anyway in this study. Nevertheless, we believe that this topic should be investigated further.

8.2 Voluntary Re-Authentication

In contrast to related work on general interruptions [22], we could not find a positive effect of deciding when to re-authenticate on reducing annoyance. For the grace period, one explanation is that participants might not have perceived the option to re-authenticate as voluntary (as re-authentication was inevitable). More generally, our results on importance, sensitivity, and interrupted tasks all point towards the conclusion that for our participants annoyance was mostly determined by the interrupted activity and not by whether it was voluntary or not.

Nevertheless, voluntary re-authentications were mentioned as positive in open comments and the interviews, and indeed accounted for a considerable proportion of 33.6% of re-authentications (11.4% excluding grace period). Moreover, users felt significantly more motivated to re-authenticate for the combined short and long term indicator. All indicators also resulted in significantly more common use of voluntary re-authentications.

Hence, a promising approach to reduce user annoyance might be to investigate concepts that provide options for users to voluntarily re-authenticate with awareness of current activities. For instance, one person suggested to allow for voluntary re-authentication when opening an app, which often coincides with the beginning of a new activity.

8.3 Grace Period

We received mixed feedback on the grace period. Many participants liked it, in particular the more predictable nature of the interruption. For example, one participant said:

> *The more sudden the interruption happened, the more annoyed I felt about it. Surprisingly, it did not depend so much on the frequency of the interrupts. It only depended on the announcement.*

However, some participants complained that they could not use the grace period to its full extent due to light conditions and wished for a longer duration. Others used our introduced option to voluntarily re-authenticate before the device was locked. In general the desired length was very different amongst the participants which implies that an option to customise this (as also suggested by Agarwal et al. [2]) might indeed be promising for future work. We also believe that there is an impact of the personal *usability-security trade-off*, as having a (longer) grace period also implies a security risk in cases where an attacker would get hold of the device. Steps to address this might be, e.g., adapting the length of the grace period to the derivative of the *DCL* (i.e., strength of change in system confidence) or the importance of the interrupted app.

In general we see the approach of gradually dimming the screen only as a first step. Moreover, as proposed by participants of our focus group, future systems could, for example, use biometrics for re-authentication. In this case, dimming the screen could be an indicator for the user to present their face to the camera or quickly put the finger on a fingerprint scanner and thus avoid a full context switch.

8.4 Interruptions

Based on the previously discussed results, we present three recommended aspects to consider with regard to scheduling re-authentication interrupts.

1. **Sensitivity of the task:** If the user is accessing non-sensitive data (e.g., while reading a book), an upcoming re-authentication could be delayed or triggered when the task is finished, as suggested by related work [1, 4] and done in practice[5]. However, while accessing sensitive data (e.g., banking app), re-authentication should be triggered instantly to restrict further access.

2. **Importance of the task:** As users found interruptions of important tasks particularly annoying, selectively delaying such interruptions could improve users' experience with the system. This assumption is further supported by Adamczyk and Bailey [1, 4].

3. **Recent changes in confidence:** Changes in device confidence level (*DCL*) over time may be used as an indication for the necessity of an immediate interruption. While a sudden decrease in confidence most likely corresponds to an intruder taking hold of the device, a slow decrease is more likely to be caused by natural variations in the legitimate user's behaviour. However, those assumptions are, as of now, speculative and further research with a functioning implicit authentication system is necessary to verify this hypothesis.

[5] e.g., Smart Lock: https://support.google.com/android/answer/9075927?hl=en, last accessed June 25, 2019

The focus of our work was on interruptions caused by a continuous authentication system. Some lessons learned may generalise to other interruptions, such as notifications. A further factor to consider in that case is the importance of the interruption itself – which we assumed to be high for implicit authentication due to the security risk.

8.5 System Design

For our study we introduced a novel method to more realistically simulate an implicit authentication system. Our approach extended previous approaches (e.g., Khan et al. [20]) and made some of our evaluations, like the *long term* indicator, possible in the first place. We believe this to be a valuable step to enable future evaluations but also acknowledge that using our system had limitations. In particular, as the system was touch-based we introduced a bias towards interrupting tasks that used many touches, such as writing, whereas very short interactions were interrupted less. One way to address this would be to track the current app and schedule interrupts to distribute re-authentication request equally over the different tasks. Due to our use of a simulated system we were also not able to remove the primary unlocking mechanism, as this would have left participants unprotected.

However, our results from the final questionnaire suggest that neither the system itself nor the introduced experience sampling had a major effect on participants' perception or behaviour. Furthermore, vibration feedback and notifications were not perceived as annoying, and the overall design was rated as very positive.

8.6 Adoption of implicit authentication

Our participants remained neutral towards using implicit authentication and only 10 of 29 agreed or partially agreed to wanting to use it. This contrasts results of previous studies: Crawford and Renaud [9] report 90% of their participants to be interested in adopting implicit authentication. Participants also generally agreed that re-authentication was more annoying than unlocking up front.

Possible reasons could be that users underestimate the actual number of authentications they perform (on average by 38% in our study) and the accompanying benefit of implicit authentication. Other explanations include authentication overhead of a simulated system, or habituation to users' traditional unlocking methods. On the other hand, studies from related work were a lot shorter (several lab studies [2, 9, 26] and shorter field studies [20]) and thus user perception in our study developed over a longer period of time (e.g., we potentially observed a lower novelty effect). Moreover, effortless fingerprint authentication in particular has become an established method in the years between some of the earlier related work and our study, potentially shifting users' views.

As a next step we suggest evaluations with a functional implicit authentication system for a more realistic scenario. In cases where such a system cannot robustly provide sufficient security, conducting the study with users that do not lock their phones anyway might be an option. Targeting this user group has also been suggested as a mayor application area for implicit authentication in related work [19, 29].

8.7 Research Questions

Regarding our initial research questions we found all our indicators being preferred to no authentication.

We found no effect of indicators on annoyance. Annoyance was rather determined by the interrupted activity (Q1). We found sensibility, importance, and the specific interrupted task to be further factors influencing the perceived annoyance of interrupts (Q2). We also found all indicators to have a positive effect on the use of voluntary re-authentications (Q3). Finally, we found that users felt particularly motivated to voluntarily re-authenticate by combined short and long term indication. They overall perceived voluntary re-authentication as positive and used it to a considerable extent (Q4).

9 Conclusion

Motivated by previous work finding unpredictability of re-authentication requests in implicit authentication systems a source of annoyance we introduced and evaluated two indicator designs. Those included a *long term* indicator constantly showing the system confidence and a *short term* indicator announcing imminent re-authentications and giving users a grace period to finish their tasks. We also introduced *voluntary re-authentications* to allow users to re-authenticate at any time and skip the grace period if desired.

From the results of our four week field study (N=32), we found that both indicators were preferred to having no indication. We also found our newly introduced conditions to be preferred over the indicator motivated by previous work and that importance and sensitivity of the interrupted task are further influencing factors on user annoyance.

We hope for our insights to provide fertile ground for designers of future implicit authentication systems with the goal of making them as usable as possible and further support the endeavour of blending authentication seamlessly with the way that users interact.

10 Acknowledgements

Work on this project was partially funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B). This research was supported by the Deutsche Forschungsgemeinschaft (DFG), Grant No.: AL 1899/2-1.

References

[1] Piotr D Adamczyk and Brian P Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–278. ACM, 2004.

[2] Lalit Agarwal, Hassan Khan, and Urs Hengartner. Ask me again but don't annoy me: Evaluating re-authentication strategies for smartphones. In *Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[3] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. Gravityspot: guiding users in front of public displays using on-screen visual cues. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 47–56. ACM, 2015.

[4] Brian P Bailey, Joseph A Konstan, and John V Carlis. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Interact*, volume 1, pages 593–601, 2001.

[5] Jakob E Bardram, Rasmus E Kjær, and Michael Ø Pedersen. Context-aware user authentication–supporting proximity-based login in pervasive computing. In *International Conference on Ubiquitous Computing*, pages 107–123. Springer, 2003.

[6] Attaullah Buriro, Bruno Crispo, Filippo Del Frari, and Konrad Wrona. Touchstroke: smartphone user authentication based on touch-typing biometrics. In *International Conference on Image Analysis and Processing*, pages 27–34. Springer, 2015.

[7] Daniel Buschek, Fabian Hartmann, Emanuel Von Zezschwitz, Alexander De Luca, and Florian Alt. Snapapp: Reducing authentication overhead with a time-constrained fast unlock option. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3736–3747. ACM, 2016.

[8] Nathan Clarke, Sevasti Karatzouni, and Steven Furnell. Flexible and transparent user authentication for mobile devices. In *IFIP International Information Security Conference*, pages 1–12. Springer, 2009.

[9] Heather Crawford and Karen Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 1(1):7, 2014.

[10] Heather Crawford, Karen Renaud, and Tim Storer. A framework for continuous, transparent mobile device authentication. *Computers & Security*, 39:127–136, 2013.

[11] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

[12] Mohammad Omar Derawi, Claudia Nickel, Patrick Bours, and Christoph Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, pages 306–311. IEEE, 2010.

[13] Tao Feng, Ziyi Liu, Kyeong-An Kwon, Weidong Shi, Bogdan Carbunar, Yifei Jiang, and Nhung Nguyen. Continuous mobile authentication using touchscreen gestures. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for*, pages 451–456. Citeseer, 2012.

[14] Joel E Fischer, Chris Greenhalgh, and Steve Benford. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*, pages 181–190. ACM, 2011.

[15] Cristiano Giuffrida, Kamil Majdanik, Mauro Conti, and Herbert Bos. I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 92–111. Springer, 2014.

[16] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on usable privacy and security (SOUPS)*, pages 213–230, 2014.

[17] Eiji Hayashi, Sauvik Das, Shahriyar Amini, Jason Hong, and Ian Oakley. Casa: context-aware scalable authentication. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 3. ACM, 2013.

[18] Eiji Hayashi, Oriana Riva, Karin Strauss, AJ Brush, and Stuart Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 2. ACM, 2012.

[19] Hassan Khan, Aaron Atwater, and Urs Hengartner. Itus: an implicit authentication framework for android. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 507–518. ACM, 2014.

[20] Hassan Khan, Urs Hengartner, and Daniel Vogel. Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 225–239, Ottawa, 2015. USENIX Association.

[21] Lingjun Li, Xinxin Zhao, and Guoliang Xue. Unobservable re-authentication for smartphones. In *NDSS*, volume 56, pages 57–59, 2013.

[22] Daniel C McFarlane. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1):63–139, 2002.

[23] Nicholas Micallef, Mike Just, Lynne Baillie, Martin Halvey, and Hilmi Güneş Kayacik. Why aren't users using protection? investigating the usability of smartphone locking. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 284–294. ACM, 2015.

[24] Kenrick Mock, Bogdan Hoanca, Justin Weaver, and Mikal Milton. Real-time continuous iris recognition for authentication using an eye tracker. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 1007–1009. ACM, 2012.

[25] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[26] Oriana Riva, Chuan Qin, Karin Strauss, and Dimitrios Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *Proceedings of the 21st USENIX Conference on Security Symposium*, Security'12, pages 15–15, Berkeley, CA, USA, 2012. USENIX Association.

[27] Hataichanok Saevanee, Nathan L Clarke, and Steven M Furnell. Multi-modal behavioural biometric authentication for mobile devices. In *IFIP International Information Security Conference*, pages 465–474. Springer, 2012.

[28] Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. Implicit authentication through learning user behavior. In *International Conference on Information Security*, pages 99–113. Springer, 2010.

[29] Hui Xu, Yangfan Zhou, and Michael R Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *Symposium On Usable Privacy and Security, SOUPS*, volume 14, pages 187–198, 2014.