

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Bachelorarbeit

**Reauthentication Concepts for Biometric Authentication
Systems on Mobile Devices**

Sarah Delgado Rodriguez
S.Delgado@campus.lmu.de

Bearbeitungszeitraum: 18. 04. 2018 bis 04. 10. 2018
Betreuer: Sarah Prange und Lukas Mecke
Verantw. Hochschullehrer: Prof. Dr. Florian Alt

Zusammenfassung

Bei den heutigen Smartphones findet die Authentifizierung durch ein einmaliges Entsperrern des Bildschirms zu Beginn jeder Nutzungssitzung statt. Da jedoch nur bei einer geringen Anzahl dieser Sitzungen auch auf sensible Daten zugegriffen wird, entsteht ein Authentifizierungsüberschuss. Auch biometrische Methoden, die den Komfort der einmaligen Authentifizierung verbessern, können diesen Überschuss nicht reduzieren, weshalb eine große Anzahl an alternativen Methoden erforscht wird. Darunter befinden sich beispielsweise kontextbasierte Ansätze oder App-beziehungswise zeitspezifische Zugriffsbeschränkungen. Diese Methoden verringern zwar die Anzahl der nötigen Authentifizierungen, können jedoch die Sicherheit nicht signifikant verbessern. Die sogenannte Implizite Authentifizierung hat jedoch das Potenzial sowohl die nötigen Entsperrungen zu verringern als auch die Sicherheit zu erhöhen. Hierbei wird die Identität des Nutzers kontinuierlich durch das Nutzungsverhalten bestimmt. Dieses System kann jedoch zu unterbrechenden Re-Authentifizierungen führen, wenn der sogenannte Authentifizierungswert zu gering und somit die Sicherheit des Geräts gefährdet ist. Vorliegende Arbeit befasst sich mit Ansätzen zur Abmilderung der negativen Auswirkungen dieser Unterbrechungen auf die Benutzerfreundlichkeit. Dabei wird (a) der aktuelle Authentifizierungswert jederzeit angezeigt, (b) eine anstehende Re-Authentifizierung angekündigt und (c) dem Nutzer eine Möglichkeit zur freiwilligen Re-Authentifizierung jederzeit zur Verfügung gestellt. Hierauf basierend haben wir einen App-Prototyp – den Authenticator – entwickelt und in einer 4-wöchigen Nutzerstudie evaluiert ($n = 11$). Wir konnten dadurch einen positiven Effekt auf das Störungsempfinden der Nutzer und eine erhöhte Anzahl von freiwilligen Re-Authentifizierungen nachweisen. Letztere sind als nicht störend empfunden worden, daher könnte eine hohe Motivation zur freiwilligen Re-Authentifizierung ein Schlüsselmerkmal für die Verbesserung der Benutzerfreundlichkeit der Impliziten Authentifizierung sein.

Abstract

Authentication on today's smartphones usually implies an explicit authentication at the beginning of each usage session. This causes an authentication overhead, as sensitive data are accessed on only a small number of these sessions. Even though biometric methods, such as fingerprint or face recognition, increased the convenience of this one-time authentication these overhead still exists. This led to a wide range of research on alternative methods such as context-aware and app- and time-based restriction systems. Although these methods decrease the number of explicit authentications, they do not include further security barriers and, therefore, do not increase the security significantly. Implicit authentication, on the other hand, has the potential to do both by continuously verifying the user's identity through their behavior. However, this method might cause mid-task reauthentication interrupts, when the device confidence level (DCL) is too low. This raises new usability concerns, which are addressed in the present thesis, by (a) showing the recent state of the DCL at any time, (b) announcing an imminent interruption and (c) enabling the user to reauthenticate voluntarily at any time. In this context, we developed a prototype application – the Authenticator – and subsequently evaluated it during a 4-week-long field study ($n = 11$). A positive effect on the user's annoyance and an increased number of voluntary reauthentication, which were perceived as not annoying, could be proved. Thus, a high motivation to reauthenticate voluntarily might be a key feature for improving the usability of implicit authentication.

Aufgabenstellung

Due to the increasing usage of mobile devices nowadays the amount of required authentication can be a burden for users. One approach to address this are continuous implicit authentication mechanisms where the user's identity is verified over time and no explicit authentication is needed prior to device usage.

In case the device is not confident about the users identity, the user is prompted with a reauthentication request instead of having to authenticate for every usage. Previous work has shown that such authentication requests can annoy users due to their unpredictable nature.

The aim of this thesis is to explore, prototypically implement and evaluate a concept to address this and reduce annoyance caused by unpredictable interruptions for continuous authentication systems.

The project comprises the following steps:

- Comprehensive survey of related work
- Design of a concept to counteract annoyance caused by unpredictable reauthentication interrupts
- Design and implementation of a prototype application
- Planning and conduct of a user study to evaluate the developed concept
- A qualitative as well as quantitative analysis of the resulting data leading to implications for future work

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, 4. Oktober 2018

.....

Contents

1	Introduction	1
1.1	Motivation and Research Question	1
1.2	Applied Methodology	2
1.3	Content Overview	2
2	Related Work	3
2.1	Security and Usability-Perception of Traditional Authentication	3
2.2	Flexible Restriction Systems	4
2.2.1	Avoiding the All-or-Nothing Access	4
2.2.2	Context-Aware Authentication	5
2.3	Implicit Authentication	6
2.3.1	Applied Behavioral Biometric-Based Authentication Systems	6
2.3.2	Usability of Implicit Authentication Schemes	8
2.3.3	Reauthentication Interrupts	10
3	Concept Development	13
3.1	Conclusions and Observations on Related Work	13
3.2	Focus Group	14
3.2.1	Course of Actions	15
3.2.2	Results and Conclusions	15
4	Authenticator - App	19
4.1	Pseudo-System Algorithm	19
4.2	Developed Activities	20
4.3	Indicator Designs	21
4.4	Homepage	23
5	Field Study	25
5.1	Recruitment Phase	25
5.2	Participants' Demographics and Smartphone Usage	26
5.3	Gathered Data	27
6	Results of the User Study	29
6.1	General Aspects of the Quantitative Results	29
6.2	Comparing the Indicators	31
6.2.1	Results of the Weekly Surveys	31
6.2.2	Final Ranking of the Indicators	34
6.2.3	Effects on Participants' Behavior	36
6.2.4	Effect on the Perception of Interrupts	38
6.3	Other Influences on Annoyance	39
6.4	General Feedback on the User Study	41
6.4.1	Results of the Last Weekly Survey	42
6.4.2	Final Interviews	44
7	Discussion	47
7.1	Effect of Indicators	47
7.2	Reasons for Annoyance	48
7.3	Voluntary Reauthentications	48
7.4	Grace Period	50

8	Conclusions	53
A	Simplified Design Space for IA Reauthentication Indicators	57
B	Indicator Design Ideas Developed by the Participants of the Focus Group	58
C	Google Material Design Icons	59
D	Selected Screenshots of the Homepage	60
E	Online Surveys	62
F	Interview Guide for the Final Semi-Structured Interviews	72
G	Top 20 Most Interrupted Apps	73

1 Introduction

Today's smartphone users store a wide range of sensitive data on their mobile devices, including photos, videos, e-mails, passwords and bank accounts. Therefore, the need for mechanisms that prevent unauthorized individuals from accessing these devices is obvious. The high number of smartphone owners worldwide emphasizes this demand, given that the Zenith Mobile Advertising Forecasts 2017 estimates an increased smartphone ownership rate for 2018 of 66% in 52 key countries, such as the USA, UK, Germany and many more [38].

1.1 Motivation and Research Question

Traditional smartphone authentication methods are based on passing through an explicit authentication mechanism such as entering a PIN, pattern, fingerprint, et cetera, at the beginning of each interaction with the device. However, they have shown to be problematic in terms of efficiency due to the high number of sessions per day. Harbach et al. [14] found that users unlocked their phone 47.8 times per day whereas they accessed sensitive data only in 25.3 % of these sessions.

Researchers have proposed several methods for reducing this authentication overhead, such as time- or app-based access restriction methods [5, 16], the user's behavioral biometrics [3, 8, 9, 24, 29, 30] or the device's context [15, 23]. Especially user behavior-based schemes, also called implicit authentication (IA), continuous authentication or transparent authentication methods, such as gait recognition [9], continuous eye-tracking [24], or the user's tap or app-execution behavior [3, 8, 30, 29], are being investigated intensively. Some of them have already been introduced to the consumer market (e.g. Android's Smart Lock [37]).

At this point, it is important to discuss the distinction between implicit and explicit authentication methods that this thesis applies. Generally used explicit methods can be divided into biometric-, knowledge- and token-based schemes [27]. Some examples of commercially used explicit biometric authentication methods are face- and fingerprint-recognition schemes. Commonly used knowledge-based authentication methods are based on the manually operated entry of a PIN, password or pattern. Furthermore, authentication can also be executed by a physical device, called a token, such as a credit card [27], or in the smartphone context a personal Bluetooth device [37]. These methods are called explicit, because the user has to perform a specific task to authenticate, such as entering a PIN or scanning a fingerprint. In contrast, on implicit methods the authentication is based on the user's behavior while using the phone [19].

However, many implicit authentication systems will trigger explicit reauthentication, for example by locking the screen, when the mechanism is unable to confirm the current user's identity [11, 18, 21]. Another approach applied in related work is to allow access only to low-security-level apps when the system was not able to implicitly authenticate the user. If the user tries to execute a high-security-level app, a reauthentication will be triggered [6, 7, 28].

In both cases, reauthentication may occur as unpredictable mid-task interrupts. Although such an interruption is necessary to prevent intruders from further accessing the device, they appear to be annoying for most users [19].

Recent research concludes that the observed annoyance is a consequence of the unpredictability of the interrupt and the sensation of not being correctly informed about the current state of the implicit authentication system [1, 7, 19]. However, the desire of the user to be able to influence the timing of the interruption in some way [1, 22] is another important aspect.

This thesis approaches the usability issues of implicit authentication systems reported in the

literature starting from the hypothesis that the user's annoyance caused by unexpected reauthentication interrupts can be reduced by using indicators that implement the following aspects:

- feedback on the recent system status illustrating the recent device confidence level,
- announcement of the reauthentication interrupt,
- the possibility of reauthenticating voluntarily at any time, to avoid forced reauthentication interruptions.

1.2 Applied Methodology

To investigate the validity of this hypothesis, we decided to execute the following main steps.

First, we hosted a focus group discussion, to learn about the possibilities of visualizing indicators and about the participants' preferences for the different approaches.

Based on the results of this discussion, we developed an prototype Android app – called Authenticator – implementing two different types of indicators.

We decided to develop this stable application to execute a long-term field study due to the high internal and external validity that this method promises. Our goal was to collect both user feedback and statistical data on the smartphone usage and participants' reauthentication behavior.

Our expectation from the results was a possible decrease in user annoyance when an indicator is shown to the user and to be able to draw conclusions on preferences for one kind of indicator or another. We also aimed to learn more about the reasons for the usability issues on reauthentication interrupts reported in the literature.

The data collected during the user study was analyzed extensively by conducting statistical parametric and non-parametric tests to find significant differences and effects. The results of these tests allow the extraction of some important conclusions for future work on this topic.

1.3 Content Overview

This thesis starts with the analysis of related work to extract stakeholder requirements and usability and security issues of different authentication methods (Section 2). This section also gathers design recommendations and interesting aspects of the investigations on the usability of implicit authentication methods. Based on the conclusions from the related work, Section 3 addresses the development of our concept, which includes the deliberations and results of the focus group that we conducted. Subsequently, in Section 4, the developed prototype app – the Authenticator – which incorporates our concept is presented and the design decisions that we made are detailed. In Section 5, we discuss the field study ($n = 11$) that was subsequently conducted to evaluate the Authenticator prototype. The results of this user study are presented in Section 6 and discussed in Section 7. The thesis finishes with Section 8, where we draw conclusions regarding our concept and the results of the study. We also discuss the limitations of our investigations and possible starting points for future work in this final section.

2 Related Work

To gain a first insight into the usability issues caused by the traditional authentication methods and the alternative approaches that have been investigated lately, we discuss related work in this section. We aimed first at extracting the stakeholder requirements and users' behavior and opinions (Section 2.1). Subsequently, we focus both on methods that reduce the number of explicit authentications (Section 2.2) and on implicit authentication mechanisms and their usability-security trade-off (Section 2.3).

The related work discussed here was used as the motivation, reference and base for the present thesis and the later development of the Authenticator App.

2.1 Security and Usability-Perception of Traditional Authentication

As an introduction to why it is necessary to perform further research on increasing the usability of reauthentication interrupts, it is important first to review users' (un)locking behavior and the reported issues caused by it when traditional authentication methods are used.

Harbach et al. [14] gained an insight into this topic by performing an online survey with 260 participants evenly split between the operating systems Android and iOS. The results indicated that only 42.7% of the participants use some kind of secure lock screen, such as a PIN, password or pattern. Although 46.8% of them felt that unlocking their smartphone can be annoying, 95.5% liked the idea that their smartphone is protected. However, the participants who were not using a secure lock screen stated that their reasons for this decision were mostly due to inconvenience and their low risk perception. These results suggest that emotion-based factors such as annoyance and inconvenience must be kept in mind when researching smartphone authentication methods.

The subsequently conducted 4-week field study included 57 Android users, who were using PIN, pattern and slide-to-unlock mechanisms. The collected data led to the conclusion that users unlocked their phones on average 47.8 times per day with session durations of 104.1 seconds. Even though only 12 of 52 participants felt annoyed with their lock screen, the efficiency of this method is far from ideal, owing to the high number of authentications required, in particular when taking into account that only 25.3% of the sampled sessions accessed sensitive data.

Egelman et al. were able to confirm these results with their study on the same topic. They started with qualitative interviews on the participants' smartphone locking behavior [10]. Out of the 28 interview participants, 20 used a secure locking mechanism to protect their phones against online identity theft, online impersonation and general privacy concerns, such as accessing photos or contacts. Eight participants reported that they felt that unlocking their phone can be annoying.

Egelman et al. subsequently conducted a large-scale online survey with 2518 participants, of whom 58% used a secure locking method whereas 42% did not. Most participants who used a lock screen did so to restrict both strangers and their family and friends from using their device. Most of the non-lock users suggested inconvenience and a low risk perception as their main reasons.

Motivated by these results, an online experiment with 995 participants was designed to evaluate the real risks that unauthorized access could cause. 35% of the participants found at least one e-mail containing sensitive information on their device. Therefore, even disregarding the personal data stored on the device, the unauthorized access to one's smartphone could represent a real risk for many smartphone users. In this regard, Egelman et al. proposed increasing users' awareness and improving the usability of explicit authentication mechanisms and, therefore, also reducing the number of persons not using any lock mechanism for reasons of inconvenience.

The most intuitive approach based on the above investigations is to decrease the number of

necessary explicit authentications and, thus, reduce the user's effort and resulting annoyance.

2.2 Flexible Restriction Systems

Today's smartphones can be used in all kind of environments and situations, hence flexibility is one of the most important attributes of these devices. However, this quality has not been fully achieved by the currently customized authentication methods. Nevertheless, several investigations have been carried out on providing such authentication methods on an app or context basis. Their ultimate goal is to establish new methods for reducing the number of explicit authentications.

2.2.1 Avoiding the All-or-Nothing Access

To overcome the recognized disadvantages of all-or-nothing access, a survey conducted by Hayashi et al. [16] suggested restricting access by explicit authentication to only some apps of smartphones. They interviewed 20 smartphone and tablet owners for their opinions and preferences regarding the possibility of accessing some applications on their device without having to unlock it.

The participants wanted to protect 45% of their apps by an explicit authentication mechanism, 35% of them to be directly accessible and 20% to be split. Split applications have some functionalities that are directly accessible and others that are not accessible without unlocking the phone. Applications containing personal information were likely to be restricted, whereas entertainment apps were selected to be directly accessible.

Thus, the commonly used all-or-nothing access mechanisms are not fulfilling users' needs. The possibility of configuring access on an app basis might be a better approach to fulfill these needs. In their thesis studies, Winkler's [31] and Griesbeck's [13] app-based authentication mechanisms were able to verify the benefits of this concept. Winkler [31] achieved a 62% reduction in the necessary authentications with her prototype of an app-based authentication system. The developed app offers the user the possibility of selecting the apps they want to be able to access without authentication or only after entering a PIN. However, access to all secure apps during a usage session – from screen on to screen off – was possible after entering the PIN once. 7% of the participants found the approach easy to use and secure and 65% suggested that it was a very useful idea. Griesbeck [13] implemented a similar system, achieving a 66.5% reduction in explicit authentications.

Buschek et al. [5] conducted a survey on another approach to reduce the number of necessary explicit authentications. They proposed a mechanism called SnapApp where apps can be accessed during a specific period of time (Snaptime) without unlocking the phone. These short, direct accesses are called Snaps. Certain apps could be excluded from this direct access by putting them on a blacklist. The number of possible snaps between two authentications is limited and the Snaps could expire 10 minutes after the last unlock if the corresponding setting was activated. To obtain unrestricted access to the phone or when the number of Snaps has expired, the user has to unlock it using a PIN.

To collect data on the usability, effectiveness and user preferences of this authentication mechanism, Buschek et al. conducted a 30-day field study with 18 participants. 17 of these users used snaps on 20% of all unlock sessions, hence the mechanism was able to reduce the number of explicit authentications. It was rated as almost as fast as their usual authentication mechanism and especially for former swipe-to-unlock users more secure. The time-constriction principle was received positively, as the users did not feel stressed during the Snaps. Nevertheless, most of the participants felt annoyed, when they were interrupted during their task execution, due to

expiration of the Snaptime. The possibility to configure most of the parameters was used by most of the participants.

2.2.2 Context-Aware Authentication

Micallef et al. [23] further investigated the question of why some users will not use any lock screen by developing a context-sensitive authentication mechanism. They used location data and environmental sensor data to decide whether an explicit authentication is necessary or not, depending on the current context.

After developing the mentioned system as an application, it was evaluated in a user study with two groups of participants: former lock-screen users (only PIN or pattern) and former non-lock users. The field study contained the following three phases, lasting one week each. During the first phase, data were gathered to create a sensor-driven profile. The goal of the second phase was the evaluation of the application prototype. During this phase, the authentication mechanism was preset per default in every context. The participants were not able to disable it. In the third phase, the user had the possibility to decide in which contexts the systems should be used and in which the traditional explicit authentication methods.

This context-sensitive approach reduced the average number of explicit authentications per day in phase two to 29% (lock group) and 34% (non-lock group). In phase three, where the participant could chose whether or not to use the system in certain contexts, the number of authentications in the lock group was still reduced to 26%. Most of the participants in the non-lock group chose to disable the system at home and at work. In total, 18 of 20 participants adopted the system in at least one context. Both groups of participants felt secure using the developed system and, therefore, ranked the security as similar to that of PIN or pattern lock. Regarding annoyance, its rank was significantly better than with PIN/pattern lock. In terms of inconvenience, the system was ranked similar to pattern lock but better than the PIN system.

Seventeen of the 20 participants expressed interest in using the system developed by Micallef et al. if it was commercialized. Hence, such a context-based mechanism seems promising, based on its popularity in the user study and the low level of annoyance it caused.

Another context-based approach, called Context-Aware Scalable Authentication (CASA), was developed by Hayashi et al. [15]. Similarly to Micallef et al.'s [23] system, CASA decides which kind of authentication screen is ideal based on the current context. It gathers location data and takes into account if users recently used their computer nearby.

Hayashi et al. performed three user studies to evaluate their approach in detail. In the first study, they gathered location data to evaluate its usefulness as a passive factor for the authentication and found it to be effective. Second, they conducted a 1-week field study with 32 participants to collect users' feedback on a prototype using context-aware authentication based on the location. In the 10-day third field study they evaluated a new context-based authentication prototype based on the results of the second study. This system also took into account the information on the recent usage of a personal computer nearby in addition to the location data. The data for this passive factor were obtained by bluetooth communication with the computer. This connection was also used to create a pop-up notification showing on the PC whenever the phone is activated, accompanied by the recent location of the phone.

The modified prototype was evaluated in the third study after finishing a 5-day training period, to profile the different locations. The evaluation period took 4 - 9 days and was executed by 18 participants, 7 of whom had been using a lock screen prior to the study. The results indicated that participants found the approach of changing their authentication mechanism based on their location and proximity to their computers to be useful and easy to understand. They also felt that the system was not less secure than the traditional authentication methods and were positive in

being interested in using such a system if available. Even former non-lock participants showed interest in further using CASA suggesting that the correct balance of security and usability is a main reason to use such a system.

The above presented investigations show that there is an urgent need for authentication systems other than the traditional approaches. The presented flexible systems allow access to be graduated in different manners and depending on the device's context, time restrictions and users' preferences [5, 15, 16, 13, 23, 31]. Although these systems are able to reduce significantly the number of necessary explicit authentications and users' annoyance, they do not provide any further security once the phone is unlocked. Therefore, the above-mentioned systems increase the usability of authentication mechanisms but not the security, as it is mostly rated similar to the traditional approaches. This thesis aims at contributing to enhanced security with means of implicit authentication.

2.3 Implicit Authentication

Many researchers recommend the usage of implicit authentication as a second layer of security owing to its property of continuously reauthenticating the user [12, 21, 32]. In this regard, implicit authentication could provide protection against unauthorized access when the phone is left unlocked or the explicit authentication system is compromised. Also, the effect of shoulder-surfing attacks [20] or smudge attacks [2] could be reduced.

As mentioned above, the concept of implicit authentication promises authentication without explicit user interaction and thus little intrusiveness. Hence there has been a wide range of investigations on different systems realizing continuous authentication, based on different methods, such as gait recognition [9], continuous eye tracking [24] and the user's tap behavior or app-usage [3, 30, 29, 8]. To our knowledge, most studies on the topic have implemented behavioral biometric systems, and that is why we focus on this approach in the next section.

2.3.1 Applied Behavioral Biometric-Based Authentication Systems

Many kinds of behavioral biometrics have been used by researchers to authenticate a user. Some approaches require specific hardware, for instance the eyetracking-system developed by Mock et al. [24]. For the sake of clarity, their system did not run on a smartphone but on a laptop computer connected to an eye tracker system. Their mechanism was able to authenticate 37 participants with an equal error rate of 11% in a performed user study. The equal error rate (EER) denominates both the rate of false accepts, where the authentication system falsely accepts an unauthorized user, and the rate of false rejects, where the system incorrectly rejects the device's owner. If the system is configured to achieve the same value for both rates, this rate is called the equal error rate.

However, there are also implicit systems based on the onboard hardware of today's smartphones. These techniques can read and evaluate sensor data, usage statistics and location data [3, 4, 8, 9, 12, 21, 29, 30]. Subsequently, they compare this information with previously recorded behavior patterns of the device's owner. Depending on the degree of consistency, a rejection or acceptance is triggered.

Derawi et al. [9] developed such a system based on gait recognition using the onboard accelerometer sensor of a Google G1 smartphone. To collect the necessary data, the device was placed on the hip of 51 volunteers using a belt mounting case. The participants were told to walk as normally as possible in a straight line. Subsequently, the resulting signals were analyzed using cycle-detection and recognition analyses. Derawi et al. determined an equal error rate of 20.1%

which is 50% higher than EER achieved by similar methods [17]. This difference is mostly due to the fact that previous studies were executed with dedicated accelerometers with much higher sampling rates.

Shi et al. [30] developed a very different approach to implicit authentication using data gathered from app usage and content, where the browser history, location data, SMS and phone call data were recorded and analyzed. Their aim was to prove that combining multiple sources of information can increase the reliability of implicit authentication systems. Starting from these features, an authentication score is calculated that represents the system confidence in the owner's identity. Therefore, the habits of 50 participants were tracked for at least two weeks, to model the users' behavior. Attacker models have also been developed to prove the efficiency of this method against informed attackers contaminating one of the features. The developed system was able to lock out an adversary after at least 16 usages with 95% probability if the mechanism is configured such that the device owner can use the device roughly 100 times without experiencing any false rejects.

As today's smartphones are mostly operated by manual input on their touchscreens, behavioral biometric methods based on touchscreen readings have been intensively considered [3, 4, 8, 12, 21]. We present here two of these approaches as examples.

Frank et al. [12] proposed a classification framework for touch-based behavioral biometric authentication. They suggested 30 behavioral characteristics that can be extracted from the touchscreen readings. To evaluate their concepts, they executed different experiments gathering touchscreen data, such as the touch-point coordinates, timestamps, finger pressures and the covered screen surface.

Before starting the evaluation phase, Frank et al. executed an enrollment phase to create touch profiles for different users. They logged horizontal and vertical slides over the screen and found 20 stroke properties that could be used for authentication. Combining these 20 features they reached an equal error rate (EER) ranging between 2% and 3% with 11-20 strokes as inputs. With only one stroke, the EER is approximately 13%. Therefore, to achieve a high confidence, the number of strokes necessary for authentication cannot be too low. Thus, an attack executed with small strokes cannot be prevented. Depending on the kind of task executed on the phone, the algorithm could take 11-43 seconds until the first decision is available.

To validate these results, further experiments based on different scenarios were carried out, such as inter-week authentication, inter-session authentication and short-term authentication. The median EER for all of these scenarios ranged between 0% and 4%. Therefore, the chance of a false acceptance or a false rejection is low, but nevertheless has to be considered. Frank et al. suggested further investigations on reducing the EER and proposed the use of implicit authentication as a second layer of security instead of replacing the traditional systems. They also suggested combinations with other concepts, such as context-aware authentication.

Buschek et al [4] focused on another kind of touch-based behavioral authentication. Instead of using stroke inputs to authenticate the user, they suggested using the keystrokes that a user performs when inputting a text. This idea had already been investigated in the early 2000s on physical keyboards [25, 26]. To increase the accuracy, Buschek et al. combined spatial and temporal features of the keystrokes. They also discussed the importance of the hand postures and of handling changing hand postures.

Furthermore, they performed a user study with 28 participants to evaluate their approach. The users had to enter passwords given to them first in a training phase and then in an evaluation phase. Their probabilistic framework handled posture changes better than systems that ignored them, reducing the EER by 36.4-64.4%. Their idea of combining spatial and temporal features decreased the EER by 8.5-36.8% on the best case compared with other systems using only one

type of feature.

Finally, Buschek et al. expressed the following three main challenges for the application of keystroke-based behavioral biometrics in authentication: (a) mobile typing behavior changes with time, (b) entering data from different users improves the accuracy of such systems, which is not possible for password-hardening systems, and (c) mobile typing biometrics vary greatly depending on the hand posture.

Although behavior-based authentication systems are being widely investigated, the usability of these approaches has been studied on only a few occasions. False rejects and false acceptances can occur when using implicit authentication and therefore create new usability issues.

2.3.2 Usability of Implicit Authentication Schemes

To our knowledge, the first research on the usability aspects of implicit authentication was performed by Clarke et al. [6] in 2009. To analyze the stakeholder requirements on authentication methods for smartphones, they first conducted a survey with 297 participants and then a focus group discussion with 12 participants. Using this methodology, they were able to gather quantitative and qualitative data on the needs of possible stakeholders. Subsequently, they developed a framework for Non-Intrusive and Continuous Authentication (NICA) and implemented this concept in a server client prototype to evaluate it later with a laboratory study.

Their system continuously calculated the authentication confidence level and, based on this level, restricted or permitted access to different parts of the mobile device. It was not implemented to run on classical mobile devices such as smartphones but on small notebooks. The concept was based on the idea of balancing security and usability, triggering more intrusive authentication when the authentication confidence level is too low. In the worst case, the system was locked. Their goal was to reduce the inconvenience that the user perceives.

The resulting NICA-system used facial and voice recognition in addition to keystroke analysis for implicit authentication. For final evaluation of their mechanism, Clarke et al. conducted a laboratory-study with 27 participants. The users' perception was slightly skewed to the system being convenient. Face recognition as an authentication method was not very popular, owing to the greater intrusiveness of this method. The participants had to keep facing the camera to authenticate. The results indicated that users prefer convenience and usability over security. Generally, 70% of the participants favored using a continuous authentication mechanism over one-time authentication at the beginning of each session; 81% reported that they would feel more protected using such a system. Clarke et al. also performed intrusion tests and 81% of the users suggested that the system locked the invader out in a timely manner, and 86% felt that the system was secure or very secure. Nevertheless, these results may have been influenced by delays with the prototype and the use of some more intrusive authentication methods, such as face recognition. When evaluating these results, we have to consider that Clarke et al.'s system was developed in 2009 and today's implicit authentication methods are able to run with less intrusion.

So as not to influence the participants due to possible limitations of real implicit authentication systems, Crawford and Renaud [7] conducted a laboratory-study with a pseudo-prototype. This prototype supposedly used keystroke and voice biometrics for authentication. Their main goal was to find users' degree of acceptance of continuous authentication systems and gather their opinions and attitudes towards such an approach. They also wanted to evaluate a more granular approach permitting access only when the recent confidence level matches the level of security that a task has. This granular approach (some condition) was compared with two other conditions where the users had access to all or no applications without explicitly answering a challenge question regardless of their device confidence level. Figure 2.1 illustrates the connection between

these three conditions and the directly executable tasks.

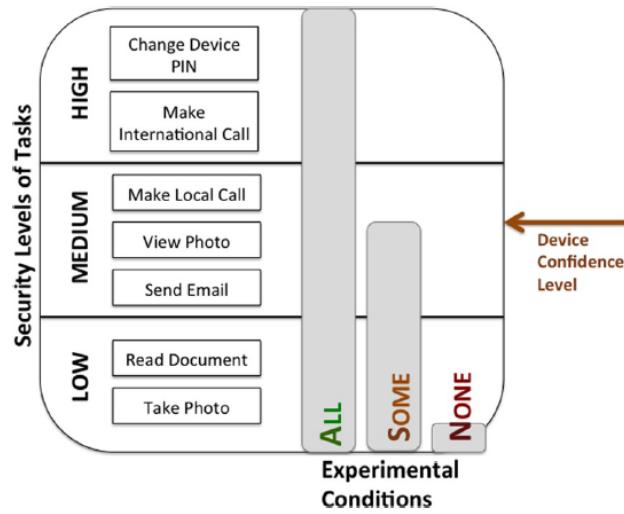


Figure 2.1: Crawford and Renaud’s [7] system: balance between the security level of each task and the accessibility due to the experimental conditions.

Crawford and Renaud evaluated their approach by carrying out a laboratory-study dividing the 30 participants randomly into three groups corresponding to the three mentioned access restriction conditions: none, some and all. All participants started the study with a low security level. Correctly answering the challenge question increased this level by one. This could be done voluntarily at any moment. All participants also had the possibility of turning the system off, when desired. To be able to evaluate the system later, the users had to execute seven tasks, starting with low-security tasks and ending with high-security ones. Following the task execution, a semi-structured interview gathered the opinions and perceptions regarding the system.

The results from this study led to the conclusion that even though barriers to accessing specific data increase the feeling of security, they are perceived as annoying or frustrating. Users even turned the system off when it became annoying. Most users felt that the mechanism is at least as secure as the authentication systems that they were currently using on their own devices. Therefore, the system could at least partly satisfy their expressed need for using more secure authentication methods on their devices. Nevertheless, the users expressed the desire to be able to assign different security levels for different tasks or applications. The users also preferred the idea of maintaining their behavioral data on their own device and not having to send it to the provider of an authentication service.

Summarizing, the requirements from the users were basically to be able to perceive some visible effects of the implicit system and to minimize the number of interruptions for explicit authentication, as they were perceived as annoying.

Khan et al. [19] further investigated the usability issues that implicit authentication can create. They focused on more detailed investigations on the possible issues, for example, on the effects of interrupts for reauthentication, false rejects or false acceptances. They performed a two-part user study consisting of a laboratory experiment and a field study evaluating two apps implementing either the traditional authentication method (EA-App) or implicit authentication (IA-App). To the best of our knowledge, Khan et al. conducted the only field study on the usability of behavioral biometric-based authentication systems.

Their prototype on implicit authentication was supposedly based on touch behavior, however

they did not implement a real authentication system but rather a pseudo-system. By using a pseudo-system, they avoided annoyance due to system failures or problems caused by specific implicit authentication schemes. They could also control better such parameters as the false reject rate or the operating threshold. The operating threshold denotes a specific value of the device confidence level. When the device confidence level drops below the operating threshold, the authentication failed either because the current user is unauthorized or because of a false reject. If this happens, in Khan et al.'s system an explicit reauthentication was triggered, by showing a lock screen.

The first phase of the user-study was a controlled laboratory experiment with 37 participants, who executed tasks first with their preferred traditional authentication method (EA-App) and subsequently with the pseudo-implicit authentication system (IA-App). Each task represented a usage session with the phone. Thus, in the EA part of the session, the user had to pick the device up, unlock it and execute the task. When the IA-App was being evaluated, the participant picked up the phone and was able to execute the task without explicitly unlocking it. The IA-App randomly interrupted the participant mid-task, showing an explicit reauthentication screen. During the following 3-day field study, the 34 participants used their normal explicit authentication system, due to security concerns. They were also randomly interrupted by the IA-App, triggering another explicit authentication.

The quantitative results showed that implicit authentication interruption did not increase the error rate of the task performance but it did increase the overall task completion time.

Qualitatively, the overall usability of implicit authentication was rated similar to that of the traditional methods. More specifically, the participants preferred implicit authentication over the traditional authentication methods in terms of ease of use, but nevertheless they felt annoyed by the interruptions. Therefore, showing the importance of the interruptions to the user is important. This could reduce the annoyance and increase the perceived security. Khan et al. recommended an indication of the status of the implicit authentication system also, even though they expressed concerns about giving clues to possible intruders with such an indicator. The security of the implicit scheme was perceived as at least as secure as the purely explicit authentication mechanisms. However, the detection delay and the false accepts that such a system includes concerned 27% and 22% of the users, respectively. A total of 63% of the participants were interested in adopting the system if it were to become available on the consumers' market. The participants were also interested in the possibility of configuring the operating threshold. Nevertheless, these results may have been affected by the short duration of the field study and the pseudo-state of the system. Not integrating physiological biometrics, such as fingerprint or face recognition, is another limitation to this research.

The above-presented investigations on the usability confirm that implicit authentication methods provide a meaningful approach that offers possibilities of being accepted as a secure authentication system by smartphone users. Nevertheless, the annoyance due to reauthentication interruptions might be an important issue and therefore was deemed a subject for further investigation.

2.3.3 Reauthentication Interrupts

Agarwal et al. [1] specifically investigated this topic. Their approach was based on the idea of first modifying the transparency of the lock screen and second announcing the authentication interrupt by gradually fading a dark screen in. They explored the effects of these visual features on the annoyance caused by the reauthentication interrupts. To do so, they developed applications implementing four different conditions based on these two ideas:

- **Imm-Dark-Imm-Lock:** When a reauthentication is necessary, the screen immediately

turns dark and is directly locked. This is the baseline condition, as explicit authentication interrupts have been implemented in this way in related work [19].

- **Imm-Trans-Imm-Lock:** On triggering the explicit authentication, the phone is immediately locked but with a transparent lock screen.
- **Grad-Dark-Imm-Lock:** In this condition, the phone is immediately locked, but the lock screen is transparent at first, then gradually turns dark.
- **Grad-Dark-Grad-Lock:** This condition allows the user to have a 4 second grace period, as the screen gradually starts to turn darker and becomes locked after the expiration of this period. During this 4 second period, the user can still interact with the phone and finish their tasks.

The reauthentication was triggered by a pseudo-implicit authentication system, supposedly based on the user touch-input and keystroke behavior. The lock screen that appeared could show either a four-digit PIN entry or a pattern lock as illustrated in Figure 2.2.

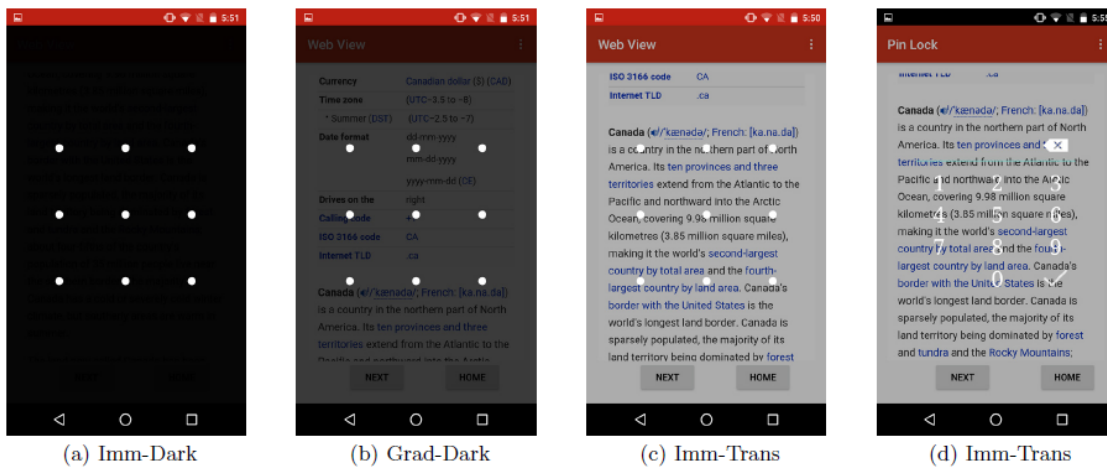


Figure 2.2: Agarwal et al.’s [1] conditions of the screen transparency. Parts (a)-(c) show the three different transparency values; (d) shows the same condition as (c), but with a PIN entry instead of the pattern lock.

The prototype was subsequently evaluated during a within-subject laboratory study with 30 participants. The users had to execute four text-entry and two e-mail reading tasks per round and were repeatedly interrupted mid-task. In the first round, the users were not interrupted, but each of the following rounds showed one of the four above-mentioned conditions, interrupting the user six times per round. Agarwal et al gathered quantitative and qualitative data during the study.

Quantitatively, there were no significant differences in the efficiency of the task executions or the task execution error rates. The qualitative results indicate that the most secure condition is Imm-Dark-Imm-Lock; however, this condition is also considered to be the most obstructive and most annoying. The perceived security of the transparent or gradually darkened lock screen was the lowest, because an intruder would be able to see the content below. The Grad-Dark-Grad-Lock condition was ranked as the least annoying. Most participants appreciated the possibility of taking advantage of the grace period that facilitates finishing a task. On the other hand, some participants reported annoyance about not being able to skip this period.

Agarwal et al. also asked the participants to suggest how the ideal reauthentication should behave. The participants noted that they would like to have a small timer or notification at the top

of the screen indicating an upcoming reauthentication interrupt. They also wanted to be able to configure the parameters of the system, such as the lock-screen background or the length of the grace period. The possibility of voluntary reauthentication to avoid forced interruptions was also mentioned. Therefore, the users wished to be able somehow to influence the interruption.

McFarlane et al. [22] obtained similar results when investigating the reduction of the usability issues of interrupts on a human machine basis. They conducted a laboratory-based user study with 36 participants, who had to play a video game on a laptop. The participants were interrupted during this task to execute a graphical matching task. When the matching task appeared, the video game was totally obscured but kept running nevertheless. To compare different types of interruptions, they defined six different experimental conditions:

- **Game only:** In this condition, the users only had to play the video game. This is the first baseline condition.
- **Match only:** Here, the user only had to execute the matching task. This is the second baseline condition.
- **Immediate:** During this condition, the matching tasks show directly, regardless of the game state.
- **Negotiated:** When a matching task occurs, it is announced by a white flash. The participant then has to decide when to execute the matching task.
- **Mediated:** Here, the matching task is shown when the mental workload of the user is low. This value is estimated by an algorithm depending on the state of the game. Thus, the matching tasks might be queued.
- **Scheduled:** In this condition, the interruption was shown once every 25 seconds.

Owing to the within-subject design of this study, participants had to pass through all six conditions.

The qualitative results proved that the preferred solution is the negotiated interrupt. The participants felt less interrupted and felt like making fewer mistakes during this phase. They also reported being able to execute the matching-task at more convenient moments in this condition. Therefore, McFarlane et al. made the following general recommendations for improving the positive user experience of interruptions:

- User interfaces that allow the user to be the most efficient, effective and precise are preferred.
- The efficiency of a user is reduced when they feel highly interrupted or distracted.
- Designs that allow the user to predict interrupts increase the efficiency and speed of the task execution and decrease the keying error rates. Nevertheless, the performance of the interrupt task, in this study the matching-task, is less efficient.
- Timing the interrupts on low-workload phases enables the user to be more efficient.

In summary, users favor less sudden interrupts. Therefore, an approach based on the possibility of influencing the exact timing of an interruption and announcing the interruption could potentially reduce the annoyance of reauthentication interrupts, as concluded by McFarlane et al. [22] and Agarwal et al. [1].

3 Concept Development

Based on the observations and suggestions in the relevant literature, we developed a new approach to improve the user experience of implicit authentication systems, which trigger reauthentication interrupts. In Section 3.1 we discuss the conclusions we extracted from the above-presented investigations and their relation to the indicators we developed. We then conducted a brainstorming and discussion with a focus group to gather and evaluate ideas on the visualization of the indicators. This is summarized in Section 3.2.

3.1 Conclusions and Observations on Related Work

The studies on the usability of the traditional, exclusively explicit authentication systems conducted by Egelman et al. and Harbach et al. [10, 14] demonstrated a need for alternative or complementary authentication mechanisms. On the one hand, the high number of explicit authentications increases the annoyance caused and leads to a large number of smartphone owners not adopting any secure lock screen [14]. On the other hand, there is a need for increased security owing to the sensitivity of the data stored on most smartphones [10].

Flexible restriction systems, such as context-aware techniques [15, 23] and time- or app-specific restriction methods [5, 13, 16, 31], mostly focus on reducing the number of explicit reauthentications and have been shown to achieve this goal properly. Nevertheless, they do not increase security, as they are reported to be only at least as secure as the traditional approach.

On the other hand, implicit authentication has the potential to improve both usability and security compared with exclusively explicit authentication systems. When used as the only authentication method, implicit authentication can reduce the number of explicit authentications [19]. Therefore, this method can be an alternative for users who currently do not use any lock screen. Furthermore, if used as a second line of defense, the security increases owing to the continuous reauthentication that implicit systems execute in the background. Therefore, intruders can be detected during the use of the phone, even when the phone has already been unlocked [3, 4, 8, 9, 12, 21, 29, 30].

Thus, if an unauthorized access, due to an intrusion or to a false reject, is detected, the implicit method needs to suspend access to the phone. This is normally done by triggering an explicit authentication. However, the interruption caused by these reauthentications is perceived as annoying by many users [19]. Our main idea was to develop indicators to reduce this annoyance by combining the following conclusions that we found promising in related work:

- **Show current state:** In Crawford and Renaud's [7] paper, some users disliked the idea of a totally invisible authentication, because they were not able to see if it is working correctly. Khan et al. [19] suggested the indication of the current system status as a solution to similar concerns from the participants of their study. We interpreted this need as a general desire to receive feedback from the system. Nevertheless, Khan et al. [19] also mentioned that such an indicator could provide a hint on the functionality of the system for intruders.
- **Announce interrupt:** The investigations of Agarwal et al. and McFarlane et al. [1, 22] revealed that predictable interruptions make the user feel less annoyed. McFarlane et al. [22] were even able to show a positive effect of the predictability on the performance of the user.
- **Delay interrupt:** McFarlane et al. [22] gave the participants the possibility of executing the interrupting task when they felt like doing so in their "Negotiated Interrupt" condition. This was perceived as positive and less interrupting. Nevertheless, delaying the reauthentication interrupt raises security concerns, because an attacker would be able to prolong access to

the device. Agarwal et al. [1] tested a trade-off between usability and security of this approach. They gave the participants a fixed 4 second grace period to finish their tasks, which was positively perceived by most. Nevertheless, some participants in Agarwal et al.'s [1] study complained about not being able to authenticate directly and felt less secure using this feature.

- **Configurable parameters:** Users usually like the possibility to personalize the configurations of the authentication system, when given the possibility [1, 5, 19]. Specifically, settings to personalize design choices or the operating threshold were mentioned [1, 19].

We discussed these suggestions with an expert group of three collaborators of the chair of media informatics at the Ludwig-Maximilians University of Munich and decided to focus on announcing the interruption and showing the current state of the authentication. However, by announcing the authentication a grace period occurs naturally and, hence we also adapted a trade-off of the “Delay interrupt” feature, similarly to Agarwal et al.'s [1] approach. Giving the user control over the actual timing of the reauthentication seemed unwise to us because it would eliminate the possibility of excluding of an intruder. Furthermore, we chose not to develop a system with configurable parameters, because our main goal was to reduce the annoyance of reauthentication interrupts by using indicators and, therefore, wanted to exclude the influence of personalized settings on this value.

Moreover, we decided to integrate the “Announce interrupt” feature in our concept, owing to the strong indications of the positive effects of increasing the predictability of an interrupt in related work [1, 22]. We found this feature to be very promising for increasing the usability of reauthentication interrupts. We also decided to inform the user about the current state of the system to increase the comprehension of the functionality of the system. Egelman et al. [10] suggested that a better informed user may also feel less annoyed about having to authenticate explicitly because in that way they are able to understand the system and see the improved security by themselves. Furthermore, the need for a feedback is comprehensible and a fundamental rule on human-machine interactions. We decided in indicating the device confidence level as the systems state, as its changes confirm the correct functionality of the system. Nevertheless, we kept the security concerns raised by Khan et al. [19] in mind and later re-evaluated them while developing our prototype.

We then convened a focus group to evaluate this basic concept and gather ideas on the possible implementations of indicators.

3.2 Focus Group

Preparing the focus group, we compiled and discussed general ideas on how to indicate reauthentications. Therefore, we divided the information that we would like to indicate into the following two categories: on the one hand, there are sudden events, such as the announcement of a necessary reauthentication; on the other hand, gradually changing information such as the current device confidence level should also be displayed. Subsequently, we compiled general possibilities of representing the changes of these values on today's smartphones. Therefore, we gathered ideas on graphical feedback, such as symbols or texts, and other visual indicators like turning on the phone's flashlight. Additionally, we also included ideas on audible and haptic feedback such as vibration and sounds. The resulting list of indication possibilities is shown in Appendix A.

Based on this very simplified indicator design space, we developed four different approaches, trying to achieve different levels of intrusiveness, and sketched them as shown in Figure 3.1. Our goal was to group and subsequently classify our indicator designs with the participants' ideas and extract the most promising conclusions as outlined in Section 3.2.2. We also aimed to identify the

accepted level of intrusiveness for such an indicator. The methodology we applied is discussed in Section 3.2.1.

3.2.1 Course of Actions

We were able to recruit five participants for the focus group by distributing an invitation on a mailing list for scientific studies of the Institute for Informatics of the University of Munich. The participants were informed about the applied data acquisition and were asked to fill in a consent form and questionnaire about their demographics and their smartphone lock screen. Table 3.1 provides the acquired information on the demographic distribution of the participants. Four of the five participants were iOS users and had configured the fingerprint unlocking method with a PIN entry as backup mechanism, as shown in Figure 3.2. They quoted ease of use, security and speed as their reasons for using this method. The remaining participants' was an Android user and did not have any secure unlock mechanisms activated. All participants were bachelor students of the University of Munich.

Gender		Age		Field of Study		Operating System	
Female	4	Minimum	18	Media informatics	5	iOS	4
Male	1	Maximum	25			Android	1

Table 3.1: Demographic data and smartphone operating system of the five focus group participants.

At the beginning of the session, we introduced ourselves and asked the participants to do so also. Subsequently, the topic was introduced with a brief presentation about reauthentications for implicit authentication and the usability issues of the interrupts. Then, we asked them to think about how an interruption could be announced and how the device confidence level could be indicated. The participants were provided with printed smartphone paper templates, different kinds of pens and markers and adhesive labels. Each participant was asked to think of their own ideas without critiquing other ideas, but they were encouraged to talk to each other and inspire each other. We also asked them to write down notes and non-visual indications on the template.

After 20 minutes, the participants had to present their ideas one by one and their thoughts about them. The other participants were encouraged to ask questions and critique the presented ideas. Subsequently, we also presented our ideas, and asked the participants for their opinions. We did not reveal our ideas in advance, to avoid biasing the participants. We then started a discussion about the similarities and differences between all the ideas and whether the participants could see any possible groupings. The last topic included possible mixed solutions and the overall preferred and disliked features. The total session took approximately 90 minutes and was held in German, because all participants were native German speakers.

3.2.2 Results and Conclusions

During the discussion with the participants, it became very clear that they would prefer an indicator that is as unintrusive as possible. The participants wished to reauthenticate rapidly and easily, for example by using fingerprint recognition. They even suggested voluntary reauthentication without an interruption or an additional click by using the onboard fingerprint sensor. Therefore, they favored an indicator integrated in the status bar, as they felt this to be the least intrusive approach (Appendix B). They disliked the idea of a floating button or even a color-changing status bar, because these approaches would be too annoying or distracting. The participants recommended



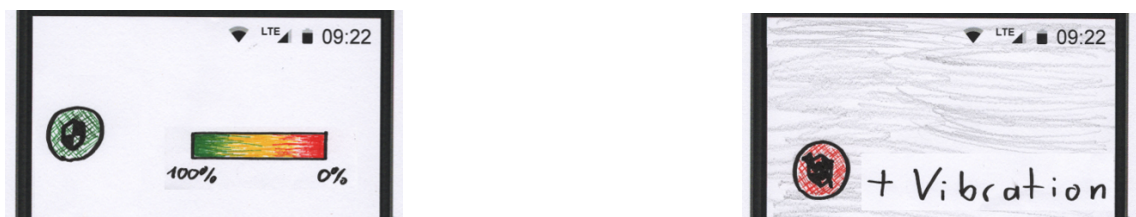
(a) Symbol and device confidence level (DCL) in the status bar. When a reauthentication is necessary, the symbol changes and a notification with vibration appears.



(b) Semi-transparent floating button with the current DCL. The transparency depends on the current DCL. When a reauthentication is triggered, the floating button shows a countdown and the phone vibrates.



(c) Color-changing status bar with a color range based on a traffic lights approach. The color depends on the current DCL. When a reauthentication is necessary, the screen starts to dim-out gradually and the phone vibrates.



(d) Color-changing floating button, with a behavior similar to the status bar in (c). When a reauthentication is triggered, the symbol changes and the smartphone vibrates. The screen also starts to dim-out gradually

Figure 3.1: Our four indicator designs. The sketches on the left show the indicator if no authentication is necessary and those on the right illustrate the announcement of an immediate reauthentication.

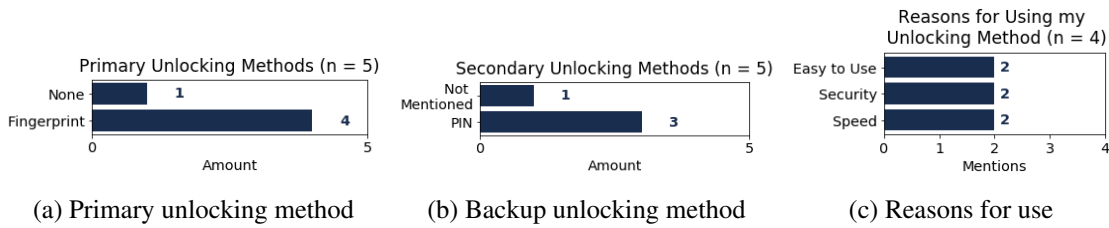


Figure 3.2: Parts (a) and (b) illustrate the reported unlocking methods from the five participants. Part (c) shows the reasons for using the unlocking mechanism for the four participants who used one. It was possible to report more than one reason.

using a progress bar approach in the status bar to visualize the current device confidence level. However, they mentioned that the status bar is not always visible and the symbol could be easily overlooked there. In this regard, the participants noted that this could be seen as positive or negative.

Regarding the announcement of an upcoming reauthentication interrupt, the participants preferred the idea of gradually dimming the screen. In their opinion, the gradual dimming of the screen is smoother and less intrusive than a notification. Some participants even reported feeling annoyed by notifications. However, the other participants did not feel annoyed by notifications and some recommended that a notification might be a more intuitive way of announcing an interruption, because this approach is the commonly used one for most other applications. The possibility of voluntary reauthentication through a notification was also mentioned and perceived very positively. Overall, they felt that both approaches are acceptable, especially in combination with a possibility to reauthenticate at any time. Most participants would combine these approaches with a configurable vibration.

We were also able to gather opinions on the general perception of an implicit authentication scheme. The participants made a point of the fact that the annoyance of a reauthentication interrupt would depend on the task that they are executing at the moment. Therefore, they wanted to be able to delay interrupts in some specific applications, such as games, camera, music or reading apps. The participants also stated the need to be able to configure this blacklist by themselves additionally to setting on and off vibration and notifications. Independently of the possible implementations of an implicit authentication method, the concern about the collected behavioral data was high. The participants were particularly worried about these data leaving their phones. Some participant even mentioned preferring the traditional authentication methods, owing to this potential risk.

Based on these findings from the focus group, we developed a first design for our prototype as shown in Figure 3.3. This design includes the indication of the device confidence level by showing a symbol and a percentage in the status bar. With this approach, an interruption is announced by a gradually dimming screen and an optional vibration. Furthermore, the possibility of voluntary reauthentication is granted through the application and a non-dismissable notification. Starting from this design, we decided to denote the announcement of an interruption as a “Short-Term Indicator” and the visualization of the device confidence level as a “Long-Term Indicator”.

As our goal is to investigate the impact of possible indicators, the dependent variable of our study is the type of indicator. Therefore, we decided to implement four different study conditions based on our indicators: No Indicator (NO), Short-Term Indicator (ST), Long-Term Indicator (LT) and Short and Long-Term Indicator (SLT). This design facilitates the comparison between the different types of indicators given that it covers all of their possible combinations. However,



(a) Long Term: Symbol and percentage of DCL in the status bar. (b) Short Term: Dimmed screen with optional vibration. (c) Notification: Can be used for voluntary reauthentication.

Figure 3.3: Part (a) shows the first design of the long-term indicator, (b) shows the announcement of the reauthentication and (c) shows the fixed notification.

the NO phase is used as a baseline, as it has been researched with similar presets in other scientific research studies [1, 19]. Our intention was to evaluate these conditions in a field-study due to the high internal and external validity that the results of this method achieve. Evaluating the Authenticator prototype in the real environment promises more realistic results. We also decided to use the within-subject study design to reduce the influence of individual user perceptions on the comparability of the different conditions as each participant has to pass through all four conditions. Compared with the between-groups study design, where each participant would evaluate only one condition, consequently a general negative perception of the system from some participants would result in poor usability results for their specific condition. This would damage the comparability of the four conditions, as one user's general perception is not efficiently detectable.

4 Authenticator - App

After developing the first design we started to implement the prototype as an Android app, called Authenticator, which is the result of a combination of the words “Authentication” and “Indicator”. We decided not to implement a real implicit authentication method, but a pseudo-system supposedly based on the user’s touch behavior. The application was implemented with German labels, as the future study participants were likely to be German native speakers. We then conducted an 8-day test study with five participants. The goal of this test study was to evaluate the prototype and find and fix possible bugs. The participants were directly recruited inside our team and our circle of friends and they were partly compensated with a 5 euro Amazon voucher. The results of this pre-study were directly integrated in the Authenticator.

As already mentioned by Khan et al. and Crawford and Renaud [7, 19], the parameters, especially the false reject rate and the increase or decrease in the device confidence level, of a pseudo-system can be controlled better. We therefore avoided specific problems of implementation, such as differing false rejection rates due to different hand postures, as mentioned by Buschek et al. [4]. This highly controllable system improves the comparability of the different user experiences, as all of the users perceive a false reject rate of approximately 10%, similar to the medium-level false reject rate evaluated by Khan et al. [19]. This means that in our system, randomly every tenth session can trigger a reauthentication interrupt. Here, a session or usage session is the period of time between unlocking the phone and locking it again.

4.1 Pseudo-System Algorithm

In a normal usage session, our system reacts to touch-inputs by triggering a one sixth chance (approximately 16.7%) either a decrease or an increase in the device confidence level (DCL). Both an increase and a decrease can have a random value from 1% to 10%. This means that in the ideal case, after performing six touches the DCL has increased once and decreased once. The sessions where a reauthentication can be triggered are called decrease sessions. In a decrease session, a decrease in the DCL is much more likely than an increase, more precisely three times more probable than in the normal session. This results in a chance of decrease of 50% and a chance of an increase of approximately 16.7% in decrease sessions. When the DCL drops below the operating threshold, here 20%, during a decrease session a reauthentication interrupt is triggered. If there is no reauthentication during a decrease session, the next session will be another decrease session, until the user performs a reauthentication. The algorithm that simulates the DCL behavior is illustrated in Figure 4.1.

When a reauthentication is necessary or voluntarily requested, the Authenticator automatically locks the screen. Even though IA as a stand-alone security system is a possible application, we did not deactivate the traditional authentication method at the beginning of each session. We took this decision because our system on a chance basis cannot guarantee reliable protection against unauthorized access. Therefore, the user has to unlock the screen first to access the screen and, if a reauthentication occurs, the user has to authenticate again. Since our aim is to compare the different study conditions, the increased number of reauthentications should have little effect on the outcome of our study. As many authors of related work have proposed the use of implicit authentication as a second line of defense [12, 21, 32], our system’s behavior gains even more validity. This recommendation is mostly due to false accepts and the detection delay of IA systems. Nevertheless, we are aware that the usage as a second-layer security system might have a negative effect on the overall annoyance caused by the system. Furthermore, it is important to mention that biometric methods, such as fingerprint and face recognition, cannot be used for reauthentication, because Android does not allow an authentication by these methods, if the phone was locked by an application. Users who use such systems

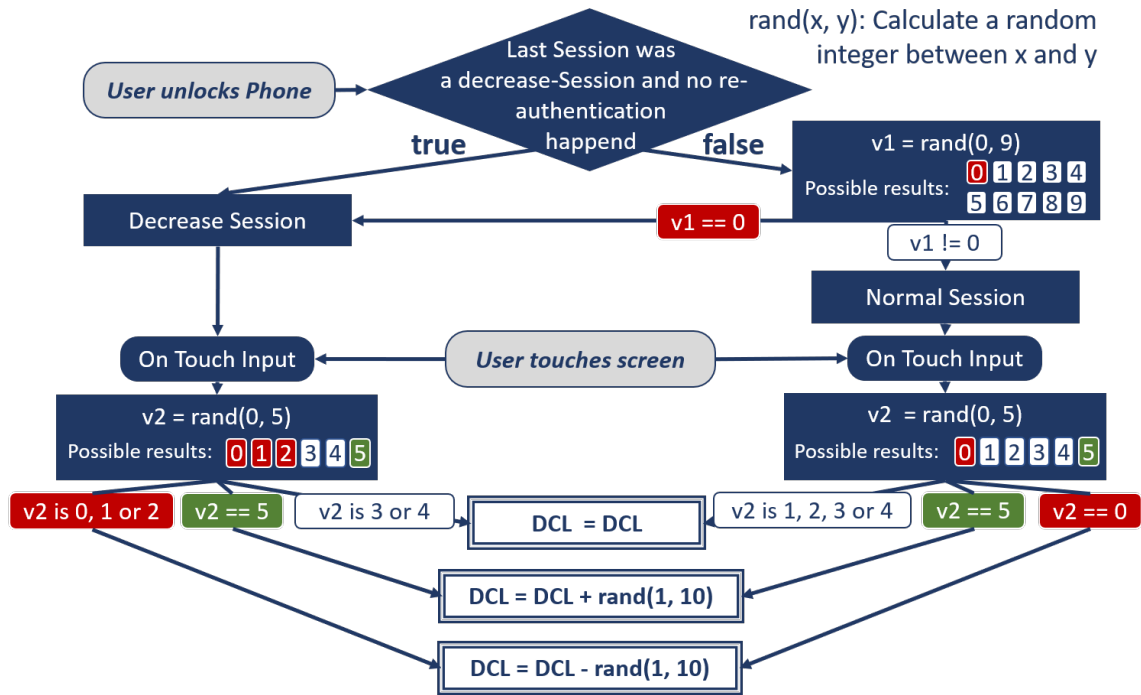


Figure 4.1: Algorithm that defines the behavior of the device confidence level (DCL). “rand(x,y)” abbreviates the calculation of a random integer value in the range from x to y.

have to reauthenticate with their backup-method such as PIN, password or pattern recognition. Nevertheless, they are still able to unlock the phone with the biometric method after a normal lock.

4.2 Developed Activities

On Android, the different screens shown by an application are called activities. The central activity of this application is the Authenticator Activity (Figure 4.2a), which illustrates information on the system, such as the current indicator, the current DCL, a plot of the DCL since the last reauthentication, the time passed since the last reauthentication, the assigned group number of the user and a randomly generated UserID. This activity also allows the user to reauthenticate voluntarily or get more detailed information by opening the Authenticator homepage, which will be presented in Section 4.4.

However, this activity is not the first one that a user will see after installing the application, because the user first has to grant some permissions necessary for the correct execution of the system. To simplify this step, we created a Permission Activity giving the user an overview over the states of the permissions, as shown in Figure 4.2b. An activated permission is marked by a green button and a deactivated permission by a red button. The user can directly access the configurations to activate these permissions by tapping on the corresponding button. The user is also directly able to gain access to information on why each permission is necessary. When all the permissions have been granted, a click on the check button will open the Authenticator Activity. If a user deactivates a permission during the study, the Permission Activity will appear again.

As our prototype was developed to evaluate the different types of indicator, we also included some in situ feedback activities. The screens appear after a reauthentication occurs, but only a maximum of three times per day. The minimal interval of time between the appearance of the feedback is 2 hours or 8 hours overnight. Therefore, we gathered qualitative feedback on the

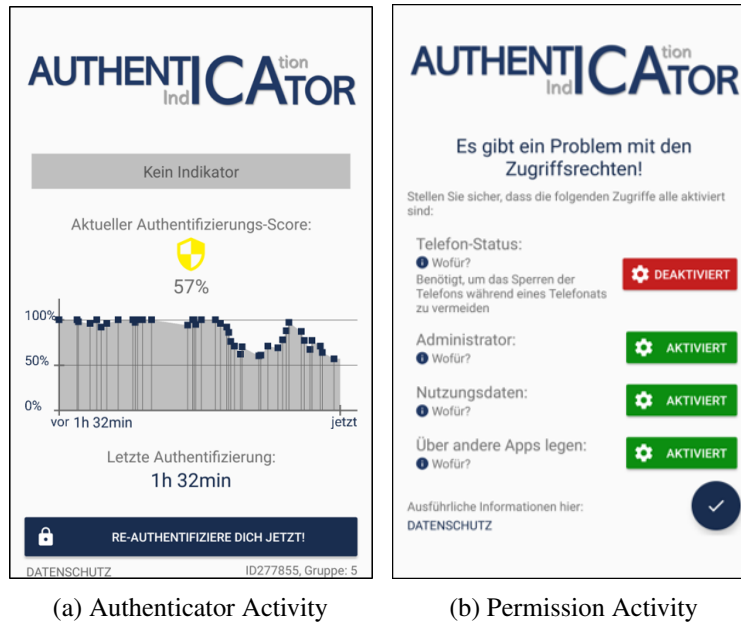


Figure 4.2: Part (a) illustrates the Authenticator Activity which is the central activity of the application. It shows information on the system. The Permission Activity in Part (b) is shown when at least one necessary permission for the application is not active.

recent location of the user, the interrupted task, its importance and sensitivity and the annoyance caused by the interruption. These activities can be seen in Figure 4.3. If the user decides to dismiss the feedback, it will appear again after the next reauthentication. Owing to the chosen within-subject study design, we had to also implement the change between the four experimental conditions. The Authenticator changes automatically between the different types of indicator when a specific number of days after the installation of the application have passed. This change is announced by a screen appearing on the first unlock after the transition. This activity also introduces the new indicator with a small pictogram and gives the user the possibility of obtaining more information by tapping on a button, which opens a description of the new indicator on the homepage.

4.3 Indicator Designs

Owing to the Android design guidelines, we decided to change the long-term indicator (LT) symbol in the status bar from our first idea, which included the device confidence level as a percentage. Instead of that, we designed an icon with different filling levels similar to a progress bar approach as illustrated in Figure 4.4a. We decided to show only five different filling levels owing to the mentioned security concerns mentioned in Khan et al.'s paper [19]. Khan et al. were worried about cluing an intruder about the behavior of the device owner, by visualizing the device confidence level. Real-time feedback on the DCL could encourage an unauthorized user to mimic the owner's behavior and, therefore, reduce the efficiency of the authentication system. By reducing the indication of the device confidence level to five levels, we avoid giving detailed feedback and consequently reduce the direct information on the correspondence of the current interaction with the saved behavior patterns.

The LT is accompanied by the Indicator-Notification (I-Notification), allowing the user to open the Authenticator Activity directly or to reauthenticate voluntarily at any moment. The user just has to open the status bar with a swipe and tap on the buttons inside the notification. Tapping

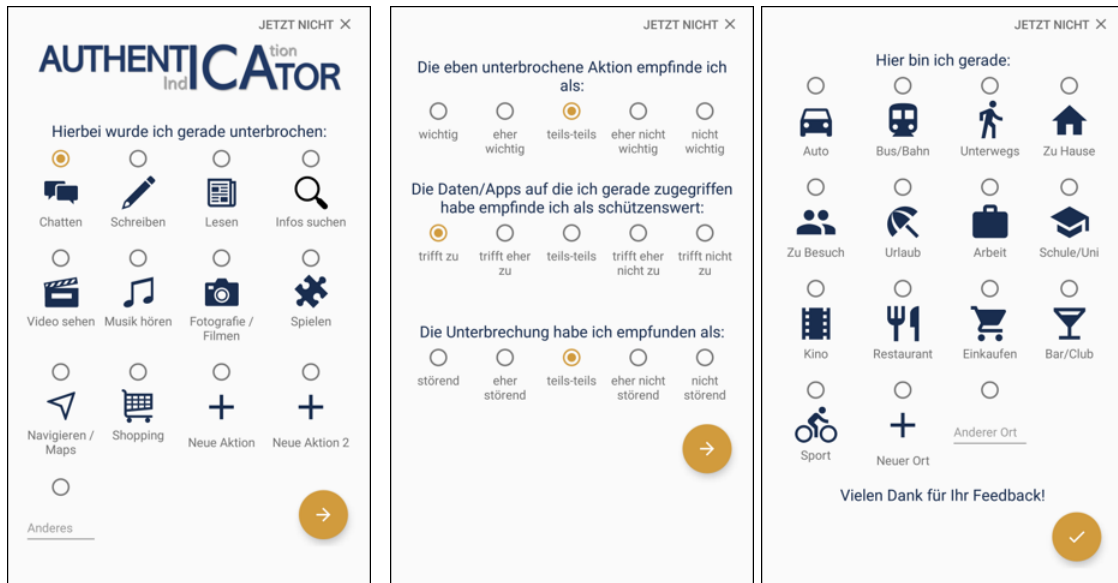


Figure 4.3: Three feedback screens gathering information on the interrupted task (left) and the recent location of the user (right). The also gathered feedback on the importance and sensitivity of the interrupted task and the annoyance due to the interruption of the user are extracted by using by a Likert scale (middle). The here used icons are based on Google Material Design Icons [34] (Appendix C).

on the notification itself will open the Authenticator Activity. The I-Notification also shows the current device confidence level, as shown in Figure 4.4b. This same notification will pop up when the short-term indicator announces an upcoming reauthentication.

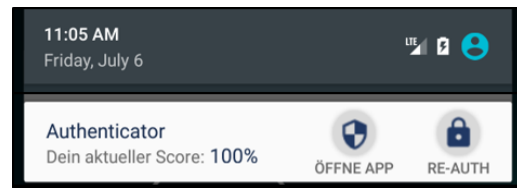
Additionally, during the ST phase, to announce the interrupt, the phone will vibrate and a shield symbol similar to the LT will appear on the status bar. When this occurs, the screen also starts to dim out. We called the period of time beginning with the announcement of the upcoming reauthentication and ending with the totally black screen and a forced reauthentication interrupt the grace period. Even though some participants of the focus group discussion that we performed felt annoyed by notifications popping up, we decided to integrate this feature on the ST. By doing so, we gave the user the possibility of reauthenticating directly and, therefore, of avoiding the grace period with the gradually dimming screen. We took this decision based on the suggestions of Agarwal et al. [1] and the experience we gathered during the test-study. The grace period for the gradual dimming was set to 8 seconds, also according to the feedback gathered during the test study.

If no immediate reauthentication is necessary, a simple notification – called No-Indicator-Notification (NO-Notification, Figure 4.4d) – and a neutral X symbol will be shown on the status bar, during the purely ST condition. This NO-Notification is necessary, because programs running in the background can be killed by the operating system. Since Android Oreo (8.0), these services will always be killed after a certain amount of time. When this occurs, our application can no longer track the touch-inputs of the user and the device confidence level will stop being calculated. As a consequence, we had to implement our system as a Foreground-Service and always show some kind of notification. To bridge the times where no indicator would be shown in the status bar, we developed the NO-Notification. This means that this notification and the neutral X symbol are shown during the whole NO condition and in the ST condition outside the grace period.

We released the finished application on Google Play Store as an internal test release and tested it during our 8-day test study. Subsequently, any remaining errors and problems were fixed.



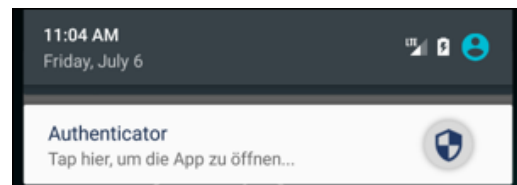
(a) Long-Term Indicator



(b) Indicator-Notification



(c) Short-Term Indicator announcing a reauthentication interrupt. The grace period starts: the screen will gradually dim-out and a shield symbol appears on the status bar. Accompanied by a vibration.



(d) No-Indicator-Notification, shown during the NO phase and the ST phase, if no reauthentication is necessary. Accompanied by a neutral X-symbol in the status bar.

Figure 4.4: Part (a) shows the LT with different filling levels. The notification when opening the status bar during the LT condition is illustrated in (b). The same notification pops up, when the ST announces an interrupt (c). The No-Notification is shown in (d). The here used shield symbol is based on an icon from the Google Material Design Icons [34] (Appendix C)

4.4 Homepage

To provide information for users, we created a homepage using the content management system Zeta Producer. The start page gave an introduction to the topic, including a video based on a vocally commented presentation, as can be seen in Appendix D.1. We also informed the user about the pseudo-state of the system in the video and on the introduction page. We decided to act in this way first to prevent users from trying to behave “correctly” to control the DCL. If their efforts have no effect, owing to the pseudo-state of the system, they could become annoyed as a result of the perceived malfunctioning. Furthermore, a real system would need a training phase and, therefore, we would also have to simulate this phase if we did not inform users about the pseudo-state of the system. In summary, we saw more disadvantages than advantages of pretending to have an actual implicit authentication system in reference to our specific evaluation goals.

The other subpages of our website gave detailed information on the installation process (Appendix D.2), the different features of the application and the phases of the user study. We also offered the possibility of contacting us through a contact form and provided instructions on resolving possible problems through different tutorials and a FAQ subpage. Furthermore, the privacy statements for the homepage and the application could also be reviewed there.

5 Field Study

As mentioned in Section 3.2.2, we decided to conduct a field study applying the within-subject design. The study participants were divided into groups with different orders regarding the types of indicators, to reduce the influence of learning and exhaustion effects. As there are four different study conditions, we developed the Latin square shown in Table 5.1 to assign the conditions to the resulting four groups. To gain a detailed insight into the topic, we decided to evaluate each condition for one week. Hence, the total duration of the field-study was four weeks.

	Week 1	Week 2	Week 3	Week 4
Group 1	NO	ST	LT	SLT
Group 2	SLT	NO	ST	LT
Group 3	LT	SLT	NO	ST
Group 4	ST	LT	SLT	NO

Table 5.1: Latin square linking the groups to the indicators.

The study procedure started with an 8-day recruitment phase where the interested parties had to register for the study (Section 5.1). The real start of the study was marked by the installation of the Authenticator app and the filling in of an online start-survey of the participants' demographics and smartphone usage (Appendix E.2). The results of this survey are further specified in Section 5.2.

As described in Section 5.3, during the field study we decided to collect both quantitative and qualitative data, to gain insights into both the perceived impressions of the participants and their actual behavior during the different study phases. Therefore, the app logged information on the smartphone usage, the answers to the feedback screens mentioned in Section 4.2 and the reauthentication interrupts. These logs were send once per day through the phone's Internet connection to a database and saved there. Additionally, the participants had to fill in weekly online surveys regarding their perception of the current indicator at the end of each week. The last survey also included questions on the general perception and usability of the Authenticator prototype. Afterwards, the participants could participate voluntarily in a semi-structured interview via telephone or a personal basis.

5.1 Recruitment Phase

We started the recruitment phase by sending an invitation through a mailing list for scientific studies of the Institute for Informatics of the University of Munich with 6075 registered participants. A similar invitation was published on Facebook groups for media informatics students of the University of Munich. The invitation included information on a 20 euro Amazon voucher reward for finishing the study. Interested Android users had 8 days to fill in an online survey to register for the study (Appendix E.1). They provided information about their smartphone, the installed Android version and the unlock method used. We received 38 registrations for the study, but had to refuse two participants because they were not using any secure unlocking mechanism. The remaining 36 persons received a welcome e-mail announcing the start date of the study and asking them to fill in a consent form. They were also added to the internal tester list on the Google Play Store console. On the start date, they received installation instructions and were asked to report their randomly generated UserID after the installation. Ultimately, 15 participants completed the installation.

After finishing the user study, we had to exclude the data sets from four of these participants. Unfortunately, we did not receive from all participants the daily logs to our database and decided to exclude the participants with fewer than 26 entries. This error may have been based on problems with the Internet connectivity of the participants' smartphones. Nevertheless, we further evaluated the collected data logged on 26 to 31 days from the remaining 11 participants.

5.2 Participants' Demographics and Smartphone Usage

After successfully finishing the installation of the application, we asked the participants to complete an online start survey about their demographic background and their smartphone usage (Appendix E.2).

Based on the results of the demographic part of this questionnaire, in the final group of 11 participants, 8 were female and the remaining 3 male. Regarding their current occupational status, 7 participants indicated that they were studying and 3 were working. The remaining participant was a houseworker. Their age distribution ranged between 20 and 61 years with a mean of 28.82 years (Figure 5.1a). The participants were also asked to rate their technical knowledge on a Likert scale ranging from "very high" to "very low". 8 of the 11 participants rated their technical knowledge as "very high" or "high", hence the median was situated on the answer "high". Figure 5.1b illustrates the whole distribution of this specific feedback.

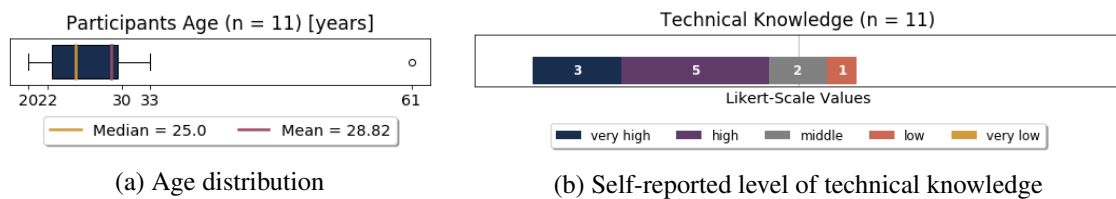
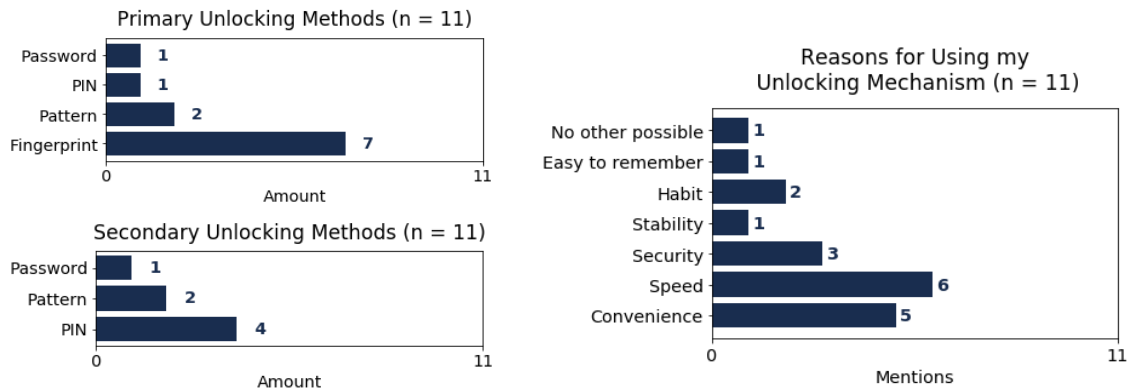


Figure 5.1: Age (a) and technical knowledge (b) of the final 11 participants.

When asked about their activated screen unlock methods, 8 of the 11 participants stated that they used a biometric explicit authentication method as their primary choice. More specifically, as shown in Figure 5.2a, 7 participants reported the use of fingerprint recognition as their primary unlocking method and one participant used face recognition. The other 3 participants used password-, PIN- and pattern-based unlocking methods. The above mentioned 8 participants who were using a biometric method had to activate a secondary unlocking mechanism to cover failures of the primary method. In this regard, the most often used backup mechanism was the PIN, as it was activated by 4 of these 8 participants. Figure 5.2a illustrates the distribution of these secondary unlocking mechanisms. We also asked the participants to express their reasons for selecting their unlocking method. As shown in Figure 5.2b, the most commonly mentioned reasons were speed – as in being able to authenticate rapidly – and convenience.

To complete the picture, we asked the participants to report their smartphone usage. First, they had to estimate how often they unlock their phones on a daily basis. The mean value of these estimate was of 40.91 daily unlocks. Subsequently, they were asked to estimate the duration of their daily smartphone usage, which resulted in a mean value of 3.28 hours daily. Both of the distributions of these values are shown in Figure 5.3. We finally asked the participants to rate the importance of the restriction of unauthorized access to their smartphone by using a 5-point Likert scale. All 11 participants agreed (6 participants) or partly agreed (5 participants) with the statement, that the restriction of access for unauthorized persons is important.



(a) Primary and secondary (backup) unlocking methods

(b) Reasons for choosing unlocking method

Figure 5.2: Part (a) shows the activated unlocking methods from the 11 study participants and (c) illustrates their reasons for using the unlocking mechanism. It was possible to report more than one reason.

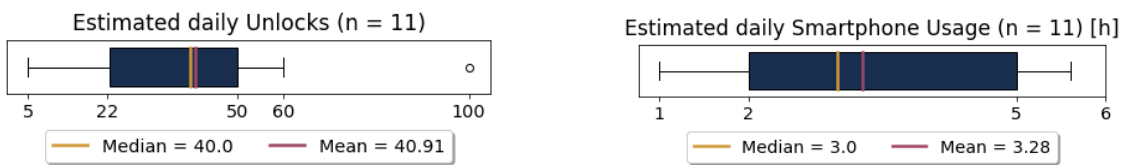


Figure 5.3: Estimated number of daily unlocks and the duration of smartphone usage, as reported by the 11 participants.

The information collected from this start survey was anonymous, as the only identification given by the participants was their randomly assigned UserID. We used this UserID to link the different quantitative and qualitative data obtained during the whole field-study and thereby to group this information per participant. This was important in order to be able to locate missing data sets from a specific participant and, thus, decide on the inclusion or exclusion of each participant in the final evaluation.

5.3 Gathered Data

During the user study, a wide range of data was gathered. First, we collected qualitative data through the above-mentioned weekly online surveys and the in situ feedback activities of the application presented in Section 4.2. The weekly online surveys accumulated data on the last shown indicator, more specifically the obstructiveness, ease of use and annoyance of the system. We also asked about the motivation for voluntary reauthentication caused by the system and the feeling of being rewarded by the increased DCL. Furthermore, we requested the participants to mention what they liked and disliked on the last indicator and if they had any problems or comments.

The last survey included some additional questions on users' general perceptions regarding the system, their general smartphone usage, their usage of the homepage, the influence of possible bugs and the influence on the feedback activities. We also asked the participants to rank the different indicators based on their personal preferences and to give reasons for their ranking.

Furthermore, we collected quantitative data that were automatically send to a Google Firebase Database once per day. These data did not include any information on the identity of the participants other than the UserID. On the one hand, we logged different counters indicating the users' smartphone usage since the last save on the database. This included the number of executed touch interactions, the number of normal screen locks, the number of reauthentications, both voluntary and forced, and the number of executed apps. A list of all executed apps since the last save was also send to the database.

On the other hand, we wanted to specify possible reasons for the annoyance felt due to the reauthentication interrupts. For this reason, we collected detailed information on the reauthentications such as a timestamp, the interrupted app, the current DCL and whether the reauthentication was voluntary or not. For reauthentications that occurred during ST or SLT conditions of the study, we also logged the time in milliseconds that passed from the beginning of the grace period to the reauthentication. This included both voluntary reauthentications during the grace period and the forced reauthentications that were triggered after the 8 seconds grace period was over. Additionally, an ID was assigned to each reauthentication, to be able to link specific reauthentications to the answers on the feedback screens that appeared three times per day after a successful reauthentication. In this regard, we also saved the responses to the feedback screen that the user gave and its timestamps. Thus, we gathered information on the interrupted task, it's importance and sensitivity, the current location of the user and the annoyance caused by the interrupt.

All these data sets were compared and evaluated under different aspects to gain a wide range of insights into the participants' opinions, perceptions and behavior. The factors influencing these points were another important focus point in the evaluation of the results.

6 Results of the User Study

In the evaluation of the results of our field study, we focused on gaining insights into different topics. Most importantly, we evaluated the influence of the dependent variable – the type of indicator – on the users’ behavior and opinions. Therefore, we compared the users’ feedback and smartphone usage during each of the four study phases. Additionally, we collected feedback on the liked and disliked features of our indicator designs and asked the participants to make a final ranking of the indicators. Furthermore, we also tried to filter possible aspects that influenced the users’ annoyance due to the reauthentication interrupts. This part of the evaluation was based on the data gathered through the in situ feedback screens, the interrupted apps and the timestamps of the interrupts.

Thus, the quantitative data collected during the user study were mostly analyzed in the context of the influence of the indicators on the smartphone usage. However, we were also able to extract some general aspects through this information.

6.1 General Aspects of the Quantitative Results

During the 4-week field study we received 323 entries to our database, which resulted in a mean value of $M = 29.36$ entries per participant with a corresponding standard deviation of $\sigma = 1.69$. As these entries were sent once per day, this mean value also represents the average duration of the user study. Correspondingly, the average number of entries per indicator is $M = 7.34$ with $\sigma = 1.10$.

Each of these database entries included the number of unlocks, touches, feedbacks, reauthentications and executed apps for the time passed since the last entry was sent to the database. Thus, the average of the daily unlocks is $M = 68.31$ ($\sigma = 40.64$), while the daily touch interactions amounted to $M = 3409.59$ ($\sigma = 2281.15$). On average, each entry contained the logged data of $M = 5.5$ reauthentications ($\sigma = 4.5$). In this context, the total number of reauthentications added up to 1776, of which 579 were executed voluntarily. Furthermore, the average number of executed apps since the last save to the database was $M = 39.98$ ($\sigma = 20.20$). Figure 6.1 illustrates the 20 most often used apps during the user study.

The algorithm that we used to simulate the behavior of the DCL on the Authenticator app (see Section 4.1) aimed at creating an equal error rate of 10%. Thus, the user should have been locked out by a reauthentication in 10% of the usage sessions. As a consequence of the above-mentioned mean values, the equal error rate of the Authenticator during this user study was 8.19%.

Finally, the entries also contained the saved responses to $M = 1.78$ ($\sigma = 1.26$) completed in situ feedback screens – as described in Section 4.2. These screens gathered information on the current location, the interrupted task, the sensitivity and importance of the interrupted task and the annoyance caused by the interrupt. In this context, the users had to select one item out of a list of locations to indicate their current location. The feedback on the interrupted task was collected by the same method. Anyway, the user had the possibility of adding new items on both screens. Figure 6.2 shows the quantitative distribution of the mentioned locations. The top 20 most mentioned tasks are also illustrated there. As we also logged the interrupted app for each reauthentication, we were able to compare the reported tasks with the real interrupted apps. The apps were mostly suitable for the reported tasks. For example, the most interrupted app was Whatsapp and the most mentioned task was chatting. An illustration of the most interrupted apps is given in Appendix G.

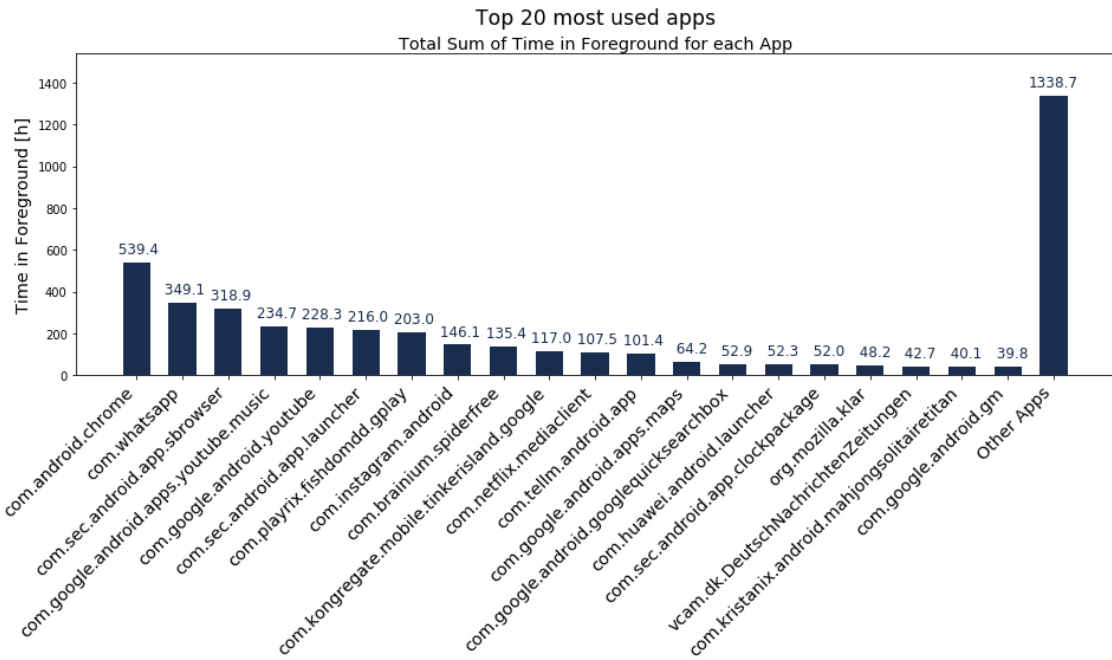


Figure 6.1: Top 20 most executed apps during the whole user study. The values correspond to the summed hours when each application was in the foreground.

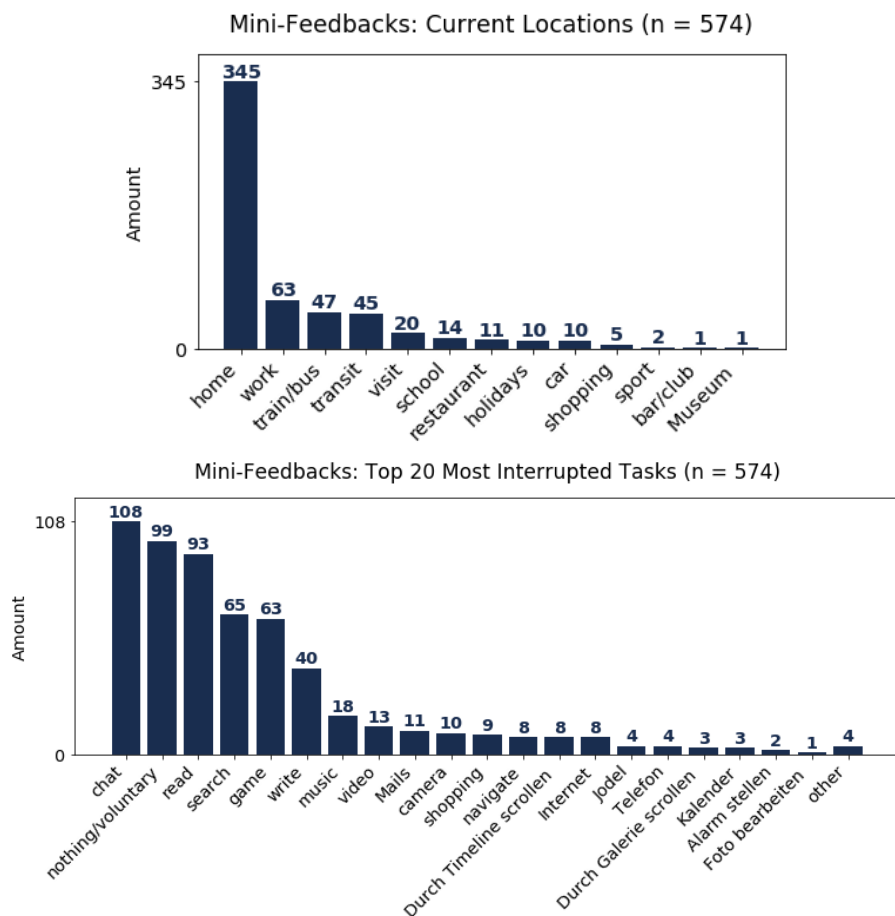


Figure 6.2: Illustration of the current locations of the users, when the in-situ feedback screens appeared and the top 20 most interrupted tasks.

6.2 Comparing the Indicators

Our evaluation of the indicators is based on the one hand on the effects of the different indicators on the user and on the other on the participants' personal preferences regarding the features of our indicator designs. We analyzed the first aspect based on both quantitative data on the smartphone usage and qualitative opinions. The latter were gathered in the online surveys through Likert scale feedbacks, open text answers and a final ranking of the indicators. In this regard, we also asked the participants to report observed buggy behavior of the application, as this might also influence their opinion.

For the Likert scale feedbacks, we tested the significance of the differences in the feedback regarding each indicator by conducting a non-parametric Friedman test of differences among repeated measures. Additionally, we applied Conover's post hoc tests for pairwise comparisons.

6.2.1 Results of the Weekly Surveys

During our 4-week field study, the participants received an e-mail at the end of each week announcing the end of the current study phase and asking them to complete an online survey about the last week's indicator. This survey's main goal was the evaluation of the usability and user experience of the specific indicator. Therefore, it began with five statements that the participants had to evaluate by choosing one item on a 5-point Likert scale. The participants could choose one following five items: 1 - "disagree", 2 - "partly disagree", 3 - "neutral", 4 - "partly agree", 5 - "agree". Figure 6.3 illustrates the distribution of these Likert scale feedbacks for the four different types of indicators. As mentioned above, these feedbacks were tested for significant differences by conducting a Friedman test with Conover's post hoc tests.

The first statement, which addressed the obstructiveness of the different conditions, was constantly rated as neutral ($Mdn = 3$). Thus, regarding the perceived obstructiveness we were not able to find significant differences between the four conditions as its chi-squared value of $\chi^2(3) = 4.463$ was not significant ($p = 0.216$).

However, the test of the feedback on the annoyance of the system during each study phase rendered a significant chi-squared value of $\chi^2(3) = 9.112$ ($p = 0.028$). Conover's post hoc tests revealed a significant difference for the comparison of the NO and LT conditions with $p = 0.029$. The other pairwise comparisons produced non-significant differences with $p > 0.14$. In this regard, the participants partly agreed to the system being annoying during the NO, ST and SLT phases ($Mdn = 4$). The LT phase was perceived neutrally ($Mdn = 3$), resulting in a lower-rated annoyance.

As shown in Figure 6.3, the participants agreed or partly agreed with the system being easy to use with a median of 4 for the ST and LT conditions and 5 for the NO and SLT conditions. We were not able to find a significant difference between the conditions ($\chi^2(3) = 7.440$, $p = 0.059$) regarding this specific feedback.

The results concerning the next statement suggested that most participants partly disagreed with the system motivating them to reauthenticate voluntarily during the NO and SLT phases of the user study ($Mdn = 2$). In the ST condition the median was 1, meaning an even higher disagreement, whereas the motivation during the LT condition was perceived as neutral ($Mdn = 3$). However, we were not able to identify a significant difference between the collected feedback on the motivation to reauthenticate voluntarily for each condition ($\chi^2(3) = 3.240$, $p = 0.356$).

The last statement addressed the users' feeling of being rewarded by the increased DCL after a reauthentication. We found a significant chi-squared value of $\chi^2(3) = 9.847$ between the four conditions ($p = 0.020$). The Conover's post hoc tests revealed a significant difference between the NO and LT conditions ($p = 0.019$) and the NO and SLT conditions ($p = 0.043$). Regarding the NO and ST conditions, the participants partly disagreed with feeling rewarded ($Mdn = 2$), while

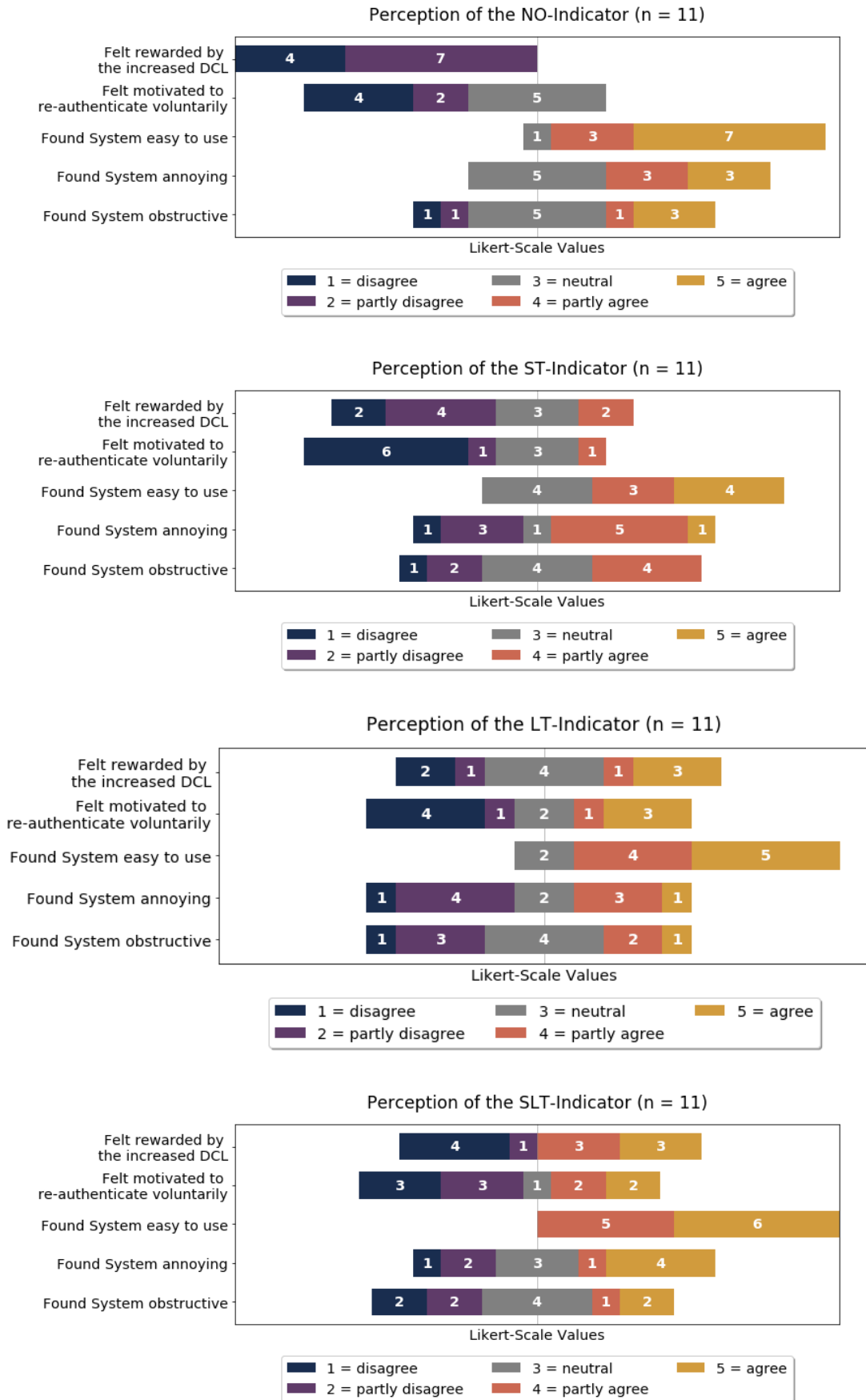


Figure 6.3: Likert scale feedback on the usability and user experience of each study phase.

the LT phase was rated neutrally ($Mdn = 3$). During the SLT condition most participants partly agreed to this statement ($Mdn = 4$). This indicates a significantly higher feeling of being rewarded during the LT and SLT phases of the study compared with the NO condition.

After this Likert scale section of the weekly online surveys, the participants were asked to highlight the features that they liked or disliked (Table 6.1) on each indicator through open text answers.

Features liked on the different Indicators									
	Fast ReAuth	ReAuth is easy	Not distracting/ Unintrusive	Announcing ReAuths makes them less sudden	Visualization of DCL at any time	Dim-Out with grace period	Motivated to do voluntary ReAuth	Pop-up of Notification enables fast voluntary ReAuth	Notification provides direct access to App
NO	3	3	2	0	0	0	0	0	1
ST	0	0	0	8	0	4	0	2	0
LT	0	1	1	5	8	0	2	0	0
SLT	0	0	0	5	5	2	2	1	0

Features disliked on the different Indicators						
	Less fast ReAuth	Sudden interrupt due to no/ poor Announcement	Vibration	Dim-Out	Overlooked DCL Symbol	Frequent popping-up of notification
NO	0	11	0	0	0	0
ST	2	3	1	2	0	0
LT	0	4	0	0	5	0
SLT	1	0	0	0	0	1

Table 6.1: Features liked and disliked on the four different indicator types extracted from open text answers in the weekly online surveys. The term “reauthentication” is abbreviated as “ReAuth”. The values are the corresponding numbers of mentions.

Regarding the NO condition, the participants most liked the possibility of fast ($n = 3$) and easy ($n = 3$) reauthentication. Other positive features mentioned were the unintrusive or not distracting design ($n = 2$) and the possibility of accessing the application easily ($n = 1$). Three participants did not report any positive features of the NO. However, all the participants mentioned features they disliked in the NO condition. In this regard, all 11 participants felt that the interrupt was very sudden due to the missing announcement.

The most liked feature during the ST phase was the announcement of the upcoming reauthentication, which made the interrupt less surprising ($n = 8$). The popping up of the notification that enabled the user to voluntarily reauthenticate fast was also mentioned ($n = 2$). Four participants specifically mentioned the dimming out of the screen and the resulting grace period as another positive aspect. Nevertheless, the dimming out of the screen was also mentioned as a negative aspect by 2 participants, and 2 other participants criticized the slower reauthentication due to the 8 second grace period. On the other hand, 3 participants felt that the short-term announcement was too sudden and did not give sufficient time to prepare for the reauthentication. One participant criticized the vibration on the appearance of the ST. All these negative aspects were reported by 8 participants, as 3 did not refer to any ST-specific disliked feature.

During the LT phase, the preferred feature was the visualization of the current DCL at any time ($n = 8$). In this regard, 5 participants reported to feeling less surprised by the interruptions due to this indicator, and 2 participants even mentioned that they used the indication of the DCL

to reauthenticate voluntarily when it was low. Other positive features were the simplicity of the reauthentication ($n = 1$) and the unintrusive design of the indicator ($n = 1$). In contrast, this same point was also the most criticized feature of the ST phase, as it seemed to be easily overlooked ($n = 5$). Thus, the interrupt was felt to be sudden for 4 participants. However, 5 participants did not mention any indicator-specific negative features.

The most liked features of the combined indicator were the announcement of the interrupt ($n = 5$) and the visualization of the DCL at any time ($n = 5$). The ST features, namely the grace period ($n = 2$) and the gradual dimming out of the screen ($n = 2$), were also received positively. The possibility of reauthenticating easily through the pop-up notification was also liked ($n = 1$). One participant did not refer to any positive aspect of the SLT condition. In contrast, 9 participants did not quote any negative feature of the SLT. The disliked aspects were the dimming out of the screen ($n = 1$) and, thus, the slower reauthentication ($n = 1$). The popping-up of the notification was also criticized by one participant.

However, some criticisms of the system were repeated during all phases of the study and cannot be seen as indicator-specific negative features (Table 6.2). In this regard, the most mentioned aspect was the general annoyance due to the interrupts ($n = 16$) and the too high frequency of the interrupts ($n = 6$). Owing to the pseudo nature of the system, the unpredictable behavior of the DCL was another constantly mentioned negative feature ($n = 6$).

Criticisms of general features		
Interrupts are annoying	Frequency of reauthentications too high	DCL behavior not predictable
16	6	6

Table 6.2: Disliked aspects mentioned during the whole study through the weekly online surveys. The values are the corresponding numbers of mentions.

To evaluate the effect of possible faulty behavior on users' perception of the last study phase, we also asked them to report observed bugs. For the NO phase of the study no errors were reported, and for the ST condition only one participant mentioned having to reauthenticate in certain cases multiple times in a short period of time. This participant reported this same behavior also in the LT phase of the study. Furthermore, 2 other participants reported some crashes of the app during this period of time and another participant mentioned a slower reaction of his smartphone during a reauthentication. One participant also observed sudden locks of the screen, even though the DCL was lower than the operating threshold. This same bug was also mentioned once regarding the SLT condition, but by another participant. During this phase, 3 participants also indicated a freezing of the screen and 2 participants also mentioned a crash of the Authenticator. Summing up, the SLT phase was the one with the most reported errors.

6.2.2 Final Ranking of the Indicators

The last weekly online survey included additional questions both on the overall perception and on the participants' opinions regarding individual features. We divided the analysis of the results provided through this questionnaire into the indicator-specific feedback and the information on other aspects of the system (see Section 6.4.1). This section describes the evaluation of the qualitative data on the indicators.

The first additional question asked the participants to rank the four different types of indicator. The participants were instructed to place the indicator they would most likely use in first place and the least likely in last place. Figure 6.4 shows the results of this ranking. From these results, most participants preferred the combination of both indicator types – the SLT condition – as it was ranked in Place 1 by 7 of the 11 participants. The indicator that was mostly placed in Place 2 was the LT ($n = 7$), followed by the SLT with 3 participants ranking it there. The ST was ranked in Place 3 by most of the participants ($n = 7$). However, the most distinguished placement was the No-Indicator condition which was ranked in Place 4 by 9 participants.

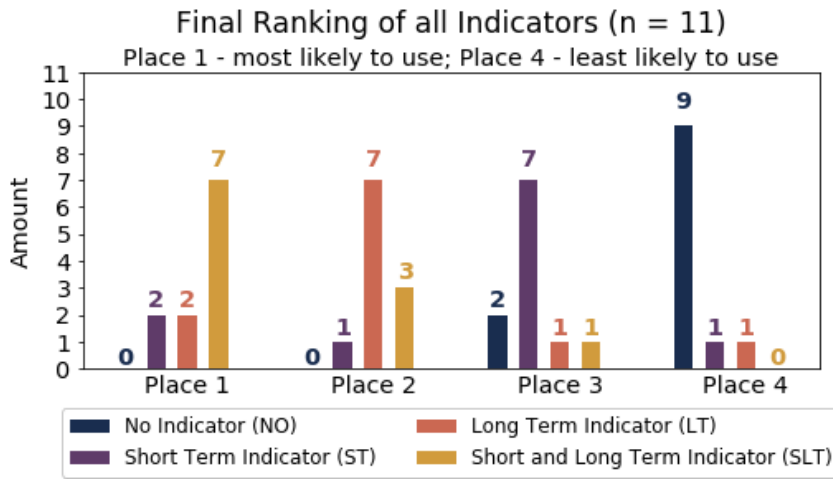


Figure 6.4: Final ranking of the four indicators. The indicator placed in “Place 1” is the most likely to be used and that in “Place 4” the least likely.

After assigning the ranking, the participants were asked to comment on the liked features of the highest rated indicator and on the disliked aspects of the lowest ranked indicator.

As mentioned before, 7 participants ranked the SLT indicator in Place 1, and therefore commented on its positive aspects. In this regard, they liked the complete overview of the state of the system, specifically the current DCL and the state of an upcoming reauthentication ($n = 6$). The possibility of reauthenticating voluntarily at any moment was also mentioned twice. One participant also referred to the unintrusive design of the SLT. As the ST and LT indicators were also placed twice in Place 1, 4 participants also commented on the positive features of these indicators. Regarding the LT condition, the participants liked the visualization of the current DCL at any time ($n = 2$) and the unintrusive design of the symbol in the status bar ($n = 1$). For the ST, this last aspect was also mentioned once. However, the grace period including the dimming out of the screen was referred to by both participants who ranked the ST highest.

The NO was the lowest ranked indicator by 9 participants and most of these participants ($n = 8$) gave the missing announcement of a reauthentication as a reason. Another disliked aspect of this indicator was the lack of an overview of the current DCL. One participant even mentioned that he thought his smartphone was experiencing an error, due to the sudden locking of the phone. The one participant who ranked the LT in the lowest place also mentioned this same perception. This participant therefore disliked the poor announcement of the upcoming interrupt. On the other hand, the one participant who disliked the ST the most referred to the bad overview of the current DCL on this indicator.

Finally, the participants were also asked to judge four statements on the design of the indicators by using a Likert scale. The scale was structured similarly to that for the weekly survey, containing the following five items: 1 - “disagree”, 2 - “partly disagree”, 3 - “neutral”, 4 - “partly agree”, 5

- “agree”. The resulting distribution of their feedback can be seen in Figure 6.5. Based on these results, most participants found the statusbar symbol of the LT to be partly helpful for predicting upcoming reauthentications ($Mdn = 4$). When asked about feeling stressed due to the dimming out of the screen in the ST and SLT conditions most participants partly agreed to feeling stressed ($Mdn = 4$). The same conditions included the popping up of the I-Notification, which was also accompanied by a vibration. As the participants of the focus group stated the annoyance of these aspects, we also asked the participants of our user study about the annoyance caused by these features. In this regard, most participants partly disagreed with both of these features being annoying ($Mdn = 2$).

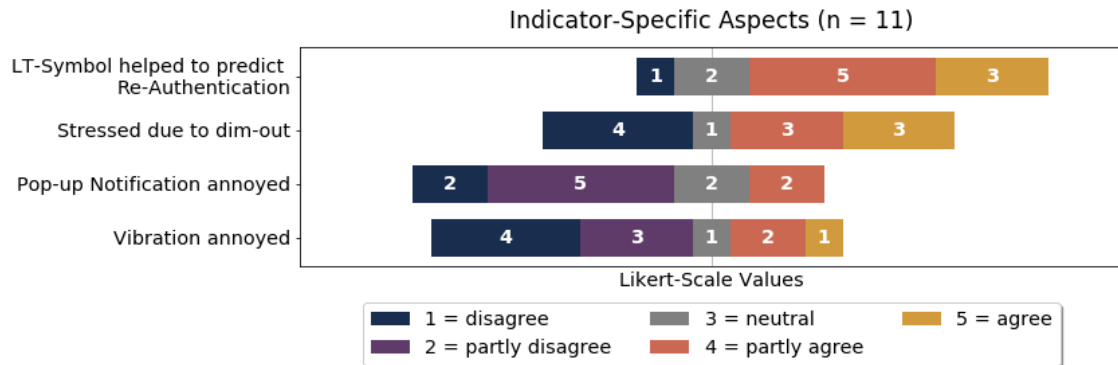


Figure 6.5: Indicator-specific Likert scale feedback on the last online survey.

Additionally to the wide range of qualitative feedback on the indicators, we also evaluated the indicators by comparing quantitative data collected during the four phases of the user study.

6.2.3 Effects on Participants' Behavior

As mentioned in Section 6.1 we received 324 daily entries to our database during the whole user study. We decided to divide these data into two categories: smartphone usage and user feedback. By doing so, we were able to analyze the effects of the indicators first on the users' behavior and second on the perception of the reauthentication interrupts. The first category contained information on the users' daily smartphone usage, such as the number of touches, unlocks, executed apps and many more. It also contained data on the reauthentications and the executed apps.

The effects of the indicators on the daily numbers of unlocks, touches, executed apps and reauthentications were tested by conducting a parametric repeated measures ANOVA test accompanied by Bonferroni post hoc tests. The test results were Greenhouse-Geisser corrected.

We were not able to find a statistically significant effect of the indicators on the number of daily touches ($F(2.786, 27.859) = 1.538, p = 0.228, \eta^2 = 0.133$), unlocks ($F(2.183, 21.829) = 0.155, p = 0.234, \eta^2 = 0.134$) and executed apps ($F(1.584, 15.840) = 0.223, p = 0.752, \eta^2 = 0.022$) between the four conditions. To analyze the daily reauthentications in depth we first tested the average numbers of all reauthentications and then only the voluntary ones. Their distributions can be seen in Figure 6.6. We were not able to find significant differences in both the average numbers of all reauthentications ($F(2.741, 27.413) = 1.514, p = 0.235, \eta^2 = 0.132$) and the voluntary ones ($F(1.466, 14.658) = 3.160, p = 0.084, \eta^2 = 0.240$).

Nevertheless, as the number of daily reauthentications varied greatly from participant to participant we decided to express the number of voluntary reauthentications in relation to all

interrupts. The repeated measures ANOVA test conducted revealed a significant effect of the indicators on these relative values ($F(1.929, 19.289) = 12.83, p < 0.001, \eta^2 = 0.562$). The Bonferroni post hoc tests suggest effects of all the indicators ($M_{ST} = 57.1\%, \sigma_{ST} = 32.4\%, M_{LT} = 18.3\%, \sigma_{LT} = 13.9\%, M_{SLT} = 35.4\%, \sigma_{SLT} = 31.1\%$) compared with the NO condition ($M_{NO} = 5.8\%, \sigma_{NO} = 5.2\%, NO - ST : p = 0.003, NO - LT : p = 0.014, NO - SLT : p = 0.049$). Furthermore, we were also able to find a significant difference between the ST and LT phases of the study ($p = 0.029$).

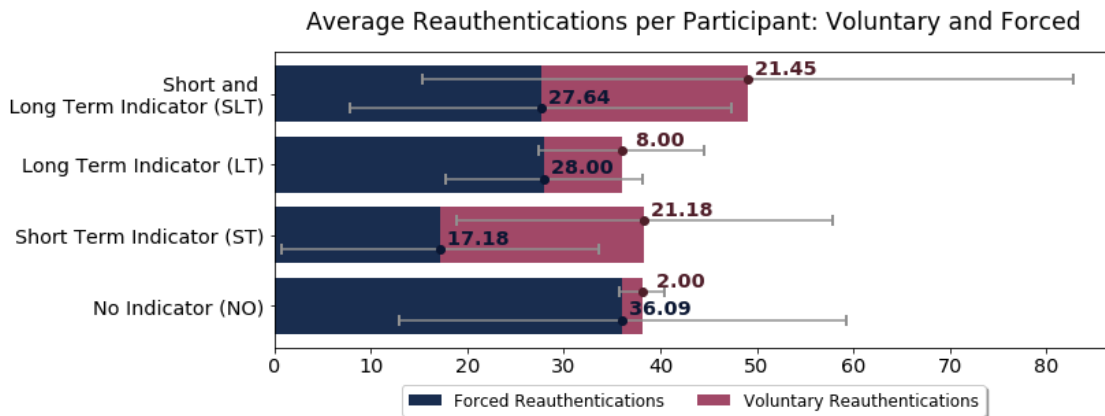


Figure 6.6: Average daily reauthentications per indicator divided into forced and voluntary.

However, as the short-term indicator included the popping up of the Indicator-Notification and therefore provided a direct shortcut to reauthenticate voluntarily during the grace period, we also divided the voluntary reauthentications into those outside the grace period and those during the grace period (Figure 6.7). By doing this, we wanted to filter specifically the influence of the I-Notification that popped up during the ST and SLT phases. The conducted tests showed a significant effect of the indicators on the relative number of voluntary reauthentications outside the grace period ($F(1.745, 17.428) = 4.877, p = 0.024, \eta^2 = 0.328$). These relative values were calculated in relation to the number of all daily reauthentications and can be seen in Table 6.3.

Voluntary reauthentications (excl. grace period)			
NO	ST	LT	SLT
5.8%	6.6%	18.4%	17.3%

Table 6.3: Relative number of daily voluntary reauthentications outside the grace period.

The post hoc tests revealed significant differences between the NO and LT conditions ($p = 0.014$) and ST and LT conditions ($p = 0.034$). Thus, the LT significantly augmented the number of voluntary reauthentications outside the grace period ($M = 18.3\%, \sigma = 13.9\%$) compared with the NO ($M = 5.8\%, \sigma = 5.2\%$) and ST ($M = 6.5\%, \sigma = 8.8\%$). Furthermore, we were able to identify a significant difference between the two indicators that allowed a voluntary reauthentication during the grace period, the ST ($M_{ST} = 50.5\%, \sigma_{ST} = 33.1\%$) and SLT ($M_{SLT} = 18.1\%, \sigma_{SLT} = 16.4\%, F(1, 10) = 5.708, p = 0.038, \eta^2 = 0.363$).

To finish the analysis on the effect of the indicators, we also evaluated the current DCL values at the moment of a voluntary reauthentication and conducted a parametric repeated measures ANOVA test on them. The Greenhouse-Geisser corrected test results suggested a significant effect of the type of indicator on the average DCL value of voluntary reauthentications ($F(1.434, 8.603) = 9.53, p = 0.009, \eta^2 = 0.614$). Bonferroni post hoc tests revealed a significant

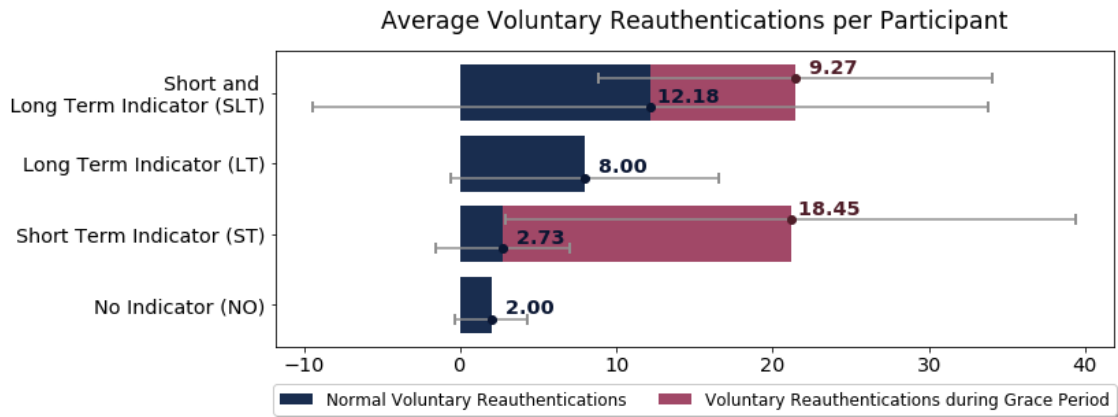


Figure 6.7: Average daily voluntary reauthentications outside and during the grace period.

difference in the values for the ST and LT conditions ($p = 0.048$). The other pairwise comparisons were not significant ($p > 0.08$). As the mean value for the ST condition was of 29.32 and for the LT phase of 40.22, the DCL for voluntary reauthentications on the ST was significantly higher than on the LT. The distribution of voluntary reauthentications per DCL values is shown in Figure 6.8.

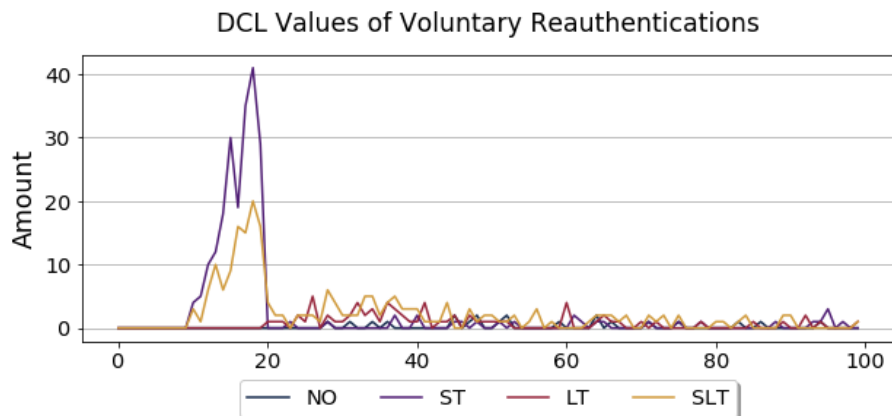


Figure 6.8: Number of voluntary reauthentications per DCL value.

6.2.4 Effect on the Perception of Interrupts

As mentioned in Section 6.1 as part of the 324 daily entries to our database, 574 responses to the in situ feedback screens were logged. The information on the importance and sensitivity of the interrupted task and the annoyance caused by the interrupt was gathered through Likert scales. The resulting distributions can be seen in Figure 6.9. Based on these data, we analyzed the effect of the type of indicator on this feedback. Similarly to evaluation of the online surveys, the results were tested for significant differences by conducting a Friedman test with Conover's post hoc tests.

The annoyance of the interrupt was perceived neutrally during the NO, ST and SLT phases of the user study ($Mdn = 3$). The interrupts in the ST condition were perceived as partly annoying ($Mdn = 2$). However, the conducted tests revealed no significant difference between the conditions ($\chi^2(3) = 1.087, p = 0.780$). We were also not able to find significant differences on the importance ($\chi^2(3) = 0.738, p = 0.864$) and sensitivity of the interrupted task ($\chi^2(3) = 0.045, p = 0.997$) for our four conditions. The interrupted tasks during the whole study

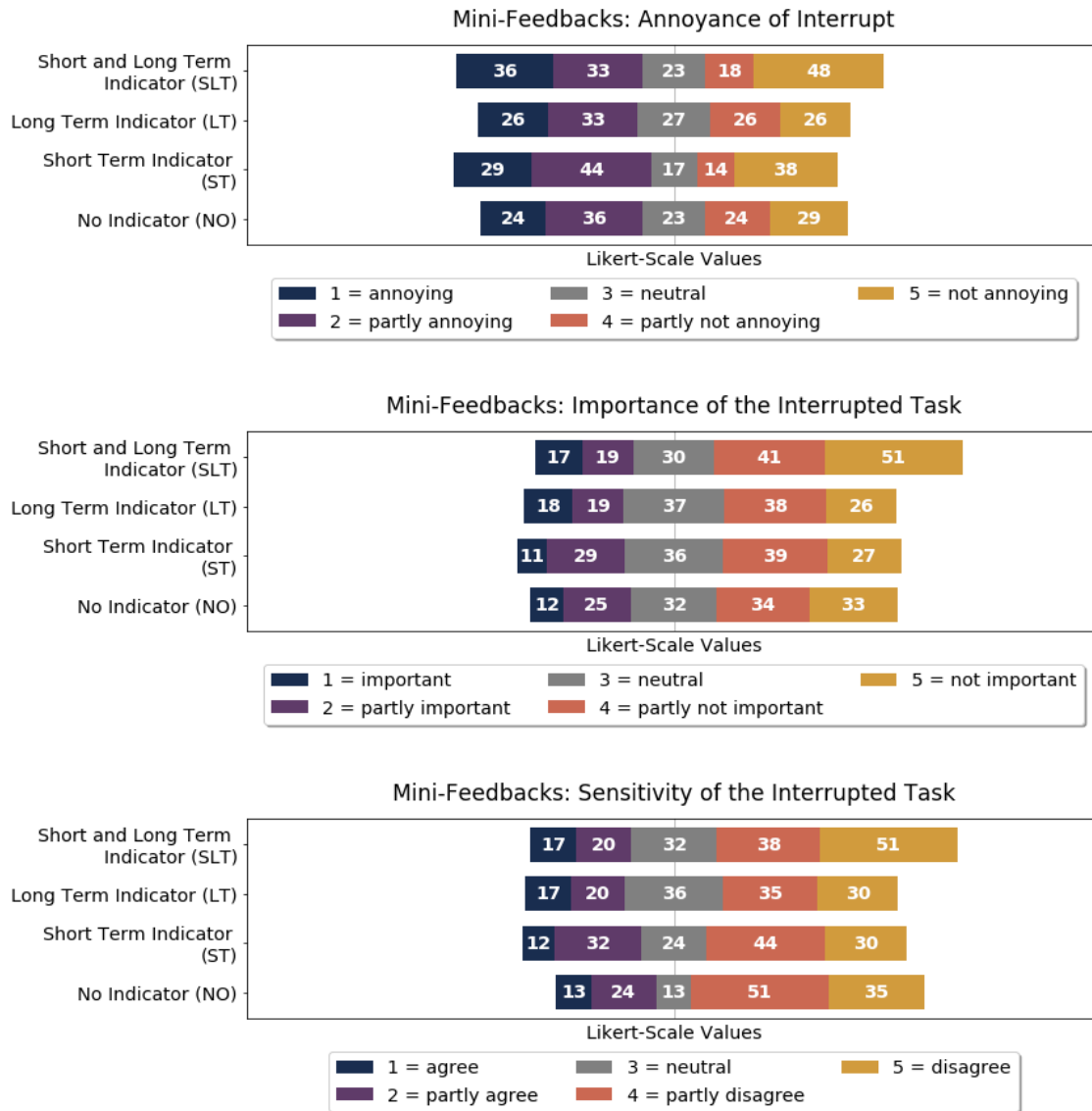


Figure 6.9: Feedback on the annoyance caused by interrupts and the importance and sensitivity of the interrupted task.

were perceived as partly not important ($Mdn = 4$). Furthermore, the participants disagreed partly on the interrupted task being sensitive ($Mdn = 4$).

6.3 Other Influences on Annoyance

As mentioned above, the effect of the indicators on the annoyance caused by the interrupts was not statistically significant. Therefore, we tried to find other aspects affecting this annoyance. In this regard, we analyzed the Spearman correlations between this annoyance and the importance of the interrupted task and this annoyance and the sensitivity of the interrupted task. Based on the results of this test, the interruption of important tasks caused a higher annoyance ($r_s = 0.667, p < 0.001$). Furthermore, the interruption of a sensitive task was also more annoying ($r_s = 0.527, p < 0.001$). These correlations can be seen in Figure 6.10.

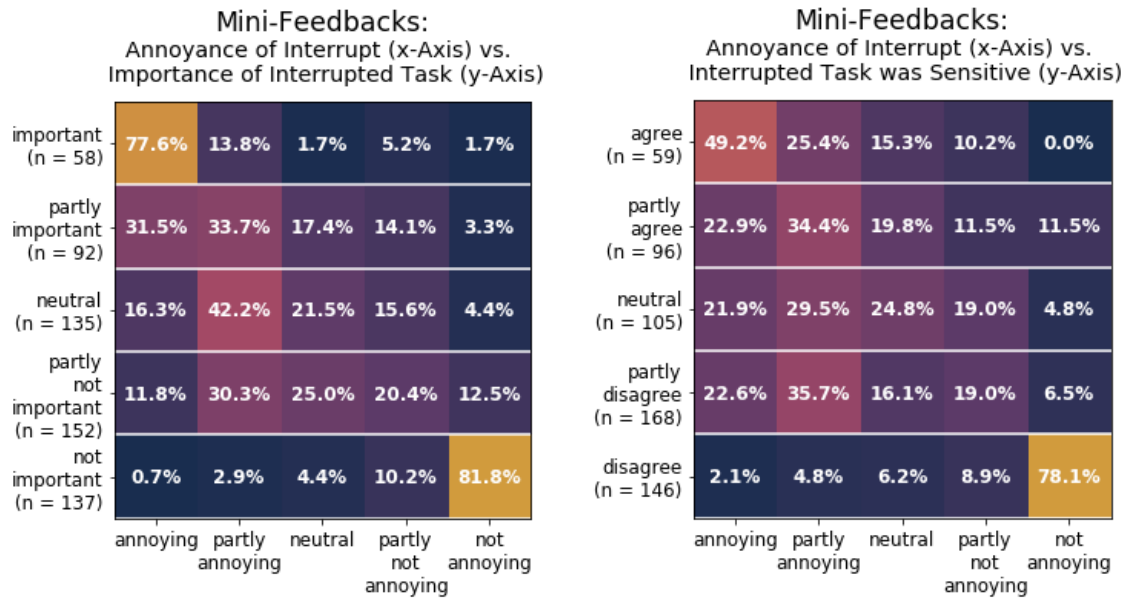


Figure 6.10: The annoyance caused by the interrupt versus the importance or the sensitivity of the interrupted task.

We furthermore investigated the effect of the day of the week on the annoyance by analyzing the Spearman correlations between these two aspects. However, we were not able to find a significant correlation ($r_s = 0.013, p < 0.759$). The effect of the location or interrupted task itself on the annoyance caused by the interrupt is another interesting aspect. To evaluate these correlations, we plotted heat maps showing the distributions of the respective annoyances (Figure 6.11). However, we did not see a clear correlation between annoyance and the current location of the interrupt. This is mostly due to the small number of mentions of the apparently most promising locations. Nevertheless, we did find an interesting trend regarding the reported task “voluntary”: 83.5% of the 79 interruptions when this task was reported were perceived as not annoying. This task was part of the initial list of items, to give the participants the possibility of reporting a voluntary reauthentication specifically. We included this as a voluntary reauthentication does not really interrupt any task. Another possible trend is the high annoyance of interrupts while the participants was chatting. In this regard, 67.6% of the 108 feedbacks while executing this task suggested that the interrupt was annoying (39.8%) or partly annoying (27.8%).

In this regard, it is worth mentioning that not all feedbacks after voluntary reauthentications were reported as such. The quantitative data on the database suggest that of the 185 responses to the feedback screens after voluntary reauthentication, only 79 were reported as such. Especially voluntary reauthentications during the grace period were not reported as such. Thus, only 16 of the 79 reported voluntary reauthentications were executed during the grace period. In this regard, the annoyance caused by the real voluntary reauthentications during the grace period and by those outside the grace period were rated very differently. The voluntary reauthentications outside the grace period ($n = 82$) were perceived as not annoying ($Mdn = 5$), while the reauthentications during the grace period ($n = 103$) were evaluated equally to the forced reauthentications ($n = 383$) as partly annoying ($Mdn = 2$). In this regard, we found a significant difference between the annoyance caused by voluntary reauthentications during and outside the grace period ($\chi^2(1) = 5.44, p = 0.02$) by conducting a Friedman test with Conover’s post hoc tests.

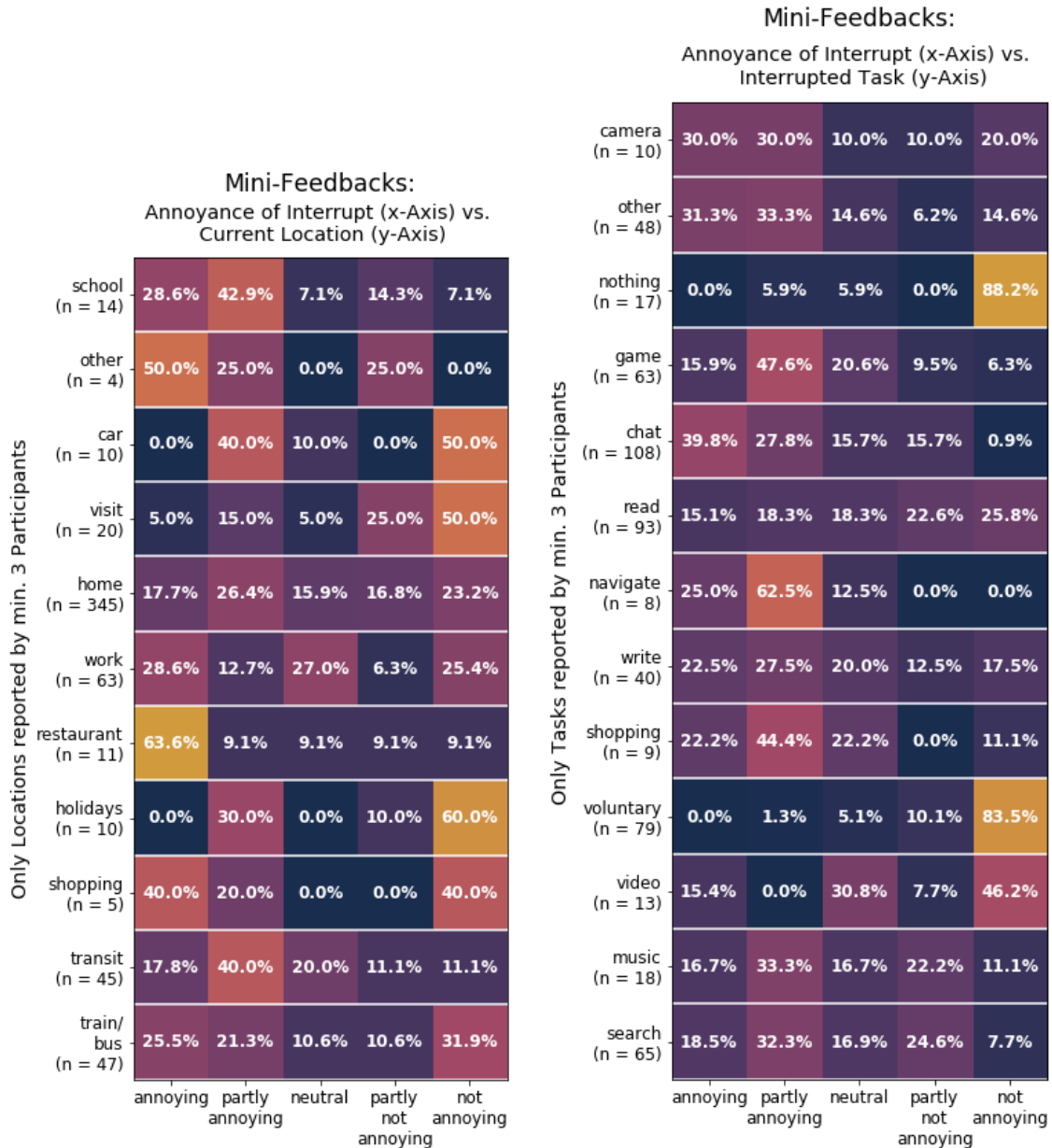


Figure 6.11: The annoyance caused by the interrupt versus the current location or the interrupted task itself. Only locations and tasks that were reported by at least three participants are included to filter personal opinions.

6.4 General Feedback on the User Study

Finally, we also gathered some qualitative feedback on general aspects of the user study and the Authenticator. We achieved this mostly through the last weekly survey, where we included Likert scale feedbacks on a wide range of topics. After finishing the study, we also invited the participants to take part in a semi-structured final interview and four of them agreed. By doing so, we were able to complete the general picture of the users' perceptions.

6.4.1 Results of the Last Weekly Survey

In addition to the above mentioned ranking of the four types of indicators, the last online survey mostly included Likert scale feedbacks, where the participants could choose one of the following five items: 1 - “disagree”, 2 - “partly disagree”, 3 - “neutral”, 4 - “partly agree”, 5 - “agree”. The first section of these feedbacks addressed the users general perception regarding the Authenticator (Figure 6.12). In this regard, most participants liked the design of the application ($Mdn = 5$) and rated the statement of being influenced by the system neutrally ($Mdn = 3$). Furthermore, they felt uninfluenced by bugs of the prototype that they might have observed during the user study ($Mdn = 1$).

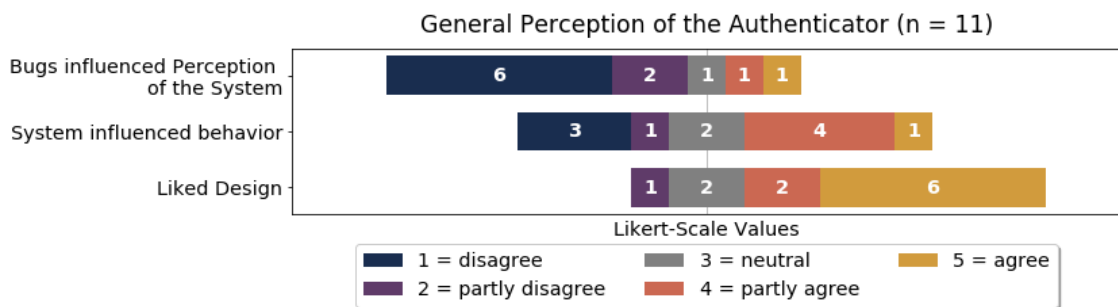


Figure 6.12: Likert scale feedback on the general perception regarding the Authenticator prototype.

The next section of the survey had the in situ feedback screens as a topic. Based on the results of this question, most participants partly disagreed with the mini-feedbacks being annoying ($Mdn = 2$) and with being influenced by them both on the perception of the system ($Mdn = 2$) and on their behavior ($Mdn = 2$). The distribution of these feedbacks can be seen in Figure 6.13.

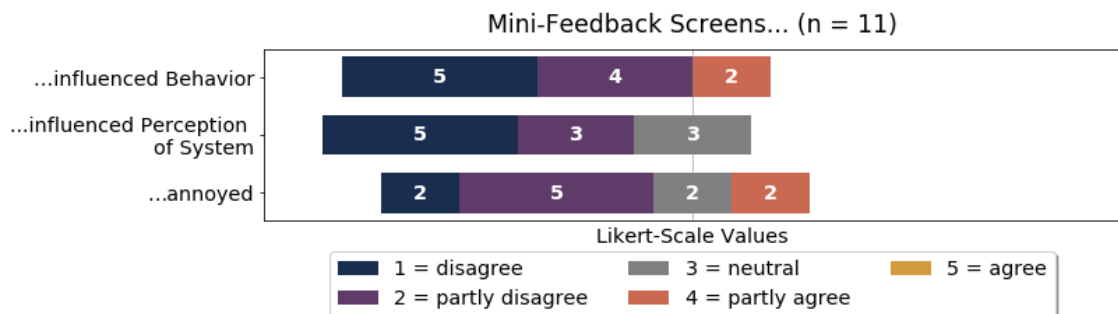


Figure 6.13: Likert scale feedback on the in situ feedback screens.

The other Likert scales of the last weekly online survey dealt with the implicit authentication. Thus, we first asked the participants about their level of knowledge and their sources of information regarding this topic. The responses are illustrated in Figure 6.14. More specifically, we also asked about their usage of the homepage. The participants disagreed with having a profound knowledge on the topic before the user study started ($Mdn = 1$) and 10 of the 11 participants did not review IA on resources other than our homepage ($Mdn = 1$). In this regard, they agreed to having read the complete introduction on the homepage ($Mdn = 5$). Nevertheless, most participants disagreed or partly disagreed with having seen the whole introduction video on the homepage ($Mdn = 2$). The participants also suggested a further usage of the homepage ($Mdn = 4$).

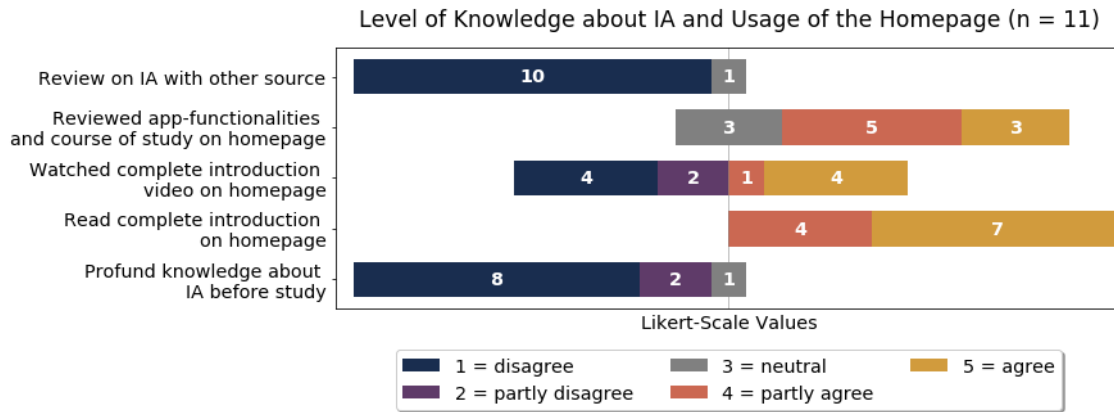


Figure 6.14: Level of knowledge of the participants regarding implicit authentication and feedback on the usage of the homepage.

Regarding the general perception of implicit authentication, the question on a possible future usage of implicit authentication was answered neutrally ($Mdn = 3$). In this regard, one participant who had answered with “partly agree” suggested, in an open text field provided for a final comment, that he would use such a system if he could choose the type of indicator. However, 7 of the 11 participants stated that they felt that IA is more annoying than the traditional approaches, owing to the reauthentication interrupts ($Mdn = 5$), even though 7 of the 11 participants always lock their phones after usage ($Mdn = 5$).

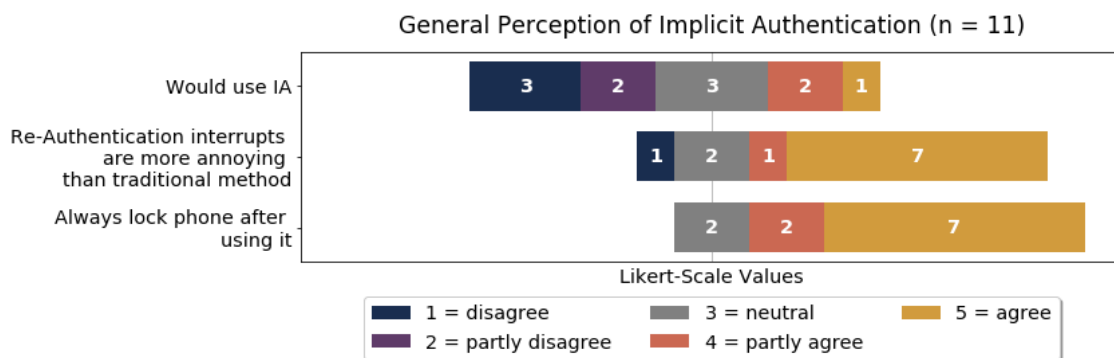


Figure 6.15: The participants’ general perception of implicit authentication.

In addition to the Likert scale feedbacks, the last weekly survey also included a question on the optimal length of the grace period of the ST and SLT conditions. In this regard, the participants were also reminded that this duration on the Authenticator is 8 seconds. The distribution of the values given by the participants ($M = 11.36, \sigma = 4.78$) regarding this question can be seen in Figure 6.16.

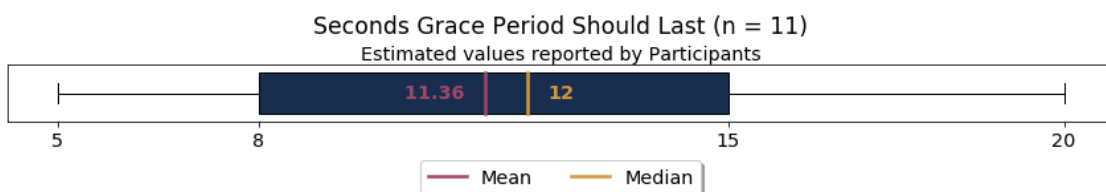


Figure 6.16: Reported values of the ideal duration of the grace period.

At the end of the last online survey, we asked the participants if they would take part in a final semi-structured interview. The participants who agreed had to choose whether they wanted to do the interview personally or via telephone. Furthermore, they had to choose items out of a list of time slots, to indicate the times when they would be available for the interview. We also informed them that participation on the interview would be rewarded with an additional 5 euro Amazon voucher.

6.4.2 Final Interviews

The four participants who were interested in the interview received a confirmation email with the final date and time of their interview. All of them wanted to do the interview via telephone. Even though we excluded one of these four participants from the data analysis owing to missing data sets, we decided to include his interview in this section, as he did not notice any unusual behavior of the application.

The interviews were recorded for later evaluation. The participants were first informed about the recording and had to consent before the interview began. Subsequently, the interview started with a question about their general perception of the user study. They were therefore asked to report spontaneously the first statement they could think of. In this regard, 3 of the 4 participants mentioned that they found the concept of the system interesting but stated that in their opinion the usefulness of an IA system depends a lot on the individual usage. One participant said in this connection:

“I think I would not use such a system, because I always lock my phone. But generally, such a system is a good idea for persons like my girlfriend who always leaves her phone unlocked.”

All participants also reported that the interrupts were at least sometimes annoying.

Subsequently, we asked them about specifically positive or negative moments that occurred in connection with the system. All participants had only negative specific memories and mostly quoted the interruption at a very stressful or important moment as such (3 of 4 participants). The following quote from one participant is an example of this situation:

“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.”

In this regard, two participants also mentioned that the reasons for the DCL behavior were not comprehensible and sometimes caused many reauthentication in a short period of time. However, the participants also used this moment to quote some positive features of the Authenticator that they enjoyed. Thus, one participant suggested the use of the status bar symbol during the LT and SLT phases to avoid forced interrupts by reauthenticating voluntarily, when the DCL was low. Another participant mentioned the feedback screens as a positive feature, as he perceived them as a gamification aspect of the app. He mentioned the grace period as another positive feature, as he used the time to finish his task before being interrupted.

After gaining this first overview of the perception of the system, we now asked the participants specifically about implicit authentication and whether they would use such a system. Two participants stated that they would use such a system if they felt that the security provided is sufficient and the number of reauthentications is not too high. One of them expressed his concern regarding the security, saying:

“I am not sure if only by using my phone, a system could gather enough information to identify me out of 8 billion persons.”

However, the other participant felt more confident in the security provided and a reduced number of reauthentications on a real IA system:

“I think it is convenient when you don’t have to authenticate at the beginning of every time you use your phone. I think, if it really works good, like it should work, you would have to reauthenticate very rarely, because the phone would recognize the owner. Then, it would be very comfortable.”

The other two participants, who stated that they would not use such a system, cited the inconvenience as a reason. Nevertheless, both said that they responded like this owing to personal preferences and that they generally perceive IA to be a useful approach. One person expressed his answer as follows:

“I think it depends a lot on the person. Personally, I never bothered about explicit authentication. So, I perceive the false rejects of implicit authentication as more negative than having to unlock my screen every time I use my phone. But I can imagine that other persons might think different.”

One participant also specifically mentioned that the detection delay would be a reason for not using such a system as the stand alone security mechanism.

In this regard, we also asked the participants how they would evaluate the security provided by an IA system compared with the traditional authentication mechanisms. Two participants rated IA as equally high but stated that it may depend on the implementation. The other two participants thought that the traditional method is more secure. Thus, one participant mentioned:

“I feel more secure, when I do the authentication on my own.”

The next topic of the semi-structured interview was the reauthentication interrupts themselves. All four participants stated, that the interrupts are at least sometimes annoying, but also sometimes neutrally perceived. In this regard, 3 participants suggested that the degree of annoyance depends greatly on the situation. 2 participants also mentioned, that the type of indicator had an effect on the annoyance, as some – especially the SLT condition – decreased it. One participant expressed his thoughts as follows:

“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.”

We then asked the participants if they could imagine some ideas on how to decrease the annoyance. Only three of the four participants had some own ideas. One suggested that the grace period should last longer, while another disliked the dimming out of the screen. As a reason for this statement, he cited the poor visibility of the screen, especially in a bright environment, which made further use of the phone during the grace period impossible. He suggested a less intrusive kind of short-term announcement, such as a light signal or the dimming out of only part of the screen. The third participant, wanted to have the possibility of increasing the DCL at any time, for example with a gesture. However, the same participant had mentioned before that he had never used the voluntary reauthentication as he did not understand its intention.

Subsequently, we asked if the reauthentication interrupts are a reason for implicit authentication not to be used. Two participants stated, that the interrupts indeed are a reason for not using an IA system. One of them mentioned that he may change his mind after a longer phase of adaptation. The other two participants answered that it depends on the frequency of interrupts.

After gathering the opinions on our system and on implicit authentication, we tried to answer some questions that arose during the course of the study. In this regard, we first asked the participants if they observed any faulty behavior of our system. Two participants did not witness

any bugs but one mentioned that the app crashed twice during the user study. The second crash occurred without any kind of notification to the user, and the system was off for several hours. The one participant who was later excluded from the evaluation of the data, observed only the continuous place switching of the status bar symbol with other symbols. This behavior was due to the frequent updating of the I-Notification as it included the current DCL.

We also asked the participants if they normally answered the appearing feedback screens directly. In this regard, two participants always answered the feedback screen directly and one other participant answered them mostly directly. The fourth participant delayed the feedback screens depending on the situation.

The next question addressed the design and implementation of the Authenticator. All participants liked the design of the app and three of them specifically mentioned liking the shield symbol on the status bar during the LT and SLT conditions. Three participants felt that the application included all necessary information, while the fourth participant suggested integrating the information on the homepage and the online surveys inside the app. In this regard, we also asked about the homepage and its implementation. All participants liked the design of the homepage and three of them felt that there was no important information missing. The fourth participant missed some data on the technical details of the application and the algorithm of the Authenticator. Regarding usage, all participants used the homepage at the beginning and two of them used it again during the study.

In summary, the opinions gathered during the semi-structured interview completed the picture on the users' opinion of our system, implicit authentication and the reauthentication interrupts. By combining these findings with all our qualitative and quantitative results, we were able to interpret conclusions on the usability of implicit authentication and reauthentication interrupts.

7 Discussion

As mentioned in Section 1.1, the hypothesis that we want to investigate in this thesis is that users' annoyance caused by unexpected reauthentication interrupts can be reduced by using indicators that implement the following aspects:

- feedback on the recent system status illustrating the recent device confidence level,
- announcement of the reauthentication interrupt,
- the possibility of reauthenticating voluntarily at any time, to avoid forced reauthentication interruptions.

In this regard, the most important point that we need to discuss is the effect of the indicators that we implemented. Nevertheless, the influence of other factors is not negligible and therefore constitutes another crucial topic. These factors can be environmental or be caused by the design of our indicators.

7.1 Effect of Indicators

The results discussed in Section 6.2.3 suggest that the type of indicator has no significant effect on smartphone usage. We were also not able to find a significant effect on the perceived importance and sensitivity of the task interrupted by reauthentications.

Furthermore, the tests that we performed with the data collected through the feedback screens also confirmed no significant effect of the indicator on the annoyance caused by the interrupt (see Section 6.2.4). Nevertheless, we detected a significant effect on the feedback regarding the general annoyance caused by the system, which we collected through the weekly surveys (see Section 6.2.1). These results give evidence that the LT is less annoying than the NO. The LT and SLT were also rated as more rewarding after a reauthentication due to the increased DCL.

When discussing these findings it is essential to keep in mind that most bugs of the system were observed during the SLT phase of the study and, therefore, the perception of this condition could have been influenced negatively. Moreover, the annoyance caused by the system during the SLT phase may have been rated higher owing to this negative influence, as the small number of reported disliked features of the SLT ($n = 2$) would lead to contradictory conclusions regarding the caused annoyance.

In this regard, the final ranking of the four indicators could represent another strong clue to the negative influence of the observed bugs, as the SLT was ranked by 63.6% of the participants as the most likely to be used (see Section 6.2.2). Thus, after completing the 4-week field study the participants preferred the SLT over the other conditions even though it was not reported as significantly less annoying than the other conditions in the weekly surveys. These contradictions could lead to the conclusion that the rating of the annoyance caused by the system during the SLT phase has been influenced negatively owing to the higher number of observed bugs. Nevertheless, this conclusion is unassured as most participants did not feel consciously influenced by the system failures.

The LT, which had the above mentioned positive effect on the general annoyance caused by the system was rated as second most likely to be used by 7 of the 11 participants. The same number of participants ranked the ST third. In this regard, it is also worth noting that all three conditions that include any kind of indication were placed higher than the No Indicator condition on this ranking by 81.8% of the participants (9 of the 11 participants). This clear preference of the indicators over the baseline condition without any indication is probably based on users' need to have a clue about an imminent interrupt, as all participants disliked the missing announcement of the NO in

the weekly surveys. This desire was also deduced by Agarwal et al. [1] and McFarlane et al. [22].

Hence, even though we are not able to verify a positive effect of all indicators on the annoyance caused by a IA system with reauthentication interrupts, we can prove a preference of our indicators over the baseline condition.

7.2 Reasons for Annoyance

In addition to the effects that our indicators have on the annoyance, we are also able to discriminate other aspects that might influence it. The results presented in Section 6.3 suggest that the importance and sensitivity of the interrupted task have an impact on the annoyance caused by the interrupt. On the one hand, the more important a task was, the more annoying was the interruption of it. The participants in the final interviews emphasized this statement, as 3 of the 4 participants reported the interruption of important or stressful tasks as particularly negative events (Section 6.4.2).

On the other hand, the interruption of a non-sensitive task was rated as not annoying. In this regard, we expected a different outcome, as the reauthentication increases the security of the system, and therefore might have been more tolerated during sensitive tasks. However, the interruption of a non-sensitive task could have been more annoying, as the increased security is not that necessary.

It is also important to keep in mind that during the study performed in this thesis, the participants had to reauthenticate after an interruption in addition to the normal unlocking of the phone at the beginning of each usage session. As the relevant literature often proposes IA as a method to provide security to users who currently use no EA system [18, 32], the increased number of authentications of our system might have increased the perceived annoyance. However, many investigations on IA also suggest its usage as a second layer of protection due to false accepts and the detection delay [12, 21, 32]. Consequently, it is not clear if future IA systems might increase or decrease the number of authentications. Therefore, as far as we understand the behavior of the Authenticator is validated.

In comparison with the 68.3 daily unlocks, the number of reauthentication interrupts was only 5.6. However, 8 of the 11 participants agreed or partly agreed with the statement that the less frequent reauthentication interrupts are more annoying than the normal unlocks at the beginning of each usage session. This statement might have been influenced by the wrong estimation of daily unlocks that the participants reported in the start survey. Based on these estimations and the data gathered on the database, most participants drastically underestimated the number of unlocks they perform daily (9 of 11 participants) by 5 to 95 daily unlocks. Thus, on average, the estimated number of daily reauthentications was only 59.2% of the real ones. Figure 7.1 shows a direct comparison of the estimated and real daily unlocks. This aspect reflects that the users' opinion regarding annoyance could also be influenced by false self-assessments or habituation effects.

In summary, the level of annoyance caused by reauthentication interrupts seems to depend on a wide range of external and internal influences. Hence, further research on this topic might be necessary to achieve a greater understanding of users' perception regarding IA.

7.3 Voluntary Reauthentications

Based on the results in this thesis, real voluntary reauthentications outside the grace period were perceived as not annoying in 67.1% of the collected feedbacks (see Section 6.3). This conclusion

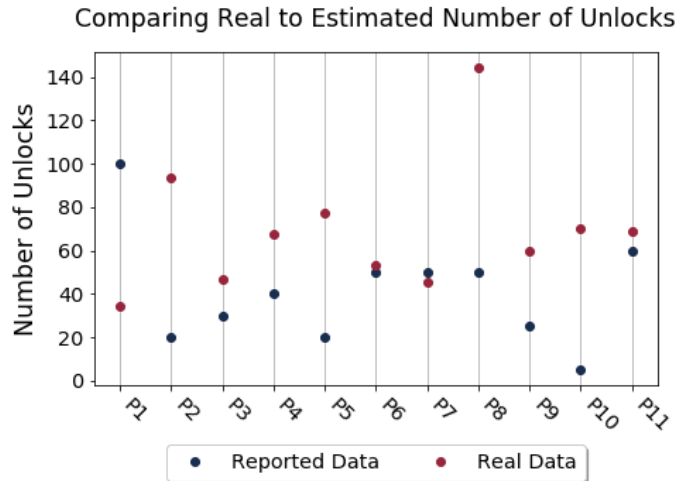


Figure 7.1: Comparison of the estimated number of daily unlocks reported by the participants with the real average numbers.

corresponds to the observations made by McFarlane et al. [22]. Hence, motivating the user to reauthenticate voluntarily to avoid forced interrupts could be a key feature to improve the usability of IA.

Even though the participants reported no significant effect of our indicators on their motivation (Section 6.2.1), the quantitative results in Section 6.2.3 indicate a significantly higher relative number of voluntary reauthentications on all phases of the user study, where an indicator was shown (Table 6.3). The LT also achieved an increase on the relative number of daily voluntary reauthentications outside the grace period.

Furthermore, we observed an influence on the moment of voluntary reauthentications probably due to the popping up of the I-Notification. During the ST phase of the study, most voluntary reauthentications occurred during the grace period (82.8%). With the SLT condition, nearly half of them were executed during this period (42.3%). However, as the voluntary reauthentications that occurred during this period were perceived as partly annoying, this effect does not reduce the general annoyance.

Thus, increasing users' motivation to reauthenticate voluntarily could be a key feature for reducing their annoyance. Moreover, it seems that the user really has to be able to choose the moment of the voluntary reauthentication, as those executed during the grace period are still rated as partly annoying. Another possible reason for this effect is that voluntary reauthentications after the beginning of the grace period of the ST and SLT phases may have been not perceived as executed voluntarily, as the screen started to dim out. This deductions could also explain why the only indicator that was rated significantly better in terms of annoyance in the weekly surveys was the LT (Section 6.2.1). As mentioned above, the LT was also the only condition that increased significantly the relative value of voluntary reauthentications outside the grace period. Therefore, we recommend further investigation on this topic.

However, the perception of voluntary reauthentications during the grace period could have been influenced by disliked features of the short-term announcement, such as the duration of the grace period and the dimming out of the screen.

7.4 Grace Period

As mentioned during the final interview (Section 6.4.2), the dimming out of the screen during the grace period degraded the visibility of the content on the screen. Therefore, the participants could not use the whole 8 seconds of the grace period to finish their current task. Nevertheless, the dimming out in connection with the grace period was mentioned six times as a positive feature during the weekly surveys in Section 6.2.1.

In addition, once the dimming out of the screen had passed a certain amount of darkness, the participants had to wait for the end of the grace period to be able to reauthenticate and continue with their task, as triggering a voluntary reauthentication without seeing the I-Notification would be difficult. This threshold depends on, among other things, the screen's brightness settings and the light conditions of the environment. The probably resulting decrease in voluntary reauthentications in the last phase of the grace period can be seen in Figure 7.2. Furthermore, the less fast reauthentication was criticized three times during the weekly surveys.

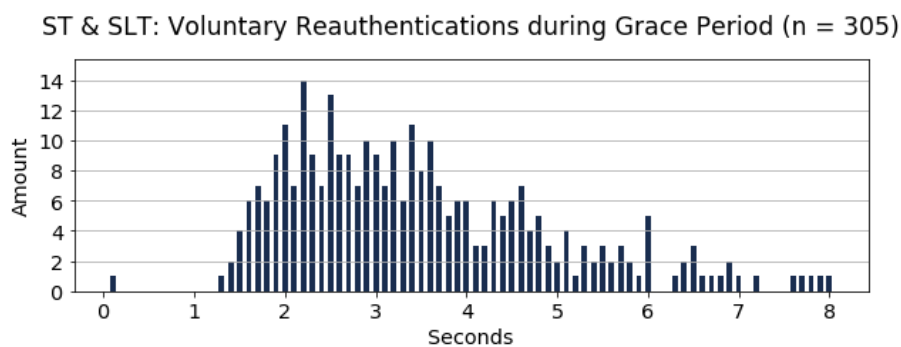


Figure 7.2: Number of voluntary reauthentications during the grace period. The timestamps were grouped in intervals of 0.1 seconds.

Hence, our chosen design of the short-term indicator could have influenced the participants' perception. In this regard, another criticized feature of the short-term announcement of an imminent interrupt was the duration of the grace period. We asked the users to report its ideal duration in the last weekly survey and 8 of the 11 participants wanted to have more time than the preset 8 seconds (Section 6.4.1). On average, they desired a period of $M = 11.36$ seconds ($\sigma = 4.78$). We compared the participants' estimations with the real average amount of time the grace period lasted after being started in Figure 7.3. However, as a reauthentication was forced after 8 seconds, this analysis can only be seen as a hint on the ideal duration of the grace period.

Only 1 of the 3 participants who preferred a shorted grace period did not use 7-8 seconds on average. However, 3 of the 4 participants who desired a grace period of 15 and more seconds used only up to 6 seconds on average. The remaining 4 participants preferred a duration of 10-12 seconds and used at least 6 seconds of the given grace period. Thus, that these participants desire a few more seconds of time seems to be confirmed by the real average values. Therefore, increasing the duration of the grace period to 10-12 seconds could be a good approach to fulfill the users needs. However, as Agarwal et al. [1] had already suggested a desire of the user to configure certain features by themselves, a configurable setting of this value might be the best solution.

In summary, concerning the above mentioned hypotheses of this thesis, we found a significant positive effect of the Long-Term Indicator on the overall annoyance caused by the implicit authentication system. Thus, we were partly able to reduce this usability issue, as postulated by the hypothesis. We furthermore identified that voluntary reauthentications, that are performed without an additional feature pressuring the user – in our case outside the grace period – are

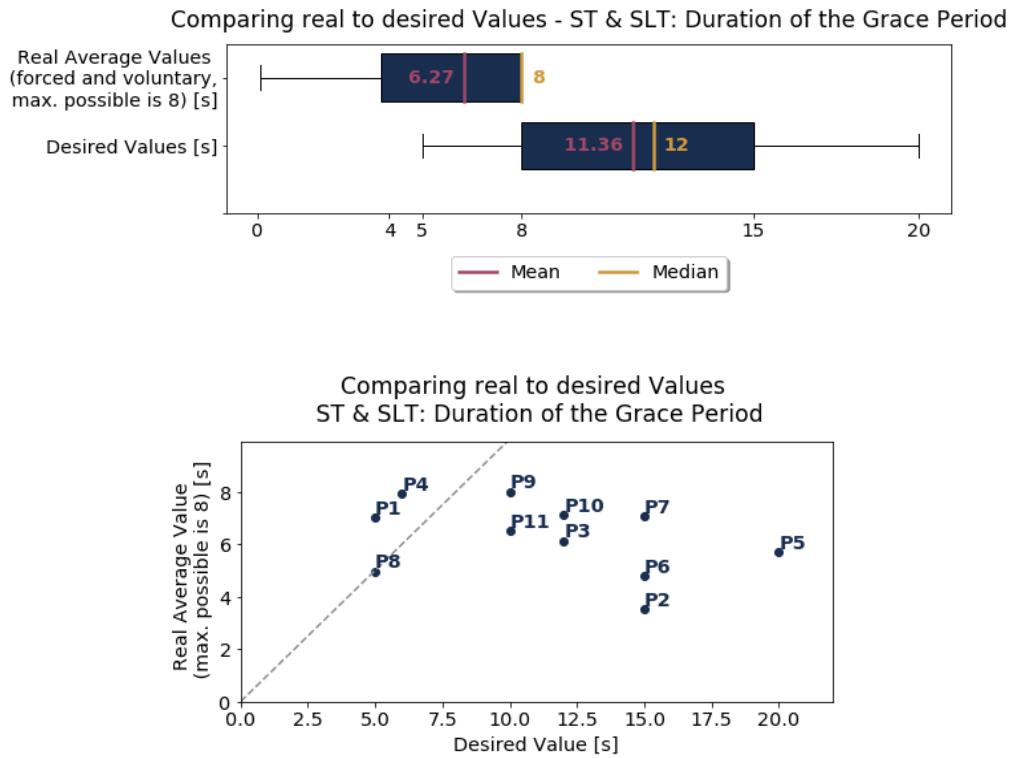


Figure 7.3: Real versus desired durations of the grace period. The real values are restricted by a forced reauthentication that occurred after 8 seconds.

perceived as not annoying. Hence, the possibility for voluntary reauthentication we defined in the hypothesis had also a positive effect on the annoyance. As a final conclusion, we did partly verify the validity of the claimed hypothesis, even though the positive effect of the Short-Term Indicator was not conclusively evidenced.

8 Conclusions

The concept of implicit authentication aims at increasing the security and reducing the number of necessary explicit authentications [12, 18, 21, 32], although, this method presumably causes mid-task reauthentication interrupts [11, 18, 21], which have been shown to raise new usability issues [19]. This thesis has analyzed the consequences that these interrupts have for the user while focusing on the emotional aspects, especially the resulting annoyance. In this regard, we conducted investigations on improving the inconvenient perception of such IA systems by (a) showing the recent state of the device confidence level (DCL) at any time, (b) announcing an imminent interruption and (c) enabling the user to reauthenticate voluntarily at any time.

We discussed these three basic aspects in a focus group session and then developed our concept in collaboration with the participants' opinions. On this basis, we generated our application prototype called Authenticator, which includes four conditions: (a) the No-Indicator state (NO), where no indication of either the current DCL or an upcoming interruption is shown, (b) the short-term announcement of an imminent reauthentication (ST) through a notification and a gradual dimming out of the screen, (c) the long-term visualization of the current DCL at any time via a symbol in the status bar (LT) and (d) the combination of ST and LT, where both the current DCL and an imminent interruption are indicated (SLT).

This prototype was subsequently evaluated in a 4-week field study with 11 participants. Based on the results of this study, we could prove a positive effect of the LT on users' annoyance and a greater number of voluntary reauthentications during this phase of the study. Furthermore, we were able to verify a low level of annoyance caused by these voluntary reauthentications. Therefore, we reason that increasing the motivation of the user to reauthenticate voluntarily might be a key feature in improving the usability of IA. However, the results of this thesis also revealed some aspects where further investigations are needed to validate our conclusions and results.

Unfortunately, the small number of participants in our user study ($n = 11$) restricts the scalability of our results to the entirety of the population. However, the age range of the participants does represent a important percentage of today's smartphone users, as 10 of the 11 participants were between 20 and 33 years of age. Based on statistical data provided by Statista.com [36] and PopulationPyramid.net [35], roughly 48% of German smartphone users in 2016 were under 35 years of age. Thus, our participants represent an important target group. Nevertheless, the observed crashes of our application and missing data sets of some participants represent additional limitations to our results. Hence, the results on the effect of our indicators on usability should be confirmed by further research and a user study with a larger number of participants.

In this regard, the modification of some design aspects of our indicators should also be evaluated. For example the status bar symbol was mostly liked by the users but also reported as easily overseen, and the dimming out of the screen degraded the visibility of the content and thus restricted the usage of the grace period. An additional possibility for configuring certain settings, such as the duration of the grace period or the turning off of the vibration of the I-Notification, could be well received by users as already suggested by Agarwal et al. [1].

Furthermore, the impact of other aspects, such as environmental influences and the importance and sensitivity of the interrupted task, is another important topic to bear in mind. As this aspect was not the main topic of our investigation, we gathered only restricted information. In this regard, a combination of implicit authentication with context-aware or app-based authentication could be a promising approach. The possible benefits of these combinations were already mentioned by the participants of our focus group and in the related work [12].

Our results on the positive perception of voluntary reauthentications suggest that motivating the user might be a key feature in reducing annoyance. Hence, enhancement of this motivation is a desirable goal when implementing an IA system. Gamification features such as a reward system could improve the positive user experience and motivation.

Another important topic that needs further investigation is the security implications that our indicators could have. As also reported by Khan et al. [19], both the long- and short-term indications on the state of our systems might be used by an unauthorized user as a cue to the remaining time they have before being locked out of the device. The visualization of the current DCL could even be used to improve the efficiency of mimicry attacks. Hence, the effects on intruder detection of the IA system are currently not predictable.

In summary, even though only 3 of the 11 participants in our user study stated that they have an interest in using an IA system in the future, other studies on this topic found possible future adoption rates of 63% and more [6, 19]. Hence, we conclude that although further investigations on usability are needed, IA is a promising concept for future authentication of smartphones, especially in combination with app- or context-aware methods.

Acronyms

DCL Device Confidence Level	16, 18–21, 23, 27–29, 31, 33–35, 37, 38, 44–47, 53, 54, 58
EA Explicit Authentication	9, 10, 48
EER Equal Error Rate	6, 7
I-Notification Indicator-Notification	21–23, 36, 37, 46, 49, 50, 53
IA Implicit Authentication	ii, 1, 9, 10, 19, 42–45, 48, 49, 53, 54, 58
LT Long-Term Indicator	17, 21, 22, 35, 37, 38, 51, 53
NO No Indicator	17, 22, 33, 35, 37, 38, 47, 53
NO-Notification No-Indicator-Notification	22, 23
SLT Short and Long-Term Indicator	17, 28, 36–38, 53
ST Short-Term Indicator	17, 22, 28, 35–38, 51, 53

Glossary

Behavioral Biometrics

Behavioral patterns that can be used to identify a person. Some examples are voice or gait recognition. 1, 6, 8

Continuous Authentication

See Implicit Authentication. 1

Device Confidence Level

Current level of the authentication confidence. Represents the level of confidence that the system has in the current user's identity. 2, 10, 14, 19, 20, 47, 57

Equal Error Rate

An implicit system can be configured to achieve that the false accept rate and false reject rate are the same. The resulting rate is called the equal error rate. 6, 7, 29

Explicit Authentication

Authentication by explicitly passing through a authentication mechanism such as PIN, password or pattern lock. This includes biometric methods such as fingerprint or face recognition. 1

False Accept

When an implicit authentication falsely accepts an unauthorized user. The number of false accepts is used to measure the level of functionality that an implicit authentication system achieves. 6, 56

False Reject

When an implicit authentication falsely rejects the device owner because it does not recognize the behavior. The number of false rejects is used to measure the level of functionality that an implicit authentication system achieves. 6, 56

Grace Period

Time that passes from the short-term announcement of an imminent reauthentication to the interruption. 11, 12, 14, 22, 23, 28, 33–35, 37, 38, 40, 43–45, 48–51, 53

Implicit Authentication

Continuous and transparent authentication based on usage patterns. 1, 10, 13

Operating Threshold

If the device confidence level falls below this value, a reauthentication is triggered. 10, 14, 19, 34

Physiological Biometrics

Biometric features that can be used to identify a user. Some prominent methods are fingerprint and face recognition. 10

Transparent Authentication

See Implicit Authentication. 1

Appendices

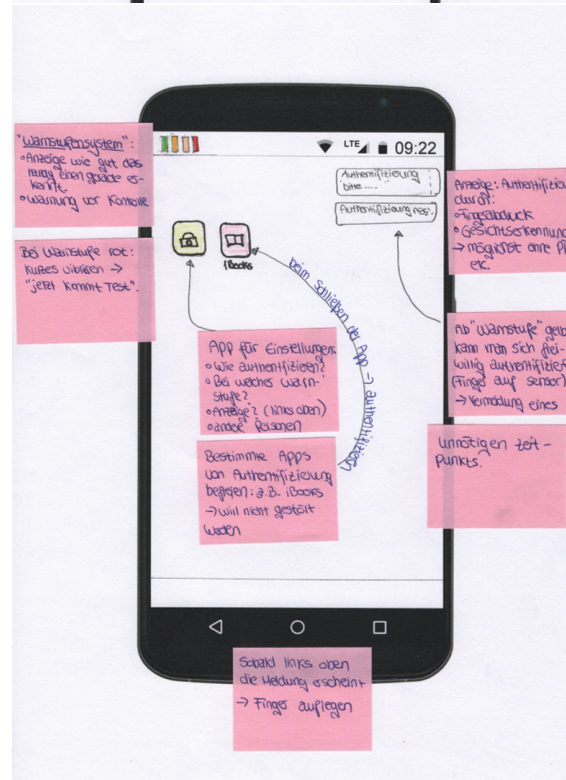
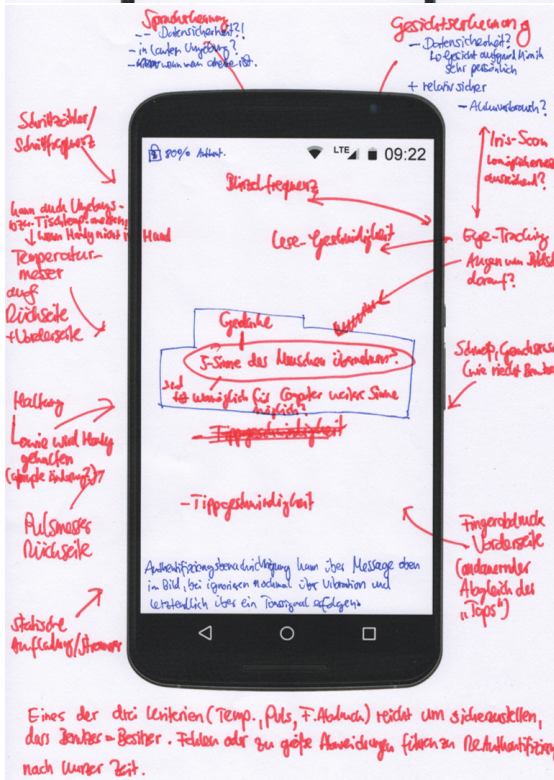
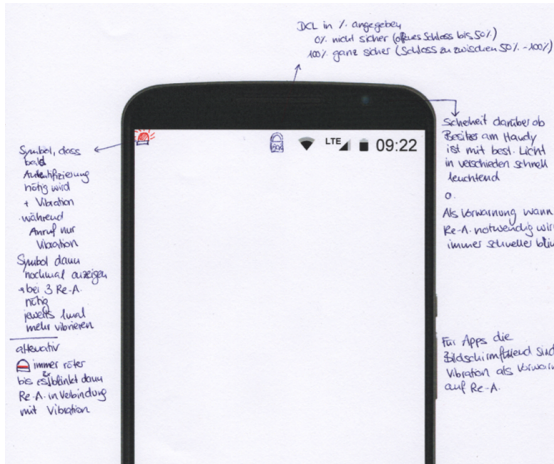
A Simplified Design Space for IA Reauthentication Indicators

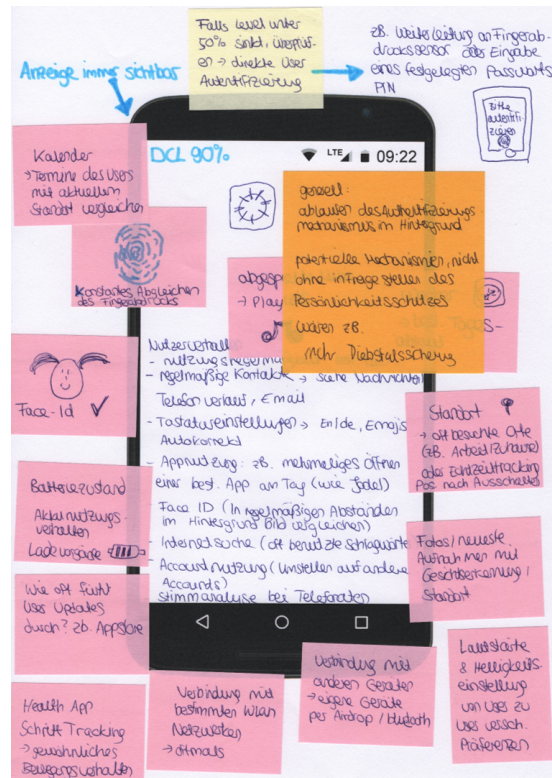
Examples of indicators for sudden events and gradually changing information on today's smart-phones

	Sudden Event	Gradually Changing Information	
Indicated Information	Reauthentication necessary	Device Confidence Level (DCL)	
Indication Technique		Gradual Value	Divided in Units
Graphical Indication e.g. Symbol, Figure, (Section of) Screen, Text	- Bright/Dark Flash - Emergence - Animation - Color - Shape (Text: Content) ...	- Color - Saturation - Luminosity - Blurring - Animation	- Countdown - Shape (Text: Content) - Color - Animation
Other Visual Feedback	- Flashlight On/Off - Notification
Haptic or Audible Feedback	- Vibration - Sound ...	- In-/Decrease Volume ...	- Vibration Pattern ...

Table A.1: Defining examples of possible indications for sudden events and gradually changing information.

B Indicator Design Ideas Developed by the Participants of the Focus Group



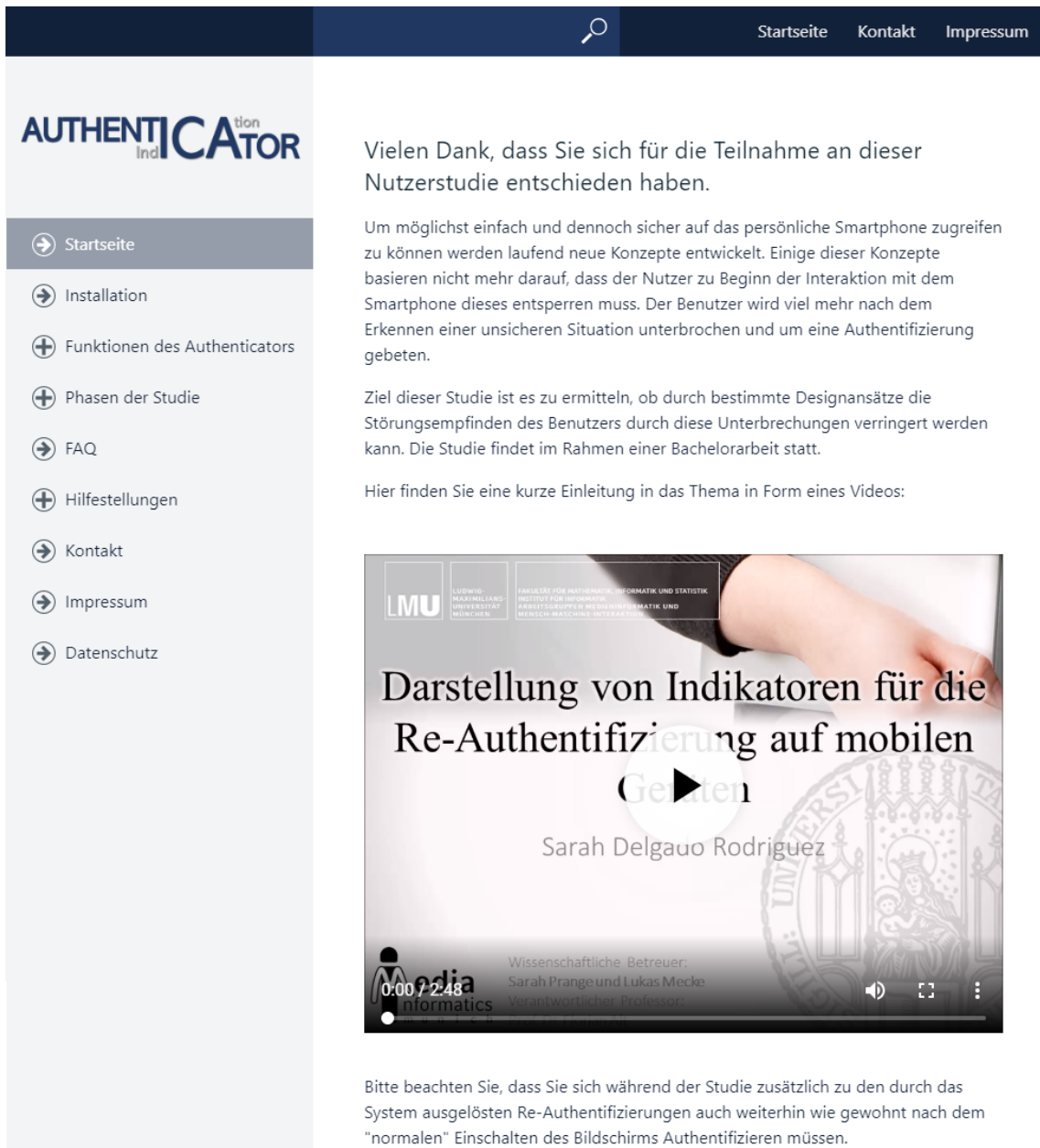


C Google Material Design Icons

The shield and neutral X symbol used for the notifications of the Authenticator are based on Google Material Icons [34]. The icons shown on the feedback screens of our application were achieved from the same source. All of them are available under the Apache License Version 2.0. [33]. In this regard, we changed the color and form of some icons.

D Selected Screenshots of the Homepage

D.1 Welcome Page



Vielen Dank, dass Sie sich für die Teilnahme an dieser Nutzerstudie entschieden haben.

Um möglichst einfach und dennoch sicher auf das persönliche Smartphone zugreifen zu können werden laufend neue Konzepte entwickelt. Einige dieser Konzepte basieren nicht mehr darauf, dass der Nutzer zu Beginn der Interaktion mit dem Smartphone dieses entsperren muss. Der Benutzer wird viel mehr nach dem Erkennen einer unsicheren Situation unterbrochen und um eine Authentifizierung gebeten.

Ziel dieser Studie ist es zu ermitteln, ob durch bestimmte Designansätze die Störungsempfinden des Benutzers durch diese Unterbrechungen verringert werden kann. Die Studie findet im Rahmen einer Bachelorarbeit statt.

Hier finden Sie eine kurze Einleitung in das Thema in Form eines Videos:

Darstellung von Indikatoren für die Re-Authentifizierung auf mobilen Geräten
Sarah Delgao Rodriguez

Wissenschaftliche Betreuer:
Sarah Prange und Lukas Mecke
Verantwortlicher Professor:

0:00 / 2:48

Bitte beachten Sie, dass Sie sich während der Studie zusätzlich zu den durch das System ausgelösten Re-Authentifizierungen auch weiterhin wie gewohnt nach dem "normalen" Einschalten des Bildschirms Authentifizieren müssen.

D.2 Installation Tutorial

[Startseite](#) [Kontakt](#) [Impressum](#)



- [→ Startseite](#)
- [→ Installation](#)
- [+ Funktionen des Authenticators](#)
- [+ Phasen der Studie](#)
- [→ FAQ](#)
- [+ Hilfestellungen](#)
- [→ Kontakt](#)
- [→ Impressum](#)
- [→ Datenschutz](#)

Installation

Nehmen Sie für die Installation der Authenticator-App bitte ihr Smartphone zur Hand. Führen Sie die nächsten Schritte bitte auf ihrem Smartphone aus.

- Klicken Sie bitte zuerst auf den Play Store Link, den Sie per E-Mail erhalten haben. Es öffnen sich Informationen dazu, dass die App derzeit eine Testversion ist. Scrollen Sie hier bitte einfach runter und drücken Sie auf den Knopf "AM INTERNEN TESTPROGRAMM TEILNEHMEN"


- Jetzt werden Sie am internen Testprogramm willkommen geheißen. Drücken Sie hier bitte auf den blau unterlegten Link "bei Google Play herunterladen."


- Nun öffnet sich die Seite des Authenticators im Google Play Store. Drücken Sie hier bitte auf "INSTALLIEREN". Die App wird nun heruntergeladen und installiert.



E Online Surveys

E.1 User Study Registration

Anmeldung Studie “Evaluation von Indikatoren für die Re-Authentifizierung auf mobilen Geräten mit Hilfe des Authenticator-Prototyps”

Die Nutzerstudie dient zu der Evaluation verschiedener Designansätze, die in einem App-Prototyp, dem sogenannten “Authenticator” umgesetzt wurden.

Herzlich Willkommen!

Wir freuen uns sehr, dass Sie sich für die Teilnahme an dieser Studie interessieren. Nachdem Sie sich angemeldet haben, werden wir uns bei Ihnen per E-Mail melden, um Sie über das weitere Vorgehen zu informieren.

Diese Umfrage enthält 4 Fragen.

Teilnehmer Anmeldung

Die Nutzerstudie dient zu der Evaluation verschiedener Designansätze, die in einem App-Prototyp, dem sogenannten „Authenticator“ umgesetzt wurden.

Herzlich Willkommen!

Wir freuen uns sehr, dass Sie sich für die Teilnahme an dieser Studie interessieren. Nachdem Sie sich angemeldet haben, werden wir uns bei Ihnen per E-Mail melden, um Sie über das weitere Vorgehen zu informieren.

Diese Umfrage enthält 4 Fragen.

Geben Sie zuerst bitte ihre Gogglemail Adresse an, die Sie für ihr Android-Smartphone benutzen. Diese wird sowohl für die weitere Kommunikation mit Ihnen benutzt, als auch dafür benötigt, Sie später für die Benutzung der App freischalten zu können.

Bitte geben Sie Ihre Antwort hier ein: [...]

Welche Android Version haben Sie auf ihrem Smartphone installiert?

Bitte geben Sie Ihre Antwort hier ein: Android [...]

Welches Smartphone-Modell benutzen Sie?

Bitte geben Sie Ihre Antwort hier ein: [...]

Um meinen Smartphone-Bildschirm wieder benutzen zu können, nachdem ich ihn zuvor gesperrt habe, muss ich...

Bitte wählen Sie zwischen 1 und 2 Antworten aus. Bitte wählen Sie alle zutreffenden Antworten aus:

- nichts tun. Ich kann mein Smartphone direkt benutzen.
- meine PIN eingeben.

- mein Passwort eingeben.
- mein Muster zeichnen.
- meinen Fingerabdruck scannen. (Markieren Sie auch die Alternative, die Sie nutzen, wenn z.B. Ihr Fingerabdruck nicht erkannt wird.)
- mein Gesicht erkennen lassen. (Markieren Sie auch die Alternative, die Sie nutzen, wenn z.B. Ihr Gesicht nicht erkannt wird.)
- Sonstiges:

Wenn Sie z.B. ihren Fingerabdruck, oder die Gesichtserkennung benutzen, um Ihr Smartphone zu entsperren, haben Sie auch eine andere Möglichkeit den Bildschirm zu entsperren, wenn z.B. der Fingerabdruck nicht erkannt wird. Wählen Sie bitte auch diesen zweiten Mechanismus mit aus. z.B. Fingerabdruck und Muster

Vielen Dank für Ihr Interesse!

Wir werden in Kürze mit Ihnen per E-Mail Kontakt aufnehmen. Vergessen Sie bitte nicht regelmäßig ihren Gmail-Posteingang und Spam-Ordner zu kontrollieren. Für Anregungen oder Rückfragen wenden Sie sich an S.Delgado@campus.lmu.de.

E.2 Start Survey

Demographische Daten und Smartphone Nutzung

Hier werden ein paar Fragen zu demographischen Daten (z.B. Alter, Geschlecht...) gestellt. Außerdem finden Sie einige grundsätzliche Fragen zu ihrer Smartphone Nutzung.

Herzlich Willkommen zu der ersten Umfrage im Rahmen dieser Nutzerstudie!

Sie wird nur wenige Minuten ihrer Zeit in Anspruch nehmen. Diese Umfrage enthält 10 Fragen.

E.2.1 Demographische Daten

Geben Sie hier bitte Ihre UserID ein:

Bitte geben Sie Ihre Antwort hier ein: Meine ID: [...]

Die UserID finden Sie in der Authenticator-App unten rechts (ID..). Eine ausführlichere Beschreibung gibt es auf der Homepage.

Welches Geschlecht haben Sie?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- männlich
- weiblich
- Sonstiges

Wie alt sind Sie?

Bitte geben Sie Ihre Antwort hier ein: [...]

Welchen Beruf bzw. welche Tätigkeit führen Sie derzeit aus?

Wenn Sie ‘Sonstiges:’ auswählen, spezifizieren Sie bitte Ihre Auswahl im entsprechenden Textfeld. Bitte wählen Sie nur eine der folgenden Antworten aus:

- Arbeitslos
- Schüler
- Student
- Rentner(in) / Pensionär(in)
- Erwerbs- / Berufstätig
- Sonstiges

Ich würde meinen Wissensstand im Umgang mit Computern, Smartphones etc. als ... einschätzen.

Bitte wählen Sie nur eine der folgenden Antworten aus:

- sehr hoch
- hoch
- mittel
- niedrig
- sehr niedrig

Geben Sie an, wie gut Sie sich mit technischen Geräten, wie PCs oder Smartphones auskennen. Beachten Sie bitte: ein “sehr niedriger”-Wissensstand bedeutet, dass Sie sich sehr wenig damit auskennen

E.2.2 Smartphone Nutzung

Um meinen Smartphone-Bildschirm wieder entsperren zu können, nachdem ich ihn zuvor gesperrt habe, muss ich...

Bitte wählen Sie zwischen 1 und 2 Antworten aus. Bitte wählen Sie alle zutreffenden Antworten aus:

- nichts tun. Ich kann mein Smartphone direkt benutzen.
- meine PIN eingeben.
- mein Passwort eingeben.
- mein Muster zeichnen.
- meinen Fingerabdruck scannen. (Markieren Sie auch die Alternative, die Sie nutzen, z.B. wenn Ihr Fingerabdruck nicht erkannt wird.)
- mein Gesicht erkennen lassen. (Markieren Sie auch die Alternative, die Sie nutzen, z.B. wenn Ihr Gesicht nicht erkannt wird.)
- Sonstiges:

Wenn Sie z.B. ihren Fingerabdruck, oder die Gesichtserkennung benutzen, um Ihr Smartphone zu entsperren, haben Sie auch einen andere Möglichkeit den Bildschirm zu entsperren, wenn z.B. der Fingerabdruck nicht erkannt wird. Wählen Sie bitte auch diesen zweiten Mechanismus mit aus. z.B. Fingerabdruck und Muster

Darum benutze ich meine (oben ausgewählte) Methode um mein Smartphone zu entsperren, und nicht eine der anderen Methoden:

Bitte geben Sie Ihre Antwort hier ein: [...]

Geben Sie hier Gründe dafür an, weswegen Sie diese spezielle Entsperr-Methode, die sie ja bereits in der vorherigen Frage angekreuzt haben, benutzen.

Ich denke ich entsperre meinen Smartphone-Bildschirm täglich ungefähr ... mal.

Bitte geben Sie Ihre Antwort hier ein: [...]

Schätzen Sie, wie oft Sie ihren Smartphone-Bildschirm am Tag entsperren. Geben Sie die geschätzte Anzahl an Entsperrungen in das Eingabefeld ein.

Ich denke ich nutze mein Smartphone täglich ungefähr ... Stunden lang.

Bitte geben Sie Ihre Antwort hier ein: [...]

Schätzen Sie, wieviele Stunden täglich Sie ihr Smartphone circa benutzen. Geben Sie die geschätzte Zeit in Stunden in das Eingabefeld ein. Sie können auch Zahlen mit Komma eingeben. (z.B. 1,6)

Ich finde es wichtig mein Smartphone vor dem Zugriff von unbefugten Personen schützen zu können.

Bitte wählen Sie nur eine der folgenden Antworten aus:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu
- trifft nicht zu

Geben Sie an, wie sehr Sie der Aussage zustimmen. Ist Ihnen dieser Schutz wichtig?

Vielen Dank für Ihre Unterstützung!

Sie werden eine E-Mail mit der Anleitung zur Installation der App bekommen, wenn die Studie beginnt.

E.3 Weekly Surveys

Fragebogen erste Woche

Am Ende jeder Woche sammeln wir mit Hilfe dieser Fragebögen die Eindrücke, Empfindungen und Meinungen der Teilnehmer.

Herzlich willkommen zu dieser wöchentlichen Umfrage!

Sie sollte nur wenige Minuten Ihrer Zeit in Anspruch nehmen. Diese Umfrage enthält 7 Fragen.

Allgemeine Angaben

Ihre UserID und Ihre Gruppennummer werden in der Authenticator-App unten rechts angezeigt. Auf der Homepage finden Sie auch eine ausführlichere Anleitung.

Das ist meine UserID:

Bitte geben Sie Ihre Antwort hier ein: ID[...]

Das ist meine Gruppennummer:

Bitte geben Sie Ihre Antwort hier ein: [...]

E.3.1 Ihre Eindrücke zu dem letzten Indikator

Wichtiger Hinweis: In der letzten Woche wurde Ihnen der folgende Indikator angezeigt. Die Fragen beziehen sich auf diesen Indikator.

Gruppe 1: Kein Indikator Die Unterbrechung fand ohne Vorwarnung statt. Es wurde nur eine einfache Benachrichtigung angezeigt.

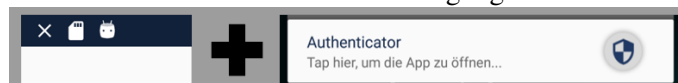
Gruppe 2: Kurz- & Langfristiger Indikator Die Unterbrechung wurde lang- und kurzfristig angekündigt:

- Langfristig durch den "Füllstand" des Schild-Symbols in der Statusleiste und durch das Anzeigen einer ausführlichen Benachrichtigung
- Kurzfristig durch das aufploppen der ausführlichen Benachrichtigung und den dunkler werdenden Bildschirm

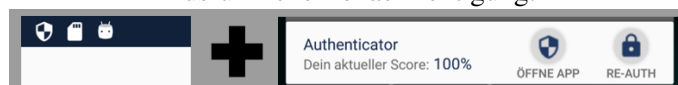
Gruppe 3: Langfristiger Indikator Die Unterbrechung wurde langfristig durch den "Füllstand" des Schild-Symbols in der Statusleiste und durch Anzeige der ausführlichen Benachrichtigung angekündigt.

Gruppe 4: Kurzfristiger Indikator Die Unterbrechung wurde kurzfristig durch das Erscheinen der ausführlichen Benachrichtigung und den dunkler werdenden Bildschirm angekündigt.

Einfache Benachrichtigung:



Ausführliche Benachrichtigung:



Likert Skalen

Beantworten Sie die folgenden Fragen bitte Ihrem persönlichen Empfinden nach. Die Fragen beziehen sich auf den Indikator der Ihnen, Ihrer Gruppe entsprechend, letzte Woche angezeigt

wurde.

Mit dem Begriff “System” beziehen wir uns auf das ganze System, also auf den letzten Indikator, die App und das automatische Sperren des Bildschirms. Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Antwortmöglichkeiten:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu
- trifft nicht zu

Aussagen:

- Ich habe das System während der letzten Woche als hinderlich empfunden.
- Ich habe das System während der letzten Woche als lästig empfunden.
- Ich habe das System während der letzten Woche als leicht zu benutzen empfunden.
- Ich habe mich in der letzten Woche durch das System dazu motiviert gefühlt mich freiwillig zu Re-Authentifizieren.
- Ich hatte das Gefühl durch den gestiegenen Score nach einer Re-Authentifizierung belohnt zu werden.

Das hat mir an dem Indikator der letzten Woche gefallen:

Bitte geben Sie Ihre Antwort hier ein: [...]

Geben Sie positive Eigenschaften des Indikators, der Ihnen in der letzten Woche angezeigt wurde, an.

Das hat mir an dem Indikator der letzten Woche nicht gefallen:

Bitte geben Sie Ihre Antwort hier ein:[...]

Geben Sie negative Eigenschaften des Indikators, der Ihnen in der letzten Woche angezeigt wurde, an.

Diese Fehler oder Probleme sind mir in der letzten Woche aufgefallen:

Bitte geben Sie Ihre Antwort hier ein: [...]

Falls Sie Probleme mit dem System hatten, erläutern Sie diese hier bitte.

Das möchte ich noch sagen:

Bitte geben Sie Ihre Antwort hier ein: [...]

Wenn Sie noch einen Kommentar zu der Studie abgeben möchten, können Sie das hier tun.

Vielen Dank für Ihr Feedback!

E.4 Additional Questions on the last weekly Survey

E.4.1 Abschließendes Ranking der Indikatoren

Sortieren Sie die 4 Arten von Indikator.

Stellen Sie den Indikator, den Sie **am ehesten benutzen würden nach ganz oben**. Der Indikator, den Sie am wenigsten gerne nutzen würden kommt auf den untersten Platz.

Bitte nummerieren Sie jede Box in der Reihenfolge Ihrer Präferenz, beginnen mit 1 bis 4

- Kein Indikator
- Kurzfristiger Indikator
- Langfristiger Indikator
- Kurz- und Langfristiger Indikator

Ganz oben sollte Ihr Favorit stehen. Je weiter unten ein Indikator steht, desto weniger hat er Ihnen gefallen. Klicken Sie auf einen Indikator links und ziehen Sie ihn rechts an die entsprechende Stelle Ihrer Rangfolge.

Das mag ich an dem Indikator, den ich auf den höchsten Platz gestellt habe:

Bitte geben Sie Ihre Antwort hier ein: [...]

Das stört mich an dem Indikator, den ich auf den letzten Platz gestellt habe:

Bitte geben Sie Ihre Antwort hier ein: [...]

Das möchte ich noch zu den Indikatoren sagen:

Bitte geben Sie Ihre Antwort hier ein: [...]

E.4.2 Erfahrungen mit dem Authenticator

Abschließend bitten wir Sie noch einige Fragen zu Ihrem allgemeinen Eindruck des Systems stellen. Beantworten Sie die Fragen bitte ihrem persönlichen Empfinden nach.

Allgemeine Erfahrungen mit der App

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Antwortmöglichkeiten:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu

trifft nicht zu

Aussagen:

- Das Design des Authenticators hat mir gefallen.
- Die Vibration der Benachrichtigungen des Authenticators hat mich gestört.
- Das Aufploppen der Benachrichtigungen des Authenticators hat mich gestört.
- Der dunkler werdende Bildschirm des kurzfristigen Indikators hat mich gestresst.
- Das Schild-Symbol des langfristigen Indikators hat mir geholfen, die Zeit bis zur nächsten Re-Authentifizierung besser abschätzen zu können.
- Das System hat mein Verhalten bei der Nutzung meines Smartphones beeinflusst.
- Fehler des Authenticators, die ich während der Benutzung bemerkt habe, haben meinen Eindruck beeinflusst.

Mini-Fragebögen

Die folgenden 3 Fragen beziehen sich auf die **Mini-Fragebögen, die Ihnen mehrmals täglich nach einer Re-Authentifizierung auf Ihrem Smartphone erschienen sind.**

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Antwortmöglichkeiten:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu
- trifft nicht zu

Aussagen:

- Die Mini-Fragebögen haben mich gestört.
- Die Mini-Fragebögen haben mein Empfinden der App beeinflusst.
- Die Mini-Fragebögen haben mein Verhalten bei der Nutzung meines Smartphones beeinflusst.

Selbständiges Informieren zum Authenticator

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Antwortmöglichkeiten:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu
- trifft nicht zu

Aussagen:

- Ich kannte den Ansatz der impliziten Authentifizierung bereits vor der Studie gut.
- Ich habe die Einführung in das Thema auf der Startseite der Homepage zu Beginn der Studie ganz durchgelesen.
- Ich habe das Video auf der Startseite der Homepage zu Beginn der Studie ganz angesehen.
- Ich habe mich gut über die Funktionen der App oder über den Studienverlauf auf der Homepage informiert.
- Ich habe mich an einer anderen Stelle über implizite Authentifizierung informiert.

Meiner Meinung nach sollte der kurzfristige Indikator ..Sekunden brauchen um den Bildschirm komplett abzudunkeln.

Der kurzfristige Indikator braucht 8 Sekunden um dem Bildschirm nach und nach komplett abzudunkeln.

Bitte geben Sie Ihre Antwort hier ein: [...] Sekunden

Geben Sie den Wert in Sekunden an, den der kurzfristige Indikator brauchen sollte, um den Bildschirm komplett abzudunkeln. Derzeit werden dafür 8 Sekunden benötigt.

Empfinden der Impliziten Authentifizierung

Die implizite Authentifizierung ist eine neue Idee zur Authentifizierung. **Der Authenticator simuliert das Verhalten eines Systems, das implizite Authentifizierung benutzt.**

Beschreibung Implizite Authentifizierung: Das Smartphone erkennt automatisch, ob die Person, die gerade das Gerät benutzt dazu berechtigt ist. Wenn das System sich nicht sicher ist, muss der Benutzer sich Re-Authentifizieren. Das Entsperren des Bildschirms, jedes mal nachdem man ihn anschaltet, ist dann nicht mehr nötig und fällt somit weg.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Antwortmöglichkeiten:

- trifft zu
- trifft eher zu
- teils-teils
- trifft eher nicht zu
- trifft nicht zu

Aussagen:

- Ich sperre meinen Bildschirm immer, wenn ich mein Smartphone aus der Hand lege.
- Angenommen ich hätte die Wahl: Es ist störender, manchmal während der Benutzung meines Smartphones für eine Re-Authentifizierung unterbrochen zu werden, als mich jedes mal Authentifizieren zu müssen, wenn ich den Bildschirm einschalte.
- Ich würde gerne implizite Authentifizierung (siehe Beschreibung oben) auf meinem Smartphone benutzen.

Das möchte ich noch sagen:

Bitte geben Sie Ihre Antwort hier ein: [...]

Wenn Sie noch einen Kommentar zu der Studie abgeben möchten, können Sie das hier tun.

Die Studie ist hiermit beendet, Sie können die App jetzt gerne deinstallieren. Vielen Dank für Ihre Teilnahme an der Studie!

Ihren Amazon-Gutschein oder Ihre MMI-Punkte werden Sie in Kürze per E-Mail erhalten.

F Interview Guide for the Final Semi-Structured Interviews

Die ausformulierten Fragen sind nur als Orientierung zu sehen.

Begrüßung

Guten Tag. Ich rufe an wegen des abschließenden Interviews zur Nutzerstudie "Indikatoren für die Re-Authentifizierung". Ich möchte Sie direkt darauf hinweisen, dass dieses Gespräch aufgezeichnet wird. Sind Sie damit einverstanden?

Allgemeiner Eindruck

Wie war Ihr allgemeiner Eindruck zu dem System? Was fällt Ihnen spontan zur Studie ein?

Situationen

Erinnern Sie sich an eine bestimmte Situation mit der App? (Positiv oder Negativ)

Vielleicht eine Situation, wo die App tatsächlich einen unautorisierten Zugriff verhindert hat? Oder eine Situation, wo die Sperrung des Smartphones besonders gestört hat?

Implizite Authentifizierung

Was halten Sie von dieser Idee? Würden Sie so ein System benutzen? Gründe - Warum?

Sicherheitsempfinden

Was halten Sie von der Sicherheit die solch ein System bietet? Im Vergleich zur traditionellen Methode?

Re-Authentifizierungs Unterbrechungen

Wie haben Sie die Unterbrechungen zur Re-Authentifizierung empfunden?

Haben Sie noch eine Idee, wie diese Unterbrechungen weniger störend sein könnten?

Eventuell: Wären unvorhersehbare Unterbrechungen zur Re-Authentifizierung für Sie ein Grund so ein System nicht zu benutzen? Und mit Indikatoren?

Themen basierend auf den Daten

Haben Sie irgendwelche Fehlfunktionen bemerkt? z.B.: An manchen Tagen nicht gesperrt worden, Fragebögen nicht erschienen, App abgestürzt.

Haben Sie die Mini-Fragebögen, die nach dem Entsperren manchmal auf Ihrem Handy erschienen sind, eher direkt beantwortet, oder auf später verschoben? ("Jetzt nicht" gedrückt)

Abschließende Fragen

Wie fanden Sie die Umsetzung der App? Hat Ihnen die Gestaltung gefallen? Informationen gefehlt?

Wie fanden Sie die Homepage? Haben Sie sie benutzt? Wofür?

Zu guter Letzt: Möchten Sie noch etwas sagen oder anmerken?

Vielen Dank, damit sind wir fertig mit dem Interview. Ich möchte mich noch einmal ganz herzlich für Ihre Teilnahme an der Studie und am Interview bedanken.

G Top 20 Most Interrupted Apps

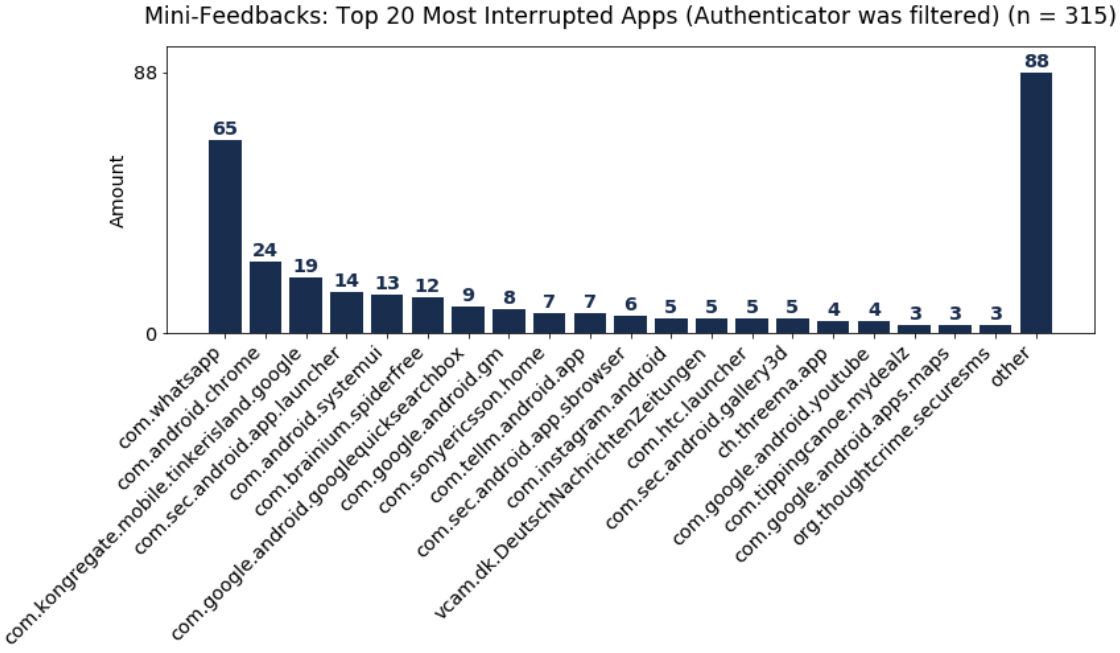


Figure G.1: Illustration of the top 20 interrupted apps given as package names.

Content of the attached CD

1. Digital version of this thesis (.pdf file)
2. Slides used for the disputation
3. Source files of this thesis including the .tex files and all images
4. Used references as .pdf files
5. Used web-references
6. Data gathered during the focus group and material used for it's execution
7. Data gathered during the user study and material used for it's execution
8. Whole project of the Authenticator app including all source code
9. Released .apk file of the Authenticator
10. Source code of the homepage including the introduction video

References

- [1] L. Agarwal, H. Khan, and U. Hengartner. Ask me again but don't annoy me: Evaluating re-authentication strategies for smartphones. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 221–236, Denver, CO, USA, June 2016. USENIX Association.
- [2] A. J. Aviv, K. L. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. *Woot*, 10:1–7, 2010.
- [3] A. Buriro, B. Crispo, F. Del Frari, and K. Wrona. Touchstroke: Smartphone user authentication based on touch-typing biometrics. In *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, pages 27–34, Genoa, Italy, September 2015. Springer International Publishing.
- [4] D. Buschek, A. De Luca, and F. Alt. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 1393–1402, Seoul, Republic of Korea, April 2015. ACM.
- [5] D. Buschek, F. Hartmann, E. von Zezschwitz, A. De Luca, and F. 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, CHI '16*, pages 3736–3747, San Jose, CA, USA, May 2016. ACM.
- [6] N. Clarke, S. Karatzouni, and S. Furnell. Flexible and transparent user authentication for mobile devices. In *Emerging Challenges for Security, Privacy and Trust, SEC 2009*, pages 1–12, Pafos, Cyprus, May 2009. Springer Berlin Heidelberg.
- [7] H. Crawford and K. Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 1(1):7, Jun 2014.
- [8] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. 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, CHI '12*, pages 987–996, Austin, Texas, USA, May 2012. ACM.
- [9] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHH-MSP '10*, pages 306–311, Washington, DC, USA, October 2010. IEEE Computer Society.
- [10] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, pages 750–761, Scottsdale, Arizona, USA, November 2014. ACM.
- [11] T. Feng, Z. Liu, K. Kwon, W. Shi, B. Carburnar, Y. Jiang, and N. Nguyen. Continuous mobile authentication using touchscreen gestures. In *2012 IEEE Conference on Technologies for Homeland Security (HST)*, pages 451–456, Waltham, MA, USA, November 2012. IEEE.
- [12] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, Jan 2013.

- [13] A. Griesbeck. Mobile app-based authentication mechanisms in the wild. Bachelor's thesis, Ludwig Maximilians Universität München, 2017.
- [14] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 213–230, Menlo Park, CA, USA, July 2014. USENIX Association.
- [15] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley. Casa: Context-aware scalable authentication. In *Proceedings of the Ninth Symposium on Usable Privacy and Security, SOUPS '13*, pages 3:1–3:10, Newcastle, United Kingdom, July 2013. ACM.
- [16] E. Hayashi, O. Riva, K. Strauss, A. J. B. Brush, and S. 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, SOUPS '12*, pages 2:1–2:11, Washington D.C., USA, July 2012. ACM.
- [17] K. Holien. Gait recognition under non-standard circumstances. Master's thesis, Gjøvik University College, 2008.
- [18] H. Khan, A. Atwater, and U. Hengartner. Itus: An implicit authentication framework for android. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, pages 507–518, Maui, Hawaii, USA, September 2014. ACM.
- [19] H. Khan, U. Hengartner, and D. 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, Canada, July 2015. USENIX Association.
- [20] A. H. Lashkari, S. Farmand, D. O. B. Zakaria, and D. R. Saleh. Shoulder surfing attack in graphical password authentication. *International Journal of Computer Science and Information Security*, 6(2):145–154, Nov 2009.
- [21] L. Li, X. Zhao, and G. Xue. Unobservable reauthentication for smartphones. In *NDSS 2013*, volume 56, pages 57–59. The Internet Society, April 2013.
- [22] D. 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] N. Micallef, M. Just, L. Baillie, M. Halvey, and H. G. 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, MobileHCI '15*, pages 284–294, Copenhagen, Denmark, August 2015. ACM.
- [24] K. Mock, B. Hoanca, J. Weaver, and M. Milton. Real-time continuous iris recognition for authentication using an eye tracker. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS '12*, pages 1007–1009, Raleigh, North Carolina, USA, October 2012. ACM.
- [25] F. Monrose, M. K. Reiter, and S. Wetzel. Password hardening based on keystroke dynamics. *International Journal of Information Security*, 1(2):69–83, Feb 2002.
- [26] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security, CCS '97*, pages 48–56, Zurich, Switzerland, April 1997. ACM.

- [27] L. O’Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, Dec 2003.
- [28] O. Riva, C. Qin, K. Strauss, and D. Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12)*, pages 301–316, Bellevue, WA, USA, August 2012. USENIX.
- [29] H. Saevanee, N. L. Clarke, and S. M. Furnell. Multi-modal behavioural biometric authentication for mobile devices. In *Information Security and Privacy Research*, pages 465–474, Heraklion, Crete, Greece, June 2012. Springer Berlin Heidelberg.
- [30] E. Shi, Y. Niu, M. Jakobsson, and R. Chow. Implicit authentication through learning user behavior. In *Information Security*, pages 99–113, Boca Raton, FL, USA, October 2011. Springer Berlin Heidelberg.
- [31] K. Winkler. Entwicklung und evaluation eines inhaltsabhängigen sicherheitsmechanismus: Performance, wahrnehmung und machbarkeit. Master’s thesis, Ludwig Maximilians Universität München, 2016.
- [32] H. Xu, Y. Zhou, and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 187–198, Menlo Park, CA, USA, July 2014. USENIX Association.

Web-References

- [33] The Apache Software Foundation. Apache License, Version 2.0, January 2004, 2018-09-26. URL: <https://www.apache.org/licenses/LICENSE-2.0.html>.
- [34] Google. Material Design Icons, 2018-09-26. URL: <https://material.io/tools/icons/?style=baseline>.
- [35] PopulationPyramid.net. Population Pyramids of the World from 1950 to 2100, Germany 2016, 2018-09-14. URL: <https://www.populationpyramid.net/germany/2016/>.
- [36] Statista. Share of smartphone users in Germany from 2012 to 2016, by age, 2018-09-14. URL: <https://www.statista.com/statistics/732504/germany-smartphone-users-by-age/>.
- [37] Google Support. Set your Android device to automatically unlock, 2018-10-02. URL: https://support.google.com/nexus/answer/9075927?hl=en&ref_topic=6168852.
- [38] Zenith The ROI agency, Publicis Groupe. Weltweite Smartphone-Penetration steigt 2018 auf 66 Prozent, 2018-07-20. Press Release. URL: <https://www.publicismedia.de/wp-content/uploads/2017/10/2017-10-16-mobile-advertising-forecasts-2017-de.pdf>.