Tactile One-Time Pad: Leakage-Resilient Authentication for Smartphones

Sebastian Uellenbeck(✉), Thomas Hupperich, Christopher Wolf, and Thorsten Holz

Ruhr-University Bochum, Bochum, Germany
{sebastian.uellenbeck,thomas.hupperich,christopher.wolf, thorsten.holz}@rub.de

Abstract. Nowadays, Smartphones are widely used and they have a growing market share of already more than 55 % according to recent studies. They often contain sensitive or private data that can easily be accessed by an attacker if the device is unlocked. Since smartphones are mobile and used as everyday gadgets, they are susceptible to get lost or stolen. To prevent the data from being accessed by an attacker, access control mechanisms like user authentication are needed. However, commonly used authentication mechanisms like PINs, passwords, and patterns suffer from the same weakness: They are vulnerable against different kinds of attacks, most notably shoulder surfing. In order to prevent shoulder surfing, a secure channel between the smartphone and the user must be established that cannot be eavesdropped by an adversary.

In this paper, we concentrate on the smartphone's tactile feedback to add a new security layer to the plain PIN-based authentication mechanism. The key idea is to use vibrations as an additional channel to complement PINs with a tactile one-time pattern. To calibrate the usability of our approach, we developed a game that more than 220 participants played to determine the shortest vibration duration most people can sense. In a security evaluation, we recorded the acoustical signal of the vibration motor of five different smartphones at four different locations with a high-end microphone to cross-correlate a login scenario with a pre-recorded acoustical fingerprint of the devices. Our evaluation results demonstrate that it is not possible for an attacker to spot the user's secret under normal conditions, e. g., in a restaurant or during a conversation, even with professional equipment. Finally, we show that the required overhead of our approach is reasonable in practice and outperforms prior work.

Keywords: Authentication · Computer security · Smartphone security · Human computer interaction · Tactile feedback

1 Introduction

Smartphones are among the most popular gadgets available on the market today. According to a study by Gartner, smartphones had a market share of 55 % in the

R. Böhme and T. Okamoto (Eds.): FC 2015, LNCS 8975, pp. 237–253, 2015.
DOI: 10.1007/978-3-662-47854-7_15

third quarter of 2013 and they are expected to grow even more in the future [12]. Such devices are not only used for taking pictures, sending text messages, or surfing the Web, but also to assist sensitive applications such as online banking by receiving *mobile Transaction Authentication Numbers* (mTANs), as electronic replacement for a purse, or as key to the office door. Hence, smartphones typically contain lots of private data like contact information, personal messages, and passwords. Obviously, they become an interesting target for attackers, who can easily access the sensitive information if the device is unlocked.

Access control mechanisms, especially user authentication, can be used to protect the data if the device is lost or stolen. Typical authentication mechanisms for smartphones include PINs, passwords, and pattern-based login mechanisms that are adapted to the screen size of mobile devices. Unfortunately, all these authentication mechanisms are cumbersome. Simultaneously users think that authentication is often required for features that should not require authentication [13,17]. In general, it is difficult to attain a usable and secure authentication approach [8]. Aviv et al. [2] show that it is feasible to utilize the accelerometer as a side channel to predict PINs and patterns, making these authentication mechanism susceptible to attacks. A related threat are so called *smudge attacks* [1] on smartphones. Furthermore, a major hurdle of all existing mechanisms is that they suffer from the same weakness: They are not resistant to *shoulder surfing attacks* [21]. Here, the adversary visually observes the login and can then easily replay the observed successful authentication.

In this paper, we introduce a novel authentication method resistant to shoulder surfing attacks. To this end, we study all available channels between a user and a smartphone without additional hardware (e. g., headsets) to determine which channels can be utilized for a secure communication. It turns out that tactile feedback suits our needs best: We demonstrate that vibrations can be used as an additional channel to complement existing, PIN-based authentication mechanisms. The key insight is that we can take advantage of vibrations to establish some kind of *one-time pad* (OTP) to generate pseudo-random numbers that can be added to an existing PIN. The combination of this tactile feedback with a PIN enables an authentication mechanism that is resistant to shoulder surfing since an attacker cannot easily intercept the vibrations. In addition, smudge attacks are dwarfed as the digits entered are now randomly distributed.

We implemented a prototype of this concept in a tool called Tactile One-Time Pad (short: Taco) for the Android operating system. In a security evaluation, we analyzed how resistant the mechanism is in practice: We recorded the acoustical signal generated by vibrations for five different smartphones at four different locations with a high-end microphone. This allows us to cross-correlate a login scenario with a pre-recorded acoustical fingerprint of the devices. It turns out that an adversary cannot perform such an acoustic attack on our authentication scheme under normal conditions, e. g., during a conversation or a modestly busy place like a restaurant. Experimental results suggest that such attacks are only feasible in a very quiet place (i. e., in an anechoic room), an attack scenario beyond our threat model.

A crucial aspect of our system is the time span of a vibration (i. e., how long we let the smartphone vibrate). To determine the optimal length, we designed a game to identify the shortest vibration duration most people can perceive. In a user study with more than 220 participants, we found that 90 % of all participants were able to notice a vibration duration of 150 ms. Combined with other insights obtained during the study, we adjust the parameters of our prototype to obtain an authentication mechanism that is usable in practice.
In summary, we make the following four contributions in this paper:

1. We introduce a novel authentication scheme that utilizes the tactile feedback available on smartphones to enhance existing, PIN-based authentication mechanisms. Vibrations are used to generate pseudo-random numbers perceivable only by a user and this channel is used as an additional input during the authentication phase.
2. We present our prototype implemented for the Android platform in a tool called Taco. Our scheme does not need special/additional hardware but only the vibration mechanism available on common smartphones.
3. For a security evaluation, we recorded a pre-defined pattern from five different smartphones at four different locations to analyze the data by means of a cross-correlation and demonstrate that the scheme is resistant to acoustic attacks. We also discuss and empirically evaluate other attack scenarios.
4. We conducted a user study with 227 participants to evaluate the usability of our approach and found that the required overhead is feasible in practice.

Note that a longer version of this paper with more technical details is available as a technical report [22].

2 Related Work

In recent years, several methods for *leakage resilient* user authentication have been proposed. In the following, we provide a brief overview of the most prominent of these methods and discuss how Taco relates to them.

Yan et al. focus on the visual channel and propose *CoverPad* [24]. Here the user has to shield the screen with the palm of his hand to hinder attackers from eavesdropping a secret. The user has to consider this secret to do simple calculations and finally he has to enter the result into the device. Although the login duration seems comparable to Taco, there is no security evaluation of *CoverPad*. Therefore, we cannot know how secure *CoverPad* performs in reality. However, we evaluate what attacks Taco resists against in Sect. 4.

In the same manner, Perkovic et al. use the visual channel to transfer a secret between the user and the device, but also propose a headset as alternative [16]. Having retrieved the secret, the user can apply two methods that are based on lookup tables, or utilize simple modulo 10 calculation [23] to authenticate. In contrast to our approach, they use additional hardware to establish a secure channel between the user and the device, while we only leverage tactile feedback.

De Luca et al. evaluated three different approaches for eye-gaze interaction to enhance PIN authentication [10]. *Cued Gaze-Points* is a system presented by Forget et al. that uses a cued-recall graphical password scheme for user authentication [11]. The user has to select points on a sequence of images with his eye-gaze as secret and later look at the desired points and hold the space bar for a few seconds. In contrast to our scheme, both approaches are only resistant to shoulder surfing if the attacker only observes the user's display. In case the attacker simultaneously tracks the eyes of the user, she can obtain the secret.

Bianchi et al. proposed *Secure Haptic Keypad* (SHK) [5] as well as PHONE LOCK [4]. Both approaches make use of tactile feedback for authentication. For the first one, the user has to touch three haptic buttons that vibrate with different frequencies to authenticate. In a round-based fashion, he has to press the button that represents his partial secret. Since no visual feedback is given to a shoulder surfer, the optical channel is secured. However, as we show in Sect. 4.2, acoustic attacks on tactile feedback are feasible so the complete secret can be obtained if only one channel is eavesdropped. As opposed to this, an attacker has to eavesdrop two channels to obtain the user's screen when using TACO. For the second approach, they implemented a virtual wheel on the smartphone's touchscreen with ten segments of the same size and a selection button in the middle of the segments. Again the login process is round-based and the user has to find his own vibration pattern. To do so, he touches the segments and tries to find his tactile pattern. Having found his pattern, he has to use the selection button. The segment's allocation to vibration patterns changes randomly so shoulder surfing is not possible. Contrary to our approach, the full secret is always transfered between the user and the device meaning that an attacker only has to eavesdrop the secure channel to retrieve the secret. In Sect. 4.2, we show that eavesdropping the tactile feedback of the smartphone is possible under some conditions.

3 Tactile One-Time Pad

In this section, we describe our approach to obtain an authentication mechanism on smartphones resilient to shoulder surfing attacks.

3.1 Potential Communication Channels

Leakage resilient authentication can be implemented by using a secure channel between the user and the device. In a nutshell, we need to make the protocol interactive, so there needs to be an information flow from the user to the device (*input*), but also in the reverse direction (*helper data*). Focusing on the reverse direction, humans only have five traditional senses to obtain stimuli: sight, hearing, touch, taste, and smell. As long as smartphones are not able to change their taste or smell controlled by an application, we cannot use taste or smell to transfer information. As a consequence, sight, hearing, and touch remain as possible

candidates. Restricting our setting further to smartphones without any additional hardware, there are only three potential channels to transfer information: the display, the speaker, and the vibration motor.

The first channel—the *display*—can show arbitrary graphical information. While this channel can transport a lot of information from the smartphone to a user, an adversary can also easily eavesdrop such a channel by utilizing a camera [3]. We therefore cannot assume that it is a secure channel, but need to treat this as an untrusted communication medium.

The second possibility is the smartphone's *speaker*, more precisely the audio output that can also transport a lot of information. However, the same drawback that holds for the display is also valid for the audio output: it can easily be eavesdropped with a normal microphone; for example, every smartphone is equipped with such a microphone. Note that this does not apply if head phones are allowed. However, they qualify as "additional hardware" and are hence excluded from our list of possible channels.

The third and most interesting channel from the smartphone to the user is the *vibration motor*. All smartphones offer it to provide tactile feedback to the user (e. g., for silent notification). Furthermore, a vibration unit is commonly available in many kinds of mobile devices, even in older ones such as feature phones. Tactile feedback has three main advantages over the display and the speaker. First, it is hard to eavesdrop by an attacker as it has only a limited visual and acoustic range. More precisely, the vibration of the smartphone can only be seen with a high-speed, high-resolution video camera that is placed near the smartphone [9]. The acoustic feedback depends on the resonating body the smartphone is fixed in. In case of a (wooden) table, the latter acts as resonating body and amplifies the oscillation. As a result, the vibration can be easily heard by an attacker. However, in the more likely case that a human holds his smartphone in his hand when entering a PIN, the game is very different: Here the hand absorbs the oscillation so the vibration can barely be heard anymore, even within a very small distance from the smartphone. Empirical measurements in different settings confirm this observation (see Sect. 4.2 for details). Second, tactile feedback is easy to identify by the user even in dark or noisy environments. Third, humans do not need special training to correctly recognize vibration. This hugely adds to the overall usability of our solution.

Despite its advantages, there is also the low bandwidth of the channel that needs to be considered. In a first feasibility study we found that it is hard to detect more than 10 events per second and for none of the participants it was possible to detect more than 15 events per second. Based on an empirical user study with 227 participants, we estimate that 90 % of all users can recognize at least four events per second (see Sect. 5.2). Even such a low bandwidth is enough since we only utilize tactical feedback during the authentication process.

In summary, we conclude that a leakage resilient authentication method suitable for mobile devices can be accomplished using the built-in vibration motor. In a nutshell, we combine a one-time pad that is information theoretically secure [19] (based on addition modulo 10) with a computer generated secret.

3.2 Attacker Model

For the rest of the paper, we assume an attacker that can eavesdrop on the screen/keypad (cf., [3]). More specifically, we assume the classical model of an eavesdropper that performs a shoulder surfing attack and, in addition, is able to observe the vibrations of the smartphone. An attacker may obtain a smartphone of the same model she wants to attack and measure the vibration unit in advance. Empirical evaluation results in several different settings demonstrate that this is actually hard in practice (see Sect. 4 for a more detailed justification that this rational is sound). Furthermore, the adversary can take notes and observe multiple rounds of the authentication process; an assumption that is stronger compared to previous work in this area [18].

3.3 Methodology and Implementation

A *Personal Identification Number* (PIN) typically consists of four to eight decimal digits—the secret—that has to be entered correctly to authenticate. PINs have the advantage that they are simple to create, to recall, to verify, and to change. The main drawback when using a plain PIN authentication schema is the relative ease to eavesdrop the secret. A prominent attack in this area is shoulder surfing, another one consists of analyzing the residue on the touch screen [1]. To use Taco, the user chooses a four to eight digit decimal PIN. As for plain PIN authentication, he needs to remember and enter it. In addition, the user needs to choose a vibration duration between 40 ms and 350 ms. This is used to establish a secret channel between the smartphone and the user. Note that users will choose larger values for the vibration duration in the beginning, but likely reduce this time span when they feel more comfortable with the scheme. We have captured and confirmed this behavior in a simple game-like user-study (see Sect. 5 for details).

To perform authentication, the user holds the smartphone in the palm of his hand and starts the authentication process by pressing the AUTHENTICATION button. After the button is pressed, the smartphone vibrates between zero and nine times (one digit of the one-time pad). The user has to count the number of vibrations. Having determined the number of vibrations, the user adds the first digit of his PIN. If the result is larger than 9, he subtracts 10 (i. e., only the last digit is used). This digit is now entered as the response to the actual challenge given by the phone. If the user was not able to sense the number of vibrations (e. g., as a result of disturbance or a lack of attention), he can press the REPEAT button to feel the same number of vibrations again. After the result is entered, the phone again starts the same cycle for all consecutive digits of the PIN.

If all digits have been entered correctly, the user is granted access to his smartphone. While the login is performed, the user can either shorten or extend the vibration duration; this allows to either speed up the authentication process or to increase the likelihood of recognizing the correct number of vibrations.

At this point, one can think that this might lead to a security-relevant side channel because an attacker could clock the time between two entered digits

to obtain the number of vibrations or at least a hint to calculate the secret. However, this is not possible because for a given vibration duration, the full time period is always equal no matter if the smartphone vibrates zero or nine times. This is accomplished by aligning the pauses between the single vibrations, so that the complete pattern always fits the same period of time. Therefore, for a given single duration of a vibration, the length of a round is always the same.

In summary, TACO is an additional security layer for the PIN authentication scheme. In case the user is sure that no shoulder surfing occurs, he can switch it off for fast authentication. In case he suspects a possible attack, he can switch it on; the price to pay is a small additional overhead in time to perform the authentication. To test our approach and verify both its efficiency and usability, we implemented a prototype in Java for the Android platform. We also developed a game to estimate a reasonable vibration duration a user can recognize.

3.4 Extensions and Discussion

There are several potential extension of our current prototype. For example, other mathematical operations like subtraction, multiplication, or integer division could be added as part of the scheme. For some people it might be easier to perform subtractions instead of additions particularly when using PINs with large numbers. When dropping the requirement of using an OTP, this could result in a more efficient scheme. Further, we need to keep the benefit that *both* channels need to be eavesdropped by an attacker to obtain the secret key.

As a potential way to increase speed, we may want to encode the digits $0 \ldots 9$ differently. This could be done by using a binary encoding, e. g., with short and long vibrations. By using this mode, we consider a short vibration as binary zero and a long vibration as binary one. Treating the concatenation of zeros and ones as a binary number, one can transform this into a decimal number. In this mode, only four binary vibrations are required to encode ten decimal digits. Albeit, switching to this mode can act like a double-edged sword: On the one hand, it leads to a decreasing overhead. On the other hand, it also leads to an increasing difficulty since users also have to do a binary to decimal transformation before entering the result of the addition modulo ten. We may envision this as the "expert mode" for TACO—where most people start with $0 \ldots 9$ vibrations and then migrate to the faster communication pattern if needed. In summary, this requires a more extensive user study to determine if this kind of encoding allows an increase of speed while still being usable and secure.

4 Security Evaluation

We now consider different attack vectors regarding their ability to attack TACO.

4.1 Timing Attacks

In our attacker model, the attacker knows the methodology of TACO including the duration of a single vibration because she might have measured it before.

To counter timing attacks, the overall length of one round is the same, no matter which number is transmitted by vibrations (cf., Sect. 3.3). If the pattern lengths would vary, an attacker could time the duration of the vibration pattern—namely the time between two keystrokes—and guess the number of vibrations. But as the pattern duration is the same for each OTP digit, an attacker cannot obtain any information by measuring the time between two keystrokes. In addition, humans do react individually on stimuli, so we can assume that the additional time a user needs to add the OTP digit to his PIN digit and enter the sum will shadow any useful information an attacker might obtain by measuring this time. To confirm this claim, we analyzed the data obtained by a usability study we describe in Sect. 5. Since we knew the number of vibrations during this study, the secret PIN, the user's input, and all timings with a granularity of milliseconds for each user, we computed the average time and the standard deviation that elapsed between two keystrokes for all users. Figure 1 shows that users on average need more time to enter the result if the number of vibrations is greater than six. However, while the differences are less than two seconds, the standard deviation is on average larger than ±2.5 s. In the end, this can lead to a timing side channel [14] if the attacker is able to measure this many times *and* the user's skill in adding two digits mod 10 does not improve. We argue that if the user's skill does not improve—what we can measure in an automatic fashion—, we can force him to change his PINs on a regular basis. Otherwise, no countermeasures against timing side channels are required.

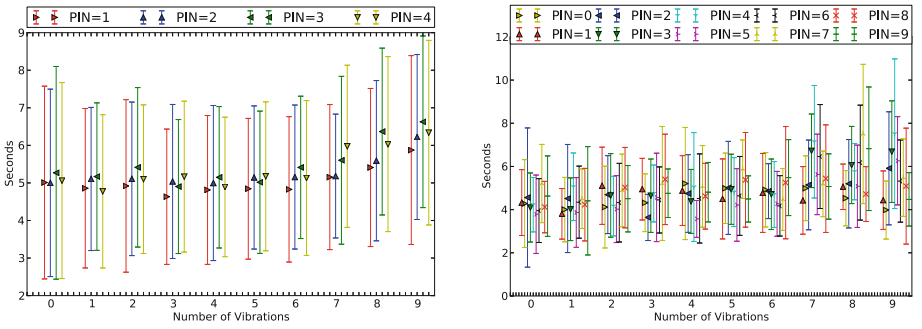


Fig. 1. Timing attacks on Taco. We analyzed the timing data from our experiments (cf., Sect. 5) to obtain the average time a user needs to enter a digit. The left figure shows the first part of our experiment, the right figure the second part of our experiment. The points are grouped by the number of vibrations.

4.2 Acoustic Attacks

In contrast to timing attacks, acoustic attacks are more severe against our scheme. If a smartphone is placed on a wooden table in a silent room, one is able to hear vibrations and most likely also to count them without any technical

equipment. In this case, the secret key of the one-time pad would be broken and in combination with classical shoulder surfing it is possible to obtain the secret. At this point, the PIN can be calculated and the authentication process can be reproduced by an attacker.

Fortunately, the mobile phone is usually held in a person's hand while entering the PIN. In addition, the human body effectively shields vibrations rather than acting as a resonator compared to a wooden table. Furthermore, environmental background noise effectively disguises any sound Taco creates. Therefore, it is likely hard to count the number of vibrations when standing next to a person using Taco as it is hard to hear the vibration signals under these circumstances. To actually gather vibration signals in a room with background noise or even outside, an attacker would need rather expensive audio recording tools which would attract too much attention. In short: In a crowded room shoulder surfing is possible, but there is too much background noise to eavesdrop the vibration channel. In a deserted room, shoulder surfing becomes suspicious and an attacker cannot read the digits entered into the device, while it might be possible to eavesdrop the OTP vibrations. Hence, using two different channels with different vulnerabilities actually leads to an authentication method that is *strengthened* against the individual attack.

To quantify this attack vector in more detail, we conducted an experiment with five different smartphones at four different locations. As smartphones we used a Google Nexus S, a Google Galaxy Nexus, a LG L7 P700, a HTC Nexus One, and a Sony Xperia S. To show that it is possible to detect vibrations from further distances, we first chose a special prepared anechoic room. In our experiments, we used a large-diaphragm capacitor microphone (Rode NT2000) with a frequency spectrum between 20 Hz and 20 000 Hz, a signal-to-noise ratio of 84 dB (1 kHz rel 1 Pa, per IEC651, IEC268-15), configured with kidney directionality.

One might think that the speaker is the ideal solution to fool the attacker by creating false sounds. To prove this assumption, we conducted a short experiment by playing-back white noise with different smartphone speakers and recoding this noise with a high-end microphone. By analyzing the obtained data, we found that the small speakers of smartphones are not capable of creating noise with a high amplitude at low frequencies. Therefore, we cannot utilize the smartphone's speaker to disturb or prevent recordings of the vibrations.

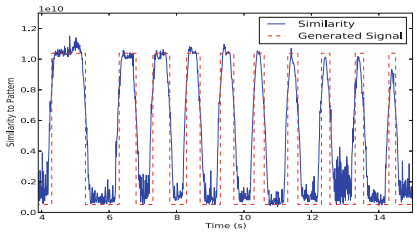
For each smartphone we recorded a self-generated, 33 s long vibration pattern containing different vibration signals at each location. On the one hand, we put a long vibration (2 s) into the pattern to find a hint of the alignment of the vibrations in noisy data. On the other hand, we also added very short vibrations (50 ms to 120 ms) to have data that matches real vibration durations in the login procedure. The cross-correlation was conducted in three steps:

1. We used a Fast Fourier Transformation (FFT) to obtain the frequency-amplitude-spectrum from the clean pattern.
2. In a loop we calculated the FFT from a slice of the recorded signal that has the same length as the clean pattern. For each iteration, we moved the starting point of this slice a predefined frame window ahead.

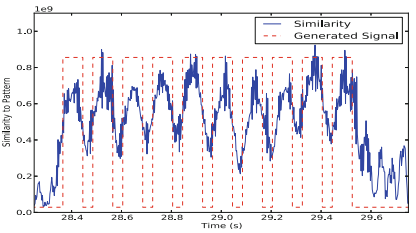
3. We cross-correlated frequency-amplitude-spectra of the clean pattern and the slice of the recorded signal over time to get the similarity of both patterns.

The result of the cross-correlation should aim at finding similarities in different audio signals to calculate the number of vibrations. This is a difficult task when the duration is short. On the one hand, the Nyquist-Shannon sampling theorem [20] concludes that a sampling rate of $2xs^{-1}$ with x as samples leads to a detection of x samples per seconds as maximum. So with a sampling rate of 44 100 Hz, we are only able to detect frequencies between 0 Hz and 22 050 Hz. On the other hand, the duration of the signal controls the resolution of the frequency-amplitude-spectrum obtained by the FFT. The shorter the duration of the signal is, the coarser-grained the result is. For a signal duration of one second, we obtain a resolution of 1 Hz. Since we can show that 90 % of all participants in our usability study can sense vibration durations of approximately 150 ms (cf., Sect. 5.2), we have to work with a resolution of 10 Hz. This coarse-grained resolution leads to a inaccurate cross-correlation especially because it is more difficult to filter out background noise. All in all, our self-generated signal should be easy to align by means of the single long vibration (2 s) and also practice-oriented because of the different short vibrations.

To compare different recordings, we conduct a cross-correlation between the two signals. As pattern we used a clean and clear vibration recorded in an anechoic room within a distance of 0.5 m. By amplifying the signal it could be easily eavesdropped by a human attacker.



(a) Using a window of 441 frames to visualize long vibrations.

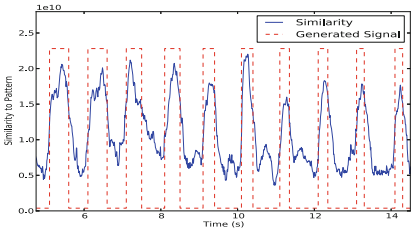


(b) Using a window of 110 frames to visualize short vibrations.

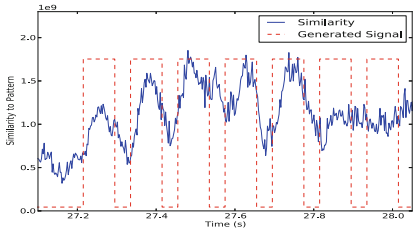
Fig. 2. Cross-correlation (solid line) between a clean vibration pattern (dashed line) of a Google Nexus S and a signal recorded in an anechoic room in a distance of 0.5 m while taking only frequencies between 150 Hz and 250 Hz into account.

As first experiment, we correlated a pattern from a Samsung Nexus S to a signal also generated in the anechoic room with the same smartphone. To visualize long vibrations we used a window of 441 frames or 10 ms to move across the signal (Fig. 2(a)). On the contrary, we used a window of 110 frames approximately 2.5 ms to obtain short vibrations (Fig. 2(b)). Note that the dashed lines delineate the clean vibration pattern while the solid lines trace the similarity we calculated. In conclusion, Fig. 2 shows that it is possible to use the acoustic

side channel the vibration motor produces to obtain the number of vibrations. In this experiment, we intentionally chose the anechoic room as location to show that this side channel can be exploited. We also evaluated other smartphones, namely a Google Galaxy Nexus, a LG L7 P700, a HTC Nexus One, and a Sony Xperia S with distances of 0.5 m, 2 m and 4 m with similar results.



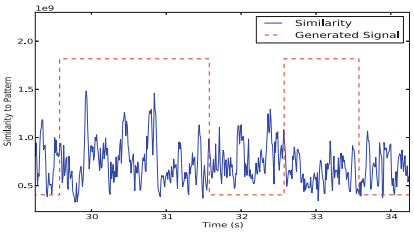
(a) Using a window of 441 frames to visualize long vibrations.



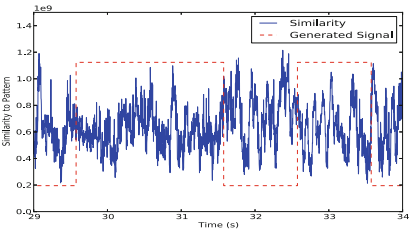
(b) Using a window of 100 frames to visualize short vibrations.

Fig. 3. Cross-correlation (solid line) between a clean vibration pattern (dashed line) of a Google Nexus S and a signal recorded on a corridor in front of an office environment in a distance of 0.5 m while taking only frequencies between 150 Hz and 250 Hz into account (Color figure online).

For the second experiment, we chose the corridor in front of an office environment as different location to record the generated signal. We started with a distance of 0.5 m between smartphone and microphone. In this setting, silent background noise was recognizable as well as keyboard noise coming from other offices and sometimes footsteps. Just like the first experiment, we were able to eavesdrop long vibrations (cf., Fig. 3(a)). Albeit, we were not able to fully reveal the generated signal for short vibrations. As one can see in Fig. 3(b), the correlation between the clean pattern and the recorded signal is not as significant as it should be to disclose the secret. Without the red bar we manually added afterwards to visualize our self generated pattern, the last two vibrations are not distinguishable from background noise. Due to the fact that the user would notice the recording of his authentication session when it is done in a distance of only 0.5 m and would be suspicious, we also recorded the self-generated pattern from a distance of 2 m and 4 m. As one can see in Fig. 4 for both correlations, no significance can be found for neither long nor short vibrations throughout to whole correlation. At this point—since we cannot even align the long vibration of 2 s—we are also not able to count the number of vibrations and are stuck. In a real attacker scenario, the attacker does not have a long vibration to align the login process. Hence, she needs to find short vibrations in the recorded signal which we were not able to find, despite the fact that we perfectly know the generated signal, but only had to align it in a range of some seconds. We repeated this experiment with all other smartphones and came to the same results for this location.



(a) The attempt to align the pattern to the generated signal with a distance of 2 m between smartphone and microphone.



(b) The attempt to align the pattern to the generated signal with a distance of 4 m between smartphone and microphone.

Fig. 4. Cross-correlation (solid line) between a clean vibration pattern (dashed line) of a Google Nexus S and a signal recorded on a corridor in front of an office environment while taking only frequencies between and 150 Hz and 250 Hz into account.

To make it even more difficult, we found another location that fits more to reality when authenticating against the smartphone being in front of our office building near a sparsely trafficked road. Outdoors, a user has to fear that attackers are shoulder surfing while walking near or behind him. Again, we generated and recorded the signals with all five smartphones having distances of 0.5 m, 2 m and 4 m. While cross-correlating the obtained signals with the clean pattern, we did not received any clue to detect the generated vibrations. Despite the highly directional kidney characteristic of the high-end microphone, the background noise in the significant frequency-range was too loud. Therefore, it was impossible to find any hints of vibrations in the signal.

In summary, we conclude that it is possible to eavesdrop the tactile feedback of Taco to attack a user's login in really silent environments. The attacker needs shoulder surfing in addition to the acoustical evaluation to obtain the user's secret. However, the requirements to gain the secrets are very high: The attacker not only needs a clean acoustic pattern of the smartphone, but also a situation where background noise is negligible *and* the distance between his microphone and the smartphone is short. We argue that an attacker with an expensive microphone trying to record the login would be suspicious for a victim.

4.3 Smudge Attacks and Shoulder Surfing

As the one-time pad effectively works as a random function in the set $\{0,\ldots,9\}^{\ell}$ for $\ell = 4\ldots 8$, all keys are equally alike. For this reason Taco is—in contrast to PINs, passwords, and patterns—secure against sophisticated smudge attacks [1]. Similarly, classical shoulder surfing does not reveal the secret. No matter how many cryptograms an attacker obtains, she cannot determine the underlying clear-text. Consequently, our scheme is secure as long as the attacker cannot read the secret key (i. e., the vibrations) at the same time as the cryptogram.

5 Usability Evaluation

In the following, we describe the usability evaluation of Taco.

5.1 Data Collection

Since Taco depends on the user's ability to perceive the number of vibrations, we investigated how many vibrations a user could differentiate in a given time interval. To accomplish this, we decided to develop a game as a smartphone application. Challenging authentication approaches encourage the user to practice the authentication a couple of times to learn it before actually using it. Therefore, the game should act as training the user's abilities on the one hand and observing the user's skill on the other hand.

We created two versions of the game: For the first version, we gave the player a predefined PIN $(1-2-3-4)$ he had to remember during the whole game before playing it. Letting a player choose his own PIN could result in two unsolicited situations: First, he could choose a random PIN that is hard to remember. In this case we would not evaluate the user's ability to utilize Taco, but to remember a sophisticated PIN. Second, he could choose a PIN that is too easy to remember and also too easy to work with like $(0-0-0-0)$ or $(1-1-1-1)$, which is more likely. Since recent studies have shown that user-chosen PINs as well as user-chosen passwords are far from being uniform distributed [6,15], we decided to give all players of this version a predefined PIN being easy enough to remember, but not too easy to require the execution of some basic calculations. This aspect was also important to have comparable results. We decided to give a player three "lives" in the game because three is the number of attempts real-world systems like debit or SIM cards and ATMs that use a PIN for authentication offer before the card is blocked for further usage.

To be able to compare also results for more sophisticated PINs, the second version of the game came with random but predefined PINs. Again, we gave the player the predefined PIN before playing the game, but we also added the PIN to the GUI. Displaying the random PIN on the GUI was important to receive meaningful data for the usability of Taco. Otherwise we would have challenged the players cognitive capabilities instead of evaluating the usability of Taco.

Both versions of the game are level-based and one game level equals one authentication attempt for Taco. Like for normal authentications, the user has to start the level by hitting a button. As a result, the smartphone vibrated randomly between zero and nine times. The player has to count the number of vibrations and add this to the first digit of the given PIN. Furthermore, he has to calculate the result modulo ten and enter the outcome into the smartphone in a round-based fashion. Afterwards the smartphone verifies the input. If it was incorrect, the player loses a life and stays in the same level. Otherwise the player reaches the next level having a decreased vibration duration. For three successful levels in a row the player obtains an additional life. By doing this we improved the player's immersion [7] and supported the learning phase so that

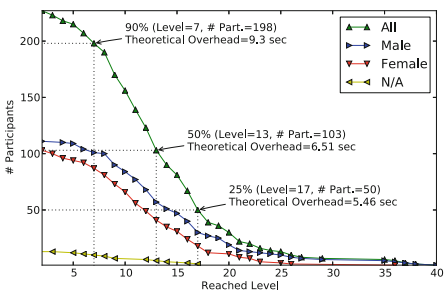


Fig. 5. Distribution of reached level grouped for male, female, n/a, and overall for both parts of the study.

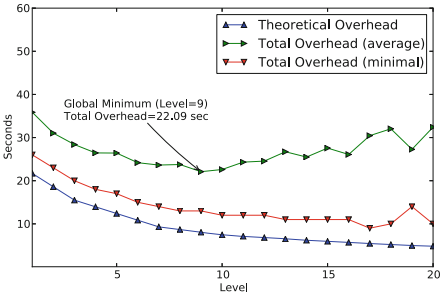


Fig. 6. Overhead for the additional security layer given as theoretical overhead and the total overhead as average and mean, ordered by level.

mistakes where punished with a loss of a life and successes were rewarded by one additional life.

Participants where recruited by simply asking them to take part. We did not gave them any reward to raise their willingness to participate. Before starting the game, we explained the details by playing a guided test level. If they did not knew how to do an addition mod 10, we told them to use a (normal) addition and use only the right digit of the result in case of a two digit result. Furthermore, we tried to implement the game as similar as possible to the actual authentication method to have comparable results. To figure out what vibration interval a participant could detect, we reduced the vibration interval with increasing level. We started with a duration d of 350 ms for a single vibration. Because the complete duration for a round should always be similar (cf., Sect. 4.1), we only modified the break between two vibrations. For a round of nine vibrations, we chose $b = d \cdot \frac{13}{20}$ as break while having a break before and after the first and last vibration as well. For a single vibration we chose $b = 7.2 \cdot d$ and for two vibration we chose $b = 4.5 \cdot d$. As a result, we always get a complete duration c for one round between $2 \cdot 7.2d + d = 15.4d$ and $10 \cdot \frac{13}{20} \cdot d + 10 \cdot d = 15.5d$. Other vibration numbers match accordingly. For the first level with $d = 350$ ms a round takes at most $350 \text{ ms} \cdot 15.5 = 5.425$ s without user interaction. We therefore get $c = 4 \cdot 350 \text{ ms} \cdot 15.5 = 21.7$ sec as complete duration for all four rounds. The duration was decreased in a stepwise fashion to exercise the player. To help the player to better detect vibrations that he had not recognized, we added a button to repeat the last vibration pattern.

5.2 Evaluation

To show the usability of our approach, we asked 187 people to play the first version of the game and 40 people the play the second version.

For Taco, the login duration takes the user's sensing capabilities into consideration. The more precise the user can feel the vibrations, the faster he can

login. Figure 5 shows that all users were able to login at least once for the longest vibration duration of 350 ms. Therefore, the overhead for vibrations and pauses for a full login is 21.7 s. Note that this does not include the user's calculation time and his response. To take this into account, Fig. 6 shows the complete overhead including the users' reactions. One can see that a full login procedure can on average be performed in less than 36 s if a user choses the longest vibration duration. Hence, every participant we asked was able to authenticate in less than 36 s without prior practice and thus our scheme outperforms existing approaches in this area.

Considering Fig. 5 again, one can see that 90 % of all participants reached Level 7. Since Level 7 uses a vibration duration of 150 ms, a full login results in a duration of approximately 22 s. As one can see in Fig. 6, the theoretical overhead decreases with decreasing vibration duration, but the average login time including user interaction decreases only till level 9 and increases afterwards. This is caused by the fact that shorter vibrations are more difficult to perceive. As a consequence, users have to reflect longer about their input.

5.3 Discussion

We conducted a usability evaluation to learn whether Taco can be used in the wild. To accomplish this, we designed a game that is very similar to the actual authentication process. We showed that Taco is usable and comprehensive since all participants were able to authenticate at least once. While we found significant advantages against comparable methods, we also have to admit that the timing overhead is the main disadvantage of Taco when compared to plain PIN authentication. However, such an overhead is inevitable when adding a secure channel to a user authentication. To the best of our knowledge, Taco has the lowest time overhead of all authentication methods that are resilient against shoulder surfing, comparable secure while staying usable to an average person.

6 Conclusion and Future Work

In this paper, we showed that the tactile feedback generated by a vibration motor of a smartphone can be used as a secure channel for user authentication. We introduced Taco, an enhancement to PIN authentication which mitigates the threat of shoulder surfing. For each digit of the PIN, Taco outputs a pseudo-random number of vibration signals. The user counts these signals, adds their number to the current digit of his PIN (mod 10), and inputs the resulting digit.

On the one hand, using this secure tactile channel causes a higher duration and more user's attention to authenticate. Even though, our usability study shows that 90 % of all participants had an authentication duration of less than 22 sonds. On the other hand, this procedure protects the user's PIN from leaking and is insusceptible to several realistic attacks which need to succeed in addition to a shoulder surfing attack. Timing attacks cannot measure the number of vibrations as we implemented Taco in such way that all vibration patterns

take the same time. However, we found that users need on average longer to add larger numbers having an even higher standard deviation so that there is no instant timing side channel. A long term study has to show whether users improve their skill over time when they get more familiar with Taco. Recording attacks require high-end audio recording equipment and are only feasible in a silent environment. But naturally in a silent environment shoulder surfing has a high risk to attract attention. Even if the user's input can be gathered (e. g., by camera) and high-end recording tools are available, we showed that it is hard to eavesdrop the vibration signals in real environments such as an office or outside a building.

References

1. Aviv, A.J., Gibson, K., Mossop, E., Blaze, M., Smith, J.M.: Smudge attacks on smartphone touch screens. In: WOOT (2010)
2. Aviv, A.J., Sapp, B., Blaze, M., Smith, J.M.: Practicality of accelerometer side channels on smartphones. In: ACSAC (2012)
3. Balzarotti, D., Cova, M., Vigna, G.: ClearShot: eavesdropping on keyboard input from video. In: IEEE Symposium on Security and Privacy (2008)
4. Bianchi, A., Oakley, I., Kostakos, V., Kwon, D.-S.: The phone lock: audio and haptic shoulder-surfing resistant PIN entry methods for mobile devices. In: Tangible and Embedded Interaction (2011)
5. Bianchi, A., Oakley, I., Kwon, D.S.: The secure haptic keypad: a tactile password system. In: CHI (2010)
6. Bonneau, J., Preibusch, S., Anderson, R.: A birthday present every eleven wallets? the security of customer-chosen banking PINs. In: Keromytis, A.D. (ed.) FC 2012. LNCS, vol. 7397, pp. 25–40. Springer, Heidelberg (2012)
7. Brown, E., Cairns, P.A.: A grounded investigation of game immersion. In: Extended Abstracts of Conference on Human Factors in Computing Systems (2004)
8. Cranor, L., Garfinkel, S.: Security and Usability: Designing Secure Systems That People Can Use. O'Reilly Media Inc., Sebastopol (2005)
9. Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G.J., Durand, F., Freeman, W.T.: The visual microphone: passive recovery of sound from video. ACM Trans. Graph. **33**(4), 79 (2014)
10. De Luca, A., Weiss, R., Drewes, H.: Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. In: Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (2007)
11. Forget, A., Chiasson, S., Biddle, R.: Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords. In: CHI (2010)
12. Gartner Research: Gartner Says Smartphone Sales Accounted for 55 Percent of Overall Mobile Phone Sales in Third Quarter of 2013 (2013). http://www.gartner.com/newsroom/id/2623415
13. Hayashi, E., Riva, O., Strauss, K., Brush, A.J.B., Schechter, S.E.: Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In: SOUPS (2012)
14. Kocher, P.C.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)

15. Murdoch, S.J., Drimer, S., Anderson, R.J., Bond, M.: Chip and PIN is broken. In: IEEE Symposium on Security and Privacy (2010)
16. Perković, T., Čagalj, M., Saxena, N.: Shoulder-surfing safe login in a partially observable attacker model. In: Sion, R. (ed.) FC 2010. LNCS, vol. 6052, pp. 351–358. Springer, Heidelberg (2010)
17. Riva, O., Qui, C., Strauss, K., Lymberopoulos, D.: Progressive authentication: deciding when to authenticate on mobile phones. In: USENIX Security Symposium (2012)
18. Roth, V., Richter, K., Freidinger, R.: A PIN-entry method resilient against shoulder surfing. In: CCS (2004)
19. Schneier, B.: Applied Cryptography: Protocols, Algorithms, and Source Code in C. Wiley, New York (1995)
20. Shannon, C.E.: Communication in the presence of noise. In: Proceedings of the Institute of Radio Engineers (IRE) (1949)
21. Tari, F., Ozok, A.A., Holden, S.H.: A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In: SOUPS (2006)
22. Uellenbeck, S., Hupperich, T., Wolf, C., Holz, T.: Tactile one-time pad: smartphone authentication resilient against shoulder surfing. Technical report, Horst Görtz Institute for IT-Security (HGI), HGI-2014-003, September 2014
23. Wilfong, G.T.: Method and Apparatus for Secure PIN Entry, 08 1999. Lucent Technologies Inc, U.S. Patent, US5940511 A
24. Yan, Q., Han, J., Li, Y., Zhou, J., Deng, R.H.: Designing leakage-resilient password entry on touchscreen mobile devices. In: Chen, K., Xie, Q., Qiu, W., Li, N., Tzeng, W.-G. (eds.) ASIACCS, pp. 37–48. ACM (2013)