Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Two-Thumbs-Up: Physical protection for PIN entry secure against recording attacks

DaeHun Nyang[a], Hyoungshick Kim[b], Woojoo Lee[c], Sung-bae Kang[a], Geumhwan Cho[b], Mun-Kyu Lee[a,*], Aziz Mohaisen[d]

[a] Department of Computer Engineering, Inha University, Incheon 22212, Korea
[b] Department of Computer Science and Engineering, Sungkyunkwan University, Suwon 16419, Korea
[c] Department of Statistics, Inha University, Incheon 22212, Korea
[d] University of Central Florida, Orlando, FL, USA

ARTICLE INFO

Article history:
Received 28 November 2017
Revised 18 May 2018
Accepted 20 May 2018
Available online 25 May 2018

Keywords:
Authentication
Personal Identification Number (PIN)
Smartphone
Recording attack
Physical shielding
User studies

ABSTRACT

We present a new Personal Identification Number (PIN) entry method for smartphones that can be used in security-critical applications, such as smartphone banking. The proposed "Two-Thumbs-Up" (TTU) scheme is resilient against observation attacks such as shoulder-surfing and camera recording, and guides users to protect their PIN information from eavesdropping by shielding the challenge area on the touch screen. To demonstrate the feasibility of TTU, we conducted a user study for TTU, and compared it with existing authentication methods (Normal PIN, Black and White PIN, and ColorPIN) in terms of usability and security. The study results demonstrate that TTU is more secure than other PIN entry methods in the presence of an observer recording multiple authentication sessions.

1. Introduction

Personal identification numbers (PINs) are a well-known and a widely utilized authentication method for many applications, including automated teller machines (ATMs), electronic door locks, and safes (Adams, 2011). Most smartphones today use PINs to protect private data against unauthorized use, lock the phone in order to control access (Chang and Ramachandran, 2016; Seol et al., 2016), or approve transactions (Mallat et al., 2004). However, their popularity and the inherent mobile feature of smartphones mean that the PIN-entry procedure is often vulnerable to various attacks, including human observation and recording attacks (Balzarotti et al., 2008; Shukla et al., 2014).

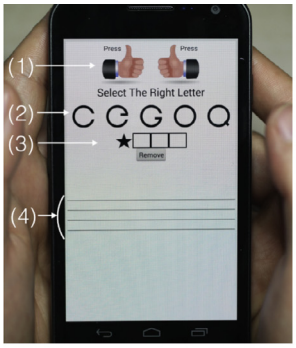
In those attacks, an adversary can observe the entry of PINs and reconstruct and reuse them later to authenticate himself, and access data or execute unauthorized transactions. Furthermore, recent advances in camera technology enabled attackers to use even small but highly accurate recording devices to enhance these attacks. For example, today's adversaries can be almost unnoticeable when executing such an attack by recording the entry of PINs in the wild. Also, attackers are today persistent, resulting in advanced attacks such as the "multiple-session recording", in which the attacker observes the entry of PIN or records it multiple times, thus improving his chances of reconstructing the PIN. Therefore, ide-

* Corresponding author.
E-mail addresses: nyang@inha.ac.kr (D. Nyang), hyoung@skku.edu (H. Kim), lwj221@inha.ac.kr (W. Lee), sbkang87@isrl.kr (S.-b. Kang), geumhwan@skku.edu (G. Cho), mklee@inha.ac.kr (M.-K. Lee), mohaisen@cs.ucf.edu (A. Mohaisen).
https://doi.org/10.1016/j.cose.2018.05.012
0167-4048/

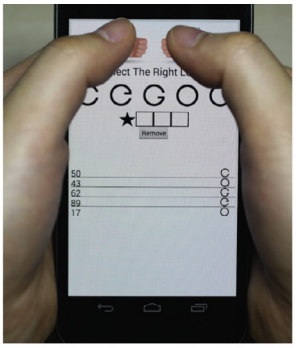
ally, a PIN-entry method should be resistant to both simple and multiple-session recording devices, as well as human observing attackers.

Prior Work. Various PIN-entry methods have been proposed (Bianchi et al., 2011; 2011; 2012; 2010; De Luca et al., 2010; 2007; 2009; Huh et al., 2015; Kumar et al., 2007; Kwon and Hong, 2015; Lee et al., 2016; Lee and Nam, 2013; Lee et al., 2016; Maeng et al., 2015; Nyang et al., 2014; Perković et al., 2009; Roth et al., 2004; Sasamoto et al., 2008; von Zezschwitz et al., 2015). These methods generally present a random challenge to the user, where the user is supposed to perform simple tasks involving this challenge and the secret PIN and calculate a response that varies between authentication sessions. Such session-specific response disables the reuse of observations in many cases although not totally addressing the observer or recorder attack (Tan et al., 2005) (see Section 2). This randomness and freshness also prevent a smudge attack, whereby finger smudges on a touchscreen are used by an attacker to track the user's secret input (Aviv et al., 2010; Kwon and Na, 2014; von Zezschwitz et al., 2013). However, previous methods do not fully satisfy various desirable requirements in PIN-entry systems (Zezschwitz et al., 2013). For example, to ensure acceptance by users and adoption by system administrators, various additional usability requirements should be met. First, it is desirable for a PIN-entry method to be compatible with the normal four-digit PIN and operable with only a multi-touch screen (i.e., without any secondary channel such as audio or vibration). Second, for an improved user experience, the speed and error rate of the input method should be within an acceptable range.

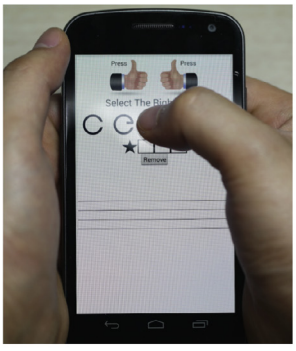
Goal and Approach. The goal of this work is to present a usable PIN-entry method for smartphones that is resistant to recording and observation attacks. To this end, in this paper we propose the Two-Thumbs-Up (TTU) PIN-entry method for smartphones. TTU requires only visual and touch interactions with a multi-touch screen. TTU cleverly utilizes multiple design ideas including 1) divided challenges protected by enforced handshield to thwart an observing and recording attacker and 2) a simple and straightforward challenge and response protocol for high usability and low mental demand. First, different from methods used in the literature that require an extra channel for communicating challenge and response or inputting PINs (Bianchi et al., 2011; 2011; 2012; 2010; De Luca et al., 2010; 2009; Kwon and Hong, 2015; Lee et al., 2016; Perković et al., 2009; Roth et al., 2004; Sasamoto et al., 2008), TTU asks a user very simple questions without using any extra channel. That is, given five (PIN digit, PIN digit, a letter) tuples that are randomly chosen for each PIN digit check, a user is given a question: "what is the matched letter for your PIN digit?". For example, assume that five random tuples (5, 0, C), (4, 3, Q), (6, 2, G), (8, 9, e), and (1, 7, O) are given. If a user's PIN digit is either 4 or 3, then a user's answer should be 'Q'. For 5 or 0, it should be 'C', which will be repeated to check all the four PIN digits. Thus, the answer is quite straightforward and no complex calculation or combination with information from extra channel is demanded. Second, TTU uses the "shielding" offered by the user's hands to hide "a divided challenge". In TTU, five tuples are placed from top to bottom at the lower part of the screen, but PIN digits are placed in the down left corner of the screen and the corresponding letters are placed in the down right corner as shown in Fig. 1(b). This separated challenge combined with TTU's handshield prevents an attacker from observing both sides of the five tuples, and thus frustrates the attack. To see the challenge, a user is asked to cover the phone with both hands, which prevents a shoulder-surfing/recording attacker from seeing both PIN digits and response letters. Depending on where the attacker is standing, it can see either PIN digit part or response letter part but not both, which will be shown by experiment in Section 5. Third, an activation mode is proposed where the challenge is only shown upon pressing an activation button by both hands, where the challenge is hidden by the hands of the user naturally from the observer. Lastly, a repeated challenge of the same PIN is produced to increase the security of the PIN entry system against the guessing attacks.



(a) Challenge inactive: Touchscreen is not touched.



(b) Challenge activated: challenge is seen only while both TTU buttons are pressed.



(c) User response: User chooses the right answer.

Fig. 1 – A high-level visual interface of the TTU system: From a first-person perspective (the user) highlighting the four different parts of the system, the challenge activation interface (providing shield from an observing adversary), and the user response interface.

Note that the approach proposed in this work is not suggested as a replacement to existing approaches to authentication, and can perhaps be utilized along with those approaches; e.g., fingerprint and facial authentication. Furthermore, the security of our approach may come at cost. For example some users who prefer to "utilize" only one hand to enter PIN, maybe because they do not want to drop whatever else they are doing

at the time may find our technique inconvenient, albeit at the risk of reduced security.

Security Features of the Proposed Work. Even without extra channels, such as audio or vibration, TTU is very secure against multiple-session recording attacks. It remains compatible with normal PINs consisting of an arbitrary number of digits. Not requiring any extra channel is a desirable feature for both usability and security. For example, it is well known in cognitive science that users perform worse when working with multiple sensory channels (Spence et al., 2001). In addition, such channels provide adversaries with an additional attack vector: a highly motivated attacker may be able to recover some information transmitted over an audio or vibration channel using only visual analysis (Davis et al., 2014).

TTU is Highly Usable. We conducted a user study for TTU compared with three existing authentication methods, Normal PIN, Black and White PIN (BWPIN) (Roth et al., 2004), and ColorPIN (De Luca et al., 2010), in terms of security and usability. The results of the study show that TTU provides a significantly greater security advantage over the three systems. As for the usability, the shortest average authentication time was achieved with Normal PIN (1.92 s), whereas the longest was for BWPIN (average 16.50 s). TTU had an average time of 10.42 s, and ColorPIN had an average of 8.03 s. Because the authentication times of BWPIN, TTU, and ColorPIN are significantly longer than that of the Normal PIN, we recommend that these systems of secure PIN entry should be limited to critical applications such as smartphone banking. With a method compatible with the normal four-digit PIN, users may switch their preferred PIN entry method in different security contexts without any memory burden to memorize a new type of PIN. For example, a user may use a Normal PIN at home, whereas in public places such as a coffee shop, she might prefer to use a secure PIN entry method such as TTU with the same PIN. Although ColorPIN requires slightly less time for authentication than TTU, it is not appropriate for the above scenario because it requires colors in addition to numbers and thus not compatible with the normal four-digit PIN.

Organization. The remainder of this paper is organized as follows. In Section 2 we review the related work. In Section 3 we explain the threat model and assumptions. In Section 4 we describe the TTU concept and implementation. In Section 5 we outline our user study. In Section 6 we provide a comparative evaluation of the normal PIN, Black and White PIN, ColorPIN, and TTU across multiple evaluation criteria, including the success rate, authentication time, and the number of trials for a successful authentication. In Section 7 we provide a comparative security analysis outlining the security level of the four PIN-entry methods. We conclude in Section 8 by outlining future directions.

2. Related work

Various PIN-entry methods have been proposed Bianchi et al. (2011, 2011, 2012, 2010); De Luca et al. (2010, 2007, 2009); Huh et al. (2015); Kumar et al. (2007); Kwon and Hong (2015); Lee et al. (2016); Lee and Nam (2013); Lee et al. (2016); Maeng et al. (2015); Nyang et al. (2014); Perković et al. (2009); Roth et al. (2004); Sasamoto et al. (2008); von Zezschwitz et al. (2015). For example, (Roth et al., 2004) proposed BWPIN, a method that presents the user with PIN digits as two sets by randomly coloring half in black and the other half in white. BWPIN requires users to enter which set the digit is in by pressing black and white key (thus the name BWPIN). The method relies on a binary decision for a challenge response, is very intuitive and is compatible with normal four-digit PINs. However, the BWPIN has an authentication time of over 23 s, affecting its usability. Also, it is vulnerable to the aforementioned recording attacks. Sasamoto et al. (2008) presented Undercover, which integrates visual challenges and tactile cues, and rely on the user's ability to simultaneously process multiple sensory inputs for authentication. However, Undercover requires a special haptic device for its operation, and uses a set of pictures instead of digits for a PIN (making Undercover incompatible with all traditional PIN-based authentication systems). Furthermore, Undercover has been shown to be insecure (Yan et al., 2012), where an attacker may obtain some information about the PIN from the challenge, rather than from the user's response.

De Luca et al. (2009) presented Vibrapass, an obfuscation technique that allows users to input an incorrect PIN digit on purpose when the phone vibrates. However, because the response supplied by Vibrapass always includes the correct PIN as a subsequence, attackers are able to mount an intersection attack to recover the correct PIN through multiple sessions. De Luca et al. (2010) also presented ColorPIN, a PIN-entry method that has a similar layout to a normal PIN pad and minimizes the additional overhead. However, because its PIN is composed of (PIN digit, color) pairs, users are required to remember significantly more information. In addition, because the observation attacks narrow down the number of PIN candidates from 531,441 to 81, recording two or more sessions uniquely identifies the correct PIN.

Recently, (Bianchi et al., 2011) proposed PhoneLock, a novel unimodal and non-visual PIN-entry method. Because PhoneLock directly transmits the challenge PIN digits through physically secured channels, using tools such as earphones and vibrations, it is resistant to recording attacks over multiple sessions. However, it may not be convenient to prepare the earphones or interrupt other activities (e.g., stop listening to music) whenever an authentication is required. Moreover, the haptic version of PhoneLock uses 10 distinct vibration patterns, presenting a very challenging task to the user. Its preliminary haptic version, Haptic Wheel (Bianchi et al., 2010), also showed a long task completion time and high error rate. Bianchi et al. (2011, 2012) presented counter-based variants to the haptic and audio cues, which require rapidly determining the number of cues presented in rapid temporal succession. However, the counter-based variants are also partially vulnerable to observation attacks because they leak partial information, such as the dial direction (Spinlock Bianchi et al., 2011; Bianchi et al., 2012), color (Colorlock Bianchi et al., 2012), and touch order (Timelock Bianchi et al., 2012). We note that there are other methods using eye-tracking (De Luca et al., 2007; Kumar et al., 2007), 3D visual interface (Lee et al., 2016; Lee and Nam, 2013), head-mounted display (Yadav et al., 2015), visual obfuscation (Lantz et al., 2015; Luca et al., 2013), brain-computer interface (Thorpe et al., 2005), biometric pattern analysis (De Luca et al., 2012) and a back-of-device touch panel (De Luca et al., 2013).

3. Threat model and assumptions

This section describes our threat model and assumptions. Traditional user authentication methods using finger movement on smartphones could be vulnerable to observation attacks such as shoulder surfing attack or recording attack. According to previous studies (Chiasson et al., 2012; Wiedenbeck et al., 2005), a single observation can be enough to disclose a password to a bystander. In this paper, we consider two types of observation attack: (1) shoulder-surfing attack and (2) camera recording attack since the effectiveness of the observation attack depends on the attacker's visibility of the entered PIN on the victim's mobile device. That is, the attacker's goal is to obtain a victim's PIN used to unlock the victim's mobile device (e.g., smartphone) by directly looking over the victim's shoulder or recording the entire unlock process through a high quality camera. In performing such an attack, attackers were able to use any tools or resources they wished during the attacks.

4. Two-Thumbs-Up: Concept and operation

In this section we present the concept and implementation issues of Two-Thumbs-Up (TTU). We start by outlining the main design (Section 4.1) and its rationale (Section 4.2), outline the response parameters (Section 4.3) and design improvements for better security guarantees (Section 4.4).

4.1. *Two-Thumbs-Up's design and operation*

The TTU system, shown in Fig. 1, consists of four parts: (1) TTU buttons (which look like "two thumbs up", thus is the name of the design) at the top to trigger a challenge, (2) five response buttons (labeled C, e, G, O, Q) in the middle of the system interface, (3) a PIN progress guide, showing which PIN digit the user is entering as an asterisk, and (4) the challenge area, where a challenge is divided into two parts and separately placed on the left and the right sides on the screen.

The challenge area consists of five rows. In this area, a challenge is only displayed while both the TTU buttons are being pressed by the user (challenge-activated mode), as shown in Fig. 1(b). The challenge disappears when either of the TTU buttons is released, as shown in Fig. 1(a). Fig. 1(b) shows the challenge-activated mode, where the challenges appear in the challenge area.

4.1.1. *Operation*

In the challenge, each row has two parts: on the left side are candidate PIN digits (two candidate PIN digits in each row, e.g., 50 in the first row representing two PIN digits of 5 and 0), and on the right side are the candidate response letters ('C' in the first row indicates that a user should tap the 'C' button if her PIN digit is either 5 or 0). Among the 10 candidate PIN numbers shown on the left side, there is only one correct PIN digit. The user finds a matching response letter in the same row, and presses the correct letter among the five buttons in the middle, as shown in Fig. 1(c). In Fig. 1(b), for example, assuming that the first PIN digit is 2, the corresponding correct letter is 'G', because both 2 and 'G' are on the third row. Thus, the user should press the letter 'G' in the middle for authentication (indicated by (2) in Fig. 1(a)). Similarly, if the user's first PIN digit was '7,' for example, the correct letter would be 'O.'

4.1.2. *Repeated challenges*

Because a PIN is composed of multiple digits, the user is asked to respond correctly to multiple challenges to complete one authentication session. For example, for a four-digit PIN, the user is asked to respond to the four challenges concerning the first to the fourth digits, and then to respond again for the first and second digits to match the guessing attack probability with the normal PIN entry method. Thus, a total of six responses per four-digit PIN are required. For example, if the PIN is 4083, the user will be challenged for the digits with the values of 4, 0, 8, 3, 4, 0. We note that this repeated authentication by requiring the user to input (again) responses for another challenge of the first and second PIN digits is rather for improving the security level: it reduces the success probability of a random guessing attack ($1/5^6 = 0.000064$) to less than that of the normal four-digit PIN entry method ($1/10^4 = 0.0001$). A detailed analysis of this improved security is presented in Section 7.

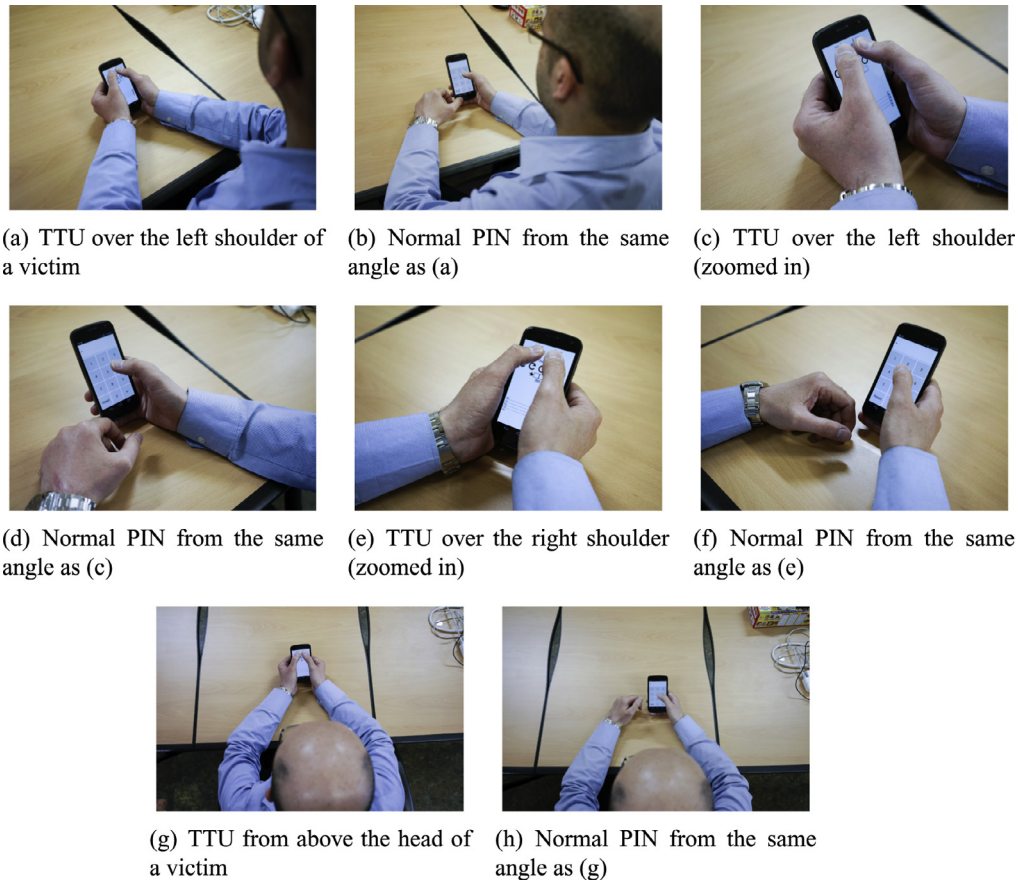
Our description of TTU in this paper focuses on four-digit PINs, which are widely as a convention in the literature (Bonneau et al., 2012; De Luca et al., 2010; Roth et al., 2004). Furthermore, four-digit PINs are still the most common choice for PIN authentication, although six-digit PINs are getting more acceptance (Wang et al., 2017). However, TTU can be easily adopted to cope with longer PINs. If TTU is used to enter an N-digit PIN, the number of required responses is $R \geq N\log_5 10 \approx 1.43N$ so that $1/5^R \leq 1/10^N$. For example, for a six-digit PIN, nine responses are required.

4.1.3. *"Physical Shield" and limited exposure*

Both the user and an attacker can see the challenge only for the moment that TTU is in the challenge-activated mode. Without the challenge, the user cannot identify the correct letter for a PIN digit, and therefore she must activate the challenge before PIN entry. After finding the matched letter in the challenge, the user presses the corresponding letter in the middle using her thumb. To press the correct letter button, the user must release either or both of the activation buttons, which makes the challenge disappear. The arrangement of PIN digits and candidate response letters is randomly determined for every PIN digit for a security reason, whereas the arrangement of the letter buttons is fixed for usability.

4.2. *Design rationale and ideas*

To prevent an observation attacker from learning a PIN in a challenge-and-response protocol where the PIN has a limited entropy, either the challenge or response must be hidden from the attacker. Many PIN-entry methods attempt to hide the challenge by sending it through an additional channel such as audio or haptic interaction, as highlighted in Section 2. Using these extra channels introduces a degree of inconvenience. The first idea in TTU, and different from methods in the literature, is that instead of using an extra channel we can indeed take advantage of the "shielding" offered by the user's hands to create "a divided challenge". In TTU, the challenge is



(a) TTU over the left shoulder of a victim

(b) Normal PIN from the same angle as (a)

(c) TTU over the left shoulder (zoomed in)

(d) Normal PIN from the same angle as (c)

(e) TTU over the right shoulder (zoomed in)

(f) Normal PIN from the same angle as (e)

(g) TTU from above the head of a victim

(h) Normal PIN from the same angle as (g)

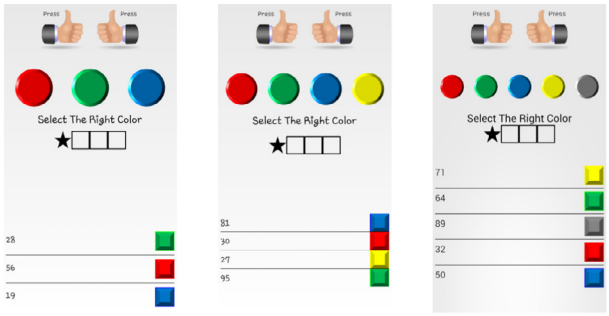
Fig. 2 – A visual comparison between the use of our method and the normal PIN entry method from the point of view of the adversary. The TTU in the challenge-activated mode vs. normal PIN entry from the perspective of an attacker highlight the benefits of the physical shield by the user to prevent observing the PIN digit value, even when zooming or alternating the observation angle.

divided into two parts: the candidate PIN digits on the left and the candidate response letters on the right. Two partial challenges comprise one complete challenge, and they are separately and randomly placed on the left and right sides of the screen. This separation prevents an attacker from observing one complete challenge.

The second idea behind TTU is that placing the activation (or TTU) buttons at the top of the screen enforces the user to use both thumbs to press them. This causes the thumbs and the backs of the hands to shield the divided challenge naturally and effectively. The left hand shields the PIN digits, and the right shields the candidate response letters. An observing attacker who would like to obtain the PIN must be able to see three components at the same time: the PIN digits and candidate letters (challenge) as well as the letter button pressed in response. Fig. 2 shows the attacker's view of TTU from various angles. Note that it is very challenging to see both the PIN digits and candidate letters in a challenge at the same time, whereas the PIN digits for normal PIN entry can be observed easily. Recent studies by (Yan et al., 2013) and (Kim et al., 2010) also used the position of the hands to cover the challenges. (Kim et al., 2010) proposed ShieldPIN and CuePIN, which require users to cover some part of the screen to hide PIN entry or a challenge. This "handshielding" is not forced but requested, whereas TTU's handshield is naturally compelled in order to see the challenges. These previous schemes were designed for desktop computers (Kim et al., 2010) or tablets (Yan et al., 2013), and are not directly applicable to the smartphone interface. This is because one hand is used for hiding the challenges and the other for pressing response buttons. Thus, the device cannot be held by hand, but must be laid on a table. Additionally, TTU's shielding of a divided challenge is in a stark contrast to that of Yan et al. (2013) in that, even when the attacker stands within one "visual cone," he cannot possibly see both parts of the challenge in TTU. The third idea is to minimize the duration of exposure of the challenge. For the user to press the corresponding letter button, she is compelled to release either or both of her thumbs, which makes the challenge disappear.

4.3. *Design of response letters*

In a pilot study, and in an attempt to reach an optimal design that strikes a balance between security and usability, we tested a number of designs for TTU, including those shown in Fig. 3. In those design options, the response buttons used colors instead of letters. To make the recognition of response letters by the attacker more difficult, specially designed



(a) 3-button prototype (b) 4-button prototype (c) 5-button prototype

Fig. 3 – The TTU system prototype with various design options, including replacing responses with colors instead of letters.

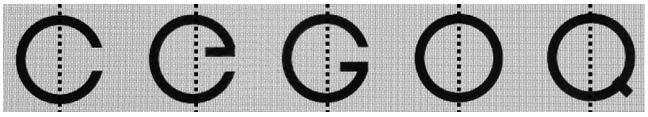


Fig. 4 – Left half of the TTU response letters are the same shape.

alphabetic letters were chosen for the response buttons. As shown in Fig. 4, the five letters (C, e, G, O, Q) used as response letters have the same shape on their left-hand side (half circle) and are the same color (black). As a result, we notice that even the partial exposure of the left half of the letter in a challenge does not help the attacker in recognizing the letter, because all five letters are the same color and have the same left-side shape.

4.4. *Design considerations for security*

For the prototype TTU design shown in Fig. 3, we have defined two parameters, N_r and N_d. N_r is the number of rows in the challenge area, which is equivalent to the number of response letters. On the other hand, N_d is the number of digits in each row. For example, the three prototypes shown in Fig. 3 have $(N_r, N_d) = (3, 2)$, (4, 2), and (5,2), respectively. In this section, we explain the rationale for our final choice of $(N_r, N_d) = (5, 2)$, as shown in Fig. 1(b).

Our choice is based on a theoretical security analysis considering two major attacks, the random guessing attack and the recording attack. For simplicity, here we only explain a simplified model in which the user enters one response for a one-digit PIN, assuming that the PIN digit has been randomly chosen from the set $\{0, 1, 2, \ldots, 9\}$. For a more precise security analysis as in Bonneau et al. (2012), Wang et al. (2017), Kim and Huh (2012), and Wang et al. (2017), a more realistic PIN distribution should be considered. We will provide this analysis in Section 7.3. In a random guessing attack, an attacker tries to pass an authentication test by guessing either a PIN digit or a valid response, i.e., a response letter button. In a recording attack, the attacker tries to record all challenge–response pairs with a camera.

Table 1 – Security of TTU for a single-digit PIN against (A) random guessing attack, and (B) multiple-session recording attack: *(2/10 x 2/3+8/10 x 1/3), **(6/10 x 2/4+4/10 x 1/4), † the best choices against each attack.

		$N_r = 3$	$N_r = 4$	$N_r = 5$
Success prob.	$N_d = 1$	1/3	1/4	†1/5
of attack A	$N_d = 2$	1/3	1/4	†1/5
	$N_d = 4$	*2/5	**2/5	2/5
# required sessions	$N_d = 1$	2	3	4
for attack B	$N_d = 2$	5	11	†∞
to reach prob.=1	$N_d = 4$	†∞	†∞	†∞

We note that TTU was designed to prevent an attacker from recording both PIN digits (left part) and candidate response letters (right part) in a challenge, although the user's responses may be fully recorded. Therefore, the attacker can choose which part to record, either the PIN digits or candidate response letters. It is easy to see that it is meaningless to record the candidate response letters. Therefore, the attacker's best strategy is to record the PIN digits in the challenge, and to extract some useful information from them. This task may be done across multiple PIN-entry sessions to allow attackers to narrow down the candidate PINs.

Table 1 shows the evaluated security of TTU for various values of N_r and N_d. As for random guessing attacks, we listed the probabilities that an attacker may guess a correct response. As for recording attacks, obviously the probability of an attacker's successful login becomes greater after recording than that of a random guessing attack, but it may vary according to the number of recorded sessions even for the same N_r and N_d. Therefore, to quantify the amount of advantage an attacker takes through recordings, we computed the numbers of sessions that an attacker has to record in order to uniquely identify a PIN digit and pass the authentication test with probability 1. In a nutshell, the smaller this number, the more advantage the attacker takes from the observation. The ∞ symbol implies that the attacker gains no information through recordings and never reaches the probability 1. In the following we elaborate on how the results in this table are constructed.

The figures for random guessing attacks in Table 1 were computed as follows.

1. If $N_r \times N_d \leq 10$, as in all three settings shown in Fig. 3, there should be only one occurrence of the correct PIN digit. Thus, there should be only one correct response letter button. In this case, the success probability of a random guessing attacker is $1/N_r$.
2. If $N_r \times N_d > 10$, there may be more than one correct response letter button, because the correct PIN digit may appear in multiple rows in the challenge. This requires a slightly more involved calculation to evaluate the success probability. For example, when $N_r = 3$ and $N_d = 4$, each row in a challenge contains four one-digit numbers, and there are three rows. Therefore, twelve numbers appear in total. Because a number cannot appear twice in the same row, eight numbers in the set $\{0, 1, 2, \ldots, 9\}$ appear in only one row, but the remaining two numbers appear twice, each time in a distinct row. If the correct PIN digit is one of the

eight numbers, the success probability is 1/3. If the correct digit is one of the two numbers that occur twice, the success probability is 2/3. Summing up the two cases, we may compute the success probability of guessing one response as $2/10 \times 2/3 + 8/10 \times 1/3 = 2/5$. The remaining elements in Table 1 have been calculated in a similar manner.

The figures for multiple-session recording attacks were computed as follows:

1. After recording the PIN digits in a single challenge session, the attacker may reduce the candidate set size for each PIN digit to $\min(10, N_r \times N_d)$, because the correct digit must always appear in the challenge.
2. After observing the second session, the attacker will obtain another candidate set with the same number of elements, and may intersect this with the first set.

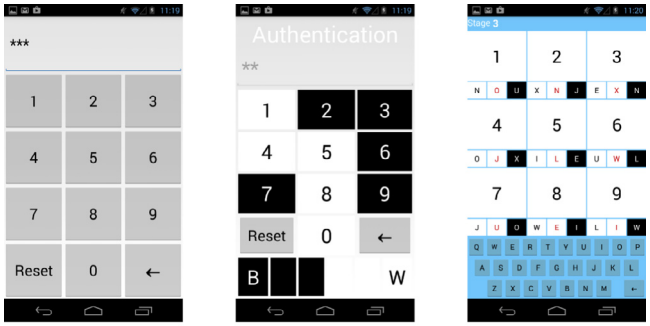
The figures in Table 1 are the break-even points when the expected number of elements in this intersection becomes less than or equal to 1. For example, for $(N_r, N_d) = (3, 1)$, the candidate set contains only three elements. Because the probability of a certain digit being selected in a set of size 3 in both of two independent random trials is $(3/10)^2 = 9/100$, the expected number of elements in the intersection is $9/100 \times 10 < 1$. However, if $N_r \times N_d \geq 10$, the candidate set always has ten elements, and thus the intersection of candidate sets for multiple sessions still has ten elements, regardless of how many sessions the attacker records. These cases are represented as ∞ in Table 1, and imply that a recording attacker cannot find any information about the PIN digit. Through the above analysis regarding the two kinds of attacks, we conclude that $(N_r, N_d) = (5, 2)$ is the best choice among the combinations in Table 1 from a security viewpoint.

5. User study

The user study and experiments were designed to demonstrate that an implementation of TTU is not only effectively secure against multiple recording attacks, but also usable. In a typical recording attack, an attacker discovers a password by looking and recording over the user's shoulder during a login process. The overall study procedure was designed similar to Zakaria et al. (2011), examining how secure the entered PINs are against recording attack.

5.1. *Study design*

To evaluate the applicability of TTU, and to make the study as realistic as possible, we simulated a typical password setup and login processes in public places (e.g., airport, cafe, and train). We assumed that the user does not move from a standard chair while typing. Note that this sedentary setting enables the attacker to monitor and record the user's PIN entry much more easily. In practice, however, users are able to input their PINs while moving around as they use the smartphone keyboard with both hands. We used a mixed-method design, an experiment and a follow-up questionnaire, to measure the security and usability of TTU. The study was conducted in a controlled laboratory environment to avoid any distractions. We compared TTU with the three representative authentication methods for mobile devices shown in Fig. 5: Normal PIN, BWPIN (Roth et al., 2004), and ColorPIN (De Luca et al., 2010).



(a) Normal PIN (b) BWPIN (c) ColorPIN

Fig. 5 – A visual illustration of the interfaces of three authentication methods that were used for a comparison with TTU.

A within-subject design was used. In this design, each participant was given a series of four authentication tasks (TTU, Normal PIN, BWPIN, and ColorPIN) to complete using an Android application on a Galaxy Nexus with a 4.65 inch ($\approx$11.8 cm) display screen. The order in which the authentication methods were tested within each group was varied over all 24 (=4!) possible combinations.

5.2. *Participants*

For a lab study, we recruited 24 participants who were familiar with smartphone by posting fliers about our study on bulletin boards in a campus. We clarified the academic motivation behind the experiment to encourage participants to focus fully on the study. After investigating the validity of their responses, each participant also received an $8 honorarium for completion of the user study. The participants were assigned the role of "victim" and were asked to enter each of the four assigned PINs in a different order in sequence.

The majority of respondents were males (92%) and were aged 18–29 (83%). All but one of the participants were right-handed. Half of participants were heavy smartphone users who spent more than 5 hs per day interacting with a phone. Thirteen of the participants (54%) had a university degree, and the remainder had a high school diploma. Table 2 provides further details on the demographics of the participants.

Two graduate students who were familiar with the tested authentication schemes were recruited to play the roles of "shoulder-surfers" and "recording-attackers" in attempting to identify the PINs entered by the victims (see Fig. 6). The attackers were trained for the task of observing PINs and strongly encouraged to remember the PINs by a $1 reward for every PIN that was successfully recovered.

5.3. *Procedures*

In a controlled laboratory environment (while sitting), each participant (i.e., victim) used an Android application to

Table 2 – The demographics of the participants in the lab study.

Age group	
18–24	29% (7)
25–34	67% (16)
35–44	4% (1)
Gender	
Male	92% (22)
Female	8% (2)
Hand	
Right-handed	96% (23)
Left-handed	4% (1)
Smartphone usage	
Over 5 hours	50% (12)
Under 5 hours	50% (12)
Highest level of education completed	
High school	46% (11)
Bachelors degree	42% (10)
Masters degree	8% (2)
Doctorate degree	4% (1)

Fig. 6 – Setup for shoulder-surfing and recording attacks.

perform a series of four authentication tasks (TTU, Normal PIN, BWPIN, and ColorPIN). All participants performed the tasks with an identical device (Galaxy Nexus with a 4.65 inch touchscreen) to marginalize the impact of screen size variation on performance, awareness, and user satisfaction.

5.3.1. *Training*

For each authentication task, before the actual test sessions, we briefly explained the procedure of the authentication scheme used in the task to each individual victim, assigned a randomly generated PIN to him/her, and allowed the victim to complete ten training sessions to practice logging in to the proposed system. For each training session, a maximum of three login trials using a given authentication method were allowed. Completing a session was defined as either succeeding within three trials or failing.

5.3.2. *Testing*

After the training sessions, each victim was asked to continuously perform the actual tests; ten test sessions were completed in the same manner. Note that the test sessions were identical to the training sessions. At the end of the ten test sessions, the participants were asked to share their experiences and opinions in an interview. After completing the ten sessions, the next authentication task was performed without a break.

5.3.3. *Attack*

Before the victim entered a PIN in each test session, the shoulder-surfers were asked to move to their best viewing position (e.g., to the left or right of the participant). While the participants entered each of the four PINs, the two shoulder-surfers observed the victim's login process; note taking was also encouraged. For the *recording attack*, the process was fully recorded with a digital camera directed toward the best viewing spot of the screen. The playback of the recorded video clip was totally under the attackers' control, allowing them to play and rewind the clip and to control the playback speed without any time limit.

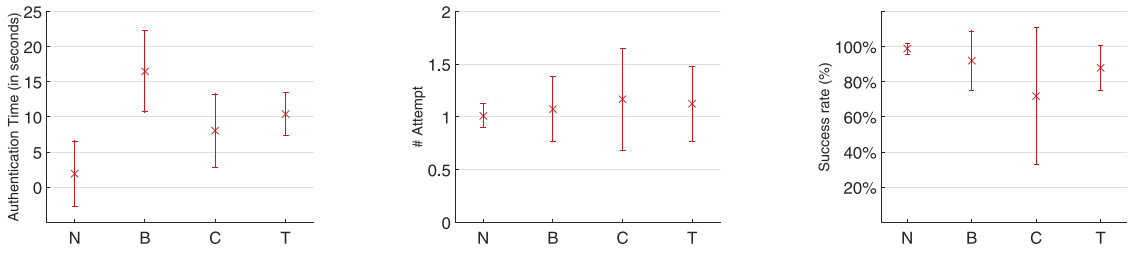
During the shoulder-surfing stage, each shoulder-surfer was asked to guess the victim's PIN. The attack was considered successful when the PIN given by the shoulder-surfer was identical to the victim's PIN. Shoulder-surfers who failed to successfully identify the victim's PIN in this stage moved to the recording attack. In this second stage, they were provided with an HD video recording showing a close-up of the entire login process. The video was shot without visual obstructions from less than 1m away from the user with a Canon EOS 6D camera, and the angle between the camera lens and the touchscreen was approximately 40° (see Fig. 6). The camera recording was employed because the use of a camera is the optimal shoulder-surfing technique in a public place, where the attacker's optimal position could change dynamically depending on the victim's position and/or location.

6. Usability evaluation

We evaluate the performance of four authentication methods by comparing them across multiple evaluation metrics: authentication time (Section 6.1), the number of attempts for successful authentication (Section 6.2), and successful authentication rate (Section 6.3).

6.1. *Authentication time*

We first measured the authentication time (i.e., the duration from the first to last touch when performing an authentication session). Fig. 7(a) shows the average authentication time and standard deviation of each method. To calculate the mean authentication time, we included only the successful sessions, and excluded the first session to reduce the bias from the initial generation. The shortest average authentication time was achieved with Normal PIN (1.92 s, standard deviation 4.63 s), whereas the longest was for BWPIN (average 16.50 s, standard deviation 5.77 s). TTU had an average time of 10.42 s with a standard deviation of 3.04 s, and ColorPIN had an average of 8.03 s with a standard deviation of 5.21 s. TTU ranked third in terms of the average authentication time, but had the smallest standard deviation. The authentication times of BWPIN, TTU, and ColorPIN are significantly longer than that of the Normal



(a) Authentication time (s) (b) Number of attempts (c) Authentication rate

Fig. 7 – Usability comparison between Normal PIN (N), BWPIN (B), ColorPIN (C), and TTU (T).

PIN, and so their application should be limited to security-critical areas such as smartphone banking.

We compared the authentication time of the four methods using a linear mixed model. This model includes one categorical factor variable with four levels (Normal PIN=1, BWPIN=2, ColorPIN=3, and TTU=4), two continuous variables (the session order (2, . . . , 10) and the order of four experimental conditions (1,2,3,4)) as fixed effects, and each individual's ability to implement the authentication as random effects. The random effects were introduced to explain the possible correlation between repeated measurements for each individual. We put a variable for the experimental order into the model to adjust for the learning effect. As the four authentication methods were statistically significant in the model (all p-values were less than 0.0001), post-hoc t-tests for all pairwise comparisons were performed using the Bonferroni correction to determine which pairs were different. The fastest authentication completion time was achieved with Normal PIN, followed by ColorPIN, TTU, and BWPIN in that order. This result is consistent with the visual impression in Fig. 7(a), which shows the average authentication time for the four methods.

6.2. *Number of attempts for successful authentication*

The number of attempts required for successful authentication was measured. Note that the number of attempts ranges from 1–3, because we limited the number of trials to three in the study. Thus, if a participant failed three times, the session was recorded as a failure and aborted. The results are shown in Fig. 7(b). With Normal PIN, the participants averaged 1.0125 attempts with a standard deviation of 0.1113. ColorPIN required an average of 1.1692 authentications with a standard deviation of 0.4838; for BWPIN, the average was 1.0733 with a standard deviation of 0.3069, and TTU averaged 1.1245 attempts with a standard deviation of 0.3559. We compared the mean number of attempts until authentication success using a Poisson mixed model with the categorical factor variable for the four authentication methods, two continuous variables (the session order and the order of four experimental conditions) as fixed effects, and each individual's ability to perform the authentication process as random effects.

As before, the random effects were introduced to explain the possible correlation between repeated measurements for each individual, and we included the experimental order variables to take into account the learning effect. Because a maximum of three attempts were allowed, some censoring issues might be of relevance to our situation, and our mean estimates could be underestimated. However, as the number of individuals who failed the authentication process were 0, 2, 3, and 9 for Normal PIN, BWPIN, ColorPIN, and TTU, respectively, the current situation would not be favorable for TTU, but advantageous to ColorPIN. Thus, with this consideration, the results obtained from the Poisson model can still be meaningfully interpreted. The four authentication methods were not statistically significant in the Poisson model (all p-values greater than 0.1), and post-hoc t-tests for all pairwise comparisons using Bonferroni correction showed no different pairs.

6.3. *Successful authentication rate*

We measured the successful authentication rate, defined as the number of successful authentication sessions normalized by the total number of trials. Fig. 7(c) shows the results for all participants. We noticed that Normal PIN achieved an average rate of 98.86% and a standard deviation of 3.07, which is the best in this experiment. In contrast, ColorPIN performed worst, with an average rate of 71.82% and standard deviation of 38.91. With TTU, the average success rate was 88.05% with a standard deviation of 12.70. Finally, BWPIN achieved an average success rate of 92.00% with a standard deviation of 0.1686.

We compared the successful authentication rates of the four methods using a binomial mixed model with the same explanatory variables as in the Poisson mixed model. Because the four authentication methods were statistically significant in the model (all p-values less than 0.005), post-hoc t-tests were performed for pairwise comparisons. Normal PIN showed the highest success rate, BWPIN and TTU achieved comparable results, and ColorPIN exhibited the lowest success rate. This result is consistent with the visual impression in Fig. 7(c), which shows the authentication success rate for the four methods.

6.4. *On the PIN compatibility*

TTU and BWPIN are fully compatible with standardized numeric PINs. However, ColorPIN is not compatible with a normal PIN because it uses a different type of PIN consisting of a combination of four-digit numbers and four colors. In addition, ColorPIN does not accept the digit '0.' The use of colors as part of a PIN also raises a memorability issue. That is, colors should be memorized in the ColorPIN user's long-term memory. Although TTU also uses colors, they are only used

temporarily as an input interface. Users do not need to memorize the colors.

PIN compatibility is important from several aspects. First, it is closely related to deployment costs. If we use the same data format as in existing systems, the underlying databases will not have to be changed. Therefore, we can use a new authentication method without increasing deployment costs. Second, the same password can be used for multiple purposes without any additional memory burden placed on the user. According to recent studies (Hayashi et al., 2013; 2012; Riva et al., 2012), many smartphone users want to use more than one authentication mechanism in different security contexts. For example, at home (which is generally secure against observation attacks), a typical user prefers to use a normal PIN entry method, whereas in public places such as a coffee shop, she might prefer to use a secure PIN entry method. As it imposes no extra memory burden for security, TTU can easily be applied in this scenario. TTU users can freely switch their preferred PIN entry method between normal PIN and TTU with the same PIN.

6.5. *Participants' Feedback*

After conducting the lab experiment, we asked participants to freely give their opinions about the advantages and disadvantages of the authentication methods.

Many participants felt that both ColorPIN and TTU provide high security compared with the other methods; 22 (91.7%) and 15 (62.5%) participants, respectively, chose security as one of the advantages of these approaches, whereas only nine (37.5%) participants believed BWPIN to be secure. For normal PIN, 13 (54.1%) participants chose the ease of input as their most-preferred feature, and nine (37.5%) participants chose memorability as the reason to choose the normal PIN method. However, many participants were concerned about the security of normal PIN; 15 (62.5%) participants felt that PIN information can easily be revealed by eavesdroppers; four (16.7%) participants were worried that the PIN could easily be guessed. Interestingly, one participant believed that PIN information could be correctly guessed from the victim's hand position alone, without observing the smartphone screen, as the position of the PIN keypad is predictable.

Unsurprisingly, most participants were concerned about the usability of the authentication methods designed for security against observation attacks; 23 (95.8%), 19 (79.2%), and 15 (62.5%) participants complained about the inconvenience of using ColorPIN, TTU, and BWPIN, respectively. For TTU, four (16.7%) participants were concerned that both hands must be used simultaneously to perform operations; three (12.5%) participants were dissatisfied with the slow authentication time. However, only one participant was concerned about memorability, unlike ColorPIN, for which 11 (45.8%) participants were worried about the memory burden of remembering both numbers and colors. On the contrary, two participants found it easy to use TTU because there is no additional memory burden when using TTU for authentication. To address the concern that both hands must be used for TTU, we recommend that TTU should be used for security-critical applications (e.g., banking), rather than every-day use, including for simply unlocking a smartphone.

Table 3 – Success probabilities of attacks for all PIN items.

Attack Type	Normal PIN	BWPIN	ColorPIN	TTU
PIN space	10,000	10,000	531,441*	10,000
Guessing	**1/10,000**	**1/10,000**	1/6,561	1/2,500
SSA	100.00%	41.30%	**0.00%**	4.35%
Recording	—	92.59%	78.95%	**27.27%**

* ColorPIN requires $2(\simeq \log_{10}53.1)$ more PIN digits.

7. Security analysis

In this section, we analyze the security of TTU and compare it with that of three existing schemes. The security is analyzed in terms of the theoretical PIN space size and the success probabilities of a random guessing attack, a shoulder-surfing attack (without a camera), and a recording attack (shoulder-surfing with a camera). The results are summarized in Table 3. In this table, Guessing, SSA, and Recording represent the success probabilities of random guessing attacks, shoulder-surfing attacks without a camera, and recording attacks with a camera, respectively.

7.1. *PIN space size*

The PIN space size can be easily calculated from the description of each authentication method. Because BWPIN and TTU use the same PIN spaces as Normal PIN, their space sizes are 10,000 for a four-digit PIN. Each PIN item of ColorPIN is composed of a pair of (digit, color). Because each digit is selected from $\{1, 2, \ldots, 9\}$ and each color is selected from {black, white, red}, the number of possible combinations for a PIN item is 27. Therefore, ColorPIN has $27^4 = 531,441$ possibilities for a four-item PIN, which is obviously advantageous against shoulder-surfing attacks because the attacker must memorize more information. Instead of {black, white, red}, three numbers (e.g., {0, 1, 2}) could be used for ColorPIN, but the effect of such alternatives has not been investigated.

7.2. *Random guessing*

Regarding random guessing attacks, an attacker may choose the better among two possible approaches. The first approach is to randomly select a PIN from the PIN space and try to pass the authentication test with this guess. However, this strategy is not always optimal. For example, ColorPIN asks the user to select a response from nine letters, not from 27 letters, to prevent shoulder-surfing attacks. Thus, it would be better for an attacker to guess a correct response than to guess a correct PIN item, as the success probabilities are 1/9 and 1/27, respectively. However, in BWPIN, the probability of an attacker successfully guessing four correct responses for a single digit is $1/2^4$. In this case, it would be better for the attacker to directly guess a PIN item with a probability of 1/10. Considering both approaches, the success probability of a random guessing attack is $max(P_1, P_2)$, where P_1 and P_2 are the success probabilities of guessing a PIN item and guessing a response, respectively. e only exception is TTU, where the first and second items are entered twice. In this case, an attacker may adopt

a hybrid approach. That is, for the third and fourth PIN digits, the attacker would guess correct responses, with a success probability of $1/5^2$. However, because each of the first and second digits requires two responses, the attacker may try to guess the correct PIN digit, not the response, which has a success probability of $1/10^2$. As a result, the success probability of this hybrid guessing attack is 1/2,500. The data in Table 3 were calculated in this way.

7.3. Dictionary attack

In the user study of Section 5, each participant was given a machine-generated random 4-digit PIN, and the analysis result shown in Table 3 is also based on this setting. To evaluate the security of TTU PINs against guessing attacks where a non-uniform PIN distribution is considered, we performed another lab study with 60 participants. The user study was designed to specifically measure the guessing entropy of user chosen TTU PINs. We recruited 30 participants from two universities, respectively. All participants provided their informed consent and were compensated about $2 (USD) for their participation. In the user study, we asked participants to choose 4-digit TTU PINs for an imaginary service they would frequently use on their smartphones.

With the collected TTU PINs, we calculated *partial guessing entropy* estimates (Bonneau, 2012) which is a popularly used technique for estimating the average number of trials needed to successfully guess a *fraction* (α) of an entire password set. For $0 \leq \alpha \leq 1$, let $\mu_\alpha = min\left\{ j | \sum_{i=1}^{j} p_i \geq \alpha \right\}$ where p_i is the probability of ith element occurring in non-increasing order, and let $\lambda_{\mu_\alpha} = \sum_{i=1}^{\mu_\alpha} p_i$, which is the actual fraction covered. With those notations, partial guessing entropy is defined as $G_\alpha(\chi) = (1 - \lambda_{\mu_\alpha}) \cdot \mu_\alpha + \sum_{i=1}^{\mu_\alpha} i \cdot p_i$, where χ is the probability distribution of PINs. In a nutshell, a larger value of $G_\alpha(\chi)$ for a given α indicates that the PIN distribution according to χ is more secure against a dictionary attack. The traditional guessing entropy is a special case of partial guessing entropy with $\alpha = 1$.

Because our collected set of 4-digit TTU PINs only represents a small portion of the theoretically possible password space, we employed the 2-gram Markov model to estimate the occurrence likelihood of every possible 4-digit TTU PIN. To cover rare N-gram cases, we particularly used the *Laplace smoothing* approximation technique – the frequency of each N-gram is incremented by one. The Markov model is one of the most representative probabilistic password models to evaluate the guessability of passwords (Ma et al., 2014). For more intuitive comparison of entropy estimates, entropy estimates can be represented in "bits of information." This conversion can be done as follows:

$$\tilde{G}_\alpha(\chi) = \log\left(\frac{2 \cdot G_\alpha(\chi)}{\lambda_{\mu_\alpha}} - 1\right) + \log\frac{1}{2 - \lambda_{\mu_\alpha}}$$

For comparison purposes, we also calculated the partial guessing entropy of the existing authentication methods: 4-digit PINs, 6-digit PINs and Android patterns. As for the traditional 4-digit PINs, we used a PIN dataset consisting of 204,508 PINs that was collected through an iPhone application (Kim and Huh, 2012). As for the traditional 6-digit PINs, we extracted 383,914 6-digit PINs from the popularly known "RockYou" (14 million) and "Yahoo" (0.5 million) password datasets. We constructed a 5-gram Markov model with those PINs to estimate the guessing entropy of 6-digit PINs. As for the Android patterns, we used an Android pattern dataset (Cho et al., 2017).

The partial guessing entropy (in bits of information) results are shown in Table 4. 4-digit TTU PINs showed higher guessing entropy estimates than the traditional 4-digit PINs. As α increases, the entropy estimate difference between TTU PINs and traditional 4-digit PINs decreases, but we can still see that TTU PINs are more secure against guessing attacks even when α is large. However, we remark that the partial guessing entropy of 4-digit TTU PINs were lower than those of traditional random 4-digit PINs, i.e., 13.29. For small α, they were also lower than the partial guessing entropy of random 4-digit TTU PINs considered in the previous subsection, which is 11.29. This implies that, in particular for small α, dictionary attacks will be more effective than random guessing.

Interestingly, when $\alpha = 0.1$, 4-digit TTU PINs showed a higher guessing entropy estimate than the traditional 6-digit PINs. This implies that our entropy estimates through the N-gram Markov model tends to be quite overestimated due to the limited size of 60 TTU PINs. Therefore, we note that the number of samples for the N-gram Markov model should be increased for obtaining more robust guessing entropy results.

7.4. Shoulder-surfing and recording

Regarding shoulder-surfing attacks and recording attacks, two participants played the role of attackers to successfully capture victims' PINs (see Section 4). Their attack results are shown in Fig. 8, and indicate that TTU and ColorPIN are secure against shoulder-surfing attacks and TTU is secure against recording attacks. For Normal PIN, the attackers did not need to perform recording attacks, because all 24 victims' PINs were successfully recovered by shoulder-surfing attacks alone. For BWPIN, the first attacker recovered seven out of 23 PINs through shoulder-surfing attacks and mounted recording attacks for the remaining 16 PINs. (One of the 24 participants failed to pass the authentication task for BWPIN. Therefore, we could not collect valid session data for these attacks.) As a result, 15 of the 16 PINs were recovered. Thus, the success probabilities for the first attacker were $7/23 = 30.43\%$ and $15/16 = 93.75\%$ for shoulder-surfing attacks and recording attacks, respectively. The second attacker recovered 12 PINs in the first stage and 10 of the 11 remaining PINs in the second stage. Thus, the success probabilities were $12/23 = 52.17\%$ and $10/11 = 90.91\%$ for shoulder-surfing attacks and recording attacks, respectively. These results are plotted in Fig. 8. The probabilities in Table 3 were then computed as $(7+12)/(23+23) = 41.30\%$ and $(15+10)/(16+11) = 92.59\%$ by merging the results of the two attackers. The figures for ColorPIN and TTU were evaluated in the same way.

7.5. Comparison

The analysis results presented in Table 3 intuitively show that TTU is significantly more secure against recording attacks than the other authentication methods. In particular, Normal PIN is very insecure, i.e., PINs are revealed by shoulder-surfing

Table 4 – Comparison of partial guessing entropy estimates (in bits of information) with α.

Authentication scheme	α									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
4-digit TTU PINs	10.76	11.08	11.32	11.51	11.69	11.85	11.99	12.12	12.23	12.30
4-digit PINs Kim and Huh (2012)	5.19	7.04	8.37	9.38	10.08	10.63	11.08	11.44	11.70	11.83
Random 4-digit PINs (U_{10000})	13.29	13.29	13.29	13.29	13.29	13.29	13.29	13.29	13.29	13.29
6-digit PINs	10.71	13.32	14.03	14.50	14.92	15.36	15.86	16.49	17.14	17.53
Random 6-digit PINs ($U_{1000000}$)	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93
Android patterns Cho et al. (2017)	5.04	5.82	6.54	7.19	7.86	8.50	9.20	9.97	11.00	12.71
Random Android patterns (U_{389112})	18.57	18.57	18.57	18.57	18.57	18.57	18.57	18.57	18.57	18.57

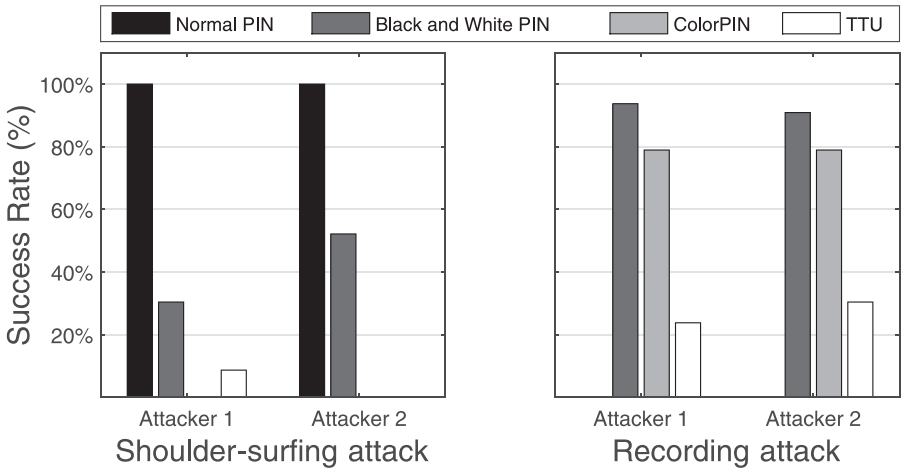


Fig. 8 – Attack performance of two attackers.

attacks, even without a camera. BWPIN is also very insecure. Almost all PINs were recovered by recording attackers, and almost half of the shoulder-surfing attacks without a camera were successful, which coincides with the observations of several previous studies, e.g., (Kwon et al., 2014; Lee, 2014; Maggi et al., 2011; Raguram et al., 2011).

As for ColorPIN, no human-eye-only attacks were successful, because the attackers found it very difficult to deal with the large amount of information given in a challenge screen, such as that shown in Fig. 5(c), in real time. However, the video recording dramatically enhanced the attack performance, and enabled the attackers to recover almost 80% of the PINs. We can also see that intersection attacks, in which data are accumulated across multiple sessions, are powerful, although recording a single authentication session of ColorPIN only gives a success probability of $1/3^4 \approx 1.23\%$. However, TTU is quite resistant to the same attack. It is very encouraging that TTU significantly enhances the resistance to recording attacks, because it is reasonable to assume that a well-prepared attacker would always have a camera available: currently, even low-end smartphones are equipped with cameras whose functionality is sufficient to record the full details of a PIN-entry session.

To verify whether the above intuitive interpretation of the attack results is reasonable, a statistical analysis was performed to compare the success rates of attacks against the four authentication methods. Our analysis was based on pairwise comparisons between pairs of attack success proportions with multiplicity correction. Two attack types, shoulder-surfing and recording attacks, were considered. Table 5 lists the pairs that exhibit significant differences after Bonferroni correction. The most notable point is that TTU permits the fewest successful recording attacks, regardless of attacker and evaluation type. However, for shoulder-surfing attacks, ColorPIN has the lowest successful attack rate, although this is obtained by sacrificing PIN compatibility and demanding more memory of the user, followed by TTU, BWPIN, and Normal PIN, generally in that order.

Table 5 – Comparing the attack success rates for the four authentication methods.

Attacker ID	Attack type	Statistical test (p-value < 0.05)
R	SSA	$N > T, N > B, N > C$
R	RA	$B > T, C > T$
L	SSA	$N > T, N > B, N > C, B > T, B > C$
L	RA	$B > T, C > T$

Normal PIN = N, BWPIN = B, ColorPIN = C, and TTU = T. SSA = Shoulder-surfing attack and RA=Recording attack. In the last column, "$x > y$" denotes that the attack success rate against authentication method x was significantly greater than that against authentication method y (at $\alpha = 0.05$).

To check whether the performance of the two attackers was comparable, simple logistic regression models were employed to examine the attack success rates. The null

hypothesis is that the two attackers do not show any difference in attack success rate for the four authentication methods. We did not observe any statistically significant difference between them for any combination of attack type and evaluation type.

8. Conclusion and future work

We have presented a new PIN entry method called Two-Thumbs-Up. In terms of security against recording attacks, TTU is significantly better than normal PIN, Black and White PIN, and ColorPIN. In terms of usability metrics such as success rate and authentication time, normal PIN performs best, but TTU represents a viable alternative in security-critical settings.

In summary, TTU can be a reasonable solution when users expect a high level of security for their applications (e.g., banking). Unlike existing authentication methods, TTU is highly secure against multiple-session recording attacks. TTU's higher level of security is achieved without sacrificing PIN compatibility and without requiring any extra hardware. Considering that TTU is designed for a smartphone in the form of a software application, TTU and normal PIN can be used selectively depending on the security context. We leave an investigation of a secure alphanumeric password entry version of TTU and research on the memorability for future work. In addition, it would be interesting to examine how the learning effect may improve the user performance of TTU over time.

Acknowledgments

This work was supported in part by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. B0717-16-0114 (Development of Biometrics-based Key Infrastructure Technology for On-line Identification), and No. 2017-0-00380 (Development of next generation user authentication)), in part by the ICT R&D program (No.2017-0-00545), and in part by Global Research Lab. (GRL) Program of the National Research Foundation (NRF) funded by MSIT (NRF-2016K1A1A2912757).

REFERENCES

Adams C. Personal identification number (PIN). Encyclopedia of cryptography and security. Springer; 2011. p. 927.

Aviv, A. J., Gibson, K., Mossop, E., Blaze, M., & Smith, J. M. (2010). Smudge attacks on smartphone touch screens. Proceedings of USENIX conference, WOOT 2010.

Balzarotti D, Cova M, Vigna G. Clearshot: eavesdropping on keyboard input from video. Proceedings of IEEE symposium on security and privacy, 2008.

Bianchi, A., Oakley, I., Kostakos, V., & Kwon, D. S. (2011). The phone lock: audio and haptic shoulder-surfing resistant pin entry methods for mobile devices. Proceedings of TEI 2011. ACM, (pp. 197–200).

Bianchi, A., Oakley, I., & Kwon, D. S. (2011). Spinlock: a single-cue haptic and audio PIN input technique for authentication. In Proceedings of HAID 2011, Series LNCS (pp. 81–90). Springer (vol. 6851).

Bianchi A, Oakley I, Kwon DS. Counting clicks and beeps: exploring numerosity based haptic and audio pin entry. Interact. Comput. 2012;24(5):409–22.

Bianchi, A., Oakley, I., Lee, J. K., & Kwon, D. S. (2010). The haptic wheel: design & evaluation of a tactile password system. Proceedings of CHI 2010 extended abstracts, ACM (pp. 3625–3630).

Bonneau J. The science of guessing: analyzing an anonymized corpus of 70 million passwords. Proceedings of the IEEE symposium on security and privacy, Series SP '12; 2012. p. 538–52.

Bonneau J, Preibusch S, Anderson R. A birthday present every eleven wallets? The security of customer-chosen banking PINs. Proceedings of financial cryptography, Series LNCS. Springer; 2012. p. 25–40.

Chang V, Ramachandran M. Towards achieving data security with the cloud computing adoption framework. IEEE Trans Serv Comput 2016;9(1):138–51.

Chiasson S, Stobert E, Forget A, Biddle R, Oorschot PCV. Persuasive cued click-points: design, implementation, and evaluation of a knowledge-based authentication mechanism. IEEE Trans Depend Secur Comput 2012;9(2):222–35.

Cho, G., Huh, J. H., Cho, J., Oh, S., Song, Y., & Kim, H. (2017). Syspal: system-guided pattern locks for android. Proceedings of the IEEE symposium on security and privacy, Series SP '17, (pp. 338–356).

Davis A, Rubinstein M, Wadhwa N, Mysore GJ, Durand F, Freeman WT. The visual microphone: passive recovery of sound from video. ACM Trans Graph 2014;33(4) pp. 79:1–79:10.

De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. Touch me once and I know it's you!: implicit authentication based on touch screen patterns. Proceedings of CHI 2012. ACM; 2012. p. 987–96.

De Luca, A., Hertzschuch, K., & Hussmann, H. (2010). ColorPIN: securing PIN entry through indirect input. Proceedings of the CHI 2010. ACM, (pp. 1103–1106).

De Luca, A., Weiss, R., & Drewes, H. (2007). Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. Proceedings of the 19th Australasian conference on computer-human interaction: entertaining user interfaces. ACM, (pp. 199–202).

De Luca A, von Zezschwitz E, Hußmann H. Vibrapass: secure authentication based on shared lies. Proceedings of the CHI 2009. ACM; 2009. p. 913–16.

De Luca A, von Zezschwitz E, Nguyen NDH, Maurer ME, Rubegni E, Scipioni MP, Langheinrich M. Back-of-device authentication on smartphones. Proceedings of the CHI 2013. ACM; 2013. p. 2389–98.

Hayashi E, Das S, Amini S, Hong J, Oakley I. CASA: context-aware scalable authentication. Proceedings of the ninth symposium on usable privacy and security, Series SOUPS '13, 2013.

Hayashi E, Riva O, Strauss K, Brush AJB, Schechter S. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. Proceedings of the eighth symposium on usable privacy and security, Series SOUPS '12, 2012.

Huh JH, Kim H, Bobba RB, Bashir MN, Beznosov K. On the memorability of system-generated PINs: can chunking help?. Proceedings of the symposium on usable privacy and security, 2015.

Kim D, Dunphy P, Briggs P, Hook J, Nicholson JW, Nicholson J, Olivier P. Multi-touch authentication on tabletops. Proceedings of the CHI 2010. ACM; 2010. p. 1093–102.

Kim H, Huh JH. PIN selection policies: are they really effective? Comput Secur 2012;31(4):484–96.

Kim H, Huh JH. PIN selection policies: are they really effective? Comput Secur 2012;31(4):484–96.

Kumar M, Garfinkel T, Boneh D, Winograd T. Reducing shoulder-surfing by using gaze-based password entry. Proceedings of the SOUPS '07; 2007. p. 13–19.
Kwon T, Hong J. Analysis and improvement of a pin-entry method resilient to shoulder-surfing and recording attacks. IEEE Trans Inf Forensics Secur 2015;10(2):278–92.
Kwon T, Na S. Tinylock: affordable defense against smudge attacks on smartphone pattern lock systems. Comput Secur 2014;42:137–50.
Kwon T, Shin S, Na S. Covert attentional shoulder surfing: human adversaries are more powerful than expected. IEEE Trans Syst Man Cybern: Syst 2014;44(6):716–27.
Lantz P, Johansson B, Hell M, Smeets B. Visual cryptography and obfuscation: a use-case for decrypting and deobfuscating information using augmented reality. Proceedings of the 1st workshop on wearable security and privacy (In association with Financial Crypto 2015), 2015. Paper 11
Lee MK. Security notions and advanced method for human shoulder-surfing resistant PIN-entry. IEEE Trans Inf Forensics Secur 2014;9(4):695–708.
Lee MK, Kim JB, Franklin MK. Enhancing the security of personal identification numbers with three-dimensional displays. Mob Inf Syst 2016;2016:9pages. Article ID 8019830
Lee MK, Nam H. Secure and fast PIN-entry method for 3D display. Proceedings of the SECURWARE 2013. IARIA; 2013. p. 26–9.
Lee MK, Nam H, Kim DK. Secure bimodal PIN-entry method using audio signals. Comput Secur 2016;56:140–50.
Luca AD, von Zezschwitz E, Pichler L, Hussmann H. Using fake cursors to secure on-screen password entry. Proceedings of the CHI 2013. ACM; 2013. p. 2399–402.
Ma J, Yang W, Luo M, Li N. A study of probabilistic password models. Proceedings of the IEEE symposium on security and privacy, Series SP '14; 2014. p. 689–704.
Maeng Y, Mohaisen A, Lee K, Nyang MD. Transaction authentication using complementary colors. Comput Secur 2015;48:167–81.
Maggi F, Volpatto A, Gasparini S, Boracchi G, Zanero S. A fast eavesdropping attack against touchscreens. Proceedings of the international conference on information assurance and security (IAS); 2011. p. 320–5.
Mallat N, Rossi M, Tuunainen VK. Mobile banking services. Commun ACM 2004;47(5):42–6.
Nyang D, Mohaisen A, Kang J. Keylogging-resistant visual authentication protocols. IEEE Trans Mob Comput 2014;13(11):2566–79.
Perković T, Čagalj M, Rakić N. SSSL: shoulder surfing safe login. Proceedings of the international conference on software, telecommunication and computer networks 2009; 2009. p. 270–5.
Raguram R, White AM, Goswami D, Monrose F, Frahm Jm. ISPY: automatic reconstruction of typed input from compromising reflections. Proceedings of the CCS 2011, 2011.
Riva O, Qin C, Strauss K, Lymberopoulos D. Progressive authentication: deciding when to authenticate on mobile phones. Proceedings of the 21st USENIX conference on security symposium, 2012. Ser. Security'12.
Roth V, Richter K, Freidinger R. A PIN-entry method resilient against shoulder surfing. Proceedings of the CCS 2004. ACM; 2004. p. 236–45.
Sasamoto H, Christin N, Hayashi E. Undercover: authentication usable in front of prying eyes. Proceedings of the CHI 2008. ACM; 2008. p. 183–92.
Seol J, Jin S, Lee D, Huh J, Maeng S. A trusted IAAS environment with hardware security module. IEEE Trans Serv Comput 2016;9(3):343–56.
Shukla D, Kumar R, Serwadda A, Phoha VV. Beware, your hands reveal your secrets!. Proceedings of the ACM SIGSAC conference on computer and communications security, 2014.
Spence C, Nicholls MER, Driver J. The cost of expecting events in the wrong sensory modality. Percept Psychophys 2001;63(2):330–6.
Tan DS, Keyani P, Czerwinski M. Spy-resistant keyboard: more secure password entry on public touch screen displays. Proceedings of the OZCHI 2005. ACM, 2005.
Thorpe J, van Oorschot P, Somayaji A. Pass-thoughts: authenticating with our minds. Proceedings of the NSPW 2005. ACM; 2005. p. 45–56.
Wang D, Cheng H, Wang P, Huang X, Jian G. Zipf's law in passwords. IEEE Trans Inf Forensics Secur 2017;12(11):2776–91.
Wang D, Gu Q, Huang X, Wang P. Understanding human-chosen PINs: characteristics, distribution and security. Proceedings of the ASIACCS 2017. ACM; 2017. p. 372–85.
Wiedenbeck S, Waters J, Birget JC, Brodskiy A, Memon N. Passpoints: design and longitudinal evaluation of a graphical password system. Int J Hum-Comput Stud 2005;63(1–2):102–27.
Yadav DK, Ionascu B, Ongole SVK, Roy A, Memon N. Design and analysis of shoulder surfing resistant PIN based authentication mechanisms on Google Glass. Proceedings of the 1st workshop on wearable security and privacy (In Association with Financial Crypto 2015), 2015. Paper 8.
Yan Q, Han J, Li Y, Deng RH. On limitations of designing leakage-resilient password systems: attacks, principles and usability. Proceedings of the NDSS 2012. Internet society, 2012.
Yan Q, Han J, Li Y, Zhou J, Deng RH. Designing leakage-resilient password entry on touchscreen mobile devices. Proceedings of the ASIACCS 2013. ACM, 2013.
Zakaria NH, Griffiths D, Brostoff S, Yan J. Shoulder surfing defence for recall-based graphical passwords. Proceedings of the seventh symposium on usable privacy and security, ser. SOUPS '11, 2011.
von Zezschwitz E, De Luca A, Brunkow B, Hussmann H. SwiPIN: fast and secure pin-entry on smartphones. Proceedings of the CHI 2015. ACM; 2015. p. 1403–6.
von Zezschwitz E, Koslow A, Luca AD, Hussmann H. Making graphic-based authentication secure against smudge attacks. Proceedings of the IUI 2013; 2013. p. 277–86.
Zezschwitz, E. V., Dunphy, P., & De Luca, A. (2013). Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. Proceedings of the 15th international conference on human-computer interaction with mobile devices and services. ACM, (pp. 261–270).

DaeHun Nyang received the B.Eng. degree in electronic engineering from Korea Advanced Institute of Science and Technology in 1994, and the M.S. and Ph.D. degrees in computer science from Yonsei University, South Korea, in 1996 and 2000, respectively. He was a Senior Member of the Engineering Staff with the Electronics and Telecommunications Research Institute, South Korea, from 2000 to 2003. Since 2003, he has been a Full Professor with the Computer Information Engineering Department, Inha University, South Korea, where he is currently the Founding Director of the Information Security Research Laboratory. His research interests include cryptography and network security, privacy, usable security, biometrics, and their applications to authentication and public key cryptography. He is a member of the Board of Directors and Editorial Board of the Korean Institute of Information Security and Cryptology.

Hyoungshick Kim received his B.S. degree from the Department of Information Engineering, Sungkyunkwan University, his M.S. degree from the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, and his Ph.D. degree from the Computer Laboratory, University of Cambridge, United Kingdom, in 1999, 2001, and 2012, respectively. He is currently an assistant professor with the Department of Software, Sungkyunkwan University. His current research interests include usable security and security engineering.

Woojoo Lee received his B.S. degree in physics, and his Ph.D. degree in statistics from Seoul National University, in 2003 and 2010, respectively. He is currently an Associate Professor with the Department of Statistics, Inha University. His current research interests include medical epidemiology and biostatistics.

Sung-bae Kang received the B.Eng. and M.S. degrees in computer engineering in 2012 and 2014, respectively, from Inha University, Incheon, Korea, where he is currently pursuing the Ph.D. degree. His research interests lie in network security, mobile security, and security in computer-human interaction.

Geumhwan Cho is a graduate student in the Department of Computer Science and Engineering, College of Information and Communication Engineering, Sungkyunkwan University. He received the B.S. degree from the Department of Communication Engineering at Cheongju University, the M.S. degree from the Department of Computer Engineering, College of Electronics and Information at Kyunghee University in 2011 and 2013, respectively. His research interests include usable security, user privacy and user authentication.

Mun-Kyu Lee received the B.S. and M.S. degrees in computer engineering, and the Ph.D. degree in electrical engineering and computer science from Seoul National University, in 1996, 1998, and 2003, respectively. From 2003 to 2005, he was a Senior Engineer with the Electronics and Telecommunications Research Institute, Korea. He is currently a Professor with the Department of Computer Engineering, Inha University, Korea. His current research interests include information security and theory of computation.

Aziz Mohaisen is an Associate Professor in the Department of Computer Science, with a joint appointment in the Department of Electrical and Computer Engineering, at the University of Central Florida. His research interests are in the areas of systems, security, privacy, and measurements. His research work has been featured in popular media, including MIT Technology Review, the New Scientist, Minnesota Daily, Slashdot, The Verge, Deep Dot Web, and Slate, among others. He obtained his Ph.D. from the University of Minnesota. He is a senior member of IEEE.