Available online at www.sciencedirect.com

ScienceDirect

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

Computers & Security

Utilizing binary code to improve usability of pressure-based authentication

Zhangyu Meng, Jun Kong, Juan Li*

Department of Computer Science, North Dakota State University, USA

ARTICLE INFO

Article history:
Received 9 April 2020
Revised 8 December 2020
Accepted 5 January 2021
Available online 8 January 2021

Keywords:
Pressure-based authentication
Experiment
Human-computer interaction
Usable security
Mobile authentication

ABSTRACT

Due to its invisibility feature, pressure is useful to enhance the security of authentication, especially preventing the shoulder surfing attack. However, users are more familiar with digital passwords than pressure-based passwords. In order to improve the usability of pressure-based authentication, this paper instantiates a pressure-based password (i.e., a sequence of pressures) to a decimal number. In addition, our approach features personalized pressure detection. The personalization further enhances security since an attacker must have a pressure habit that is consistent with the user. We conducted a series of user studies to compare the traditional four-digit password with our pressure-based password. The empirical result indicates that a pressure-based password is more resistant to the shoulder surfing attack than a four-digit password. However, it takes more time to input a pressure-based password on the first-time usage. The slowdown is caused by a modality change from vision to pressure. A field study that lasted for 10 days revealed that the side effect of modality change can be overcome through regular usages.

1. Introduction

Smartphones have become pervasive in daily life. Since many sensitive data are stored on a smartphone, authentication is necessary to access a smartphone. However, traditional usernames and passwords are not mobile-friendly since a smartphone lacks a tactile keyboard, which makes data entry tedious and error-prone. Accordingly, lock pattern was proposed to replace typing with drawing on smartphones, but it is vulnerable to smudge attack that analyzes the reflective properties of oily residues (Aviv et al., 2010). In addition, both username/password and lock pattern suffer from the shoulder surfing attack.

To address the above issues, fingerprint or face ID has been implemented on smartphones. However, those biometric authentications require some specialized sensor/software, which is not available on all types of mobile devices, especially low-end smartphones. In addition, the biometrics hacking team of the Chaos Computer Club (CCC) successfully bypassed the biometric security of Apple's TouchID [Fra13].

With the fast development of pressure-enabled sensors, pressure provides an alternative solution to enhance the security of authentication. Especially, pressure has been combined together with traditional PINs, such as Force-PINs (Krombholz et al., 2016). The combination was proved to mitigate the shoulder surfing attack due to the invisible feature of pressure and increase the password space (Krombholz et al., 2016). On the other hand, a combined password is more time-consuming than a pure digital PIN, as suggested by previous studies (Arif et al., 2014; Krombholz et al., 2016), since it requires a user to handle two modalities (i.e., pressure and vision) simultaneously. Several studies found out that the

* Corresponding author.
E-mail addresses: Zhangyu.meng@ndsu.edu (Z. Meng), jun.kong@ndsu.edu (J. Kong), j.li@ndsu.edu (J. Li).
https://doi.org/10.1016/j.cose.2021.102187

authentication time is a key factor that affects the usability of authentication (Harbach et al., 2016; De Luca et al., 2015; De Luca et al., 2014). Different from previous studies that focused on security, this paper intends to improve usability of pressure-based authentication. Since digital numbers are used much more frequently than pressure in our daily life, we expect a digital PIN is easier to memorize and use than a pressure-based password (i.e., a sequence of deep/shallow pressure levels) due to familiarity. In order to reduce the familiarity gap, we instantiate a pressure-based password to a digital number. For example, for a decimal number 6, its equivalent binary number is 0110, indicating a four-bit pressure-based password of "Shallow Deep Deep Shallow" by mapping a binary bit of 1 or 0 to a deep or shallow pressure level, respectively.

To the best of our knowledge, existing pressure-based authentication, in general, used a predefined threshold to classify pressure levels (Krombholz et al., 2016). However, a previous study revealed that finger pressure is more discriminative than a keystroke to identify a user (Saevanee and Bhattarakosol, 2009), which implied that users have different patterns to press a touch screen. Therefore, a static threshold may not fit an individual user's pressure pattern. Different from previous studies, our approach features personalized detection that fits each user's pressure pattern, and thus improves usability. In addition, personalized detection enhances security by adding a second layer of protection. In other words, even if a pressure-based password is leaked, an attacker may not bypass the authentication if his/her pressure pattern is different from the user's pattern.

To evaluate the security and usability, we compared the traditional four-digit password with the above pressure-based password. The empirical evidence indicated that the pressure-based password was more secure than the four-digit password, especially more resistant to the shoulder surfing attack. On the other hand, the four-digit password was faster and easier to remember than the pressure-based password on the first-time usage. The slow efficiency of a pressure-based password is mainly caused by a modality change (i.e., from vision to pressure). A 10-day field study further revealed that the side effect of modality change can be overcome through regular uses. In summary, this paper presents a usable and secure pressure-based authentication. The contributions are summarized as follows.

- We instantiated a pressure-based password to a decimal number, which is justified to be easy to memorize and use after a simple training.
- Personalized pressure detection improves security since an attacker's input must be consistent with the user's pressure habit.
- Our approach collects three types of information, i.e., pressure (if applicable), pressing duration, and pressing size, to detect a pressure level. Therefore, our approach is also applicable to smartphones that do not have a pressure sensor.

The remainder of the paper is structured as follows. Section 2 reviews the related work. Section 3 overviews our approach. Sections 4 and 5 present two user studies and the empirical results. Section 6 analyzes the results, followed by the conclusion and future work.

2. Related work

Since each user has a unique pattern in his/her daily behaviors, such as keystroke dynamics, mouse movement or speech, behavioral biometrics has been used to identify and authenticate users. For example, Clarke et al. used the latency between consecutive keystrokes and the time taken to press and release a key to recognize users on mobile devices (Clarke and Furnell, 2007). Ali et al. proposed a keystroke pressure-based typing biometrics authentication system on a physical keyboard to verify authorized users (Ali et al., 2009). Saevanee et al. conducted a feasibility study that combined finger pressure and keystroke dynamics to authenticate a user and revealed that finger pressure is more discriminative than keystroke dynamics (Saevanee and Bhattarakosol, 2009). Luca et al. introduced an implicit authentication approach that enhanced password patterns with an additional security layer, i.e., the way a user performed the input (Luca et al., 2012). Khan et al. systematically compared six implicit authentication schemes based on behavioral biometrics to authenticate mobile device users and concluded that touch-behavior-based scheme including pressure as a feature outperformed other schemes in terms of accuracy and detection delay (Khan et al., 2014). Buschek et al. combined the temporal typing features with spatial touch-specific features in keystroke biometrics to improve authentication accuracy (Buschek et al., 2015). In summary, the above studies proved that keystroke biometrics is discriminative to recognize users. Based on the previous work, this paper focuses on applying the pressure feature to enhance security and implemented personalized pressure detection to observe the unique pressure pattern of each user.

Pressure is an integral component in many daily gestures (e.g., holding or touching). Therefore, it is natural to extend the design space of Human-Computer Interaction with pressure. For example, Steward et al. (Stewart et al., 2010) investigated the fundamental characteristics of pressure interaction, while some studies focused on applying pressure in a specific context, such as typing (Brewster and Hughes, 2009) or security (Krombholz et al., 2016). Especially, pressure has been integrated with a knowledge-based password to enhance security. Sen and Muralidharan (Sen and Muralidharan, 2014) proposed a mobile authentication that used the pressure as an additional authentication attribute, in addition to a passcode. Krombholz et al. (Krombholz et al., 2016) implemented the force-PINs on iPhone that combined the traditional digit password with finger pressure and justified that force-PINs were resistant to the shoulder surfing attack. Different from Krombholz's work, our approach features personalized detection that adjusts pressure detection for each user. In addition, we convert a pressure-based password to a digit number in order to reduce the learning time. Instead of traditional textual passwords, Chang et al. used a sequence of user selected photos as a graphical password, which is combined with keystroke dynamic to implement an authentication system for mobile devices, and concluded that the pressure feature of keystroke can reduce the error rate (Chang et al., 2012). Orozco

et al. proposed a graphical password as connected nodes on a grid and used pressure as an extra feature in characterizing the password (Orozco et al., 2006; Malek et al., 2006).

Different techniques have been applied to detect pressure when a user pressed the touch screen. For example, Sen and Muralidharan (Sen and Muralidharan, 2014) used the WEKA classifier to classify the keystroke pattern of a user and achieved 14.06% false rejection rate. On the other hand, our approach has a 13.2% error rate in Study 1 (refer to Section 4.6.1), which includes both false negatives and user operational errors. Arif et al. (Arif et al., 2014) used predetermined ranges to detect pressure levels, and achieved a 72.7% accuracy, while our approach yielded a better accuracy of 86.8% in study 1. Brewster and Hughes (Brewster and Hughes, 2009) applied a predefined threshold to differentiate two pressure levels, and Dwell pressure interaction, with which a user had to apply force for 0.5 s before a selection was made, yielded a 2.8% error rate. Our approach has a 0.9% error rate after 10-day use in Study 2 (refer to Section 5.6.3).

Pressure has been applied to address the shoulder surfing issue on different platforms, such as drawing a shape on the back of a mobile device (De Luca et al., 2013), a pressure-grid on multi-touch tabletops (Kim et al., 2010) or Force-PINs on the touch screen of mobile devices (Krombholz et al., 2016). Being consistent with previous studies, our research concluded that pressure was effective to prevent the shoulder surfing attack. Our study also implied that attackers can still guess a pressure level based on a user's finger behavior, even if pressure is invisible. In addition to shoulder surfing, pressure has been proven to reduce the threat of smudge attack, in which attackers discern a password based on the oily smudges left on a touch screen by the user's fingers (Arif et al., 2014).

The knowledge-based password requires a long and random sequence of characters to enhance security, but it is challenging for a human brain to memorize such a sequence. Therefore, there is a tradeoff between security and memorization. Different approaches have been proposed to improve memorization without compromising security. For example, Weiss et al. proposed PassShapes, which converted a complex sequence of characters to an easy-to-remember shape (Weiss and De Luca, 2008), while Takada and Koike proposed an image-based authentication system for mobile phones, which used user's favorite images to define and memorize a password (Takada and Koike, 2003). To improve the memorization, our approach uses a decimal number to represent a sequence of pressure levels.

3. A pressure-based authentication system

3.1. Overview

Pressure has received more attention with the release of Apple' 3D touch, while digital numbers were used every day and everywhere in our daily life. Due to the familiarity, we expect a digital PIN is easier to memorize and faster to input than a pressure-based password. In order to reduce the learning curve, our approach concretizes an invisible and unfamiliar pressure-based password as a visible and familiar digital number. Specifically speaking, when classifying pressure into two

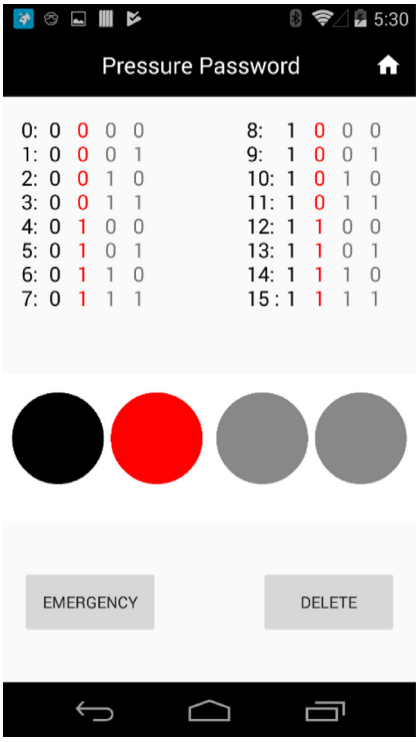

Fig. 1 – Concretizing a pressure-based password.

levels (i.e., deep or shallow), it is natural to map the binary bit 1 (or 0) to a deep (or shallow) press. Correspondingly, each number's binary code visually defines a unique pressure sequence. For example, the binary code "1101" (i.e., equivalent to the decimal number 13) defines a pressure sequence of "deep, deep, shallow, deep".

In order to input a pressure-based password, a user needs to convert a decimal number to its equivalent pressure sequence. However, the conversion is time-consuming and error-prone. In order to mitigate the conversion effort, our approach divides the authentication interface into two areas, as shown in Fig. 1. The top area displays a set of decimal numbers, each of which is followed by its corresponding binary code, while the bottom area presents four circle buttons that accept a user's pressure input. In the authentication, a user first skims the numbers displayed in the top area to identify the defined password. Then, he/she follows the binary code to press the circle buttons sequentially with appropriate pressure levels. In order to provide visual feedback on the current step in a pressure sequence, our approach highlights the current binary bit with red (See Fig. 1). In addition, a black circle indicated a pressed button, and a gray circle indicated a button to be pressed.

3.2. Threat/adversarial models

Traditional PIN is vulnerable to key logging attacks (Maheshwari and Mondal, 2016). In order to deduce a user's input on the touch screen of a mobile device, a mobile keylogger needs to capture both the coordinates of a user's touch and the screenshot of an interface. Since a pressure-based password is determined by a finger's pressure level, instead

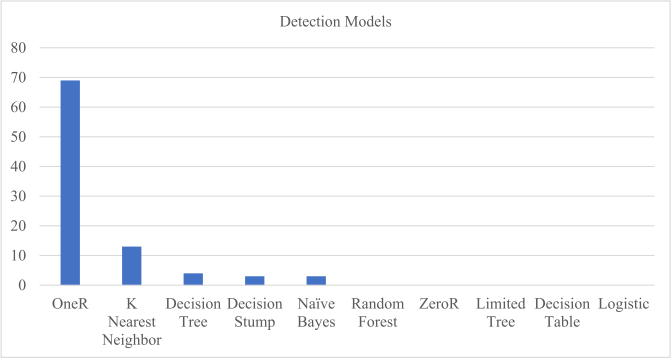

Fig. 2 – Detection models (i.e., x-axis) and the number of participants adopting a model (i.e., y-axis).

of what a user pressed on a touch screen, pressure-based passwords are resistant to key logging attacks.

Smartphones are equipped with a set of zero-permission sensors to enhance user experience. However, those sensors may cause unintentional leakage of user private data (Berend et al., 2018). Based on the assumption that the movement of a finger press on a touch screen is specific to a pressure level, attackers may potentially access and analyze the readings from smartphone's accelerometer and gyroscope sensors to derive a pressure level.

A pressure-based password can efficiently defend against the smudge attack, in which an attacker analyzes the oily smudges left behind by a user's finger when operating a mobile device (Aviv et al., 2010). In our approach, the input area includes four horizontal circle buttons, and a user presses each circle button exactly once. Therefore, the oily smudges are left the same on all buttons.

3.3. *Detection models*

Our approach builds a personal model to detect pressure for each individual user, because users have different pressure habits and a constant threshold may not reflect the actual feeling of pressure for a specific user. For example, a user with a great muscle strength probably performs a shallow press with more force than another user's deep press. In order to build a personal model. A user needs to first complete a training process by pressing the touchscreen 20 times, which includes both deep and shallow presses. For each press in the training process, the application prompts a user to press the touch screen with a specific pressure level, i.e., deep or shallow. After the user's press, the application collects three pieces of data, i.e., pressure (if applicable), pressing size, and pressing duration, and accordingly labels the press with the pressure level based on the system prompt. Therefore, our approach is also applicable to touch screens without a pressure sensor.

Based on the data collected during the training, we used the Weka[1] library to calculate a personal model for each user from 10 popular classification algorithms (see Fig. 2). Specifically speaking, 20 pressings were randomly divided into two groups, 15 to the training set and 5 to the test set. The training set was used to build a model of each classification algorithm, and the test set to compare the 10 models. If two models have the same accuracy, we chose the model with the shorter execution time. Both the data collection and the calculation of a personal model are only performed once in the beginning.

Based on the data collected from 92 participants (i.e., 55 from Study 1 and 37 from Study 2), we found that the simple OneR algorithm works best for a majority of participants, i.e., 69 out of 92, which implies that pressure can be predicted by one single predictor.

With a personalized detection model, a zero-effort impostor submits his/her own pressure feature, which is compared against the pressure pattern of a genuine user. Since the impostor and the genuine user can have different pressure patterns, a personalized detection can potentially reduce false positives with zero effort attack.

4. User study 1

This section compares a pressure-based password with a traditional four-digit password. This study focuses on the user experience of the first-time usage.

4.1. *Research hypotheses*

Goal Question Metric (GQM) (Basili et al., 1994) was used to define the goals of our study.

Goal 1: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to completion time.

Hypothesis 1: There is no difference in completion time between the two treatment methods.

Goal 2: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to the error rate.

Hypothesis 2: There is no difference in error rates between the two treatment methods.

Goal 3: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to the subjective feedback.

Hypothesis 3: There is no difference in the subjective feedback between the two treatment methods.

[1] https://github.com/rjmarsan/Weka-for-Android.

Goal 4: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to security.

Hypothesis 4. The pressure-based password is more secure than the four-digit password.

4.2. *Participating subject*

55 participants were recruited by email at a mid-west university for this experiment. 83.6% of the participants identified themselves as males, and 16.4% as females. 87.3% of the participants were between 18 and 24 years old, 10.9% were between 25 and 34 years old, and the remaining 1.8% were 35 or older. 54.5% of the participants identified themselves as iPhone users, 49.1% as Android users, and 1.8% as Windows Phone users. A participant can choose multiple authentication methods he/she was using to access his/her phone. Among 55 participants, 47.3% participants were using 4-digit PINs, 12.7% for 6-digit PINs, 1.8% for a character and digit password, 18.2% for unlock patterns, 60% for fingerprint, and 7.3% participants did not use any password. In addition, 45.5% of the participants had not used the 3D Touch technique.

4.3. *Apparatus*

In order to test the backward compatibility feature of our approach, Google Nexus 5 that does not have a pressure sensor was selected in our lab study.

4.4. *Experiment design*

Since the four-digit password was commonly used in our daily life, it was selected as a benchmark in the experiment. In order to make a fair comparison, the pressure-based password was limited to four presses, which imply a password space from 0 to 15.

The study was conducted as a pretest-posttest, repeated-measures experiment. Participants were randomly divided into two invisible groups in order to minimize the learning effect of the treatment order. Each subject in group 1 first used the four-digit password, followed by the pressure-based password. Conversely, each subject in group 2 used the authentication methods in a reverse order. Our study started with a pre-study questionnaire. Then, each treatment method starts with a training session, followed by defining a password and inputting the defined password. In the end, the participants completed a post-study questionnaire to measure their subjective satisfaction. After the first treatment method, participants repeat the above procedure with the second treatment method. The researchers recorded the time each participant took to input a password and counted the errors. In summary, the experiment included the following steps.

Step 1: *Pre-Study Survey*. The first step was to collect background information from the participants regarding their reading-comprehension skills and prior knowledge about mobile authentication. The information gathered during the pre-study was used to gain additional insight into the participants' performances during the experiment.

Step 2: *Training*. Following the pre-study survey, the participants were trained for an authentication method (i.e., pressure-based password or four-digit password) by the same researcher. In addition, in the pressure-based password, a participant was asked to press the touchscreen 20 times with a predefined sequence of different pressure levels to collect the training data.

Step 3: *Define a Password*. Since a participant chose his/her own password in the real life, we instructed a participant to define a password that he/she thought as secure as possible in the study.

Step 4: *Input a Password*. In this step, a participant was asked to input the password he/she defined in the previous step. A participant had three chances to input the password. If a participant inputted a wrong password, he/she was asked to retype it until a successful input is received or all the three chances were used up.

Step 5: *Post-study Questionnaire*. After inputting a password, participants were asked to complete a post-study questionnaire to give their subjective feedback. Then, participants repeated steps 2 to 5 with the second authentication method.

Step 6: *Shoulder Surfing*. Shoulder surfing happens when an attacker is close to observe the typing behavior of a user. In order to evaluate shoulder surfing, we invited four volunteers (two female and two male), who were neither the researchers nor participants of this study, and each volunteer shot two videos, one using the pressure-based password and one using the four-digit password. Each participant was asked to watch those 8 videos with a random order. After watching a video, a participant guessed the password and wrote it down on a piece of paper.

Step 7: *False Acceptance Test*. A volunteer who is not related to the research was invited to build a personal detection model and define a pressure-based password. Then, participants were given the defined pressure-based password and pretended to be an attacker to input this password.

4.5. *Data collection*

We recorded the completion time and error rate for each authentication method. The completion time was measured as the duration from the first touch to the last touch, and only successful authentication attempts were recorded. The error was distinguished to be either a basic error or a critical error. The basic error counted the number of the failed login attempts, and the critical error was calculated as the number of completely failed authentication sessions, i.e., a participant failed to input the password three times. In addition, we collected the number of correct guesses in the shoulder surfing attack and calculated the false acceptance rate.

We also gathered the subjective, self-reported data in the pre-study survey and the post-study questionnaire. Using the 5-point Likert scale (ranging from "1- strongly disagree" to "5-strongly agree"), each participant was asked to rate both authentication methods on three different characteristics, i.e., security, ease of use, ease of memorization.

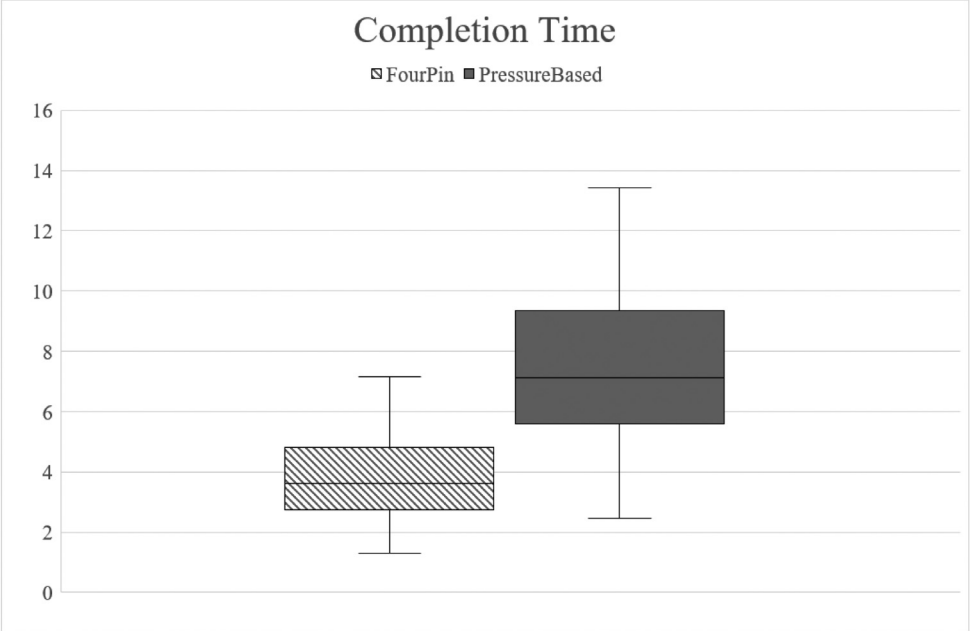

Fig. 3 – Box Plot of Completion Time (Measured in Second).

4.6. *Results*

This subsection presents the evaluation results.

4.6.1. *Completion time and error rate*

Fig. 3 presented the completion of two treatment methods that were measured in second. The four-digit password is significantly more efficient than the pressure-based password (four-digit=4.54 s vs pressure=7.69 s; $p<0.001$).

The error rate indicates the average percentage of failed login attempts per treatment. In each treatment, a participant has at most three opportunities to correctly enter a password. In the treatment of four-digit password, 5 participants among 55 in the first round did not pass the authentication with an error rate of 9%; and all 5 participants in the second round correctly entered the password with an error rate of 0. Therefore, the average error rate in the four-digit password is 3%. In the treatment of pressure-based password, 8 participants among 55 in the first round did not pass the authentication with an error rate of 14.5%; 2 participants among 8 in the second round did not pass the authentication with an error rate of 25%; and all 2 participants in the third round correctly entered the password with an error rate of 0. Therefore, the error rate in the pressure-based password is 13.2%. In summary, the pressure-based password has a higher error rate than the traditional four-digit password. Two reasons trigger a failed login in the pressure-based password. First, the underlying detection model does not recognize a correct pressure-based password, i.e., a false negative. Second, a login is denied by the authentication due to a participant's operation error, e.g., muscle is not controlled well to press the touch screen at an appropriate level. Based on the improvement of error rate from Day 1 to Day 10 in the second study (Refer to 5.6.3), a high error rate in the pressure-based password is attributed to a participant's operation, and the errors can be reduced by practice.

The critical errors in both treatment methods are 0.

4.6.2. *User subjective feedback*

Participants perceived the pressure-based password significantly more secure than the four-digit password (pressure=4.09 vs four-digit=2.87; $p<0.001$). On the other hand, the four-digit password is significantly easier to memorize (four-digit=4.65 vs pressure=3.93; $p<0.001$) and easier to use (four-digit=4.69 vs pressure=3.85; $p<0.001$) than the pressure-based password.

4.6.3. *Shoulder surfing*

In the shoulder surfing attack, we collected a total of 440 guesses (=4 volunteers × 2 videos each volunteer × 55 participants), and each method has exactly 220 guesses. In the evaluation, 96% guesses are correct in the four-digit passwords while only 45% guesses are correct in the pressure-based passwords ($p<0.001$). In summary, the evaluation result showed that the pressure-based password is more resistant to the shoulder surfing attack than the four-digit password.

In the pressure-based password, among 99 correct guesses (i.e., 45% of 220 total guesses), Table 1 presents the distribution of correct guesses among 4 volunteers. The result showed large difference among 4 volunteers. Only 14 participants correctly derived volunteer 1's password, while the number is more than double in volunteer 3. The difference reveals that both a user's pressure entry behavior and a defined password can significantly affect the resistance to shoulder surfing. For example, volunteer 3's pressure-based password is "deep shallow deep shallow", i.e., any two continuous pressures have opposite levels. In addition, volunteer 3 has a much harder press on a deep level than on a shallow level. Therefore, an attacker can capture the difference between two continuous pressings to derive a password.

Table 1 – Distribution of Correct Guesses among 4 Volunteers.

	Volunteer 1	Volunteer 2	Volunteer 3	Volunteer 4
Correct Guesses	14	28	34	23

4.6.4. *Enhanced security with personalization*

Among 55 attempts, only 34.5% attempts successfully passed the authentication when a participant was given a defined pressure-based password. This low false acceptance rate indicated that personalization enhances security by providing a second layer of protection. In other words, an attacker can 100% successfully pass the pressure-based authentication without personalized detection, while an attacker only has 34.5% chance to pass the authentication with personalization. Therefore, personalization enhances security by verifying the consistency of pressure habits between a user and an attacker.

However, the evaluation on the false acceptance rate was limited to one user, and the result can be significantly affected by the pressure behavior of the user. In other words, a good user whose pressure behavior is hard to replicate can achieve a better false acceptance rate. In the future, we will conduct a more conclusive study by asking a few volunteers to simulate attackers and replicate the pressure-based passwords of participants.

5. User study 2

The first user study revealed that both the usability and performance of the pressure-based password are lower than that of the four-digit password. However, participants are already familiar with the four-digit password before the first study, but they are new to the pressure-based password. It is interesting to find out whether the performance and usability of the pressure-based password are improved over a period of regular use. Therefore, we conducted a second user study that asked participants to use the pressure-based password for a 10-day period.

5.1. *Research hypotheses*

Goal Question Metric (GQM) (Basili et al., 1994) was used to define the goals of our study.

Goal 1: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to efficiency after 10-day usage.

Hypothesis 1: There is no difference in completion time between the two treatment methods after 10-day usage.

Goal 2: Analyze a pressure-based password and a traditional four-digit password for the purpose of their evaluation with respect to subjective feedback after 10-day usage.

Hypothesis 2: There is no difference in subjective satisfaction between the two treatment methods after 10-day usage.

5.2. *Participating subject*

37 participants were recruited by email at a mid-west university for this experiment. None of the participants in the second experiment attended the first experiment. 81.1% of the participants identified themselves as males, and 18.9% as females. 75.7% of the participants were between 18 and 24 years old, and 24.3% between 25 and 34 years old. 43.24% of the participants used iPhones and 56.76% used Android. 45.95% participants reported they were using 4-digit PINs, 16.22% for 6-digit PINs, 5.4% for FaceID, 24.32% for unlock patterns, 2.7% for Android Smartlock, 70.27% for fingerprint, and 10.81% did not use any authentication method. In addition, 43.24% of the participants did not use the 3D Touch technique.

5.3. *Apparatus*

Participants used their personal Android phones to complete the study. If a participant did not have an Android phone, Google Nexus 5 was lent to the participant.

5.4. *Experiment design*

The objective of this experiment was to compare the pressure-based password with the traditional four-digit password after 10-day usage. The study was conducted as a pretest-posttest, repeated-measures experiment. All participants went through each of the authentication methods each day for 10 continuous days. To encourage participants to complete the 10-day trial, we only ask participants to input a password once every day. In order to prevent participants from memorizing passwords in a certain pattern, we randomized the order of two authentication methods each day. The experiment's operations included the following steps.

Step 1: *Pre-Study Survey*. The first step was to collect background information from the subjects.

Step 2: *Training*. Following the pre-study survey, the subjects were trained on each authentication method by the same researcher. In the pressure-based password, a user also completed a training process to build up a personal detection model.

Step 3: *Define a Password*. The participants were asked to define a password as secure as possible for each authentication method. Then, participants were asked to practice entering the defined password of each method for 15 times and ask questions if any. The purpose of the practice is to let participants familiar with each treatment method and to prevent potential problems that a participant may encounter during the 10-day trial.

Step 4: *10-day Usage*. In this step, each participant was asked to use each authentication method once per day for 10 continuous days. Participants could retrieve the password if they forgot the password. Each participant completed a post-study questionnaire on both the first day and the last day.

5.5. Results

5.5.1. Completion time

Table 2 shows the average completion time on both the first and the last days. In the pressure-based password, the completion time on the 10th day is significantly improved over that on the 1st day, which indicates efficiency is significantly improved over a period of regular use. However, the four-digit is still significantly faster than the pressure-based password after 10-day use. To discover the trend of the completion time during the 10-day usage, we calculated a fitted line plot for each treatment method to visualize the relationship between the completion time and the day. As presented in Fig. 4, the completion time of the pressure-based password decreased much faster from day one to day ten than that of the four-digit password.

5.5.2. User subjective feedback

User subjective feedback is measured from the perspectives of security, ease of memorization and ease of use. Being consistent with the results in the first study, the pressure-based password has a higher perceived security rating than the four-digit password, as presented in Table 3. Furthermore, the comparison on the security rating between the first and the last days in the pressure-based password indicated that users enjoyed the security feature of pressure more after a 10-day usage. In the ease of memorization, though users felt that the four-digit password was easier to memorize than the pressure-based password on the first day, there was no difference after a 10-day usage between two treatment methods. A similar result was revealed on ease of use. In summary, after a 10-day usage, user subjective feedback on the pressure-based password is significantly improved.

5.5.3. Error rate

Table 4 presents the error rate on the first and last days in the second study. Each participant has at most three opportunities each day to input a password.

On the 1st day in the treatment of pressure-based password, 3 participants among 37 in the first round did not pass the authentication; 1 participants among 3 in the second round did not pass the authentication; and the only 1 participant in the third round correctly entered the password. On the other hand, in the treatment of four-digit password, only 1 participant among 37 did not pass the authentication; and the only 1 participant correctly entered the password in the second round. Accordingly, the pressure-based password has an error rate of 13.7%, and the four-digit password has an error rate of 0.9%. This result is consistent with that of the first study that pressure-based password caused a higher error rate than the four-digit password on the 1st day.

On the 10th day in the treatment of pressure-based password, only 1 participant among 37 did not pass the authentication; and the only 1 participant correctly entered the password in the second round. On the other hand, in the treatment of four-digit password, all 37 participants entered the password correctly in the first round. Accordingly, the pressure-based password has an error rate of 0.9%, and the four-digit password has a zero-error rate.

By comparing the error rates between 1st and 10th days, we found that the error rate in the treatment of pressure-based password is greatly reduced by 10-day practice. In other words, continuous practice can help users to improve muscle control and thus reduce operational error. In addition, the error rate of pressure-based password is very close to that of traditional four-digit password after 10-day usage.

The critical errors in both treatments are 0.

Table 2 – Average Completion Time of the First and Last Days.

		1st day	10th day	
Completion Time	Pressure	3.25	2.3	$p<0.001$
	Four-digit	2.06	1.7	$p = 0.066$
		$p < 0.001$	$p < 0.001$	

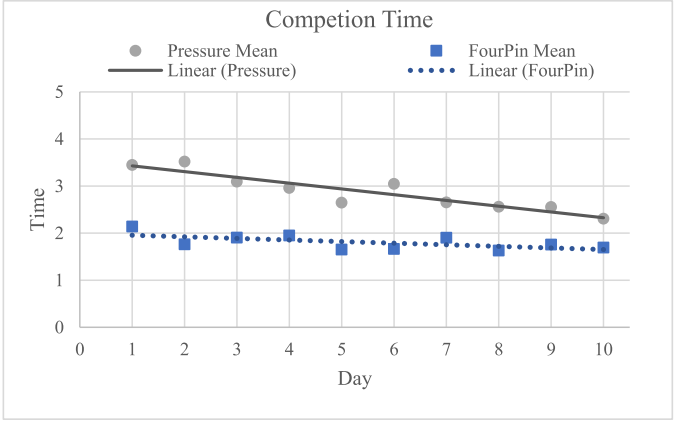

Fig. 4 – Fitted Line Plot of Two Methods.

Table 3 – User Subjective Feedback.

		1st day	10th day	
Security	Pressure	4.38	4.59	$p<0.001$
	Four-digit	3.05	3.0	$p = 0.073$
		$P<0.001$	$p<0.001$	
Ease of memorization	Pressure	4.3	4.57	$p = 0.031$
	Four-digit	4.75	4.49	$p = 0.058$
		$P = 0.005$	$p = 0.584$	
Ease of use	Pressure	4.08	4.38	$p = 0.078$
	Four-digit	4.54	4.43	$P = 0.524$
		$p = 0.017$	$p = 0.744$	

Table 4 – Error Rate.

	1st Day				10th Day			
	1st Round	2nd Round	3rd Round	Error Rate	1st Round	2nd Round	3rd Round	Error Rate
Pressure	3 out of 37	1 out of 3	0 out of 1	13.7%	1 out of 37	0 out of 1	0 out of 0	0.9%
Four-digit	1 out of 37	0 out of 1	0 out of 0	0.9%	0 out of 37	0 out of 0	0 out of 0	0%

6. Discussion

This section discusses the evaluation results and presents the findings from the study.

6.1. *Shoulder surfing and personalized detection*

Both the subjective and objective measures justified that a pressure-based password is more secure than a traditional four-digit password, which is consistent with previous studies (Arif et al., 2014; Krombholz et al., 2016). The improved security is mainly caused by the invisible feature of pressure. Therefore, a pressure-based password is more resilient against the shoulder surfing attack for mobile authentication in a public environment. In a previous study (Krombholz et al., 2016), the shoulder surfer was not able to guess a single force-PIN. However, in our study, 70 out of 152 guesses were correct in a pressure-based password. Our study indicates that although pressure is invisible, a shoulder surfer can still guess a pressure level through a user's typing behavior. Therefore, a user needs to be wary of the shoulder surfing risk even when he/she is using a pressure-based authentication.

Our approach features personalized pressure detection. With personalized detection, even if a pressure-based password is leaked, an attacker may not pass the authentication if he/she has a different pressure habit from the user. However, an attacker can simply press touchscreen extremely lightly or extremely hard to bypass personalized detection when a single threshold is used to differentiate pressure levels. Therefore, in practice, personalized detection is useful when a scope is used to define a pressure level (See Fig. 5). In other words, if pressure is normalized as a value between 0 (i.e., shallowest) and 1 (i.e., deepest), each pressure level is defined as a scope with personal lower and upper bounds. With a scope-based definition, an attacker must assure his/her pressure falls within appropriate boundaries that are consistent with the user.

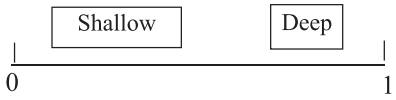

Fig. 5 – Use a scope to define a pressure level.

6.2. *Efficiency*

People intend to process familiar information faster. Therefore, our approach converts a pressure-based password to a digit number that a user is familiar with. Compared with a traditional four-digit PIN, our approach reduces the finger travel distance since the next button is always adjacent to the previous button. We expect that this benefit may offset the familiarity of a digit PIN and thus our approach is as efficient as the four-digit PIN. However, the result of the first study was beyond our expectation. This surprise is caused by a modality change. Specifically speaking, our approach converts a pressure modality to a vision modality in the password definition and then requires a reverse conversion in the password authentication. The modality change causes inconsistency between password definition and authentication, which reduces the perceived easiness. In addition, the modality change triggers an additional skimming operation in the password authentication. In other words, a user must scan the top area of the interface to identify a predefined decimal number and its corresponding pressure sequence. The skimming operation can slow down the overall efficiency. The second study indicated that the modality change can be overcome by a 10-day usage, and thus both the efficiency and perceived easiness are significantly improved. According to a previous study, users unlock their phone on average 47.8 times a day (Harbach et al., 2014), while our second study only required a participant to use a pressure-based password once a day. Therefore, we expect that users can overcome modality change much faster in practice. Another consideration is to update the interface

to improve the skimming operation. Our approach uses 0 to indicate a shallow pressure and 1 to a deep pressure, which originates from binary code. However, only users with a strong computer background are familiar with binary code. On the other hand, users are using color every day. Therefore, we expect color (such as white and black) may potentially speed up the skimming operation over the binary code.

6.3. Usability and deployment

A set of criteria have been proposed to systematically and objectively compare authentication schemes. For example, Bonneau et al. (Bonneau et al., 2012) proposed the usability-deployability-security framework for evaluating Web authentication schemes. Recently, Wang et al. (Wang and Wang, 2018) defined 12 independent criteria in terms of user friendliness and security for evaluating two-factor authentication schemes. Since previous sections have already discussed security, this section focuses on usability and deployment of the proposed pressure-based authentication.

Based on Wang's framework (Wang and Wang, 2018), our proposed approach is *password-friendly* since a pressure-based password is memorable by converting it to a decimal number and it can be chosen freely and changed locally by a user. Based on the Bonneau's framework (Bonneau et al., 2012), our approach is not *memorywise-effortless* and *scalable-for-users* since a user has to memorize each account password, while it is *Quasi-Nothing-to-Carry* since a user only needs to carry a smartphone and is *Physically-Effortless* since a user only needs to press a button. Since we convert a sequence of pressure levels to a decimal number, our approach is *Easy-to-Learn*. The user study results indicated that our approach is *Efficient-to-Use* and *Infrequent-Errors*. Finally, our approach is *Easy-Recovery-from-Loss*.

Our approach features personalized detection, which increases security. However, personalized detection requires a training process in the beginning to calculate a detection model. Consequently, our approach increases the effort of deployment.

6.4. *Multi-Level pressure*

Our study investigated a four-bit two-level pressure-based password, which limits the password space to only 16 combinations and is vulnerable to the guessing threat. If four-bit two-level pressure-based passwords are evenly distributed, the theoretical entropy is 4 bits. Previous studies indicated that over 50% of every PIN dataset can be accounted for by just the top 5%˜8% most popular PINs (Wang et al., 2017). We expect the same feature is applied to pressure-based passwords. Consequently, the practical entropy will be smaller than 4 bits. Wang et al. (Wang et al., 2017) calculated four PINs database and concluded that four-digit PINs provided about 4.81–6.62 bits of security and six-digit PINs about 4.35–7.24 bits of security against the guessing threat. Therefore, four-bit two-level pressure-based passwords are more vulnerable to the guessing threat than 4- or 6-digit PINs. To increase the security against the guessing threat, one solution is to increase the pressure levels. According to a previous study (Mizobuchi et al., 2005), users can control 6 ± 1pressure levels without major difficulties. If pressure levels are increased from 2 to 7, the theoretical entropy is increased to 11.23 bits. Our approach is especially useful to memorize multi-level pressure-based passwords. For example, a four-bit seven-level pressure-based password can be visualized as a four-digit decimal number, where each digit is ranging from 1 to 7 that indicates a unique pressure level.

In order to evaluate the feasibility of a multi-level pressure detection, we conducted a pilot test to detect 3-level pressures, i.e. deep, medium and shallow. The pilot test invited 10 volunteers. Each volunteer pressed the touch screen 45 times, which included 15 pressures of each level. The collected data of each volunteer is divided into two sets, i.e., the training and test sets. The training set, which includes 30 pressures (i.e., 10 pressures each level), is used to build personalized detection models. On the other hand, the test set, which includes 15 pressures (i.e., 5 pressures each level) is applied to evaluate the detection accuracy. Table 5 presents the detection accuracy on 10 detection algorithms per volunteer, while Table 6 displays the average accuracy per detection algorithm. Though the pilot study did not provide enough evidence to support seven different pressure levels, it revealed some practical issues that are worth of future investigation. The preliminary data showed that the features of a medium level are very close to that of a deep or shallow level. This observation implied that the detection errors in the pilot study were attributed to participants' pressing operations. In other words, when a participant intends to press the touch screen at a specific level, his/her actual pressure falls in the scope of an adjacent level. Based on the preliminary data, our future work includes designing an efficient training process that helps a user to build a personal and consistent pressing pattern to avoid overlapping between adjacent levels. First, visualizing the pressure force in the training process may facilitate users to concretize an abstract pressure level and thus help them to memorize the muscle operation on a specific pressure level. Second, to avoid fatigue, the training process should terminate once a consistent pressing pattern is formed. The challenge is to identify a set of features (e.g., the standard deviation in a pressure level or the degree of overlapping between the maximal pressure force of a lower level and the minimal pressure force of an adjacent upper level) to determine the termination of a training process.

6.5. Limitations

Our study was limited to comparing a 4-bit pressure-based password with a 4-digit PIN. We expect similar results can be applied to 6-digit PINs based on the following two situations.

- **Comparing a 6-bit pressure-based password with a 6-digit PIN.** Due to the invisible feature of pressure, our study justified that a 4-bit pressure-based password is more resistant to shoulder surfing attacks and had a higher rate on perceived security. Since 6-bit pressure-based passwords have the same feature as 4-bit pressure-based passwords, we expect that 6-bit pressure-based passwords are more resistant to shoulder surfing attacks than 6-digit PINs and

Table 5 – The detection accuracy of each volunteer.

	1	2	3	4	5	6	7	8	9	10
J48	93.3%	100%	60%	100%	93.3%	93.3%	93.3%	60%	66.7%	86.7%
KNN	86.7%	93.3%	66.7%	86.7%	93.3%	86.7%	86.7%	93.3%	80%	80%
OnR	80%	100%	60%	100%	93.3%	80%	100%	73.3%	66.7%	86.7%
ZEROR	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%
Decision Stump	60%	66.7%	40%	60%	53.3%	60%	60%	60%	66.7%	53.3%
Random Forest	80%	100%	73.3%	100%	93.3%	80%	100%	80%	80%	86.7%
Limited Tree	93.3%	100%	73.3%	93.3%	86.7%	93.3%	100%	86.7%	73.3%	80%
Naïve Bayes	86.7%	100%	73.3%	80%	93.3%	86.7%	100%	73.3%	73.3%	86.7%
Decision Table	80%	100%	60%	100%	93.3%	80%	100%	66.7%	73.3%	53.3%
Logistic	80%	100%	86.7%	86.7%	93.3%	80%	100%	80%	73.3%	86.7%

Table 6 – The average accuracy of 10 models.

J48	84.67%
KNN	85.33%
OneR	84.00%
ZEROR	26.67%
DecisionStump	58.00%
RandomForest	87.33%
Limited Tree	88.00%
NaiveBayes	85.33%
DecisionTable	80.67%
Logistic	86.67%

have a higher perceived security rating. Since the empirical study indicates that 4-bit pressure-based passwords are as efficient as 4-digit PINs after a period of usage and the increased operation efforts from 4 bits to 6 bits are almost the same between pressure-based passwords and digital PINs, we expect 6-bit pressure-based passwords are as efficient as 6-digit PINs.

- **Comparing a 4-bit pressure-based password with a 6-digit PIN**. Since the invisible feature of pressure mainly contributes to the resistance of shoulder surfing attacks, we expect 4-bit pressure-based passwords are more resistant to shoulder surfing attacks even though 6-digit PINs have a much larger password space. Furthermore, since 6-digit PINs requires more user operations than 4-digit PINs, we expect it is faster to input a 4-bit pressure-based password than a 6-digit PIN. However, since 6-digit PINs have a much larger password space, it is not clear which method may achieve a higher rating on perceived security.

In summary, it is worth of future investigation to compare 4/6-bit pressure-based passwords with 6-digit PINs.

6.6. *Threats to validity*

We faced the following threats to validity in the study. The threat due to the heterogeneity of participants was not controlled because participants were not drawn from the same course and were a mix of undergraduate and graduate students (i.e., having different education levels). While we use a 5-point Likert scale, there remains a threat that we treated the scale as an interval scale rather than an ordinal scale. This practice means that the p-value needs to be treated with care when interpreting the results. In addition, there remains a threat because the participants were all undergraduate and graduate students in an educational setting, and they did not represent typical smartphone users.

To increase the internal validity, we did not inform the participants about the study goals. Therefore, they should not have been biased with their evaluations. Participants volunteered to take part in this study and were not graded on their performance.

7. Conclusion and future work

This paper proposes a pressure-based password and compares it with the traditional 4-digit password. Our approach features personalized detection, which proves to improve security. In addition, empirical studies revealed that a pressure-based password is more resistant to the shoulder surfing attack than a traditional digit password. On the other hand, it takes more time to input a pressure-based password than a digit password on the first-time usage. The slowdown is mainly caused by a modality change. Fortunately, a field study indicated that the modality change can be overcome through regular usages, which make a pressure-based password as efficient as a digit password. The future work includes improving the visualization to reduce the completion time and investigating a large password space.

Author statement

Zhangyu Meng implemented the pressure-based password and conducted the user studies.
Jun Kong supervised and design the whole study.
Juan Li supervised the detection models.

Declaration of Competing Interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgment

This work is in part supported by NSF under grant #1722913.

REFERENCES

Ali Hasimah, Wahyudi, Salami MJE. Keystroke pressure based typing biometrics authentication system by combining ANN and ANFIS-based classifiers. In: 2009 5th International Colloquium on Signal Processing & Its Applications; 2009. p. 198–203.

Arif AS, Mazalek A, Stuerzlinger W. The use of pseudo pressure in authenticating smartphone users. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services; 2014. p. 151–60.

Aviv AJ, Gibson KL, Mossop E, Blaze M, Smith JM. Smudge attacks on smartphone touch screens. Woot 2010;10:1–7.

Basili VR, Caldiera G, Rombach HD. The Goal Question Metric Approach. Department of Computer Science, University of Maryland; 1994. Technical Report ftp://ftp.cs.umd.edu/pub/sel/papers/gqm.pdf.

Berend D, Jungk B, Bhasin S. There goes your PIN: exploiting smartphone sensor fusion under single and cross user setting. Proceedings of the 13th International Conference on Availability, Reliability and Security, 2018.

Bonneau J, Herley C, Oorschot PC, Stajano F. The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. Proc. IEEE Symp. On Security and Privacy, 2012.

Brewster SA, Hughes M. Pressure-based text entry for mobile devices. Proc. MobileHCI 2009;09:73–6.

Buschek D, De Luca A, Alt F. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM; 2015. p. 1393–402.

Chang T-Y, Tsai C-J, Lin J-H. A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices. J. Syst. Softw. 2012;85(5):1157–65.

Clarke NL, Furnell S. Advanced user authentication for mobile devices. Comput. Secur. 2007;26(2):109–19.

De Luca A, Von Zezschwitz E, Nguyen NDH, Maurer M-E, Rubegni E, Scipioni MP, Langheinrich M. Back-of-device authentication on smartphones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2013. p. 2389–98.

De Luca A, Hang A, Von Zezschwitz E, Hussmann H. I feel like I'm taking selfies all day!: towards understanding biometric authentication on smartphones. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM; 2015. p. 1411–14.

De Luca A, Harbach M, von Zezschwitz E, Maurer M, Slawik B, Hussmann H, Smith M. Now you see me, now you don't: protecting smartphone authentication from shoulder surfers. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2014. p. 2937–46.

Harbach M, Zezschwitz EV, Fichtner A, De Luca A, Smith M. It's a hard lock life: a field study of smartphone (Un)Locking behavior and risk perception. Proc. SOUP, 2014.

Harbach M, De Luca A, Egelman S. The anatomy of smartphone unlocking: a field study of android lock screens. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM; 2016. p. 4806–17.

Khan H, Atwater A, Hengartner U. A comparative evaluation of implicit authentication schemes. In: International Workshop on Recent Advances in Intrusion Detection. Springer; 2014. p. 255–75.

Kim D, Dunphy P, Briggs P, Hook J, Nicholson JW, Nicholson J, Olivier P. Multi-touch authentication on tabletops. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2010. p. 1093–102.

Krombholz K, Hupperich T, Holz T. Use the force: evaluating force-sensitive authentication for mobile devices. In: Symposium on Usable Privacy and Security; 2016. p. 207–19.

Luca DA, Huang A, Brudy F, Lindner C, Hussmann H. Touch me once and I know it's you! Implicit authentication based on touch screen patterns. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2012. p. 987–96.

Maheshwari A, Mondal S. SPOSS: secure pin-based-authentication obviating shoulder surfing. In: Information Systems Security. Springer International Publishing; 2016. p. 66–86.

Malek B, Orozco M, El Saddik A. Novel shoulder-surfing resistant haptic-based graphical password. Proc. EuroHaptics 2006;6:1–6.

Mizobuchi S, Terasaki S, Keski-Jaskari T, Nousiainen J, Ryynanen M, Silfverberg M. Making an impression: force-controlled pen input for handheld devices. Ext. Abstracts CHI'05 2005:1661–4.

Orozco M, Malek B, Eid M, El Saddik A. Haptic-based sensible graphical password. Proceed. Virtual Concept 2006;56:1–4.

Saevanee H, Bhattarakosol P. Authenticating user using keystroke dynamics and finger pressure. In: Proceedings of Consumer Communications and Networking Conference. IEEE; 2009. p. 1–2.

Sen S, Muralidharan K. Putting 'pressure'on mobile authentication. In: Proceedings of 2014 International Conference on Mobile Computing and Ubiquitous Networking. IEEE; 2014. p. 56–61.

Stewart C, Rohs M, Kratz S, Essl G. Characteristics of pressure-based input for mobile devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2010. p. 801–10.

Takada T, Koike H. Awase-E: image-based authentication for mobile phones using user's favorite images. In: Proceedings of International Conference on Mobile Human-Computer Interaction. Springer; 2003. p. 347–51.

Wang D, Gu Q, Huang X, Wang P. Understanding human-chosen PINs: characteristics, distribution and security. Proceedings of ASIA CCS'17 2017:372–85.

Wang D, Wang P. Two birds with one stone: two-factor authentication with security beyond conventional bound. IEEE Trans. Dependable Secure Comput. 2018;15(4):708–22.

Weiss R, De Luca A. PassShapes: utilizing stroke based authentication to increase password memorability. In: Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges. ACM; 2008. p. 383–92.

Zhangyu Meng is a Master student at Department of Computer Science, North Dakota State University. His-research interests include Usable Security and Human Computer Interaction.

Jun Kong received the B.S. degree from the Huazhong University of Science and Technology in 1998, the M.S. degree from Shanghai Jiao Tong University in 2001, and the Ph.D. degree from the University of Texas at Dallas in 2005, all in computer science. He has been a Professor of computer science with North Dakota State University, Fargo, USA. His-research and teaching interests include human–computer interaction, software engineering and brain-computer interface.

Juan Li received the B.S. degree in Computer Science from the Beijing Jiaotong University, Beijing, China, in 1997, the M.S. degree in Computer Science from the Chinese Academy of Sciences, Beijing, China, in 2001, and the Ph.D. degree in Computer Science from the University of British Columbia, Vancouver, Canada, in 2008. Currently she is an Associate Professor in the Computer Science Department at the North Dakota State University, Fargo, ND, USA. Dr. Li's major research interest lies in distributed systems, intelligent systems, social networking, and semantic web technologies.