Hollow-Pass: A Dual-View Pattern Password Against Shoulder-Surfing Attacks

Jiayi Tan\textsuperscript{1(เคฟฟ์)} and Dipti Kapoor Sarmah\textsuperscript{2(เคฟฟ์)}

\textsuperscript{1} EEMCS, Universiteit Twente, Drienerlolaan 5, 7522 Enschede, NB, The Netherlands
j.tan-1@student.utwente.nl
\textsuperscript{2} SCS/EEMCS, University of Twente, Drienerlolaan 5, 7522 Enschede, NB, The Netherlands
d.k.sarmah@utwente.nl

Abstract. This paper presents Hollow-Pass, a developed solution that strengthens the security of pattern passwords against shoulder-surfing attacks. It is a novel approach to graphical password (GP) schemes that utilize a dual-view technology known as the global precedence effect, which eliminates the need for external devices and makes the grid and pattern invisible to potential shoulder surfers. The usability of Hollow-Pass was evaluated through an online as well as an offline user test. We recruited 30 participants from varied backgrounds, ranging in age from 20 to 80 years, for the online user test. An offline small-scale sampling test was conducted among 19 undergraduates from the Universiteit of Twente. The developed solution successfully demonstrated its ability to effectively resist shoulder-surfing attacks for simple patterns at various viewing angles (front, left-front, and right-front) and different distances (1.0 m, 1.5 m, and 2.0 m).

Keywords: Pattern password \cdot Graphical password \cdot Shoulder-surfing \cdot Dual-View \cdot Global precedence

1 Introduction

Graphical passwords (GP) are gaining widespread popularity as a method of authentication. GP can be classified into four categories: recognition-based, recall-based, cued-recall-based, and hybrid schemes \cite{7,31}. One widely used GP system is the Android Pattern Unlock \cite{40}(illustrated in Fig. 1). This is a recall-based GP that is based on the Asian game Go and is known as the Pass-Go scheme \cite{44}. During the registration and authentication process, users are asked to create a pattern by drawing a series of lines on a grid, which serves as their password \cite{37}. According to a 2014 study \cite{9}, 40% of Android users use patterns as their password instead of a PIN.
Several studies have identified shoulder surfing as a major security risk for GP systems [34,41,42]. This type of attack involves an attacker observing a user’s device screen, keyboard, or mouse in a public place to steal their login credentials [19]. A study [6] reported that 6-length Android patterns have a high attack success rate of 64.2% with a single observation and 79.9% with multiple observations. To address this concern, various countermeasures have been proposed, such as increasing the complexity of patterns or obscuring part of the pattern using external hardware. However, a 2018 systematic literature review [4] found that of the 84 countermeasures proposed for pattern locking on smartphones. Only 10 pattern-based locking schemes were known, and the majority of them emphasized increasing pattern complexity. Only one of ten techniques, called “XSide” [13,14] focused on obscuring the pattern with external hardware. Since then, new approaches such as eyes-free [43], SysPal [12], Pass-O [38], TinPal [39], gaze tracking [15,22], and swipe behavior-based mechanisms [20] have been proposed, however, they are typically designed for small-screen devices or require specialized hardware. This research aims to tackle the issue of shoulder surfing on both mobile devices and monitors without the need for specialized hardware by focusing on the processing of the pattern password itself.

According to previous research [29], shoulder-surfing attacks can be successfully resisted by using pattern passwords that incorporate distinguish features, such as colored images in a grid format. Researchers [16,24] have investigated and proved how the human visual system processes objects over time, with recognition of general or global objects preceding that of detailed or local features. This phenomenon, known as “global precedence”, can be utilized in the design of pattern passwords. By presenting the pattern password to users at a local level but obscuring it at a global level, it can be safeguarded against observation by shoulder surfers.

This study intended to establish a new pattern password mechanism, called Hollow-Pass, by combining the View manipulation and Image degradation techniques as described in Aris and Yaakob’s review of anti-shoulder-surfing techniques [4]. The efficacy of Hollow-Pass in terms of both usability and security was also assessed through a limited-scale user experiment.
The structure of this research is as follows: Sect. 2 provides an overview of previous studies that are relevant to the research. Section 3 details the methodology that was employed in the research. Section 4 analyzes the results and limitations of the research. The research concludes with a section on Acknowledgements, References.

2 Related Work

This section focuses on various aspects of visual perception, usability, and security in the context of Hollow-Pass. We reviewed the relevant criteria of the dual-view technology and shoulder surfing resistance by studying several research papers from reputed journals and conferences. The systematic search was conducted using the formulated search string below:

(`"shoulder surf" OR "shoulder surfing" OR "shoulder surfer" OR "shoulder surfers" OR shoulder-surf*) AND ("dual-view" OR "hybrid")

After evaluating 50 search results and examining 11 relevant papers, we have identified five important criteria, namely spatial frequency, visual acuity, global precedence, pattern grid layout, and CIE color system. These criteria are essential for developing a pattern-based password scheme that is both secure and user-friendly.

(i). Spatial Frequency
Spatial frequency (SF), as defined in [10], is a measurement of the number of repeating patterns within a specific distance. It is commonly expressed in cycles per unit distance, such as cycles per degree (cpd). In image processing and computer vision, spatial frequency is utilized to determine the level of detail and texture in an image. An image with high spatial frequency exhibits intricate details and textures, while an image with low spatial frequency features larger and more prominent structures [35].

(ii). Visual Acuity
The visual acuity (VA) [18] is a measure of the capacity of the human visual system to perceive fine details in objects. It is commonly assessed using the Snellen chart, which is viewed from a distance of 6 m (20 feet). A standard VA, represented as 6/6 or 20/20, corresponds to a line of letters on the chart that subtend an angle of 5 min of arc. The Snellen “E” letter, made up of three strokes and two gaps with each stroke and gap subtending 1 min of arc, is often used as a reference. Normal visual acuity is equivalent to 30 cycles per degree (cpd), with each stroke of the “E” letter representing a peak of a sine wave and the white gap between strokes representing a trough (as illustrated in Fig. 2b).
(iii). Global Precedence

Studies conducted by Flevaris [16] and Navon [24] have demonstrated that lower spatial frequencies tend to promote global perception [23], while higher spatial frequencies tend to encourage local perception. When individuals process both low and high spatial frequencies from an image, their visual processing follows a “coarse-to-fine” strategy, where they identify larger or general objects more rapidly and accurately than smaller or specific features. For instance, one might first identify a tree before noticing its branches and leaves. This effect referred to as global precedence, implies that humans have the ability to perceive the global aspect of a scene, but cannot immediately jump to local perception in a single step.

In this study, we refrained from exploring further psychological theories and instead applied the principle in processing the pattern password. This was achieved through the use of a dual-view mechanism, which displayed the “local” grid layout. Through this mechanism, a potential shoulder surfer may be able to capture the user-drawn lines on the screen, but would not be able to see the “local” grid layout, which is only visible to the user.

(iv). Pattern Grid Layout

The conventional Android pattern grid size is $3 \times 3$, however, previous research has indicated that this size is easily guessable and susceptible to attacks [32]. In an effort to enhance security, larger grid sizes have been suggested. However, a study found that there was little improvement in security from changing the grid size to $4 \times 4$ [5]. An alternative approach is to employ 9 points to make the $3 \times 3$ grid more intricate rather than increasing the size.

While various patterns can be created using a 9-point layout, a random layout may prove difficult for users to remember, leading them to choose simple patterns that are vulnerable to attacks. To address this issue, researchers have developed new grid layouts, such as the trapezium [44], circle [38,44], and house [44], which have been shown to improve security. The password space size of the circle and house layouts is larger than the original Android grid, while their overall recall
success rates do not have significant differences [44]. This suggests that different grid layouts can offer improved security while maintaining user-friendly usability.

(v). CIE Color System and $\Delta E$

Color plays a significant role in human visual perception and is an important aspect of digital images. As such, one of the research objectives is to enhance the global precedence effect by modifying the color difference in images.

The International Commission on Illumination’s CIE color system provides a numerical means of describing all colors that are visible to the human eye. Unlike the RGB color model, the color definitions in the CIE color system are absolute, unambiguous, and not influenced by device or display specifications. The CIE LAB (Lab*) model [30], published in 1976, is widely accepted as a means of quantitatively measuring perceived color. It consists of three components:

- $L^*$ represents the lightness, ranging from 0 to 100, where 0 is black and 100 is white.
- $a^*$ represents the green to red axis, ranging from $-128$ to $+127$.
- $b^*$ represents the yellow-to-blue axis, ranging from $-128$ to $+127$.

To understand how the human eye perceives color difference, the CIE color system employs the metric $\Delta E$ [28,30]. For CIE LAB, a $\Delta E \approx 10$ indicates a color difference that is visible at first glance.

3 Methodology

In this section, the research questions and the methods used to address the research questions are outlined. A web development framework [2] and a pattern unlock grid template [45] were utilized to create a website specifically for the user test. The website was made accessible to all participants through deployment on a web hosting platform [1].

Considering the situation mentioned in previous sections, we formulate the research question as follows. The main research question (MRQ) of this study is: To what extent do pattern complexity and dual-view technology impact the usability and security of Hollow-Pass? This is further broken down into three sub-research questions (SRQs):

**SRQ 1:** To what extent does a pattern drawn on a distorted grid in preventing shoulder surfing at distances of 1.0 m, 1.5 m, and 2.0 m?

**SRQ 2:** To what extent do color contrast and global precedence in preventing shoulder surfing from identifying the pattern at 1.0 m, 1.5 m, and 2.0 m?

**SRQ 3:** To what extent do the color contrast between the password pattern and background image, and the distorted grid layout impact the usability and security of Hollow-Pass?
3.1 Proposed Methods

In this section, the detailed methods we proposed to answer the SRQs are discussed. We defined (i) pattern drawing rules for users to create the patterns, (ii) notations to label the nodes and patterns. In (iii) and (iv), keyspace, and pattern complexity are outlined, respectively. From (v) to (vii), the detailed approach for addressing SRQ1, SRQ2, and SRQ3 is outlined.

(i). Pattern Drawing Rule
The pattern drawing rules were developed based on the design presented by Tupsamudre et al. [38].

(a) The pattern must be created by drawing straight lines without lifting the hand,
(b) The pattern must connect a minimum of 4 and a maximum of 9 nodes,
(c) A node cannot be linked more than once,
(d) Unlike conventional $3 \times 3$ patterns, a node that is not connected may be bypassed if it is situated on the pattern’s path. For example, a line segment can be drawn from node 1 to node 3 without visiting node 2.

(ii). Notations
For ease of understanding, the following definitions have been made for node labeling and pattern shape.

(a) Node labeling: For convenience, all nodes in the original $3 \times 3$ grid are numbered from 1 to 9 in a row-major format, with the upper-left node being labeled 1 and the bottom-right node labeled 9.
(b) Pattern representation: A pattern can be expressed as a sequence of nodes in a specific order, such as 123698745 (refer to Fig. 3).

![Fig. 3. (a) distorted grid layout; (b) conventional grid layout.](image)

![Fig. 4. (a) Intersection example 125846: one intersection point at node 5; (b) Overlap example 2564: one overlapped part 56.](image)
(iii). Key Space
Keypoint, pertains to the number of possible combinations, \( P \), a password can have. The number of valid \( r \)-node patterns in a \( 3 \times 3 \) grid can be determined by using a mathematical formula:

\[
P(n, r) = \frac{n!}{(n-r)!} = \frac{9!}{(9-r)!}
\]  \hspace{1cm} (1)

where \( n \) is the node count of the grid. For a \( 3 \times 3 \) grid, \( n \) is 9. To sum up valid patterns from 4 to 9 nodes, we used the formula:

\[
\sum_{r=4,9} P(n, r) = \sum_{r=4,9} \frac{n!}{(n-r)!} = \sum_{r=4,9} \frac{9!}{(9-r)!}
\]  \hspace{1cm} (2)

(iv). Pattern Complexity
The visual complexity of a pattern is determined by various factors such as the number of connected nodes, the length of the pattern, and the number of intersections and overlaps in the pattern (as illustrated in Fig. 4). To quantify this complexity, we employed the formula presented by Sun et al. [33] which is as follows:

\[
PS_P = S_P \times \log_2(L_P + I_P + O_P)
\]  \hspace{1cm} (3)

where \( PS_P \) is the strength score of the pattern, \( S_P \) is the number of nodes in the pattern, \( L_P \) represents the length of the pattern, \( I_P \) is the number of intersections in the pattern, and \( O_P \) is the number of overlaps in the pattern. By dividing the range of scores into five equal segments, the patterns can be classified into five levels: very weak, weak, medium, strong, and very strong.

(v). Increase Grid Layout Randomness
To address SRQ1, our approach was to increase the randomness of the grid layout. This would prevent shoulder surfers from being able to authenticate using the observed pattern, as the distribution of nodes and grid layout would vary each time. We implemented the following two steps to achieve this goal:

(a) The \( 3 \times 3 \) grid was divided into nine equal sections and each node was randomly placed within its section.

To balance usability and security, we implemented a solution that involves scattering the nodes within their designated grid squares while maintaining the visibility of the grid borders. This helped users recognize the node locations. However, as the nodes were randomly placed and did not align with the grid borders, users were able to access non-adjacent nodes on the same line without having to connect through intermediate nodes: in the distorted grid layout, node 1 can reach node 7, 3, and 9 directly while this is not allowed in the conventional grid layout (as shown in Fig. 5). This approach increased the range of reachable nodes and the complexity of the patterns, without making it overly difficult for users to recall the pattern.
The grid was then rotated by 45° in either a clockwise or counterclockwise direction.

In order to address the potential issue of users having difficulty recognizing patterns that have been rotated more than 45°, three possible grid layout rotations (45° counterclockwise, original, and 45° clockwise) were defined, and one of these variations were randomly displayed each time (as shown in Fig. 6). To help users identify the rotation direction, node 1 was circled as an indicator. When users draw their patterns, they follow the same node sequence as the original grid, while shoulder surfers face difficulty in identifying the correct pattern without knowledge of the current grid orientation due to the random rotation.

(vi). Utilize Dual-View Technology
The approach for addressing SRQ2 was to create and strengthen the global precedence effect in Hollow-Pass. We accomplished this by implementing the following two steps:

(a) Converted the foreground grid into a dashed-line format and adjusted the SF of the gratings in the background.
The research in [16] highlights that the human visual perception system is more sensitive to low spatial frequencies for global processing and more sensitive to high spatial frequencies for local processing. As per Kalloniatis and Luu [18], the representation of sinusoidal gratings in terms of SF can be done and vice versa. The background image of the study was generated using the Python open-source package PsychoPy, consisting of four layers of sinusoidal gratings, each corresponding to an orientation of 0°, 4°, 90° and 135° to cover all borders of the foreground grid (see Fig. 7). The spatial frequency in cpd of every layer was in the range of (1, 3). To establish the optimal spatial frequency for the background sinusoidal gratings, we conducted experiments with spatial frequencies ranging from 0 to 30 (which aligns with normal visual acuity as shown in Fig. 2b). The comparison results are presented in Fig. 8. Based on the results from our perceptual performance evaluations, we selected a lower bound of 1 and an upper bound of 3 (excluding 3) as
Fig. 7. An example of background gratings. The spatial frequency of this example is $sf = [1, 2, 1, 2]$.

Fig. 8. (a) $sf = 0.3, 0.3, 0.3, 0.3$, the pattern is easily observed; (b) $sf = 1, 1, 1, 1$, the pattern is somehow disguised; (c) $sf = 1, 2, 1, 2$, the pattern is somehow disguised; (d) $sf = 3, 3, 3, 3$, the background starts to have packed of holes, which may cause revulsion; (e) $sf = 30, 30, 30, 30$, the pattern is easily observed.

the optimal spatial frequency for each layer of gratings. The gratings in each layer were masked using a 2-D Gaussian filter ($sd = 3$) in PsychoPy. Additionally, to conceal the outline of the foreground at a global level, we fragmented the pattern and grid into local features by depicting them in dashed-line.

(b) Modified the color difference ($\Delta E$) between the foreground grid and background.

We hypothesized that a noticeable color difference between the foreground grid and the background may make it easier for shoulder surfers to identify the grid orientation and node positions. To mitigate this, we limited the color contrast between the foreground grid and background gratings to a specific level. Firstly, we converted the grid’s RGB color to CIE LAB using an algorithm from Manoj Pandey [25]. And then we calculated the color difference ($\Delta E$) between the grid and background image (as expressed in Eq. 4):

$$\Delta E_{Lab} = \sqrt{(L^*_{grid} - L^*_{bg})^2 + (a^*_{grid} - a^*_{bg})^2 + (b^*_{grid} - b^*_{bg})^2} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

where $bg$ represents the background image.

We limited the generation time of background images during user testing by utilizing five pre-generated images. The average color difference, as indicated by $\Delta E$, between the grid color and the background image was 10.817 ($sd = 3.998$), while the average color difference between the pattern line color and the background image was 31.841 ($sd = 1.646$). During the test, one of
the five pre-generated images was randomly displayed on the user interface each time the users created a pattern password on the grid. This method was expected to have the intended visual effect, as demonstrated in Fig. 8 through Fig. 8c.

(vii). Evaluate usability and security
The approach for addressing SRQ3 was to evaluate the usability and security of Hollow-Pass. Usability was measured through the recall error rate and the adapted System Usability Scale (SUS), while security was measured through observed pattern accuracy in the online and offline simulated attacks.

![Fig. 9. (a) Online user test procedure; (b) Offline user test setting.](image)

(a) Online test.
A 7-section online user test was conducted (as depicted in Fig. 9a). The prototype is available for trial through the provided link during the first 20 days of each month: https://hollow-pass.up.railway.app/ [36]. Participants were given the option to perform the test on either a computer or a mobile phone. This study was designed to specifically test the usability of Hollow-Pass and eliminate potential confounding variables. Participants were asked to provide consent and instructed on how to use Hollow-Pass. They were then asked to register a valid pattern (4 to 9 nodes) on the grid and demographic information was collected. Anti-shoulder-surfing resistance was evaluated by asking participants to identify 6 simulated pattern passwords at viewing distances of 1.0 m, 1.5 m, and 2.0 m. Memorability and usability were assessed by recall error rate and adapted SUS questions, respectively. The online test took an average of 10–15 minutes to complete.

(1) **Objective of Informed Consent:** Participants were made aware of and agreed to the user test at the start of the test. This included information on the purpose of the test, alternatives to the procedure or intervention, and potential risks.

(2) **Instructions:** Participants were provided with clear instructions on how to draw a pattern on either the desktop or mobile version of the site during the test.
(3) **1st time drawing (Registration):** Participants were instructed to draw their first Hollow-Pass patterns.

(4) **Collection of Demographic Information:** Participants were requested to furnish information about their demographics, including gender, age, eyesight quality, device used for testing, and prior experience with pattern-based passwords.

(5) **Simulated attack:** In this phrase, participants were asked to act as shoulder surfers to perform an online simulated attack. They were given six pattern identification questions and were asked to select the correct pattern from among four options. Each question presented a Hollow-Pass pattern (of either weak or medium complexity) that had been simulated at a viewing distance: 1.0 m, 1.5 m, or 2.0 m. The complexity of the patterns was determined using Eq. 3. These stimuli were created to assess the anti-shoulder-surfing effect of Hollow-Pass, taking into account factors: perceived resolution and size.

* **Perceived resolution.** The perception of resolution by the human visual system closely resembles a low-frequency filter. Research by Pappas and Neuhof [27] has determined that the impulse response of the 1-D eye filters matches a Gaussian shape with an appropriate standard deviation. At 300 dots per inch and a 30-in. viewing distance, the eye filter’s impulse response matches a Gaussian filter with $\sigma = 1.5$ and $\tau = 0.0095^\circ$. To simulate the perceived resolution at different viewing distances, we used a Gaussian filter, adjusting the standard deviation.

* **Perceived size.** Based on Emmert’s law [8], which states that the perceived image size changes proportionately with its distance from the observer while controlling for the visual angle. The perceived size of the image was simulated by scaling the original stimuli size ($340 \times 340$ px) for different viewing distances (1.0 m, 1.5 m, and 2.0 m). This simulation was done under the assumption that the observer had normal visual acuity and the default viewing distance was 0.5 m from the screen to the user’s eye.

(6) **2nd time drawing (Authentication):** Participants were asked to repeat their initial pattern on a grid. In case a participant failed to recall their initial pattern, they were allowed to proceed to the next question. The accuracy of the participant’s redrawn pattern in relation to their original pattern was regarded as their recall rate and was used to assess the usability of Hollow-Pass.

(7) **Adapted system usability scale(SUS):** A standardized questionnaire is a widely-used 10-item questionnaire that measures usability, learnability, efficiency, and overall user satisfaction [17]. We adapted the System Usability Scale (SUS) for user testing. The purpose of using adapted version of the SUS was to gather detailed feedback about participants’ experiences with Hollow-Pass in three dimension: reliability, feasibility and affinity. Participants were asked to evaluate their experience with the Hollow-Pass pattern using a 9-item, 5-point Likert scale
questionnaire, covering five levels of agreement. This aimed to measure their perceptions of the pattern and their overall preference towards it.

(b) Offline test.
An offline small-scale sampling test was conducted to evaluate the anti-shoulder-surfing effect of the Hollow-Pass system in practicality. Typically, pattern passwords are used on smaller mobile phone screens where the patterns can be more difficult to observe by observers. However, when pattern passwords are used on larger screens, such as laptop screens, they become more visible and easier to observe. Therefore, the test used an Acer TravelMate P2 laptop with a 14-in. screen and a pixel density of 164.64 pixels per inch to examine the system’s ability to resist shoulder surfing on a relatively large display.

The center of the laptop screen was used as the origin, and the viewing distance was calculated as the radius extending from the screen to the participant’s position. The semicircle was divided into four equal parts, each considered a distinct viewing angle: left-front, front, and right-front, as depicted in Fig. 9b. The researcher acted as a legitimate user and sat at a distance of 0.5 m from the laptop screen, using a mouse to draw six weak patterns. The password strength was evaluated using Eq. 3. Participants were asked to stand at three viewing angles (front, left-front, right-front) at specified distances from the screen (1.0 m, 1.5 m, and 2.0 m), and to draw what they observed on a test form. This was to examine the mechanism’s resistance to shoulder surfing at various viewing angles and distances for weak patterns. Participants were allowed to ask the researcher to redraw the pattern and make multiple attempts.

4 Results and Discussion

This section entails the presentation and analysis of the user test results. Section 4.1 analyzes the strength score of Hollow-Pass patterns with regards to their complexity and key space. In Sect. 4.2, participant demographic information is provided. Section 4.3 discusses the patterns chosen by participants. The usability of Hollow-Pass is discussed in Sect. 4.4, focusing on three aspects: reliability, feasibility, and user affinity. The security of Hollow-Pass is discussed in Sect. 4.5 in relation to simulated online and offline attack results. Lastly, any limitations or future research directions are examined. The purpose of this section is to offer a comprehensive overview of the mechanism and the user test outcomes.

4.1 Password Strength

In this section, we evaluated the strength of the passwords in two aspects, namely: (i). key space and (ii). pattern complexity.

(i.) Key space. We conducted a comparison between the conventional $3 \times 3$ grid used in the typical Android pattern unlock system (as illustrated in
Fig. 1) and that of the Hollow-Pass system. The key space calculation was done using Eqs. 1 and 2. The results, as presented in Table 1, show that the key space of the Hollow-Pass system is larger than that of the conventional method, increasing from 389,112 to 985,824, particularly for node sizes 4, 5, 6, and 7. This implies that users can create a greater diversity of patterns, making it harder for shoulder surfers to recognize them, without necessarily increasing the number of nodes.

Table 1. Key space comparison between conventional GP and Hollow-Pass

<table>
<thead>
<tr>
<th># of nodes</th>
<th>Conventional GP</th>
<th>Hollow-Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1,624</td>
<td>3,024</td>
</tr>
<tr>
<td>5</td>
<td>7152</td>
<td>15,120</td>
</tr>
<tr>
<td>6</td>
<td>26,016</td>
<td>60,480</td>
</tr>
<tr>
<td>7</td>
<td>72,912</td>
<td>181,440</td>
</tr>
<tr>
<td>8</td>
<td>140,704</td>
<td>362,880</td>
</tr>
<tr>
<td>9</td>
<td>140,704</td>
<td>362,880</td>
</tr>
<tr>
<td>Total</td>
<td>389,112</td>
<td>985,824</td>
</tr>
</tbody>
</table>

(ii.) **Pattern complexity.** The password strength of all acceptable patterns was determined using Eq. 2 by taking into account factors such as the length of the pattern (depicted in Fig. 10a), the number of intersections in the pattern (depicted in Fig. 10b), and the amount of overlap in the pattern (depicted in Fig. 10c). The results showed that as the number of nodes increased, the length of the pattern increased, and there were more intersections and overlaps, making the pattern more complex. The password strength score varied from 6.34 to 46.81, as illustrated in Fig. 10d.

### 4.2 Participant Demographic Information

In the online user test, 30 participants (15 female) between the ages of 20–80 took part, and in the offline sampling user test, 19 undergraduates (10 female) between the ages of 20–23 from the Universiteit of Twente participated. All participants gave informed consent, which was approved by the Universiteit of Twente’s ethics committee. As shown in Table 2, the majority of participants (70%) were in the 20–29 age group, with 2 participants in the 30–49 age group, 6 in the 50–59 age group, and 1 over 60. The eyesight quality of participants was recorded using decimal Snellen notation, with 11 participants having a score less than 0.8, 6 between 0.8–1.0, 5 above 1.0, 6 unsure of their score but wearing glasses, and 2 unsure and not wearing glasses. Most participants were familiar with pattern passwords, with 17 having used it before and 11 knowing of it but
not having used it. Only 2 participants were unfamiliar with pattern passwords before the test. The participants also indicated the device used for the test, which could be a desktop, phone, or tablet.

### 4.3 Pattern Selection

In the online user test, participants were suggested (but not required) to create a pattern with at least 4 and at most 9 nodes, as Hollow-Pass was a new pattern password that might present challenges to some users. 83% of the participants drew suggested patterns while only 4% of these participants failed to confirm the registered pattern. 59% of the participants chose a very weak pattern as their password, 22% chose a weak pattern, and only 3% chose a strong pattern (as illustrated in Fig. 11a). Patterns with 3 or fewer nodes were deemed invalid. The password strength was calculated using Eq. 3. The most commonly used valid pattern was 12369874 with a strength of weak (as shown in Fig. 11b), followed by 1235789 with a strength of very weak (as shown in Fig. 11c). Most participants used 5–6 nodes to create their patterns. To evaluate participants’ memorability, the number of successful authentication and participants’ recall rate was calculated by comparing their registered patterns (first-time drawing) and authenticated patterns (second-time drawing). 10% of participants, across all age groups, were unable to redraw their patterns during the authentication stage (as shown in Fig. 11d). Because we recruited a limited number of participants in the age group 30–49 and 60+, we focused on discussing the recall rate of the age group 20–29 and 50–59. In Fig. 11e, we observed that the recall rate was 90% of the age group 20–29, and 83% of the age group 50–59.

![Fig. 10. Distribution of valid patterns attributes.](image)
4.4 System Usability

In the online user test, participants were asked to evaluate the usability of the system, using an adapted version of SUS with 5 points, where 1 represents “Completely disagree,” 2 represents “Somewhat disagree,” 3 represents “Neither agree nor disagree,” 4 represents “Somewhat agree,” and 5 represents “Completely agree”. To ensure the adapted SUS validation, we calculated the factor loading with Varimax rotation and conducted Cronbach’s alpha to assess questionnaire validation and reliability. We chose 3 factors that had high loadings (> .4) to assess system usability based on the scree plot and the eigenvalue, which were perceived reliability, perceived feasibility, and user affinity. Furthermore, the calculated Cronbach’s alpha was .825, suggesting that the adapted SUS questionnaire had good internal consistency.

Table 2. Demographic Information of the Participants

<table>
<thead>
<tr>
<th>Demographic information</th>
<th>Desktop</th>
<th>Phone</th>
<th>Tablet</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Non-binary</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–29</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>30–49</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>50–59</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>60+</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Eyesight quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below 0.8</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>0.8–1.0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Over 1.0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>unsure, wears glasses</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>unsure, does not wear glasses</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Pattern experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>used</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>known but not used</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>not known</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>15</td>
<td>14</td>
<td>1</td>
<td>30</td>
</tr>
</tbody>
</table>
(i). **Reliability.** 90% of the participants expressed positive opinions regarding the mechanism reliability of Hollow-Pass, with 63% completely agreeing and 27% somewhat agreeing (as depicted in Fig. 12a). They reported a greater sense of security with Hollow-Pass in comparison to the conventional pattern password and concurred that it makes it more challenging for shoulder surfers to discern the credentials from a distance.

(ii). **Feasibility.** Over 70% of the participants gave a favorable assessment of the mechanism feasibility, with 44% indicating complete agreement and 27% indicating somewhat agreement (as shown in Fig. 12b). Participants reported that they could easily view the grid and nodes, observe their pattern while drawing, and complete the pattern drawing without any difficulties.

(iii). **Affinity.** Approximately 70% of the participants expressed positive views on the grid layout and background design (44% completely agreed and 26% somewhat agreed, as illustrated in Fig. 12c). Participants found the mechanism easy to learn and use during the test. They concurred that the system-generated background design was more secure than a customizable background design. Nevertheless, they also desired the option of customizing their background image, should this feature be made available.

### 4.5 System Security

(i). **Online simulated attack.** Each participant was asked to identify two different patterns for each viewing distance (30 participants, totaling 60
patterns per viewing distance). The mean accuracy of identifying stimuli that simulate perceptual effects at different viewing distances (1.0 m, 1.5 m, 2.0 m) was 55.55% in the online simulated attack, as shown in Fig. 13a. However, the stimuli accuracy was not significantly affected by the viewing distance.

(ii). **Offline simulated attack.** In the offline simulated attack, each participant observed 6 patterns per viewing distance at three viewing angles (2 patterns per angle), resulting in a total of 114 patterns collected from 19 participants. The accuracy in the offline test was lower than that of the online test, with a mean accuracy of 20% compared to 56%, as shown in Fig. 14a. The accuracy decreased with increasing viewing distance and was highest when participants were directly in front of the screen (41%) and lowest when positioned at the right-front (25%). The average accuracy of observing Hollow-Pass at various viewing angles was 32%, as illustrated in Fig. 14b.

The difference in results between the online and offline tests may be due to the consistent option placement in the online test. A systematic review [21] has shown that test takers perform better when the correct answer is consistently placed in a specific location, as opposed to randomly placed. The study also found that placing the correct answer at the top or bottom of the options may increase the likelihood of the test taker choosing it, leading to potential bias. This aligns with the result of our research, where the correct answer was usually located at options (a) and (b), as shown in Fig. 13b. We used the chi-squared test of independence to determine whether there was a
significant association between participants’ answers and the correct answer placement. We found that the p-value was <.001, indicating participants’ accuracy was affected by the option placement. Another factor contributing to the difference is the format of the tests, with online participants being presented with multiple-choice questions and offline participants having to draw the full patterns without any options. Furthermore, offline participants reported difficulties in recognizing the grid orientation and skipped nodes, as they had limited time to identify and memorize the pattern while drawing.

It is noteworthy that the correctness of the observed patterns decreased when the participants conducted multiple attempts. This outcome is consistent with the conclusions of Adam et al. [6], as illustrated in Fig. 15.

![Fig. 15. Correctness of single attempt and multiple attempts.](image)

### 4.6 Comparative Analysis

The purpose of this section is to compare and contrast Hollow-Pass with two existing methods that also employed image processing techniques to resist shoulder-surfing attacks: IllusionPIN [26] and HideImage [11].

As shown in Table 3, all three methods have robust protection against shoulder-surfing attacks. IllusionPIN, as a PIN-based textual password, uses a shuffled keypad and hybrid images technique to provide robust protection against shoulder-surfing attacks. However, it simplifies its visual algorithm by converting its keypad’s color into black and white, and the used parameters, SF, need to be further tuned. Moreover, the conducted user test disregarded viewing angles and the scale was limited. HideImage, as a recognition-based graphical password, downgrades static images and converts them into grayscale. This results in the loss of color and high SF in the images. On the other hand, HideImage is independent of user interaction. This means it may not be as effective when it comes to dynamic image frames that require user interaction, such as drawing a pattern password. In contrast, Hollow-Pass, as a dynamic, recall-based graphical password, is designed to protect passwords that are based on user interaction without information loss. It can be a better option for applications that require dynamic image frames without latency. It also has a relatively larger user scale among ages 20 to 60+. However, because Hollow-Pass uses a simplified visual algorithm, SF needs to be tuned to achieve better protection.
### 4.7 Limitations

The Hollow-Pass mechanism has a few limitations as follows:

(i). One of these is the potential for a biased perceptual effect when users draw patterns on displays with different specifications or at greater viewing distances, as the mechanism employs a background image created with a default display pixel density of 164.54 pixels per inch and a viewing distance of 0.5 m.

(ii). The user test had a restricted scope, and the offline evaluation solely included university students between the ages of 20 and 29. Younger adults tend to be more receptive to novel technology than middle-aged and older adults [3]. Therefore, a more comprehensive study that includes all age groups is necessary to investigate the effect of a background image and color contrast on other demographic categories.

### 5 Conclusion and Future Work

We present Hollow-Pass, a novel pattern password mechanism utilizing global precedence and color difference $\Delta E$, allowing users to draw patterns on a dynamic $3 \times 3$ grid as their authentication credentials. An online user test was conducted with $n = 30$ participants aged between 20–80, and an offline user test was conducted with $n = 19$ undergraduate students, viewed from three distances (1.0 m, 1.5 m, 2.0 m) and three angles (front, left-front, right-front). The
results indicate that Hollow-Pass enhances the security of weak pattern passwords against shoulder-surfing attacks while maintaining usability, as the simulated shoulder surfer observed 20% of tested patterns on average in the offline user test. Over 70% online participants gave positive feedback, suggesting that Hollow-Pass effectively resists shoulder-surfing attacks and balances security and usability.

We found that a significant percentage of participants (57%) had pattern password experience. As our experimental design does not allow for a direct comparison of the usability of the developed grid with a conventional grid, future studies could include a conventional grid as a control group to more directly compare the usability of different grid designs. Moreover, we intend to carry out a comprehensive offline study that focuses on both the usability and security of Hollow-Pass, covering all age groups. This research also opens a door for the researchers to consider human eye adjustments [18] during the online test instead of having an average value of visual acuity and a viewing distance. This study aims to investigate any potential tradeoffs that may arise due to the application of the mechanism. Hollow-Pass may provide defense against automatic online guessing attacks, as the grid layout changes randomly with each use, akin to the Captcha scheme [46].

References