RubikAuth: Fast and Secure Authentication in Virtual Reality

Citation for published version:

Digital Object Identifier (DOI):
10.1145/3334480.3382827

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
RubikAuth: Fast and Secure Authentication in Virtual Reality

Abstract
There is a growing need for usable and secure authentication in virtual reality (VR). Established concepts (e.g., 2D graphical PINs) are vulnerable to observation attacks, and proposed alternatives are relatively slow. We present RubikAuth, a novel authentication scheme for VR where users authenticate quickly by selecting digits from a virtual 3D cube that is manipulated with a handheld controller. We report two studies comparing how pointing using gaze, head pose, and controller tapping impacts RubikAuth’s usability and observation resistance under three realistic threat models. Entering a four-symbol RubikAuth password is fast: 1.69 s to 3.5 s using controller tapping, 2.35 s to 4.68 s using head pose, and 2.39 s to 4.92 s using gaze and highly resilient to observations: 97.78% to 100% of observation attacks were unsuccessful. Our results suggest that providing attackers with support material contributes to more realistic security evaluations.

Author Keywords
Usable Security; Authentication; Virtual Reality

CCS Concepts
- Human-centered computing → Human computer interaction (HCI); - Security and privacy → Authentication;
Introduction and Related Work

The surge of new immersive virtual reality applications [2, 7, 23], and the availability of high-end untethered head-mounted displays (HMDs) [6, 27], has made VR ubiquitous. However, the ability to experience VR almost anywhere comes with security implications. Users are often required to authenticate in VR to, for example, make in-app purchases [17] or to verify their identity [16]. Recent research indicates established authentication methods such as PINs or 2D graphical PINs [11, 30] are prone to observation attacks when used in VR. The problem is exacerbated by the fact VR users are often unaware of bystanders [9, 24].

<table>
<thead>
<tr>
<th>System</th>
<th>Authentication Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>RubikAuth</td>
<td>2.39 s-4.92 s</td>
</tr>
<tr>
<td>eye gaze</td>
<td></td>
</tr>
<tr>
<td>head pose</td>
<td>2.35 s-4.68 s</td>
</tr>
<tr>
<td>tapping</td>
<td>1.69 s-3.5 s</td>
</tr>
<tr>
<td>RoomLock [10]</td>
<td>8.58 s-14.33 s</td>
</tr>
<tr>
<td>HoloPass [14]</td>
<td>16.69 s</td>
</tr>
<tr>
<td>LookUnlock [8]</td>
<td>≈6 s</td>
</tr>
<tr>
<td>2D PINs [11]</td>
<td>2.38 s-3.84 s</td>
</tr>
<tr>
<td>2D/3D PINs [30]</td>
<td>≈10.5 s-19 s</td>
</tr>
<tr>
<td>VRPursuits [21]</td>
<td>21.40 s</td>
</tr>
</tbody>
</table>

Table 1: RubikAuth improves entry times and observation resistance over many existing schemes for VR and AR. All systems above use four-symbol PINs. (*) Based on the implementation of LookUnlock [8], we estimate that entering a four-symbol PIN takes at least 6 s (4 × 1.5 s).

RubikAuth: Concept and Implementation

RubikAuth is a knowledge-based authentication scheme, where users verify their identity by inputting digits on a virtual 3×3×3 cube (Fig. 1). The digits 1-9 are displayed on five of its six uniquely-coloured surfaces; we omitted the rear face as it is not easily reachable. The cube pose is directly linked to the sensed pose of an HTC VIVE controller held in the non-dominant hand. RubikAuth's efficiency derives from the use of Guiard's kinematic chain model for human asymmetrical bimanual cooperation [12, 13], its resistance to observation by splitting input on multiple coordinated input modalities [4, 29], and it is faster to select symbols from polygon 3D shapes than from 2D grids [18].

The advantage of using a manipulable 3D object for authentication in VR is threefold: 1) it gives quick access to many targets in high speed using minimal wrist movements, 2) it complicates attacks by requiring attackers need to observe both the cube manipulations, and the positions of the selected targets, and 3) the intuitiveness of cube manipulations makes it easier for users to anticipate actions that improve observation resistance.

To authenticate, the user points at the target digit on the desired surface (Fig. 2), and then presses the trigger button on the dominant-hand HTC VIVE controller. All RubikAuth pointing methods use explicit selection by pressing the trigger button. Compared to dwell time, the use of a separate trigger has several advantages: a) it gives users more control [15], b) adds an additional channel that attackers must observe [19, 20], and c) significantly decreases best-case authentication time; reliable dwell selection requires at least 350 ms per selection [26, 28], implying a minimum of 1.4 seconds to enter a four-symbol PIN.

Threat Models

RubikAuth addresses three realistic threat models that ensure optimal conditions for the attacker (Fig. 3). These depict the scenarios where users are in a public space and are not aware of potential attackers. In all threat models,
the attacker: a) has an optimal view of the user’s interactions, b) can move freely, c) knows the beginning and the end of the authentication process, d) knows which pointing method will be used, and e) knows that the user will enter a four-symbol PIN. The attacker’s knowledge of this information is realistic as previous work showed that bystanders are able to identify the user’s task in VR [9].

**Threat Model 1: Pen and Paper**
The attacker observes the user during authentication. They note down observations on a paper on which an abstract 2D form of RubikAuth is drawn with labels showing the surface colours (Fig. 3-1).

**Threat Model 2: 3D Replica**
In recent work, attackers came up with ways to help them note down observations (e.g., folding paper to form a 3D version of a virtual environment [10]). Motivated by these strategies, in addition to the material used in threat model 1, the attacker uses a real-world replica of the 3D cube: a Rubik’s cube with overlaid digits (Fig. 3-2).

**Threat Model 3: Video Recordings**
Motivated by the ubiquity of smartphones, here the attacker uses a smartphone (S7 EDGE, 12 MP Camera) to record and freely play back authentications, in addition to all material used in threat models 1 and 2 (Fig. 3-3). The attacker has the advantage of choosing the recording angle as the user is not aware of their presence due to the HMD [9,24].

**User Studies**
We conducted two user studies (2×1h) to study RubikAuth’s usability and observation resistance. Both studies were designed as repeated measures lab experiments. Conditions were counter balanced using a Latin Square. All participants were compensated with an £8 online shop voucher. Both studies complied with university’s ethics procedure.

**Study 1: Usability Evaluation**
We recruited 23 participants (13 females, 10 males) aged between 18 and 54 years (M=27.65, SD=8.26). 11 (47.83%) had never used VR before. There were two independent variables: **IV1 Pointing Method**: we compared pointing via gaze, head pose and controller tapping (three conditions, Fig. 2), and **IV2 Required Switches**: we studied the impact of the number of times the user switches from one surface to another while authenticating.

A four-symbol PIN in RubikAuth has either 0-switches, 1-switch, 2-switches, or 3-switches (four conditions). Entering a 0-switches PIN is equivalent to a classical PIN-pad, so we treat 0-switches as a baseline.
Switches | 0 | 1 | 2 | 3 | Σ
---|---|---|---|---|---
Threat 1 | 0 | 0 | 0 | 0 | 0
Threat 2 | 0 | 0 | 0 | 0 | 0
Threat 3 | 0 | 0 | 0 | 0 | 0
Head pose | 0 | 0 | 0 | 0 | 0
Tapping | 3 | 0 | 1 | 0 | 4
Switches | 0 | 1 | 2 | 3 | Σ
---|---|---|---|---|---
Threat 1 | 2 | 0 | 0 | 0 | 2
Threat 2 | 1 | 0 | 0 | 0 | 1
Threat 3 | 1 | 0 | 0 | 0 | 1
Σ | 4 | 0 | 0 | 0 | 4

Table 2: Attacks against RubikAuth are rarely successful. Attacks were only successful against head pose and controller tapping.

Procedure
After filling a consent form and a demographics questionnaire, participants were introduced to VR and RubikAuth. They went through a training session by entering 3 PINs each with eye gaze, head pose and controller tapping. We excluded training runs from analysis. Participants then went through one block per pointing method, entering predefined PINs. In each block, participants entered 2 PINs × 4 switches × 4 repetitions = 32 PINs/block. Before each PIN entry, we showed participants which targets they should select directly on the cube. The order of the digits was highlighted with white numbers on a black background (Fig. 1).

Usability Evaluation Results
We logged 8 PINs × 3 pointing methods × 4 repetitions × 23 participants = 2208 authentications. We excluded 87 outliers due to tracking issues, such as moving out of the tracking range, or accidentally pressing the menu button on the HTC VIVE controller.

VR savvy vs. non-VR savvy participants
We compared the performance of participants who used VR before with those who did not. VR savvy users authenticated in \( (M=3.27 \text{ s}, SD=0.623 \text{ s}) \) and made 9.56% errors. For non-VR savvy participants, the values were \( (M=3.15 \text{ s}, SD=0.567 \text{ s}) \) and 10.25% respectively. None of the differences were significant \((p > .05)\). This highlights the naturalness of interacting with RubikAuth gained from the use of Guiard’s kinematic chain model [12, 13].

Authentication Time
We measured authentication time from the moment the first entry is made until the fourth symbol is selected. When analysing input time of these successful authentications, a two-way repeated measures ANOVA with Greenhouse-Geisser correction (due to violation of the sphericity assumption) revealed a statistically significant main effect of pointing method \((F_{1,619,35,617} = 38.894, p < .05)\) and number of switches on authentication time \((F_{2,477,54,497} = 309.887, p < .05)\). It also showed a significant two-way interaction between pointing method and number of switches on authentication time \((F_{3,619,79,621} = 5.096, p < .05)\).

Further analysis was conducted to distinguish the impact of each independent variable. Individual ANOVAs for each switches condition and post hoc t-tests with Bonferroni correction showed that across all switches, authentication time using controller tapping \((M=2.60 \text{ s}, SD=0.90 \text{ s})\) is significantly faster \((p < .05)\) than when using eye gaze \((M=3.60 \text{ s}, SD=1.35 \text{ s})\) or head pose \((M=3.44 \text{ s}, SD=1.07 \text{ s})\). We found no significant differences between gaze and head pose \((p > .05)\). Results are summarised in Figure 4. We also found that authentication time is significantly different across switches \((p < .05, \text{Figure 4})\).
Entry Accuracy
Entry accuracy is the number of correct entries during authentication. The average successful accuracy was 90.80% across all conditions with 88.2% (SD = 32.2%) for eye gaze, 92.3% (SD = 26.8%) for head pose, and 91.9% (SD = 27.3%) for controller tapping. This is inline with prior work on authentication in VR with 82% [21] and 93% [11].

Study 2: Security Evaluation
We invited 15 participants (5 females, 10 males) aged between 17 and 44 years (M = 26.6, SD = 6.79) with the objective of role play bystanding attackers and observe the experimenter during authentications. To motivate participants to perform well, they took part in a lottery for an additional £8 voucher where the chance of winning increases as they correctly guess more PINs.

Design and Procedure
We added IV3) Threat Model with three conditions: Pen and Paper, 3D Replica, and Video Recordings as additional independent variable. We trained the participants by: a) introducing them to the arrangement of the digits and surface colours of RubikAuth, b) allowing them to enter multiple PINs using all pointing methods, and c) running training attacks on all pointing methods. PINs were entered by the experimenter, while we simulated the three threat models with the participant as the attacker. Each participant performed 36 attacks against: 1 PIN × 4 switches × 3 pointing methods × 3 threat models. This results in overall 540 observation attacks. Attacks were performed on 36 predefined unique PINs to ensure fairness of comparisons. Participants were told which pointing method will be used and the beginning and end of the authentication process.

Successful Attack Rate
We measured the successful attack rate, i.e., the percentage of times the correct PIN was guessed. Attacks were successful 8 out of 540 times (1.48%): 0 against eye gaze (0%), 4 against head pose (2.22%), and 4 against controller tapping (2.22%). 7 out of 8 (87.5%) successful attacks were on 0-switch PINs. Results are summarised in Table 2.

Attack Accuracy
To gain better insights on how close the guesses are to the entered PINs, we calculated the Euclidean distance between the centre of the entered PIN symbol (users’ inputs) and the centre of each guessed PIN symbol (attackers’ guesses). While previous work used Levenshtein distance to measure similarity of guesses [5, 11, 22], we opted for Euclidean distance (ED) because it better reflects spatial distances between targets on different surfaces. An attack is considered more successful if the resulting ED between the guess and the actual PIN is shorter.

To study the effect of the independent variables on similarity of their guesses to the correct PINs, we ran a three-way repeated measures ANOVA. No significant three-way interaction was found (p > .05). We ran subsequent two-way ANOVA tests where two-way interaction effects were found, and followed those by pair-wise comparisons using t-tests. We used Bonferroni for controlling familywise errors.

In case of threat model 1, where attackers used pen and paper to note their observations, attacks against controller tapping are significantly (p < .05) more successful when PINs contain 0-switches (M = 0.147, SD = 0.102), compared to 2-switches (M = 0.248, SD = 0.070) and 3-switches (M = 0.259, SD = 0.056), where 0 is a perfect match to the correct PIN and 0.37 is an unsuccessful attack. We also found that RubikAuth PINs that contain 0-switches are significantly more secure when entered using eye gaze (M = 0.261, SD = 0.075) compared to controller tapping (M = 0.147, SD = 0.102) (Fig. 5). When attackers use a smart phone to record and play back the authentications
(threat model 3), entering PINs using gaze ($M=0.251$, $SD=0.057$) is significantly more secure ($p < .05$) than head pose ($M=0.218$, $SD=0.079$) and tapping ($M=0.204$, $SD=0.046$) (Fig. 5).

To understand if the advanced threat models resulted in more successful attacks, we compared the accuracy of guesses by running multiple ANOVAs. We found a significant main effect of threat model on ED when using head pose ($F_{2,28}=4.317$, $p < .05$) and tapping ($F_{2,28}=5.576$, $p < .05$). Post hoc analysis using t-tests with Bonferroni correction confirmed the significant differences between threat model 3 ($M=0.218$, $SD=0.079$) and threat model 1 ($M=0.257$, $SD=0.044$) when using head pose, and between threat model 3 ($M=0.204$, $SD=0.045$) and threat model 2 ($M=0.239$, $SD=0.059$) when using tapping (Fig. 6).

**Discussion and Future Work**

**Using Manipulable 3D Objects for Authentication**

Our two user studies highlight the benefits of leveraging natural two-handed interaction for authentication in VR. Authentications with RubikAuth are fast and highly resilient to observation attacks, even in advanced threat models (100% for gaze-based interaction). This is also attributed to the high cognitive effort required to observe the manipulations and multiple visual channels, such as hand movements, at the same time [4, 19, 22, 29]. In a future work, we plan to conduct an in-depth analysis of users’ and attackers’ cognitive effort and plan to incorporate additional aspects of the human body such as foot-tapping for selection in RubikAuth as this could overwhelm attackers even more [25].

For high observation resistance when using controller tapping, we recommend to include at least one switch in RubikAuth PINs. Gaze performs well against all studied threat models even without switches but at the expense of longer authentication time. Qualitative feedback from the security study revealed that some poses allow selection from multiple surfaces without explicitly rotating the cube. This can be particularly effective against observations when combined with gaze pointing, and could potentially counteract the increased authentication time caused by rotating the cube. We recommend to leverage manipulable 3D objects for frequent authentications in VR as authentications are fast (1.69 s to 4.92 s) and highly secure (97.78% - 100%).

**Employing Suitable Threat Models**

Existing work focused mostly on one-time shoulder surfing attacks, and video attacks recorded using a stationary camera [8, 10, 11]. We employed three threat models that simulate a best case scenario for attackers. While successful attack rates did not differ significantly across the threat models, the accuracy of guesses increased significantly. This allowed us to gain a better understanding of the impact of switches and pointing methods on observation attacks. We argue that future evaluations of authentication schemes should employ advanced threat models like the ones considered in this paper to ensure realistic results.

**Conclusion**

We investigated authentication in VR using a manipulable 3D cube. We compared pointing using eye gaze, head pose, and controller tapping. We conducted two within-subjects experiments, a usability study (N=23) and a security study (N=15). We found that entering a four-symbol PIN using controller tapping is significantly faster (2.60 s) than head pose (3.44 s) and eye gaze (3.60 s). In terms of observation resistance, eye gaze outperformed head pose and controller tapping with a observation resistance of 100%, 97.78%, and 97.78% respectively. Our results suggest that providing attackers with support material contributes to more critical security evaluations.
REFERENCES


