Towards Secure and Usable Authentication for Augmented and Virtual Reality Head-Mounted Displays

Reyhan Duezguen*, Peter Mayer*, Sanchari Das†, Melanie Volkamer*
* SECUSO - Security, Usability, Society, Karlsruhe Institute of Technology
† University of Denver, Indiana University Bloomington
* firstname.lastname@kit.edu, † sanchari.das@du.edu

Abstract
Immersive technologies, including augmented and virtual reality (AR & VR) devices, have enhanced digital communication along with a considerable increase in digital threats. Thus, authentication becomes critical in AR & VR technology, particularly in shared spaces. In this paper, we propose applying the ZeTA protocol that allows secure authentication even in shared spaces for the AR & VR context. We explain how it can be used with the available interaction methods provided by Head-Mounted Displays. In future work, our research goal is to evaluate different designs of ZeTA (e.g., interaction modes) concerning their usability and users’ risk perception regarding their security - while using a cross-cultural approach.

1 Introduction
New-age technologies help to connect people despite geographical constraints. However, such technological evolution brings new risks. Augmented and virtual reality (AR & VR) are such technologies that have expanded considerably and are projected to reach $114 billion and $65 billion, respectively, by 2021 [44]. AR & VR systems like the Oculus and Google Glass increasingly promise to provide social activities like interactive gaming, virtual shopping, or attending virtual meetings [46]. Many of these activities happen in so-called shared spaces, i.e., places not strictly public, but where multiple people are present at the same time [22]. However, these technologies also introduce new security challenges in AR & VR [24], including authentication challenges. Nowadays, authentication on AR & VR systems is neglected or carried out on the smartphone or PC [6]. Yet, if authentication is required during a VR experience, e.g., paying for a product or entering a virtual conference, the user must take off the Head-Mounted Display (HMD), interrupting the virtual experience. Such challenges motivated our research direction to implement more secure and usable authentication strategies for AR & VR devices.

A naive approach using voice recognition technology of the HMD as an authentication strategy might put users at serious security risks, especially in public and shared spaces. Another method could be to use the available sensors for biometric authentication, e.g., gait recognition [18]. Such authentication schemes are designed for continuous authentication. The goal of our research is to focus on authenticating services when needed. Additionally, biometric-based approaches would also hamper authenticating with someone else’s HMD (as it would first need to be trained) and may have several privacy concerns. Thus, what is needed is a secure (especially in shared spaces) and usable authentication scheme, which only uses the sensors of the HMDs while being privacy-preserving.

Therefore, we are proposing a shoulder-surfing resistant authentication scheme that relies only on the equipment of the AR & VR HMDs.

The proposed authentication scheme is based on our previous research: the Zero-Trust Authentication (ZeTA) protocol [23]. In this paper we describe how ZeTA can be applied to the AR & VR context. Our future research goal is to implement the proposed authentication scheme using a user-centred development approach and conduct user studies to evaluate its usability and users’ risk perception. Note, since organizations aim to provide their products and services worldwide, it is in particular interesting to understand the cultural differences in the use and perception of upcoming technologies like AR & VR.

The importance of social and cultural aspects when investigating the acceptability and appropriateness of technology are shown in many papers [3, 12, 28, 49]. Hofstede’s [48] five cultural dimensions (namely power distance, individualism, masculinity, uncertainty avoidance and long-term orientation) are widely used to quantify national differences. These cultural dimensions showed many times an association towards technology use [1,13,50]. Some studies also discovered differences on perceived usability among different cultures [40,43]. The impact of cultural aspects on the use and acceptance of HMDs and authentication schemes has yet to be determined.
Thus, the study is going to be conducted in Germany and the U.S. for cross-cultural analysis.

2 Related Work

Prior research has proposed and developed different authentication schemes on HMDs. Yu et al. [53] and George et al. [20] investigated well-established concepts for the VR context, such as PINs or 2D and 3D sliding patterns within VR environments. These concepts, though helpful for authentication, have some security concerns. For example, bystanders can observe or even record the movement which can help them to guess the password from the controller’s action.

Additionally, for AR devices like Google Glass, Islam et al. [25] proposed tapping gestures on the glasses’ temple and use tapping patterns as a means to authenticate. Winkler et al. [52] introduced an authentication method that is more resistant to observations by using AR glasses in combination with the smartphone. The glasses show a randomly created PIN pad on the private display according to which the user can input password through their smartphone. Other proposals include biometric authentication based on head and body movement [31,37,39] or the human visual system [29,32,34]. These proposals require either additional hardware (such as a smartphone) or a training phase to capture the user’s biometric pattern. In contrast, our proposal requires neither.

For any proposal aiming to advance authentication for AR & VR devices, investigating societal and cultural aspects in technology adoption is critical. Prior studies have shown that authentication behaviour, usage, and experience is influenced vastly by age [11], cultural differences [2], and geographical locations [42,45,51]. Riley et al. investigated regional differences in the perception of biometric authentication in India, South Africa, and the United Kingdom [45]. Volkamer et al. observed in a field study PIN usage at ATMs and in various electronic payment scenarios in Germany, Sweden, and the United Kingdom [51]. Given prior evidence, it is essential to evaluate the impact of different countries when designing a new authentication scheme, especially for new-age technologies. These technologies, such as AR & VR are used worldwide where the demographical, societal, and cultural impact can play a critical role.

Yet, in the AR & VR space, we found very little research on cross-cultural aspects. Jung et al. [27] and Lee et al. [30] explored the cultural differences in the adoption of mobile AR in South Korea and Ireland. Few studies investigated the effect of web-based AR on online shopping and compared results from different countries inside of Europe [19, 41]. These studies identified differences in the use and perception of mobile and web-based AR applications between countries. Despite such critical research, to our knowledge, there are no cross-cultural studies in AR & VR with HMDs. Thus, comparing HMD usage in different countries in the context of authentication will be novel and, therefore, very valuable.

3 Proposed Solution

The goal of this work is to propose an authentication scheme for the AR & VR devices, which is resistant to observation and only relies on the sensors integrated into the most AR & VR HMDs. Our proposed authentication scheme is based on our previous research on observation resistant authentication: the Zero-Trust Authentication (ZeTA) protocol [23]. Here, we first provide a summary of the ZeTA protocol and explain how it could be applied in the AR & VR context.

3.1 Zero-Trust Authentication (ZeTA)

ZeTA is a knowledge-based authentication protocol, i.e., the user has to memorize a secret analogously to text passwords. In this section we describe its working principle, which is also illustrated in fig. 1.

The general idea of ZeTA is to expand upon the human capacity to build up semantic networks of related concepts and is thus based on innate human-based computation. To that end, ZeTA requires a knowledge base of concepts (e.g., words or symbols) and their semantic relations. The users’ secrets in ZeTA consist of two or more concepts and logical connections between them (i.e., AND, OR, NOT), e.g., “yellow OR wheel”. This secret is generated and assigned to the user by ZeTA during the enrolment of the user. The enrolment has to be performed through a private channel between the system and the user.
The authentication is based on a challenge-response interaction. The user has to determine whether a specific attribute is related to their secret or not, e.g., if the secret was “yellow OR wheel” and the challenge was “sunflower”, then the correct answer would be “yes”. Note that all challenges are pre-generated as part of the creation of the user secret and stored as described in [23]. Thereby, the secret is chosen such that it partitions the knowledge base equally in yes and no challenges (i.e., half of the attributes are related to the secret and half of the attributes are not related to the secret).

Due to its design, ZeTA can allow errors in responses by the users to compensate for innate differences in users’ interpretations of the semantic relations between concepts. This can potentially increase ZeTAs usability but might impair security if the two are not carefully balanced. It also highlights the importance of cultural effects. The system repeats the challenge-response protocol until the desired certainty threshold is achieved; i.e., the probability of the user being an imposter is sufficiently small. Consequently, ZeTA can be scaled seamlessly to arbitrary security levels. When user errors are not allowed during an authentication attempt, according to [23] ZeTa can easily reach PIN-level security with 14 challenges. The usual online guessing threshold of $10^6$ [16] can be achieved as easily using 25 challenges, while even allowing for one error by the user [23].

As stated above, the enrolment procedure of ZeTA relies on a private channel. In contrast, after the enrolment, ZeTA was designed with the threat model as introduced by Matsumoto and Imai [35] in mind. The attacker can compromise the communication channels and even the user’s device. Thus, ZeTA relies only on the server being secure. Proofs for lower bounds on the number of observations required to learn a secret based on a probably approximately correct learning model are presented in the original publication [23].

### 3.3 Future Design and Implementation

The proposed authentication scheme will be implemented as mock-up for both AR & VR HMDs, as well as for each interaction method. The development is based on a human-centered design approach: the mock-ups are tested and improved iteratively by evaluating different design variations of the outputs and inputs with users to maximize the authentication scheme’s usability. Options for the output to show the challenges are text, image, and audio. Options for the input of the responses are: voice, head movement, and buttons/touch controls (cf. section 3.2). There might also be different approaches to give feedback to the user after answering each challenge or to proceed from one challenge to the next one.

### 4 Proposed Methodology for User Evaluation

As future work, we will evaluate the three interaction methods of the proposed authentication scheme through in-lab user studies. We are planning to use Google Glass for the AR application and the Oculus Rift S for the VR application. The study design is built upon our research on shoulder-surfing resistant authentication using gamepads [36].

### 3.2 Application in the AR & VR context

Augmented and Virtual Reality HMDs provide various interaction methods depending on the capabilities of the device. Examples of input systems are controller, head movement, gesture, and voice recognition. The core output system is the private display (i.e., optics that create the virtual image) combined with audio. The idea underlying the usage of ZeTA in the AR & VR context is that the challenge is shown on the display of the HMDs. The user responses are entered using input options, which can be found in most of the AR & VR HMDs. Thus, we avoid dependencies on additional hardware. Concerning the entry of the response to the system’s challenge by the user, the following interaction options can be used: 1) voice control, 2) head movement, and 3) buttons on the VR controller or touch controls on the AR glasses. Additionally, finding the right number of challenges as a trade-off between usability and security while considering the specifics of the AR/VR context is an important aspect of the development of the ZeTA implementations for our user study.

The advantage of using ZeTA as an authentication scheme for the AR/VR context is that shoulder-surfing resistance does not need to be empirically evaluated due to the aforementioned security proofs. Therefore, the lower bound of needed observations holds no matter whether the attacker observes the communication channel, the user’s interaction, or even the private display of the HMD. The user input can even be processed by the web server and does not need to burden the capacity of the HMD without impairing the user’s privacy. The only time the user is required to use a non-compromised device and a private channel is when being assigned the secret by ZeTA during enrolment.

### 4.1 Research Goal

The evaluation of the authentication scheme for each of the AR & VR HMDs and each of the three interaction methods (voice, head movement, and touch/press) will be based on usability criteria and users’ risk perception regarding the authentication protocol security. Usability is measured by users effectiveness, efficiency and satisfaction with the authentication scheme. Thus, our research goal for future work is: Identifying the best interaction method for authenticating through ZeTA on both, AR & VR HMDs, i.e., the method that provides the highest effectiveness, efficiency, and satisfaction as well as the lowest perceived risk by users regarding the security of the authentication process.

We aim to inspect the cross-cultural influence by conducting identical studies in Germany and the United States. Germany and the U.S. are interesting cultures to compare because of their global influence in the field of technology [21, 38]. Both of these nations share much in common (democratic governments, similar linguistic roots), they also have some interesting differences (ethical heterogeneity, capitalistic versus socialistic approach) [47]. Additionally, it is predicted that the AR/VR market will rise globally, especially in U.S. (96.1% Compound annual growth rate (CAGR)) and in Western European countries (104.2% CAGR), including Germany [8].
4.2 Study Protocol

After completing the implementation, a pre-study is planned to pilot and refine the study protocol of the main study, which is described below. The authentication scheme is tested with each combination of the device (i.e., AR, VR) and interaction method (i.e., voice, head gestures, button/touch controls) regarding its usability and users’ risk perception regarding its security mechanism. The study will be conducted in both, Germany and the United States. Therefore, 12 (2x3x2) groups are used to collect data as visualized in Fig. 3.

Each participant will test all three interaction methods. To avoid first-order carryover effects, the allocation of the participants will be specified with the Latin Square Design [7] that counterbalances sequential effects. The procedure of the main study is presented in figure 2. We will ask two participants to come to the lab simultaneously. Both of the participants will receive an explanation of the ZeTA scheme and will be given a user scenario with three different randomly generated passwords. Then, we will run a 3-step evaluation process:

1. Participant-1 authenticates on the HMD three times. Participant-2 observes the process.
2. Now they change roles: participant-2 authenticates on the HMD three times. Participant-1 observes the process.
3. Both participants answer questions in a survey as well as we conclude with a short semi-structured interview.

By having two participants in the lab simultaneously, we aim to create a higher validity setting with respect to evaluating users’ risk perception. Secrets will be assigned to the participants by the system. Each of them will have time alone to memorize their secret. As a baseline for the configuration, we propose to use the online guessing resistance threshold of $10^6$ [16]. This is in line with the envisioned types of accounts used on the HMDs (e.g., purchasing media content from an online service). Before conducting the study, we will ask for ethical approval. Participants will be compensated based on the minimum wage regulations in the U.S. and Germany.

The effectiveness will be measured by the ratio of correct password entries among the three. Efficiency will be assessed by the average time needed for authentication across the three passwords. Satisfaction will be measured with the System Usability Scale (SUS) that covers users’ subjective reactions to using the scheme [4]. To examine the user’s risk perception, the scales proposed by Fischhoff et al. [15], Liang & Xue [33], and Das [9] will be adapted to our use case. The risk perception metric is defined by nine characteristics of the risk: 1) voluntariness, 2) immediacy, 3) knowledge of the exposed, 4) knowledge of experts, 5) control, 6) newness, 7) common-dread, 8) chronic-catastrophic, and 9) severity. Offline, this framework informed four decades of research in risk perception and public policy in a diversity of risk domains, e.g., environmental risk [17] and health risk [26]. Online, this framework has been used to explain perceptions of technical security risks [5, 10] and insider threats [14].

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) in the Competence Center for Applied Security Technology (KASTEL), Karlsruhe Institute of Technology; Secure and Privacy Research in New-Age Technology (SPRINT) Lab, University of Denver; and Human and Technical Security (HATS) Lab, Indiana University. Any opinions, findings, and conclusions or recommendations expressed in this material are solely those of the author(s).

References


