Dynamic Keypad – Digit Shuffling for Secure PIN Entry in a Virtual World

Andrew Holland and Tony Morelli(✉)

Central Michigan University, Mt Pleasant, MI 48859, USA
{holla1aa,morel1a}@cmich.edu

Abstract. As virtual reality becomes more mainstream there is a need to investigate the security of user level authentication while in the virtual world. In order for authentication methods to be useful, they must be secure, not allow for any external observers to determine the secure data being entered by the user, and also not break the immersion that the virtual world provides. Using head mounted virtual reality displays, users can interact with the world by using gaze, that is selecting objects by what the user is focusing on. This paper analyzes the security issues involved with utilizing gaze detection for secure password entry. A user study finds security issues with standard gaze based PIN input, and as a result a solution to this problem is presented. The solution shuffles the numbers on the PIN pad and finds that method to be more secure while maintaining accuracy and speed.

Keywords: Virtual reality · Security · Password

1 Introduction

Head mounted displays (HMD) for virtual reality environments allow for users to experience a virtual world without being able to see anything within their real physical environment. Typical HMDs are devices like the Oculus Rift or HTC Vive that require the device to be tethered to a computer. Other devices such as Google Cardboard or Samsung Gear VR do not require a tethered PC and are powered by a standard cell phone. While in a virtual environment it is difficult for a user to utilize a physical keyboard. In order to use a virtual keyboard within the virtual environment, standard virtual object selection is used where the user moves his head around to line up a cursor with the desired target, then holds still for a period of time until the target is selected. Although this works, it also can reveal what direction the user is looking at to other people located in the same physical space. These physical head movements could reveal how the user is interacting with a virtual keyboard and as a result secure information such as PIN codes or passwords could be captured. This paper presents a user study that reveals observers can guess the correct 5 digit PIN code within 1 digit 72% of the time, and it also provides a results of a Dynamic Keypad concept that eliminates this possibility with no significant decrease in speed or accuracy.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 102–111, 2018.
https://doi.org/10.1007/978-3-319-91581-4_8

2 Background

This paper focusses on user authentication in Virtual Reality. Virtual Reality is an environment where the user cannot see the real world. A fixed display is located in a headset very close to the user's own eyes and the user is completed immersed in the world presented on the display. Commercially available Virtual Reality headsets include the Oculus Rift and the HTC Vive. These headsets are required to be tethered to a computer that is generating the content for the display.

The Oculus Rift contains a screen resolution of 1080×1200 per eye, a 90 Hz refresh rate and a 110° field of view. The Rift has built in headphones and built in rotational tracking. In addition to sensors built into the headset, the Oculus Rift also contains optional external USB sensors (known as the Constellation sensors) for more accurate positional tracking. The tracking units track the headset as well as optional hand held controllers called the Oculus Touch. The touch controllers are tracked by the Constellation sensors and also have sensors to determine hand gestures made by the user.

The HTC Vive contains a similar architecture as the Oculus Rift. The Vive headset also contains a 1080×1200 resolution display for each eye. Unlike the Oculus Rift, the Vive contains a front facing camera that permits the user to see the real world through a video representation. The VR tracking system for the HTC Vive is a series of Vive Base stations that give a 3 dimensional view of the area and include tracking the headset and the handheld controllers. These base stations, known as the Lighthouse tracking system are larger than the sensors used in the Oculus Rift.

The PC based systems such as the Rift and Vive are a little costly and also require a decent computer to run them. For a more budget friendly VR experience devices such as Google Cardboard and Samsung Gear VR allow users to experience VR by inserting a regular smart phone into a head mounted display. These types of VR devices rely on many of the sensors, display, and processing power built into the phone. The common feature between these virtual reality devices is that the user cannot directly see the surrounding environment as the display is limited to the screen located directly in front of the user's eyes.

Virtual Reality differs from augmented reality or mixed reality in that as opposed to the user seeing only from a projected screen as in virtual reality, the user can also see the environment with his own eyes and additional information is projected on a transparent screen. The main difference is that a user can still easily see items such as keyboards to input information. Augmented devices such as Microsoft's HoloLens also allow the user to interact with the device through the use of hand gestures. These hand gestures may pose their own security risks, but it is beyond the scope of this paper. This paper focuses on data entry using head motions on virtual reality devices. The rest of the paper is organized by analyzing the current state of the art when it comes to object selection in a virtual space, then current authentication methods are analyzed, and finally a new method with some experimental results is presented.

3 Related Work

In order to properly evaluate potential methods of user level validation in virtual reality, different types of validation and object selection both inside and outside of virtual reality were reviewed. Some of the most relevant examples are listed below and they are broken down into three different categories: Interaction Techniques, Security, and Representation.

3.1 Interaction Techniques

Manipulating objects virtual using a pointing device such as a Nintendo WiiMote was common several years ago. However, the issue with the WiiMote is that it senses its position by using the controller based Infrared camera to locate LEDs in the room. In order for it to function properly, the LEDs must be set up and visible by the controller. A method (Chuah and Lok 2015) for manipulating objects with a standard phone was investigated. They were able to successfully integrate navigating in a virtual reality based environment with a standard phone by utilizing the built-in sensors on the phone.

Using a user's own hands are a method of interacting within virtual reality if the virtual reality system is equipped with sensors capable of sensing where the user's hands are located in 3D space. Although this is a common method, it still has many issues (Argelaguet and Andujar 2013) with real world target selection. That study found issues with occlusion, visibility mismatch and depth perception in stereoscopic displays. It also found that interactions in virtual reality is more physically demanding than that of normal computer interactions. The study also found pointing limitations within the human motor system.

Selection techniques for wearable virtual reality systems were evaluated in another study (Brancati et al. 2015). In this study three different types of object selection tasks were analyzed including wait to click, air tap and thumb trigger. In this study it was shown that wait to click was the most effective. In this type of interaction, the user will point with his finger at an object and hold the finger there for a period of time before the system recognizes the gesture as a desired selection action. The implementation used in this paper is based on this wait to click process, however instead of using cameras to track the position of the hand, gaze detection was used.

More target selection research (Velloso et al. 2015) analyzed the difference between gaze based selection and hand based selection in both 2D and 3D virtual environments. The results of that study show that gaze selection is not only faster but more preferred for users tasked with selecting objects in virtual reality.

Looking at how people best interact with objects in virtual reality, it was shown (Geiger et al. 2017) that people interact best in a virtual environment when additional feedback is given to them. In this case, additional hand color feedback was given in order to maximize the performance.

3.2 Security

Computer security is a very important topic and it directly translates to virtual reality. It has been shown (Das et al. 2014) that social factors can have an effect on how users

perceive security behaviors. Most importantly, it was shown that the observability of security features was a key factor in socially motivated behavior change. Because of that the user study described here involves an observer phase. This helps both with the determination of the security of entering PINs in virtual reality as well as showing users how they might best interact with such an environment.

It has been shown (Fiebig et al. 2014) that a user's interaction with a standard smartphone can be determined by examining the user's face through the built in smartphone camera. Modern smartphone cameras are equipped with a high enough resolution camera such that the users actions and even reflections off of the cornea can be determined through the front facing camera. In that Fiebig user study, users were tasked with entering PIN numbers on to a smartphone while the smart phone snapped pictures of them with the front facing camera. Different users looked at the snapshots and were able to determine the PIN numbers being entered by the users.

Augmented or mixed reality systems allow the user to see with his own eyes and that may result in a different type of authentication. One approach (Roesner et al. 2014) utilized a chrome based plugin for secure web browsing authentication. In this case, the user is authenticating to a website through the augmented reality system. When a password is required, the browser will popup a unique QR code and the augmented reality system will then display the password on the user's display. As the user is already authenticated with the augmented reality device, all lookups for the password are done completely outside the computer requesting the authentication.

3.3 Representation

Determining how to design the virtual key pad for PIN based entry, design techniques were analyzed. One project (Ragan et al. 2015) found field of view and scene complexity as two factors leading to better performance. The larger the field of view, and the lower the complexity resulted in better performance. These two factors were taken into consideration when designing the keypad and its interactions.

Commercially, (Chang and Gupta 2017) a patent has been granted dealing with object manipulation in virtual reality. In this patent, users can interact in a virtual reality based inter-net browser by spinning a virtual representation of a three dimensional graphical representation of internet search results. In this example, although the user is viewing the data in three dimensions within the virtual world interactions are performed using a computer mouse and pointer very similarly to how interactions are performed on a standard computer. Although this is a novel approach, the system described in this paper utilizes a 2D PIN pad located in three dimensional space.

4 Methods for Validation

In order to properly authenticate a user in a virtual world there are several options. The user could authenticate before entering the virtual world. The user could authenticate while in the virtual world, or the user could leave the virtual world mid-session in order to perform the authentication. Each of these are discussed below.

4.1 Before Entering the World

Using this authentication method, a user would authenticate herself by utilizing a standard authentication method before putting on the virtual reality headset. The advantages of this method include the access to any existing or future standard computer based authentication techniques.

If the virtual reality headset is attached to a standard computer, the user could authenticate herself through a standard user name and password entry through the keyboard. A successful validation will allow the user to put on the headset and remain in the virtual world as the authenticated user.

If the virtual reality headset is not attached to a standard computer, but is using a mobile phone as the screen, the user also has validation options prior to putting the mobile device into the headset. The user can be validated through a text username/password as mentioned above, or the user could use any type of biometric validation methods provided by the device's operating system such as fingerprint, iris, or facial recognition.

Although using existing validation methods has its advantages, there are also some disadvantages. One issue using these methods is that it is not a seamless experience for the user. The user must perform all the validation actions, then get into the virtual environment. Many times this ends up with the user having the headset halfway on to view the computer screen while the credentials are entered. Or the user will enter the credentials onto the mobile device, but then accidentally hit the power button on the phone as it was attempted to be inserted into the headset. This results in a frustrating experience.

In addition to a frustrating experience, it also allows for credentials to be shared. For example, a user will authenticate with the system, then the headset could be passed around from person to person without requiring a new authentication thus leaving the first person logged into the system for the duration of the experience no matter how many different people use the headset.

4.2 External Mid-Session Validation

Another approach would be to use the standard validation equipment for the platform, i.e. Mobile or Computer based, in the middle of the experience. When credentials are required, the user is requested to remove the headset and enter the correct information through a keyboard or through biometric sensors on a mobile device. This type of approach is good because it is using standard authentication methods, and can be used for multiple users.

The downside of an external mid-session validation is that it breaks the immersion of the experience. If a person is in a virtual world, and then travels to a different section or a completely different virtual world, it is likely that a subsequent validation will have to take place. Taking off the headset and potentially removing a mobile device from the headset is very time consuming and not natural for a person who is in a virtual world. This type of interruption could not only break the immersion for a single user in a single session, but it could jeopardize the entire virtual reality industry as user authentication

is necessary for all computing, and immersion is necessary for virtual reality. This approach makes it difficult for both of these to ever be true.

4.3 Internal Mid-Session Validation

An alternative approach is to have the user perform the authentication within the virtual environment itself. This will address the security concerns, as well as allow the user to remain immersed and preserve the virtual reality experience. The direct path to internal validation is to use a virtual keyboard that a user will select keys on using some kind of virtual pointer. The rest of this paper analyzes this topic as there are issues surrounding this type of user authentication. There are also many benefits. User Study.

4.4 Design

In order to investigate if characters entered from a virtual keyboard could be recognized from another person observing the user in the virtual world, a user study was created. The first part of the user study was to familiarize the user with virtual keyboard entry within a virtual world using two different types of keypads. In the virtual world a standard 9 digit key pad was shown on the screen. The Static Keypad was organized in 3 columns and 4 rows with the numbers appearing in sequence starting with the keys 1, 2, 3 in the top row, 4, 5, 6 in the second row, 7, 8, 9 in the third row and in the bottom row the keys were B, 0, E. The two additional buttons B (backspace) and E (enter) allowed for the participants to correct any errors and to signify that the PIN entry was complete. The Dynamic Keypad was identical to the Static Keypad, except that the location of the numerical buttons was shuffled. Participants were instructed to select a particular virtual key by staring at the desired key for 1 s. Participants in the user study were given a random 5 digit pin code before going into the virtual environment and once in the virtual environment they were tasked with entering the 5 digit code followed by selecting the E virtual key to complete the task. Each participant entered the pin code twice. Times for each character entry, total time to enter the full pin, and any errors were recorded.

After the user entered the PIN codes within the virtual environment, the user was tasked with observing two videos of a person performing the pin entry task. All participants viewed the same 2 videos and were shown each video twice. After each video was complete, users were asked to write down the PIN code being performed in the video. The first video was of a person using the Static Keypad. The second video showed a user performing the PIN entry using the Dynamic Keypad, where the digits on the key pad were shuffled and shown in a random order. Both videos were taken from the rear perspective of the User to simulate a more realistic shoulder surfing scenario. The subjects were instructed to deduce the 5 digit PIN of the subject within the video by observing the orientation of their head. Upon completion of the video observations the users were queried on their observations from the video, and what strategies they used to deduce the PINs.

4.5 Implementation

The Pin-Based virtual reality authentication system consists of a back-end server, and a front-end user interface.

Back End Server
The Windows Operation system, Apache HTTP server, MySQL database, and PHP server-side scripting (WAMP) web development platform provides several functions for the study: (1) it contains the scripts that provide verification and authentication within the system, (2) it houses the two log systems; (i) the primary log system tracks every authentication attempt and outcome as well as the user identification number, and duration of the authentication session, (ii) the secondary log system is used to track pin positions for inputs on dynamic PIN pad interfaces, and input time in all sessions, (3) maintains a database table that contains the records of all users, and user PINs which is referenced to authorize a user's authentication attempt, and send the unlock screen notification.

Front-end User Interface
The two user interfaces are composed of a static user log-in screen and a dynamic user log-in screen.

- Static Log-in Screen
  The statically created log-in screen (Fig. 1) posted within the virtual reality environment was crafted as a three by four grid with each index holding a single numeric or command value. The model was based on standard PIN pads for automatic teller machines or credit card readers and is depicted below. The log-in screen was developed using the C# programming language within the Unity IDE version 5.5.1f1.

Fig. 1. Static screen showing standard digit placement.

- Dynamic Log-in Screen
  The dynamically generated log-in screen (Fig. 2) displayed within the virtual reality environment generates a three by four grid with each index randomly assigned a numeric value of zero to nine or a command value of enter or backspace; represented

as “E” and “B” respectively, which is depicted below. The log-in screen was developed using the C# programming language within the Unity IDE version 5.5.1f1.

Fig. 2. Dynamic screen showing random digit placement

In addition to the specific PIN pads shown above, the users would be able to see the current series of numbers selected, as well as the completed pin once the E key was pressed (Fig. 3).

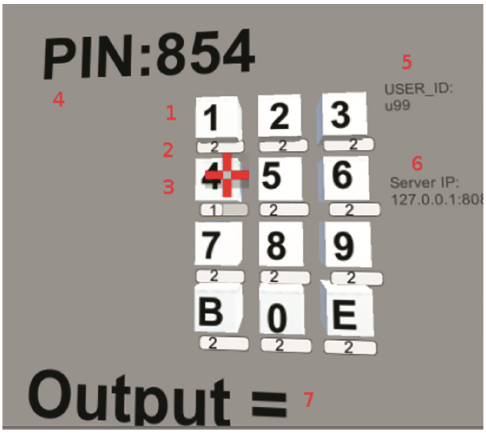

Fig. 3. User display with additional information. Numerical identifiers are as follows: (1) Static Keypad, (2) Timer counter indicating the number of seconds until that digit is selected, (3) Current selected number, (4) Current digits of the PIN, (5) User identifier, (6) IP Address of the server, and (7) Output label indicating whether or not the PIN was successfully validated.

5 Results

The user study consisted of 16 participants. It found no significant difference in speed or accuracy for users entering PIN data using the Static Keypad vs the Dynamic Keypad.

Participants observing a person entering a PIN using the Static Keypad were able to accurately determine the exact PIN 31% of the time, and with only 1 digit incorrect 41% of the time. That is 72% of the time the PIN was either exactly known or known within 1 digit. Users observing someone using the Dynamic Keypad were never able to guess the PIN within 2 digits.

The total results of the initial session showed that PIN entry speed decreased between the first 2 trials with the second PIN entry being on average 57.5% of the first (SD = 0.279). The average time to enter a complete accurate password for the second trial was 20.42 s (SD = 4.7). This is the time to enter the 5 digit PIN code and press enter with a 1 s wait time to acknowledge the user was really intending to select a digit. It is also noteworthy that none of the participants in their second attempt entered any incorrect digits. The Static Keypad was on average 2.4% faster than the Dynamic Keypad. An analysis of variance (ANOVA) shows no significant variation between the time to enter a PIN on the Static vs the Dynamic keypad, $F(1, 12) = 0.88$, $p > 0.5$.

6 Conclusion

Virtual reality presents many great possibilities. However, security in a virtual environment may have holes depending on its implementation. Object selection has restrictions based on the hardware in use. Some devices contain external controllers, others utilize cameras, and others (mixed reality) allow the user to interact with a standard keyboard. However there are many scenarios where a user is wearing a head mounted display with no access to external devices and must perform a user level authentication within the virtual space. This paper investigated the security of using head position to select a pin number on a standard keypad and found that it was possible for an external observer to determine the PIN number entered by the user by simply watching the head movements. In order to overcome that limitation, a possible solution was presented that shuffled the digits. An identified potential downside to this method could be more errors and less speed. The results did not identify any significant downside from the digit shuffling method. Due to no significant speed increase and no significant decrease in accuracy, yet an extreme increase in security, it is recommended that Dynamic Keypads be used for PIN entry in virtual environments.

References

Alhadidi, B., Arabeyat, Z., Alzyoud, F., Alkhwaldeh, A.: Cloud computing security enhancement by using mobile PIN code. JCP **11**(3), 225–231 (2016)

Argelaguet, F., Andujar, C.: A survey of 3D object selection techniques for virtual environments. Comput. Graph. **37**(3), 121–136 (2013)

Brancati, N., Caggianese, G., Frucci, M., Gallo, L., Neroni, P.: Touchless target selection techniques for wearable augmented reality systems. In: Damiani, E., Howlett, R.J., Jain, L.C., Gallo, L., De Pietro, G. (eds.) Intelligent Interactive Multimedia Systems and Services. SIST, vol. 40, pp. 1–9. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19830-9_1

Chang, S., Gupta, A.: U.S. Patent No. 9,720,562. U.S. Patent and Trademark Office, Washington, DC (2017)

Chuah, J.H., Lok, B.: Experiences in using a smartphone as a virtual reality interaction device. Int. J. Virtual Real. (IJVR) **11**(3), 25–31 (2015)

Das, S., Kim, T.H.J., Dabbish, L.A., Hong, J.I.: The effect of social influence on security sensitivity. In: Proceedings of SOUPS, vol. 14, July 2014

Fiebig, T., Krissler, J., Hänsch, R.: Security impact of high resolution smartphone cameras. In: WOOT, August 2014

Geiger, A., Bewersdorf, I., Brandenburg, E., Stark, R.: Visual feedback for grasping in virtual reality environments for an interface to instruct digital human models. In: Ahram, T., Falcão, C. (eds.) AHFE 2017. AISC, vol. 607, pp. 228–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60492-3_22

Huang, H., Lin, N.C., Barrett, L., Springer, D., Wang, H.C., Pomplun, M., Yu, L.F.: Analyzing visual attention via virtual environments. In: SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments, p. 8. ACM, Chicago, November 2016

Ragan, E.D., Bowman, D.A., Kopper, R., Stinson, C., Scerbo, S., McMahan, R.P.: Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. IEEE Trans. Vis. Comput. Graph. **21**(7), 794–807 (2015)

Roesner, F., Kohno, T., Molnar, D.: Security and privacy for augmented reality systems. Commun. ACM **57**(4), 88–96 (2014)

Velloso, E., Turner, J., Alexander, J., Bulling, A., Gellersen, H.: An empirical investigation of gaze selection in mid-air gestural 3D manipulation. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 315–330. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22668-2_25