AirAuth: Towards Attack-Resilient Biometric Authentication Using In-Air Gestures

Md Tanvir Islam Aumi
UbiComp Lab, DUB Group
University of Washington
Seattle, WA 98195, USA
tanvir@cs.washington.edu

Sven Kratz
FX Palo Alto Laboratory
3174 Porter Drive
Palo Alto, CA, 94304, USA
kratz@fxpal.com

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2581157

Abstract

AirAuth is a biometric, gesture-based authentication system based on in-air gesture input. We describe the operations necessary to sample enrollment gestures and to perform matching for authentication, using data from a short range depth sensor. We present the results of two initial user studies. A first study was conducted to crowd source a *simple* gesture set for use in further evaluations. The results of our second study indicate that AirAuth achieves a very high Equal Error Rate (EER-)based accuracy of 96.6% for simple gesture set and 100% for user-specific gestures. Future work will encompass the evaluation of possible attack scenarios and obtaining qualitative user feedback on usability advantages of gesture-based authentication.

Author Keywords

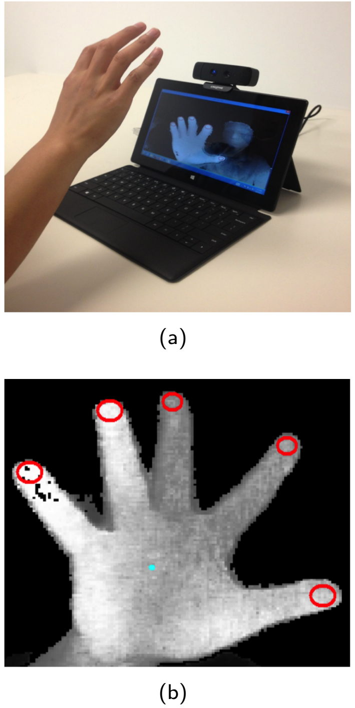
in-air gestures; authentication; shoulder surfing; user experience; acceptability; usable security

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction

A growing number of users store and manipulate important and sensitive information online, on their


(a)

(b)

Figure 1: (a) A user is performing an authentication gesture in front of a short range depth camera. (b) Hand features (3D locations of finger tips and hand center) captured and used by AirAuth.

personal computers and mobile devices. As such, finding methods of secure and easy-to-use authentication is of increasing importance, since tradeoffs exist between the users' desire for security and the compromises in user experience they are willing to take [5].

At present, passwords and PIN numbers are the most widely-used authentication methods for gaining access to PCs, mobile devices and online accounts, and they are well-understood by the users. However, such knowledge-based systems can have disadvantages, such as requiring the user to learn complex passwords (for increased security vs. "trivial" passwords). Also, as the number of accounts or devices the user needs to access grows, the mental burden increases for the user to remember multiple passwords or variants thereof. That is one reason why a number of users resort to using only a few passwords (perhaps with simple variants) for all of their authentication activity [15], which puts the users at significant risk if one of these passwords is compromised.

Traditional password entry can be prone to *shoulder-surfing attacks* [13]. The growing amount of video surveillance in public spaces and the potential of misusing it compounds this danger. In addition to shoulder-surfing attacks, mobile devices equipped with touch screens are prone to *smudge attacks* [2]. Our prototype is resistant to smudge attacks since it is entirely touchless. We also believe it to be highly resistant towards shoulder-surfing, since beyond the knowledge based component required for authentication (the gesture) our system implicitly uses two traits that are unique to the user: the way the users enter their gesture (made possible by the time-based data stream during gesture entry) and the users' unique hand geometry (made possible by capturing the location of multiple distinct points on the user's hand).

To address the issues mentioned previously, we are developing a novel authentication system, *AirAuth*, that uses hand gestures made in the vicinity of a computing device (Figure 1(a)). Using a short-range depth camera, our system tracks hand gestures input by the user. In contrast to longer-range systems such as Kinect, the short-range depth camera allows us to track individual finger tips and the center of the user's hand in 3D space (Figure 1(b)). This data provides us with an abundance of features allowing the decoding the user's authentication secret, classifying biometric properties of the user's hand, and classifying the movement style of the user.

In this paper we present two initial results: (1) a crowd-sourced gesture set that is used in subsequent evaluations of our system, and (2) an initial user study to determine the accuracy of our system.

Related Work

For the purposes of this abstract we will only cover the most relevant of a large body publications related to the subject: [9, 15] highlight problems with traditional password entry, and [5] describes the tradeoff between usability and security. A wide range of input devices and sensors have been used to gather biometric data for authentication. Jorgensen et al. explore mouse dynamics as a biometric for behavioral authentication [4]. Similarly, keystroke dynamics have also been explored [7]. Alpcan et al. propose signature-like strokes on a touchpad as a biometric [1]. This concept has been extended to multi-touch surfaces with good results [10]. Motion gestures on mobile phones have been proposed as an authentication biometric. During a gesture input, the user moves the entire phone through the air. The movement is

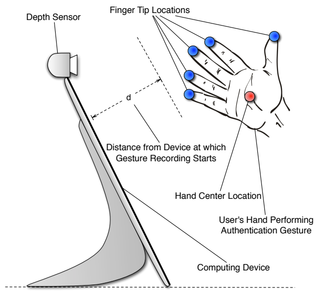

Depth Sensor
Finger Tip Locations
d
Distance from Device at which Gesture Recording Starts
Hand Center Location
User's Hand Performing Authentication Gesture
Computing Device

Figure 2: Prototype sketch of our system. A depth sensor is arranged on a computing device so that it can observe the user's hand while it makes in-air gestures for authentication. The depth sensor measures the 3D locations of the user's fingertips and of the hand center.

recorded by the mobile phone's motion sensors (typically an accelerometer and a gyroscope). The results presented so far in the literature appear promising [3, 6, 8]. A relatively closely related touch-less biometric based on magnetic tracking has been proposed by Sahami et al. [11]. In contrast to that work, AirAuth tracks multiple points in free space without the requirement for additional user instrumentation. corollary

The AirAuth System

Our hardware prototype (Figure 2) consists of a Creative Senz3D[1] short range depth sensor placed on a computing device, in our case a Microsoft Surface Pro Tablet (Figure 1(a)), in such a way that the sensor can image the user's hand while she makes authentication gestures in the vicinity of the device. When the user's hand passes a previously set threshold distance from the device, gesture recording starts. The recording stops when the user's hand moves beyond the threshold distance.

Data Processing

We used the Intel Perceptual Computing SDK[2] to extract the 3D locations of the users fingertips and also the hand center from the depth image of the camera. The extracted features (tip and hand center locations) are a biometric that AirAuth uses implicitly for user authentication. However, our system is not limited to these specific features. Other types of persistent features can also be used given a means to extract them reliably. Since a gesture entry typically takes around 1–2 s to perform, our biometric consists of a time-based stream of the aforementioned features.

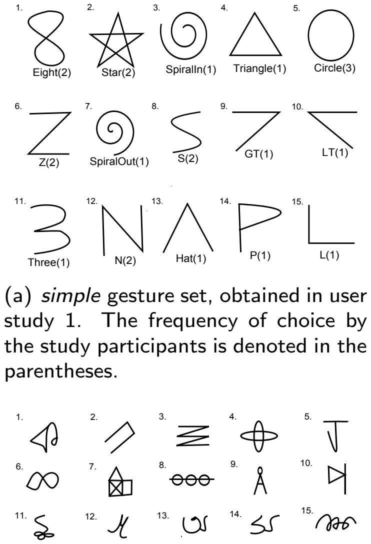
From the depth sensor, we obtain raw 3D coordinates (x, y, and z) of all (detected) fingertips of the user and also the user's hand center. However, the location of user's hand with respect to the depth sensor can be anywhere and in any orientation in the air and differ one instance to another. Hence, location and orientation of the hand need to be normalized before applying any matching algorithm. Therefore, we first normalize the data to a $[0, 1]$ interval by scaling the raw 3D points, then mean-shift the data [14].

Enrollment and Authentication Phases

New users undergo an enrollment phase where they are each required to enter three samples of their desired authentication gesture. These three gestures are averaged and stored as a gesture template for the particular user. Together with the gesture template, we also store an authentication threshold ε, which is calculated as the

[1] The pixel resolution of the depth image of this sensor is 320x240 and its diagonal field of view is $73°$. We configured the sensor to capture images images at a rate of 30 Hz.

[2] http://software.intel.com/en-us/vcsource/tools/perceptual-computing-sdk



(a) *simple* gesture set, obtained in user study 1. The frequency of choice by the study participants is denoted in the parentheses.

(b) *Complex* gestures defined by individual users, obtained in user study 2.

Figure 3: Simple (a) and complex (b) user-defined gestures

maximum pairwise distance between the three initial authentication gestures. For authentication, our system compares the gesture data input by the user with the template stored by the system and calculates a distance score. If this distance score is $> \varepsilon$ the input is rejected, if it is $\leq \varepsilon$ the user is authenticated. We use Dynamic Time Warping (DTW) [12] to calculate the distance between gesture inputs both in the enrollment and authentication phases.

Initial User Studies

We present the results of two initial user studies. In the first one, we aimed to find a set of simple gestures users would be willing to use. The second study was conducted to test the enrollment and authentication phases, with the goal of determining initial accuracy values for our system.

User Study 1: Finding a Set of Simple Gestures
One of the research goals for AirAuth is to test the accuracy and vulnerability of a predefined, *simple*, set of in-air gestures. The gestures to be identified were intended to be simple and easy to reproduce for the users, and to be performable with one or multiple fingers.

Therefore, as a pilot study, we conducted a short interview session with 10 participants from an industrial research lab. We asked participants to come up with up to three different in-air gestures, that they thought were simple and would be easy to reproduce. Participants were free to choose any gesture they wanted, as long as the gestures could be performed with both one and multiple finger(s). As a result of this study, we obtained a set of 15 *simple* gestures. Figure 3(a) shows the 2D sketches of the gesture suggestions we obtained from the study.

If we examine the gestures closely, we can see that some participants chose their first name initial (gesture IDs 6, 12, 14, and 16), favorite number (gesture IDs 1, 5, and 11), shape, symbol, etc. Participants chose certain gestures because, according to them, they were easy for them to perform and also to remember later without any effort. It should also be noted that some of the gestures were chosen more than once by different participants. This shows that there was at least some consensus regarding the choice of gesture amongst the participants. Some people also chose the existing 2D gesture they use to unlock their touch-based mobile phone since they were already comfortable with the 2D version of the gesture.

User Study 2: Initial Study AirAuth's Accuracy
In this study, we investigated the accuracy of our system across multiple users. For this study, we recruited 15 participants. 11 were male and 14 of them were right handed. All but one of them was fairly experienced with smartphones and computers. 4 of them were familiar with in-air gestures (mostly using Kinect). The participants' age ranged from 18 to 46 ($\mu = 28.3$, $\sigma = 6.5$). We note that this study does not reflect the accuracy under real-world attack scenarios. Rather, the goal was to verify the viability of the enrollment process and the ability of the system to discern between gesture entries of different users. Before each user study session, we explained the concept of our system briefly to the participants. After that, they were asked to fill out a short questionnaire with demographic information and then proceeded to the gesture trials. To test our system with more complex and personalized gestures, we decided to add the participant's signature (performed as a hand movement in the air) and one custom complex gesture for each participant. So, each participant was asked to enter a total of 17 gestures: the 15 fixed gestures from Figure 3(a), one signature, and one self-defined complex gesture. Figure 3(b) portrays 2D sketches of the self-defined *complex* gestures each

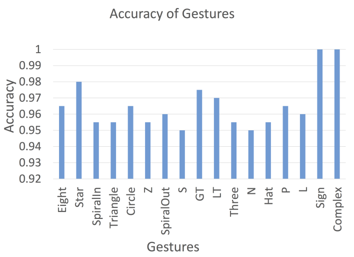


Figure 4: Achieved accuracy of the gestures used in the study. Note that the user-defined gestures (*sign* and *complex*) achieved 100% accuracy.

participant came up with. We note that each of the complex gestures were thought to be more secure and less vulnerable to attack from the participants' point of view. Before entering each gesture, the participants were asked to practice the gesture a few times. Once comfortable, we asked them to perform the gesture 20 times: 10 times with single finger and 10 times with multiple fingers. In case of multiple fingers, participants were free to choose any number of fingers they wanted. To get rid of the carryover effect, we counterbalanced the order of gestures to enter between participants.

Accuracy Analysis

To determine the accuracy of our system across multiple users, we calculated the EER, a point where (at a certain threshold setting) the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. FAR and FRR are inversely related. Thus, decreasing one measure will result in an increase in another. Calculating the EER is a standard method in the literature of measuring the accuracy of authentication systems [1]. We measured FAR and FRR using the following formulae:

$$\text{FAR} = \text{number of accepted forgeries / total number of forgery cases} \tag{1}$$

$$\text{FRR} = \text{number of rejected genuine authentications / total number of genuine authentications} \tag{2}$$

We measured the EER of single and multiple finger gestures separately. For each gesture, we had 10 single and 10 multiple finger samples from each user. As mentioned previously, our prototype requires users to provide 3 samples of their password gesture during the enrollment stage and the average of those are used as the reference template. Therefore, we decided to do the same for EER calculation and used the average of the first 3 samples for each gesture as a reference template for that gesture. Generating the template gesture in this way reflects the state of a hypothetical deployed AirAuth system that has been newly enrolled. The last 7 samples were used as a test case. Trials with the same gesture from other participants were used as forgery cases. So, for each gesture performed with single or multiple fingers, we had $7 \times 15 = 105$ gestures to calculate FRR and $10 \times 14 \times 15 = 2100$ gestures to calculate FAR. Table 1 shows the value of EER for our different gestures. Figure 4 also shows a graphical version of the result where we averaged the single and multiple finger accuracy of each gesture. From Table 1, we can see that average EER for single and multi finger gestures are 0.038 and 0.029, respectively. Thus, gestures using multiple fingers appear to be more secure than single finger gestures. This trend is also observable in the EER of individual gestures, which is always lesser or equal for the multi finger version of a gesture. From our results, we can thus conclude that multi finger gestures are at least as secure as single finger ones. A possible explanation for this is that a larger amount of information usable for authentication can be extracted from the participants during multiple finger gestures.

Gesture	**EER (Single Finger)**	**EER (Multiple fingers)**
01 (Eight)	0.04	0.03
02 (Star)	0.02	0.02
03 (SpiralIn)	0.04	0.04
04 (Triangle)	0.05	0.04
05 (Circle)	0.04	0.03
06 (Z)	0.05	0.04
07 (SpiralOut)	0.04	0.04
08 (S)	0.05	0.05
09 (GT)	0.04	0.01
10 (LT)	0.04	0.02
11 (Three)	0.06	0.03
12 (N)	0.06	0.04
13 (Hat)	0.04	0.04
14 (P)	0.05	0.02
15 (L)	0.04	0.04
16 (Sign)	**0.0**	**0.0**
17 (Complex)	**0.0**	**0.0**
Average (all gestures)	0.038	0.029

Table 1: Equal Error Rate for each gesture, for single finger input and multiple finger input. Note that the *sign* and *complex* gestures have an EER of 0.

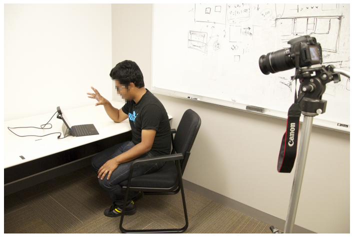


Figure 5: Experimental setup for a shoulder surfing study. Participants acting as legitimate users are videotaped from one or multiple angles while entering authentication gestures, so that other subjects, acting as *attackers*, can later try to forge gesture entries.

Discussion and Future Work

Our initial results indicate that AirAuth, our novel gesture-based authentication technique achieves a significantly higher EER-based accuracy than previous systems (e.g.,[10, 11]). AirAuth has the following benefits: the sensor instrumentation required is relatively simple compared to related systems and the proposed type of gesture-based authentication could significantly improve the usability as well as user experience of authentication tasks on mobile as well as fixed devices. In the future we thus intend to also conduct more qualitative user evaluations, for instance on location appropriateness of gesture entry and also on accuracy–usability tradeoffs for simple vs. complex gestures. Furthermore, for an evaluation of real-world performance, we will need to test how the system performs under a realistic attack scenario, for instance shoulder surfing, one of the most obvious real-world threat scenarios for AirAuth. Figure 5 shows a possible experimental setup of such a study: *legitimate* users are filmed from one or multiple angles when entering their authentication gestures. A disjoint set of *attacker* users will be shown the videos and asked to imitate the gestures based on the videos.

References

[1] Alpcan, T., Kesici, S., Bicher, D., Mihçak, M. K., Bauckhage, C., and Çamtepe, S. A. A lightweight biometric signature scheme for user authentication over networks. In *Proc. of the 4th international conference on Security and privacy in communication networks*, ACM (2008), 33.

[2] Aviv, A. J., Gibson, K., Mossop, E., Blaze, M., and Smith, J. M. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX conference on Offensive technologies*, USENIX Association (2010), 1–7.

[3] Farella, E., OModhrain, S., Benini, L., and Riccó, B. Gesture signature for ambient intelligence applications: a feasibility study. In *Pervasive Computing*. Springer, 2006, 288–304.

[4] Jorgensen, Z., and Yu, T. On mouse dynamics as a behavioral biometric for authentication. In *Proc. of the 6th ACM Symposium on Information, Computer and Communications Security*, ACM (2011), 476–482.

[5] Kainda, R., Flechais, I., and Roscoe, W. A. Security and usability: Analysis and evaluation. In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*, IEEE (2010), 275–282.

[6] Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing 5*, 6 (2009), 657–675.

[7] Monrose, F., and Rubin, A. D. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems 16*, 4 (2000), 351–359.

[8] Patel, S. N., Pierce, J. S., and Abowd, G. D. A gesture-based authentication scheme for untrusted public terminals. In *Proc. UIST 2004*, ACM (2004), 157–160.

[9] Riley, S. Password security: What users know and what they actually do. *Usability News 8*, 1 (2006).

[10] Sae-Bae, N., Ahmed, K., Isbister, K., and Memon, N. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In *Proc. of the 2012 ACM annual conference on human factors in computing systems*, ACM (2012), 977–986.

[11] Sahami Shirazi, A., Moghadam, P., Ketabdar, H., and Schmidt, A. Assessing the vulnerability of magnetic gestural authentication to video-based shoulder surfing attacks. In *Proc. CHI 2012*, CHI '12, ACM (New York, NY, USA, 2012), 2045–2048.

[12] Sakoe, H., and Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing 26*, 1 (1978), 43–49.

[13] Tari, F., Ozok, A., and Holden, S. H. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on Usable privacy and security*, ACM (2006), 56–66.

[14] Vatavu, R.-D., Anthony, L., and Wobbrock, J. O. Gestures as point clouds: a $p recognizer for user interface prototypes. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, ACM (2012), 273–280.

[15] von Zezschwitz, E., De Luca, A., and Hussmann, H. Survival of the shortest: A retrospective analysis of influencing factors on password composition. In *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, 460–467.