My Thoughts Are Not Your Thoughts

Benjamin Johnson

Carnegie Mellon University johnsonb@andrew.cmu.edu

Thomas Maillart

University of California, Berkeley thomas.maillart@ischool.berkeley.edu

John Chuang

University of California, Berkeley chuang@ischool.berkeley.edu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

UbiComp '14, September 13 – 17 2014, Seattle, WA, USA Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-3047-3/14/09 ... \$15.00.

http://dx.doi.org/10.1145/2638728.2641710

Abstract

Authenticating users of computer systems based on their brainwave signals is now a realistic possibility, made possible by the increasing availability of EEG (electroencephalography) sensors in wireless headsets and wearable devices. This possibility is especially interesting because brainwave-based authentication naturally meets the criteria for two-factor authentication. To pass an authentication test using brainwave signals, a user must have both an inherence factor (his or her brain) and a knowledge factor (a chosen passthought). In this study, we investigate the extent to which both factors are truly necessary. In particular, we address the question of whether an attacker may gain advantage from information about a given target's secret thoughts.

Author Keywords

brainwave signals, impersonation, mobile security, wearable authentication, EEG, passthoughts

Introduction

Users look for both security and usability in the mobile computing devices that they use every day. We want our devices to protect access to our private data; but we largely refuse time-consuming efforts to make it happen. We rather prefer that the software (or hardware) makes the security part easy for us. In short, we want our computing security to also be usable.

This goal of usable security has motivated¹ a variety of technologies in recent years that promise to change the way we authenticate ourselves to our computing devices. For example, the most recent iPhone's fingerprint scanner authenticates users by reading a biometric signal. Other means of authentication are made possible by the availability of wearable computing devices that incorporate accelerometers and gyroscopes, ECG and EMG sensors, and even brainwave signal readers.

Among these examples, brainwave signal readers stand out to many as an exciting option, because the idea of an authentication system based fundamentally on our thoughts seems at once both secure and convenient. It seems secure because our thoughts are perceived intuitively as private; and it seems convenient because we always have our thoughts available and can usually access them efficiently.

Moreover, brainwave signals don't seem to suffer from the same vulnerabilities as other biometric data used for authentication, such as fingerprints and irises. These vulnerabilities are often dramatized in popular films, in which fingerprint scanners and iris readers are readily fooled through clever uses of wax moldings or printing technologies. Our thoughts, in contrast, seem secure because we control our own minds; and imaginations for changing this are grounded squarely in the science fiction category.

Brainwave signal authentication itself is not science

fiction. Over the last two years, a number of researchers have investigated the serious possibility of using consumer-grade single-channel brainwave signal readers to authenticate users into a computer system. The first such project [3] involved a recruited user base of 15 students, who gave samples of their brainwave signals for a set of 7 tasks. The authors developed a rubric for matching the data from those signals to their originating subjects, and designed an authentication system based on task customization with a failure rate as low as 1%. That effort demonstrated for the first time the plausibility of using single-channel brainwave signal readers for authentication.

The case for authentication using brainwaves is made yet stronger by the availability of consumer-grade brainwave sensors and their integration with wearable devices (e.g., integration with augmented reality glasses [4]), as these typically do not have keyboards, and often times do not even have a touchscreen, making password-based and PIN-based authentication difficult.

In this paper, we examine the system from [3] for robustness against deliberate attacks from thought impersonators. While the previous work showed that a closed authentication system can be designed to work effectively, our interest in this study is attuned toward the extent to which that system can be broken. If an attacker knows your passthought, and puts on the brain reader, what is now the likelihood that they can successfully authenticate as you? Our overall goal is to investigate the susceptibility of brainwave signal authentication to impersonation attacks; and a key technical objective is to measure the extent to which the attacker is aided by more information about the defender's passthoughts.

Our methodology and results can be summarized as follows. First, we calibrate the original authentication

¹The motivation for new authentication technologies is grounded in data from numerous experimental studies showing how security is degraded by some heuristics that users actually employ when dealing with text-based passwords.

system from [3] using a subset of their signal data, to get a new system that correctly authenticates all of its own users when they authenticate as themselves, and repels authentication attempts from unauthorized users (or attackers) within the same user base with a false acceptance rate of 2%. We then test the strength of this system using more than 6000 new impersonation attacks. The impersonation efforts succeed more than the baseline system average, but the total false acceptance rate for impersonators remains under 5%. To assess the effects of information on these new attacks, we group all impersonation attacks according to the information conditions of the attacker relative to the defender, and we examine the aggregate trends in terms of subjects and in terms of tasks. We find that while there is a slight overall trend favoring more information in the aggregate, these results are inconsistent across different tasks and different subjects. The two main takeaways are that:

- 1. the authentication system is relatively robust against impersonation attacks in general, and
- the extent to which passthought knowledge aids an attacker is strongly dominated by the extent to which not having exactly the defender's brain hinders the attacker.

The rest of the paper is organized as follows. After reviewing related work, we describe the experimental setup for collecting brainwave signals. Then we describe our general methodology for analyzing the brainwave data, and extend this methodology to address our research questions about impersonation attacks. Finally, we discuss implications of our results before concluding.

Related Work

A series of studies since 2002 have demonstrated that brainwave signals can be used to both identify and to authenticate users with high accuracy. Earlier work employed clinical-grade multi-channel EEG sensors [6, 9, 10, 11], while more recent work extended the results to consumer-grade multi-channel [2] and consumer-grade single-channel [3] EEG technologies.

In most of these studies, the experimental subjects performed identical mental tasks, and so the authentication protocols had to differentiate the brainwave "signatures" that are distinct to each subject. This is in line with classical biometric authentication where individuals are identified based on their distinctive physiological characteristics such as fingerprints, iris patterns, and heart rate variability [1].

The passthoughts-based approach [3, 12] allows users to select their own personal secret (e.g., a particular song to sing in their head) as their authenticator. A key advantage of this approach is that the passthoughts can be more easily changed when desired than the user's inherent brainwave signatures or fingerprints or iris patterns. Furthermore, the passthoughts approach offers the possibility of two-factor authentication, where the two factors are the user's brainwaves ("who you are") and their chosen secret ("what you know").

Authentication based on keystroke dynamics [8] is similar to authentication based on passthoughts. Users are authenticated based on their typed password as well as their typing rhythm. Killourhy and Maxion evaluated a range of anomaly detectors and found that even if an attacker has gained knowledge of the password, they can still be detected, based on their typing rhythms, up to approximately 90% of the time [5]. Martinovic et al. undertook the first study of security attacks on consumer-grade brain-computer interface (BCI) technologies [7]. They demonstrated that it is feasible to launch a side-channel attack to gain private information regarding, for example, the user's bank card PIN or date of birth. Specifically, the entropy of the private information can be reduced by 15-40% compared to random guessing. However, they did not consider an impersonation attack against a BCI-based authentication system.

Experiment

Existing Brainwave Data

Our research extends prior experimental work [3] involving human subjects, in which brainwave signals were collected from 15 university students. In that study, the subjects performed a series of 7 mental tasks. Signals were collected ten times per subject per task. The data analysis in that study yielded an authentication protocol which correctly authenticated each brainwave trial from a given test set with 99% accuracy.

Impersonation Data

To examine the robustness of this system to impersonation attacks, we collected additional brainwave signals from three researchers (the authors of this paper). The experimental methodology followed all guidelines approved by an Institutional Review Board.

We collected data from each impersonator in multiple sessions, selecting tasks primarily to correspond to the best tasks (for authentication purposes) for subjects in the original study. Each impersonator recorded ten trial samples for each selected task. Additional samples were also collected to assess the effectiveness of various attack strategies – for example, attempting an impersonation without knowing the subject's task, intentionally using the wrong task, using the correct task with an unknown secret, or using the correct task with the wrong secret.

Description of Mental Tasks

The 7 tasks used in the previous study – and also for our impersonation efforts – are described below. For all but one of the tasks, a single trial lasts for a duration of ten seconds. A trial for the color task lasts 30 seconds. Here we describe each task in terms of its instructions for the subjects.

Breathing Task (breathing)

Close your eyes and focus on your breathing for 10 seconds.

Simulated Finger Movement (finger)

Imagine in your mind that you are moving your right index finger up and down in sync with your breathing, without actually moving your finger, for 10 seconds.

Sports Task (sport)

Select a specific repetitive motion from a sport of your choosing. Imagine moving your body muscles to perform this motion, for 10 seconds.

Song/Passage Recitation Task (song)

Imagine that you are singing a song or reciting a passage for 10 seconds without making any noise.

Eye and Audio Tone Task (audio)

Close your eyes and listen for an audio tone. After 5 seconds, the tone will play; upon hearing the tone, open your eyes and stare at the dot on the piece of paper in front of you for an additional 5 seconds.

Object Counting Task (color)

Choose one of four colors – red, green, blue, or yellow. You will be shown on a computer screen a sequence of six images. Each image contains a 5x6 grid of colored boxes. As each grid appears, count, silently in your mind, the number of boxes corresponding to your chosen color. A new grid will appear after each 5 seconds. This will continue 6 rounds for a total of 30 seconds.

Pass-thought Task (pass)

Choose your own pass-thought. A pass-thought is like a password; however, instead of choosing a sequence of letters and numbers, one chooses a mental thought. When instructed to begin, focus on your pass-thought for 10 seconds.

Brainwave Signal Collection

Brainwave signal data from each trial was transmitted via a bluetooth network connection from the Neurosky MindSet headset to a computer. The raw data includes single-channel EEG signals in both the time and frequency domains. Our analysis focuses on the power spectrum data, a two-dimensional matrix which gives the magnitude of the signal for every frequency component at every point in time.

The original study involved fifteen subjects, seven tasks, and ten trials per task, for a total of 1050 trials. The current work supplements this corpus with brainwave signals from three impersonators, each performing a varying selection of task genres, 44 genres in total, with ten trials per genre, for a total of 440 additional impostor trials.

Data Analysis

To use brainwave signal trials for authentication, we need both a clear method of representing trials, and a mechanism for comparing them. From the onset this is not a trivial task, as the signal itself can be represented in multiple forms, and each individual trial may have a different length due to slight discrepancies in the recording times. Our analysis thus requires us to first compress the data in a systematic way, that allows a consistent method of comparison.

Data Compression

We begin our data analysis by processing the power spectrum data to compress each trial. The method is as follows. For each trial not corresponding to the color task, we extract the middle five seconds in the temporal dimension. For color trials, we extract a five-second temporal component corresponding to the transition between the first and second image. In the frequency dimension, we extract data corresponding to the alpha wave (8-12 Hz) and the beta wave (12-30 Hz) ranges.²

Finally, we compress the signal in the time dimension by taking the median magnitude of each frequency over all time. This compression yields a one-dimensional column vector with one entry for each measured frequency.

Signal Differentiation

Our authentication system relies fundamentally on the notion of signal similarity. We expect signals coming from the same subject to be similar, and for signals coming from different subjects to be dissimilar. Our chosen metric for capturing this notion is the cosine similarity metric on vectors. For two vectors u and v, their cosine similarity is given by the equation:

similarity
$$(u, v) = \frac{u \cdot v}{\|u\| \|v\|}.$$

²Beta and alpha waves are associated with normal waking consciousness and wakeful relaxation, and are the most well-studied brainwave patterns in neuroscience research.

Similarity gives a value between 0 and 1. If two signal vectors are perfect scalar multiples of one another, then their similarity is 1. If the signals have non-intersecting support in their frequency components, then their similarity is zero. In practice all our non-identical signals fall between these two extremes.

Subject Authentication

To develop an authentication system we must extend the comparison of data trials to a comparison among subjects. Our methodology for this step relies on one fundamental observation – that two trials from the same subject tend to have higher similarity than trials between that subject and another subject. The authentication system derived from this observation reports its result for a given trial by asking one basic question. Does this trial look more like me on average, than it does everyone else?

Formally, for a fixed task, we define the test similarity between a test trial and a subject to be the average similarity between the test trial and all trials from the subject for that task. We define the test cross similarity between a test trial and a subject to be the average similarity between the test trial and all trials from other subjects for the same task. Our authentication method assigns to each trial a score, which is the difference between the test self similarity and the test cross similarity, normalized by the test self similarity. A test trial is authenticated if this score is above a certain threshold. The thresholds in our system were chosen to minimize the total error rate of the authenticator, but could also be adjusted to account for preferences between false rejects and false accepts. Formally, the authenticator accepts a test trial for a given subject if and only if

 $\frac{\text{test self similarity} - \text{test cross similarity}}{\text{test self similarity}} > \text{threshold.}$

Impersonation

Prior efforts [3] give evidence that an authentication system can be built based on EEG signals recorded by consumer-grade single-channel devices, but many questions concerning the security of such a system remain open. In particular, the existence of a secure closed system does not imply robustness against various forms of attacks from outside the system. The results in this section make progress in addressing this issue.

Authentication System Optimization

Our first step was to develop a closed 15-subject authentication system using the data from [3]. For each subject, we determined a best task and a customized threshold for the trial authentication protocol. The protocol was then further optimized through the elimination of outlier trials. Outlier trials were determined based on two criteria. Either the trial was specifically coded by the recording software as having poor signal quality, or the trial was sufficiently dissimilar to other trials from the same subject/task pair to cause a degradation in the authentication system.

Note that the authentication system we developed here for impersonation attacks differs from the one developed in [3]. Our system optimizes parameters to minimize the authentication error for every trial in the dataset, whereas the earlier authentication system used a random set of training trials to determine authentication parameters, and tested these parameters on the subset of remaining trials. Our system takes advantage of more trials for training. We then use a new set of impostor trials for testing. The impostor trials were not used in configuring the authentication system.

Table 1: Best Tasks and Optimal Thresholds

	Best	Optimal	Baseline	Baseline	Baseline
Subject	Task	Threshold	HTER	FRR	FAR
subject 0	color	0.10	0.0071	0	0.0143
subject 1	sport	0.08	0.0357	0	0.0714
subject 2	eye	0.12	0	0	0
subject 3	pass	0.13	0.0071	0	0.0143
subject 4	color	0.11	0.0286	0	0.0571
subject 5	color	0.17	0.0143	0	0.0286
subject 6	base	0.12	0.0286	0	0.0571
subject 7	sport	0.09	0	0	0
subject 8	finger	-0.02	0.0074	0	0.0147
subject 9	song	0.13	0.0143	0	0.0286
subject 10	base	0.10	0	0	0
subject 11	song	0.08	0	0	0
subject 12	eye	0.08	0	0	0
subject 13	sport	0.13	0	0	0
subject 14	pass	0.13	0.0071	0	0.0143
totals			0.0100	0	0.0200

Table 2: Impersonation attack success rates grouped by task

Best	Subjects	Baseline	Impostors	Unknown	Wrong	No	Unknown	Wrong	Correct
Task	w/ Task	FAR	FAR	Task	Task	Secret	Secret	Secret	Secret
base	2	0.0286	0.0943	0.0333	0.1013	0.0667	NA	NA	NA
finger	1	0.0147	0.0705	0.0667	0.0684	0.1	NA	NA	NA
eye	2	0	0.0148	0.0167	0.0132	0.0333	NA	NA	NA
song	2	0.0143	0.0295	0.0167	0.0297	NA	0	0.03	0.05
sport	3	0.0238	0.0455	0.0333	0.0414	NA	0.0333	0.0565	0.0857
color	3	0.0333	0.0348	0.0111	0.0381	NA	NA	0.0333	0.0167
pass	2	0.0143	0.0432	0.0167	0.0458	NA	NA	0.04	0.04
totals	15	0.0200	0.0450	0.0244	0.0465	0.06	0.025	0.044	0.05

Table 3: Impersonation attack success rates grouped by subject

	1	Best	Baseline	Impostors	Unknown	Wrong	No	Unknown	Wrong	Correct
	Subject	Task	FAR	FAR	Task	Task	Secret	Secret	Secret	Secret
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	0	color	0.0143	0.0659	0.0333	0.0743	NA	NA	0.025	0.05
2 eye 0 0.0182 0 0.0184 0.0333 NA NA 3 pass 0.0143 0.0114 0 0.019 NA NA 0.0 4 color 0.02571 0.0295 0 0.0286 NA NA 0.0 5 color 0.0258 0.0991 0 0.0114 NA NA 0.0 6 base 0.0258 0.0991 0 0.0114 NA NA 0 7 sport 0 0.0384 0 0.0144 NA NA NA 9 song 0.0286 0.0273 0.0667 0.0584 0.1 NA NA 9 song 0 0.0150 0.0333 0.027 NA 0.02 0.021 11 song 0 0.0118 0.0333 0.027 NA NA NA 12 eye 0 0.0144 0.0733 0.0778 </td <td>1</td> <td>sport</td> <td>0.0714</td> <td>0.0386</td> <td>0.0667</td> <td>0.0310</td> <td>NA</td> <td>0.05</td> <td>0.05</td> <td>0.05</td>	1	sport	0.0714	0.0386	0.0667	0.0310	NA	0.05	0.05	0.05
3 pass pass 0.0143 0.0114 0 0.0139 NA NA 0 4 color 0.0571 0.0295 0 0.0286 NA NA 0.0775 5 color 0.0226 0.0031 0 0.0114 NA NA 0.0775 6 base 0.0376 0.0386 0 0.0447 0 NA NA 7 sport 0 0.0386 0.0343 0.0897 NA NA NA 8 finger 0.0147 0.0705 0.0667 0.0548 0.1 NA NA 9 song 0.0273 0 0.0312 NA 0 0.014 10 base 0 0.0518 0.0533 0.0281 NA NA 11 song 0 0.0318 0.0333 0.0079 0.0333 0.079 0.0333 0.079 0.0333 NA NA 12 eye 0 </td <td>2</td> <td>eye</td> <td>0</td> <td>0.0182</td> <td>0</td> <td>0.0184</td> <td>0.0333</td> <td>NA</td> <td>NA</td> <td>NA</td>	2	eye	0	0.0182	0	0.0184	0.0333	NA	NA	NA
4 color 0.0571 0.0295 0 0.0286 NA NA 0.0775 5 color 0.0286 0.0991 0 0.0114 NA NA 0 6 base 0.0286 0.0991 0 0.0114 NA NA 0 7 sport 0 0.0932 0.0333 0.0897 NA 0.05 0.114' 8 finger 0.0126 0.0273 0 0.0312 NA NA NA 9 song 0.0286 0.0273 0 0.0331 NA NA 0.026 11 song 0 0.0114 0.0333 0.027 0.0333 NA NA 12 cyc 0 0.0118 0.0333 0.027 NA 0.0414 13 sport 0 0.0045 0 0.0077 NA NA 0.0122 14 spars 0.0143 0.0750 0.0333 0.077	3	pass	0.0143	0.0114	0	0.0139	NA	NA	0	0
5 color 0.0286 0.0091 0 0.0114 NA NA 0 6 base 0.0571 0.03866 0 0.0447 0 NA NA 7 sport 0 0.0932 0.0333 0.0897 NA 0.055 0.0147 8 finger 0.0147 0.0705 0.0667 0.0684 0.1 NA NA 9 song 0.0286 0.0273 0 0.0312 NA 0 0.020 10 base 0 0.1500 0.0667 0.1579 0.1333 NA NA 11 song 0 0.018 0.0333 0.0281 NA 0 0.0414 12 eye 0 0.0144 0.0333 0.0278 NA NA NA 13 sport 0 0.00454 0 0.00778 NA NA 0.0021 14 pass 0.01950 0.07444 0.0	4	color	0.0571	0.0295	0	0.0286	NA	NA	0.075	0
6 base 0.0571 0.0386 0 0.0447 0 NA NA 7 sport 0 0.0932 0.0333 0.0897 NA 0.05 0.114' 8 finger 0.0147 0.0773 0 0.0584 0.1 NA NA 9 song 0.0286 0.0273 0 0.0312 NA NA NA NA 11 song 0 0.0518 0.0333 0.0279 0.1333 NA NA 12 eye 0 0.0114 0.0333 0.0279 0.0333 NA NA 13 sport 0 0.0454 0.0778 NA NA 0.012 14 sport 0.00454 0.00540 0.00778 NA NA 0.0012	5	color	0.0286	0.0091	0	0.0114	NA	NA	0	0
7 sport 0 0.0932 0.0333 0.0897 NA 0.05 0.11 8 finger 0.0147 0.0705 0.0667 0.0584 0.1 NA NA 9 song 0.0286 0.0273 0 0.0312 NA 0.0 0.012 10 base 0 0.1500 0.0667 0.1579 0.1333 NA NA 11 song 0.0318 0.0333 0.0281 NA 0.04 12 eye 0 0.0114 0.0333 0.0271 NA NA 13 sport 0 0.0451 0.0333 0.0791 NA NA 14 pass 0.0150 0.0333 0.0778 NA NA 0.0012 14 pass 0.0150 0.0333 0.0778 NA NA 0.050	6	base	0.0571	0.0386	0	0.0447	0	NA	NA	NA
8 finger 0.0147 0.0705 0.0667 0.0684 0.1 NA NA 9 song 0.0286 0.0273 0 0.0312 NA 0 0.02 10 base 0 0.1500 0.0667 0.1579 0.1333 NA NA 11 song 0 0.0181 0.0333 0.0279 0.0333 NA NA 12 eye 0 0.0114 0.0333 0.0279 0.0333 NA NA 13 sport 0 0.0045 0 0.0079 0.0333 NA NA 14 pass 0.0143 0.0079 0.0333 NA NA 0.0012 14 pass 0.0045 0 0.0033 NA NA 0.0012 14 pass 0.0045 0.0056 0.0044 0.0162 0.0046 0.0046 0.0046 0.0046 0.0046 0.0046 0.0046 0.0046 0.0046	7	sport	0	0.0932	0.0333	0.0897	NA	0.05	0.1143	0.1663
9 song 0.0286 0.0273 0 0.0312 NA 0 0.02 10 base 0 0.1500 0.0667 0.1579 0.1333 NA NA 11 song 0 0.018 0.0333 0.0281 NA 0 0.04 12 eye 0 0.0114 0.0333 0.0291 NA NA NA 13 sport 0 0.0454 0 0.0034 NA NA 13 sport 0 0.0454 0 0.0037 NA NA 0 0.012 14 pass 0.0150 0.0333 0.0778 NA NA 0.0054 0.0150 0.0354 0.0450 0.0450 0.0667 0.067 0.067 0.0667 0.0667 0.0667 0.0667 0.0667 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666 0.0666<	8	finger	0.0147	0.0705	0.0667	0.0684	0.1	NA	NA	NA
10 base 0 0.1500 0.0667 0.1579 0.1333 NA NA 11 song 0 0.0318 0.0333 0.0281 NA 0 0.04 12 eye 0 0.0114 0.0333 0.0079 0.0333 NA NA 13 sport 0 0.0045 0 0.0078 NA 0 0.0122 14 pass 0.0143 0.0750 0.0333 0.0778 NA NA 0.0012 14 pass 0.0150 0.0333 0.0778 NA NA 0.0012	9	song	0.0286	0.0273	0	0.0312	NA	0	0.02	0.0333
11 song 0 0.0318 0.0333 0.0281 NA 0 0.04 12 eye 0 0.0114 0.0333 0.0079 0.0333 NA 13 sport 0 0.0455 0 0.0034 NA 0 0.0122 14 pass 0.0150 0.0050 0.00778 NA NA 0.0012 14 pass 0.0150 0.0054 0.0050 0.00778 NA NA 0.0050	10	base	0	0.1500	0.0667	0.1579	0.1333	NA	NA	NA
12 eye 0 0.0114 0.0333 0.0079 0.0333 NA NA 13 sport 0 0.0045 0 0.0034 NA 0 0.0122 14 pass 0.0143 0.0750 0.0333 0.0778 NA NA 0.0605 read-	11	song	0	0.0318	0.0333	0.0281	NA	0	0.04	0.0663
13 sport 0 0.0045 0 0.0034 NA 0 0.0125 14 pass 0.0143 0.0750 0.0333 0.0778 NA NA 0.0066 total 0.0200 0.0450 0.02044 0.0465 0.055 0.054	12	eye	0	0.0114	0.0333	0.0079	0.0333	NA	NA	NA
14 pass 0.0143 0.0750 0.0333 0.0778 NA NA 0.0667	13	sport	0	0.0045	0	0.0034	NA	0	0.0125	0
totals 0.0200 0.0450 0.0244 0.0465 0.06 0.025 0.044	14	pass	0.0143	0.0750	0.0333	0.0778	NA	NA	0.0667	0.1
101215 0.0200 0.0430 0.0244 0.0403 0.00 0.023 0.044	totals	-	0.0200	0.0450	0.0244	0.0465	0.06	0.025	0.044	0.05

Table 1 shows each subject's best task, the optimal thresholds, and the error rates for this authentication system. Note that the false rejection rates (FRRs) are all zero, meaning that every subject properly authenticates each of his or her own trials. The false acceptance rates (FARs) average 2%, implying that relatively few subject

trials are able to authenticate as any other subject. The tasks and thresholds were selected to minimize the half total error rate (HTER) which is defined as the average of the two aforementioned error rates.

$$HTER = \frac{FRR + FAR}{2}$$

Overall, the average HTER across the 15 subjects is 1.0%.

Impersonation Attacks

Our next step was to study the robustness of this system against outside attacks. For each of the 440 impostor trials, we tried to authenticate the trial against every subject using that subject's best task. The authentication system has 15 subjects, giving a total of 6600 impersonation attempts.

A motivating question for this research is to determine whether knowledge of a subject's chosen task or secret has an effect on the success rate of an impersonation attack. To address this question, we categorized each impersonation attempt based on the answers to a sequence of five yes-or-no questions.

First, is the impostor performing one of the specific 7 tasks, or not? If yes, then is this specific task the best task for the subject being impersonated, or not? If yes, then does the task have an additional secret, or not? If yes, then does the impostor have a particular secret in mind, or not? If yes, then does the secret match that of the subject being impersonated, or not?

These questions divide impersonation attempts into 6 distinct categories: *unknown task*; *wrong task*; correct task *no secret*; correct task *unknown secret*; correct task *wrong secret*; and correct task *correct secret*. By

evaluating the success probabilities within each category, we gain information about exactly how much it helps an impostor to know a particular task or secret.

Impersonation Results

The results of our impersonation attacks are presented in Tables 2 and 3. Table 2 groups the results by task, and Table 2 groups according to subject. In each table, the column "Baseline FAR" refers to the FAR of our closed 15-subject authentication system. The column "Impostors FAR" refers to the overall FAR aggregated over the six different information conditions of the impersonation attacks. The remaining columns represent the breakdown, by information condition, of the FAR achieved by the attackers.

From the task-centric view, we identify three trends.

First, for every task, the authentication rate for impostors is an improvement over the baseline FAR observed in the closed authentication system. Overall the improvement over the baseline rate was 125%. Some improvement should be expected because the parameters for the system were optimized specifically against subjects in the original dataset. Moreover, the acceptance rate for impostors is still fairly low at 4.5%.

Second, the impostor trials in categories of *unknown task* and *unknown secret* authenticated only about half as often as trials in the other categories. In the *unknown task* trials, impostors were free to choose any thought for the purpose of impersonation; and for the *unknown secret* trials, impostors were given a specific task category within song, sport, color, or pass, but were free to choose any thought within that category. The lower success rates of impersonation within these categories give some evidence that customized thoughts are more difficult to

impersonate when the customization is kept secret.

Third, within all the categories in which a specific task was known and performed, there was no clear evidence to indicate that an attacker benefits from having more information about the specifics of the task. When attempting to authenticate using the sport task, impostors could succeed more often when thinking of the correct sport, but the increase in success rate was mild, and moreover the opposite was true for the color task. In fact, for the color task, the highest rate of authentications was reached by impostors performing the wrong task. Finally, for the generic passthought task, knowing the secret thought did not make any difference in the acceptance rates.

We can separately identify two trends from the subject-centric view.

First, although every subject experienced some successful attacks, some subjects were more robust against attacks than others. Subjects 3, 5, 12, and 13 all had less than 1.2% false acceptance rates against impersonation attempts. Many of the successful attacks came against subject 7, against whose defenses impostors performing a silent rendition of "Freestyle swimming" authenticated nearly 17% of the time.

Second, there are very few observable trends in terms of dominance of information conditions that persist across all subjects. The closest candidate seems to be that no subject does the worst against impostor trials in the category of *unknown secret*. The other categories all have at least one subject who fares poorly against attacks of that sort. Subject 1 performed worst against impostor trials in the category of *unknown task*. Subjects 0, 3, 5, 6, 8, 10, 13, and 14 all fared worst against trials from the

wrong task category. Subject 2 did worst against trials with no secret. Subject 4 did worst against trials with the correct task but the wrong secret; while subjects 7, 9, and 11 did worst against impostor trials having the correct secret. Subject 12 did equally bad with trials from the categories of unknown task and no secret.

Discussion and Conclusion

A key premise underlying our investigative study is the notion that individual thoughts can be impersonated. The notion seems plausible for generic mental tasks such as breathing, and also for a number of more customized mental tasks such as jogging, swimming, or singing the happy birthday song. In the course of our investigations, however, we came across a few examples of customized thoughts that proved to be quite difficult to impersonate.

We illustrate this difficulty with a single example, in which we tried to impersonate subject 9, whose best identifying task was the "song" task, and whose secret choice was the "Serbian National Anthem". While we understood the subject to be familiar with this song, we had difficulty impersonating the thought on our own, due both to lack of familiarity and to the language barrier. (The language of the song is Serbian). Our solution to this dilemma involved playing a video of the Serbian National Anthem on YouTube, and pretending to sing along as we wore the brain reader. Interestingly, in this case, our attacks against this subject were more effective than average, compared to the overall success rate of impersonations in this study. However, if YouTube had not come to our rescue, the very notion that we could effectively impersonate this subject's thought would have been cast into doubt. This example serves to illustrate one of many challenges an attacker may face to impersonate a given thought. In particular, thoughts may be especially difficult to impersonate when

arising from a highly individualized experience or when involving an unfamiliar language. More generally, mental thoughts may be difficult to impersonate not just from a data comparison perspective, but also from considerations of more basic conceptual feasibility.

Features of a system that are challenging for an attacker are generally good for the system's security. To this extent, our observations support the feasibility of using brainwave signals as the basis for an authentication system, although many limitations remain. For example, the error rates for our system are still high compared to authentication systems based on other biometric data such as fingerprints. Our data set is relatively small compared to what we would likely need for a usable implementation. We only addressed one type of attack and showed that some secret knowledge can make these attacks more effective. Another potential limitation is perceived bias from having the paper's authors serving as attackers.

For future work it remains essential to develop data processing enhancements that further reduce the authenticator's error rates in general, to increase the number of test subjects, to use recruited attackers, and to integrate the data processing methodology with a real-time authentication framework. It also remains to address the susceptibility of brainwave authentication to other plausible types of attacks such as digital signal cloning.

Our goal in this paper was to measure the robustness of brainwave signal authentication systems against direct impersonation attacks. Building on the authentication system from [3], we simulated several thousand impersonation attacks and categorized the results to examine how knowledge conditions about the defender's secret thoughts might impact the attacker's success rates. We found that, in general, the success rate of impostor authentications is low, which bodes well for the feasibility of brainwave authentication systems in general. We also found that knowing the subject's task, or task secret, provided at best modest improvement over just thinking in a focused way about something completely different. Evidence for most of the trends observed in our results are at best mildly supported statistically. Our most robust result is that brainwave signal authentication systems appear to be fundamentally grounded in their more classical biometric components. That is, the extent to which my thoughts are not your thoughts is based less on "what you know" than "who I am".

Acknowledgements

This research was supported in part by the National Science Foundation under award CCF-0424422 (TRUST), by the Army Research Laboratory under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA) and by the Swiss National Science Foundation (Grant Nr. PA00P2-145368).

References

- [1] Bionym Nymi. http://www.bionym.com/.
- [2] Ashby, C., Bhatia, A., Tenore, F., and Vogelstein, J. Low-cost electroencephalogram (eeg) based authentication. In *Proceedings of 5th International IEEE EMBS Conference on Neural Engineering* (April 2011).
- [3] Chuang, J., Nguyen, H., Wang, C., and Johnson, B. I think, therefore i am: Usability and security of authentication using brainwaves. In *Proceedings of* 2013 Workshop on Usable Security (2013).
- [4] Devices, P. N. Press release: Pnd wearable environment. http://personalneuro.com/, 2014. [Online; accessed 30-May-2014].

- [5] Killourhy, K. S., and Maxion, R. A. Comparing anomaly detectors for keystroke dynamics. In *Proceedings of the 39th Annual International Conference on Dependable Systems and Networks* (DSN-2009), IEEE Computer Society Press (June 2009), 125–134.
- [6] Marcel, S., and del R. Millan, J. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 4 (April 2007).
- [7] Martinovic, I., Davies, D., Frank, M., Perito, D., Ros, T., and Song, D. On the feasibility of side-channel attacks with brain-computer interfaces. In *Proceedings of 21st Usenix Security Symposium* (Usenix Security) (2012).
- [8] Monrose, F., and Rubin, A. Authentication via keystroke dynamics. In Proceedings of the 4th ACM conference on Computer and communications security, ACM (1997), 48–56.
- [9] Palaniappan, R. Electroencephalogram signals from imagined activities: A novel biometric identifier for a small population. In *IDEAL 2006, LNCS 4224* (2006), 604–611.
- [10] Palaniappan, R. Two-stage biometric authentication method using thought activity brain waves. *International Journal of Neural Systems 18*, 1 (2008), 59–66.
- [11] Poulos, M., Rangoussi, M., Alexandris, N., and Evangelou, A. Person identification from the eeg using nonlinear signal classification. *Methods of Information in Medicine* (2002).
- [12] Thorpe, J., van Oorschot, P., and Somayaji, A. Pass-thoughts: Authenticating with our minds. In Proceedings of the New Security Paradigms Workshop (NSPW) (2005).