Pass-thoughts: Authenticating With Our Minds

Julie Thorpe P.C. van Oorschot Anil Somayaji

Digital Security Group, School of Computer Science
Carleton University, Canada
{jthorpe,paulv,soma}@scs.carleton.ca

April 18, 2005

Abstract

We present a novel idea for user authentication that we call *pass-thoughts*. Recent advances in Brain-Computer Interface (BCI) technology indicate that there is potential for a new type of human-computer interaction: a user transmitting thoughts directly to a computer. The goal of a pass-thought system would be to extract as much entropy as possible from a user's brain signals upon "transmitting" a thought. Provided that these brain signals can be recorded and processed in an accurate and repeatable way, a pass-thought system might provide a quasi two-factor, changeable, authentication method resilient to shoulder-surfing. The potential size of the space of a pass-thought system would seem to be unbounded in theory, due to the lack of bounds on what composes a thought, although in practice it will be finite due to system constraints. In this paper, we discuss the motivation and potential of pass-thought authentication, the status quo of BCI technology, and outline the design of what we believe to be a currently feasible pass-thought system. We also briefly mention the need for general exploration and open debate regarding ethical considerations for such technologies.

1 Introduction

The goal of user authentication is to establish a user's identity using one or more mechanisms, e.g. what the user is, knows, or has, or where they are. While textual passwords are by far the most commonly used method for user authentication in computer systems, the use of textual passwords for user identification has several well-known limitations: passwords have low entropy in practice (making them susceptible to dictionary attacks [28]), are often difficult to remember, and are vulnerable to "shoulder surfing," or observation by nearby third parties [45]. While proposed replacements for passwords typically do not have these same limitations, most schemes have other limitations or requirements. For example, biometric systems rely upon unchanging features that have a lifetime as long as the individual; this characteristic, combined with the threat of theft leaves biometrics (on their own) unsuitable for remote authentication. In contrast, smart cards can be used to securely authenticate users to remote servers, but at the cost of per-user hardware tokens.

Now imagine if we could authenticate by *thinking* a password. We could avoid the shoulder surfing problem associated with most "what you know" schemes by simply "transmitting" some chosen thought, authenticating with our minds. This type of authentication might also provide a "who we are" by virtue of the uniqueness of our individual brains (see §3). While such an idea might appear to lie in the realm of science fiction, recent advances in Brain-Computer Interface (BCI) technology give evidence that authenticating with our minds is within our technological reach.

The main goal of BCI research is to provide an alternative communication and control channel that does not depend on the brain's normal output pathway of peripheral nerves and muscles [53]. The driving application for BCI research is the communication and control of prosthetic devices for disabled patients. BCIs have been a hot topic for the past few years, making notable progress such as using a monkey's thoughts to control a robotic arm [37], and allowing paralyzed patients to communicate (albeit slowly) [5]. In general, these BCIs work by observing a brain signal S, extracting its features F, and then translating or classifying

these features into some recognizable command C through the use of signal processing and machine learning techniques. While BCI technology is still very immature, current efforts have demonstrated that the electrical signals generated by our brains can be recorded and interpreted by man-made sensors.

In this paper we propose an authentication method for access to computing devices, whereby a user *thinks* a password. We call this method a *pass-thought*, the general concept of which is illustrated in Figure 1.

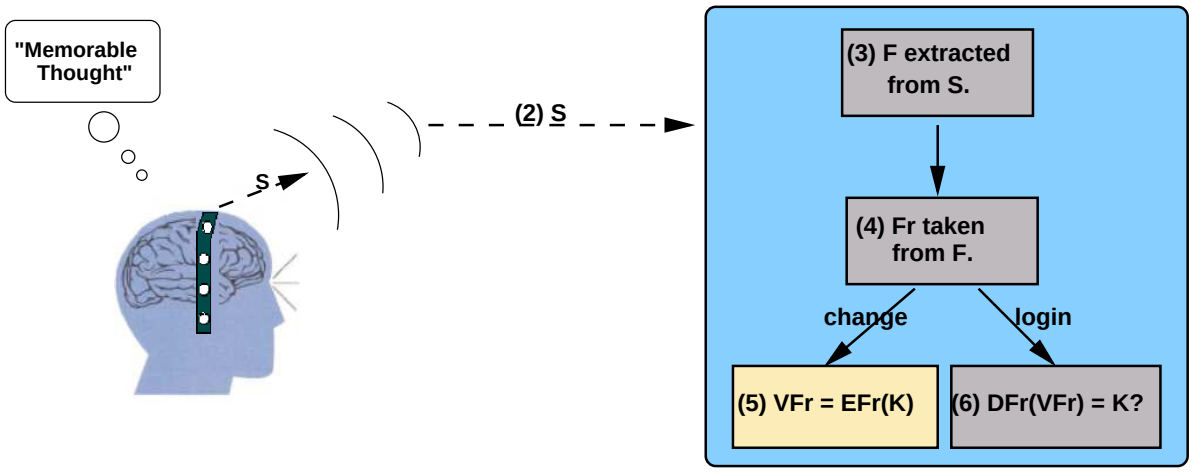

Figure 1: General concept of a pass-thought. 1) The user presses a key when ready, thinks of their previously chosen pass-thought, and presses the key when done. 2) Electrodes record the signal S emitted during the time between the start and stop key presses. 3) S is processed into signal features F. 4) The subset of features $F_r \in F$ are those that best capture the user's thought and are consistent over time (determined by trials). 5) F_r is used to encrypt a key or checkword in such a way that permits some small level of error (see §3.4); this value is called V_{F_r}. 6) V_{F_r} is used for user authentication to a computing device; login success is indicated by the user's pass-thought entry ${F_r}^{\ell}$ successfully unlocking V_{F_r}, which occurs provided that F_r is approximately the same as ${F_r}^{\ell}$.

Steps 2-5 from the general pass-thought description in Figure 1 could be performed using a BCI. For an example of a pass-thought scheme that elaborates on each of these steps, see §3.2.

There is a significant difference between what the BCI research to date can offer, and the BCI requirements of pass-thoughts. BCI research has been focussed on enabling a user to control something external (e.g. movement of a cursor) using their thoughts. For a user to provide control commands using their thoughts, their thoughts must undergo translation (interpretation). Pass-thought input should undergo feature extraction to filter out the non-repeatable parts, but there is no need to translate the brain signals. Such translation is not only unnecessary but undesirable as it might reduce the entropy the user's brain signals provide. Given that the translation of signals is one of biggest challenges for BCI research, our proposed pass-thought application might be better suited to BCI technology than the current communication and control applications under research.

The sequel is organized as follows. §2 discusses related work and elaborates on the problems of other authentication methods. §3 presents what we believe to be the potential of pass-thoughts, the current state of BCI technology, an example pass-thought framework that is currently feasible under this current state, a security analysis, and a discussion of the future of pass-thoughts. §4 discusses the requirements for future work, in particular to allow higher-entropy pass-thoughts. Concluding remarks are made in §5.

2 Motivation and Related Work

A wide variety of research to date has explored authentication methods and their associated problems. Textual passwords enjoy popularity due to low cost, user familiarity, and the lack of persuasively better

alternatives. Nonetheless, as explained below textual passwords and existing alternatives all have well-known limitations.

One of the most important problem with textual passwords is their vulnerability to *dictionary attacks*. A dictionary attack is a brute-force guessing attack where an attacker draws candidate guesses from a dictionary of "likely passwords" [28, 47]. If a password is correctly guessed or otherwise obtained, the attacker can easily masquerade as a legitimate user with minimal threat of being caught. To reduce the effectiveness of dictionary attacks, many techniques for improving password quality have been developed, including proactive password checking [56], automatically generated passwords, password construction rules, and passphrases [57]. However, for various reasons these measures do not guarantee uniformly strong passwords in practice; thus, researchers and developers have sought ways to prevent dictionary attacks from being successfully executed. Online dictionary attacks may be thwarted by introducing a small delay after a failed login attempt, by restricting the number of failed login attempts, and by using human-in-the-loop verification methods [40] (see also [48]). Schemes such as Encrypted Key Exchange (EKE) [3] and Strong Password Exponentiated Key Exchange (SPEKE) [21] can help prevent off-line analysis of captured network traffic. However, operating system vulnerabilities [9] and related access control failures can lead to the inadvertent disclosure of entire password databases.

A password database can contain passwords in plaintext, one-way hashed passwords, or encrypted checkwords using the password as a key. Once stolen, the only defense against off-line dictionary attacks is the computational infeasibility of a correct guess (which may involve transformations to match the password file entry format). As processor speeds increase, and with advances in distributed computing, the feasibility of such attacks will only increase.

Graphical passwords have been proposed as a plausible alternative to textual passwords, motivated in part by findings that people have a remarkable memory for pictures. A graphical password can be described as a password that requires the user to remember a picture (or something related to a set of visual representations) in lieu of a word. A wide variety of graphical password schemes have emerged in the last six years. Recognition-based graphical password schemes include Déjà Vu [13, 38], Passfaces [42], and Story [12]. Recall-based graphical password schemes include Draw-A-Secret [24] and others that require a user to click on parts of a presented image [6, 46, 23]. Unfortunately, graphical password schemes are vulnerable to shoulder-surfing [45] (more so than textual passwords), and possibly also to guessing attacks [12, 51]. Further, the size of the effective password space of graphical password schemes may also be less than originally expected [52].

Now more than ever, shoulder-surfing is a problem for passwords. The ubiquity of cell phone cameras and wireless video cameras brings a new ever-present threat upon us: recorded shoulder-surfing. As users, we can no longer simply look over our shoulder to be aware of an adversary observing our password. To combat such a threat, we require an authentication method that is unobservable under any circumstance.

Shoulder surfing can be addressed in the context of "what you know" authentication methods, for example through the use of *human interactive proofs* [7] or *cognitive trapdoor games* [45]. These methods require cognitive processing by the user for each bit of information, and thus unfortunately would require a large amount of a user's input time to increase the size of the password space. Although they are unlikely to become a ubiquitous authentication scheme for that reason, they may be useful in high-security settings such as banking machines.

A user may also pose a threat to a "what you know" authentication system by writing down their password, or sharing it with others. Both textual and some types of graphical passwords are susceptible to this threat. Users often share passwords to bypass the administrative overhead of setting up the proper group access control. Sharing passwords may be more convenient for the user; however, it defeats the purpose of user authentication. Writing down a password makes it available to anyone with access to the medium (e.g. if paper, co-workers, cleaners, and dumpster-divers all have access). When a user can describe their password to others, the user is also susceptible to social engineering methods whereby the user is tricked by an adversary into providing their password [19]. Thus, we currently have a paradox: we want a scheme whereby we *can* write a password down as a reminder, yet we want to ensure the password cannot be used by others.

The susceptibility of "what you know" authentication methods to guessing attack is largely created by patterns in user choice: their entropy is limited by human memory. Authentication schemes that are based on "what you have" (physical tokens, e.g. smart cards [1]) have the advantage of high entropy, and if they

are lost or stolen, they can be changed and replaced. However, they are not necessarily always in the owner's possession. During the period before a user's lost token(s) is revoked, their system and the information on it may be vulnerable [26]. An idea to solve this problem is to have users wear their physical token, as in *zero-interaction authentication* (ZIA) [10]. ZIA is an authentication scheme that provides security against physical attacks by continuously polling the token to ensure it (and thus presumably the user) is present. ZIA appears to be a useful scheme to protect against physical attacks, particularly for mobile devices. However, it suffers from the same scalability problem as smart cards and other physical token methods: the tokens become burdensome if required for many different domains, resulting in a stack of such tokens for the user to bear.

Biometrics attempt to solve the problem of "what you know" and "what you have" authentication methods by the use of an appealing concept: authentication by using the unique physical or behavioral characteristics of users, e.g. fingerprints [43], the iris [11], voice recognition [35], and keystroke dynamics [36]. Biometrics suffer from a major drawback: they cannot be (easily) changed. Because biometric information is valid for the lifetime of the user and risks being stolen, such information cannot be used as keying material for remote authentication purposes. Furthermore, even when performing local identification, certain types of biometric readers cannot detect fraudulent inputs, e.g. for fingerprints, it has been shown that a gelatin finger that models a legitimate user's fingerprint (e.g. lifted from a glass) can fool many commercial fingerprint readers [32].

The following set of authentication method requirements emerge from the unresolved problems outlined in this section; several known methods meet a subset of these, but none meet all of these desiderata.

1. *Changeability.* If the user's authentication information is compromised, we must be able to replace this information (and revoke any old password or access credential).

2. *Shoulder-surfing resistance.* The scheme must not be vulnerable to shoulder-surfing, particularly in the presence of ubiquitous visual recording devices.

3. *Theft protection.* This includes physical theft and the computational infeasibility of guessing attacks. If we must rely on the entropy of an authentication scheme for protection against off-line dictionary attack, we require an authentication method whose entropy can scale with processor speeds, Moore's Law, and increasing distributed collaboration.

4. *Protection from user non-compliance.* To discourage unintended transfer to other parties, the user should not be able to write down (in a manner useful to an attacker) or share their authentication information "too easily".

Pass-thoughts have the potential to satisfy all of these requirements.

3 Pass-thoughts

Our authentication problem is one of extracting high-entropy information from a user to prove that they are who they claim. This essentially means extracting something that makes a person unique. It is interesting to consider how people recognize one another: aside from appearances, we recognize a person's movements, actions, and expressions, all of which are initiated by thought in the brain.

There are a number of reasons to believe that there is uniqueness (given genetic and environmental differences) within our brains: certain areas of our brains are developed more depending on our training and experience. For example, string musicians are known to have larger somatosensory cortical areas associated to the fingers than the average person [18]. Also, the alpha frequency (a signal feature in an electroencephalographic (EEG) signal) has been found to have considerable variability between subjects [16]. These results may imply that the signals emitted from our brains are different upon thinking "the same thing". Thus, it is plausible that if two people think of what they could best describe as the same thing, the brain signals emitted would be distinguishable. Similarly, we also expect two different thoughts by the same person to result in distinguishable signals.

If we assume that a user's pass-thoughts could be recorded with enough accuracy to distinguish between different thoughts and distinguish the differences between different user's "same" thoughts, pass-thoughts

may be a natural two-factor (what we know and what we are), changeable authentication scheme. A pass-thought is changeable (the thought itself; the "what you know" portion), and the physiological uniqueness of a user's brain that emits the pass-thought would act as a second, biometric factor.

The theoretical entropy of pass-thoughts is enormous. A pass-thought could belong to a language (as in textual passwords), an image (as in graphical passwords), a type of (imagined) movement, an abstract thought, an emotion, a memory, etc. An entire sentence, picture, or memory (or sequence thereof) can be represented by a simple thought. Even pieces of music can be represented by a thought. There is also a significant amount of variation within the same type of thought. For example, a user could think of their first dog in countless ways through combinations of a variety of factors including the dog's name, breed, smell, bark, colour, visualizing the dog doing activities such as running through the park, sleeping, eating, licking one's face, etc. (not to mention the places, movements, and emotions associated with each of these actions). It is impossible for us to know the size of the pass-thought space; however, we have some hope based on the sheer number of neurons that exist in a typical adult brain (approximately 100 billion [50]). Even if we assume that each neuron could only store one bit of information, pass-thoughts could produce keys of 2^{36} bits in size. The actual theoretical entropy is much greater, though, since individual neurons are so complex that we do not yet have a complete model of individual neuronal behavior.

While pass-thoughts have the potential to be an authentication scheme with an extremely large password (pass-thought) space, in practice there would likely be boundaries on the size of the pass-thought space that correspond to the encoding scheme. There may also be a more probable subset of pass-thoughts (see §4) as with other "what you know" authentication schemes that correlate with the strengths of human memory. While these factors may limit the theoretical strength of pass-thoughts, we do not believe they would compromise their security in practice.

Despite this promising potential, we recognize that BCI technology is still in its infancy (see §3.1), and thus the accuracy of signal recording and processing is unknown. However, research interest in this topic is increasing [20] (and likely will continue to do so given the myriad of applications it could enable). Towards the goal of a currently plausible pass-thought-based authentication method, we review the status quo of BCI technology (§3.1), how a pass-thought-based system could make use of existing BCI technology (§3.2), and some future possibilities for pass-thoughts (§3.4). We also provide a security analysis of pass-thoughts (§3.3) and address some ethical considerations (§3.5).

3.1 Current Status of BCI Technology

Although the first research relating to BCI's appeared in the 1960's [8], it is still in its infancy for a variety of historic reasons. First, the chance of extracting a user's intended message (i.e. a yes/no answer to a question) from brain signals appeared to be extremely remote. Second, it is only in recent years that the cost of computers with sufficient processing power to analyze electroencephalography (EEG) signals in real-time has become affordable. Third, there was not much resulting interest in the limited applications that a first generation BCI was likely to offer [55]. Recent changes in technology and advances in research have changed the environment for BCI research. Recent advancements include enabling owl monkeys to control a robotic arm and a computer cursor through their thoughts alone in 2000 [37], and enabling a paralyzed 25-year old man to do the same in 2004 [17].

The basic model of a BCI system is shown in Figure 2. This diagram will be the basis of the rest of this section, describing the brain signals, signal acquisition, signal feature extraction, signal feature translation, and some possible outputs.

BCIs are capable of monitoring various brainwave phenomena. Examples of brainwave phenomena include slow cortical potentials (SCPs), P300 potentials (positive peaks after 300ms), and mu or beta rhythms recorded from the scalp. Methods to observe these phenomena include EEG, magnetoencephalography (MEG), positron emission topography (PET), functional magnetic resonance imaging (fMRI), and optical imaging [55]. Other signal features that are considered to contaminate the user intent are electromyography (EMG), and electroocuology raphy (EOG), which result from muscle and eye movement respectively [55]. There are also invasive BCI recording methods (e.g. implanted electrodes [37, 54]).

Only the EEG and related methods (EMG and EOG) can use portable hardware and require relatively simple and inexpensive equipment (i.e. on the order of hundreds of dollars) [41]. Most BCIs use EEG signals, which represent the electrical activity in the brain as measured from outside of the skull. EEGs are normally

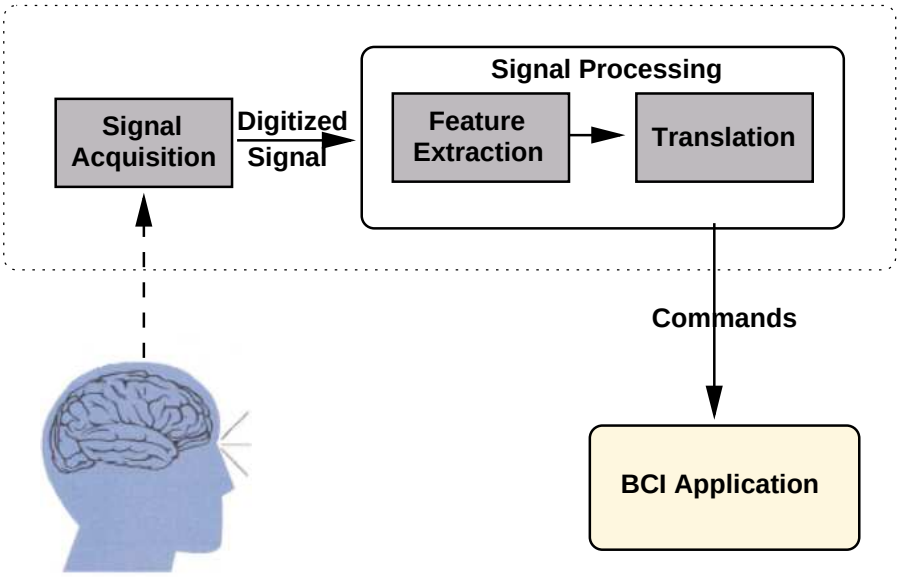

Figure 2: Basic design of a BCI system [55].

acquired by using a set of electrodes that must make contact with the skin on the scalp. Each electrode then directly measures the signals read at its location. These signals are amplified and digitized.

The EEG is the summation of the synchronous signal activity of thousands of cortical neurons, thus the signals read at a given location may not originate from the area under the electrode. To obtain a more accurate reading from a particular location, often a type of spatial filtering is used during the signal feature extraction phase (e.g. large Laplacian, small Laplacian, ear reference, and common average reference [55]). The general idea behind each of these spatial filters is to focus on activity with a particular spatial distribution.

The goal of signal feature extraction is to help with the signal-to-noise ratio, identifying signal features such as rhythm amplitudes and neuronal firing rates. Hopefully, feature extraction will help distinguish the parts of the signal that encode the user's thought. A variety of EEG signal features can be extracted; they can be characterized as time-domain or frequency-domain (or both). This information should to help the BCI system to distinguish the parts of the signal that encode the user's intent. In addition to spatial filtering, a variety of options for feature extraction are currently under study by others, including spatial and temporal filtering techniques, signal averaging, voltage amplitude measurements, and spectral analyses [55] (e.g. by using the Fast Fourier Transform).

In comparing feature extraction methods, a statistical measure called r^2 can be useful. This measure is the proportion of the total variance in the signal feature that is accounted for by the user's intent. Temporal filtering can measure certain parts of a signal by using a band-pass filter [39], or by amplitude in specific spectral bands of Fourier or autoregressive analysis [31]. A set of band-pass filters can be used to remove signals such as the EMG (this is considered a contaminant in most BCI research as it may indicate the user is using their facial muscles to effectively control the BCI). Due to the real-time requirements of BCIs on signals that can change rapidly, frequency analysis techniques that only require short time segments can achieve better performance (e.g. band-pass filters and autoregressive analysis). Other signals, such as the P300, can be enhanced by averaging where the loss in communication rate can be minimized by overlapping the trials [15].

Signal translation algorithms take signal features as input, and output some dependent variable such as a device control command. Translation algorithms used include linear equations, discriminant analysis, and neural networks. If only a single signal feature is used, the output of a translation algorithm can be a simple linear function of the feature value (e.g. a linear function of the amplitude of a mu-rhythm). More complex translation algorithms include supervised learning approaches such as linear discriminate analysis

(e.g. [22]), and non-linear discriminate analysis (e.g. adaptive logic networks [29]). Due to its simplicity, the most popular architecture for artificial neural networks is the multilayer perceptron [2].

BCI prototypes have been developed over the past few years that have enabled severely disabled people to communicate (e.g. the Thought Translation Device [5] and Spelling Device [4]); however, the bit rates for these BCIs are quite low (the maximum is approximately 25 bits/minute). While this is discouraging, this field has only recently begun making large advances, which we believe are likely to pave the way for significantly better technology in the future. It is unclear how far this technology can and will advance, but it appears that the areas that will determine this include the selection of signal acquisition methods, signal feature extraction methods, and translation algorithms [55].

3.2 Feasible Pass-thoughts Based System

In this section, we discuss a feasible (in the sense it could be built) pass-thoughts system given current BCI technology, guided by the 6 steps in Figure 1. We assume that the same capability from the BCI used as that shown in recent BCI prototypes such as the Thought Translation Device [5]. We also assume that the user is logging onto a desktop PC, where the pass-thought will be used exactly as a password.

To begin, the user presses a special key sequence when ready (e.g. CTRL-ALT-DEL). Given the current status of BCI technology for communication, we can only assume a poor signal to noise ratio, and thus that we can only extract one or two bits of information from each thought. For a currently possible pass-thought system, we propose a scheme similar to that which uses evoked P300 potentials for a spelling device for the disabled [4]. A P300 potential is a positive potential that is evoked about 300ms after a surprising or exciting event. By randomly highlighting the components (either textual or graphical) on the desktop's monitor, when the user sees the part of their "pass-thought" highlighted (e.g. see Figure 3), they presumably generate a P300 spike as for the spelling device [4]. The results of the P300 potential spikes are silently recorded and determine whether the user's P300 firing matched the expected template that represents the account's password. This type of scheme could be used in conjunction with either textual or graphical passwords, where a sequence of letters, pictures, or points on a picture are highlighted at random times (thus randomly generating the user's P300 potential spikes, where the verification side of the system knows when to expect which bits).

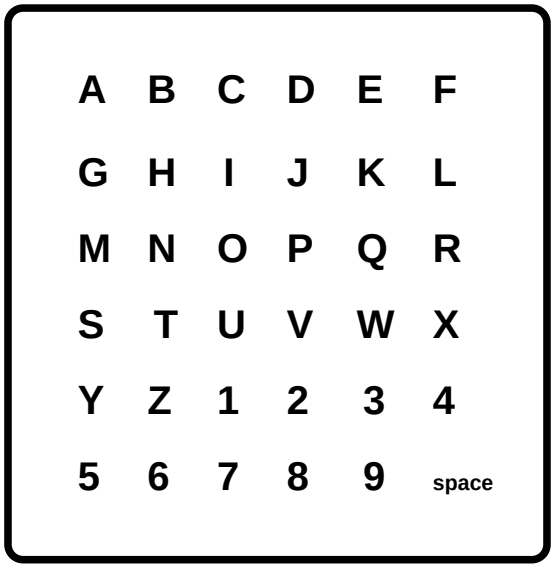

Figure 3: Illustration of a screen intended to randomly highlight letters (from the spelling device in [4]). The technology exists such that this approach could be used for a pass-thought system.

The size of the pass-thought space for this scheme depends on the number of components on the screen and the number of screens presented to the user. If we assume a textual password scheme where all 95

printable ASCII characters are displayed on each screen, and the user must select a sequence of 8 characters, the size of the full pass-thought space is 95^8, approximately 52 bits. Of course, we would not expect 52 bits of security for the same reasons that we do not for textual passwords. This pass-thought based system is shoulder-surfing resistant and otherwise provides the same amount of security as textual passwords. One disadvantage is that given the current state of BCI technology (P300-based approaches have shown a bit rate of 4.8 characters/minute to obtain 90% accuracy on a 36-character grid [15]), this process would take 1 minute and 40 seconds if the performance did not degrade with a larger character grid.

Electrodes record the P300 spikes generated by the user. The number and placement of electrodes required to currently obtain an acceptable signal affects the size of the hardware interface. An acceptable number of electrodes has experimentally been found to be as low as 8 [34]. The placement of electrodes can affect the signal-to-noise ratio, depending on the signal features the BCI wishes to focus upon (e.g. SCPs or mu-rhythms) [55]. A hardware interface that has the look and feel of headphones may feel less awkward than the electrode caps used in many BCIs (e.g. picture in [33]). A wireless headphone-style electrode interface might have electrodes lining the band going over the head, around the ears, and/or on a small disc that is attached to the hardware that extends towards the back of the skull (to place electrodes at the top of the skull if necessary). For example, recall Figure 1, the user is wearing such an interface that wirelessly transmits S. Due to a possible lack of user tolerance for such devices, pass-thoughts may be most appropriate for use in specialized environments.

In this scheme only one signal feature is used, since without future studies to determine the variation and repeatability of brain signals, we only have the BCI literature to date as a guide of what is possible. The results for BCI communication so far have low bit-rates, thus we can only assume a yes/no answer. Thus, in the context of this scheme F (recall §1) is a set of P300 potentials, and $F = F_r$.

Depending on the rate of highlighted screen components, the user may miss a few P300 potentials; however, the algorithm should only record those P300 spikes that occur, not the ones that were missed. This allows the system to be exactly repeatable, and we could use a one-way hash function h to store the pass-thought instead of a fuzzy vault (for a discussion on fuzzy vaults, see §3.4).

Authentication using this basic pass-thought system can then be performed exactly as with textual passwords. Input completion occurs either after a certain number of P300 potential spikes have been received, or by the user pressing the key sequence again. The hashed pass-thought $h(F_r)$ is compared upon input completion to the stored pass-thought file hash for the user, and login success occurs if they match.

In addition to the system described above, there are other pass-thought systems that could be implemented using today's technology. For example, a pass-thought system could be built that is based on touch. In this type of system, a user might place their palms down on a surface that raises pins that cannot be observed, but can be felt by the user. The user would have a subset of raised pin patterns that they should answer "yes" to upon feeling; other raised pin patterns should evoke a "no" answer from the user. The "yes" answers could be obtained using the user's P300 potential spikes. Lack of a P300 spike could be interpreted as a "no" answer.

3.3 Security Analysis

In this section, we discuss how pass-thought-based systems would respond to different theft and guessing attacks. As first noted in §1, a pass-thought system is unobservable and thus resilient to shoulder-surfing attacks. Interception attacks (i.e. when using a pass-thought for remote authentication) can be avoided using the same protection mechanisms as for traditional passwords (e.g. encrypted channels and challenge-response protocols).

Social engineering attacks would succeed for our currently feasible scheme discussed in §3.2 since users could communicate these just as in traditional passwords. Similarly, our currently feasible scheme as described in §3.2 (but not the general approach) is as susceptible to dictionary attack as traditional passwords. However, if signal recording and processing methods advance such that they are able to capture thoughts in detail such that the thought itself could be used, a pass-thought would be quite difficult to communicate to a social engineer. Using such a scheme, even if a particular pass-thought is successfully communicated, a social engineer's brain signal may be different than the user upon thinking "the same thing" (recall §3). For these reasons, we (perhaps optimistically) hope that the size of the pass-thought space might be sufficiently large to protect against most dictionary attacks.

A pass-thought based system cannot avoid the threats of phishing attacks [49] and of recording the pass-thought through either a hardware or software tap (i.e. when the input device has been compromised). It is unclear whether any authentication system could be entirely immune to phishing attacks. If pass-thoughts could be constructed such that each transmitted pass-thought was used only once, though, we could potentially defeat passive interception attacks. For example, it might be possible to build a kind of challenge-response system by taking advantage of the complex computations that the human brain performs automatically (and subconsciously), or by having a remote server model a user's brain processes. Given the current state of brain research, however, such systems are even further in the future than pass-thoughts.

3.4 Future of Pass-thoughts

The ultimate goal of a pass-thought system is to extract (in as short of a time as possible) as much repeatable entropy as possible from a user's brain signals upon "entering" a thought. Until more information is gathered regarding which signal features are best, a signal S recorded from a BCI might be processed into as many features F (e.g. parts of the signal that belong to a certain frequency and/or amplitude range) as possible. From a series of repeated trials of entering the pass-thought, the largest matching number of features F_r will be considered repeatable. A pass-thought might be represented as F_r, the repeatable subset of brain signal features. This subset could be stored on the system in a fuzzy vault [25] (discussed further below), where the number of matching features from this subset is some predefined threshold value, lower than $|F_r|$.

We acknowledge that it would be difficult to equip the vast number of existing computer systems with an electrode-based hardware interface to record brain signals. A possible solution could be to assign such a hardware interface to every user, which would act as an authentication token on behalf of the user for access to a device D. This token could record a user's pass-thought T, then use a wireless interface and Station-to-Station Protocol [14] as done in the *zero-interaction authentication* scheme [10] for authenticating to D. In this case, the pass-thought could be used in place of a PIN number to enable the device. A plausible candidate token might be a cell-phone or PDA (which could also use T for access). Using such a token would reduce the risk of the user's pass-thought being recorded by a hardware (or malicious software) tap. The tapping problem is not solved by this solution; it simply moves this type of attack from being a per-system attack to a per-user, physical possession attack, which is presumably more difficult.

A pass-thought system ideally would obtain as much information as possible from the BCI. Given the focus of BCI research to date (communication and control), it does not appear that the documented approaches to processing brain signal features would be applicable to the approaches required for pass-thoughts. Thus, at this point it is difficult to make reasonable assumptions about issues such as which signals features are best, which signal features are "pollutants", and what electrode placements are best for spatial filters.

We do not know if the signals will be exactly repeatable as in the system discussed in §3.2. Thus, we must assume that some level of error tolerance will be required. One feasible method for achieving this is for F_r to encrypt a key or checkword using a fuzzy vault: this value is called V_{F_r}.

Fuzzy vaults are cryptographic constructions described by Juels and Sudan [25]. They can be thought of as a form of error-tolerant encryption, where the keys consist of sets. To apply fuzzy vaults to pass-thoughts, we will use the set of repeatable signal features F_r as the key for "locking" V_{F_r}, and a pass-thought login attempt ${F_r}^{\ell}$ as the key to try "unlocking" V_{F_r}. Juels and Sudan [25] prove that fuzzy vaults are provably secure against a computationally unbounded attacker.

We discuss fuzzy vaults in the context of pass-thoughts (recall §1). A secret value k (e.g. a cryptographic key) is "locked" using the set of repeatable signal features F_r; in essence this would be the template for the pass-thought. We only want F_r to be unlocked if the user's pass-thought input from a login attempt, ${F_r}^{\ell}$, overlaps with F_r substantially. The following procedure describes the how "lock" and "unlock" for k works:

1. Choose a polynomial p in a single variable x such that p encodes k (e.g. by embedding k in its coefficients).

2. Compute evaluations of p on elements of F_r (we can think of this as "projecting" elements of F_r onto points in p).

3. Create a number of random *chaff points* that are not in p. Chaff points are points added to provide random noise, and the more chaff points there are, the greater the noise to conceal p from an attacker.

Let V_{F_r} be the chaff points and p together. V_{F_r} is called a *commitment* of p (which is really the secret k). F_r identifies the points in V_{F_r} that lie on p.

4. Extracting k using ${F_r}^{\ell}$ requires that it overlaps substantially with F_r. Since F_r identifies the points in V_{F_r} that lie on p, a number of points corresponding to the amount of overlap can be recovered. If there is only a small amount of noise, k can be recovered using error correction. Otherwise, it is infeasible due to the many chaff points that will serve as noise.

V_{F_r} is used for user authentication to a computing device; login success is indicated by the user's pass-thought entry ${F_r}^{\ell}$ successfully unlocking the fuzzy vault V_{F_r}, which occurs provided that F_r is approximately the same as ${F_r}^{\ell}$.

Another method to achieve error tolerance (originally proposed for keystroke dynamics) is described by Monrose et al. [36]. Using our terminology, this method takes the hamming distance of the two sets of features F_r and ${F_r}^{\ell}$. ${F_r}^{\ell}$ is sufficiently similar if the hamming distance is less than a pre-specified error tolerance threshold.

To address the challenge that F_r may change over time due to memory drift and other changes in mental state, an occasional change of the features in F_r might be required. This should only be performed if the features in the password that unlocked the fuzzy vault are similar enough to the F_r in the fuzzy vault. Alternatively, a signal translation algorithm could be applied to improve the repeatability of each feature. Until the features F_r are known (and their properties), we cannot reasonably suggest a translation algorithm (although some are discussed by Wolpaw et al. [55] for communication and control purposes).

3.5 Ethical Issues

We expect ethical and moral contention to arise with the advancement of BCI-based technologies. We note the existence of new types of authentication that differ quite substantially from pass-thoughts (e.g. invasive technologies such as embedding chips [27]). Ensuring privacy in the face of BCI advancements will likely be a challenging issue with the advent of new applications.

An example of where BCI technology could be abused includes tricking a user into inputting a pass-thought, where the device interprets the meaning of the user's thought instead of using it for authentication as intended. Although BCI technology for such meaningful interpretation does not exist (and may never exist if users' brain signals differ upon inputting the same thoughts), we do not know whether it would ever be possible. If it were, many potential problems would arise, e.g. an interpretation of a user's thought being used against them (e.g. for bribery or in a court of law). Another useful application with the potential for abuse is a BCI-based lie detector [44]. We note the importance of open debate for such issues, and strongly encourage such debate.

4 Discussion

There is a clear need for shoulder-surfing proof user authentication, especially given the ubiquity of cell phone cameras and wireless video cameras. One of the primary benefits of pass-thoughts (see §3) is that they are visually unobservable and thus are resistant to shoulder-surfing. Another idea for unobservable authentication method might be to make use of eye-gaze tracking (using e.g. the LC Technologies Eyegaze Systems [30]). Such an eye-gaze based method could permit unobservable passwords of the same strength provided by textual or graphical password schemes by allowing the user to select parts of the password with their eyes (e.g. by eye fixation for a specified period denoting selection), and not echoing the input on the screen. While the security of such a system is comparable to the pass-thought systems that can be built with today's technology, advances in BCI technology could lead to much more powerful pass-thought authentication systems (recall §3.4).

The ultimate feasibility of pass-thoughts is dependent upon the accuracy of a BCI in recording the repeatable parts of a brain signal. For a brain signal to be repeatable, the signal features that represent the intended pass-thought must be extracted. Given the BCI research to date, we know that at least a binary response can be evoked, and would thus be at least as effective (if not considering input time) as an authentication method that makes use of an eye-tracking device (e.g. Eyegaze Systems). If BCI technology

advances towards enabling accurate and repeatable brain signal input (which we believe is likely given the recent advances), we might have a new authentication method that solves many of the problems associated with current systems. The flexible nature of pass-thoughts could allow for strong authentication to scale with increasing processor speeds. Increasing the complexity of a pass-thought might be as simple as recording a longer thought. A pass-thought system would be unobservable, defeating the threat of recorded (and unrecorded) shoulder-surfing. It would be difficult for users to share or write down exact thoughts, as the pass-thought is bounded by communication mediums such as language, drawings, and demonstrated movement. Users could still write down a note for themselves to remind them of their pass-thought, which would presumably be of little value to an attacker due to the user's signal feature variations.

The many areas for future work arising from our proposal include: understanding brain phenomenon, the acquisition of brain signals, extraction of features, and algorithms to aid in repeatability of a "transmitted" thought. Since the goals of a pass-thought application are different than previously researched BCI applications, many of these topics have yet to be explored in this context.

We believe it is likely that certain mental processes will interfere with a pass-thought such that parts of the signal may need to be filtered out. The current BCI literature has not answered the question of what parts of a signal represent parts of what is going on in our mind, at least at the granularity required for pass-thoughts. We must be careful in processing and extracting parts of the signal, since this decreases the amount of information provided by the pass-thought. This is arguably the most important area for future BCI work that would increase the success of a pass-thought-based system.

Perhaps the problem of choosing repeatable signal features will only be solved by user training and using an authentication method that does not require exact repeatability (e.g. using a fuzzy vault – see §4). It would be interesting to determine how difficult it is for most people to control their brain state enough to reduce noise upon entering a pass-thought. A low training time is very important for user acceptance, and is yet another important area for future work. Perhaps during training, a method for providing the user some feedback about their brain signals would be useful, e.g. a real-time changing image that represents the signal features, where each feature is shown in different colours.

As with any "what you know" authentication method, user studies would be required to determine whether patterns in user choice occur. If given free selection, users may choose pass-thoughts that are easily categorized. Such high-level categories might include abstract, verbal, or symbolic thought. At a lower-level granularity, we might find that many users e.g. think of the chorus to a popular radio song. User choice can be positively or negatively influenced by providing recommendations or wording instructions differently. An area for future work might be to determine how much variation in user choice can be achieved under instruction or no instruction, and the associated memorability.

If users did think similar things, we have reason to believe (recall §3) that the signal encoding a transmitted thought would be dissimilar from one user to another. No particular experiments to verify this appear to exist; thus it would be interesting to confirm and quantify the differences between the brain signals generated by people who are thinking "the same thing".

5 Concluding Remarks

We present (what to our knowledge is) a novel idea for user authentication called pass-thoughts, whereby a user authenticates to a device by "transmitting" a thought. This transmission would occur through a Brain Computer Interface (BCI), tailored specifically for this purpose. BCI technology to date has been focused on interpreting brain signals for communication and control for the disabled [55]. The BCI requirements of a pass-thought system are entirely different: they require no interpretation of the brain signals, but the use of as much signal information as possible. Pass-thought's goal of extracting as much information as possible from a signal has the opposite goal from the filtering and many-to-one signal translation that must occur for interpretation of brain signals.

The advantages of pass-thoughts over many of the existing authentication technologies include changeability, shoulder-surfing resistance, and protection against theft and user non-compliance (recall §2). Disadvantages of pass-thought authentication include the requirement for a new hardware component (including electrodes) to record the user's brain signals. For this reason, a pass-thought system may not be accepted for widespread use, but perhaps for high-value or high-importance applications or environments (e.g. within

banks and governments).

There are many unknowns to resolve before pass-thoughts might become the method we envision. It is our hope that this idea for a pass-thought system will inspire research into the area of signal processing and translation algorithms that retain as much repeatable information as possible. If the recording and processing of brain signals can be accurate and repeatable, pass-thoughts might become a viable and useful new form of authentication.

References

[1] Martin Abadi, Michael Burrows, C. Kaufman, and Butler W. Lampson. Authentication and Delegation with Smart-cards. In *Theoretical Aspects of Computer Software*, pages 326–345, 1991.

[2] M. A. Arbib, editor. *The Handbook of Brain Theory and Neural Networks*, pages 178–181. The MIT Press, second edition, 2003.

[3] S. Bellovin and M. Merritt. Encrypted Key Exchange: Password-Based Protocols Secure Against Dictionary Attacks. In *IEEE Computer Society Symposium on Research in Security and Privacy*, pages 72–84, 1992.

[4] N. Bierbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub, and H. Flor. A Spelling Device for the Paralyzed. *Nature*, 398:297–298, 1999.

[5] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, and B. Kotchoubey. The Thought Translation Device (TTD) for Completely Paralyzed Patients, 2000.

[6] J.-C. Birget, D. Hong, and N. Memon. Robust Discretization, With an Application to Graphical Passwords. Cryptology ePrint Archive, Report 2003/168, 2003. `http://eprint.iacr.org/`, site accessed Jan. 12, 2004.

[7] M. Blum and N. J. Hopper. A Secure Human-Computer Authentication Scheme, 2000. `http://www.aladdin.cs.cmu.edu/papers/pdfs/y2001/manuel_blum.pdf`, accessed Mar. 16, 2005.

[8] V. Brower. When Mind Meets Machine. *EBMO Reports*, 6(2):108–110, 2005.

[9] CERT Coordination Center. Vulnerabilities, Incidents, and Fixes. `www.cert.org`.

[10] M. D. Corner and B. D. Noble. Zero-Interaction Authentication. In *International Conference on Mobile Computing and Networking*, pages 1–11, 2002.

[11] J. Daugman. How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[12] D. Davis, F. Monrose, and M.K. Reiter. On User Choice in Graphical Password Schemes. In *13th USENIX Security Symposium*, 2004.

[13] R. Dhamija and A. Perrig. Déjà Vu: A User Study Using Images for Authentication. In *9th USENIX Security Symposium*, 2000.

[14] W. Diffie, P.C. van Oorschot, and M. Weiner. *Authentication and Authenticated Key Exchanges*, volume 2 of *Design Codes and Cryptography*, pages 107–125. Kluwer Academic Publishers, 1992.

[15] E. Donchin, K. M. Spencer, and R. Wijesinghe. The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface. *IEEE Transactions on Rehabilitation Engineering*, 8:174–179, 2000.

[16] M. Doppelmayr, W. Klimesch, T. Pachinger, and B. Ripper. Individual Differences in Brain Dynamics: Important Implications for the Calculation of Event-Related Brain Power, 1998.

[17] D. E. Duncan. Implanting Hope. *Technology Review: MIT's Magazine of Innovation*, 108(3):48–54, 2005.

[18] T. Elbert, C. Pantev, C. Wienbruch, B. Rockstroh, and E. Taub. Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science*, 270:305–307, 1995.

[19] S. Granger. Social Engineering Fundamentals, Part I: Hacker Tactics, 2001. `http://www.securityfocus.com/infocus/1527`, site accessed Mar. 22, 2005.

[20] ISI Web of Knowledge. Analysis: Brain Computer Interface Search Results, 2005.

[21] D. P. Jablon. Strong Password-Only Authenticated Key Exchange. *ACM SIGCOMM Computer Communication Review*, 26(6):5–26, 1996.

[22] A. K. Jain, P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:4–37, 2000.

[23] Wayne Jansen, Serban Gavrila, Vlad Korolev, Rick Ayers, and Ryan Swanstrom. Picture Password: A Visual Login Technique for Mobile Devices. National Institute of Standards and Technology Interagency Report (NISTIR) 7030, 2003. `http://csrc.nist.gov/publications/nistir/nistir-7030.pdf`, site accessed Mar. 22, 2004.

[24] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin. The Design and Analysis of Graphical Passwords. *8th USENIX Security Symposium*, 1999.

[25] A. Juels and M. Sudan. A Fuzzy Vault Scheme. In *IEEE International Symposium on Information Theory*, 2002.

[26] M. Just and P.C. van Oorschot. Addressing the problem of undetected signature key compromise. In *NDSS*, 1999.

[27] I. Kerr. So Trendy, So Convienient – So Dangerous to our Privacy, July 31, 2004. Vancouver Sun, available at: `http://anonequity.org/en3/July31-Van_Sun-Baja_Beach_Club.pdf`.

[28] D. Klein. Foiling the Cracker: A Survey of, and Improvements to, Password Security. In *The 2nd USENIX Security Workshop*, pages 5–14, 1990.

[29] A. Kostov and M. Polak. Parallel Man-Machine Training in Development of EEG-Based Cursor Control. *IEEE Transactions on Rehabilitation Engineering*, 8:203–204, 2000.

[30] LC Technologies Inc. Eyegaze Systems. `www.eyegaze.com`, site accessed Mar. 22, 2005.

[31] S. L. Marple. *Digital Spectral Analysis with Applications*. Prentice-Hall, 1987.

[32] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of Artificial "Gummy" Fingers on Fingerprint Systems. In Rudolf L. van Renesse, editor, *SPIE Optical Security and Counterfeit Deterrence Techniques IV*, volume 4677, pages 275–289, April 2002.

[33] J. R. Millan. Adaptive Brain Interfaces. *Communications of the ACM*, 46(3):75–80, 2003.

[34] J. R. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni. A Local Neural Classifier for the Recognition of EEG Patterns Associated to Mental Tasks. *IEEE Transactions on Neural Networks*, 13(3):678–686, 2002.

[35] F. Monrose, M. K. Reiter, Q. Li, and S. Wetzel. Cryptographic Key Generation From Voice. In *IEEE Conference on Security and Privacy*, 2001.

[36] F. Monrose, M. K. Reiter, and S. Wetzel. Password Hardening based on Keystroke Dynamics. *International Journal of Information Security*, 1(1):69–83, 2001.

[37] M. A. L. Nicolelis and J. K. Chapin. Controlling Robots with the Mind. *Scientific American*, 289(4):46–53, 2002.

[38] A. Perrig and D. Song. Hash Visualization: a New Technique to Improve Real-World Security. In *International Workshop on Cryptographic Techniques and E-Commerce*, pages 131–138, 1999.

[39] G. Pfurtscheller and A. Aranibar. Evaluation of Event-Related Desynchronization (ERD) Preceding and Following Voluntary Self-Paced Movement. *Electroencephalography and Clinical Neurophysiology*, pages 138–146, 1979.

[40] B. Pinkas and T. Sander. Securing Passwords Against Dictionary Attacks. In *9th ACM Conference on Computer and Communications Security*, pages 161–170. ACM Press, 2002.

[41] Pulse Medical. EEG Supplies: EEG Electrode Cap. `http://www.pulsemedical.co.uk/productslink3.html`, site accessed Mar. 20, 2005.

[42] Real User Corporation. About Passfaces. `http://www.realuser.com`, site accessed May 24, 2004.

[43] A. R. Roddy and J. D. Stosz. Fingerprint Features - Statistical Analysis and System Performance Estimates. *Proceedings of the IEEE*, 85(9):1390–1421, 1996.

[44] P. Ross. Mind Readers. *Scientific American*, 289(3):74–77, 2003.

[45] V. Roth, K. Richter, and R. Freidinger. A PIN-Entry Method Resilient Against Shoulder Surfing. In *Conference on Computer and Communications Security*, pages 236–245, 2004.

[46] Leonardo Sobrado and J.-C. Birget. Graphical Passwords. The Rutgers Scholar: An Electronic Bulletin of Undergraduate Research, Volume 4, 2002. `http://rutgersscholar.rutgers.edu/volume04/sobrbirg/sobrbirg.htm`, site accessed Mar. 22, 2004.

[47] E. Spafford. Crisis and Aftermath (The Internet Worm). *Comm. of the ACM*, 32(6):678–687, 1989.

[48] S. Stubblebine and P.C. van Oorschot. Addressing Online Dictionary Attacks with Login Histories and Humans-in-the-Loop. In *Financial Cryptography'04*. Springer-Verlag LNCS 3110, 2004.

[49] G. Tally, R. Thomas, and T. Van Vleck. Anti-Phishing: Best Practices for Institutions and Consumers, March 2004. `http://www.networkassociates.com/us/_tier2/products/_media/mcafee/wp\_a%ntiphishing.pdf`, site accessed Mar. 22, 2005.

[50] The Editors of Scientific American. *The Scientific American Book of the Brain*. New York: The Lyons Press, 1999.

[51] J. Thorpe and P.C. van Oorschot. Graphical Dictionaries and the Memorable Space of Graphical Passwords. In *13th USENIX Security Symposium*, 2004.

[52] J. Thorpe and P.C. van Oorschot. Towards Secure Design Choices for Implementing Graphical Passwords. In *20th Annual Computer Security Applications Conference*, 2004.

[53] T. M. Vaughan, W. J. Heetderks, L.J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw, and J. R. Wolpaw. Brain-computer interface technology: A review of the Second International Meeting, 2003.

[54] K. Warwick, M. Gasson, B. Hutt, I. Goodhew, P. Kyberd, H. Schulzrinne, and X. Wu. Thought Communication and Control: a First Step Using Radiotelegraphy. *IEEE Proc. Commun.*, 151(3):185–189, 2004.

[55] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-Computer Interfaces For Communication and Control. *Clinical Neurophysiology*, 113:767–791, 2002.

[56] J. Yan. A Note on Proactive Password Checking. ACM New Security Paradigms Workshop, New Mexico, USA, 2001. `http://citeseer.nj.nec.com/yan01note.html`, site accessed Jan. 12, 2004.

[57] Jianxin Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. The Memorability and Security of Passwords – Some Empirical Results. Technical Report No. 500, Computer Laboratory, University of Cambridge, 2000. `http://www.ftp.cl.cam.ac.uk/ftp/users/rja14/tr500.pdf`, site accessed September 6, 2004.