Deceptive and Counter-Deceptive Machines
Papers from the AAAI 2015 Fall Symposium

Mind ID: A Psychologically Inspired Approach to Secure Authentication Based on Memory for Faces

Alexei V. Samsonovich
Krasnow Institute for Advanced Study, George Mason University
asamsono@gmu.edu

Abstract

The identity of every human subject is imprinted in the subject's mind. This work explores one possible approach to authentication of the user identity using implicit long-term memory, specifically, memory for faces. New elements include social-emotional associated actions similar to episodic memories used in the training and authentication procedure. Also in contrast with related studies, the method is designed to defeat shoulder-surfing attacks: even prolonged observation of the authentication procedure does not allow a third party to learn how to pass the challenge without special training. Potential applications, problems and ideas of alternative approaches are discussed.

Introduction

Secure authentication research is currently a hot topic in science and engineering. In addition, it has a huge potential practical value. Passwords and pin codes as a method of authentication are incompatible with rapidly developing modern technology, modern lifestyle, and techniques of phishing and hacking (Sasse et al., 2001). They need to be replaced with something that cannot be easily stolen or faked. A number of popular approaches that address this challenge are based on biometrics (fingerprint recognition, retina scan, iris recognition, heartbeat pattern, and so on). This solution, however, needs to be complemented with an alternative method that will not be prone to physical alteration or faking. The present work expands on the idea that the identity of a person is the mind of that person; therefore, the mind, specifically, implicit long-term memory, should be used for secure authentication.

Human memory is divided into a number of memory systems (e.g., Parkin, 1993; Goldstein, 2015), including short-term and long-term memory, explicit (or declarative) and implicit memory. Implicit long-term memory by definition cannot be communicated verbally, which makes it an ideal candidate. Among many kinds of implicit long-term memory, memory for faces has been recently a target of research in the secure authentication domain (e.g., Brostoff & Sasse, 2000; Jenkins et al., 2014). The present study differs from the previous attempts, in particular, by utilizing social and emotional aspects of memory for faces. This is done by implicitly associating learned faces with actions, reflecting social roles or positions of the imaginary subjects in a virtual social hierarchy, or by creation of episodic-like memories involving familiar faces.

Methods

Procedures used in this study were based on learning and recognition of faces. One set of faces is shown in Figure 1. The pool of subjects and procedures are described below.

Figure 1. An example of a set of faces used in the study. The material is taken from an online public domain.

Participants

A total of 12 George Mason University students participated in the study. Fifty percent of the students were female and 88% reported that English is their native language. The ethnic breakdown was as follows: 38% White, 38% Black,

12% Asian, and 12% reported other. One student did not report her ethnicity. A total of 50% of the students were from Northern Virginia while 25% reported that they were from a different part of Virginia. In terms of their student status, 38% were freshmen, 25% were sophomore, 13% were juniors, and 25% were seniors. All of the students were full time students. The students were majoring in Psychology, Computer Game Design, Economics, Criminology, Tourism and Event Management, Social Work, and Music Education. Participants were recruited via the Sona system.

Procedure 1

There are N virtual actors represented by familiar faces, connected to each other by a circular sequence of actions. Each actor is associated with actions in two roles: as an agent and as a patient. The associated actions are selected form the set of A actions (examples: hit, yield, greet, follow, lead). The goal for the participant during training is to learn faces representing actors and the actions associated with them. This is done in 7 levels of training sessions, each consisting of up to 500 trials.

At the first level, faces are learned individually, each together with one action associated with the virtual actor.

At the second level, in each turn of the training session, the subject is presented with a pair of faces and an arrow connecting them. The arrow is annotated with the associated action. The subject is asked to hit the key matching the action name.

At the third level, the same procedure repeats, only the arrows are not annotated: the subject must guess or recall the associated action, then hit the corresponding key. After that the correct action is displayed: the annotation emerges next to the arrow.

At the fourth level, the subject is still presented with pairs of familiar faces, only the arrow connecting them is not displayed. The subject must recall the roles of actors in this given pair, as well as the action connecting them. After the subject hits the action key, the annotated arrow is displayed.

At the fifth level the same procedure continues, however, in addition to the pair of familiar faces, four randomly chosen unfamiliar distractor faces are displayed. All six faces are placed at random locations in a 2x3 matrix.

At the sixth level, the same procedure continues, however, the arrow is not displayed at all. Instead, mistakes are signified by sound. Moreover, in some turns, instead of two, there may be a different number of familiar faces on the screen: 0, 1, or 3. In such cases, the subject is asked to hit a random key, and the response is ignored. Insertion of these 'wrong' turns is necessary to make learning over the shoulder (based on statistical learning) practically difficult.

The seventh level corresponds to the actual authentication challenge. Individual mistakes are not signified. In order to be authenticated, it is necessary to complete a sequence of turns without mistake. The length of the sequence depends on the desired security level.

Finally, shoulder surfing was either performed with one subject watching another during authentication, or simulated by displaying the correct action key in every turn.

Procedure 2

Among many explored variations of the method, the following one was used.

Paradigm, Settings and Rules

There are A+1 actions (e.g., hit, yield, greet, and ignore, with 'ignore' being reserved for skipping a trial).

There are N abstract human faces in total, all faces are unique to each test material.

Each of randomly selected M faces is associated with a randomly picked action; however, the number of faces associated with each action is kept constant.

Training

Goal: Learn face-to-action mapping for all M faces, be able to recognize each face

Procedure: repeated trials with feedback and multiple levels of difficulty.

Test

A sequence of T=20 non-ignored trials, pass if no mistakes and less than 3 ignores.

If fail, can try again once.

The paradigm for each trial is the following:

- L=6 faces are displayed on the screen,
- K of those faces were learned and associated with actions – must recognize them all before responding;
- Must respond following the rules.

Rules of response:

- If K=1, do the associated action;
- If K=2, do any action not associated with the 2 faces, try not to repeat yourself;
- If K=0 or K>2, do any action, try not to repeat yourself in any obvious way;
- If not sure, then "ignore" (try not to).

Constraints for automated generation of trials

- Probabilities of faces to appear on display at any location are uniform and independent; the only restriction is that there should be no duplicate faces on display. Any of the N faces (including learned and not learned) has the same probability to be displayed anywhere on the screen during each trial. No information about the subset of learned faces and the face-to-action mapping should be used in any way in selection and allocation of faces on display.

- The chances to see two learned faces on display should be twice higher than the chances to see only one learned face: P(K=2) = (A-1)*P(K=1). This, together with the first constraint, ensures that there is no individual face-to-action correlation visible in the sequence of correct responses.

Results

Procedure 1

Results for this procedure show that a typical subject can learn to be able to pass the challenge with high reliability in less than an hour (for N<10, A<6). The process of reliable authentication ensuring the probability of passing by chance below 10^{-9} may take up to two minutes. Most importantly, subjects are unable to pass the challenge on unfamiliar material better than at the chance level, even after an hour of "watching over the shoulder" another real or virtual subject doing the authentication.

Procedure 2

Similar qualitative observations were made in the alternative procedure. As the Figure 2 shows, after approximately 30 minutes of training, subjects reliably pass the test on familiar material (solid color areas in Figure 2). However, "watching over the shoulder" allows selected subjects to learn how to pass the test on an unfamiliar material as well (pink area in Figure 2).

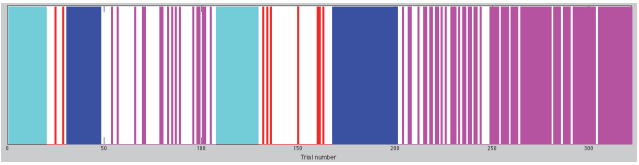

Figure 2. The figure represents successes and failures of all subjects in all trials during the first session. The color coding is as follows. Cyan: Subject 1 on own material. Red: Subject 2 on the material learned by Subject 1. Blue: Subject 2 on own material. Magenta: Subject 2 on the material learned by Subject 1. White: failures.

Therefore, it is important to take into account the strong human statistical learning ability when designing the procedure. Increasing the number of faces and associated actions strengthens the security, while at the same time makes the authentication challenge more difficult.

Discussion

Strengths of the method

- Watching one or several test procedures in principle does not allow one to break the code.
- The knowledge cannot be easily transferred in words, which excludes possibilities of sharing the code with friends or disclosing it at a gun point.
- With the new level of security, time requirements practically do not change. The training may not take longer than opening a bank account; the test procedure may not take longer than entering a credit card number.

Weaknesses

- **Weakness:** The code can be broken by learning pair-to-action associations from the display-response statistics. If two faces in a pair have different associated actions, then one and the same action will be produced whenever this pair appears on display and there are no other familiar faces. The easiest first step in code breaking is to identify pairs that have significant correlations with actions. This will require a pair to appear on the screen a sufficient number of times (there is a presumably uniform "background noise" coming from K=0 and K>2 cases). As a consequence, it may be necessary to accumulate statistics of responses over hundreds of trials, which may be practically difficult.
- **Mitigation:** The test material can be altered or expanded once in a while, thereby making this (and any response-statistics-based) method of code breaking useless. Of course, this would require regular re-training, which may be necessary for other reasons as well (see below).
- **Weakness:** Over time, after a long (how long?) period of inactivity, the subject may forget some of the faces or their associated actions.
- **Mitigation:** The subject should be required to use the system regularly (not necessarily for real transactions), and if necessary take additional training. Additional experiments are necessary to determine how long the memories last.
- **Weakness:** The subject may not sufficiently randomize responses in the K=0 and K>2 cases. E.g., if in all such cases the subject uses one standard response, then breaking the code may be possible based on only a few observed tests.
- **Mitigation:** The subject can be penalized for not sufficient randomization and required to take additional training.
- **Weakness:** The subject may react psychologically differently to different trial cases (K=0 vs. K=1 vs. K=2, etc.), which may be possible to detect visually or with other natural senses for somebody who learns to detect such differences.

• **Reply:** This needs testing. Modifications of the paradigm may be necessary.
• **Weakness:** Using Google Glasses to record the eye fixation point on the display during the test may allow one to break the code.
• **Reply:** This needs testing. An additional element in training may be necessary.
• **Weakness:** The subject, in principle, can describe the learned faces and associated actions. While it is practically difficult to confidently identify a face during the test based on a limited verbal description, in combination with the first approach this method may work, if a verbal description can be somehow obtained from the subject.
• **Mitigation:** The pictures of faces, and the appearance of faces themselves, can be altered over time, while keeping the faces recognizable. No specific features (such as text on t-shirts, unique objects in the background, etc.) should be used in photographs. Moreover, faces can be artificially generated on computer, and optimized in their mutual dissimilarity and uniqueness.
• **Weakness:** Photographs of the test material can be obtained using Google Glasses (or screen capture, if the test is available over the Internet). Then the subject (kidnapped or at a gun point, etc.) can be presented with the captured photographs and forced to identify learned faces and the associated actions.
• **Reply:** True; however, this problem is not adding to other possibilities of, e.g., kidnapping a daughter and demanding a ransom. The test material, including its frequencies (without subject's responses), contains no information that would allow one to break the code.

Alternative Paradigms

A=1, L=4.
If there is exactly one familiar face on display, say "yes", otherwise say "no".
Suppose all individual probabilities and correlations can be balanced.
Any pair of familiar faces will guarantee a "no" response (may be easy to notice).
Modification of response rule: 0 even, 1 odd, 2 even, 3 odd, 4 even.
All faces appear with natural uniform frequencies (no information about selection).
Can make zero correlations for individual faces
Can we also eliminate correlations for pairs?
Probably cannot eliminate bias of variances?

General Discussion

Imagine a future without passwords and credit cards. Everybody has the ability to be securely authenticated in order to authorize transactions or do business virtually anywhere at any time, and this ability cannot be stolen or compromised by bad guys. Moreover, the person does not need to remember a secret code, carry a special object or device, swallow a pill, get an implant, etc. Biometrics can be added to enhance security, but they are not a part of the solution. In the proposed approach, a part of the mind that is not transferrable becomes a secure ID of the person. While this scenario sounds too unrealistic even for a science fiction, it can be achieved in the nearest future, as argued below.

First of all, let us assume that a Global Authentication System is created by the US Government, based on a cloud of secure servers that are not accessible to the general population and cannot be all destroyed by a disaster. The System stores a personal key (explained below) for each individual client. The System communicates with businesses that require authentication of clients via a cell phone network or the Internet. The first step in authentication is to get a public key of the person, which can be a reserved for this purpose phone number. The second step is to validate the knowledge of the private key, which by itself is never explicitly transmitted or displayed during the authentication. The validation can be done using a smart phone, a computer, or a special device owned by one of the two parties: the client or the business (the device does not store or receive the key).

The private key becomes imprinted in the client's mind during training in a bank or another secure facility, when the client opens an account. The nature of this memory does not allow the client to communicate it to another person without using the special client-specific data plus the software and equipment used for training or validation.

There are multiple possible approaches to implementation of this scenario. The choice depends on the balance between the required level of security and the time necessary for authentication. Two approaches are described below that differ in these parameters. The first approach ("secure authentication") will guarantee higher reliability, e.g., up to 10-9 and higher, while may require longer time for authentication, possibly on the scale of several minutes. The second approach would be characterized by relatively lower reliability (which is difficult to estimate a priori) and will take shorter time for authentication, estimated to be on the scale of a few seconds.

Conclusions

Secure authentication can be based on memory for faces. Preliminary experimental study shows that the key can be learned in less than 30 minutes. Authentication may require up to two minutes or possibly more, if a high level of security is required. The method has a potential for improvement. The study also shows that repeated demonstrations without using verbal comments (i.e., "watching over the shoulder"), as well as detailed verbal descriptions with-

out demonstrations and without usage of the test material, do not allow one person to transfer the key to another person.

Risks and mitigation: In principle, videorecording, etc. of multiple authentication sessions of a given client allows breaking the code using a computer. To mitigate this risk, the key can be changed frequently.

Fast authentication can be based on the ability to control the balance of a virtual object using a touch screen. This fast authentication method can be combined with a face memory test. The advantage of this combination is twofold: (i) reinforcement of the face memory without compromising the key, and (ii) enhancement of the fast authentication security. An experimental study is necessary to test this approach.

Acknowledgments

The author is grateful to Walter H.C. Drakeford, who inspired him to investigate the problem. The funding for this research was provided by the Russian Science Foundation, Grant RSF 15-11-30014.

References

Brostoff, S. & Sasse, M.A. (2000). Are passfaces more usable than passwords? A field trial investigation. *People and Computers XIV - Usability or Else. Book Series: BCS Conference Series*, pp. 405-424.

Goldstein, E.B. (2015). *Cognitive Psychology, Fourth Edition.* Cengage Learning. ISBN 1285763882.

Jenkins, R., McLachlan, J.L., and Renaud, K. (2014). Facelock: familiarity-based graphical authentication. *PeerJ*. 2014 (2): e444, doi: 10.7717/peerj.444.

Parkin, A.J. (1993). *Memory: Phenomena, Experiment and Theory*. Oxford, UK: Blackwell.

Sasse, M.A., Brostoff, S., and Weirich, D. (2001). Transforming the 'weakest link' - a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19 (3): 122-131.