

USING A COMBINATION OF SOUND AND IMAGES TO AUTHENTICATE WEB USERS

Jim Liddell¹, Karen V. Renaud¹ Antonella De Angeli²

¹University of Glasgow

²NCR FSD - Advanced Technology and Research, Dundee

ABSTRACT

Passwords and pin numbers are ubiquitous even though they are basically flawed and cause problems for users due to memory workload. There is a need for other mechanisms to be investigated. This paper explores a mechanism for web-based authentication which does not require any expensive hardware or extra software. Two evaluations were conducted using a combination of sound and images to authenticate web site users – exploiting users' associative memory strengths. The results and findings are presented, and conclusions drawn.

Keywords

Authentication, Associative Memory, Memorability, Sound and Image.

1. INTRODUCTION

User authentication is the process required to provide an individual with secure access to confidential or personal information or services. As well as a user possessing knowledge of their own unique user-id, they must provide data that verifies their identity. Such data can be categorized as something the user knows, something the user possesses, or something the user is [7].

In today's technological world, one cannot use a computer, access a bank account, or use a mobile phone without having to recall a sequence of characters and or digits. Despite such a widespread use, Personal Identification Numbers (PINs) and passwords have a number of well-known drawbacks regarding security and memorability. Alternatives are currently being explored.

The Visual Identification Protocol (VIP) is a solution based around the fact that humans have a vast memory for pictures [5]. VIP uses a similar interface to that of a standard Automatic Teller Machine (ATM) PINpad, but replaces numbers with images (Fig 1). An evaluation of this scheme

has demonstrated that VIP is a promising and easy-to-use alternative to the PIN approach, but that the benefits of using pictures instead of numbers may be easily disrupted by a bad design.

A similar alternative, known as Passfaces, makes use of the human brain's remarkable ability to recognise individual faces [3]. The system uses a 3x3 grid to display 9 faces

from which the user must select those which they have been assigned. Real User's own long-term trial has operated successfully for over 15,000 users, some being able to immediately recognize their Passfaces after two years of non-use. Nevertheless, a field evaluation of this scheme revealed controversial results and did not fully support the expected superiority of faces against passwords [3].

Biometrics is a different branch of authentication concerned with the use of physiological, anatomical, or behavioural information to uniquely identify an individual [4]. It provides authentication without the need to possess any verification data such as a token or password. This resolves the issues of memorability and security. For example, the use of keystroke dynamics has proved extremely successful, with a false-reject rate of only 4% and a false-accept rate of less than one in 10,000 attempts [1]. However, no mention is made of how factors such as a change of hardware, or damaged fingers, would affect these results. The main setback in the widespread adoption of biometrics in user authentication appears to be the relative infancy and expense of the additional hardware required by most biometric techniques.

This paper reports the design and evaluation of the *Audio-Visual Associative Protocol (AVAP)* an authentication scheme relying on the previously-proven efficacy of pictorial passwords and on the benefits of non-speech audio, thus exploiting previously untapped human associative-memory strengths.

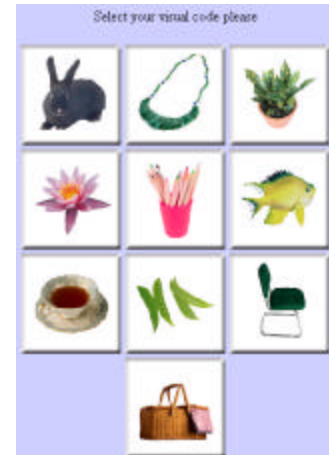


Figure 1: VIP

2. WEB AUTHENTICATION

To provide personalised services and access to confidential information remotely from a user's home computer, effective authentication techniques must be specifically tailored for use on the Internet. Most authentication on the Internet currently involves knowledge-based verification, as this technique is cheap and easy to implement, and suitable hardware is not generally available on a user's computer to verify any token-based authentication with a high enough degree of security. Similarly, biometric techniques are not widely used over the Internet, as these generally require specialised hardware and/or software on the user's machine.

Since the most common authentication scheme currently encountered on the Internet is knowledge-based, users are often required to recall a large number of different passwords. This leads to a high percentage of users who are unable to recall their password/s, which in turn forces users to record their passwords, creating a potential security flaw. Passwords are often chosen for memorability rather than security. A recent survey in our department revealed that over 10% of passwords were easily broken by a password cracking program.

One alternative verification method uses two separate channels to authenticate users [8]. Registration involves the submission of the user's phone number, and the server then initiates a call to the user, who responds with a unique digital authentication ID and transaction ID. This technique is specifically aimed at e-commerce-based applications. Although the system appears to work, it has a number of drawbacks in terms of widespread use. Firstly the system optimistically assumes that the user possesses a mobile. Secondly the system assumes mobile signal coverage wherever the user is – obviously an unrealistic assumption. Furthermore, this system involves a considerable amount of complex interaction which does not improve usability.

Both VIP and Passfaces prove that taking advantage of the human brain's ability to recognise faces or objects provides a promising alternative to the use of a sequence of digits or characters. These results provide strong motivation for the use of a pictorial system. Furthermore, research into the integration of sound into graphical user interfaces by Brewster [2] has shown that sound can improve usability by increasing performance and reducing time to recover from errors.

It was therefore decided to develop AVAP, an authentication mechanism involving both audio and visual information, thus exploiting previously untapped human associative-memory strengths. Association in this context is defined following Paivio's dual coding theory [6]. It is a functionally defined relation reflecting the probability that different units within the same memory system will activate each other.

3. AVAP AUTHENTICATION

In considering how both audio and visual information can be used to authenticate a user, it was assumed that an individual would make a visual association when a particular piece of music is heard. Since such associations would theoretically be based on each individual's personal life experience, it was hoped that such associations would vary sufficiently to authenticate users uniquely. To expedite this scheme, a number of these associations would be recorded for an individual at enrolment. Subsequent authentication is achieved by having the user recall the same associations. AVAP is based on the following hypothesis: *mnemonic associations between audio and visual information can be exploited to authenticate a user.*

A prototype was developed to authenticate users entering a particular website. The prototype records a number of associations during enrolment, and requests those associations to authenticate users for subsequent website accesses. Five image-sound associations were required. Users were given a randomly-selected sound and required to relate it to one of a corresponding set of 10 images. These associations have to *all* be recalled at subsequent site entries. The beauty of this scheme is that it is harder for the user to record their password, thus it increases the security of the scheme.

Audio controls were displayed to facilitate repeated audio activation. These controls were placed above the display of 10 images, in an ATM-like format (see Fig 3). Audio clips were chosen deliberately to provoke an association within a particular image set, and tended to mirror the general mood of a category (e.g., epic orchestral music corresponding to dramatic imagery of the cosmos). Nine audio clips were associated with groups of 10 images, six of which were semantically similar (i.e., same subject matter) and three random. We expected that grouping semantic images together would increase security by reducing the predictability of an association. For instance, it may be trivial to guess that an individual may choose an image of plant life for a given piece of music, but it may not be so easy to select what *type* of plant life an individual would select from a set of ten semantically-similar images. The three random categories were included to measure the relative performance of associations made using a random image set to those made using a semantically-similar set.

No attention was paid to the order in which users calibrated their associations when prompting recall. The system deliberately avoids emphasising sequence, which may be difficult to recall. The idea is that a user need not necessarily *remember* their associations, but that they would be able to make the same association later by relying on the audio cue, which should invoke the same association as it did originally.

4. FIRST EVALUATION

AVAP was evaluated in a large-scale longitudinal experiment comparing different types of authentication systems and involving almost 100 third year Computing Science students at the University of Glasgow. The authentication mechanism controlled access to a personalised module web page – thus creating the need for subjects to be authenticated. The experiment ran for three months and all accesses were logged. In this paper we will report a sub-set of results involving 65 users randomly assigned to one of the following conditions: PIN (5 digit PIN number), VIP (5 pictures in sequence), AVAP (5 audio-visual associations).

4.1 Results and discussions

Two main objective variables were analysed to compare the three authentication systems: (a) frequency of use (as defined by the number of correct authentication trials); (b) number of authentication errors.

A one way ANOVA was run to compare the difference between frequency of use in the experimental conditions. A significant difference emerged, $F_{(2,64)} = 3.06$, $p < 0.05$. Analysing the value of the means and the Post-Hoc comparisons, it appears that this is mainly due to the difference between VIP and AVAP. VIP was the most used condition (mean = 14.95), while AVAP was the least used (mean = 9.82).

Error occurrence was found to vary significantly by experimental conditions, $\chi^2_{(2)} = 5.94$, $p < 0.05$. Error percentages are illustrated in Figure 2. The worst condition is AVAP, with 24% errors.

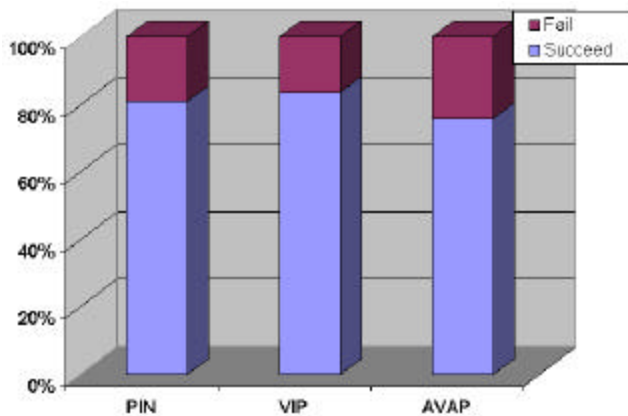


Figure 2: Failure Rates

In general, users being authenticated with AVAP tended to make more mistakes than others and tended not to make use of the web site as much as the other users. This can be attributed to two factors – firstly problems with the use of sound in the student lab (caused by difficulties with sound and different browsers on the Linux platform). The other factor appears to have been the time-consuming nature of the authentication process.

Another finding that caused concern was that users tended to make the same associations as others and that the hoped-for spread of associations across images did not seem to materialise. For one group of semantically similar images, only half the images were associated with the sound and in the best case, where 9 of the 10 images were chosen, one particular image was chosen by a third of the users. Furthermore, the following was noted:

- Users did not always listen to the audio, but only chose pictures they liked, which reverts AVAP to VIP.
- Users became irritated with the time taken for the images to download at each site access.
- Users struggled to distinguish between semantically-similar images and thus made frequent errors.
- Images were deemed to be too abstract.

To conclude, the initial evaluation was plagued by technical limitations which caused users to make less frequent use of the system, and to encounter a higher number of errors whilst doing so. The conditions under which the experiment was performed were not adequate to test the hypothesis in a proper manner as too many confounding variables were introduced. This led to the decision to run a revised evaluation which attempted to create a more realistic environment.

5. SECOND EVALUATION

The second evaluation was aimed specifically at home and office users, who would exhibit a more varied spread of age and gender. A web site was created which was updated daily with new jokes, links, facts, stories and illusions. It was hoped that this would provide a suitable incentive for users to log in on a frequent basis.

To reduce error occurrence, it was decided to abolish the semantic grouping of images. Images for the revised prototype were chosen to be less abstract and more meaningful than those originally used. An example is shown in Figure 3. Enrolment again required the user to make five associations, although this time an entirely random set of ten images was presented, alongside an unrelated audio clip. The user was asked to recall only a random two from these five associations to authenticate, in an effort to make the system easier, less time-consuming and more enjoyable to use.

The evaluation involved a total of 26 voluntary participants, 10 female and 16 male. Participants were emailed occasionally to remind them to log in on a daily basis. They were asked to complete a short questionnaire upon enrolment, and once again at the conclusion of the evaluation in order to test reactions to AVAP.

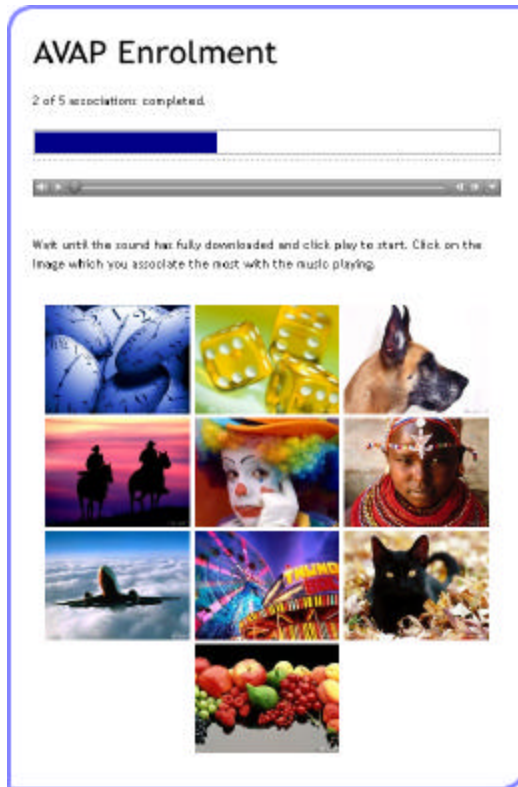


Figure 3: AVAP Enrolment

5.1 Results and discussions

The revised AVAP prototype appeared to be more popular with the users. On average, participants made far more use of the site, with 7 accesses per month versus 3 accesses in the first evaluation. This is particularly relevant given the fact that participants in the second evaluation did not have any obligation to access the web site. However, no significant difference was observed in terms of error frequency.

The user reaction to AVAP was very positive. It was evaluated as being significantly more exciting and relaxing than traditional authentication mechanisms, such as PINs and passwords. Users were only dissatisfied with the speed of AVAP, but no differences were reported as regards security, ease of use or efficiency.

The following user comments also emerged from the post evaluation questionnaire.

- Users mentioned that they would have liked access to a greater variety of images in choosing their associations.
- Some users expressed concerns about the predictability of the AVAP mechanism with respect to the security of access.
- Some users disliked the use of popular music in the experiment.
- Users appeared to enjoy the AVAP authentication more than PIN or password-based authentication.

6. CONCLUSIONS

AVAP provides an alternative approach to authentication that, while exciting and relaxing to use, does not appear to provide a more effective mechanism than PINs, passwords or alternative visual solutions. The original hypothesis – that mnemonic association between audio and visual information can be exploited to authenticate a user – has not been conclusively proven. The survey conducted at the end of the revised experiment revealed that on average users rated AVAP at the same level of security as PIN, and considered AVAP to be more enjoyable and relaxing. Despite these praises however, the results of the evaluation show that AVAP was not a success in terms of improving memorability.

This study has shown that image sound associations are not individual enough to be used as an authentication mechanism in a highly secure environment. However, the very positive subjective reaction to AVAP stresses the potential for pleasurable authentication mechanisms to be used in leisure-based web sites. We believe that the interaction context should be considered when choosing an authentication mechanism. The enjoyment factor is something that is often ignored by security experts and this study emphasises the importance of it as a major determinant of the whole user experience.

7. REFERENCES

- [1] Bergadano, F Gunetti D & Picardi C. User authentication through keystroke dynamics. (2002). *ACM Transactions on Information and System Security (TISSEC)*. volume 5, number 4, pp367—397.
- [2] Brewster, S A. (1997). Using non-speech sound to overcome information overload. *Displays, Special issue on multimedia displays*, 17, pp 179-189.
- [3] Brostoff, S., & Sasse, A. 2000. Are Passfaces more usable than passwords? A field trial investigation. In S. McDonald (ed.) *People and Computers XIV - Usability or Else!* Proceedings of HCI 2000, (Springer) 405-424
- [4] Coventry, L., De Angeli, A., and Johnson, G.I. (2003). Usability and Biometric Verification at the ATM Interface. *CHI 2003 Proceedings*, ACM Press.
- [5] De Angeli, A. *et al.* (2002) VIP: a visual approach to user authentication. *Proceedings of the Working Conference on Advanced Visual Interfaces AVI*. 2002. ACM Press. pp316-323
- [6] Paivio, A. (1971) *Imagery and Verbal Processes*. New York. Holt, Rinehart & Winston.
- [7] Renaud, K. & Smith, E. Helping Users to Remember Their Passwords. (2001) *Proc SAICSIT Conference*. Pretoria. South Africa. pp 73-80.
- [8] Resolving webuser on the fly (2002) *ACM SIGCAS. Computers and Society*, Volume 32, Number 2