USABILITY AND USER AUTHENTICATION: PICTORIAL PASSWORDS VS. PIN

Antonella De Angeli, Lynne Coventry, Graham Johnson & Mike Coutts

NCR-FSD Advanced Technology & Research,
Discovery Centre, 3 Fulton Road
Dundee DD2 4SW

This paper presents the design and evaluation of the Visual Identification Protocol (VIP) an innovative solution to user authentication based on pictures and visual memory. Three authentication systems were prototyped and compared with the PIN approach in a longitudinal evaluation (N=61). The study revealed important knowledge about attitudes towards and behaviour with different authentication approaches. VIP was found to be easier to remember and preferred by users, but its usability can be easily disrupted by an inappropriate solution. A detailed error analysis is presented to help understand the limits and constraints of visual memory. This knowledge is instrumental in designing innovative authentication system.

Introduction

User verification is a crucial component of secure systems that provide access to valuable information or offer personalised services. You cannot withdraw money at the ATM, log in to your computer, or place a call on your mobile phone without remembering a sequence of numbers or letters. Despite such a wide diffusion, Personal Identification Numbers (PINs) and passwords have a number of well-known deficiencies reflecting a difficult compromise between security and memorability (Adams and Sasse, 1999). Secure codes correspond to random selection of alphanumeric strings being as long as the system allows but humans struggle to remember meaningless strings. Thus, people choose passwords that are related to their everyday life and are often lax about the security of this information, writing it down or deliberately sharing it.

The approach to user authentication so far has been very technical in nature, concentrating on encryption methods and transmission protocols. Few proactive actions have been proposed to raise security awareness and drives secure behaviour (Adams and Sasse, 1999). Yet, current systems still suffer from a neglect of the human factors. Biometrics techniques have been proposed as a solution to memory limitations, but there are still many open issues with respect to adopting them in the ATM environment (Coventry *et. al.*, 2003).

This paper reports on the design and evaluation of the Visual Identification Protocol (VIP) a concept aimed at improving user authentication in self-service technology by replacing the precise recall of a numerical code with the recognition of previously seen images, a skill at which humans are remarkably proficient. Limits and potentialities of the approach are discussed based on a literature review and our experimental findings. In particular, the paper concentrates on errors and mnemonic interference, proposing guidelines to exploit the advantage of visual memory as a means for user authentication.

Exploiting visual memory

Humans have a vast, almost limitless memory for pictures. Classic cognitive science studies have shown that images are usually remembered far better than words or numbers (Madigan, 1993). Furthermore, it has been argued that visual memory is less affected by the general decline of cognitive capabilities associated with ageing than other types of memory (Park, 1997). Picture superiority over alphanumeric material has been attributed to encoding, storing and retrieval differences (Madigan, 1993). In particular, free recall suits alphanumeric stimuli; recognition suits visual ones.

The idea of exploiting visual memory in user authentication is not new. In 1996 Blonder patented a graphical password requiring users to touch predetermined areas of an image for authentication. The theory was advanced and implemented on a PDA, thus exploiting the input capabilities of graphical device (Jermyn *et al*, 2000). Drawings proved to be harder to crack than passwords and were assumed to be easier to remember. An user evaluation demonstrated that when an exact match is required, drawings are more difficult to replicate than passwords (Goldberg *et al*, 2002).

Graphical codes are becoming increasingly popular in personal technology. A common design solution requires the user to select target pictures among a set of distractors. An example is Passfaces by Real User Corp., based on face recognition. Users are given ‘five faces’, which represent their visual code. Each ‘face’ is displayed on a separate screen amongst different distractors. A field evaluation of this scheme revealed controversial results and did not fully support the expected superiority of faces against password (Brostoff and Sasse, 2000). Following a similar paradigm, Dhamija and Perrig (2000) investigated the memorability of abstract and photographic pictures against passwords and PINs. They showed that creating passwords and PINs is much faster than selecting an image portfolio, with photographic pictures requiring the longest time, but pictures are less error prone after a week interval.

A different approach to graphical authentication requires the user to simulate familiar actions on a graphical interface. An example is V-go by Passlogix, where users can ‘mix a cocktail’ or ‘cook a meal’ and the authentication code corresponds to the sequence of objects they clicked on. Despite the growing interest generated by different approaches to visual authentication systems, we are still far away from robust solutions, which could stand up to the challenge of public technology. Current proposal concentrates on maximising security of personal technology and may overestimate visual-memory potentialities. In particular, little attention has been devoted to understanding memory failures and use of this knowledge to improve graphical authentication systems. This paper is a preliminary contribution in this direction.

The Visual Identification protocol

VIP is intended to improve user authentication in self-service technology, supporting easy, fast and secure interaction. The objective is to find the best compromise between security and usability, following a user centred design process. VIP consists of a self-enrolment and an authentication. At enrolment, the user is automatically given an image portfolio, without the need for a printout. Different authentication processes were designed and evaluated (Table 1). All of them displayed detailed, colourful and meaningful photos of objects on a touch screen interface.

Table 1. Visual authentication systems

Prototype	Type of code	Location	Security scoring
VIP1	4 fixed order images from 10	Fixed	Same as PIN
VIP2	4 fixed order images from 10	Random	Intermediate
VIP3	Portfolio based	Random	Maximum

VIP1 is the pictorial equivalent of the PIN (Figure 1). It requires memorising a sequence of 4 pictures, which are always displayed in the same location on the visual keypad and must be entered in a fixed order. The interface resembles the PIN keypad but a new set of distractors is extracted from the database whenever the user makes an authentication attempt. Images in the visual database are clustered in semantic categories sharing common subject matter (e.g., flowers, animals, rocks, landscapes, etc). Each picture from the authentication code belongs to a different semantic category and the distractors are selected from the remaining categories. From a security point of view, VIP1 add to the traditional PIN all of the intrinsic advantages of pictures. Images are more difficult to describe verbally which should prevent users from revealing their code; they are easier to remember that should decrease the need to write down codes.

VIP2 differs from VIP1 in that the 4 pictures forming the authentication code are displayed in new random positions around the set of 10 locations of the visual keypad at each authentication attempt. Because of that, VIP2 is more secure than VIP1. It minimises the risks related to shoulder surfing since the position of the code is always different. VIP3 is a different concept, designed to investigate the limits of the visual paradigm. The user is assigned a portfolio of 8 pictures. At every authentication attempt, four of these pictures are randomly displayed in the challenge set together with 12 distractors (Figure 1). The distractors are selected from the database, avoiding the categories of the targets currently displayed in the challenge. To authenticate, the users select their images in any order. A new code is presented at each authentication trial. This minimises the risks related to several types of PIN thefts at the ATM.

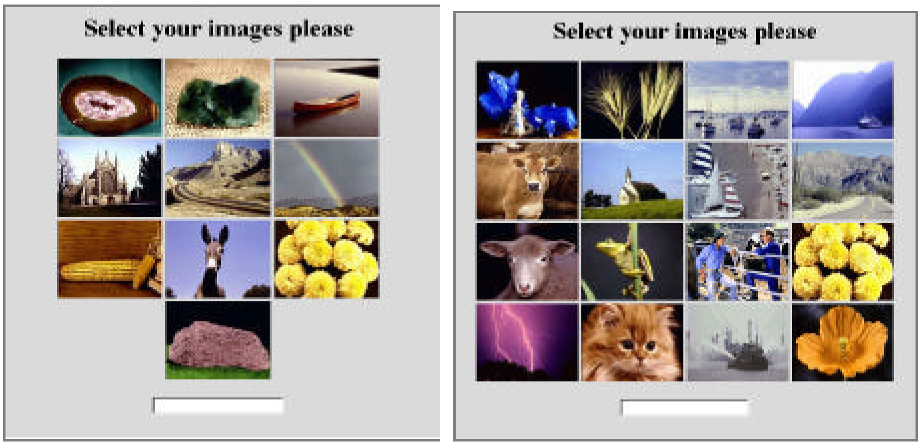

Figure 1. VIP1 and VIP2 (left hand side); VIP 3 (right and side)

The evaluation

A longitudinal evaluation addressed attitudinal, cognitive, and usability issues related to the VIP approach in comparison with PIN. The experiment involved 61 ATM users (29 males and 32 females), covering a broad range of ages (from 16 to 66 years, mean = 30), and education levels. A touch screen implementation of the traditional PIN approach was used as a control condition. Participants were randomly assigned to one of the four experimental conditions corresponding to different prototypes (PIN, VIP1, VIP2, and VIP3). Data were collected at three stages: learning, test1 and test2. After swiping an ATM card participants underwent the automatic enrolment and performed 10 authentication trials (learning). The first memory test took place 40 minutes later, after a unrelated task. The second memory test took place a week later. During both tests, participants had to swipe their card and entering the code 10 times in a row, as fast and accurately as possible. In case of erroneous entry, they were automatically given up to 3 attempts, as in a normal ATM transaction.

The evaluation provided an insight into cognitive constraints of visual and numerical memory in the context of self-service, addressing both performance criteria (number of errors and entry speed) and subjective evaluation (satisfaction relative to current PIN). Principal findings are summarised below, a more detailed report can be found in (De Angeli et al. 2002).

VIP was found to provide a promising and easy-to-use alternative to the PIN approach. The comparison between VIP1 and PIN demonstrated that pictures are less error prone than numbers after a week interval without compromising the speed of the transaction. The users' reaction to the VIP concept was promising. Overall, users liked VIP better and perceived it as more secure and easy to remember than PIN. Although these results need to take into account the novelty factors, they suggest potential acceptance of the VIP paradigm.

No difficulties emerged with respect to sequence retrieval in visual recognition. As a matter of fact, sequence errors were more frequent when participants had to retrieve sequences of numbers than when they had to recognise sequences of pictures. According to follow-up interviews, users supported sequence retrieval in visual configurations, by incorporating them into a mnemonic story.

The 'best' visual condition as regards objective and subjective evaluation was VIP1. VIP1 differed from VIP2 in that code retrieval was supported by an implicit memory related to the hand movement and to the position of the objects. These cues appeared to be important: they decreased the amount of errors and speeded up the action of entering the code. Participants were also more positive when evaluating VIP1 than any other visual conditions. The 'worst' design condition as regards performance criteria was VIP3. Participants who used VIP3 constantly achieved the slowest and more error prone performance. Nevertheless, VIP3 is the more secure solutions. The following error analysis can help us in the challenge of improving the usability of secure solutions.

Error analysis

None of the participants forgot their visual code throughout the duration of the experiment. However, in almost 6% of the authentication trials (100/1661) the users could not enter the correct code. These errors were not homogeneously distributed among the three conditions, $\chi^2_{(2)}= 46.95$, $p <.001$. Rather, they concentrated in condition VIP3 that accounted for 62% of all the errors (Figure 2). A slight advantage of VIP1 over VIP2 was also observed, $\chi^2_{(1)}= 3.42$, $p = .07$, supporting the idea that fixed location improves code retrieval.

To better understand the factors triggering errors, every wrong selection was tabulated according to its type. Multiple errors occurring in the same selection were coded as independent entries, leading to a sample of 120 errors. The following four error categories emerged from the analysis.

- *Sequence*: the correct code is retrieved but entered in a wrong order.
- *Double click*: the same item is unintentionally selected two consecutive times (the prototype did not allow corrections).
- *Double selection*: the same item is selected twice in non-consecutive positions.
- *Wrong selection*: one or more of the selected items does not belong to the authentication code.

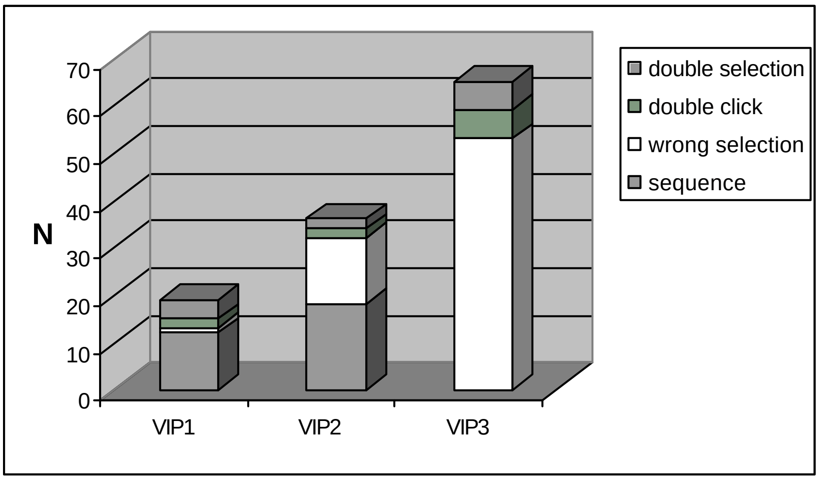

Figure 2. Error types as a function of experimental conditions

Analysing the graph in Figure 2, it emerges that different system configurations trigger specific error types. The poor performance of VIP3 was mainly due to wrong selections. All but four of these errors were due to intra-category confusions: participants tended to falsely recognise distractors belonging to the same category of items from their portfolio, which were not displayed in the current challenge set. Note that this type of errors could not occur in conditions VIP1 and VIP2 because of the selection algorithms implemented. It is worth noting, that some 63% of intra-category errors involved flowers, so that if participants had a flower in their portfolio, they were very likely to identify other flowers in the challenging set as 'their flower.

The wrong-selections occurred in condition VIP 2 (N=14) can be explained by physical proximity to the target and/or similarity. Physical proximity appears to reflect a slip in movement planning, whereas similarity reflects a different type of visual interference. It occurred particularly when the target was visually complex and difficult to label verbally (e.g., rocks and minerals, or skies view). In this case targets were confused with distractors which had very similar visual configurations (same shape or colour) even if they belonged to other semantic categories (e.g., a yellow flower mistaken for a yellow mineral).

VIP1 and VIP2 were mainly affected by sequence errors. However, 95% of them occurred at the very beginning of the interaction, when the user had to build an understanding of the way the system worked.

Conclusion

Visual memory is sensitive to interference (Goldstein and Chanche 1970). This is a natural consequence of the constructive nature of visual memory: a general concept rather than details is stored. For example, if we are asked to remember a room, our memory will not hold all the details of that room. It will, however, remember the important details and piece together the rest of the details from what makes sense. Details that are not held in memory can be added later to that memory. This explains why we tend to forget details and can be blind to small differences.

The contribution of this paper is in the identification of factors, which trigger interference. This phenomenon may affect the accuracy of code selection and hamper the picture superiority effect. We have demonstrated that those pictures sharing common subject matter (i.e., belonging to the same semantic category) interfere with each other. Fine discrimination among items of the same category is difficult, particularly if those items belong to a familiar category such as flowers. Another important factor is inter-stimulus similarity.

Applying this knowledge to the design of novel visual approach to user authentication, we following two guidelines: Firstly use concrete, nameable, clear, coloured, and distinctive images, since they are easier to remember and preferred. Secondly control the visual configuration of the challenge set. In particular, avoid displaying distracters from the same semantic category of items belonging to the user portfolio.

References

Adams A. and Sasse M.A. 1999, Users are not the enemy, *Communications of the ACM*, 42, 41-46

Blonder, G. E. 1996, Graphical password. *United States Patent 5559961*

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

Coventry, L., De Angeli, A. and Johnson, G. 2003, Usability and biometric verification at the ATM interface. *Proceedings of the ACM Human Factors in Computer Systems CHI'03 Conference*, Fort Lauderdale, (ACM Press).

De Angeli, A., Coutts, M., Coventry, L., Johnson, G.I, Cameron, D. and Fischer M. 2002. VIP: a visual approach to user authentication. *Proceedings of the Working Conference on Advanced Visual Interfaces AVI 2002*, (ACM Press), 316-323

Dhamija, R. and Perrig, A. 2000, Déjà vu: A User Study Using Images for Authentication. In *Proceedings of 9th USENIX Security Symposium*, 45-58.

Goldberg, J., Hangman, J. and Sazawal, V. 2002, Doodling our way to better authentication. Poster presented at *ACM Human Factors in Computer Systems*: *CHI'02*, Minneapolis (ACM Press)

Jermyn, I., et al. 2000. The design and analysis of graphical passwords. *Proceedings of the 9th USENIX Security Symposium*.

Madigan, S. 1993, Picture memory. In J.C. Yuille (ed.) *Imagery, memory, and cognition: Essays in honor of Allan Paivio*. (LEA Hillsdale, NJ), 66-89

Park, D.C. 1997, Ageing and memory: Mechanisms underlying age differences in performances. In *Proceedings of the 1997 World Congress of Gerontology.*