Etudes expérimentales d'un système d'authentification graphique basée sur la catégorisation sémantique

Pascal Salembier, Moustapha Zouinar, Robin Héron, Christophe Mathias, Guirec Lorant, Jean-Philippe Wary

Etudes expérimentales d'un système d'authentification graphique basée sur la catégorisation sémantique

Pascal Salembier
ICD-TechCICO (UMR 6281)
Université de Technologie de Troyes
10001, Troyes, France
pascal.salembier@utt.fr

Moustapha Zouinar
Robin Héron
Orange Labs
Code postal, Ville, Pays
moustafa.zouinar ;
robin.heron@orange.com

Christophe Mathias, Guirec Lorant, Jean-Philippe Wary
Orange Labs
Code postal, Ville, Pays
christophe.mathias ;
guirec.lorant ;
jeanphilippe.wary@orange.com

RÉSUME
Cet article présente une synthèse des résultats de deux études expérimentales menées dans le cadre d'un projet de recherche portant sur la conception et l'évaluation d'un système d'authentification graphique (HSA® pour Human Semantic Authentication). Les deux expérimentations ont porté sur plusieurs dimensions de l'utilisabilité du système (la performance) et ses exigences au plan cognitif (mémorisation et catégorisation). Les résultats mettent en évidence que si ce type de système présente un intérêt potentiel en termes d'usage lorsque certaines conditions sont remplies, il soulève un ensemble de difficultés d'ordre perceptif et cognitif qui obère son utilisation et nous interroge sur la généralisation de son usage en situation réelle. Le système ayant été principalement imaginé pour renforcer la sécurité, ces résultats illustrent par ailleurs les tensions qui existent entre sécurité et utilisabilité.

Mots Clés
Authentification graphique ; utilisabilité ; sécurité.

ABSTRACT
This article sums up results from two experimental studies conducted in the context of a research project on the design and assessment of a graphical authentication system (HSA® for Human Semantic Authentication). The three experiments address different issues related to the usability of the system (performance), and its cognitive demands (memorisation and categorisation). The results put into evidence that, despite the fact that this type of system presents some advantages in terms of use (providing some requirements are observed), it raises nonetheless different problems which put uncertainty on its potential acceptabilty and use in real world situations. The results illustrate the traditional tension between security and usability.

Author Keywords
Graphical password ; usability ; security.

ACM Classification Keywords
K.6.5 Security and Protection : Authentication ; H.5.2 Information interfaces and presentation (e.g., HCI): Evaluation/methodology.

INTRODUCTION

Les problèmes de sécurité soulevés par l'usage des systèmes d'authentification ont suscité un développement important de travaux associant sécurité informatique et interaction Homme-Machine [2]. Le système d'authentification le plus couramment utilisé (mot de passe alphanumérique) s'avère en effet de plus en plus vulnérable face à des attaques de plus en plus sophistiquées, qu'elles soient automatiques ou d'origine humaine. Cette vulnérabilité est en grande partie liée aux usages de ce type de système (p. ex. la tendance des utilisateurs à choisir des mots de passe facilement identifiables). Il est en effet maintenant reconnu que la question de la sécurité des systèmes informatiques ne peut être réduite à un problème de nature purement technique et qu'elle doit intégrer la composante humaine [11]. La multiplication et la disparité des mesures de protection mises en place par les concepteurs conduisent parfois les utilisateurs, qui ne disposent pas forcément d'une représentation adéquate des enjeux, à mettre en œuvre des stratégies de contournement de ces mesures de manière à relâcher les contraintes qui rendent leurs activités quotidiennes plus difficiles, au détriment de la sécurité [8]. La recherche de solutions qui optimisent à la fois la sécurité et l'utilisabilité [2] a conduit les chercheurs du domaine à explorer de nouveaux systèmes d'authentification. L'authentification graphique a ainsi été proposée comme une alternative à l'authentification alphanumérique. La justification de cette innovation est double [2] : un renforcement du niveau de résistance aux attaques (les mots de passe traditionnels sont prédictibles lorsque c'est l'utilisateur qui les choisit) et une facilité d'utilisation accrue, en particulier du point de vue de la rétention mnésique (les mots de passe traditionnels posent des problèmes de mémorisation). Les deux objectifs sont en fait directement liés dans la mesure où un niveau d'utilisabilité moindre peut induire des comportements non souhaitables qui vont impacter directement le niveau de sécurité du système (trop grande facilité à inférer les

mots de passe, réplication des mots de passe pour des services différents, utilisation de notes de type « pense-bête », etc.).

Cet article présente de façon synthétique deux études ergonomiques d'un système d'authentification graphique (HSA pour Human Semantic Authentication®) qui questionnent l'intérêt potentiel et les inconvénients de ce type de système d'un point de vue « utilisateur » (utilisabilité, contraintes cognitives, usages).

L'AUTHENTIFICATION GRAPHIQUE

On distingue dans la littérature, trois grandes catégories de systèmes d'authentification graphique selon leurs caractéristiques en termes de mémorisation [2] : (1) dans les systèmes basés sur la reconnaissance, dits *cognométriques*, l'utilisateur doit mémoriser un ensemble d'images, par exemple des visages comme dans le système *Passfaces* [6] ; la procédure d'authentification consiste alors à les reconnaître parmi un autre ensemble plus large de leurres. (2) Les systèmes basés sur le rappel nécessitent la mémorisation d'un graphique ou dessin que l'utilisateur doit reproduire à l'aide d'une souris ou d'un stylet ; par exemple, *Draw-A-Secret* [5] constitue un exemple canonique de ce type de système appelé *drawmetrics*. (3) Dans les systèmes basés sur le rappel indicé, l'utilisateur dispose d'indices qui facilitent la tâche de récupération en mémoire du mot de passe. C'est par exemple le cas de *PassPoints* [11] ou *Passhints* [4].

LE SYSTEME HSA®

Ce système a été conçu pour rendre plus difficiles les attaques automatiques par espionnage (via des outils malveillants installés sur le terminal client, tels que des *spywares* ou *key loggers*) qui opèrent par captation des données ou par tentatives aléatoires d'authentification (pour plus de détails sur la sécurité, voir [9]). Le principe de fonctionnement de ce système du point de vue de la procédure d'authentification est le suivant : un mot de passe constitué de quatre concepts (par exemple, Jaune, Outil, Animal, Nourriture) est attribué à un utilisateur qui doit les retenir. Ces concepts sont « codés » dans des images dans lesquelles des instances de ces concepts sont présentes. Le codage consiste à définir les zones qui correspondent à ces concepts et à les rendre ainsi actives (figure 1). Ce codage est invisible pour l'utilisateur final. La phase d'authentification consiste à cliquer sur les zones qui correspondent au mot de passe (c'est-à-dire à chacun des quatre concepts), sachant que l'image change à chaque nouvelle authentification pour des raisons de sécurité. La tâche d'authentification est réussie lorsque l'utilisateur clique ou pointe sur les zones qui ont été pré-codées comme instances des concepts, et ce dans l'ordre défini au moment de la création du mot de passe. La tâche d'authentification peut être ainsi décrite comme une tâche d'association de concepts à des éléments picturaux, qui implique la mise en œuvre de processus de recherche visuelle et de catégorisation à partir de concepts). Comparé aux autres systèmes, avec HSA® l'authentification repose à la fois sur le rappel (il n'y pas d'indices explicites dans l'image) et la reconnaissance (l'utilisateur doit reconnaître des éléments visuels, comme des objets ou des couleurs, à partir de mots). Cependant, une caractéristique spécifique de ce système est qu'il n'est pas entièrement graphique puisqu'il inclut des mots ; c'est l'authentification qui s'effectue de manière graphique.

Figure 1. Exemple d'image codée pour le concept « Outil ». Les zones codées apparaissent en rouge sur l'image.

Caractérisation en termes d'utilisabilité

Du point de vue « utilisateur », ce système soulève un ensemble de questions liées à :

- l'utilisabilité du dispositif d'authentification : le système permet-il à l'utilisateur de s'authentifier de façon efficace et efficiente ? Le mode d'interaction par pointage est-il adapté à différents types de terminaux : tablette, smartphone ? Quelles sont les exigences en termes de mémorisation ? ;
- l'intelligibilité de la logique d'associations concepts-images : quels types de concepts/mots faut-il fournir aux utilisateurs ? Comment les choisir ? ;
- la compatibilité entre utilisabilité et critères de sécurité.

Suite à une première expérimentation[1] (qui est décrite plus en détail dans [9]) plusieurs éléments critiques ont pu être mis en évidence :

- Le mode d'authentification du système HSA ne présente pas de difficultés quant à sa compréhension ; les participants ont très rapidement compris la tâche.
- Les résultats globaux de la tâche d'association concept/instance peuvent être vus comme globalement satisfaisants (82% d'associations correctes).
- L'association correcte de 4 concepts à une image, qui approxime une procédure d'authentification basée sur le principe du mot de passe graphique telle qu'instanciée dans le démonstrateur HSA®, est insuffisante (40% d'authentifications réussies au premier essai2).

[1] Référencée comme « expérimentation 1 » dans la suite du texte.

[2] Dans cette expérimentation, l'objectif se limitait à analyser l'association entre concepts et images. Il n'était donc pas demandé aux sujets de faire plusieurs essais pour un concept au sein d'une même image. Dans la littérature, l'évaluation des

- Le temps mis par les participants pour identifier quatre instances (28 secondes en moyenne) paraît également peu satisfaisant au regard des seuils de performance considérés comme acceptable dans la littérature (généralement moins de 20 secondes [2]). Des temps d'authentification plus satisfaisants ont été relevés (13 secondes), mais ce résultat doit être nuancé dans la mesure où les participants n'étaient pas soumis à la contrainte de mémorisation (les concepts étaient présentés en clair), dont on peut supposer qu'elle est susceptible d'influencer la performance.
- Trouver des instances visuelles de concepts dans une image riche peut être particulièrement ardu dans certaines conditions (par exemple, lorsque les concepts sont trop abstraits ou nécessitent un travail inférentiel important, lorsque les instances ne sont pas suffisamment saillantes ou quand le système catégoriel du sujet et celui du codeur ne sont pas congruents).
- La procédure de codage des images doit être améliorée de manière significative pour limiter certaines associations valides mais non reconnues par le système par défaut de codage.

EXPERIMENTATION 2

Dans le continuité de la première expérimentation l'objectif de cette seconde expérimentation était de documenter la capacité d'un utilisateur à correctement mettre en œuvre la procédure d'authentification, c'est-à-dire associer des concepts (par exemple relatifs à des couleurs, des objets, des formes, des substances, ou des textures) à des zones d'une image qui « dénotent » ou renvoient à ces concepts.. Un objectif plus général était de poursuivre l'évaluation de l'utilisabilité du système dans sa version courante. Cette expérimentation tient compte des résultats de l'expérimentation précédente (choix des concepts et des images, modalités de codage des images). Plus précisément, nous avons repris les concepts qui nous sont apparus les plus satisfaisants du point de vue du compromis entre utilisabilité et de sécurité.

Méthode

Sujets

Le système a été testé avec 22 sujets familiarisés avec l'usage régulier de systèmes d'authentification alphanumériques, de langue maternelle française et ne présentant pas de dyschromatopsie avérée (daltonisme, perception sélective des couleurs,…) et autres troubles non corrigés (presbytie, myopie, etc.). La population était constituée de 12 femmes et 10 hommes.

Matériel expérimental

Le terminal utilisé pour l'expérimentation était une tablette tactile Android (Samsung GT-N8020, Android 4.1.2).

Les 16 concepts retenus (Tableau 1) sont extraits d'une liste établie sur la base des résultats de l'expérimentation précédente et sur un ensemble de nouveaux concepts. Les principes suivants ont été systématiquement appliqués :

- les instances doivent être suffisamment visibles (netteté, taille et position spatiale des éléments visuels) ;
- les concepts qui peuvent impliquer un travail inférentiel complexe, trop incertain ou un jugement qui fait appel à une autre modalité sensorielle que la vision ont été écartés (par exemple, « Lourd » ou « Léger ») ;
- les concepts de couleur ont été abandonnés car ils étaient susceptibles de fragiliser la sécurité du système par une analyse automatique de l'image (trop grande facilité d'identification des couleurs).

ACCROCHE	ANIMAL	APPAREIL	BOIS
ECLAIRAGE	ENFANT	JEU	LIQUIDE
METAL	NOURRITURE	PLASTIQUE	POINTU
PROTECTION	TRANSPORT	VEGETAL	VERRE

Tableau 1. Liste des concepts utilisés dans l'expérimentation 2

Les 16 images utilisées ont été collectées sur internet (photographies et peintures), et éventuellement retouchées pour les besoins de l'expérimentation (Figure 2).

Figure 2. Exemples d'images utilisées lors de l'expérimentation 2.

Comme dans la première expérimentation la recherche des images devait respecter un certain nombre de critères de sécurité qui sont les suivants : utiliser des images avec le moins de surface vide possible ; présence de plusieurs instances d'un même concept dans une image ; pourcentage de la surface totale de l'image occupée par une instance compris entre 2 et 20%[3]. Concernant le dernier critère, il inclut également un aspect « utilisabilité » puisqu'il définit un seuil en deçà duquel les concepteurs ont considéré que l'identification visuelle des instances deviendrait difficile[4]. La sélection finale a été réalisée conjointement par deux expérimentateurs. Du

systèmes d'authentification se centre en effet principalement sur l'authentification réussie au premier essai.

[3] Ces critères ont également été appliqués dans la troisième expérimentation décrite plus loin.

[4] Ce seuil est intuitif et mériterait sans doute des études expérimentales propres qui n'ont pas été réalisées dans le cadre de cette étude car il dépassait les objectifs de celle-ci. On peut toutefois noter que ce seuil n'a pas été une source de problèmes pour les participants dans les expérimentations présentées ici.

point de vue du codage, la zone sensible est constituée d'une zone focale (codée) et d'une zone de tolérance de 16 pixels autour de la zone focale pour tenir compte des marges d'erreurs relatives au toucher sur les écrans tactile.

Le codage des instances a été réalisé indépendamment par deux personnes (les deux expérimentateurs) avec un éditeur dédié qui permet de définir directement à l'aide de la souris les zones sensibles associées aux concepts et de vérifier en temps réel la surface relative occupée par les zones codées pour un concept donné. Les codages étaient dans un second temps comparés et les éventuelles divergences traitées. Cette procédure a été mise en œuvre afin d'éviter d'éventuelles omissions et limiter les possibles divergences entre d'une part les interprétations des concepts et des contenus des images par les « utilisateurs » et d'autre part les codages réalisés (par exemple, si les deux expérimentateurs constataient une divergence entre eux du point de vue du lien sémantique entre une instance et un concept dans une image soit ils la rejetaient soit ils codaient cette instance pour couvrir le maximum d'associations possibles, tout en respectant les critères de sécurité et d'utilisabilité mentionnés plus haut).

Procédure

Après une phase de familiarisation avec le système, il était demandé aux participants de réaliser par pointage tactile une tâche d'association entre un groupe de 4 concepts et une image à chaque fois différente (16 images). Chaque groupe de concepts était constitué d'une combinaison de 4 concepts tirés de la liste indiquée plus haut (tableau 1). Les concepts étaient affichés en même temps que l'image. L'ordre de passage des images et des concepts associés à chaque image a été tiré aléatoirement puis maintenu constant pour l'ensemble des sujets. Après chaque pointage, le participant appuyait sur un bouton de validation pour passer au concept suivant. Les sujets ne réalisaient qu'un seul essai par couple « concept/image » et n'avaient pas de retour sur la validité du choix car à ce stade de la recherche l'intérêt était principalement focalisé sur l'association concept/image.

A l'issue de chaque phase d'association entre un groupe de 4 concepts et une image on demandait aux participants d'évaluer le niveau de difficulté de la tâche d'association pour chacun des 4 concepts sur une échelle subjective de Likert-Osgood en 5 points (de « très facile » = 1 à « très difficile » = 5). On demandait également aux sujets de préciser, lorsque cela était pertinent, les difficultés qu'ils avaient éprouvées et les raisonnements mis en œuvre dans la tâche d'association. Lors d'un entretien global réalisé à la fin de la session de test (représentant 64 associations par sujet), les participants étaient invités à expliciter leur vécu de la réalisation de la tâche et de l'interaction avec le système. L'ensemble de la session faisait l'objet d'un enregistrement vidéo.

Recueil et traitement des données

Les actions des participants sur l'interface du système étaient automatiquement horodatées et enregistrées dans un fichier log. Elles ont constitué la base pour renseigner plusieurs indicateurs : taux de réalisation de la tâche d'association ; taux de correspondance entre le nombre de pixels activés lors de la sélection par le sujet d'une zone de l'écran associée à un concept, et le nombre de pixels codant la zone associée à ce concept ; temps de réalisation. Bien que le test n'ait pas été initialement conçu pour l'obtenir, nous avons également calculé un taux d'authentification correcte approximatif relatif à un premier essai (c.à.d. le nombre d'identifications correctes de groupes de 4 concepts associées à une image par un participant). Les verbalisations consécutives obtenues au cours de chaque session ont fait l'objet d'une analyse qualitative centrée sur différentes composantes en particulier, raisonnements mis en œuvre, difficultés éprouvées, catégorisation.

Résultats

Tâche d'association

Le taux moyen de réalisation de la tâche d'association entre un concept et une image est de 98%. Les quelques abandons constatés s'expliquent essentiellement par des difficultés ressenties à identifier des associations. Le taux moyen d'associations correctes entre zone cliquée et zone codée est de 88% (maximum : 99% ; minimum : 58%) dont 86% en zone focale et 2% en zone de tolérance (Figure 3).

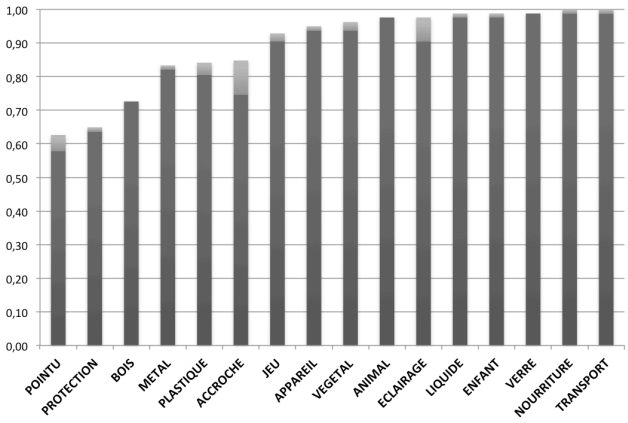

Figure 3. Taux d'association correcte entre zone codée et zone cliquée par concept (en noir : zone focale ; en gris : zone de tolérance).

Le temps de réalisation moyen des associations concept-image est de 5,81 sec. (écart-type : 4,18, avec un maximum de 43 sec. et un minimum de 2 sec. ;). A noter que l'on n'observe pas de corrélation entre la taille de la surface occupée par les instances des concepts et le temps d'association.

Pour ces deux indicateurs de performance on retrouve le phénomène de variabilité observé lors de l'expérimentation précédente ; les concepts « Pointu » et « Protection » sont ainsi correctement associés en moyenne dans 60% des cas alors que les concepts « Nourriture » et « Verre» atteignent les 100% d'associations correctes.

Les concepts « Enfant » et « Nourriture » nécessitent moins de 4 sec. pour être correctement associés à une

zone d'image, alors que les concepts « Accroché » et « Protection » requièrent plus de 9 sec. (Figure 4).

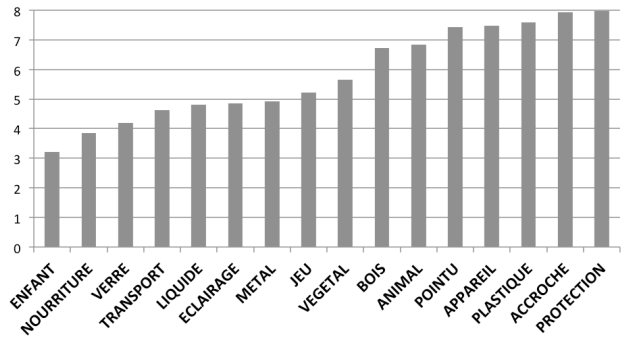

Figure 4. Temps moyen de réalisation de la tâche d'association par concept.

De même que dans la première expérimentation, les résultats semblent montrer l'existence d'une interaction entre images et concepts dans la tâche d'association. Ainsi le taux d'association correcte du concept « Bois » varie de 18% à 95% selon l'image. De même, le temps de réalisation de l'association pour le concept « Animal » varie de 4,45 sec. à 10,82 sec. en moyenne. Dans le cas du concept « Bois » on peut évoquer une explication en termes d'ambiguité perceptive : il s'est révélé plus difficile pour les sujets dans une image particulière de décider s'ils avaient à faire à du « Bois » ou du « Métal ». Pour le concept « Animal » une explication évoquée par les sujets tient à la saillance insuffisante de l'instance du concept dans une image donnée (difficulté à trouver le concept dans l'image).

Authentification

A titre indicatif nous avons estimé la performance des sujets en termes d'authentification au premier essai (associations entre une image et un groupe de 4 concepts. Les résultats obtenus mettent en évidence un taux d'association correcte moyen (approximant une authentification au premier essai) de 60% (taux supérieur à celui de la première expérimentation : moins de 50%) avec une variabilité inter sujets importante (de moins de 20% à plus de 80% de réussite). Le temps moyen d'authentification est de 24,19 secondes (maximum de 61 sec. et min. de 8 sec.) avec également une variabilité inter sujets notable (de moins de 20 sec. en moyenne à près de 35 sec.).

Estimation de la difficulté de la tâche par les sujets

La moyenne des scores d'évaluation de la difficulté perçue de la tâche d'association par sujet est de 1,77 (écart-type : 0,31) sur une échelle de 1 à 5. Les résultats obtenus varient assez sensiblement en fonction des concepts, d'un peu plus de 1 pour le concept « Enfant » à près de 2,5 pour le concept « Protection ». Ces résultats sont similaires à ceux enregistrés dans la première expérimentation..

L'analyse des verbalisations des sujets et de leurs actions a permis d'identifier les mêmes difficultés que celles observées dans la première expérimentation : décalage de catégorisation entre le codeur et le sujet ; défaut de saillance de certaines instances ; ambiguïté perceptive.

En conclusion, outre le fait qu'elle nous a permis de tester de nouveaux concepts avec de nouvelles images, cette expérimentation montre un progrès sensible du point de vue de la performance par rapport à l'expérimentation 1. On observe notamment une amélioration du taux d'association correcte (88% vs. 82%) et du temps de réalisation moyen (5,8 sec. vs. 7,2 sec.). Mais de nombreuses questions restent posées, en particulier les conséquences sur la mémorisation et l'usage dans un cadre d'authentification réaliste. La procédure de codage demeure également perfectible. C'est pour traiter ces points qu'une troisième expérimentation a été réalisée.

EXPERIMENTATION 3

Cette troisième expérimentation poursuivait plusieurs objectifs : évaluer le processus d'association entre concepts (tels que retenus suite à la seconde expérimentation) et nouveau jeu d'images ; poursuivre l'évaluation de l'utilisabilité du système HSA dans une situation d'usage plus réaliste, comme suggéré dans la littérature [1] (utilisation en environnement « naturel » ; tâche « réelle » d'authentification ; déroulement de l'expérimentation sur un temps plus long que lors des phases précédentes) documenter les pratiques de mémorisation des mots de passe mises en œuvre par les sujets et évaluer le niveau de rappel ; et examiner la projection dans un usage réel du système par les participants.

Méthode

Sujets

La population retenue pour cette expérimentation était constituée de 22 sujets familiarisés avec l'usage régulier de systèmes d'authentification alphanumériques. L'échantillon était composé d'étudiants en cycle d'ingénieur et en sciences humaines et sociales, d'enseignants-chercheurs, d'ingénieurs et de personnel administratif. La population était constituée de 12 femmes et 10 hommes ; l'âge moyen était de 33 ans (minimum : 17 ; maximum : 61).

Matériel expérimental

Le choix des concepts utilisés tient compte des résultats des deux premières expérimentations par rapport notamment aux taux d'associations correctes observés et aux difficultés d'association pointés par les sujets.

Trois sites internet fictifs ont été créés pour les besoins de l'expérimentation (figure 7) : réseau social (AMICO), banque en ligne (BANCO), achats en ligne (SHOPPING).

A chaque site était associé un mot de passe constitué de 4 concepts (tableau 2).

Le choix de tester trois sites a été principalement guidé par le souhait de nous rapprocher des situations réelles dans lesquelles les utilisateurs ont souvent recours à plusieurs mots de passe. Il est donc important d'en tenir compte pour analyser les effets sur la mémorisation.

Figure 7. Exemple de site fictif créé pour l'expérimentation 3.

Sites	Concept 1	Concept 2	Concept 3	Concept 4
AMICO	Transport	Eclairage	Végétal	Ecriture
BANCO	Nourriture	Accroché	Verre	Appareil
SHOPPING	Liquide	Animal	Métal	Bois

Tableau 2. Liste des concepts utilisés dans l'expérimentation 3.

Les concepts ont été sélectionnés sur la base des expérimentations précédentes. Chaque concept d'un mot de passe devait être associé à des zones d'une image tirée d'une liste de 48 images conçues et sélectionnées pour les besoins de l'expérimentation (24 images pour AMICO, 16 pour BANCO et 8 pour SHOPPING). Ces images étaient de 3 types : photographies (33), planches de bandes dessinées (14) ou dessin (1). La variation du nombre d'images est liée au fait que nous voulions étudier les effets de la fréquence de connexion et donc d'utilisation des mots de passe sur la mémorisation. Des critères relatifs à l'utilisabilité ont également été pris en compte suite aux deux premières expérimentations : éviter la superposition d'images ; optimiser la visibilité des instances (contraste, netteté, contour) ; minimiser l'ambiguïté des instances.

Les images ont été codées par les trois expérimentateurs en charge de la construction et de la passation de l'expérimentation, selon les modalités suivantes : une première phase de codage individuel suivie d'une confrontation collective pour vérifier l'accord inter-codeurs, et/ou réaliser un codage collectif.

Procédure

L'expérimentation a été menée sur une période d'environ 3 mois et demi. La consigne était dans un premier temps envoyée par courrier électronique aux participants qui se familiarisaient avec la tâche (avec l'aide d'un des expérimentateurs). Ils recevaient ensuite les mots de passe par messagerie électronique.

Après cette phase, les participants devaient se connecter et s'authentifier 48 fois sur les 3 sites conçus pour l'expérimentation (24 fois pour AMICO, 16 fois pour BANCO et 8 fois pour SHOPPING ; soit respectivement pour chaque site : 6 fois, 4 fois et 2 fois par semaine). Ils recevaient à échéance fixe un message électronique leur indiquant les connexions à réaliser durant la journée courante et le nombre de connexions réalisées/restantes pour chacun des 3 sites. Lors d'une connexion et après avoir indiqué leur adresse électronique, les sujets devaient saisir leur mot de passe en associant les 4 concepts composant ce mot de passe à une image (à chaque fois nouvelle), et en respectant l'ordre prédéfini ; ils disposaient de 3 essais. Il leur était ensuite demandé de remplir un court questionnaire en ligne qui visait à collecter des éléments sur les difficultés rencontrées (oubli du mot de passe, problème d'association, etc.). En cas d'échec, le site leur rappelait leur mot de passe. Les participants pouvaient utiliser un écran de PC ou une tablette mais pas un smartphone pour réaliser le test dans la mesure où le système n'était pas encore adapté à ce type de terminal. A l'issue des 4 semaines d'expérimentation, un entretien était réalisé avec les sujets sur la base des données collectées et du matériel expérimental utilisé.

Recueil et traitement des données

Pour cette expérimentation, les données recueillies étaient les suivantes : logs horodatés, réponses aux questionnaires en ligne, et verbalisations recueillies au cours des entretiens semi-directifs post test. Le traitement de ces données était de même nature que dans la première expérimentation.

Résultats

Tâche d'association

Le nombre total d'associations réalisées par les sujets au cours de l'expérimentation est de 4134. Comme dans la première expérimentation, la tâche a été correctement comprise et réalisée par l'ensemble des participants.

Le taux de réussite (associations correctes) est de 91%. On observe une certaine variabilité selon les concepts : d'un peu plus de 80% pour Métal et Bois à plus de 95% pour Transport en moyenne ; Figure 5).

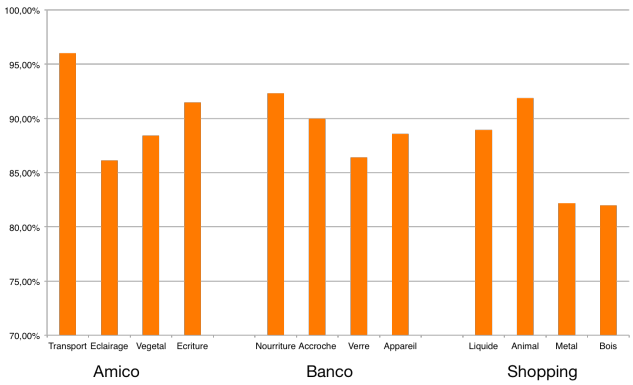

Figure 5 : Taux de correspondance au premier essai par concept.

Lorsqu'on examine le temps mis par les sujets pour trouver les instances par concept (Figure 6), on constate que les premiers concepts des mots de passe prennent significativement plus de temps que les concepts suivants (en moyenne14 sec. pour le 1er ; 4 sec. pour le 2ème ; 4,5 sec. pour le 3ème et 4,2 sec. pour le 4ème). Plusieurs explications sont possibles : le sujet regarde l'image dans son ensemble, il cherche tous les concepts avant de cliquer, la recherche du mot de passe en mémoire ou dans des supports externes prend du temps.

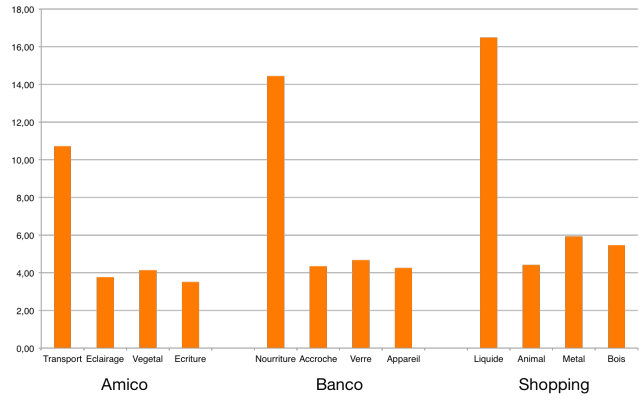

Figure 6 : Temps moyen d'association par concepts au premier essai par site (1005 connexions)

Les sujets ont là aussi rencontré des difficultés dans l'association concepts-instances mais beaucoup moins que dans les expérimentations précédentes.

Enfin, notons que nous avons observé beaucoup moins de phénomènes d'interaction entre concepts et images que dans les deux premières expérimentations du point de vue de l tâche d'assocciation. Ceci est probablement dû à la procédure collective de codage qui a permis d'éviter les erreurs de codage et de choix des instances observés dans les deux premières expérimentations.

Les sujets ont là aussi rencontré des difficultés dans l'association concepts-instances mais beaucoup moins que dans les expérimentations précédentes.

Enfin, notons que nous avons observé beaucoup moins de phénomènes d'interaction entre concepts et images que dans les deux premières expérimentations du point de vue de l tâche d'assocciation. Ceci est probablement dû à la procédure collective de codage qui a permis d'éviter les erreurs de codage et de choix des instances observés dans les deux premières expérimentations.

Authentification

Les résultats mettent en évidence un taux moyen d'authentification correcte de 94% toutes tentatives confondues. Ce résultat se décompose de la manière suivante : 81,5% de réussite au 1er essai ; 90,5% de réussite au 2ème essai ; 94% de réussite au 3ème essai. On peut constater que la réussite de l'authentification varie selon les sites (et donc le nombre de connexions) : comme on pouvait s'y attendre plus le nombre de connexions est important, plus le taux d'authentification correcte augmente (Figure 7). Ce phénomène est plus particulièrement notable si l'on s'en tient aux premiers essais.

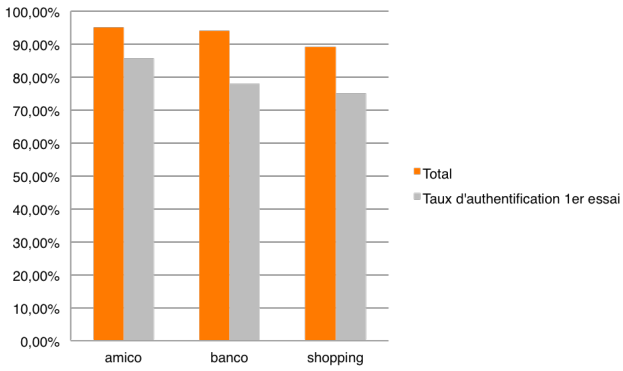

Figure 7 : Taux moyen d'authentification pour les 3 sites (1005 connexions).

Concernant le temps d'authentification par connexion, la moyenne observée est de 31,25 sec. toutes tentatives confondues (m : 23, écart-type : 18,4, min 4, max 220). Le temps moyen obtenu au premier essai est de 24,75 sec. On peut remarquer que la valeur minimum obtenue semble montrer que l'authentification peut être très rapide.On observe également que, comme pour le taux d'authentification correcte, il existe une différence sensible selon les sites, ici inversement proportionnelle au nombre de connexions réalisées (Figure 8) : le temps moyen d'authentification passe ainsi de 26 sec. pour Amico, à 37 sec. pour Banco et 40 sec. pour Shopping. Un lien entre le nombre de connexions et la performance d'authentification semble donc se confirmer.

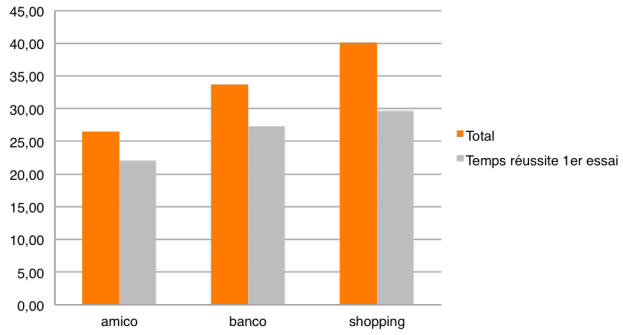

Figure 8 : Temps moyen d'authentification pour les 3 sites (en sec.)

Mémorisation

Le taux de rappel des concepts en fin de test lors des entretiens, était de 73% pour le site AMICO, 55% pour BANCO et 50% pour SHOPPING (figure 9). Ce résultat, qui doit être pris avec précaution (situation expérimentale, rappel effectué plus ou moins longtemps après la fin du test, mots de passe imposés), indique l'existence d'une relation entre taux de rappel et fréquence de connexion, l'usage régulier des mots de passe semblant favoriser leur rétention. Les difficultés de mémorisation se sont traduites par des oublis et des confusions (mots de passe ou ordre des concepts). Ce problème s'est surtout posé aux sujets qui ont fait l'effort de mémoriser les mots de passe (une part importante des sujets a eu recours de manière ponctuelle ou régulière à des supports externes).

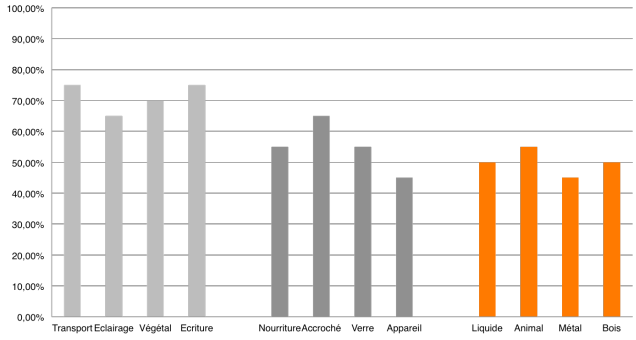

Figure 9. Taux de rappel des concepts constituant les mots de passe (expérimentation 3) . De gauche à droite sites AMICO, BANCO, SHOPPING.

De façon générale, les entretiens montrent que c'est cet aspect mémorisation qui a été vécu comme le plus problématique par les sujets.

Projection dans l'usage

Les entretiens mettent en évidence que la plupart des sujets ne s'opposeraient pas à l'idée d'utiliser HSA dans le cadre d'un usage quotidien, mais à certaines conditions : pouvoir choisir son mot de passe (9 sur 16) et n'avoir qu'un seul mot de passe (6 sur 16). Ils montrent aussi qu'une partie des participants envisagent ce mode d'authentification plutôt pour des services à risque impliquant des données personnelles (service bancaire par exemple) et un usage peu fréquent. Enfin, notons que nous avons observé beaucoup moins de phénomènes d'interaction entre concepts et images que dans les deux premières expérimentations. Ceci est probablement dû à la procédure collective de codage qui a permis d'éviter les erreurs de codage et de choix des instances observés dans les deux premières expérimentations.

DISCUSSION

Comparée aux deux premières expérimentations, la troisième expérimentation montre une nette amélioration des résultats au plan de l'utilisabilité, en particulier le taux d'association réussie pour quatre concepts au premier essai : 40% pour la 1ère, 60% pour la 2ème, 81,5 % pour la 3ème (figure 10). Ce résultat est en grande partie lié à la démarche en trois étapes qui a permis d'affiner la procédure de codage et la sélection progressive de concepts plus adaptés à la tâche.

Comparés à ceux des études antérieures sur les systèmes d'authentification graphique, les taux d'authentification obtenus à la 3ème étude sont plutôt encourageants (dans la littérature ces taux varient entre 57 et 100% [2]). Le temps moyen d'authentification (24,7 sec.) reste cependant relativement élevé par rapport à ce qui est généralement considéré comme acceptable (moins de 20 sec.) même s'il s'est amélioré par rapport à la 1ère étude. Il est à noter que dans la 3ème expérimentation, seuls 5 sujets ont estimé que l'authentification prenait trop de temps.

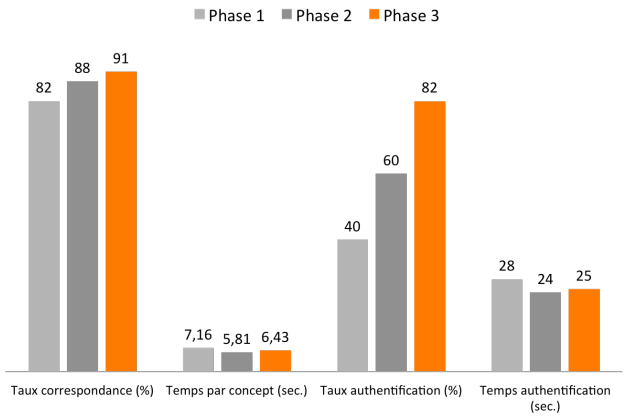

Figure 10. Synthèse des résultats pour les 3 expérimentations ; de gauche à droite : taux d'associations correctes global ; taux d'associations correctes en zone focale ; temps d'association par concept ; taux d'authentification ; temps d'authentification (les pourcentages sont arrondis à l'unité supérieure)

Si les résultats obtenus sont donc globalement plutôt encourageants, il reste que le système soulève un ensemble de difficultés qui questionnent sa viabilité du point de vue de l'utilisabilité. En l'état, il ne paraît pas adapté à l'utilisation de plusieurs mots de passe, en particulier du point de la vue de la mémorisation lorsque les mots de passe sont imposés. Outre qu'il reste à vérifier que permettre le choix du mot de passe aurait un effet positif sur la mémorisation, le risque associé en termes de fragilisation de la sécurité ne peut être écarté.

Une autre difficulté, liée à cette tension entre sécurité et utilisabilité, concerne les concepts et les images : si l'on suit les critères de sécurité, les instances ne doivent pas être facilement identifiables (éviter les objets bien délimités, limiter les surfaces vides), et les concepts ne doivent pas être trop spécifiques (par exemple « Chien »). Or, au plan de l'utilisabilité, l'étude montre que les meilleurs scores sont obtenus lorsque les instances sont bien délimitées et/ou lorsque les concepts n'exigent pas un travail inférentiel important ou ne génèrent pas de l'incertitude (contrairement, par exemple, au concept « Lourd » qui peut impliquer des opérations complexes de comparaison entre les différents objets présents dans l'image). Ceci pose la question du compromis idéal à trouver entre sécurité et utilisabilité, et de la nature de ce compromis dans le cas particulier du système HSA®. Un autre résultat notable de ces expérimentations est que l'interprétation par les utilisateurs des concepts composant le mot de passe peut être différente de celle des concepteurs, d'où l'importance de la systématisation de la procédure de codage qui a été mise en œuvre dans la 3ème expérimentation.

CONCLUSION ET PERSPECTIVES

Dans cet article nous avons présenté les résultats d'une étude en trois phases d'un nouveau système d'authentification graphique. L'objectif était d'en analyser l'utilisabilité, les implications du point de vue des mécanismes cognitifs mis en œuvre, et l'adéquation potentielle à des situations d'usage réel.

Bien qu'elle ait fournit des résultats qui permettent de nourrir la réflexion sur la conception du système et sur son intérêt potentiel, cette étude présente cependant quelques limites. Premièrement, l'analyse des conséquences en termes de mémorisation a été rendue difficile en raison du choix méthodologique réalisé. Dans la 3ème expérimentation, nous avons en effet privilégié l'aspect écologique qui s'est traduit par l'utilisation ponctuelle ou régulière par une partie des participants d'aide-mémoires alors qu'il leur était demandé par consigne d'éviter d'y faire appel. Il a donc été difficile de pouvoir disposer d'éléments quantitatifs précis sur les effets sur la mémorisation. Nous aurions pu retenir la méthode classiquement utilisée dans ce champ d'étude (en insérant des moments « distracteurs » entre sessions d'authentification), ce qui aurait nécessité de rester dans le contexte d'une expérimentation en laboratoire. Mais une telle méthode nous apparaissait trop éloignée des usages réels des mots de passe. Malgré cette limite, les résultats, notamment au plan de l'expérience vécue des participants, nous apportent des éléments significatifs sur la mémorisation.

Deuxièmement, nous n'avons pas pu explorer l'utilisabilité du système sur d'autres dispositifs de type smartphone alors que la mise en œuvre du principe de l'authentification graphique sur ces terminaux donne lieu à un nombre croissant d'études [4,11]. Ceci suggère une première perspective de poursuite de l'étude qui intégrerait l'état actuel des connaissances sur l'interaction tactile [7]. Il serait en effet intéressant de documenter les effets d'une réduction plus ou moins importante de la taille de l'écran, et donc de l'image, sur la performance et l'expérience utilisateur, par rapport à des écrans de PC ou des tablettes. Dans cette perspective, la possibilité de zoomer sur l'image pourra se révéler nécessaire. Dans le même ordre d'idée, l'adaptation du système à des utlisateurs atteints de déficience visuelle nécessitera le recours à un dispositif d'affichage spécifique permettant d'identifier des zones pertinentes de l'écran (traduction dans une modalité non-visuelle par exemple).

Une seconde perspective, liée au volet sécurité, consisterait à étudier la résistance du système aux attaques de type *shoulder surfing* (observation par-dessus l'épaule) et des attaques automatiques basées sur l'utilisation d'algorithmes de reconnaissance d'images.

Enfin, les questions relatives d'une part au processus de génération et de codage des images, et, d'autre part, au processus de recherche de concepts restent ouvertes. Dans l'état actuel des choses ces processus sont totalement manuels et restent donc lourds à mettre en œuvre. La recherche d'images prend notamment un temps considérable peu compatible avec une industrialisation du procédé. Ce point n'est pas négligeable car il conditionne la viabilité globale du système : il faut trouver un nombre important d'images pour instancier les concepts et garantir un niveau de sécurité satisfaisant (le nombre d'images nécessaires n'est pas encore complètement défini mais il pourrait avoisiner au minimum le millier). Ceci signifie qu'il faudrait trouver une solution permettant de produire rapidement des images (si possible de manière automatique). De même, la mise en œuvre du système à grande échelle nécessite d'enrichir la base de concepts (au moins une cinquantaine serait nécessaire), en respectant les critères de sécurité (nécessité de retenir des concepts présentant un niveau d'abstraction élevé) sans (trop) compromettre l'utilisabilité (concepts présentant un niveau d'abstraction limité).

REMERCIEMENTS

Cette étude a été réalisée dans le cadre d'un contrat de recherche collaborative entre Orange Labs et l'Université de technologie de Troyes.

BIBLIOGRAPHIE

1. Alt, F., Schneegass, S., Shirazi, A. S., Hassib, M. and Bulling, A. Graphical Passwords in the Wild – Understanding How Users Choose Pictures and Passwords in Image-based Authentication Schemes. 2015. In Proceedings of the MobileHCI '15.. http://dx.doi.org/10.1145/2785830.2785882
2. Biddle, R., Chiasson, S. and Oorschot, P. C. v. 2011. *Graphical Passwords: Learning from the First Twelve Years*. Technical Report TR-11-01, School of Computer Science, Carleton University.
3. Chiang, H., and Chiasson, S. 2013. Improving user authentication on mobile devices: A Touchscreen Graphical Password. *Proceedings of MobileHCI'13*, (2013). http://dx.doi.org/10.1145/2493190.2493213
4. Chowdhury, S., Poet, R., and Mackenzie, L. Passhint: Memorable and secure authentication. 2014. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'14), 2917-2926. http://dx.doi.org/10.1145/2556288.2557153.
5. Jermyn, I., Mayer, A. J., Monrose, F., Reiter, M. K. and Rubin, A. D. 1999.The Design and Analysis of Graphical Passwords. *USENIX Security Symposium.*
6. Davis, D., Monrose, F. and Reiter, M. K. 2004. On User Choice in Graphical Password Schemes. *13th USENIX Security Symposium.*
7. Motti, L. G., Vigouroux, N., & Gorce, P. 2013. Interaction techniques for older adults using touchscreen devices: a literature review *Proceedings of the 25th Conference on l'Interaction Homme-Machine,* pp. 125).
8. Norman, D. A. THE WAY I SEE IT When security gets in the way. 2009. *Interactions*, 16, 6 : 60-63.
9. Salembier, P., Zouinar, M., M., Mathias, C., Lorant, G., & Wary, J.-P. 2015. Evaluation ergonomique d'un système d'authentification graphique *Actes de la conférence EPIQUE 2015*, Aix-en-Provence, 8-10 juillet: Arpege Publishing.
10. Schaub, F., Walch, M., Könings, B., and Weber, M. 2013. Exploring the design space of graphical passwords on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS'13, 1-15.

11. Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A. and Memon, N. 2005. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63, 1:,102-127.