On the User Acceptance of Graphical Passwords

A.M. Varka and V. Katos

Information Security and Incident Response Unit, Department of Electrical and
Computer Engineering, Democritus University of Thrace, Greece
e-mail : a.varka@ihu.edu.gr; vkatos@ee.duth.gr

Abstract

In this paper we investigate options for improving the user acceptance of graphical passwords. We conducted a survey with a dual purpose. Firstly, we explored the users' reluctance to adopt graphical passwords. Secondly, we treated the graphical password authentication process as a biometric. By doing this, we proposed a distance metric to compare the user authentication response with the right answer. Although we inherited some drawbacks of biometric authentication, we established that this tradeoff in security can result in higher user acceptance and therefore can be used in contexts and environments with flexible security policies.

Keywords

Picture password, ANOVA tests.

1 Introduction and motivation

With the prevalence of smart phones and pervasive, portable computing devices in general, a wide range of user authentication technologies has been proposed, most of them relating to knowledge-based authentication, with biometrics (Clarke and Furnell, 2007; Furnell et al. 2008) being the runner up. The constraints on the user interface especially in the earlier portable devices encouraged the research in novel knowledge based authentication technologies such as graphical passwords, which encompass a wide variety of approaches (Suo et al., 2005).

Although there is some level of consensus in the literature that graphical passwords may exhibit high usability and user acceptance (Eljetlawi and Ithnin, 2008), the claim is based upon the hypothesis that users are more effective in memorizing pictures over numbers. However if we also consider the order/sequence of images against the sequence of numbers as part of the correct answer, we intuitively may agree that sequence of numbers are easier to remember than sequence of pictures. As there is no evidence to our knowledge that specifically puts this hypothesis into test, the objectives of this paper are twofold. First, we confirm by empirical means that the sequence of pictures accounts for the first level of mistakes a user may make during a user authentication process. Second, in order to improve user acceptance we propose the adoption of authentication practices inspired by biometric based approaches, and more specifically to introduce tolerance sensitivity levels to potential failures and reduce false negatives. In addition we present some side conclusions and findings relating to general behavioural aspects on graphical passwords future graphical authentication design initiatives may take into consideration.

2 Related work and methodology

A direct consequence of treating graphical authentication by biometric terms is the need to establish a distance metric between an observed (or a user provided) value and the correct data. The metric would need to maintain properties that are desirable and suitable for the underlying context. In the case of picture based authentication, we considered a number of published distance metrics, used in string and number comparison. The rationale behind this is the fact that the user's answer is encoded as a number sequence and comparison is performed on this basis. Candidate distance metrics involved the Levenshtein distance (Levenshtein, 1966) and the Jaro–Winkler distance (Winkler, 1990). Considering that the hypothesis required testing must differentiate between (a) the right answer, (b) correct identification of all pictures but in a wrong order, and (c) a wrong answer, we can see that the published distances are not suitable, as there can be distance collisions between cases (b) and (c). Consider for example the answer strings *abced* and *abcef*. If the right answer is *abcde* then for both cases the Levenshtein distance will be equal to 2. Therefore the need to construct a suitable distance capable of not only discriminating between the two cases above, but also consider (b) *closer* to the right answer arose.

The requirement for considering case (b) close to the right answer can be captured in the following proposed metric. Let *a* be the answer during the user's registration and *b* the answer provided during the user's authentication attempt. Then we define the distance from the correct answer, the nominal metric *d*(·,·) such that:

$$d(a,b)=\begin{cases} 0,\text{iff } a=b \\ 1,\text{for correct pictures, mistake in order} \\ k+1,\text{for } k \text{ mistakes} \end{cases}$$

The validation of the metric was performed with the following hypothesis:

H_{10}: Users who have a positive attitude toward graphical passwords make fewer mistakes.

More specifically, assuming that users behave rationally we intuitively expect that when exposing users to a picture based authentication test, those users who are more positively positioned toward this technology will make fewer mistakes, or alternatively, the users that make more mistakes will be in principle more frustrated by the technology.

3 Empirical work and findings

We conducted a survey consisting of a graphical/picture password simulation at log-in devices, in order to measure user's experience on graphical password usage, and a questionnaire, to assess the acceptance of graphical passwords.

Initially the participants were requested to create and register their own passwords. For that reason, a registration environment was simulated. The simulation was written in PHP programming language and was hosted on a publicly available site, for the period the survey was running. The picture options were based on Jansen's (2004) technique 'picture password 1' – individual selection. This approach results to a password space of S possible combinations where

$$S=A\cdot(M\cdot N)^X$$

and A denotes the number of different thematic categories, M,N the table (grid) dimensions, and X the number of password digits (pictures). Jansen used a password space of $A\cdot(5\cdot6)^5$. In our example, for usability reasons we made the following changes:

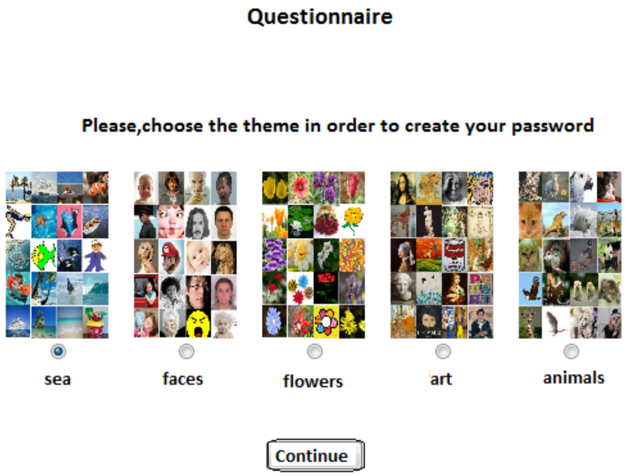
- 4x5 templates (20 pictures in total per category) were used instead of 5x6 (30 pictures in total per category)
- Only 5 thematic categories were available to the participants: sea, flowers, animals, art, faces.

This yielded a password space of 16 million passwords. All pictures were retrieved through Google, by using various key words, related to the thematic categories, such as sea, flowers, cats, dogs etc. Every picture corresponded to a number (from 1-20), so a 5-picture password was stored in the equivalent "passwords" database as a 5 digit password.

The questionnaire consisted of 21 questions and the duration of the survey was 16 days, from 10/5/2010 to 26/5/2010. The users were informed for this survey via a personal Facebook account and emails, with the exhortation to forward this survey. There was an explanatory message that accompanied the site address, containing information about text, pin and graphical passwords, the advantages and the problems of each method and a brief explanation of the procedure that users should follow. The instructions also brought to the users' attention that the order of the selection is also important. This message was sent to approximately 1300 people, from 16 to 55 years old, and about 270 valid results were collected and used in the analysis.

3.1 The survey and test

Figure 1 shows the first form of the registration phase where the users were requested to submit their thematic preference.

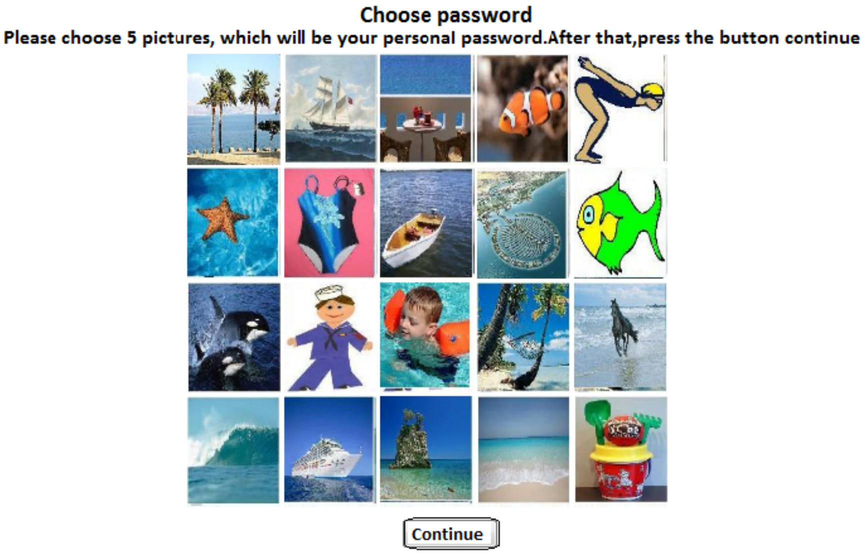


Questionnaire

Please,choose the theme in order to create your password

Figure 1: The initial registration screen

Upon submission, the registration proceeded by presenting a 4x5 template (Figure 2)and the users were asked to create their own graphical password. It was mentioned that they should pay attention to the password selection and try to memorize not only the pictures, but the sequence as well.



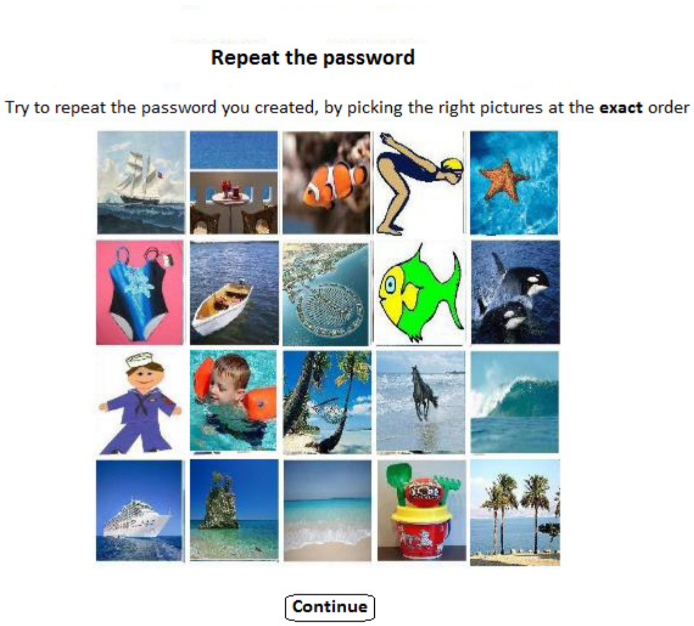
Choose password

Please choose 5 pictures, which will be your personal password.After that,press the button continue

Figure 2: The specific theme selection

Upon completion of the registration, the user was directed to the questionnaire procedure, at first with some demographic questions and then generic questions on text, pin and graphical passwords, the usage of them in their daily routine and their awareness on graphical passwords.

Approximately halfway through the procedure, the participants were requested to authenticate with their password. They were presented with the 5x4 template with the same thematic category they choose during the first step. In order to be more accurate and realistic and to avoid any shoulder surfing problems at this stage, the pictures were in a different order.



Repeat the password

Try to repeat the password you created, by picking the right pictures at the **exact** order

Irrespectively of their answer (correct or not), users continued and completed the questionnaire.

All passwords (both those created at the first step and the authentication attempts of), were stored in the database as 5 digit representation together with the questionnaire answers and were compared, correlated and analysed.

3.2 Demographics and user profile

The 270 valid questionnaires corresponded to persons of whom 57% were women and 43% men, whose majority (84%) was between 21-30 years old.

The majority of participants have 1-3 passwords (49%), and they use passwords consisted of numbers and letters (66%).

They are wary of the security of their passwords, so the majority create passwords of more than 7 digits that consist mostly of letters and characters. The users do not reveal their passwords, although they do not change them often.

In Figure 3 we show the breakdown of the users ways of memorising their passwords and their intention to use graphical passwords instead of text passwords.

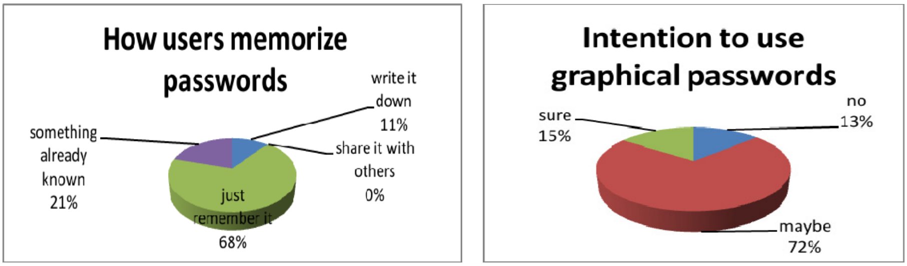


Figure 3: User practices on text passwords and intention to use graphical passwords

Lastly, Figure 4 summarizes the comparative evaluation and ranking of the three types of passwords (text, PIN, graphical). The 3 methods were evaluated and ranked in terms of aesthetics, security, time consumption, ease of memorizing and friendliness. We initially performed χ^2 homogeneity test in order to determine whether the five distributions were the same (H_0). The result was $\chi^2(4)=27.864[p=0.000]$ revealing that the distributions were not the same (hence rejecting H_0).

We also performed z-tests to identify whether there are statistically significant differences in the preferences for the attributes with close answers. For a 5% significance level ($a=0.05$), we obtained that there is no difference between the friendliness of text and graphical passwords ($z=1.07<1.96$), no difference between the security of text and graphical passwords ($z=1.85$) and no difference between graphical and text passwords with respect to time consumption. However, although the friendliness is higher for graphical passwords compared to pins ($z=2.78$), there is no differentiation between text passwords and pins ($z=1.81$) in terms of friendliness.

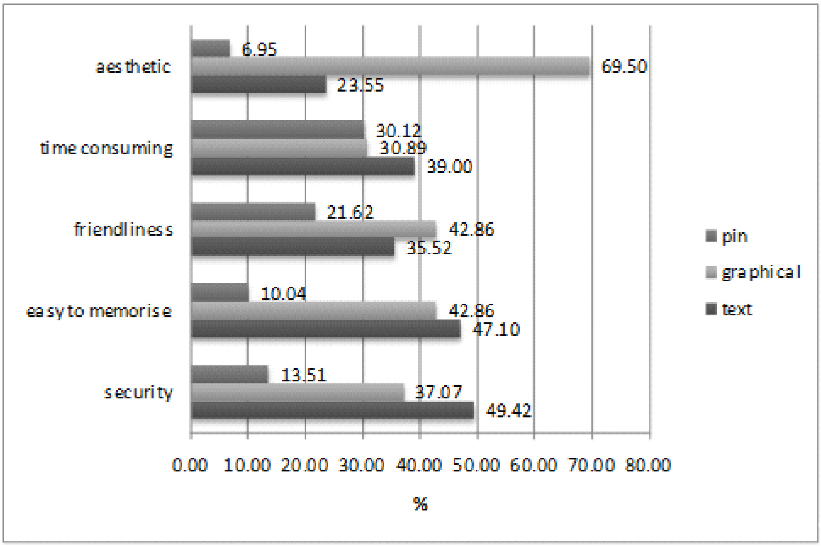


Figure 4: User evaluation of text, graphical passwords and PINs

3.3 Results

With a high degree of confidence, we see that users who had positive attitude to graphical passwords or they probably use them in the future, had also less mistakes comparing to them who had a negative attitude or were negative to future use (Table 1). Hence, the H_{10} hypothesis which we intuitively expect to hold is also statistically confirmed. We use these result as a basis for validating and consequently accepting the proposed distance metric.

Position_gp	N	Subset for alpha= 0.05	
		1	2
Positive	117	.8291	
Skeptical	135	1.1481	1.8750
Negative	16		
Sig.		**.247**	**1.000**

Use_of_gp	N	Subset for alpha= 0.05	
		1	2
Yes	40	.6000	
Maybe	187		1.1123
No	35		1.2571
Sig.		**1.000**	**.548**

Table 1: Anova grouping results for H_{10} (1=right answer(low *d*), 2=wrong answer(high *d*)).

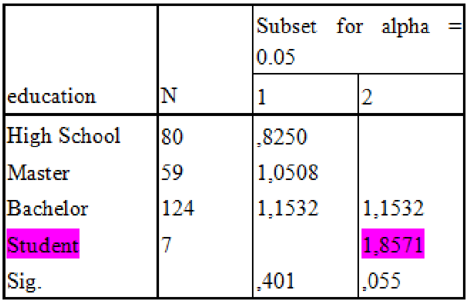
Another test reinforcing H_{10} was on the ability to remember graphical and text passwords (Table 2). Nonsurprisingly, those who find difficult or more difficult to remember graphical passwords, made more mistakes. In addition, participants who write down their passwords in order to remember them made more mistakes than the others.

remembering _gp	N	Subset for alpha= 0.05	
		1	2
Very easy	35	.4857	
Easy	92	.7717	
Neutral	84	1.3214	1.3214
Difficult	41	1.4146	1.4146
Very difficult	14		1.7143
Sig.		**.053**	**.392**

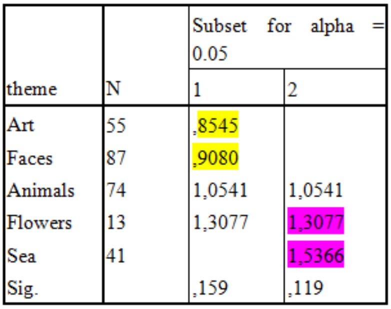
remembering	N	Subset for alpha= 0.05	
		1	2
Something known	56	.8036	
Just remember it	184	1.0380	
Write it down	30		1.6000
Sig.		**.319**	**1.000**

Table 2: Anova grouping results: ease of remembering graphical passwords and practice of remembering text passwords (1=no or less mistakes, 2=more mistakes).

Having confidence on the distance *d* meeting the representation condition, we run some further tests with the results captured in Table 3. There is differentiation on mistakes by education level. Quite surprisingly, although younger participants were expected to make fewer mistakes, in fact they made most of them. There was also significance on the differentiation of mistakes by theme. Users who choose art and faces made fewer mistakes than the others.



education	N	Subset for alpha = 0.05	
		1	2
High School	80	,8250	
Master	59	1,0508	
Bachelor	124	1,1532	1,1532
Student	7		1,8571
Sig.		,401	,055



theme	N	Subset for alpha = 0.05	
		1	2
Art	55	,8545	
Faces	87	,9080	
Animals	74	1,0541	1,0541
Flowers	13	1,3077	1,3077
Sea	41		1,5366
Sig.		,159	,119

Table 3: Anova grouping results: education and theme discrimination (1=no or less mistakes, 2=more mistakes).

4 Conclusions and future work.

We argue that user acceptance of graphical passwords can be improved if these are treated as biometric type of passwords. This is because although there is a positive attitude toward using graphic passwords, people tend to get the order of pictures wrong, rather than the actual pictures themselves. We defined a distance metric which considers the event <*right answer, wrong order*> to be closer to the <*right answer*>, whereas <*one mistake*> or higher number of mistakes are further away from the right answer. This convention was tested against user's attitude, expectations and behaviour and provided statistically significant results. As such, in applications where the authentication level can be decreased within acceptable levels, the optimum trade-off between security and user acceptance would prioritize to accept the right selection of passwords irrespective of order, as opposed to one or more mistakes. This will have a significant impact on false negatives, whereas the false positives can be filtered out or further reduced by adding a second layer of authentication, depending on the policy.

The distance metric can be further refined by increasing the granularity in the order of the pictures selection. Provided that the user responds with the right pictures but in the wrong order, distances like number of permutations needed to get to the right answer can be considered. However, these will also need to be validated to get the most suitable metric.

As for the survey results, a number of interesting conclusions were reached. The majority of participants realize the importance of having a strong password, with more than 7 digits and combining letters and numbers. Also, none of them reveal or share their passwords to others. However, they do not change them often and have the same passwords for different applications. In addition, it was clear that there is a strong correlation between the successful application of graphical passwords and users' attitude on remembering passwords. Those who found difficulties in memorizing them, made more mistakes. More mistakes were also made by the users who use ways to remember their passwords other than learning them by heart. Nevertheless, the majority of them have a positive attitude toward graphical passwords and they will probably use them in future.

For future research a better security level assessment model is considered. It was noticed that some categories and pictures were more popular than others, creating the so-called hotspots, resulting to a non-uniform distribution of graphical password selection. This means that the effective password search space can be smaller than the actual space. Research on factors leading to a distribution closer to the uniform distribution is also planned for future work.

5 References

Clarke, N., Furnell, S. (2007). "Advanced user authentication for mobile devices", *Computers & Security,* 26, pp. 109-119

Eljetlawi, A. M., Ithnin, N. (2008). "Graphical Password: Comprehensive study of the usability features of the Recognition Base Graphical Password methods". *Third 2008 International Conference on Convergence and Hybrid Information Technology*, pp. 1137-1143.

Furnell, S., Clarke, N., Karatzouni, S. (2008). "Beyond the PIN: Enhancing user authentication for mobile devices". *Computer Fraud & Security*, August, pp.12-17.

Jansen, W. (2004). "Authenticating Mobile Device Users Through Image Selection". *Data Security*.

Levenshtein, V. 1966, Binary codes capable of correcting deletions, insertions and reversals. In Soviet Physics Doklady, Vol. 10.

Suo, X., Zhu, Y., Owen, S. (2005). "Graphical Passwords: A Survey". *Proceedings of Annual Computer Security Applications Conference*, pp. 463-472.

Winkler, W. E. (1990). “String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage”. *Proceedings of the Section on Survey Research Methods*, pp. 354–359.