On the Necessity of User-Friendly CAPTCHA

Christos A. Fidas and Nikolaos M. Avouris
Interactive Technologies Lab., HCI Group
Electrical and Computer Engineering Dept.
University of Patras, GR-26504, Patras, Greece
fidas@ece.upatras.gr, avouris@upatras.gr

Artemios G. Voyiatzis
Industrial Systems Institute/RC 'Athena'
Patras Science Park, GR-26504, Patras, Greece
bogart@isi.gr

ABSTRACT
A "Completely Automated Public Turing test to tell Computers and Humans Apart" (CAPTCHA) is a mechanism widely used nowadays for protection of web applications, interfaces, and services from malicious users. A questionnaire-based survey combined with a real usage scenario of a native-language CAPTCHA mechanism was conducted in order to investigate several aspects that affect end-user perceptions related to the quality of CAPTCHA. A total of 210 participants of age between 19 and 64 participated during May and July 2010. The survey results validate the common belief that CAPTCHAs are still difficult for humans to solve. They also provide insights that can be applied to improve users' experience on interacting with CAPTCHA systems.

Author Keywords
User Perceptions, CAPTCHA, Usable Security, Usability, Security, Native Language CAPTCHAs

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces.

General Terms
Human Factors, Design, Security

INTRODUCTION
The web has become gradually a platform for deployment of complex applications of increased interactivity. Within this highly dynamic and evolving context, security issues and concerns have shifted lately to the center of attention. Indeed, the consequences of a security breach can harm the credibility and legal liability of an organization, leading to loss of users' trust. In this context, a particularly important security concern is that malicious users or software agents try to automate the misuse of system resources having as a side effect the degradation of service quality for normal users.

Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) is a type of a challenge-response test to defend against such threats. These tests try to ensure that it is a human interacting with a web system. CAPTCHA tests are commonly based on open problems of Artificial Intelligence like recognizing distorted text embedded in images. These open problems cannot yet be easily solved by computers but can be easily solved by a human [2].

Two important quality dimensions of a good CAPTCHA are its security and usability. The security level of a CAPTCHA determines its strength in resisting adversarial attacks. At the same time, a good CAPTCHA protecting a web application must also be user-friendly and transparent, aiming to minimize the added cognitive effort of a casual user interacting with the application.

Research on CAPTCHA has received lately significant attention. A lot of work focuses on the security of various CAPTCHA mechanisms, mainly on how to break and how to improve them [1,8,10,12,13]. Another direction of work focuses on evaluating widely-used implementations and proposing methods to improve their security or usability [4,5,9,14]. Lately, there is an increased interest on non-text CAPTCHA mechanisms, like audio and video [3,6,7,11]. In any case, it is acknowledged that the design of CAPTCHA mechanisms is still more an art than science [14].

The goals of this paper are to investigate the user views related to **perceptions, usage, and user preferences** related to CAPTCHA. Specifically, we investigate (i) users perceptions relating to the usability and security aspects of the CAPTCHA technology, (ii) usage statistics like frequency of solving a CAPTCHA and number of attempts needed to solve a CAPTCHA, and (iii) users experience and preferences relating to the usage of CAPTCHA consisting of characters in their native language (non-Latin alphabet) and conditions under which users would prefer it rather than a Latin-based alphabet one. To the best of our knowledge, this is among the first studies to report actual user views for CAPTCHA systems.

METHOD OF STUDY

Procedure
A survey was conducted to develop empirical evidence on the important of several quality factors relating with the usage of CAPTCHA. We formed an online questionnaire consisting of fifteen questions and statements. The

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

questions and statements were grouped in 4 categories: **Demographics** (gender; age group; and familiarity with computers, Internet, forum, blogs, and social networks), **CAPTCHA usage statistics** (importance of security and usability; frequency of solving a CAPTCHA; and number of attempts needed to solve a CAPTCHA), **Usage preferences of CAPTCHA** (fixed or random length and random characters or dictionary words), and **Perceptions related to the use of non-Latin CAPTCHA** (preference to Latin-based CAPTCHA; efficiency for protecting the system; perceived usability for the user; and comparison with Latin-based CAPTCHA) A five-point Likert scale was used to rank the level of importance from "Not Important" (1) up to "Most Important" (5).

The participants were recruited by using an invitation announced on the web site of the University and on social networking sites, by sending invitation through email at known mailing lists, and by lectures and invited talks at conferences and meetings. The aim of this selection process was to recruit a representative sample of participants of varying profiles and age, familiarized already with CAPTCHA challenges prior to the study. Given the subjective self-reporting character of the study, it was our intention to increase internal validity by involving rather experts and average users than novice users with respect to CAPTCHA challenges.

Initially, we asked the participants to visit a web page which entailed a short introduction about the CAPTCHA technology aiming to create a common ground among the participants relating to its usage. Next, we asked participants to solve a non-Latin CAPTCHA challenge (Figure 1). Then, we asked participants a set of questions related to the adequacy of utilizing a non-Latin CAPTCHA and their perceptions related to both drawbacks and benefits of using it. A limitation of the aforementioned self-reporting approach is that is actually measures perceptions of constructs rather than measuring them directly. The latter would be the case in a controlled experimental study which would lack ecological validity.

Demographics of Participants

A total of 210 people participated in the survey held between May and July 2010. Their age was between 19 and 64 years old. The distribution in age groups is depicted in Table 1. Males outnumbered females by 65.2%. Most people (N=192, f=91.4%) reported that they are not aware of having any kind of vision problem that hampered their effort to identify colors, shapes, or patterns.

	Male	Female	19-24	25-34	35-49	50-64
N	137	73	82	72	43	13
%	65.2	34.8	39.0	34.3	20.5	6.2

Table 1. Demographics of the sample.

Figure 1. A Latin-based CAPTCHA (left) and a native language CAPTCHA based on the Greek alphabet (right).

The participants consider themselves quite familiarized with computers (average rating 4.42 on a 5-point scale, 1 being "Very low" and 5 being "Very high"), Internet (4.47), forum (3.59), blogs (3.30), and social networks (3.36).

The participants are often exposed to CAPTCHA challenges: 65.7% once per week or more often; 33.8% once per day or more often (included); and 12% multiple times per day.

Hypotheses

The following null hypotheses were formulated for the purpose of this research: a) there is no significant difference in users' perception of the importance of security vs. usability relating to the usage of CAPTCHA, b) users consider CAPTCHA a hard problem that hampers usability and productivity, c) character distortion is the main obstacle on solving a CAPTCHA, and d) CAPTCHA containing non-Latin alphabet characters are acceptable by the users.

RESULTS AND DISCUSSION

We present and discuss the main findings of our study in the next. For our analysis, we separated users in two categories based on age: group O (adults) consisted of participants with age 25 or greater and group Y (teenagers) consisted of participants with age 24 or less. The first group contained mostly employees of some kind, while the second group consisted mostly of undergraduate students.

Security vs. usability

A common belief is that users will opt for usability when faced with the dilemma of usability versus security. We asked the participants: "*What is more important for you, security or usability?*" with possible answers "*Equally important*", "*Something else*", "*Security*", "*Usability*", and "*Depending on the system*". Nobody chose "*Something else*", which means that security and usability are the main concerns of the users. The majority of the participants take a neutral position by selecting "*Equally important*" (N=92) or "*Depending on the system*" (N=70). An interesting finding lies in answers "*Security*" (N=42) and "*Usability*" (N=6). Contrary to common belief, among participants selecting one of the two factors as most important (22.9% in total) the majority chooses "*Security*" (20%) over "*Usability*" (2.9%). A possible explanation could be that users come mainly from an academic background and are more aware of the various threats in the Internet. There is no significant association between age and the choice of the participants $x^2(3)=1.898$, $p>0.05$ or between gender and their choice $x^2(3)=8.065$, $p>0.05$. We seek to further

explore and understand this rather interesting position in the future.

Number of tries needed to solve a CAPTCHA

The sample consists of users who consider themselves quite familiarized with Internet technologies and that are faced often with CAPTCHA challenges. Still, only a 48.5% reports that manages to solve a CAPTCHA challenge on first try, while the rest 51.5% require two or more tries on average. The findings indicate that users are already having a great difficulty in solving CAPTCHA challenges. There is a significant association between age and number of tries, since adults participants (Group O) are more likely to solve CAPTCHA at first try while teenager ones (Group Y) need two or more tries $x^2(1)=5.039$, $p<0.05$.

Main obstacle when solving a CAPTCHA

There are multiple techniques for hardening a CAPTCHA image, with varying difficulty. We asked participants to identify the main obstacle they face when trying to solve a CAPTCHA by choosing one of "*Character distortion*", "*Font size*", "*Font shape*", "*Character coloring*", "*Background patterns*", and "*Other, explain*" (open question). The majority of the users (61.4%) identified character distortion as the main obstacle. There is no statistically significant correlation between the two age groups, or gender, or number of tries needed to solve a CAPTCHA (grouped as "one try" and "two or more tries") with choosing either a character feature or the background as the main obstacle for solving a CAPTCHA $x^2(5)=4.528$, $x^2(5)=8.486$, $x^2(5)=3.309$, $p>0.05$ respectively. This indicates that the result is consistent across all users.

It is interesting to note that there is a significant percentage of 21.4% which considers the background patterns as the main obstacle. However, the background is known to offer little or no more additional security to a CAPTCHA, since it can be easily and automatically removed by cracking software [2]. Thus, users are exposed to a feature that hampers usability and their effort on solving CAPTCHA, which on the other hand offers little if any security value to the system. This is clearly an indication of bad design and an area for improvement of future CAPTCHAs.

A native language CAPTCHA

Finally, we tried to explore user views on solving a CAPTCHA composed by non-Latin characters. We asked participants two relevant questions. The first one was: "*Would you prefer to solve Native Language CAPTCHA challenges instead of Latin-based ones?*" with possible answers (and response percentage): "*Yes, everywhere*" (19%), "*Yes, only in Native Language content sites*" (11%), "*No, I prefer only Latin*" (18.1%), and "*No preference*" (51.9%). Again, the majority of the participants take a neutral position. Yet, there is a significant percentage (aggregate 30%) that clearly prefers Native Language CAPTCHA. A possible explanation could be that users are already having great difficulties solving Latin-based CAPTCHA and feel like a CAPTCHA in their native language would be easier to solve. This view is also supported by anecdotal evidence provided in [2]. Another interesting finding is that 18.1% of the participants prefer only Latin-based CAPTCHA challenges. These participants identified themselves as expert users in solving Latin-based CAPTCHA challenges. There is no significant correlation between the two age groups, gender, or number of tries needed to solve a CAPTCHA (grouped as one try and two or more tries) with the preference on solving Native Language CAPTCHA $x^2(5)=2.670$, $x^2(3)=3.408$, $x^2(5)=1.727$, $p>0.05$ respectively. This indicates that the result is consistent across all users.

The participants were required to use a research prototype of a Native Language CAPTCHA mechanism we developed. We asked participants to provide their early rating on "*How do you compare the Native Language CAPTCHA you solved with the other (Latin) CAPTCHA you are usually faced?*" (1="*definitely more usable*" up to 5="*definitely less usable*"). The answers (N=191, mean=2.31) indicate that users are positive to this option (definitely more usable: 31%, more usable: 20.3%) with only a small aggregate fraction of 9.7% evaluating it as less and definitely less usable. There is a significant correlation between gender and rating, since male participants rate the system higher (M=2.02, SE=0.142) than female participants (M=2.47, SE=0.090) $t(195)=2.775$, $p<0.05$. There is no significant correlation between the two age groups (M=2.19, SE=0.116) and (M=2.39, SE=0.104) respectively $t(195)=-1.234$, $p>0.05$.

In order to correlate with the previous question, we form two groups. The first group, A, consists of those participants that prefer Native Language CAPTCHA, either in every occasion or only on sites with content in their native language. The second group, B, consists of those participants that clearly prefer the Latin-based CAPTCHA. There is also a significant correlation between clearly preferring a Native Language CAPTCHA and grading positively our research prototype, since group A participants rate positively our research prototype (M=1.78, SE=0.126), while group B participants rate it more neutral (M=3.22, SE=0.183) $t(94)=-6.671$, $p<0.05$. The findings of the two last questions indicate that users consider a CAPTCHA composed of native language characters as a viable and probably usable alternative to a Latin-based one.

CONCLUSIONS AND FUTURE WORK

The findings of our study validate the common belief that CAPTCHAs are still difficult for humans to solve. We emphasize that the actual purpose of CAPTCHA existence is to protect the system and provider resources, like bandwidth and storage, or even its brand name, rather than to contribute to a positive user experience on interacting with the system.

We note that from a technology point of view, there are solutions albeit harder to implement that can achieve the same level of system resource protection. For example, rate limiting techniques (e.g., limiting the number of connections or queries per second or per IP address or both) can protect both a system and its users while not obligating the users to solve CAPTCHA challenges.

A second conclusion concerns the usability of the CAPTCHA mechanisms. Every other participant in our study needs on average two or more tries to solve one CAPTCHA challenge. This is even more worrying since study participants are challenged with CAPTCHA multiple times per week and consider themselves quite familiar with blogs and forums. These are the places where usually CAPTCHAs are installed. A user comment was: "*[...] if my post is not that important, I give up trying to post my [blog] comments if I don't solve CAPTCHA on first try*". This signifies that the usability problems caused by CAPTCHA can be deeper than a temporal user distraction and should be further examined.

A third conclusion relates to the background patterns found in CAPTCHA. It is known that these patterns offer little or no additional security to a CAPTCHA mechanism. Yet, one out of five participants ranks the background patterns as the main obstacle for solving a CAPTCHA. Clearly, this is an area of improvement and future CAPTCHA designs should not insert such patterns.

Our research underpins that the use of a native language CAPTCHA is a viable alternative, especially for users who are not proficient with English and the Latin alphabet. Also, in cases when they are using a web application in their native language. This is further supported by the fact that from a web design point of view, it is preferable to have a uniform presentation, including the actual site content and the CAPTCHA challenge. The native language CAPTCHA used in this study was based on the Greek alphabet and we plan to further explore and compare these findings with other native language mechanisms and participants aiming to increase external validity of the study.

Finally, from a usability point of view, our findings raise the question whether CAPTCHA mechanisms are still an acceptable trade-off solution between security and usability or whether we should work, in the short term, towards more user friendly CAPTCHAs and, in the long term, towards more sophisticated solutions.

REFERENCES

1. A. El Ahmad, J. Yan, and L. Marshall. The Robustness of a New CAPTCHA. In *Proc. of the Third European Workshop on System Security (EUROSEC '10)*. ACM, New York, NY, USA, 36-41.
2. L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. *Commun. ACM* 47, 2 (February 2004), 56-60.
3. S. Bohr, A. Shome, and J. Simon. Improving auditory CAPTCHA security. ISR Tech. rep., University of Maryland, College Park, MD 2008.
4. R. Chow, P. Gollé, M. Jakobsson, X. Wang, and L. Wang. Making CAPTCHAs clickable. In *Proc. of the 9th workshop on Mobile computing systems and applications* (HotMobile '08). ACM, New York, NY, USA, 91-94.
5. C.J. Hernandez-Castro and A. Ribagorda. Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study. *Computers & Security*, 29, 1 (2010), 141–157.
6. J. Holman, J. Lazar, J. Feng, and J. d'Arcy. Developing Usable CAPTCHAs for Blind Users. In *Proc. of the 9th SIGACCESS conference on Computers and Accessibility* (Assets '07). ACM, New York, NY, USA, 245-246.
7. K.A. Kluever and R. Zanibbi. Balancing usability and security in a video CAPTCHA. In *Proc. of the 5th Symposium on Usable Privacy and Security* (SOUPS '09). ACM, New York, NY, USA, Article 14 , 11 pages.
8. S. Li, S.A.H. Shah, M.A.U. Khan, S.A. Khayam, A.-R. Sadeghi, and R. Schmitz. Breaking e-Banking CAPTCHAs. In *Proc. of 26th Annual Computer Security Applications Conference (ACSAC 2010)*, Austin, Texas, USA, December 6-10, 2010, 171-180.
9. M. Motoyama, K. Levchenko, C. Kanich, D. McCoy, G.M. Voelker, and S. Savage. Re: CAPTCHAs-Understanding CAPTCHA-solving services in an economic context. In *Proc. of the 19th USENIX Security Symposium*, USENIX Association, 435-462.
10. B. Pinkas and T. Sander. Securing Passwords Against Dictionary Attacks. In *Proc. of the 9th ACM conference on Computer and communications security* (CCS '02), ACM, New York, NY, USA, 161-170.
11. J. Tam, J. Simsa, S. Hyde, and L. Von Ahn. Breaking Audio CAPTCHAs. In *Proc. of the 22nd Annual Conference on Neural Information Processing Systems 2008*, MIT Press, 1625-1632.
12. J. Yan and A.S. El Ahmad. Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms. In *Proc. of. ACSAC 2007.* IEEE Computer Society, 279-291.
13. J. Yan and A.S. El Ahmad. The Robustness of CAPTCHAs: A Security Engineering Perspective. Newcastle Univ. Tech. rep. CS-TR-1180, Nov. 2009.
14. J. Yan and A.S. El Ahmad. Usability of CAPTCHAs or usability issues in CAPTCHA design. In Proc. of SOUPS '08. ACM, New York, NY, USA, 44-52.