[Position Paper] Motivating the Need for Evaluation Criteria for Captchas

Gerardo Reynaga
School of Computer Science
Carleton University
Ottawa, Canada
gerardor@scs.carleton.ca

ABSTRACT

We argue that a set of usability heuristics are needed for easy and quick evaluation of Captchas implementations. With this set of heuristics we contribute to sustain the Captcha Mantra: "Easy for humans, hard for machines". In particular, the usability of Captcha schemes change radically when utilized on mobile environments. We are developing a set of heuristics for use by practitioners wishing to evaluate which Captcha scheme is most appropriate for their website.

1. POSITION

We argue that a set of usability heuristics are needed for easy and quick evaluation of Captcha implementations. A Captcha is a program that generates and grades challenges that are human solvable, and should be unsolvable by current computer programs [19, 21]. They are typically used on websites to deter automatic programs (*e.g.,* bots) from abusing web applications, to prevent of e-mail harvesting, to avoid automated voting in Internet polls, and other applications that may require online automatic human verification [18, 3, 17, 1, 9, 15]. A *challenge* refers to a single Captcha puzzle to be solved by the user.

Attacks on Captcha schemes and new proposals are frequent and common [20, 13, 6, 4, 5, 11]. Diligent site administrators may want to update their Captcha challenges based on news of such attacks, but it can be difficult to choose an appropriate replacement. Our goal is to provide an evaluation methodology to help administrators make such decisions. However, heuristic evaluation cannot quantify the security of Captcha schemes. Therefore, an acceptable level of security of a Captcha scheme has to be evaluated as part of the overall decision process.

Small changes to Captcha schemes may not cause obvious problems, nevertheless these changes may affect their overall usability in ways that are unexpected. In particular, changes may be acceptable for desktop or laptop usage, but may cause difficulties for other modalities such as smartphones. With the increase use of mobile devices [12], website designers must consider the usability impact of design choices in this growing segment of users.

The strength of heuristic evaluation is that it is cheap, quick and easy to carry out. It does not require a user study with a large number of users. A few people, knowledgeable in both the domain area and interaction design, are recruited to conduct the evaluation and no special facilities are needed for the heuristic evaluation.

While existing heuristics, such as Nielsen's [16], Jaferian's [14], and Zhou's [22], provide a good set of heuristics, these are insufficient to evaluate Captchas. Nielsen's heuristics are too general, and in a mobile environment they may find more cosmetic problems rather than critical problems. In fact, Nielsen suggests developing domain-specific heuristics that apply to a specific category of products. Jaferian's and Zhou's, although developed for the security domain, evaluate security management tools and intrusion detection systems, respectively. As opposed to these software programs, Captchas are used in a variety of environments, including mobile, causing a set of problems that are substantially different. The usability work done for Captchas focuses mainly on challenge design betterment, or design of innovative scheme proposals other than text-based Captchas [7, 8, 10, 21, 2].

We are developing a set of domain-specific heuristics for evaluating Captcha schemes. The main goal of the proposed heuristics is to assess Captcha scheme deployment targeting smartphones. The proposed heuristics cover the usability and deployability of Captcha schemes. For example, usability heuristics may include *Input mechanisms* and *Solvability*. Deployability includes *Consistency with user's localization and environment*. Usability heuristics evaluate issues such as challenge obstruction, typing, restricted screen space. Deployability deals with language, culture and universality.

We have started an evaluation of Captchas on mobile devices using our proposed set of heuristics with evaluators having expertise from the usability and security. In addition, we are running a small user study, not involving the heuristics, to compare results between the usability study and the expert evaluation.

2. ACKNOWLEDGMENTS

I am working on this research in collaboration with my advisors Sonia Chiasson and Paul van Oorschot. The author acknowledges the NSERC Internetworked Systems Security Network (ISSNet) funding.

Workshop on Usable Privacy & Security for Mobile Devices (U-PriSM) 2012, July 11-13, 2012, Washington, DC, USA.

3. REFERENCES

[1] M. Alsaleh, M. Mannan, and P. C. van Oorschot. Revisiting defenses against large-scale online password guessing attacks. *IEEE Trans. Dependable Sec. Comput.*, 9(1):128–141, 2012.

[2] H. S. Baird and J. L. Bentley. Implicit CAPTCHAs. In E. H. B. Smith and K. Taghva, editors, *Document Recognition and Retrieval XII, 16-20 January 2005, San Jose, California, USA, Proceedings*, volume 5676 of *SPIE Proceedings*, pages 191–196. SPIE, 2005.

[3] A. Basso and F. Bergadano. Anti-bot strategies based on human interactive proofs. In P. P. Stavroulakis and M. Stamp, editors, *Handbook of Information and Communication Security*, pages 273–291. Springer, 2010.

[4] J. Bau, E. Bursztein, D. Gupta, and J. C. Mitchell. State of the art: Automated black-box web application vulnerability testing. In *IEEE Symposium on Security and Privacy*, pages 332–345. IEEE Computer Society, 2010.

[5] E. Bursztein, R. Beauxis, H. Paskov, D. Perito, C. Fabry, and J. Mitchell. The failure of noise-based non-continuous audio captchas. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 19 –31, may 2011.

[6] E. Bursztein and S. Bethard. Decaptcha: breaking 75% of ebay audio CAPTCHAs. In *Proceedings of the 3rd USENIX conference on Offensive technologies*, WOOT'09, Berkeley, CA, USA, 2009. USENIX Association.

[7] K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski. Building segmentation based human-friendly human interaction proofs (HIPs). In H. Baird and D. Lopresti, editors, *Human Interactive Proofs*, volume 3517 of *Lecture Notes in Computer Science*, pages 173–185. Springer Berlin / Heidelberg, 2005.

[8] K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski. Designing human friendly human interaction proofs (HIPs). In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 711–720, New York, NY, USA, 2005. ACM.

[9] M. Dailey and C. Namprempre. A text graphics character CAPTCHA for password authentication. In *TENCON 2004. 2004 IEEE Region 10 Conference*, volume B, pages 45 – 48 Vol. 2, nov. 2004.

[10] A. El Ahmad, J. Yan, and W. Ng. Captcha design: colour, usability and security. *Internet Computing, IEEE*, PP(99):1, 2011.

[11] A. S. El Ahmad, J. Yan, and M. Tayara. The robustness of Google CAPTCHAs. Technical report, School of Computer Science, Newcastle University, UK, May 2011.

[12] R. Gartner. Mobile devices grew 5.6 percent in third quarter of 2011; Smartphone sales increased 42 percent. Available from `http://www.gartner.com/it/page.jsp?id=1848514`, 2011. Last visited: May 2012.

[13] P. Golle. Machine learning attacks against the Asirra CAPTCHA. In *Proceedings of the 15th ACM conference on Computer and communications security*, CCS '08, pages 535–542, New York, NY, USA, 2008. ACM.

[14] P. Jaferian, K. Hawkey, A. Sotirakopoulos, and K. Beznosov. Heuristics for evaluating IT security management tools. In D. S. Tan, S. Amershi, B. Begole, W. A. Kellogg, and M. Tungare, editors, *CHI Extended Abstracts*, pages 1633–1638. ACM, 2011.

[15] S. Li, S. A. H. Shah, M. A. U. Khan, S. A. Khayam, A.-R. Sadeghi, and R. Schmitz. Breaking e-banking captchas. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 171–180, New York, NY, USA, 2010. ACM.

[16] J. Nielsen. Ten usability heuristics. Available from `http://www.useit.com/papers/heuristic/heuristic_evaluation.html`, 2012. Last visited: May 2012.

[17] B. Pinkas and T. Sander. Securing passwords against dictionary attacks. In *Proceedings of the 9th ACM conference on Computer and communications security*, CCS '02, pages 161–170, New York, NY, USA, 2002. ACM.

[18] L. von Ahn, M. Blum, N. Hopper, and J. Langford. CAPTCHA: Using hard AI problems for security. In E. Biham, editor, *Advances in Cryptology EUROCRYPT 2003*, volume 2656 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2003.

[19] L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. *Commun. ACM*, 47:56–60, February 2004.

[20] Y. Xu, G. Reynaga, S. Chiasson, J.-M. Frahm, F. Monrose, and P. C. Van Oorschot. Security and usability challenges of moving-object CAPTCHAs: Decoding codewords in motion. In *Proceedings of the 21st USENIX Security Symposium*, Berkeley, CA, USA, 2012. USENIX Association.

[21] J. Yan and A. S. El Ahmad. Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Proceedings of the 4th Symposium on Usable Privacy and Security*, SOUPS '08, pages 44–52, New York, NY, USA, 2008. ACM.

[22] A. Zhou, J. Blustein, and N. Zincir-Heywood. Improving intrusion detection systems through heuristic evaluation. In *Electrical and Computer Engineering, 2004. Canadian Conference on*, volume 3, pages 1641 – 1644, may 2004.