Accessible Voice CAPTCHAs for Internet Telephony

Anu Markkola
Helsinki University of Technology
P.O. Box 5400
FI-02015 TKK, Finland
anu.markkola@iki.fi

Janne Lindqvist
Helsinki University of Technology
P.O. Box 5400
FI-02015 TKK, Finland
janne.lindqvist@tml.hut.fi

ABSTRACT

CAPTCHAs have become a pervasive method for protecting against automated submissions to web forums and registration to web based email services. The CAPTCHAs are usually image-based, but voice CAPTCHAs have also emerged as an alternative. In this short note, we discuss our ongoing efforts on designing accessible voice CAPTCHAs for Internet Telephony. We have implemented a testbed for Skype to assess the usability of the approach, and conducted preliminary usability tests with 10 users.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: Information Interfaces and Presentation—*User Interfaces*; K.4.1 [**Computers and Society**]: Public Policy Issues—*Abuse and crime involving computers*

General Terms

Design, Human Factors, Security

Keywords

CAPTCHA, Internet Telephony, accessibility

1. INTRODUCTION

Image-based CAPTCHAs are a common way to prevent undesirable behavior in Web based forums and Web emails. Usually, a CAPTCHA requires the user to interpret a word from a distorted image, and type it to the web form. This method reduces the possibility of automated web email account registrations and spamming of web forums. Unfortunately, the method is not accessible for users with eyesight disabilities. As a new alternative, a voice CAPTCHA can be presented to the user.

In this short note, we present our ongoing work on accessible voice CAPTCHAs for Internet Telephony. The work is motivated by emergent open Internet Telephony services, where users could be reachable anywhere with VoIP, by contrast to closed systems such as Skype. However, even though the motivation for voice CAPTCHAs is in open systems, we implemented our approach for Skype since it is familiar to users worldwide. We argue that even though accessible voice CAPTCHAs in general require careful design, the setting of Internet Telephony makes it even harder compared to a web based voice CAPTCHA.

Symposium on Accessible Privacy and Security (SOAPS) 2008, July 23, 2008, Pittsburgh, PA USA

2. VOICE CAPTCHAS FOR WEB AND INTERNET TELEPHONY

On the web, the voice CAPTCHAs are usually presented as an alternative for image-based CAPTCHAs. These have been adopted by services such as Google Mail, Microsoft Live and LinkedIn, among others. Instead that the users need to figure out the distorted text in an image, they can listen to the text pronounced e.g. letter-by-letter.

With Internet Telephony, the situation where CAPTCHAs are presented and solved is fundamentally different. Even though the user might be using a desktop computer for calls, the CAPTCHAs are presented in real-time, when the user is actively trying to reach someone. Further, the calling device might be a mobile phone, a PDA, in addition to the desktop computer. Thus, the only input device the user might have, is the common telephony keypad, consisting of numbers from 0 to 9 and the signs $*$ and #. Thus, the CAPTCHAs need to be designed to support only the most basic input device available, the numeric keypad. Alternatively, voice could be used as input, however, voice recognition software can significantly increase the cost and complexity of the system.

One interesting point is that with telephony based voice CAPTCHAs, we cannot assume any auxiliary interfaces for presenting information about the CAPTCHA. Everything we need to inform the user about the CAPTCHA needs to be told during the call setup. Thus, we have an intrinsic additional delay (and potential pitfall for accessibility) for the call, in addition to the time needed for solving the CAPTCHA.

3. SKYPE IMPLEMENTATION

We implemented the voice CAPTCHA mechanism as a Skype plugin. The motivation for a Skype implementation was that there are many users familiar with Skype, and we can reduce the effect of unfamiliarity to VoIP in usability tests. Further, we are interested in deploying the approach in real use, and Skype is the predominant VoIP service. Even though Skype is closed, and has strong central authen-

tication, there have been reports on spam in Skype, too. We also believe that some users might be interested just to try out the approach for fun. Since Skype can be used with mobile phones and handheld devices, we could conveniently also test a scenario where the user has only a keypad as the input device.

So far, we have implemented a simple version of a voice CAPTCHA. When a user calls a protected user, the caller is redirected to the CAPTCHA service. The CAPTCHA service presents information for the user how to proceed and presents the CAPTCHA by saying 5 random digits. Although our implementation of the CAPTCHA is clearly not secure enough for wide adoption, we believe it is sufficient enough to gain insight on further steps towards accessible and secure voice CAPTCHAs for Internet Telephony.

The implementation follows the architectural design principles outlined and published before by the second author [11]. One of the key principles is that an unknown caller should be bothered only once with a CAPTCHA. After a CAPTCHA has been solved, the user is registered as a known caller in the system, and can make further calls without solving CAPTCHAs.

4. RELATED WORK

The inaccessibility of CAPTCHA on the web is a well-known problem [12]. There is a body of work that have looked into the usability of image-based CAPTCHAs [1, 2, 3, 5, 6, 7, 8, 13, 15, 16]. On voice CAPTCHAs, there has been work on quantifying how background noise affects the processing of synthesized speech between humans and computers [4, 10], and how voice CAPTCHAs can be used on the web [9, 14]. However, to the best knowledge of the authors, there is not work available on developing accessible voice CAPTCHAs for Internet Telephony.

5. CONCLUSIONS

We have outlined some problems that are intrinsic for voice CAPTCHAs in Internet Telephony. Our preliminary usability tests confirmed the above issues presented. At first, the users were confused what is actually happening, when they were presented a CAPTCHA. Second, when users were more familiar with the concept, they started to get annoyed of the time that is needed to listen to all the information. Interestingly, some users were annoyed by the fact that they did not understand why the CAPTCHA was presented to them on the first place. When explained, all of the users agreed that if spam was a similar problem in VoIP as it is today in email, they would adopt the system to use, although some questioned the security of the implemented CAPTCHA. The important point was that the CAPTCHA would be presented only once during the first connect, if successfully solved. Further work includes designing secure CAPTCHAs keeping in mind the underlying limitations, and further usability tests for assessing the accessibility of the approach.

6. REFERENCES

[1] Baird, H., and Bentley, J. Implicit CAPTCHAs. *Proc. SPIE 5676* (2005), 191–196.

[2] Baird, H., Moll, M., and Wang, S. A Highly Legible CAPTCHA That Resists Segmentation Attacks. *Human Interactive Proofs: Second International Workshop, HIP 2005, Bethlehem, PA, USA, May 19-20, 2005: Proceedings* (2005).

[3] Baird, H., Moll, M., and Wang, S. ScatterType: A Legible but Hard-to-Segment CAPTCHA. *Proceedings of the Eighth International Conference on Document Analysis and Recognition* (2005), 935–939.

[4] Chan, T.-Y. Using a text-to-speech synthesizer to generate a reverse turing test. *Proceedings of 15th IEEE International Conference on Tools with Artificial Intelligence, 2003* (Nov. 2003), 226–232.

[5] Chellapilla, K., Larson, K., Simard, P., and Czerwinski, M. Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs). *Conference on Email and Anti-Spam* (2005).

[6] Chellapilla, K., Larson, K., Simard, P., and Czerwinski, M. Designing human friendly human interaction proofs (HIPs). *Conference on Human Factors in Computing Systems* (2005), 711–720.

[7] Chew, M., and Baird, H. BaffleText: a Human Interactive Proof. *Proc., 10th IS&T/SPIE Document Recognition & Retrieval Conf* (2003).

[8] Elson, J., Douceur, J. R., Howell, J., and Saul, J. H. J. Asirra: a captcha that exploits interest-aligned manual image categorization. In *CCS '07: Proceedings of the 14th ACM conference on Computer and communications security* (New York, NY, USA, 2007), ACM, pp. 366–374.

[9] Holman, J., Lazar, J., Feng, J., and D'Arcy, J. Developing usable CAPTCHAs for blind users. *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility* (2007), 245–246.

[10] Kochanski, G., Lopresti, D., and Shih, C. A Reverse Turing Test Using Speech. *Seventh International Conference on Spoken Language Processing* (2002).

[11] Lindqvist, J., and Komu, M. Cure for Spam Over Internet Telephony. *4th IEEE Consumer Communications and Networking Conference* (Jan. 2007), 896–900.

[12] May, M. Inaccessibility of CAPTCHA. Alternatives to Visual Turing Tests on the Web. Web page. URL: http://www.w3.org/TR/turingtest/.

[13] Rui, Y., and Liu, Z. ARTiFACIAL: Automated Reverse Turing test using FACIAL features. *Multimedia Systems 9*, 6 (2004), 493–502.

[14] Schlaikjer, A. A Dual-Use Speech CAPTCHA: Aiding Visually Impaired Web Users while Providing Transcriptions of Audio Streams. CMU-LTI-07-014, CMU, Nov. 2007.

[15] Shirali-Shahreza, M., and Shirali-Shahreza, S. Online Collage CAPTCHA. *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS'07. Eighth International Workshop on* (2007), 58–58.

[16] Wang, S., and Bentley, J. CAPTCHA Challenge Tradeoffs: Familiarity of Strings versus Degradation of Images. *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)-Volume 03* (2006), 164–167.