available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony

Yannis Soupionis[*], *Dimitris Gritzalis*

Information Security and Critical Infrastructure Protection Research Group, Dept. of Informatics, Athens University of Economics & Business (AUEB), 76 Patission Ave., Athens GR-10434, Greece

ARTICLE INFO

Article history:
Received 13 September 2009
Received in revised form
4 December 2009
Accepted 7 December 2009

Keywords:
SPIT
Audio CAPTCHA attributes
VoIP
Authentication
Evaluation
Speech Recognition
Turing Test

ABSTRACT

SPam over Internet Telephony (SPIT) is a potential source of future annoyance in Voice over IP (VoIP) systems. A typical way to launch a SPIT attack is the use of an automated procedure (i.e., bot), which generates calls and produces unsolicited audio messages. A known way to protect against SPAM is a Reverse Turing Test, called CAPTCHA (Completely Automated Public Turing Test to Tell Computer and Humans Apart). In this paper, we evaluate existing audio CAPTCHA, as this type of format is more suitable for VoIP systems, to help them fight bots. To do so, we first suggest specific attributes-requirements that an audio CAPTCHA should meet in order to be effective. Then, we evaluate this set of popular audio CAPTCHA, and demonstrate that there is no existing implementation suitable enough for VoIP environments. Next, we develop and implement a new audio CAPTCHA, which is suitable for SIP-based VoIP telephony. Finally, the new CAPTCHA is tested against users and bots and demonstrated to be efficient.

1. Introduction

With the rapid worldwide growth of VoIP services, the spam issue in VoIP systems becomes increasingly important (Rosenberg et al., 2006), which is the reason why important companies, like NEC and Microsoft, have already developed mechanisms (Quittek et al., 2007; Graham-Rowe, 2006) to tackle SPam over Internet Telephony (SPIT). A serious obstacle when trying to prevent SPIT is identifying VoIP communications, which originate from software robots (''bots''). Alan Turing's ''Turing Test'' paper (Turing, 1950) discusses the special case of a human tester who wishes to distinguish humans from computer programs. Nowadays, there has been a considerable interest in applying an alternate form of the Turing Test, the so called Reverse Turing Test. The term ''Reverse Turing Test'' is used to describe that the tester is not a human but a machine. In the spam protection world this kind of computer administrated Reverse Turing Test is also called CAPTCHA (Completely Automated Public Turing Test to Tell Computer and Humans Apart). The research interest in this subject has spurred a number of relevant proposals (Blum et al., 2000; von Ahn et al., 2003, 2004; Chellapilla et al., 2005; Yan and El Ahamad, 2009). Commercial examples include major stakeholders in the field, such as Google and MSN, which require CAPTCHA (visual or audio), in order to provide services to users. However, there exist computer programs, which can break the CAPTCHA that have been proposed so far.

In this paper, an audio CAPTCHA was developed that is suitable for use in VoIP systems. In specific, first we present the background and related work and explain the main aspects of SPIT and CAPTCHA. Then, we provide the basic requirements of a CAPTCHA, briefly explain why an audio

[*] *Corresponding author.*
E-mail addresses: jsoup@aueb.gr (Y. Soupionis), dgrit@aueb.gr (D. Gritzalis).

0167-4048/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2009.12.003

CAPTCHA is suitable for VoIP systems, and present an algorithm for selecting a suitable CAPTCHA. In Section 3, a classification of the characteristics/attributes of audio CAPTCHA is proposed. In Section 4 a number of popular CAPTCHA is introduced. In Section 5, the procedure to be followed for testing a CAPTCHA is described; this includes a bot and a speech recognition tool. In Section 6 we demonstrate that the existing audio CAPTCHA implementations are not adequate enough for a VoIP system. In Section 7, the experimental environment which was used for testing the proposed CAPTCHA is presented. The VoIP experimental environment was based on the Session Initiation Protocol (SIP), because one of the most known and deployed multimedia protocols for VoIP infrastructures. In Section 8, the new audio CAPTCHA is presented, which is based on the attributes selected in Section 3. Finally, we provide the reader with the results of the tests performed with the proposed CAPTCHA.

2. Background

SPIT constitutes an emerging type of threat in VoIP systems. It illustrates several similarities to email spam. Both spammers and ‘‘spitters’’ use the Internet, so as to target a group of users and initiate bulk and unsolicited messages and calls. Compared to traditional telephony, IP telephony provides a more effective channel, since messages are sent in bulk and at a low cost. Individuals can use spam-bots to harvest VoIP addresses. Furthermore, since call-route tracing over IP is harder, the potential for fraud is considerably greater.

A CAPTCHA is a method that is widely used to uphold automated SPAM attacks. The same technique can be used to mitigate SPIT. According to this, each time a callee receives a call from an unknown caller, an automated Reverse Turing Test would be triggered. The ‘‘spit-bot’’ needs to solve this test in order to complete its attack. Integrating such a technique into a VoIP system raises two main issues. First, the CAPTCHA module should be combined with other anti-SPIT controls, i.e., not every call should pass through the CAPTCHA challenge, since each CAPTCHA requires considerable computational resources. A simultaneous triggering of several CAPTCHA challenges can soon lead to denial of service. Challenges would also cause annoyance to users, if they had to solve one CAPTCHA for every call they make. Second, a CAPTCHA needs to be friendly and easy to solve (‘‘pass’’) for a human user.

2.1. *CAPTCHA*

A CAPTCHA is a test that most humans should be able to pass, but computer programs should not. Such a test is often based on hard open AI problems, e.g., automatic recognition of distorted text, or of human speech against a noisy background. Differing from the original Turing Test, CAPTCHA challenges are automatically generated and graded by a computer. Since only humans are able to return a sensible response, an automated Turing Test embedded in a protocol can verify whether there is a human or a bot behind the challenged computer. Although the original Turing Test was designed as a measure of progress for AI, CAPTCHA is rather a human-nature-authentication mechanism.

This paper is focused on audio CAPTCHA. These were initially created to enable people that are visually impaired to register or make use of a service that requires solving a CAPTCHA. Today, an audio CAPTCHA would be useful to defend against automated audio VoIP messages, as visual CAPTCHA are hard to apply in VoIP systems, mainly due to the limitations of end-user devices. For example, nowadays not many people have a home telephony device with a screen capable of displaying a proper (high resolution) image CAPTCHA. If an adequate CAPTCHA is used, it should be hard for a spit-bot to respond correctly and thus manage to initiate a call. Also, audio CAPTCHA seems attractive, as text-based CAPTCHA has been demonstrated breakable (Chew and Baird, 2003; Mori and Malik, 2003; Defeated CAPTCHA; Yan and El Ahmad, 2007; Yan and El Ahamad, 2008).

2.2. *Related work*

As the audio CAPTCHA technology is practically in its infancy, the relevant research work is currently limited.

Bigham and Cavender demonstrated that existing audio CAPTCHA are clearly more difficult and time-consuming to complete as compared to visual CAPTCHA (Bigham and Cavender, 2009). They created a comparison between the existing CAPTCHA implementations, but they do not reach to any conclusion on how their characteristics affect the user success rate. They developed and evaluated an optimized interface for non-visual use, which can be added in-place to an existing audio CAPTCHA. In their published CAPTCHA evaluation they mentioned that Facebook, Veoh, and Craigslist use different CAPTCHA; today, all three of them use Recaptcha (Recaptcha Audio CAPTCHA).

Tam et al. (2008a,b) described a number of security tests of audio CAPTCHA. The authors used machine learning techniques, which are similar to the ones used for breaking visual CAPTCHA. They analyzed three audio CAPTCHA taken from popular websites (Google (Google Audio CAPTCHA), Recaptcha (Recaptcha Audio CAPTCHA), Digg (DIGG)). In some cases they reached correct solutions with an accuracy of up to 71%. The main issue with this work is that they only tested the audio CAPTCHA implementations and did not analyze what is the impact of audio CAPTCHA characteristics on its performance.

Yan and El Ahmad (2008) worked on the usability issues that should be taken into consideration when developing a CAPTCHA. Their work does not specifically focus on audio CAPTCHA, with the exception of a few characteristics (i.e., character set). Their work was concluded with a framework referring to CAPTCHA usability.

Bursztein and Bethard (2009) developed a prototype audio CAPTCHA decoder, called *decaptcha*, which is able to successfully break 75% of the eBay audio CAPTCHA. They described an automated process for downloading audio CAPTCHA, training the decaptcha bot and finally solving the eBay CAPTCHA.

Finally, Markkola and Lindqvist (2008) proposed a number of ‘‘voice’’ CAPTCHA for Internet telephony. However, they did not explain in detail how this could be integrated into an Internet telephony infrastructure. Also, their work lacks experimentation results.