Chapter 10
Distinguishing Human Users from Bots: Methods and Assessments

M. Hassan Shirali-Shahreza
Amirkabir University of Technology, Iran

Sajad Shirali-Shahreza
University of Toronto, Canada

ABSTRACT

Human Interactive Proof (HIP) systems have been introduced to distinguish between various groups of users. CAPTCHA methods are one of the important branches of HIP systems, which are used to distinguish between human users and computer programs automatically and block automated computer programs from abusing Web services. The goal of these systems is to ask questions, which human users can easily answer but current computer programs cannot. In this chapter, the authors collect different pioneering works, which are done on CAPTCHA systems and create a complete survey of them. They collect more than 100 published works and classify them into 3 categories. This chapter contains different works, which are done for creating CAPTCHA methods and assessing CAPTCHA methods from different aspects, including the attacks done against CAPTCHA methods. This chapter can be used by researchers in CAPTCHA domains to quickly find previous works.

INTRODUCTION

Many aspects of human life are affected by the expansion of the World-Wide Web (WWW). In particular, in developed countries, many daily affairs from daily shopping to education and commerce can be done using the Internet. A common action which is done in most websites, especially those belonging to commercial and administrative applications, is to fill out registration forms for certain purposes. After filling out the forms by entering the required information, the individuals are allowed to login into that website and carry out certain jobs.

Unfortunately, there are persons who abuse Web services by writing programs which do automatic false registration in websites. These programs automatically fill out forms with incor-
Distinguishing Human Users from Bots

rect information in order to enroll in the site. This wastes a large volume of the resources of the site in favor of the profit-seeking programmers and reduces the performance of the system.

Various methods have been presented in order to prevent such attacks, aiming at distinguishing between human users and computer programs. The main characteristic of these methods should be their automaticity so as to be run only by using the computer; because examination of a large bulk of registration on the Internet Web sites by human forces requires a great deal of time and expense and in some cases, such as email services Web sites, using human force for examining the registration forms is practically impossible. Therefore, it is necessary to use automatic systems to distinguish human users from computer programs.

In the discussions of artificial intelligence (AI), a test known as the Turing test (Turing, 1950) is proposed for testing the intelligence of a computer. In this test, a human person and a computer are put in two different rooms and a human interrogator in a third room asks them questions. If the interrogator cannot recognize which room the computer is in and which room the human, it is said that the computer has passed the Turing test.

A similar method to the Turing test can be used to distinguish human users from computer programs with the difference that the human interrogator is replaced with a computer. The computer interrogator asks questions from the applicant to distinguish between the human user and the computer program. These methods are known as CAPTCHA (Completely Automated Public Turing test to tell Computers and Human Apart), after the CAPTCHA project which is done at Carnegie Mellon University (von Ahn et al., 2004). Therefore, the main focus of these methods is on questions that the human user can easily answer but the present computer programs are hardly likely to be able to answer.

The CAPTCHA methods were successful in blocking automated programs from abusing Web services. The developers of first known CAPTCHA method, which is developed in AltaVista, said that their method could successfully block more than 95% of spam URL’s added to AltaVista search engine during its first year usage in 1997 (Baird & Popat, 2002). With the increase of spammers during past years, more websites are using CAPTCHA in their Web services to protect their services from abusing by spammers.

Nowadays, we are facing CAPTCHA on many websites: whenever we want to register for an account in a website, when we want to post a comment to a blog, when we are reporting a bug for a product, or even if we are using a service more than usual, such as when we are sending too much search queries to a website or trying more than 3 times to enter our email password. Their success made this field an attractive field for researchers and many works are done for designing CAPTCHA methods and assessing their robustness to attacks.

The aim of this chapter is to provide a complete survey of works done on CAPTCHA. As we will see in section 2, there are a number of surveys of CAPTCHA methods, but they are not broad. This survey tries to be comprehensive, which it means that we are collecting nearly all of the published works done on CAPTCHA including different papers reporting a work and also technical reports which are describing the work in more details. A shorter version of this survey is presented in Shirali-Shahreza & Shirali-Shahreza (2008d). This chapter is an expanded version which covers more methods with more detail discussion. We are also adding a new section for covering methods which are especially designed for disabled people, because we think that it is an important topic and is an open field for further research.

The structure of this chapter is as follows: In section 2 we will mention the history of CAPTCHA methods and provide a classification for different types of CAPTCHA methods. We divided the CAPTCHA methods into three categories: OCR-Based, Visual Non-OCR-Based and Non Visual. Methods of each type are reviewed in sections 3, 4 and 5 respectively. A number of CAPTCHA
Related Content

Using Statistical Texture Analysis for Medical Image Tamper Proofing
[www.igi-global.com/article/using-statistical-texture-analysis-medical/2484?camid=4v1a](www.igi-global.com/article/using-statistical-texture-analysis-medical/2484?camid=4v1a)

False Alarm Reduction Using Adaptive Agent-Based Profiling
[www.igi-global.com/article/false-alarm-reduction-using-adaptive-agent-based-profiling/111276?camid=4v1a](www.igi-global.com/article/false-alarm-reduction-using-adaptive-agent-based-profiling/111276?camid=4v1a)

eHealth Governance, A Key Factor for Better Health Care: Implementation of IT Governance to Ensure Better care through Better eHealth
[www.igi-global.com/chapter/ehealth-governance-key-factor-better/52361?camid=4v1a](www.igi-global.com/chapter/ehealth-governance-key-factor-better/52361?camid=4v1a)

Protecting User Privacy Better with Query l-Diversity
[www.igi-global.com/article/protecting-user-privacy-better-query/46100?camid=4v1a](www.igi-global.com/article/protecting-user-privacy-better-query/46100?camid=4v1a)