

THE VALUE OF CAPTCHA

Humans vs Robots

Beginners' Kaffee Klatch
Presented by Bill Wilkinson
April 4, 2009

A CAPTCHA is a program that protects websites against bots (robots) by generating and grading tests that humans can pass but current computer programs cannot. For example, humans can read distorted text as the one shown here, but current computer programs can't.



The term CAPTCHA (for Completely Automated Turing Test To Tell Computers and Humans Apart) was coined in 2000 by four computer engineers at Carnegie Mellon University. At the time, they developed the first CAPTCHA to be used by Yahoo.

CAPTCHA is rather odd sounding, but it's something we all have dealt with on our computers from time to time. It's basically a kind of response test used with computers to determine if a user who is attempting to register for access to a website is a human or a robot.

In other words, when you're on certain Web sites, have you ever had to type in a series of convoluted letters and numbers in order to continue? If so, that's a CAPTCHA test! The letters and numbers are distorted so as to make it impossible for a computer robot to read. CAPTCHA tests are usually used on sites that allow you to do your shopping online. Web sites like MySpace and Ticketmaster, among others, use them as well.

CAPTCHA tests are sometimes hard to solve, but they are there for your own good. They are mainly used for security, especially on Web sites that require you to enter in your personal information. Hackers use what are called bots (short for robots) to attack users. The bots are generated by computers which are unable to solve the CAPTCHA tests. Only humans can type in the right code and continue

on, so that really helps in keeping you safe. So, while it's true that these tests can be quite a nuisance, seeing these boxes of “gibberish” near the end of a registration page should make you feel more confident that none of your personal data is being compromised should a robot attempt to access one of your accounts as you.

Applications of CAPTCHAs

CAPTCHAs have several applications for practical security, including (but not limited to):

Preventing Comment Spam in Blogs. Most bloggers are familiar with programs that submit bogus comments, usually for the purpose of raising search engine ranks of some website (e.g., "buy penny stocks here"). This is called comment spam. By using a CAPTCHA, only humans can enter comments on a blog. There is no need to make users sign up before they enter a comment, and no legitimate comments are ever lost!

Protecting Website Registration. Several companies (Yahoo!, Microsoft, Google, MyWay) offer free email services. Up until a few years ago, most of these services suffered from a specific type of attack: "bots" that would sign up for thousands of email accounts every minute. The solution to this problem was to use CAPTCHAs to ensure that only humans obtain free accounts. In general, free services should be protected with a CAPTCHA in order to prevent abuse by automated scripts.

Protecting Email Addresses From Scrapers. Spammers crawl the Web in search of email addresses posted in clear text. CAPTCHAs provide an effective mechanism to hide your email address from Web scrapers. The idea is to require users to solve a CAPTCHA before showing your email address.

Online Polls. In November 1999, <http://www.slashdot.org> released an online poll asking which was the best graduate school in computer science (a dangerous question to ask over the web!). As is the case with most online polls, IP addresses of voters were recorded in order to prevent single users from voting more than once. However, students at Carnegie Mellon found a way to stuff the ballots using programs that voted for CMU thousands of times. CMU's score started growing rapidly. The next day, students at MIT wrote their own program and the poll became a contest between voting "bots." MIT finished with 21,156 votes, Carnegie Mellon with 21,032 and every other school finished with fewer than 1,000 votes. Can the result of any online poll be trusted? Not unless the poll ensures that only humans can vote.

Preventing Dictionary Attacks. CAPTCHAs can also be used to prevent dictionary attacks in password systems. The idea is simple: prevent a computer from being able to iterate through the entire space of passwords by requiring it to solve a CAPTCHA after a certain number of unsuccessful logins. This is better than the classic approach of locking an account after a sequence of unsuccessful logins, since doing so allows an attacker to lock accounts at will.

Search Engine Bots. It is sometimes desirable to keep webpages unindexed to prevent others from finding them easily. There is an html tag to prevent search engine bots from reading web pages. The tag, however, doesn't guarantee that bots won't read a web page; it only serves to say "no bots, please." Search engine bots, since they usually belong to large companies, respect web pages that don't want to allow them in. However, in order to truly guarantee that bots won't enter a web site, CAPTCHAs are needed.

Worms and Spam. CAPTCHAs also offer a plausible solution against email worms and spam: "I will only accept an email if I know there is a human behind the other computer." A few companies are already marketing this idea.