

01-13  
Rev. - 200

# SPURGEON

The right use of knowledge

sponsored by:

**TIMES**  
COMPUTER EDUCATION



ELECTRONICS AND COMMUNICATION DEPARTMENT  
C. U. SHAH COLLEGE OF ENGINEERING & TECHNOLOGY  
WADHWAN CITY.

**STAFF COMPOSITION**

Entitled by      Prof. A. C. Suthar  
EC Deptt. CCET

**DEFENDING E-COMMERCE WITH CAPTCHAS**

CAPTCHAS can provide an easy programmable way to Tell computers from human being & keep spammer and Bots away from e-commerce system.

**ABSTRACT**

CAPTCHAS is an acronym for completely automated public turing test to tell computers and human a part. It is a test that human can pass but computer can't. Turing test is conversational where as captcha test works on variety of inputs. They are made up of pseudorandom letters and numbers which are placed in front of obfuscating background or run through some degradation algorithms to make OCR of the final image impractical. The use of these captchas to defend e-commerce system by spammers, heckers or spoolers. It is a type of challenge response test used in computing to determine whether the user is human or not. If we don't use this type of alternative then spammers can write simply automated scripts to create hundreds of free e-mail accounts with a single command. There are so many types of captchas like Gimpy, Bongo, Fix, Sound, Buffle text etc. So far as technology is getting better and better every year, spammers and hackers can also be able to solve the captchas. Captchas development on both the side (programs and system that defeat the captchas). Both the sides are using advanced image processing as well as dictionary style attack because captchas are made up of letters and numbers. Researchers are developing audio captchas. That will be more difficult to solve by spammers.

**INTRODUCTION**

A CAPTCHA is acronym for "Completely Automatic Public Turing test to Tell Computer and humans Apart." It is a test that human being can pass but computer program can not, such tests are becoming key to defending e-commerce systems. If we don't use these type of alternative then spammers can write simply automated scripts to create hundreds of free email accounts with a single command. And e-mail service provider cannot choose validity. It's a type of challenge response test used in computing to determine whether the user is human or not.

By inserting a captcha, the system administrators can defeat these automated scripts and have some assurance that an actual human is associated with the account. Moreover, they are also useful in defending online shopping or auction sites by preventing spammers from posting irrelevant or bogus bids to prevent other buyers from purchasing the products. They are modern implementation of the Turing test, which asks a series of questions to two players: a computer and a human and both mislead the judge and decides which one is human. Captchas are similar to the Turing test that they distinguish computers from humans, except that with a captcha, the judge is also a computer. Captchas also differ from the Turing test because they work on a variety of inputs, whereas the Turing test is conversational. Captchas come in several types. Most of the captchas are made up of pseudorandom letters and numbers placed either in front of an obfuscating background or run through some degradation algorithms to make OCR of the final image impractical.

**HISTORY**

Alta Vista was the first to use a simple Captcha that generated images of random text. It used the captcha to prevent users from abusing its free URL services. Andrei Broder, Alta Vista chief scientist patented this technology in 2001. The Alta Vista Captcha reduced abuse by 95%.

In 2000, Udi Mamber of Yahoo was looking for ways to prevent bots from joining the online chat room posting advertisement & irritating people. But how can bots be refused? Researchers at Carnegie Mellon University took up the problem and proceeded to quantify desirable characteristics of captchas as well as generate several types. Xerox Palo Alto Research Center (PARC) also actively continues to study captchas.



Meanwhile, Captchas are enjoying increasing adoption by Internet service providers, free e-mail servers, online tickets agent, online auction sites, and file sharing sites.

#### How are the images integrated?

Both your Webserver and our CAPTCHA server need a common secret key. The picture is integrated quite similar to page counters. As a parameter you have to send a random string within the link. The secret key and the random string are used to generate the password. Both strings are needed to compute the password. The password is computed by the CAPTCHA server to generate the image and by your server to validate the users input. A random string should not be used multiple times.

#### What does it cost?

For non-commercial use this service is free. To register send an email with your preferred user name to mail@captchas.net. We will send you the generated secret key by e-mail. Inquiries regarding commercial should be sent to the same address.

#### How does that work technically?

The following demonstration in JavaScript shows the computation of the password and the image link (Note: JavaScript is not necessary on your web pages). You can use this form to check your own implementation. For this demo please use the username "demo" and the secret key "secret".

#### How is the password computed?

- Concatenate the secret key and the random string.
- Take the md5-sum of the resulting string.
- Take the first 6 bytes of the resulting 16-byte-long MD5 value.
- Every byte modulo 26 encodes a character from the range "a" to "z".
- This string of 6 small letters is the password.

#### How does it work?

A CAPTCHA image shows a random string which the user has to type to submit a form. This is a simple problem for (seeing) humans, but a very hard problem for computers which have to use character recognition, especially, because the displayed string is alienated in a way, which makes it very hard for a computer to decode.

#### TYPES OF CAPTCHAS:

- **GIMPY** : It is a type of captcha based on Optical Character Recognition. It was developed in collaboration with Yahoo! to protect chat rooms from spammers. It works by selecting several words from a dictionary and displays them. Users must enter the words in the image to get entry into the service.

Bongo is a program that asks the user to solve a visual pattern recognition problem. In particular, Bongo displays two series of blocks, the left and the right series. The blocks in the left series differ from those in the right, and the user must find the characteristic that sets the two series apart.

After seeing the two series of blocks, the user is presented with four single blocks and is asked to determine whether each block belongs to the right series or to the left. The user passes the test if he or she correctly determines the side to which all the four blocks belong.

- **PIX** : This uses a large database of photographic and animated images of everyday objects. Captcha system then presents a user with a set of images, all associated with a same object or concept. The user must then enter a object to which all the images belong. For example, the program might Bongo is a program that asks the user to solve a visual pattern recognition problem. In particular, Bongo displays two series of blocks, the left and the right series. The blocks in the left series differ from those in the right, and the user must find the characteristic that sets the two series apart.

After seeing the two series of blocks, the user is presented with four single blocks and is asked to determine whether each block belongs to the right series or to the left. The user passes the test if he or she correctly determines



the side to which all the four blocks belong.

- **P1X**: This uses a large database of photographic and animated images of everyday objects. Captcha system then presents a user with a set of images, all associate with a same object or concept. The user must then enter a object to which all the images belong. For example, the program might present picture of a globe, volleyball, planet, and base ball, & user to correctly associate all these type with the the word 'boll'.

- **SOUND**: Sounds can be thought of as a sound version of Gimp. The program picks a word or a sequence of numbers at random, renders the word or the numbers into a sound clip and distorts the clip. It then presents the distorted, sound clip to its user and asks the user to type in the contents of the sound clip.

- **BUFFLE TEXT**: Sounds can be thought of as a sound version of Gimp. The program picks a word or a sequence of numbers at random, renders the word or the numbers into a sound clip and distorts the clip. It then presents the distorted sound clip to its user and asks the user to type in the contents of the sound clip

#### CHARACTERISTICS

- Regardless of the type, good captchas share many common characteristics.
- They are amenable to completely automated processes for generating tests. It requires intervention or involvement.
- The code, data, and algorithm must be public.
- Good captchas rely on a completely random system of generation based on choosing files from a database consisting of many names, images, and other files.
- The data base used to create the Captchas should not contain the solutions, because hackers could crack the database and obtain the test solutions. It is also important that the computer program generating the Captchas not be able to also solve them. If it did, it would be possible for hackers to exploit the program to solve its own Captchas.

#### APPLICATIONS

- Captchas have been used to prevent web crawlers and bots from participating in online polls. System administrators accomplish this by inserting the captcha into the vote submission process. This mechanism, of course, doesn't prevent individuals from voting multiple times.
- Many free e-mail service providers like Yahoo and MSN Hotmail use captchas to prevent spammers from creating accounts using automated scripts.

#### ACCESSIBILITY

Captchas based on reading text or other visual-perception tasks prevent visually impaired users from accessing the protected resource. However, captchas do not have to be visual. Any hard artificial intelligence problem, such as speech recognition, can be used as the basis of a captcha. Some implementations of captchas permit users to opt for an audio captcha. The development of audio captchas appears to have lagged behind that of visual captchas, however, and presently may not be as effective. Other kinds of challenges, such as those that require understanding the meaning of some text (e.g., a logic puzzle, trivia question, or instructions on how to create a password) can also be used as a Captcha. Again, there is little research into their resistance against countermeasures.

For non-sighted users (for example blind users), visual captchas present serious problems. Because captchas are designed to be unreadable by machines, common assistive technology tools such as screen readers cannot interpret them. Since captchas are often used in initial registration processes (for example eBay and Yahoo!, and some other sites), this challenge can completely block access. In certain jurisdictions site owners could become target of litigation if using Captchas that discriminate against certain people with disabilities.

Even for perfectly sighted individuals, new generations of captchas, designed to overcome sophisticated recognition software, can be very hard or impossible to read. Even some of the demo captchas at the software sites listed below are indecipherable to many if not all humans.



## EXAMPLE

In 1997, Andrei Broder and his colleagues at DEC Systems Research Center created the first Captcha to prevent abusive and automated URL submissions to the Alta Vista search engine. Yahoo! uses the EZ-Gimpy captcha (developed at Carnegie Mellon University) to protect the online services, including free e-mail accounts registrations. Ticket master uses captchas to prevent scalpers from generating automated on high value tickets. Ear Mink's Spam Blockers uses captchas to challenge e-mail senders trying to gain access to a recipient's "allowed senders" list.

Many file download sites use captchas to prevent bots from downloading and archiving the entire file library.

## ADVANTAGES AND DISADVANTAGES

The main advantage of Captcha is that they are effective at defeating spammers, spoofers, search engine crawlers and virtually all automated programs that might try to access a site services. They do this in a relatively automatic way that is considerably cheaper than the available alternatives, such as requiring users to call a human to obtain access to a resource.

However, the use of Captchas has several disadvantages. Captchas are unfriendly for the disabled and visually impaired, though research continues on audio Captchas to alleviate this problem, such systems also require a large image library, server and software to generate the Captchas. The time to generate, display and grade the Captchas increases the load on the server and presents delay to the user. Captchas are moderately difficult to work around, for example, a hacker can program a bot to log all sites presenting Captcha so that a human user can later solve the Captchas. Captchas impose an accessibility problem and annoyances on genuine users.

## CREATING & USING CAPTCHAS

Someone told me about an ingenious way that spammers were cracking "captchas" the distorted graphic words that a human being has to key into a box before Yahoo and Hotmail and similar services will give her a free email account. The idea is to require a human being and so prevent spammers from automatically generating millions of free email accounts.

The ingenious crack is to offer a free porn site which requires that you key in the solution to a captcha which has been inlined from Yahoo or Hotmail before you can gain access. Free porn sites attract lots of users around the clock, and the spammers were able to generate captcha solutions fast enough to create as many throw-away email accounts as they wanted.

Now, chances are that they didn't need to do this, since optical character recognition has been shown to be readily tweakable to decode captchas without human intervention that which a computer can generate, a computer can often solve.

## FUTURE OF CAPTCHAS

Researchers at University of California, Berkeley, have developed computer programs that can automatically solve the Captcha with 83% accuracy. Researchers are also developing the audio captchas independent of visual perception. The current focus seems to be on human generated captchas. These type of captchas are much difficult to solve.

Captchas development continues on both the offensive (programs and systems that defeat Captchas) and defensive (improved captchas) sides. Both sides are using advanced image processing as well as dictionary-style attacks. This sort of "arms race" between researchers seeking more secure captchas and the crackers, spoofers, spammers trying to defeat the captchas is likely to continue for some time.