



Strength and Weakness of Various CAPTCHA Schemes

Mohinder Kumar

Panjab university Regional Centre, Sri Muktsar Sahib

Abstract

HIP Human Interaction Proofs or CAPTCH are very popular these days for the web site authentication as well as authorization. It is a Reverse Turing Test to be judged by the computer for verification of a human being. These HIPs are available in a number of forms like text based, image based, audio based, puzzle based, video based and also some invisible types of schemes are also launched by the Google. All these schemes have been used by the web sites in a number of ways to improve the security of web sites from bot programs. A brief but useful effort has been made in this paper to highlights the strengths as well as weaknesses of all these types of CAPTCHA schemes. It should be very beneficial for all researchers who want a brief and impact information about all the available CAPTCHA schemes.

Keywords: CAPTCHA, Strengths, Weaknesses, Invisible CAPTCHA

1. Introduction

Today the use of web sites is increasing rapidly because every service is available on the web sites. Shopping, eating, ticket booking, travelling, banking, education and many more tasks of our day to day life are done online. The world is going to be a paperless world. It is very convenient to do all these tasks with just some of the finger taps or mouse clicks. Life becomes very comfortable. But the solutions generate new problems. A bot program is attacking these web sites for false authentication and the actual purpose is not solved. The need of security from these attacks is a question for the software engineers. The simple and very quick solution is CAPTCHA. It places a test to a bot program s to prove itself a human being. Because these web services are made for human beings not for these bot programs. So a number of CAPTCHA schemes are already discussed in [1]. The various strengths and weaknesses are not discussed of all these types in any paper at one place. So in this paper I have put my efforts to bring all these types of CAPTCHA schemes under one tree along with the positive aspects as well as limitations. In the following all these types are discussed one by one along with the strengths and limitations.

2. Text Based CAPTCHA

These are the classic schemes and even very popular. The user is given a simple to understand intuitively words or random characters that is hard to understand by a computer program. An example of such scheme is given in the Fig 1.



Fig 1. Text Based CAPTCHA

1.1 Strengths-

- 1.1.1 Text based CAPTCHA are the simplest type
- 1.1.2 Text CAPTCHA are designed in every language
- 1.1.3 The loading speed of these schemes is very fast
- 1.1.4 These are very easy to use and easy to understand for humans

1.2 Weakness -

- 1.2.1 To break these CAPTCHA is very easy most of the time.
- 1.2.2 A number of times simple OCR techniques are enough to crack them.
- 1.2.3 Generally these are designed in English language [2]. Although in recent times these are available in Chinese as well.
- 1.2.4 To make these challenges hard distortion is added that also creates problems for humans also. Some alphabets and digits have very different shapes, but when they are distorted it becomes very difficult to recognize them[4]

Sr. No.	CAPTCHA	Problem
1		There is confusion first 2 characters are "cl" or 'd'
2		Another confusion of "cl" and 'd'
3		Whether 2 nd and 3 rd character are 'l' and 'v' respectively or it 'w'
4		First two characters are 'm' or it is continuous m
5		A real headache: is the first part "m" or "rn,the middle part inv" or "nw"?

Table 1 Confusing characters Text CAPTCHA

1.2.5 Sometimes letters like cl can confuse as d , nn can confuse with letter m or rn or vv can confuse with w.

1.2.6 It is found in many cases 8 may look like 6 or 9 and 7 may look like 1 or vice versa. It can happen due to some specific type of fonts or too much distortion.

2. Image Based CAPTCHA

In these schemes images are shown to the user for identifying. Images can belongs to particular category like traffic signals, buildings or buses etc. all the similar pictures are to be selected to pass the test. The images are not very clear and also very damaged (in some cases). Fig.2 is an example of Image Based CAPTCHA.

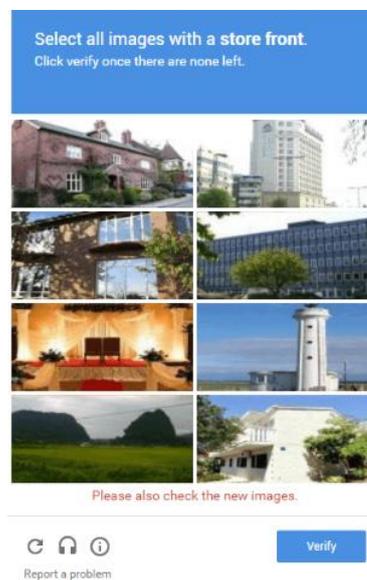


Fig. 2 Image Based CAPTCHA

2.1 Strengths-

2.1.1 These are still hard to recognize by a computer program. Image based CAPTCHA therefore very hard to break.

2.1.2 These images make the challenge very interesting for the user.

2.1.3 For improvement in network speed these schemes are also very fast

2.2 Weaknesses-

2.2.1 Mostly Image based CAPTCHA scheme are available in only English language. So it is necessary that the end user must know the English vocabulary. According to a survey there are only 25% Internet users are English speaking [5] as shown in the following table.

Rank ↕	Language ↕	Internet users ↕	Percentage ↕
1	English	1,052,764,386	25.3%
2	Chinese	804,634,814	19.4%
3	Spanish	337,892,295	8.1%
4	Arabic	219,041,264	5.3%
5	Portuguese	169,157,589	4.1%
6	Indonesian / Malaysian	168,755,091	4.1%
7	French	118,626,672	2.9%
8	Japanese	109,552,842	2.8%
9	Russian	108,014,564	2.7%
10	German	84,700,419	2.2%
11-36	Others	950,318,284	22.9%
Total		4.16 Billion	100%

Table 2 Language Used on the Internet

- 2.2.2 When too fewer images are given to the user then the scheme is vulnerable to attack.
- 2.2.3 Image CAPTCHA poses a problem to low vision users or having learning disability [6]
- 2.2.4 Image CAPTCHA generates heavy load on the server side.
- 2.2.5 These schemes become IQ test in the presence of ambiguity in objects of image.
- 2.2.6 In case an image contains text for passing challenge then simple OCR program can break the scheme very easily.

3. Audio Based CAPTCHA

Text and Image based schemes are not very useful for those persons who are have poor eye vision. For such online users a new scheme is proposed by a number of researchers that is based on sounds. The user is given a sound that is a pronunciation of the letters in the image of text. The Fig. 3 is an example of audio CAPTCHA

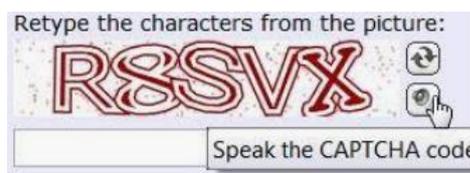


Fig. 3 Audio CAPTCHA

3.1 Strengths-

- 3.1.1 Audio CAPTCHA are a better choice for users who have poor eye sight.
- 3.1.2 Audio CAPTCHA becomes more strong and secure after adding the noise.

3.2 Weaknesses-

- 3.2.1 These schemes are not very beneficial for hearing disabled users.
- 3.2.2 Audio CAPTCHA are designed for visually impaired user but proves less useful if some pointing device interaction is needed.
- 3.2.3 Audio CAPTCHA frustrates the users by consuming more time.
- 3.2.4 The knowledge of English language vocabulary is must for using this CAPTCHA [7].
- 3.2.5 Similar sounds alphabets create confusion in the user mind.
- 3.2.6 Sometimes audio based CAPTCHAs becomes more problematic as compare to visual CAPTCHAs, particularly for visually impaired users who utilization screen-reader software.
- 3.2.7 When too much noise is added the usability of audio CAPTCHA decreases.
- 3.2.8 English listening skills are also required proficiently as most of the audio CAPTCHA is available in English.
- 3.2.9 Specific alphabets like C & K, G & J have similar sound so it decreases usability [3] [8].
- 3.2.10 Most of the audio files being used in these schemes are of very poor quality. Inability to control volume.
- 3.2.11 Audio schemes are also issues regarding accessibility.

4. Video Based/ Animation CAPTCHA

Video or animation based CAPTCHA is also available in these days. In such tests a video is showing to the user and asked to identify the object or words in a video or animation. For example Fig. 4 is showing a animated text in specific color and the user is instructed to type that particular text to access the web site or create an account.



Fig. 4 Animation CAPTCHA

4.1 Strengths-

- 4.1.1 It is difficult to break a video CAPTCHA by a bot program.
- 4.1.2 The user feels it very interesting type of CAPTCHA.

4.2 Weaknesses

- 4.2.1 In case of low internet speed the scheme is not usable at all.



4.2.2 Like audio schemes these video schemes also requires the proficiency in English as most of these schemes are available in English.

4.2.3 These CAPTCHA again not very beneficial for users having low visions.

4.2.4 Although video CAPTCHA are interesting to see but very time consuming.

5 Puzzle based CAPTCHA

In this scheme a game like puzzle or a simple math question is asked to the user. It is more interactive and takes a more time to pass the test. It is not very popular type of CAPTCHA scheme. Sometimes a simple questions like $233+500$ is asked to the user, but sometimes a harder questions Series like question is asked.

5.1 Strengths-

5.1.1 Puzzles are liked by everyone so these schemes are very interesting.

5.1.2 Easy to design.

5.1.3 In most of these schemes simple mathematics is used.

5.2 Weaknesses-

5.2.1 Although simple math is used but these takes more time solve as compare to text base schemes.

5.2.2 In some puzzle based CAPTCHA the use of images is more. it makes the schemes less usable for those who have low vision same as image CAPTCHA.

6. Invisible CAPTCHA

In recent times Google provides an invisible CAPTCHA that utilizes the mouse movements and some other details to prove that the user is human being not a bot.

6.1 Strengths-

6.1.1 This invisible scheme require no audio, video files .

6.1.2 It requires just mouse click and even no after continuing use of this scheme.

6.1.3 Google distributes this scheme freely.

6.1.4 It takes a little time to test.

6.2 Weaknesses-

6.2.1 This new scheme is not very reliable till date.

6.2.2 Most of the times image based scheme is given to the user that makes is too much time consuming.

6.2.3 If images are given then it makes it a language dependent scheme (English in particular).

6.2.4 It is again not comfortable for low vision users.

Conclusion

The paper is very beneficial for those researchers who want to start their research in CAPTCHA schemes. Here all the strengths and weaknesses are discussed that make it easy to decide that what type of CAPTCHA can be improved. It also helps to design some new kind of security schemes. The paper does not include some non English CAPTCHA schemes that may make the paper more useful for multi language CAPTCHA researchers.

