

## Recent Trends in CAPTCHA Security and Bot Challenges

Shivank Singh<sup>1\*</sup>, Rahul kumar Chawda<sup>2</sup>

<sup>1,2</sup>Dept. of Computer Science, Kalinga University, Naya Raipur, Raipur, Chhattisgarh 492101, India

\*Corresponding Author: shivanksingh622@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v8i5.123127> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Received: 13/May/2020, Accepted: 20/May/2020, Published: 31/May/2020

**Abstract-** CAPTCHA is a security mechanism which is basically used to distinguish between human and bots in automated online systems. Here bot means software robots that are used for brute force attacks or webmail bombarding or DOS attack to the system. CAPTCHA is also useful for protected web services from various types of dynamic attacks every day. Generally, it is seen in online registration form for preventing spam applications. In this paper, we have studied the various types of CAPTCHA like picture CAPTCHA, video CAPTCHA, content CAPTCHA, sound CAPTCHA and their uses in present environments.

**Keywords:** CAPTCHA; Security; Security by Bot Attack.

### I. INTRODUCTION

CAPTCHA stands for Completely Automated Public Turing Test to Tell Computers and Humans Apart. It is a smart program that is designed to test human solvable problems and it also differentiates between bot and human. In CAPTCHA technology, audio or visual contents are generated [1] [3]. Text-based CAPTCHA is widely used. It is most acceptable CAPTCHA form. Now-a-days CAPTCHA is widely used in web security with various kinds like picture, video, alphabets and hybrid of the CAPTCHAs [2].

The graphical security is like Turing Test. It can restrict automated programs. CAPTCHAs are now called “Reverse Turing Tests”: Since they are aiming to permit a computer to decide in case the client is human or not. In show its significance, their amazingly broad utilize, and a developing number of investigate thinks about there's right now no efficient strategy for planning or assessing CAPTCHAs. In truth, as we substantiate by careful ponder, numerous well known websites still depend on plans that are powerless to mechanized assaults. A novel e-CAPTCHA is adopted to use for new security challenges. It can be also used as graphical password [4]. The CAPTCHA test helps recognize which customers are genuine people and which ones are COMPUTER programs. Spammers are always endeavoring to gather computations that read the ravaged substance effectively. So, strong CAPTCHAs must be arranged and created so that the endeavors of the spammers [10]. The internet has gotten to be steadily a stage for arrangement of complex applications of expanded interactivity. Inside this exceedingly energetic and advancing setting, security issues and concerns have moved recently to the center of consideration. In fact, the results of a security breach can hurt the validity and legitimate liability of an organization, driving to misfortune of users “believe”. In this setting, especially imperative security concern is that pernicious clients or program specialists attempt to robotize the abuse of framework assets having as a side impact the debasement of benefit quality for typical clients [6].

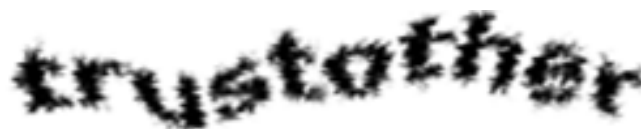
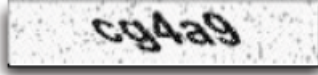



Fig.1. A typical CAPTCHA

Table 1. Sample Questions of each category of CAPTCHA

Category	Questions
Analytical	Ravi had apple and mango. Ravi ate orange. Which fruit is left?
Mathematical	What does the square root of 144 equal?
General	How many bones does an adult human have?
Text	
Image	

## II. LITERATURE REVIEW

Creators endeavor to fathom particular text-based challenges by presenting a combination of procedures for the change of the picture, character extraction and acknowledgment for each conveyed CAPTCHA conspire. Other than the clamor evacuation and division procedures, which are depended on the text-based challenge, the most contrast between the previously mentioned procedures is the classification strategy being used [1].

The various strategies were presented to utilize both CAPTCHA and Password in a client verification convention, which we call as CAPTCHA-based Password Authentication protocol, makes a difference to resist the online dictionary attacks [2].

The MSN Plot (Figure 2) appears a few test challenges created by the MSN CAPTCHA plot. We have no get to the codebase of the MSN conspire, so we collected from Microsoft's site 100 arbitrary tests that were produced in genuine time online at 16 By examining and the tests we collected, we watched that the MSN plot (as sent) has the taking after characteristics. In MSN CAPTCHA, 8 characters are used in each challenge, only capital letters as well as digits are used [3].



Fig.2. A typical MSN CAPTCHA

The motion pattern CAPTCHA is also used for better security in which user need to draw a specific pattern which is based on AI algorithms. Some CAPTCHAs use grid-box to represent graphical contents. It used two levels for eCAPTCHA. In first level, animal-grid is used. In second level, number-grid is used [4].

Software uses CAPTCHA that requires a question to identify the human or bot but it may reduce the usability of software. So, we have to balance the security by using three basic properties: 1. Easy to human to pass the test, 2. Easy for tester machine also for fast processing. 3. It should be hard for bots or software to penetrate the system [5]. There are other some features are defined in CAPTCHA: 1. Automated: A program should generate and evaluate the CAPTCHA, 2. Open: by some algorithms of the database, it should be public to grade, 3. Usability: It should be human solvable 3. Secure: It should be difficult for bots [11].

Internet technology uses CAPTCHA to overcome various challenges. Some users face great difficulties to solve CAPTCHA challenges. Still, only a 48.5% CAPTCHA reported that it fails in first try and the other 51.5% take more tries to solve its challenge [6]. A CAPTCHA should be automatically generate and test and distinguish between human and bots. A grey level increases the power of CAPTCHA. Grey level background noise harder to recognize by bots [7].

There are following 3 types of CAPTCHAs: 1. Text-based: It uses some distortion in text but human can recognize easily, 2. Sound-based: Audio clip used to solve some problem like  $2+2$  is what? and 3. Image-based: A grey scale in images with numbers are used such that human can easily recognize it [9]. By testing the various CAPTCHA only Google and ReCAPTCHA are protected by some typical attacks. Every day new algorithms are used to break CAPTCHA security, So it is consistent process to boost the security [8]. CAPTCHA technology prevents from fishing attacks. It performs a sequence of clicks on an image to identify human. A technology of multi-factor authentication is used for better security that protects thousands of fraudulent over the time [10].

The CAPTCHA application may be useful in various areas like online systems, email sign up, online purchase and sale, online reservation systems etc. [12] Almost every online commercial activities uses CAPTCHA for it security issues, however many tools are developed by hackers which tries to complete automatically commercial activities [13]. CAPTCHA Sample is novel website for testing various types of it which are designed for research and study purposes [14]. The font choice of letters, colors and background color i.e. noise with distortion boost the power of CAPTCHA. Some random parameters should be selected to avoid recent cyber-attacks on the system [15]. CAPTCHA has many forms like text, image or hybrid but the automated fraud programs is similar to reverse turing test which is already given by Alan Turing [16].

### III. Types of CAPTCHA

#### 1.1. CAPTCHAs based on Text

Text-based CAPTCHA is very simple to implement and effective with huge database of it. The number of categories and digits are small therefore the problems occur for humans to identify correct characters and digits. It is possible for recognizing the elements as the character and digits by the device OCR (Optical Character Recognition).

In text-based CAPTCHAS, the following questions may be asked:

- (i) Simple Arithmetic Addition: What is two plus two? i.e.  $2 + 2 = ?$
- (ii) Simple Arithmetic Subtraction: What is four minus two? i.e.  $4 - 2 = ?$
- (iii) Simple questions regarding universal truth: In which direction the sun rises?

##### 3.1.1. Sub-subheadings

In this method, some set of characters are selected and presented them as distorted with gray background with the help of some lines or non-uniform color pattern also. In this technique, users have to type the characters in correct way. This pattern becomes difficult for bots.

This is already implemented in yahoo to prevent form bot attacks.

##### 3.1.2. PessimialPrint

In this method, the quality of text images is degraded to make it difficult for bots. It is nearly 10 parameters of physical of machine printing. This method takes spatial sampling rate, character size, errors, and some blur thresh holdings etc. [13].

#### 1.2. CAPTCHAs based on Image

It is one the most effective and advanced CAPTCHA [12]. In this challenge users have to guess to those similar images to identify bots and humans [16]. Some words are also mixed with image-based CAPTCHA which becomes more difficult to perform bot attacks. This method uses huge database of images. Sometimes animated images are also used [15]. The image transformation technology is also used for 2D, 3D images to boot the security of CAPTCHA.

Types of Image CAPTCHA:

##### 3.2.1. Bongo

This model consists of 2 series of blocks namely the left block series and right block series. The both blocks of series are different and users have to identify those differences [15].

##### 3.2.2. Pix CAPTCHA

This model uses huge database of both simple images and animated images of easily available domestic animals like cats, dogs [13]. The users have to choose their categories of the images. This technology uses the pattern recognition based on AI algorithms. This makes it difficult for bots to crack it [16].

### 1.3. CAPTCHA based on Audio

This model of CAPTCHA is based on the sound technology. This becomes very useful for those users who cannot view properly to the images or text. It is downloaded audio clip and first user listen it properly and reply or type the same spoken words in the given textbox. This matching is done automatically by the sophisticated algorithms [15].

The main drawback of Audio-based CAPTCHA is lack of user-friendly and robustness [11].

### 1.4. CAPTCHA based on Video

This model is novel and less popular. Video-based CAPTCHA contains some predefined special colored text characters among all the text information. This is the challenge for users to recognized the specific text message and type to the textbox manually. Unfortunately such kind of CAPTCHA is vulnerable to bots also due to its specific color or pattern.

### 1.5. CAPTCHA based on Puzzles

This model uses big database of puzzles and users have to solve these puzzles. This puzzle uses text-based CAPTCHA for asking to users. Users have to identify the specific location on the images for solving puzzle [13].

## 2. Drawbacks of Different Kinds of CAPTCHAs

Each type of CAPTCHA has certain properties and power. Each has some advantages and disadvantages also which is given below in the table.

## 3. The Usability of CAPTCHAs

There are three types of characteristics of CAPTCHA:

### (1) Usability:

Usability means how user-friendly and fast CAPTCHA that can be recognized by human but not recognized by bots. It also should be quick to solve. There are various factors to affect usability of CAPTCHA namely age factor, visibility, education etc. It is given in fig. 3 bellow.

Table 2. The comparative drawbacks of different CAPTCHAs [16].

Sr. No.	Different Types of CAPTCHA	Drawbacks
1	CAPTCHA based on Text	(1) Distorted text has certain problems to identify the correct characters, (2) Multiple font styles, (3) Different font size, (4) Blurred text or Gray Background, (5) Wave Motion, and (6) There are threats to identify easily by OCR techniques.
2	CAPTCHA based on Images	Low vision users face difficulties in image-based CAPTCHA due to noise in image.
3	CAPTCHA based on Audio	The correct pronunciation is difficult due to different geographical areas on the globe. There may be communication barriers also. Generally, it is available in English language so, users must have good knowledge of English. Characters with similar sound are also a major problem.
4	CAPTCHA based on Video	Video uses large in size so, some users faces problem to download and to find correct CAPTCHA.
5	CAPTCHA based on Puzzle	The puzzle task is not comfortable for all users. It also takes more time to solve puzzle.



Fig. 3. The usability of the CAPTCHA [18]

### (2) Security

Security is nothing but how difficult for bots or computers or algorithms to break the CAPTCHA. Ideal CAPTCHA should be easily solvable by human but not bots.

## (3) Practicality

CAPTCHA creation should be easy.

Table 3. The response time of some CAPTCHA

CAPTCHA Name	Response time in Second
Text	6
Alpha-numeric	6
Number	4
Hybrid	7

#### IV. CONCLUSION

In this paper, many kinds of CAPTCHAs have been studied. A brief literature survey has been done for text-based, image-based, audio-based, video-based and puzzle-based CAPTCHAs including advantages and disadvantages of each CAPTCHA. Usability of CAPTCHAs is also discussed.

The future of CAPTCHA security depends on the user-friendliness for all types of users as well as high difficulty level for machines or bots. Bot attacks should be prevented by the strong CAPTCHA but user's constraints should be also balanced. So, we suggest that selection of CAPTCHA security according to the application and users both.

#### Acknowledgments

We thanks to Computer Science Department, Kalinga University, Naya Raipur for providing support computer lab facilities to carry out the research work.

#### REFERENCES

- [1] Korakakis, M., Magkos, E. and Mylonas, P. (2014): Automated CAPTCHA solving: An empirical comparison of selected techniques, Proceedings - 9th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2014, pp. 44–47.
- [2] Barbole, B. A. and Surywanshi, S. (2015): A Survey on to Enhance Security Approach Using Discretized Centralization for CAPTCHA as Graphical Password, International Journal of Science and Research (IJSR), 4(11), pp. 826–830.
- [3] Yan, J. and Ahmad, A. S. El (2008): A low-cost attack on a microsoft CAPTCHA, Proceedings of the ACM Conference on Computer and Communications Security, pp. 543–554.
- [4] Soni, S. and Bonde, P. (2017): E-CAPTCHA : A Two Way Graphical Password based Hard AI Problem, (June), pp. 418–421.
- [5] Chew N. and Tygar J. D. (2004): Image Recognition CAPTCHAs, UC Berkeley Computer Science Division technical report.
- [6] Fidas, C. A., Avouris, N. M. and Voyiatzis, A. G. (2011): On the necessity of user-friendly CAPTCHA, Conference on Human Factors in Computing Systems - Proceedings, (May 2016), pp. 2623–2626.
- [7] Newton, F. and Kouritzin, M. A. (2011): On grey levels in random CAPTCHA generation, Visual Information Processing XX, 8056(c), p. 80560U.
- [8] Bursztein, E., Martin, M. and Mitchell, J. C. (2011): Text-based CAPTCHA strengths and weaknesses, Proceedings of the ACM Conference on Computer and Communications Security, 2011, pp. 125–137.
- [9] Yan, J. and El Ahmad, A. S. (2008): Usability of CAPTCHAs or usability issues in CAPTCHA design, SOUPS 2008 - Proceedings of the 4th Symposium on Usable Privacy and Security, pp. 44–55.
- [10] Siva Nagalakshmi, K., Prakash, P. S. and Prem Kumar, D. S. (2015): Confident Multi-Factor Authentication on web application via CAPTCHA Technologies, International Journal of Computer Engineering in Research Trends, 876(8), pp. 2349–7084.
- [11] Javed, M. and Ranjan, N. (2013): CAPTCHA Based on Human Cognitive Factor, International Journal of Advanced Computer Science and Applications, 4(11), pp. 144–149.
- [12] Brodić, D., Amelio, A. and Draganov, I. R. (2017): Statistical Analysis of Dice CAPTCHA Usability, pp. 1–9.
- [13] Abdullah Hasan, W. K. (2016): A Survey of Current Research on CAPTCHA, International Journal of Computer Science & Engineering Survey, 7(3), pp. 1–21.
- [14] Amelio, A. et al. (2018): The {CAPTCHA} Samples Websit, {ERCIM} News, 2018(112), p. 2018.
- [15] Kaur, K. and Behal, S. (2014): CAPTCHA and Its Techniques : A Review, International Journal of Computer Science and Information Technologies (IJCSIT), 5(5), pp. 6341–6344.
- [16] Singh, V. and Pal, P. (2014): Survey of different types of CAPTCHA, International Journal of Computer Science and Information Technologies, 5(2), pp. 2242–2245.
- [17] Roshanbin, N. (2014): Interweaving Unicode, Color, and Human Interactions to Enhance CAPTCHA Security.
- [18] Yan, J. and El Ahmad, A. S. (2016): Usability Analysis of the Specific CAPTCHA Types, International Scientific Conference, pp. II-272–II-277.