

## An insight into CAPTCHA

Yash Doshi<sup>1</sup>, Ami Sangani<sup>2</sup>, Pratik Kanani<sup>3</sup>, Dr. Mamta Padole<sup>4</sup>

<sup>1,2</sup>Bachelor of Engineering (pursuing), <sup>3</sup> Assi.Professor, <sup>4</sup>Asso.Professor

<sup>1,2,3</sup> Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Mumbai

<sup>4</sup>Department of Computer Science and Engineering, MS University, Baroda

### Abstract

**Due to the increase in illegitimate use of services, offered by various organizations, there needs to be a method in order to confine it from happening. Therefore CAPTCHA came into prospect in order to control this course of action. In this paper, the role of CAPTCHA in a lucid manner is represented. Different types of CAPTCHAs exist today, starting from the text captchas to graphical captchas as per the requirements. It also explains the captchas used by several software companies. Also, flaws of Captcha and how captchas create a friction among the users has been discussed. It also highlights various applications of CAPTCHA and various forms of attack that can take place onto captcha system.**

**Keywords**—CAPTCHA, De-CAPTCHA, CAPTCHA Attacks, text captcha, graphical captcha, captcha threats, captcha usability.

### I. INTRODUCTION

With the ongoing digitization across the world, websites play an important role in maintaining the e-relationship between the users and the service providers. To personalize the user's experience, these websites ask the users to register and create their own space by giving them an account. This goes well until the services are paid, but free service providers started getting exploited because users made multiple accounts. Not only the users, but the website holders do fake registration to upsurge their site ranking. Hackers thought after this opportunity and built programs, that recursively make accounts on the website. [1] To avoid such exploitative, automated and unethical activities, CAPTCHA was introduced. CAPTCHA stands for **C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part. It's a responsive test which identifies, if the user is human or not. The term was composed in 2003 by Luis von Ahn, Manuel Blum et al. The most basic type of CAPTCHA, first invented in 1997 by two groups working together by Mark L., Martin A., Krishna B., Andrei B. and Reshef, Raanan and Solan.

In this test users are required to type the letters of an image shown which is not cleared, tilted or disorganized. This test is monitored by the machine which makes the image difficult to read and enter. Such difficulties can only be understood by humans, so CAPTCHA is sometimes described as a reverse Turing test. [2] The CAPTCHA technology is mainly used to prevent bots and computer programs to get login to emails, web sites and sending spam messages. Images given by CAPTCHA cannot be read by any software programs and hence it prevents the automatic sign-in. [3] Attackers and spammers are constantly building algorithms which can detect, read and understand the CAPTCHA images. In order to avoid such malpractices CAPTCHA needs to be stronger.

### II. TYPES OF CAPTCHAS

There are three main characteristics of CAPTCHA:

1. It should be painless for humans to pass.

2. It must be uncomplicated for a tester machine to generate.
3. It must be tough for a software bots to pass. [4]

### A. Text CAPTCHA

These are the most simple to design and implement. It presents the user with some questions which only a human user can understand and solve. Some of the examples are:

1. What is five plus ten?
2. What is the second letter in INDIA?
3. Which of Green, Friday and Michael is a boy?
4. If tomorrow is a Monday, what is today?

Such tasks are very simple to answer for a normal human being and very difficult for visually disabled like color blindness, etc. [6]. Text CAPCHAS have limited number of questions which can be recorded for longer time and then based on matching patterns bots can answer the same.

### B. Distorted text

It applies distortion on the selected word from the dictionary and asks user to identify it. Such captcha can be easily generated from the dictionary based words and only humans can understand it, no bots can get it very easily because of their tilted structures. Hence they are widely used.



Fig 1. Text Captcha

### C. ReCAPTCHA



Fig 2. ReCAPTCHA

This approach was developed by one of the most recognized CAPTCHA organization. The ReCAPTCHA approach is trustworthy but sometimes it becomes troublesome to solve the words that are distorted. This was overcome by giving the visitant the option of “reCAPTCHA” and thus the visitant receives a new CAPTCHA. This type of CAPTCHA provides with an audio option if it becomes difficult for the visitant to visually recognize the word. [5], this captcha helps actual human users to log in to the systems if they are not able to get the captchas at first site because of closely placed letters or their tilted shapes. Advantages of ReCaptcha are the audio feature in recaptcha is an alternative for the people who are visually impaired and it becomes easy to identify the captcha through audio if people couldn't recognize it through scanned text. Whereas disadvantage of ReCaptcha are the audio is difficult to interpret sometimes, takes longer time to complete and usability ratio is low and not all the users will be able to recognize the English words.

#### D. Math Problem

Few Captchas provides user, a simple and straightforward math problem, the answer to which is the solution to that Captcha challenge. Such captchas are easy to generate and it takes less average time to solve than other captchas.

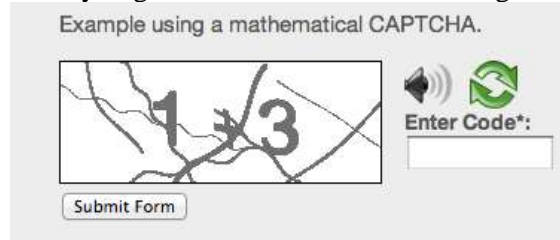


Fig 3. Math problem Captcha [5]

#### E. Image-Based CAPTCHAs

This Captcha requires users to identify a set of images which are distinguishably different from each other. Images are too difficult to process for the computers. A computer may need complex coding and set of tasks to distinguish the Elephant from other animals while it would be rather an easy task for the humans to do so.



Fig 4. Image based Captcha [11]

Advantages of Image Based Captcha are no legibility issue i.e., the user doesn't have to read the text and no specific language is required for this type of captcha. But it is not recommended for the people who are visually impaired. And adding an alternative image will distort the image.

#### F. Logic-Based

This type of CAPTCHA relies on the logical relations present in the world which are basic and requires common sense to solve it. Some examples might be:

- Identify the cars from the list: Mercedes, apple, book
- How many windows are there in a two floor building?
- What is the third letter in the word "exhibition"?
- What part of text will be remaining if you remove B from Blackboard?[12]

Such captchas are difficult to generate and most of the time they are repeated because of limited patterns present in the databases. And they take require noticeable time to solve it. This captcha is beneficial to give you multiple tries to solve the captcha and the user can easily solve the captcha by just looking at the information provided, while not recommended for the people who are visually impaired and also its time consuming.

### G. Graphic CAPTCHAs

Graphic CAPTCHAs consists of pictures or objects that has some common aspect that user has to recognize. In this type of CAPTCHA the computer program will generate a puzzle and accordingly rank the answers. Thus the above process has a disadvantage that it won't be able to solve the puzzle by itself.

In order to overcome the disadvantage it requires a two-step approach. First approach states that the user clicks on some component of the image that is comprised of variety of images and select the image in such a manner that only a single image is selected amongst the other images. Second approach is that once the image selection is done, the image is loaded. After the image gets loaded it gets magnified and the image is distorted. This process results in loading of answers on the UI. Thus, visitant should select an accurate answer from the given set of words. [6]



Fig 5. Graphic Captchas

Graphic Captcha makes it easy for the user to identify the solution from the given set of images and the users can easily interpret the relations between the objects displayed. Whereas the computer itself doesn't have the solution of how to interpret the answer and it cannot accurately determine the answer is correct or not.

### H. AUDIO CAPTCHAS

The AUDIO CAPTCHA is developed for the people who are physically challenged people such as blind people in order to ensure human verification. This captcha work very well for visually disabled person but they have linguistic barriers and they are breakable by using different machine learning techniques.



Fig 6. Audio based captcha [3]

It is meant for the people who are visually impaired and it makes becomes easy to identify the captcha through audio if people couldn't recognize it through scanned text. However, the audio is difficult to interpret sometimes, takes longer time to complete and usability ratio is low and not all the users will be able to recognize the English words.

### I. Ad-Injected CAPTCHA

This kind of Captcha is a normal captcha with just an added characteristic of displaying text/image related to the advertising entity just for the purpose of advertisements. [5]



Fig 7. Ad injected Captcha example

Disadvantages:

- Difficulty in reading.
- Difficulty in using by physically challenged people.
- Tedious process to decipher the cipher text.
- Technical issues due to certain use of Web Browsers.
- Sometimes mathematical equations can be too complex to be solved.
- Too much noise in the audio Captcha.
- The frustration level of the visitants increases if the word is not recognized correctly in spite of numerous chances. [8]

### III. CAPCHAS USED BY SOFTWARE GIANTS

Google uses the reCAPTCHA. A huge number of users can now prove that they are not bots without having to solve a CAPTCHA. Rather they just need to click on a single field which will prove they are humans.



Fig 8. Captcha used by Google

Yahoo uses type of captcha as shown in the figure below in their signup page. It picks a word from the dictionary in a random manner, adds distortion to it and allows the user to guess the correct word.



Fig 09. Captcha used by Yahoo

Microsoft use captchas which are popularly called MSN Passport CAPTCHAs. They use a combination of eight upper case characters and digits. Foreground is dark blue, and background is grey. Conditioning is used to distort the characters, to yield a curling effect, which makes computer recognition very burdensome.



Fig 10. Captcha used by Microsoft

VMware, Inc. is a collateral part of Dell Technologies that provides cloud and virtualization software and services. Image/Picture captchas from VMware consist of 4-8 black digits and uppercase/lowercase letters. To obfuscate the text, characters are compressed, bent and displaced so that they touch or overlap each other.



Fig 11. Captcha used by VMWare [14]

#### IV. CAPTCHAS IMPEDING THE USER ACTIVITIES

Captchas may be a protection against spam bots to prevent unethical usage of free services but it also hinders the user activity in today's fast digitalized world. Certain captchas take a considerable amount of time to be solved or might need a couple of attempts before getting it right. Below is the table which shown the actual time (seconds), accuracy (0-1) and expected time (seconds) taken by the users to solve the captcha on the respective website of the entity. This table clearly signifies how many seconds the user has to invest in order to solve the captcha which is not suitable in today's world.

For example, a user solving a Yahoo audio captcha, which takes on average 25 seconds, and on which users have a solving accuracy of 0.68, the expected solving time is actually 36.8 seconds which is 67% longer than the single captcha time would suggest.

Analysis of the resulting data shows that Captchas are quite troublesome to solve. We also consider the fact that non-native English speakers are comparatively slower.

- **An Easy way to stop spam bots:** It's very straightforward to cease conventional spam bots than a planned attack. There are multiple procedures and techniques to stop such bots, but most uncomplicated way is to add an extra field with some tricky different name to the form and make it hidden using appropriate CSS and HTML. No human will be able to view such a field and won't fill it in an obvious manner. But the pre-instructed bot will just look at the script and tags and fill the form completely. This is the simplest method to stop most of the spam bots without adding any friction and negative user experience.

Table I. Time taken by users to solve the Captchas of various originations

Scheme	Time	Accuracy	Expected time
Authorize	6.8	0.98	6.9
Baidu	7.1	0.93	7.6
Captchas.net	8.2	0.84	9.8
Digg	8.2	0.92	8.9
eBay	7.3	0.93	7.8
Google	9.7	0.86	11.3
mail.ru	12.8	0.7	18.3
Microsoft	13	0.8	16.3
Recaptcha	11.9	0.75	15.8
Skyrock	7.9	0.95	8.3
Slashdot	7.7	0.87	8.8
Blizzard	9.3	0.95	9.8
Yahoo	10.6	0.88	12
Authorize audio	11.9	0.59	20.2
Digg audio	14.8	0.38	39
eBay audio	11.8	0.63	18.8
Google audio	35.2	0.35	100.6
Microsoft audio	16.6	0.38	43.8
Recaptcha audio	30.1	0.47	64.1
Slashdot audio	11.7	0.68	17.2
Yahoo audio	25	0.68	36.8

[7]

## V. DE-CAPTCHA UTILITY

Nowadays, CAPTCHAS are easily conquered by many software's which doubt the acceptance of Captcha. One can just hire or pay a small amount of money to solve the Captchas instead of undertaking and developing time consuming and expensive software's including thousands of lines of code. Also, the accuracy of such Captcha solvers is high. For example, Death by Captcha provides easy usable APIs for users to have an easy way to overcome the friction caused by Captchas. [10]

## VI. APPLICATION OF CAPTCHA

### 1) Online Polls

Spam bots can build a catastrophe to any unsecured online poll. These bots can increase the vote count falsely to impersonate a wrong poll winner. This may lead to a downfall in the assurance of such polls and the trust in them. Captchas can also be used in securing website from the accessibility of the bots increasing its authenticity of such polls.

### 2) Safeguarding Web Registration

Multiple firms present policies and utilities by offering unpaid online services. Bots take advantage of such services by making accounts inordinately and use of services in an inappropriate manner. Captchas can be used to safeguard such services from Bots which may assure that only Humans are allowed to utilize their services ethically. Because such Bots may lead to overburden and ill-management of those accounts and services.

### 3) Avoid comment spamming

Many websites and bloggers try to add automatic comments and post using bots in order to uplift their search engine rank so that users can view them easily. Captchas can be used to stop such activities by obstructing the bots from spam commenting and posting and only pass the legit human users to add comments.

### 4) E-Ticketing

Agent and brokers use application which makes a huge number of ticket purchase for some event or request. This makes the legitimate users a sufferer as most or all the tickets just sell out within few seconds. Brokers then try to sell these tickets at high rates.

5) **Email spam**

Captchas provide a solution to the issue of email spams. Just to do was to use a Captcha test to find out that a human has sent the mail. [6]

## **VII. DIFFERENT MODES OF ATTACK**

### **A. Bypass attacks**

This form of attack involves any way that intercepts the requirement to solve the Captcha. Take an example of Network Replay Attack. System that sends decrypted form of captcha to a program as the component of data sent is defenseless to such an attack.

### **B. Challenge replay attacks**

Assume that a captcha system can produce only a limited number of different challenges, then the agent may present possible solutions to those challenges. A collaborator may provide a collection of such answers and then these agents may seek the correct answer to the provided challenge as and when required. Image based captchas are most vulnerable to these attacks as they are generally limited in number.

### **C. Signal processing attacks**

The distortions produced in the normal text and images by the computers have been easily reversible by the bots and attackers. This makes such a distorted captcha very easy to crack and that too with high degree of accuracy. These reversing of Captchas can be done using OCRs, heuristics and other machine learning methods.

### **D. Mechanical Turk attacks**

In this form of an attack, the difficulty of solving the captcha is redistributed to a paid human agent. They solve the challenge instantly and reply the solution to the user. One can prevent this attack by increasing the difficulty of the captchas but this may cause more frustration in the users which may finally lead to the assurance of Captchas.

### **E. Trivial guessing attacks**

If the number of challenges are more but the answer to those challenges are limited then this can be easily exploited by just guessing in any way from the available possible answers. Captchas that have a higher error tolerance for users are more susceptible to such an attack.

### **F. Brute force attacks**

If there are a finite number of achievable answers then it is possible for a distributed collection of agents to attack the captcha in different ways iteratively to exploit it in a faster manner. For example, a three digit captcha would have 1000 possible answers then the automated bots would work upon it and answer it efficiently. This is different from the previous attack, the difference is that in this attack it has a support of a great number of agents rather than a captcha designed in an ill-manner.

### **G. Hybrid attacks**

It is very much efficient to combine any of the above attacks for faster retrieval of the answer to the challenge. For example, three of the five characters are guessed correctly with a high degree of accuracy then the remaining two letters can be guessed using any guessing or brute force method. [9]

## **VIII. CONCLUSION**

After a brief description about CAPTCHA and its various types. This paper gives the reader an example of DeCAPTCHA services which are presently used around the world with its advantages as well as disadvantages, it helps the readers to understand different uses and threats to CAPTCHA. Also, helps the researchers to work on better user



friendly captchas which are less complex and more secure. Although efforts have been made to cutback the duration of the users by experimenting on different types of captchas and making it easier for the system to identify the difference between a legit user and a spam bot, there still needs advancement in making the CAPTCHAS more user friendly.

## REFERENCES

- [1]Introduction to Captcha(2017, Jan, 21). Eukhost. [Online]. Available:<https://www.eukhost.com/blog/webhosting/introduction-to-captcha/>
- [2]CAPTCHA(2017, Jan,22). Wikipedia. [Online]. Available: <https://en.wikipedia.org/wiki/CAPTCHA>
- [3]What is the purpose of CAPTCHA technology(2017, Jan, 26).searchsecurity.techtarget.[Online] Available:<http://searchsecurity.techtarget.com/answer/What-is-the-purpose-of-CAPTCHA-technology>
- [4]Different types of CPATCHA and its userability(2017, Jan, 31).ScrapeSentry.[Online]. Available: <https://www.scrapesentry.com/scraping-wiki/different-types-captcha-userability/>
- [5]examples of different typesof CAPTCHAs(2017, Feb, 3).findexamples.[Online]. Available: <http://www.findexamples.com/5-examples-of-different-types-of-captchas/>
- [6]Sarika Choudhary, Ritika Saroha, Yatan Dahiya, Sachin Choudhary” Understanding Captcha: Text and Audio Based Captcha with its Applications”IJARCSSE Volume 3, Issue 6, June 2013 ISSN: 2277 128X [https://www.ijarcsse.com/docs/papers/Volume\\_3/6\\_June2013/V3I6-0103.pdf](https://www.ijarcsse.com/docs/papers/Volume_3/6_June2013/V3I6-0103.pdf)
- [7]Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, Dan Jurafsky “How Good are Humans at Solving CAPTCHAs? A Large Scale Evaluation”[https://web.stanford.edu/~jurafsky/burszstein\\_2010\\_captcha.pdf](https://web.stanford.edu/~jurafsky/burszstein_2010_captcha.pdf)
- [8] Captcha: Advantages and Disadvantages(2017, Feb,15)blogspot[Online] Available: <http://lanena19-informationtechnology.blogspot.in/2010/10/captcha-advantages-and-disadvantages.html>
- [9]Strengthening Captcha based web security(2017, Feb,23)firstmonday[Online]. Available: <http://firstmonday.org/ojs/index.php/fm/article/view/3630/3145>
- [10]Everything You Ever Wanted To Know About CAPTCHAs But Were Afraid To Ask (2017, Feb,28)makeuseof. [Online]. Available: <http://www.makeuseof.com/tag/wanted-captchas-afraid-technology-explained/>
- [11]Another image-based Captcha method(2017, Mar,3).cnet. [Online]. Available: <https://www.cnet.com/news/another-image-based-captcha-method/>
- [12]The Story of Captcha and its Variants(2017, Mar,8). Tutorialspoint[Online]. Available: <http://www.tutorialspoint.com/articles/the-story-of-captcha-and-its-variants>
- [13]Image(2017, Mar,15)Microsoft.[Online]. Available: <https://signup.live.com/signup>
- [14] Image(2017, Mar,15)VMWare.[Online]. Available: <https://my.vmware.com/web/vmware/registration>