# A Survey on CAPTCHA as Graphical Password

Priti P. Doke<sup>1</sup>, S.A Nagtilak<sup>2</sup>

<sup>1</sup>Department of Information Technology, Smt. Kashibai Navale College of Engineering Pune, India

<sup>2</sup>Professor, Department of Information Technology, Smt. Kashibai Navale College of Engineering Pune, India

Abstract: There are various challenges to be faced to provide security to Web applications. To reach the backend problems attackers use information that is not validated before by any Web applications. Unauthorized functions are used to access to other accounts. There are various injection flaws in which malicious commands are embedded in parameters and the external system executes the commands accordingly on behalf of the Web application. Improper coding results in weak protection of web applications that use cryptographic functions to protect information and credentials. Thus, in various ways access to the Web application is provided with security. Applications are such that they are robust against input data of all forms be it user, database or any other external entity. CAPTCHA stand for Completely Automated Public Turning Test to Tell Computers and Humans. It is a test that authenticated users have to pass the test to gain access to their respective mail accounts. Malicious programs like bots, attack the accounts that are a threat to the data integrity, privacy, and confidentiality. To avoid this, CAPTCHA was introduced. Here all type of CAPTCHA and their drawbacks are going to describe.

Keywords: CAPTCHA, HIP, OCR, SPC.

### 1. Introduction

CAPTCHA was first invented in 2000 at Carnegie Mellon University by John Langford, Nicholas J. Hooper and Luis Von Ahn [1]. CAPTCHA stand for "Completely Automated Public Turning Test to Tell Computers and Humans Apart" [2]. The progress of Web security and the internet has become an important issue. There are many malicious threats across the Internet which may compromise your system in the absence of any secure application that provides protection against such threats. One of such threat is the Bot. A Bots are the malicious program that has the capability to run automated tasks over the network and thus creates a problem in the network [3]. CAPTCHA is one such shield that can be used as a protection from these malicious programs like Bot.

CAPTCHA is the defensive system that acts as a tool to check web Bots from exploiting online services on the Internet including email provider, wikis, blogs, etc. It is an HIP system that is widely used to secure the internet based applications that are also called as a challenge-response test that gives a challenge to the users when the user gives correct response he/she is considered as human otherwise a web bot. CAPTCHA is an authentication process based on challenge-response authentication.

Different web services like Yahoo, Google, and Bing, etc., use CAPTCHA to differentiate between a malicious program and an authenticated user. The sites that provide access to sensitive data, such credit card account, and banks also use CAPTCHA.

## 2. Categories of CAPTCHA

CAPTCHA means to present a challenge-response test to the users or humans. They are classified based on what is distorted that is whether it is characters, digits, or images. Some types of CAPTCHAs are given below:

1. CAPTCHAs based on the text.

2. CAPTCHAs based on audio.

- 3. CAPTCHAs based on a video.
- 4. CAPTCHAs based on an image.

### A. CAPTCHAs based on the Text:

Text-based CAPTCHAs is simple to implement. It is effective and requires a large question bank. In Text-based CAPTCHA, different classes of characters and digits are included. It included alphabets with lower and upper cases and digits from 0-9.



Figure 1: Text-based CAPTCHA

The problem with CAPTCHAs based on Text: In text images, user have some problem to identify the correct text or characters like Multiple Fonts, Font size, Blurred Letters, Wave Motion. And also the text-based CAPTCHA is possible to identify through Optical character recognition (OCR) technique

#### **B. CAPTCHAs based on Audio**

These CAPTCHAs is based on the sound-based systems. These CAPTCHAs are developed for users who are visually disabled. It contains an audio-clip. In this type of CAPTCHA, the user has to listen and after that submits the spoken word [1]. The sound-based first system name ECO was implemented by the Nancy Chan, from the Hong Kong University. This system is based on the difference in the ability between computers and humans in recognizing the language spoken. The CAPTCHA chooses a sequence of digits and words randomly and renders the number and words in a sound clip and distorts it. The distorted sound clip is then given to the user to enter the correct number or words. The user is to enter the same words as which are spoken in the audio clip



Figure 2: Audio based CAPTCHA

The problem with CAPTCHAs based on Audio: It is available in English. Therefore, the end user must have a comprehensive English vocabulary and user may get confused with a character that have a similar sound.

## C. CAPTCHAs based on Video

Video CAPTCHA is less commonly seen in CAPTCHA system. In video-based CAPTCHAs, three words are provided to the user that describe a video. The use tag must match to a set of automatically generated ground tags then test taken by the user is said to be passed. The term video CAPTCHA is used to any CAPTCHA that uses a video as its means to present information to a user. Although video CAPTCHA is limited, in both commercial and academic application that are existing.



Figure 3: Video based CAPTCHA

Problem with CAPTCHAs based on Video: Due to large size of file, users face problem to download video and find correct CAPTCHA

## **D. CAPTCHAs based on Image:**

Image-based CAPTCHA is also known as graphic based. Graphics-based CAPTCHAs are challenge tests in which the users have to guess the images that have some similarity. In graphic based CAPTCHAs user is required to identify the image. The main advantage of CAPTCHA based on the image is that pattern recognition is hard AI problem and. Therefore, it is not easy to break this test using pattern recognition technique.



Figure 4: Image-based CAPTCHA

The problem with CAPTCHAs based on image: Some users face the problem of image identification who have low vision or even due to the color blindness of images. The human mind sometimes perceives patterns even when they don't exist, a quirk we call pareidolia.

# 3. Literature Survey

A. Work by R. Biddle [5]. In this Starting around 1999, a great many graphical password schemes have been proposed as alternatives to text-based password authentication. They provide a comprehensive overview of published research in the area, which covers both usability and security aspects, as well as system evaluation. The paper first catalogs existing approaches highlights novel features of selected schemes and identifying key usability or security advantages. They then usability requirements for knowledge-based review authentication as they apply to graphical passwords, identify security threats so such systems must address and review known attacks, discuss methodological issues related to empirical evaluation, and also identify areas for further research and improved methodology.

B. Work by S. Benson Edwin Raj et al. [4]. Denial of Service is a common threat to network security and is also considered to be automated network attack. To prevent the system from such kind of attacks to identify the difference

## International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

between the legitimate user and the fake one CAPTCHA shield must be implemented in the system. These days OCR based CAPTCHAs are more vulnerable as many algorithms are available on the internet with which these CAPTCHAs can easily be broken. There is a way to make the CAPTCHA text more distorted, although it has its disadvantage as the user is not able to recognize this kind of CAPTCHAs. This problem was overcome by introducing a new CAPTCHA called as graphics or picture or image-based CAPTCHA. The CAPTCHA based on images has advantages as any malicious program can not perform any edge detection for segmentation and thresholding, shape matching and random guessing. In the security analysis process, this mechanism has shown better results.

C. Work by Aditya Raj [3]. The CAPTCHAs, which based on OCR, have exploited are thus insecure while CAPTCHAs based on non-OCR are safe as they employ natural skill of the person's eye to identify the image. They incorporated the concept of Sequencing in image CAPTCHA to build a new concept known as SPC. SPC generation can be classified into two different types that are non-inherent sequencing and inherent sequencing. SPCs have two levels of security; the first one is the recognition of objects in the image and the second one is to determine their logical sequence.

D. Work by JingTing Mei [6]. In this paper, a new CAPTCHA is a mention that is based on tracking problems and moving object identification. Based on the Edge Mutation an innovative Single-frame Zero-knowledge rule is also put forward to the CAPTCHA generation algorithm. It refers to biological motion vision model. After solving a moving object recognition problem successfully an attacker can access the service test system. This kind of animated CAPTCHAs are not only able to check the attacks based on static OCR technology but also can check the attacks against the moving object. In this paper, they have included three kinds of programs: the visual programs based on OCR problems, the non-visual programs, and visual programs based on non-OCR problems. It most widely used and applied program is based on the visual programs based on the visual programs based on the OCR problems with the advantage of implementation and operation. Whenever the user makes a request to the server, the server responds to the user with an image containing a string of random characters and numbers. The user has to identify the sequence of characters to get access to the server resources.

E. Work by Chen-Chung Hsieh [9]. In this paper, by using embedding versatile characters in the images, they proposed an innovative image-based CAPTCHA for distinguishing human and computer and in a method that they offered by automated image analysis technologies like scale-invariant feature transform it makes the characters invisible while a human can easily distinguish the location of the embedded characters. Their designed mechanism was capable of eluding such generous of attacks. For in experiments, to test the system about 15 users were invited, and found the rate of success is 86%. The success rate reached 95% if wrong operations like clicking out of text boxes were excluded. The proposed method is better and faster than the compared average logging time with reCAPTCHA and hello CAPTCHA, two methods by 32 seconds and 115 seconds, respectively. In this, they have proposed random number generation module is used to select image and verification characters for conducting man-machine distinguishes. Then size, color, style, and angle variation are applied to the verification characters. Finally, the verification image is generated.

F. Work by Gaurav Goswami.[8]. In this paper, a new image based CAPTCHA is used for distinguishing human and bots CAPTCHAs using facial features, facial expressions and face recognition are brought out. A new type of CAPTCHA called as Collage CAPTCHA is introduced in where different distorted textures, lightings, colors, poses and arranges them on a cluttered background. The user has to identify only two similar faces matching each other with the expressions and facial features. The lightings, the cluttered background, makes the image more confused, and if the dark colored cluttered background is used then, the faces on the imag.es can get a user confused or even non-recognizable.

## 4. Table of Comparison

## International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

1	CAPTCHA as Graphical	Click Text and Animal Grid	It is better than pervious	Length of this CAPTCHA is hig
1	Passwords A New	using ASP NET	text CAPTCHA as it	Length of this erth Territ is big
	Security Primitive Based on	In this CAPTCHA only conital	avoid the L O L and 7"	
	Hard AI Problems [10]	latters are used. That are 33	$\frac{1}{1}$ avoid the $\frac{1}{1}$ , 0, 3 and $\frac{1}{2}$ of	
	Hard AI Hobienis.[10]	characters capital letters excent	the digit	
		L L O and Z digits execut 0	tile digit.	
		i, J, O, and Z, digits except 0		
		and 1, and three special		
2	A Case Study of Text	Click Text is a managemitian	Tayt Daged CADTCILA	Use of various lines, use of multiple forts and
2	A Case Study of Text-	Click Text is a recognition-	is simula to implement	Cise of various lines, use of multiple fonts and
	Based CAPTCHA	based CaRP scheme that is built	is simple to implement	fond sizes of fetters and digits together, a user
	Attacks[/]	on top of text CAPTCHA.	so it is mostly used in	laced problems to enter the correct text or
			Tart has a CADTOUA	character or letter. This type of CAPTCHA is
			Text-based CAPTCHA	easily broken by OCK (Optical character
			is used for dictionary	recognition)technique
2			defect attack	
3	An Audio CAPTCHA to	Based on the sound-based	It is used for people that	The user has to understand the language and the
	Distinguish Humans from	systems. The user is asked to	have visual impairment.	accent in which the sound clip is recorded, Not
	Computers.[2]	enter the same words as spoken		working for dumb people or people that have low
		the audio clip.		listening power.
4	Balancing usability and	CAPTCHA that uses a video as	it provides greater	The size of files is large
	security in a video	a means to present information	security than Text-	
	CAPTCHA.[1]	to a user.	Based CAPTCHA and	
			Audio-Based	
			CAPTCHA.	
5	Anti-SIFT Images Based	This is a CAPTCHA scheme	There is no need of	User may fail to identify all the similar images
	CAPTCHA Using Versatile	that uses models of flowers and	typing. Can't be broken	have to select all images to have same animal or
	[6]	animals to generate the 2D	by OCR (Optical	model
		image with different, colors	character	
		textures, lightings, and poses,	recognition)technique	
		and arranges that images.		
6	Face Recognition	Face CAPTCHA is a	The user has to identify	A user clicks, all the same, faces is a challenge
	CAPTCHA[5]	CAPTCHA scheme that uses	only two same faces.	image to pass the test, lightings, and cluttered
		models of face to generate the		background can get a user confused
		2D image with different colors,		
		textures, lightings, and poses,		
		and arranges them on a		
		cluttered background.		

# 5. Application of All Types of Captcha

There are some applications of CAPTCHA on the web that are defined as follows.

## A. Registering the web forms:

There are many sites on the Internet that provide free registration to available their services. They are vulnerable to the web bots. It can come into the form of scripts that can register about thousands of email accounts on the internet, thus wasting the space of the web.

## **B. E-Ticketing.**

### C. Preventing Dictionary Attacks and E-mail spam

#### **D. Online Polling sites:**

These sites in the form of questionnaires take user's response or feedback. To ensure not other than human makes the response they make use of CAPTCHA.

#### E. To avoid web crawling:

Here the use of CAPTCHA is done if a site doesn't want to get indexed by a search engine.

- F. Preventing Comment Spam in Blogs
- G. Protecting Email Addresses From Scrapers
- H. Protecting Website Registration

## 6. Conclusion

A survey has been done on the different kinds of CAPTCHA developed till now like Text, Audio, Video, and Image CAPTCHA. The problem found in audio was that the language of Audio clip should be common, and also, user gets confused with a character that has a similar sound. In Video CAPTCHA the problem of loading that video clip that required more time. In Text CAPTCHA due to the uppercase, lowercase and also included digit the user usually get confused in some characters; the Text CAPTCHA can be cracked by bolts that again lead to a security problem. In the Image, CAPTCHA users face the problem of image identification who have low vision or even due to the color blindness of images. The Image CAPTCHA is found much secured then the audio and video CAPTCHA by this survey. The survey also states the applications of CAPTCHA.

## References

- Kurt Alfred Kluever, Richard Zanibbi "Balancing usability and security in a video CAPTCHA," SOUPS '09: Proceedings of the 5th Symposium on Usable Privacy and Security. July 2009.
- [2] Haichang Gao, Honggang Liu, Dan Yao; Xiyang Liu, Aickelin, U. "An Audio CAPTCHA to Distinguish Humans from Computers, "; Third International

Symposium on Electronic Commerce and Security (ISECS), pp. 265 – 269, 2010.

- [3] Aditya Raj, Ashish Jain, Tushar Pahwa and Abhimanyu Jain –Picture CAPTCHAs With Sequencing: Their Types and Analysis," International Journal of Digital Society, vol. 1, no. 3, pp. 208-220, 2010.
- [4] S. Benson Edwin Raj, Deepa Devassy and Jiji Jagannivas — ANew Architecture for the Generation of Picture Based CAPTCHA," IEEE, pp. 67-71, 2011.
- [5] R. Biddle, S. Chiasson, and P. C. van Oorschot, Graphical passwords: Learning from the first twelve years," ACM Comput. Surveys, vol. 44, no. 4, 2012.
- [6] JingTing Mei, JingSong Cui, Da Zhang, LiJing Wang, Xia Wang, Yang Peng, WuZhou Zhang —@APTCHA Design Based on Moving Object Recognition Problem," in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, 2012.
- [7] Xiao Ling-Zi and ZHANG Yi-Chun —ACase Study of Text-Based CAPTCHA Attacks," in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, 2012.
- [8] Gaurav Goswami, Richa Singh, Mayank Vatsa –*Face Recognition CAPTCHA*," IEEE, pp 412-417, 2012
- [9] Chen-Chung Hsieh and Zong-Yu Wu —Ati-SIFT Images Based CAPTCHA Using Versatile," IEEE, 2013.
- [10] Bin B. Zhu, Jeff Yan, Guanbo Bao, Maowei Yang, and Ning Xu —Capeha as Graphical Passwords—A New Security Primitive Based on Hard AI Problems," IEEE, pp 891-904, 2014.