CHAPTER I: INTRODUCTION

CAPTCHA—or "Completely Automated Public Turing test to tell Computers and Humans Apart" is a security feature that can prevent the use of websites that may be unlawful. It invented in 2000 at Carnegie Mellon University by John Langford, Nicholas J. Hooper and Luis Von Ahn. It is a type of test that is located on websites where a user needs to put in a series of letters and numbers or click certain parts of a picture in order to complete a task. That task might be purchasing a product, opening an account, posting a review, or any other kind of task. Web developers use this on sites to make sure the being using the website is indeed a human with good intentions.

The progress of Internet, Web security has become an important issue. There are too many malicious threats across the Internet which may compromise your system in the absence of any secure application which provides protection against such threats. One such threat is the Bot. A Bot is a malicious program which has the capability to run automated tasks over the network and thus creating problem in the network. CAPTCHA is one such shield which can be used as a protection from these malicious programs like Bot. A CAPTCHA may come in various forms like text based or image-based CAPTCHA. The Bot operation is similar to reverse "TURING TEST" where the program acts like judge and the other person acts like user. If the user fails this test then he/she is considered to be a machine otherwise the user is considered to be an authentic user or a human being. Often times, the word-solving image is distorted, which makes it difficult to read. While its purpose is to fool a robot or computer from being able to read it, it can also sometimes be difficult for humans to read. In fact, some people feel that CAPTCHAs can inhibit disabled users from being able to access certain parts of a website. While CAPTCHAs may not be able to prevent all types of spam on the internet, it certainly does a good job. Without it, there would be so much more negativity and frustration happening all over the web.

CAPTCHA is a defensive system that acts as a tool to check web Bots from exploiting online services on the internet including free email providers, wikis, blogs

etc. It is a HIP system that is widely used to secure the internet-based applications. It is also called as a challenge response test which gives a challenge to the users, when the user gives correct answer he is considered as human otherwise a web bot. CAPTCHA is an authentication process based on challenge-response authentication. CAPTCHA provides a mechanism with the help of which a user's can protect themselves for spam and password decryption by taking a simple test. In this test a user will see either an image or a text which are normally distorted. The user is supposed to enter the pattern exactly as shown to him if the CAPTCH is based on text. If the CAPTCHA is based on image the user is supposed to enter the correct name of the image which correctly symbolizes the image. CAPTCHA is used where authenticated access is the primary concern. Various web services like Yahoo, Google, and Bing etc. use CAPTCHA to differentiate between an authenticated user and a malicious program. CAPTCHAs are also used in the sites which provide access to sensitive data, such credit card accounts and banks.

CAPTCHA is important to many internet empires in order to prevent a robot from manipulating services. Some services that robots can manipulate would be opening too many accounts, getting a lot of sensitive information, or posting too many messages, which can hinder web analytic results.

There are some people out there who might want to solve CAPTCHA codes in order to participate in some of the above-mentioned acts on a website. As with anything in life, there are two sides to every story. Some companies make their money solving CAPTCHA codes and selling them to people who are interested in using them. Solving CAPTCHA is useful for people who want to manipulate websites opening multiple accounts, sending spam messages, and trolling the internet.

CHAPTER II: LITERATURE REVIEW

A SURVEY AND ANALYSIS OF CURRENT CAPTCHA APPROACHES, NARGES ROSHANBIN and JAMES MILLER(2012)

Computer programs are misusing Internet services designed for humans. A CAPTCHA, Completely Automated Public Turing test to tell Computers and Humans Apart, is a standard security mechanism to defend against such attacks. Two fundamental issues with CAPTCHAs are usability and robustness. It is important for a CAPTCHA to be both legible for humans and strong against malicious computer programs. Recently, computer vision and pattern recognition algorithms have broken many well-known CAPTCHAs. Lack of security and usability in CAPTCHAs designed to protect popular websites such as Gmail and Yahoo mail, with almost 500 million users in July 2011, would cause huge problems. Therefore, security researchers have become motivated to discover techniques to improve CAPTCHAs. Exploiting the gap in the recognition abilities between humans and computers is a key point to design a CAPTCHA that is hard-to-break for machines but easy-to-solve for humans.

Review Paper on Different CAPTCHA Techniques, Anvesh Sinha, Dr. Sandhya Tarar(2016)

CAPTCHA is a program or a system that protects against automated scripts (or bots). It generates tests that humans can pass but computer programs cannot. CAPTCHA systems are widely used nowadays for protecting and providing security to internet based services for humans from abuse by bots. Different types of CAPTCHA technologies are discussed in this paper and a detailed analysis on their reliability is performed. Subsequently, a new CAPTCHA technique is proposed which is based on facial expression detection.

Survey of captcha, Ved Prakash Singh, Preet Pal,School of Computer Science, Lovely Professional University Phagwara, Punjab(2013)

CAPTCHA is an acronym for Completely Automated Public Turning Test to tell Computers and Humans Apart . CAPTCHA is basically used as a protection from these malicious programs like Bot. Now aday's for web security we are using different type of captcha.In this paper we describe all type of captcha and also

describe their drawbacks all type of captcha. We also describe application of captcha. And review paper of different types of Captcha.

Text-based CAPTCHA strengths and weaknesses, Elie Bursztein ie Bursztein Matthieu Martin(2011)

We carry out a systematic study of existing visual CAPTCHAs based on distorted characters that are augmented with anti-segmentation techniques. Applying a systematic evaluation methodology to 15 current CAPTCHA schemes from popular web sites, we find that 13 are vulnerable to automated attacks. Based on this evaluation, we identify a series of recommendations for CAPTCHA designers and attackers, and possible future directions for producing more reliable human/computer distinguishers.

What's up CAPTCHA?: a CAPTCHA based on image orientation, Rich Gossweiler, Shumeet Baluja (2003)

We present a new CAPTCHA which is based on identifying an image's upright orientation. This task requires analysis of the often complex contents of an image, a task which humans usually perform well and machines generally do not. Given a large repository of images, such as those from a web search result, we use a suite of automated orientation detectors to prune those images that can be automatically set upright easily. We then apply a social feedback mechanism to verify that the remaining images have a human-recognizable upright orientation. The main advantages of our CAPTCHA technique over the traditional text recognition techniques are that it is language-independent, does not require text-entry (e.g. for a mobile device), and employs another domain for CAPTCHA generation beyond character obfuscation. This CAPTCHA lends itself to rapid implementation and has an almost limitless supply of images. We conducted extensive experiments to measure the viability of this technique.

Usability of CAPTCHAs or usability issues in CAPTCHA design, Ahmad Salah El Ahmad, Jeff Yan(2008)

CAPTCHA is now almost a standard security technology, and has found widespread application in commercial websites. Usability and robustness are two fundamental issues with CAPTCHA, and they often interconnect with each other. This paper discusses usability issues that should be considered and addressed in the design of CAPTCHAs. Some of these issues are intuitive, but some others have subtle implications for robustness (or security). A simple but novel framework for examining CAPTCHA usability is also proposed.

CAPTCHA: Using Hard AI Problems for Security, Luis von AhnManuel BlumNicholas J. HopperJohn Langford(2003)

We introduce captcha, an automated test that humans can pass, but current computer programs can't pass: any program that has high success over a captcha can be used to solve an unsolved Artificial Intelligence (AI) problem. We provide several novel constructions of captchas. Since captchas have many applications in practical security, our approach introduces a new class of hard problems that can be exploited for security purposes. Much like research in cryptography has had a positive impact on algorithms for factoring and discrete log, we hope that the use of hard AI problems for security purposes allows us to advance the field of Artificial Intelligence. We introduce two families of AI problems that can be used to construct captchas and we show that solutions to such problems can be used for steganographic communication. CAPTCHAs based on these AI problem families, then, imply a win-win situation: either the problems remain unsolved and there is a way to differentiate humans from computers, or the problems are solved and there is a way to communicate covertly on some channels.

Survey of Text CAPTCHA Techniques and Attacks, Simran Sharma, Nidhi Seth(2015)

CAPTCHA is a challenge-response test most often placed within web forms to determine whether the user is human. The purpose of CAPTCHA is to block automated scripts that post spam content everywhere they can. This paper presents a survey of Optical Character Recognition applications and further focuses on three important applications of Optical Character Recognition, CAPTCHAS. There is a

constant need to improve current CAPTCHAs and to develop new CAPTCHAs in order to safely secure against developing programs which can create thousands of e-mail accounts used for malicious purposes, stuff online polls with ballots, and develop worms and viruses contained in emails

,

CHAPTER III: AIM AND OBJECTIVE

Aim

Comparison between different types of Text based Captcha.

Objectives

- To identify that which captcha is more efficient.
- To know that which captcha is more secure and useful.
- Comparison of different types of text captchas.
- Success Rate in Attacks.

CHAPTER IV:MATERIALS AND METHODOLOGY

Materials Required:

Computer or laptop, different types of -text based captcha.

Methodology:

Firstly, the different types of text captchas are taken and are grouped under their own categories. Then the comparison of the various text captchas are done (CAPTCHAs with English words, CAPTCHAs with random strings, CAPTCHAs based on handwritten text).now the results are noted and what is the, success rate in attacks of ransomware and web attacks are analysed.

CHAPTER V: OBSEVATIONS

Text-based CAPTCHAs

A text-based CAPTCHA is a distorted image of a sequence of characters on which different types of degradations, background clutters and colour mixtures are applied to make it more challenging for attackers. We will introduce current text-based CAPTCHAs into different sub-groups such as:

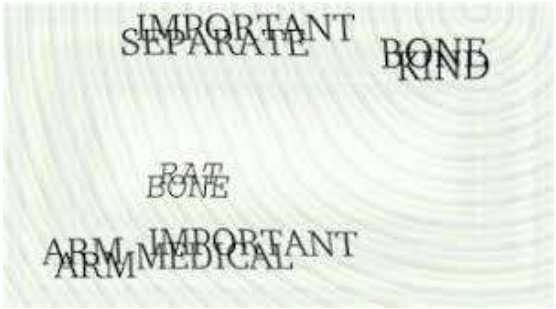
- CAPTCHAs with “English words”

- CAPTCHAs with “random strings’’

- CAPTCHAs based on handwritten text

CAPTCHAs with “English words” as their CAPTCHA text: In some CAPTCHA systems, such as Gimpy, EZ-Gimpy, CaptchaService.org, PessimalPrint and reCAPTCHA, the CAPTCHA image contains English word(s).

1.Gimpy: Gimpy (Figure 1). is one of the most famous CAPTCHAs which are primarily based on distorted text This CAPTCHA was developed in collaboration with Yahoo with the aim of protecting chat rooms from spammers to make them unable to post classified ads and write scripts to generate free e-mail addresses. Gimpy picks seven words from a dictionary; then renders a distorted image containing those words. It finally presents them to its users and asks them to type three of the words of the image to gain entry to the service



(Figure 1)

2.EZ-Gimpy : In this CAPTCHA(Figure 2). the word is rendered to an image using various fonts; and different types of distortions such as black or white lines, background grids and gradients, blurring and pixel noise are added Then, the user is asked to type the word

(Figure 2).

3.PessimalPrint: PessimalPrint (Figure 3) concentrates on degradations, such as adding noise to or blurring the images to defeat OCR techniques; the designers of this CAPTCHA argue that under the conditions of inferior image quality, there is a significant gap in pattern recognition ability between humans and machines [2]. This CAPTCHA works as follows. First, a word is pseudo-randomly selected from a fixed list containing 5-to-8-letter English words. Then, it is rendered with a typeface (from a fixed list of 5 fonts) and a fixed font size (size=8). Finally, a set of image degradations including x-scaling, y-scaling, skewness, blurriness and adding noise are applied to the image.

(Figure 3)

CAPTCHAs with "random strings" as their CAPTCHA text: using English words in some current CAPTCHAs makes them vulnerable to dictionary attacks. The solution for this issue is exploiting random strings instead of words. This technique is utilized by MSN CAPTCHA, Yahoo, Ticketmaster, Google, etc.:

1.Hotmail or MSN CAPTCHAs(Figure 4): This CAPTCHA used in the Hotmail email service registration, selects eight English characters (upper case letters and digits); then, after applying local and global warping, renders the characters with dark blue colour on a light grey background. In the next step, three types of arcs are randomly added to make segmentation difficult. The arcs include: "Very thick arcs" (the same as the characters) of foreground colour that do not intersect

characters, “Thick arcs” of foreground colour that intersect characters , and “Thin arcs” of background colour that cut characters and remove some of their pixels.

(Figure 4)

2.Yahoo! CAPTCHA (Figure 5): Starting in August 2004, Yahoo! introduced its second-generation CAPTCHA. Its characteristics include using a string of characters instead of English words, containing only black and white colours, using both letters and digits, and having connected lines and arcs as clutter.

(Figure 5)

3.Google/Gmail Captcha (Figure 6): The specifications of this CAPTCHA, used by Gmail.com, include: using only image warping for character distortion, having only two colors (one for foreground and the other for background), locating characters close to each other and following a curved baseline.

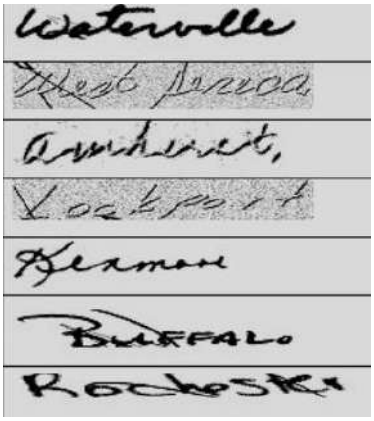
(Figure 6)

4.ScatterType (Figure 7): Scatter Type CAPTCHA selects its challenge word from a set of 15,000 English- like nonsense words. Then the algorithm applies a font (from 100 different font types) to it. The image of each character is fragmented using horizontal and vertical cuts and fragments are forced to drift apart until it is difficult to resemble them into characters. In order to achieve the human legibility, some characters with highest confusability (‘q’,’c’,’I’,’o’,’u’) are removed.

(Figure 7)

CAPTCHAs based on handwritten text: While most current text-based CAPTCHAs use machine-printed text, which makes them vulnerable to pattern recognition attacks, there are CAPTCHAs that use handwritten text in their challenges. An example is Handwritten CAPTCHA.

Handwritten CAPTCHA (Figure 8): This CAPTCHA is based on distorted handwritten text. The authors of this article discussed that according to Gestalt laws of perception and the Geon theory of pattern recognition, the interpretation of distorted handwritten text is easy and reliable for humans but difficult for automated programs In human perception of occluded images of words, decomposing a word into known and unknown visual elements allows users to recognize characters by using rules such as: “words contain specific visual elements”, “combination of letters follows specific rules” and “words convey meaning” to recognize the entire word . The designers of this CAPTCHA investigated the effects of different transformations including overlapping, adding occlusions, splitting the image in parts and displacing the parts, changing word orientation, etc. on the usability and security of their CAPTCHA They designed an algorithm to generate cursive English handwritten text synthetically. They used existing character images and performed auto-scaling, automatic baseline determination, ligature parameterization, ligature joining, skeleton perturbation and skeleton thickening to generate synthetic handwritten words


Waterville
West Seneca
Amherst,
Lockport
Clarence
BUFFALO
ROCHESTER

(Figure 8)

CHAPTER VI : RESULT AND CONCLUSION

Result

On the above observation the we can find the success rate of captchas from the attacks and the comparison of text captchas are given below

Conclusion

From the analysis we can conclude the following

1. It will be helping in the security of the websites
2. Maintaining the illegal attacks against the website
3. Efficiency of captchas
4. More useful captchas

S. No.	Type of CAPTCHA	Success rate of attacks (in %)
1	Gimpy	92
2	EZ - Gimpy	92
3	PessimalPrint	40
4	Hotmail or MSN CAPTCHAs	92
5	Yahoo! CAPTCHA	57.4
6	Google/Gmail	70.78
7	ScatterType	3.33
8	Handwritten CAPTCHA	9

CHAPTER VII: REFERENCE

1.K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski, "Designing human friendly human interaction proofs (HIPs)," in ACM Conference on Human Factors in Computing Systems (CHI 05), 2005, pp. 711-720.

2. A. Gupta, A. Jain, A. Raj, and A. Jain, "sequenced tagged CAPTCHA: generation and its analysis," in IEEE International Advance Computing Conference 2009 (IACC 2009), 2009, pp. 1286-1291.

3. A. Raj, A. Jain, T. Pahwa, and A. Jain, "Analysis of tagging variants of Sequenced Tagged CAPTCHA (STC)," in IEEE Toronto International Conference on Science and Technology for Humanity (TIC-STH 2009), 2009, pp. 427-432.

4. A. O. Thomas, A. Rusu, and V. Govindaraju, "Synthetic handwritten CAPTCHAs," Pattern Recognition, vol. 42, pp. 3365-3373, 2009.

6. B. Khan, K. Alghathbar, M. Khan, A. AlKelabi, and A. AlAjaji, "Using Arabic CAPTCHA for Cyber Security," in Security Technology, Disaster Recovery and Business Continuity. vol. 122, ed: Springer Berlin Heidelberg, 2010, pp. 8-17

7. M. Chew and H. S. Baird, "Baffletext: A human interactive proof," presented at the 10th Document Recognition & Retrieval Conference (SPIE), 2003.

8. A. Rusu, A. Thomas, and V. Govindaraju, "Generation and use of handwritten CAPTCHAs," International journal on document analysis and recognition, vol. 13, pp. 49-64, 2010.

9.. H. S. Baird, M. A. Moll, and S. Y. Wang, "ScatterType: A legible but hard-to-segment CAPTCHA," in 8th International Conference on Document Analysis and Recognition, 2005, pp. 935-939.

10. L. Ahn, M. Blum, J. Langford,“Telling Humans and Computers ApatAutomatically”, Communications of the ACM, 47(2), pp. 57–60, 2004

11. Jeff Yan, Ahmad Salah El Ahmad,“A low-cost attack on a Microsoft captcha”, Proceeding of the 15th ACM Conference on Computer and communications security, CCS ’08, pp.543-554, October, 2008

12. C.J. Hernandez-Castro, A. Ribagorda,“Pitfalls in CAPTCHA design and implementation: The MATH CAPTCHA, a case study”, Computers & Security, 29(1), pp. 141-157, 2010.

13. Ved Prakash Singh, Preet Pal,“Survey of Different Types of CAPTCHA”, International Journal of Computer Science and Information Technologies, Vol. 5, pp. 2242-2245, 2014.

14. ElieBursztein, Steven Bethard, John C. Mitchell, Dan Jurafsky, Celine Fabry,“How good are humans solving captchas? A large scale evaluation”, In Security and Privacy, 2010.

15. R. Gossweiler, M. Kamvar, and S. Baluja, "What's up CAPTCHA?: a CAPTCHA based on image orientation," in 18th International Conference on World Wide Web 2009, pp. 841-850.