PROPOSED METHODOLOGY OF TEXT CAPTCHA FOR SECURITY WITH CLOUD

Abstract

Verification strategies are wont to handle an extremely significant issue, the digital security. It assists with keeping away from the unlawful use or abuse of touchy information. Totally Automated Public Turing Test to Tell Computer and the Humans Apart (CAPTCHA) is the programmed security component which is utilized to decide if the client is a human or the robot. In this paper, we specialize in this various sorts of CAPTCHAs are often simply broken due to its style purpose. Once following a comprehensive analysis of the varied CAPTCHA, we have a tendency to address this downside by presenting a replacement CAPTCHA style principle to differentiate human (robot) from human (user). The core plan is that the challenge method of the CAPTCHA ought to contain the distinctive info with a private attribute. The notion of our plan is based on the data imbalance between humans.

Keywords: CAPTCHA, Artificial Intelligence, Machine Learning, Cloud Computing

1.0 Introduction: The internet has been playing a progressively essential role in our lifestyle, with the amount of the various web services like email and search engines. However, these are often threatened by attacks from computer programs like bots. To deal with this problem, CAPTCHA (Completely Automated Public Turing Test to inform Computers and Humans Apart) was developed to differentiate between computer drivers and social users. Although this mechanism offers good security and limits automatic record-keeping to web services, some CAPTCHAs have several weaknesses which permit hackers to break into the appliance of the CAPTCHA. This paper scrutinizes recent research on various CAPTCHA methods and their categories. Moreover it discusses the weakness and strength of those types. CAPTCHA may be a program which generates and evaluations the tests that are social answerable but is beyond the present computer virus capability. CAPTCHA is tentatively very useful within the sort of deformed text, mathematical, OTP (One Time Password), Audio, 3D, Graphical and Gaming CAPTCHA. The motive of this paper is to review the prevailing CAPTCHA techniques within the field of web security facilities and supply best level of security that can't be cracked and ready to replace the prevailing systems [1]. The planning and implementation of CAPTCHAs fall within the realm of AI. We aim to utilize CAPTCHAs as a tool to enhance the safety of Internet based applications. During this research we present a framework for a text-based CAPTCHA supported Greek words, Spanish and French words which may exploit the difference within the reading proficiency between humans and workstation programs. Our selection of script-based CAPTCHA is predicated on the very fact that it's not employed by an outsized number of individuals because these three languages having the various keywords. There's potential for an exponential rise within the applications that are likely to be developed therein script thereby making it easy to secure Indian language based applications. Unfortunately among several CAPTCHA techniques, the text CAPTCHA are often straightforwardly hacked and aren't dependable for the information security. In spite of the fact that CAPTCHA's which may be strong like confounded photo CAPTCHA, graphical CAPTCHA, iCAPTCHA are accessible, yet limit of the web site s don't place in power as they have to confide in the outsider CAPTCHA specialist organization and furthermore on account of the very reality that the size inhabitance which hundreds their database at

the server which is of less size and intention progressive activity of the site [2]. We formalize, design and implement two models on our proposed standard, a character-fundamentally based, and a datagram-basically based case, predictable with a web collaboration and secret phrase adapting to program. We task the individual to pick the secret key from the arbitrary characters that aren't in the secret phrase succession or to position the haphazardly cared for arrangements into the best possible request. an absolutely one of a kind innovation set of rules with a fluffy coordinating methodology has been proposed to highlight the limit of human blunder resistance and in this way the issue of arbitrary wager ambush. In contrast to different arrangements, our strategy would not found a workable pace essential validation convention, interface, and aptitude of the regular old web access. The few customer studies' results suggest that our proposed system is both straightforward (might be unraveled by people precisely inside anyway 20 seconds) and proficient (the Typer can handiest set up an irregular theory assault with a plainly low accomplishment rate). The flow diagram of online CAPTCHA is shown in fig 1 given below [3]
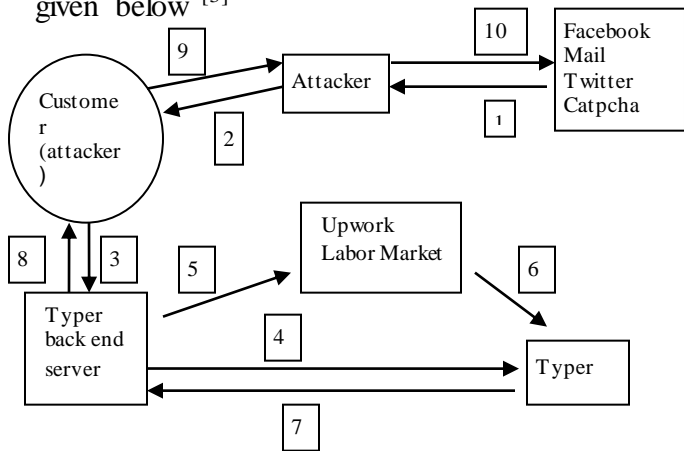

Fig 1: Flow Diagram of Online CAPTCHA

2.0 Literature Review: CAPTCHA might be a program or a framework that ensures against computerized contents (or bots). This paper will portray the new sorts of CAPTCHAs proposed in ongoing distributed papers, depict their characterizations, contrasting between

CAPTCHA upheld shortcoming and quality of them. This paper portrays the shifted classifications of the CAPTCHA frameworks, their applications and subsequently the disadvantages of each CAPTCHA procedures. It produces tests that people can pass yet PC programs can't. CAPTCHA frameworks are broadly utilized these days for protecting and giving security to web based administrations for people from maltreatment by bots. Varying sorts of CAPTCHA advancements are talked about during this paper and an inside and out investigation on their unwavering quality is performed. Subsequently, a substitution CAPTCHA system is proposed which is predicated on face location. The web contributes significantly to a few parts of human life like correspondences, training, and online business exercises and so forth. Some web administrations have online enrollment where the clients give data in order to join and use administrations like email in Yahoo, Gmail and Hotmail. In any case, numerous projects are created by programmers which naturally complete site enrollment pages with falsehood which may cause road turned parking lot , limiting the presentation of the framework and now and again, in any event, making it come up short, especially where a web webpage includes countless records. Along these lines, analysts built up a component to separate between human clients and PC programs inside the instance of online enlistment. The quality instrument utilized right now to manage this issue is CAPTCHA (Completely Automated Public Turing Test to advise Computers and Humans Apart). The idea of CAPTCHA is predicated on the intensity of individuals to endeavor to attempt to specific errands which pc programs can't, such as asking clients to a mutilated book picture or pick a particular picture from many showed pictures [4]. As of late, numerous sorts of CAPTCHAs are created. Some are bolstered Optical Character Recognition (OCR) like content CAPTCHA, though others are upheld

Non-Optical Character Recognition (Non-OCR) which utilizes sight and sound, similar to voice and video. Some of these sorts of CAPTCHAs are broken by new bot programs. for example , a book CAPTCHA are regularly broken by utilizing the component of division letters. A CAPTCHA, which means "Totally Automated Public Turing Test to illuminate Computers and Humans Apart", might be a test which will recognize human clients from PCs/robots. At the end of the day, it's a test, which most people can pass, however PC programs can't. Such tests are typically upheld hard, open AI issues like the fame of contorted content. The idea of a CAPTCHA originates from "Turing test", however it's occasionally depicted as an "opposite Turing test". the method of reasoning is that contrasting from the principal Turing test, which is controlled by an individual's and focused to a machine; CAPTCHA challenges are naturally produced and reviewed by a PC. Besides, the objective of structuring a CAPTCHA is to separate, rather than to neglect to separate, which is that the primary motivation behind Turing tests. Another distinction between a CAPTCHA and a Turing test is that the past was intended to go about as a proportion of progress for AI; nonetheless, the last might be a security component. The P for Public methods the code and along these lines the information utilized by a CAPTCHA ought to be openly accessible. Along these lines a program which will produce and grade tests that recognize people from PCs, however whose code or information are private, isn't a CAPTCHA [1]. it's significant for the difficulties to be considerably extraordinary a large portion of the occasions; else, they could be recorded, unraveled by people, at that point wont to answer future difficulties. Therefore, they should be created pseudo arbitrarily from an extremely huge space of unmistakable difficulties [2]. utilizing a CAPTCHA as a security component is critical in light of the fact that it can keep noxious projects from pursuing a great many records, posting numerous remarks in weblogs then on.

2.1 Current CAPTCHAs: Generally, CAPTCHA methods are often divided into five groups:
-Text-basedCAPTCHAs,
-Image-basedCAPTCHAs,
-Audio-basedCAPTCHAs,
-Motion-basedCAPTCHAs,
-HybridCAPTCHAs.
In content based frameworks, twisted variants of characters of a word rendered as an image and are introduced to the client. At that point, the clients are approached to type the arrangement that necessities recognizing all characters inside the right request. Since the picture contains enhanced visualizations, it's hard for a PC to recognize the words. Content based CAPTCHAs have the shortcoming of being deciphered by OCR programming. To beat this risk, different sorts of CAPTCHAs are presented. A book based CAPTCHA might be a mutilated picture of a grouping of characters on which varying sorts of debasements, foundation messes and shading blends are applied to shape it harder forattackers [5]. We'll present current content based CAPTCHAs in six sub-gatherings:
- CAPTCHAs with "English words" as their CAPTCHAtext,
- CAPTCHAs with "irregular strings" as their CAPTCHAtext,
- CAPTCHAs bolstered written by hand message,
- CAPTCHAs bolstered phonetic information,
- CAPTCHAs that require increasingly physical association with clients,
- Non-English CAPTCHAs.

CAPTCHAs with "English words" as their CAPTCHA content: In some CAPTCHA frameworks, as Gimpy, EZ-Gimpy, Captcha Service.org, PessimalPrint and reCAPTCHA, the CAPTCHA picture contains English word(s).

2.2 Gimpy: It is one of the most acclaimed CAPTCHAs which are basically founded on contorted content (Fig 2). This CAPTCHA was created in a joint effort with Yahoo with the point of shielding talk rooms from spammers to make them unfit to post ordered promotions and compose contents to produce free email addresses. Gimpy picks seven words from a lexicon; at that point renders a misshaped picture containing those words. It at long last presents them to its clients and solicits them to type three from the expressions of the picture to pick up section to the administration [6].
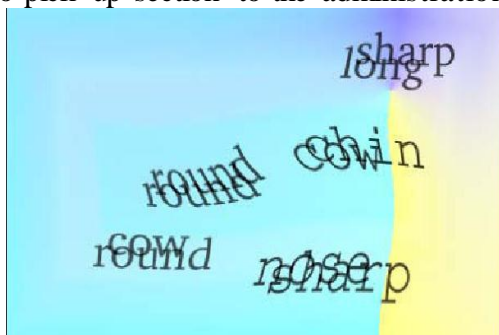

Fig 2: Gimpy CAPTCHA

2.3 EZ-Gimpy (CMU): In this CAPTCHA, from the start, a word is looked over a lexicon. In the following stage, the word is rendered to a picture utilizing different textual styles;
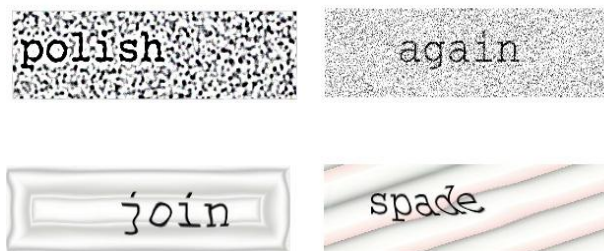

Fig 3: EZ-Gimpy CAPTCHA

and various sorts of bends, for example, dark or white lines, foundation lattices and angles, obscuring and pixel clamor are included (Fig 3). At that point, the client is approached to type the word [7].

2.4 Captchaservice.org: Right now (4), each challenge is a six-letter English word browsed a lot of 6000 words. The bending technique utilized is arbitrary shearing [8].

Fig 4: Captchaservice.org

2.5 Pessimal Print: Pessimal Print (Fig 4) focuses on corruptions, for example, adding commotion to or obscuring the pictures to vanquish OCR methods; the originators of this CAPTCHA contend that under the states of second rate picture quality, there is a huge hole in design acknowledgment capacity among people and machines [2]. This CAPTCHA fills in as follows. Initial, a word is pseudo haphazardly chose from a fixed run down containing 5-to-8-letter English words. At that point, it is rendered with a typeface (from a fixed rundown of 5 textual styles) and a fixed text dimension (size=8). At long last, a lot of picture corruptions including x-scaling, y-scaling, skewness, haziness and adding clamor are applied to the image[8].

Fig 5: Pessimal Print

Image-based CAPTCHAs usually use the prevalence of humans over computer vision organizations in recognizing the sort of an object in a picture . Although it's more appropriate for the human to unravel image-based CAPTCHAs instead of text-based ones, image-based CAPTCHAs have the problem of needing an outsized space for storing .

An audio-based CAPTCHA picks a string, extracts it to a superficial clip and offerings it to the users who are asked to acknowledge the contents of the audio clip. consistent with an outsized scale study on the usability of

CAPTCHAs, audio-based CAPTCHAs are more problematic than other types [5].

Another category is motion-based CAPTCHAs during which a movie or animation is presented to the users and that they are asked to acknowledge an action, vibrant word or image within the movie. This CAPTCHA is appropriate for users. Additionally, since the specified time interval during this CAPTCHA is comparatively high, it's safer . However, the high loading times are often an obstacle from a usability viewpoint. Another disadvantage is requiring an outsized database of animations. Finally, the term "hybrid CAPTCHA" has been nominated for a CAPTCHA that's a mixture of various types or considered for superior purposes [9].

3.0 Application of all kinds of CAPTCHA

There are numeral of solicitations of CAPTCHA on the online which are defined as follows.

1) Registration the online forms: At hand are several situates on the web which give permitted registering to avail their services. But they're vulnerable to network bots, it's going to inherit the shape of screenplays which may inventory thousands of email accounts on the web , thus worsening the valuable universe of web[10].

2) Online polling sites: These sites take user's response or feedback within the sort of questionnaires. to make sure that only human makes the response they create use of CAPTCHA.

3) To avoid web creeping: If a site doesn't want to urge indexed by an enquiry engine then they will make use of CAPTCHA.

4)E-Ticketing.

5) Thwarting Dictionary Attacks and E-mail spam.

3.1 Disadvantages in CAPTCHA

• Text Based Captcha: - during this there are different sizes font, blurring of the text is that the major issue and its OCR value is definitely guessed by bots.

• Image based CAPTCA- The users having low vision or colour blindness will face identify problems.

• Audio Based CAPTCHA- This CAPTCHA is predicated on English , so if the user won't know English Language then he's unable to seek out it and also he will learn the Language.

• Math Based CAPTCHA- It requires problem solving Skills and wish longer to unravel the CAPTCHA.

4.0 Proposed Methodology: We are proposing a methodology for introducing the CAPTCHA on cloud. The following steps are required to implement this methodology.

1. Initially the Captcha is created and stored in Cloud database.

2. User request Captcha

3. Extract IP/Mac address and check in database.

4. First dashboard frame is shown where CAPTCHA occurs having 6 keys

5. User will press the shown characters through keys which will occur on virtual keyboard.

6. Now, after pressing the keys response is recorded with time stamp.

7. If the threshold time exceeds then user ask for retry.

8. If the pressed keywords are found in database then prompt the message.

9. If the pressed keywords are not found in database then prompt the message "retry" and follow the below steps

10. Second frame 12 keys to press with 6 Greek words and 6 Spanish words

11. Follow step 5 to 8 database, IP address or Mac address is recorded after each retry.

12. After 3 retry the suspicious message is displayed.

4.1 Result Analysis: Here we have two snapshots of our result which shows the Captcha in greek words.
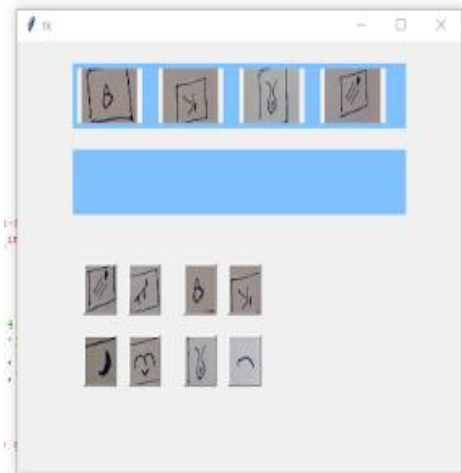
Fig 6: Proposed CAPTCHA in greek words

5.0 Comparison of CAPTCHA in Accuracy: The breakdown of the research work is completed in this segment and is tabularized in Table 1 In Table 1, the % attainment rate denotes that out of 100 attacks done on a CAPTCHA the assessment accompanying with it is the quantity of times the CAPTCHA is "broken" positively. It has been frequently rummage-sale because of easily manageable on the internet. The CAPTCHA execution is very dangerous without suspicious scheme .There is various bouts on script constructed CAPTCHA system. Most of end result from unfortunate safety and uncaring proposal superintend. For this persistence they offer good and well safety to avoid from occurrences.

S.No	Types CAPTCHA	Success rate (%)
1.	Textbased (Microsoft CAPTCHA)	92%
2.	EZ-Gimpy	97%
3.	AudioBased CAPTCHA	78%
4.	AudioBased CAPTCHA	71%
5.	Our Approach	99%

Conclusion: CAPTCHA shows a vital starring role in the life of safety wherever it precludes the Bot hackers from abusing the web service area. This research methodology proposes the CAPTCHA on cloud and provides the scheme to be safe and less defenseless.

References:
1. Niket Kumar Choudhary et al, "captchas based on the Principle- Hard to Separate Text from Background" / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014, 7501-750
2. Anju Bala and Baljit Singh Saini, "A Review of Bot Protection using CAPTCHA for Web Security", (IOSR-JCE) IOSR Journal of Computer Engine.
3. S. Ashok Kumar, N. Ram Kumar, S. Prakash and K Sangeetha, "Gamification of Internet Security by Next Generation captchas", International Conference on Computer Communication, IEEE 2017.
4. Cao Lei, "Image CAPTCHA Technology Research Based On The Mechanism Of Finger-Guessing Game", Third International Conference, IEEE 2015.
5. Nitisha Payal and Rama Krishna Challa, "AJIGJAX: A Hybrid Image Based Model for CAPTCHA/carp", IEEE 2016.
6. 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.
7. Google. 2018. Recaptcha v3The new way to stopbots.Available:https://www.google.com/recaptcha/intro/v3.html.
8. Statista, Bowman. 2018. Number of internet users in Thailand from 2015 to 2022 (in millions). https://www.statista.com/statistics/553730/number-of-internet-users-thailand/.
9. Nitirat Tanthavech, Apichaya Nimkoompai," CAPTCHA: Impact of Website Security on User Experience", ICIIT '19: Proceedings of the 2019 4th International Conference on Intelligent Information technology Pages 37–41, February 2019

10. Shadi Khawandi, Firas Abdallah, Anis Ismail," A Survey On The Different Implemented Captchas", 3rd International Conference on Computer Science and Information Technology (COMIT 2019), DOI: 10.5121/csit.2019.90101.