www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 5 May 2016, Page No. 16458-16462

A Survey on Captcha Categories

Divyashree N[1], Dr. T. Satish Kumar[2]

[1]Department of Computer Science and Engineering
RNS Institute of Technology, Bengaluru-560 098
ndivya.shree17@gmail.com
[2]Professor, Department of Computer Science and Engineering
RNS Institute of Technology, Bengaluru-560 098
Satish.savvy@gmail.com

Abstract – Authentication methods are used to tackle a very important issue, the cyber security. It helps to avoid the illegal use or misuse of highly sensitive data. Completely Automated Public Turing Test to Tell Computer and the Humans Apart (CAPTCHA) is the automatic security mechanism which is used to determine whether the user is a human or the robot. It is a program which generates and grades the tests that are human solvable but is beyond the current computer program capability. Unfortunately among several CAPTCHA techniques, the text catpcha can be easily hacked and are not reliable for the data security. Though CAPTCHA's that are robust such as complex image CAPTCHA, graphical CAPTCHA, iCAPTCHA are available, yet most of the websites do not implement as they have to rely on the 3rd party CAPTCHA service provider and also due to the fact that the size occupancy which loads their database on the server which is of less size and cause slow operation of the website. This paper describes the various categories of the CAPTCHA systems, their applications and the drawbacks of each CAPTCHA techniques.

I. INTRODUCTION

Security is an important aspect of an information security program. Authentication is a key area in security research and practice. Authentication process is used to determine whether the user should be given access to the system/resource. Apart from the user driven username and password for authentication, CAPTCHA is used nowadays to protect the user accounts in systems and websites from malicious bot attacks and spywares. Bots or robots called as "software agents" are the malicious program that sneaks into a person's computer or website accounts in many ways. They search for vulnerable and unprotected computers and websites to hack the user accounts and their valuable information. Spyware [1] is also software which gathers information about the computer use from the user perspective and sends that information back to the third party. CAPTCHA [2, 6], is basically a challenge response authentication system which provides a mechanism for users to protect their access to information from spam's and password decryptions by just taking a simple test. Usability (must be human friendly), robustness (must be difficult to break by robots) and quick and easy respond are the important aspects for designing a successful CAPTCHA system. The robustness of a CAPTCHA lies in providing a strong password for resisting the continuous attacks into real-time security applications.

The four important properties [3] that every CAPTCHA system must possess:

1. Secure: The program generated tests must be difficult for the machines to solve using any algorithms.
2. Automated: The programs (computer programs) must be capable of generating and grading the test.
3. Open: In accordance with the Kerckhoff's principle [4], the underlying
   Algorithms and databases must be made public.
4. Usable: Humans as users should be able to solve these tests in a reasonable time.

II. RELATED WORK

Different CAPTCHA techniques are designed to provide authentication for secured systems and online applications. CAPTCHA prevents the bots attack and provides highest security level and easy access to users to users for secured information. Attacks such as spyware, malicious, dictionary attacks are possible on the CAPTCHA. For example, under text CAPTCHA,

the Gimpy CAPTCHA and Ez- Gimpy CAPTCHA were cracked by dictionary attack. As the existing CAPTCHA is hacked, new and more robust CAPTCHA techniques are designed, for example, the re-CAPTCHA, iCAPTCHA, CaRP scheme. I have studied few research papers related to different types of CAPTCHA system as given below. Spyware is the software which gathers information about the computer from the user perspective and sends the information back to the third party. The paper [1] describes the 2 ways in which the password can be protected from the spyware attacks. 1st method is to embed the user password in random key strokes to confuse the spyware by leaving the actual login unaffected. 2nd method is by employing a proxy server for stripping the random keys. CAPTCHA is a security measure used to prevent malicious attacks. The paper [2] describes about the CAPTCHA token that requires soft keyboard to input which prevents the spyware attack as well. Authentication is very important in online applications because the attackers will try to hack the user credential information through keyboard hooking. To overcome the problem of keyboard hooking, mouse click or touchpad input method was designed to enter the CAPTCHA images [18]. For designing a CAPTCHA, every design must include the properties of secure, automated, open and usability. The paper [3] describes how these individual properties should be handled and also two different approaches for designing the CAPTCHA based on image. The Kerckhoff's principle [4] states the rules for sending and using cryptographic messages. An encrypted message can without key not be efficient to decipher, the communication partner may suffer harm if the encryption system is known, the key must be easy to remember and be interchangeable, the cryptograms must be transferred, which was then called "telegraphierbar" be, the Chiffrierapparat and the documents must be transportable, the system must be easy to use (without expert assistance), the encryption system should be thoroughly investigated by experts are some of the rules and requirements for a confidential erected communications. There are different types of CAPTCHA's. The paper [5] discusses about the various CAPTCHA's, their applications and their drawbacks. The paper [7] explores the testing of object recognition techniques on visual CAPTCHA's, Ez- gimpy and gimpy CAPTCHA. CAPTCHA's are more used in online web services for protecting the webpage's from the bots attack. The paper [13] discusses about some of the CAPTCHA techniques that are broken by bots and the some of the CAPTCHA's that can stand strong against the bots such as puzzle CAPTCHA, hard text recognition CAPTCHA's etc.

III. THE SURVEY

There are different types of CAPTCHA [5] system which can be categorized into five types as given below:

1. Text based CAPTCHA.
2. Image based CAPTCHA.
3. Audio based CAPTCHA.
4. Video based CAPTCHA.
5. Puzzle based CAPTCHA.

1. Text based CAPTCHA:

CAPTCHA based on text is very simple to implement. It is very effective and just requires a large question bank. Some of the examples are given below:

- What is 1 + six?
- Which of apple, mango, and carrot is a vegetable?
- What is the next day after Monday?

There are different types of text CAPTCHA's. They are:

Gimpy CAPTCHA is based on the ability of a human to read distorted texts which the computer programs cannot do. It is built by the CMU for their message services in collaboration with yahoo. Gimpy works by selecting few words from the dictionary and displaying in misshape or corrupted or in a disfigure way. The Fig 1 shows an example of Gimpy CAPTCHA. The user has to pass the gimpy test in order to attain access to the secured services. Gimpy CAPTCHA challenge has be broken by the dictionary attacks that contained limited number of words [7].

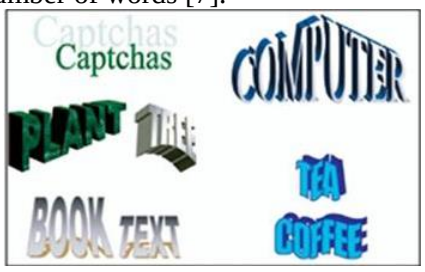

Fig 1: Example of Gimpy CAPTCHA

Ez Gimpy CAPTCHA is the simplified technique of Gimpy CAPTCHA. It randomly picks only a single word from the dictionary of words and applies distortion or misshape to that text and the user is requested to type the text correctly in order to attain access to the secured services. The Fig 2 shows an example of Ez Gimpy CAPTCHA and was broken by dictionary attacks and by Mori et al [8].

Fig 2: Example of Ez Gimpy CAPTCHA

DOI: 10.18535/ijecs/v5i5.21

Baffle Text CAPTCHA was designed by Henry Baird at California University at Berkeley. It is a modified version of Gimpy. The Fig 3 shows an example of Baffle-Text CAPTCHA. In this case a random characters or alphabets are picked from pronounceable text and the user is challenged to type the text correctly in order to attain access to the secured services.

Fig 3: Example of Baffle-Text CAPTCHA

MSN CAPTCHA makes use of digits and characters (upper case) of 8 lengths are used. The Fig 4 shows an example of MSN CAPTCHA in which wrapping is used to produce the distortion of characters and the ripple effect. MSN CAPTCHA was broken by yan [9].

Fig 4: Example of MSN CAPTCHA

Other prominent text based Captcha techniques includes the Scatter Type [10], Human Visual System masking Characteristic CAPTCHA [11] and Handwritten Word based CAPTCHA [12].

Bar CAPTCHA [19] consists of many small bars that are used to specify text as well as adds noise in the background. . The Fig 5 shows an example of Bar CAPTCHA in which the bars are used so that the bots will be unable to differentiate between the text and the noise.

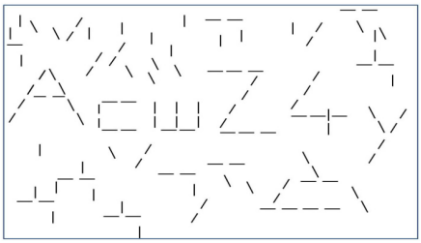

Fig 5: Example of Bar CAPTCHA

Thread CAPTCHA appears as a handwritten text in a free form style using a long thread. The Fig 6 shows an example of Thread CAPTCHA which is Similar to bar CAPTCHA; the thread is used to create noise as well as a text in the image which makes the bots difficult to identify the thread which represents the text.

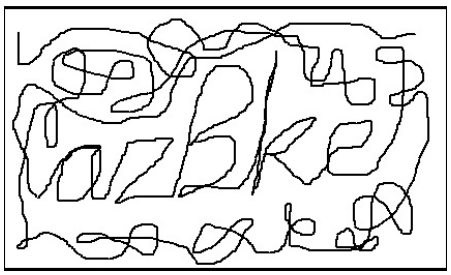

Fig 6: Example of Thread CAPTCHA

Transparent CAPTCHA is a technique in which the text is written in a transparent font over an image. The Fig 7 shows an example of transparent CAPTCHA which makes the bots very difficult to identify the pixels that represent the exact text and the pixels that represent the other objects in that image like trees, birds, sky, boat, sand etc that are visible in the image.

Fig 7: Example of transparent Captcha

2. Image based CAPTCHA:

Image based CAPTCHA consists of set of images that are associated with the same concept or object. The user is required to enter the concept or the object which all the image belonged to. Initially ESP_PIX CAPTCHA was proposed as an image based CAPTCHA by Bhem and Von Ahn. It used large databases of animated images and photographs of every object. There are different types of Image based CAPTCHA schemes which uses different concepts or patterns that can be recognized by the users but difficult for the robots to recognize. Some of the image based CAPTCHA's are Bongo and Pix CAPTCHA.

Bongo CAPTCHA is named after Mikhail M Bongard who published pattern recognition problems book. In Bongo [16] visual based pattern recognition is provided for the user to solve. The Fig 8 shows an example of Bongo CAPTCHA. It contains 2 block series namely the right block and the left block series. The series of the right block differs from the left blocks, and the user should identify the characteristic which set them apart.

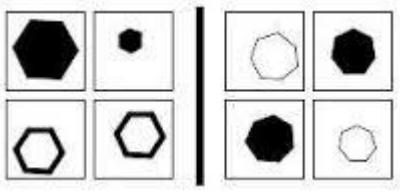

Fig 8: Example of Bongo catpcha

Pix CAPTCHA [17] uses a large volume of database containing the animated images and photographs of daily objects with labels. All the images are concrete objects (example: a flower, a baby, a horse etc). The Fig 9 shows an example of Pix CAPTCHA. In Pix, a set of objects are picked at random from the database and the user is asked a question based on the pictures displayed saying, what are these pictures of? It is very easy to write a bots program for such question as the data and the code of the CAPTCHA is displayed publicly which makes Pix to be not considered as catpcha. However, one way to consider Pix as CAPTCHA is to distort the images randomly before displaying them to the user, so that the bot programs find difficulty in searching the database for undistorted images.

3. Audio based CAPTCHA:

Audio based CAPTCHA are for visually disabled users which work on the sound-based systems. The Fig 10 shows an example of Audio CAPTCHA. It contains a downloadable audio clips which the user should listen and submit the herd word. ECO was the first audio based CAPTCHA system which was put into operation in Hong Kong at City University by Nancy Chan [13].

Fig 9: Example of Pix CAPTCHA

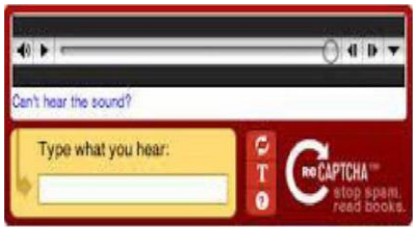

Fig 10: Example of Audio CAPTCHA

Fig 11: Example of Video based CAPTCHA

4. Video based CAPTCHA:

Video based CAPTCHA system [14] uses a technique in which the video contains few random words. The Fig 11 shows an example of Video CAPTCHA. When the video is played the user has to submit those displayed words. The users need not to wait until the video finishes for submitting the displayed words. The user passes the test only when the ground truth tags which are produced automatically matches with the user entered tags.

5. Puzzle based CAPTCHA:

Puzzle based CAPTCHA can either be a picture based puzzle or a mathematical puzzle. The Fig 12 shows an example of Puzzle based CAPTCHA. In a picture based puzzle, the picture is divided into segments and is shuffled. Each segment will have a segment number followed by the next segment. The user has to combine these segments properly to form a correct complete picture [15]. The mathematical puzzle is 100% effective and can be integrated into login, registration forms in the website for secured access. The user has to solve the math puzzle provided in order to gain the access to secured services.

Fig 12: Example of Puzzle based Captcha

Applications of CAPTCHAs

Some of the practical applications of the CAPTCHA's (but not limited to):

- Search Engine Bots: it is sometimes necessary to keep the webpage's unindexed in order to prevent the others from finding it. An html tag can be used to prevent the search engine bots from reading the webpage's, but the tags however doesn't guarantee that the robots will not read the webpage. Search engine bots, since they usually belong to large companies, respect web pages that don't want to allow them in. CAPTCHAs are needed to guarantee that the robots will not enter into a website.
- Worms and Spam: CAPTCHA system offers a plausible solution against spam and email worms
- Preventing Dictionary Attacks: CAPTCHAs are used to prevent the website from dictionary attacks in password systems. It prevents a computer from getting iterated in a password space for solving a CAPTCHA after certain amount of unsuccessful login attempts. It provides a better approach of account locking after a sequence of failed login attempts. By doing so, it will also block the attacker from locking the accounts purposely.
- Preventing Comment Spam in Blogs: Many bloggers are familiar with the malicious programs that submit false comments, which raise the search engine rank of

a website. CAPTCHA allows only humans to comment on a blog. There is no need of sign up requirement for the user to enter a comment, and legitimate comments are never lost.

- Protecting Website Registration: Several companies like Yahoo!, Microsoft, etc. offer free email services. Many of these services suffered from bots attack that signs up for thousands of email accounts. CAPTCHA provided the solution to this problem by ensuring that only humans would obtain free accounts.
- Protecting Email Addresses From Scrapers: The spammers crawl in the web, searching the email addresses which are posted in clear text. CAPTCHAs provide an effective mechanism to hide the email address from the web scrapers. The idea is to make the users solve a CAPTCHA before showing the email address. This can be found at reCAPTCHA MailHide.

IV. CONCLUSION

This paper conducts a comprehensive survey of different existing techniques of CAPTCHA systems and how they are used for providing authentication. In addition, this paper specifies some of the practical applications of the CAPTCHA's.

REFERENCES

[1] D. Florencio and C. Herley. KLASSP: "Entering Passwords on a Spyware Infected Machine Using a Shared-Secret Proxy". 22nd Annual Computer Security Applications Conference (ACSAC), 2006, pp.67-76.

[2] Ahn, L. V., Blum, M., & Langford, J. Telling Humans and Computer Apart Automatically. CACM, V47, No 2.

[3] Mehrnejad, M., Bafghi, A. G., Harati, A., & Toreini, E. SEIMCHA: A New Semantic Image CAPTCHA Using Geometric Transformations. International Journal of Information Security, 63 - 76.

[4] Kerckhoff's, A. (1883). La CryptographieMilitaire. Journal des Sciences Militaires 9, 161-191.

[5] Ved Prakash Singh, Preet Pal "Survey of Different Types of CAPTCHA" International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2242-2245

[6] L. von Ahn, M. Blum, and J. Langford, "Telling Humans and Computers Apart Automatically", Communications of the ACM, Vol. 47, No. 2, February 2004, pp. 57-60.

[7] G. Mori and J. Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA" In Proceedings of the Conference on Computer Vision and Pattern Recognition, Madison, USA, 2003, pp. 134-141.

[8] J.malik and G.mori, "Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA, "Conference on Computer Vision and Pattern Recognition, vol-1,134-141, June 2003.

[9] Ahmad Salah El Ahmad and Jeff Yan, "A low -cost attack on a Microsoft CAPTCHA," in 15th ACM conference on computer and communications security, oct-2008, 543-554.

[10] M. Chew and J. D. Tygar, "Image Recognition CAPTCHAs", In Proceedings of the 7th International Information Security Conference (ISC 2004). 2004, Springer.

[11] R. Ferzli, R. Bazzi, and L. J. Karam, "A CAPTCHA Based on the Human Visual Systems Masking Characteristics", In Proceedings of the 2006 IEEE International Conference on Multimedia and Expo (ICME'06), Toronto, Ontario, Canada, 2006, pp. 517-520.

[12] A. Rusu and V. Govindaraju, "Handwritten CAPTCHA: Using the Difference in the Abilities of Humans and Machines in Reading Handwritten Words", In Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR- 9 2004), Kokubunji, Tokyo, Japan, 2004, pp. 226-231.

[13] Anju Bala and Baljit Singh Saini, "A Review of Bot Protection using CAPTCHA for Web Security", (IOSR-JCE) IOSR Journal of Computer Engineering, Volume 8, Issue 6 (Jan. - Feb. 2013), 36-42.

[14] H. Kwak, M. chew, P. Rodriguez, S. Moon and Y.Y. Ahn, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," In Proc. IMC 2007, ACM Press, 1–14.

[15] Preet Pal and Ved Prakash Singh, "Survey of Different Types of CAPTCHA," / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2242-2245.

[16] Anju Bala and Baljit Singh Saini, "A Review of Bot Protection using CAPTCHA for Web Security,"(IOSR-JCE) IOSR Journal of Computer Engineering, Volume 8, Issue 6 (Jan. - Feb. 2013), 36-42.

[17]www.slideshare.net/kunalkiit/seminar-report-on-CAPTCHA.

[18] M. Alsaleh, M. Mannan, and P. C. van Oorschot, "Revisiting defenses against large-scale online password guessing attacks," IEEE Trans. Dependable Secure Comput., vol. 9, no. 1, pp. 128–141, Jan./Feb. 2012

[19] 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