A Survey on Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA)

[1]Sheikh Mustaq Ahmmed,[2] Muhammad Minhazul Haque Bhuiyan,[3]Md. Shakir Khan
[1]Lecturer,[2] Senior Lecturer,[3]Lecturer
[1] Department of Computer Science and Engineering,
[1]Leading University, Sylhet, Bangladesh

Abstract— **Nowadays, many services in the internet including Email, search engine, social networking are provided with free of charge due to enormous growth of web users. With the expansion of Web services, denial of service (DoS) attacks by malicious automated programs (e.g., web bots) is becoming a serious problem of web service accounts. A HIP, or Human Interactive Proofs, is a human authentication mechanism that generates and grades tests to determine whether the user is a human or a malicious computer program. Unfortunately, the existing HIPs tried to maximize the difficulty for automated programs to pass tests by increasing distortion or noise. Consequently, it has also become difficult for potential users too. So there is a tradeoff between the usability and robustness in designing HIP tests. In our propose technique we tried to balance the readability and security by adding contextual information in the form of natural conversation without reducing the distortion and noise. In the result section, a microscopic large-scale user study was conducted involving 110 users to investigate the actual user views compare to existing state of the art CAPTCHA systems like Google's re-CAPTCHA and Microsoft's CAPTCHA in terms of usability and security and found our system capable of deploying largely over internet.**

Keywords- CAPTCHA; HIPs; Usability ; OCR; Web Services; Context; Cognitive Psychology; Conversation.

I. INTRODUCTION

A CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) or HIP (Human Interactive Proof) is an automatic security mechanism used to determine whether the user is a human or a malicious computer program .It is a program that generates and grades tests that are human solvable, but intends to be beyond the capabilities of current computer programs [1]. It has become the most widely used standard security technology to prevent automated computer program attacks. With the expansion of Web services, denial of service (DoS) attacks by malicious automated programs (e.g., bots) are becoming a serious problem as masses of Web service accounts are being illicitly obtained, bulk spam e-mails are being sent, and mass spam blogs (splogs) are being created. Thus, the Turing test is becoming a necessary technique to discriminate humans from malicious automated programs [1].

The CAPTCHA is usually a simple visual test or puzzle that a human can complete without much difficulty, but an automated program cannot understand. The test usually consists of letters, numbers or their combination with overlapping and intersection. A typical example of a text-based CAPTCHA challenge is shown in Figure 1. The CAPTCHA images may be distorted in some way or shown against an intricate background to keep them from being easily read by Optical Character Recognition (OCR) software) or other image recognition systems.

Currently, in order to defend malicious programs from issuing advertisements or other useless information recklessly, message boards of BBS, blog and wiki have widely used CAPTCHA challenges as a defense mechanism [1],requiring that users must input the correct letters to leave a message. CAPTCHs have a wide variety of applications on the web such as:

1. Worms and Spam
2. Web crawler
3. Online Polls
4. Free Email Services
5. Preventing Dictionary Attacks

Moreover, some spammers have found a creative way to provide their bots with CAPTCHA solving capabilities using pornographic sites – outsource the CAPTCHA-solving task to humans. For example, when a bot is faced with a CAPTCHA, it might place that CAPTCHA onto the entrance page of a pornographic site [12], and the next visitor to that site solves the CAPTCHA for the bot, in exchange for (free or price-reduced) entrance to the porn site[12].

Researches on CAPTCHA mechanism have gained significant attention recently. Researchers encourage images, audio and video as a possible alternative to text-based CAPTCHA. These tests are easier for humans to solve and tough for automated bots. Unfortunately, the existing CAPTCHA techniques tried to maximize the difficulty for automated programs to pass tests by increasing distortion or noise. Consequently, it has also become difficult for humans too. A recent study which investigated user's perceptions towards CAPTCHA challenges and also highlights the necessity for user friendly CAPTCHA challenges [8]. Results have shown that even experienced users face difficulties in solving a CAPTCHA challenge [8] need few tries or solving time is more than 15 seconds.

In this context, the work presented in this paper constitutes an effort towards focusing mainly on user's cognitive styles and performance related to CAPTCHA challenges within Web-based environments. Cognitive styles represent an individually preferred and habitual
approach to organizing and representing information [9]. Various research attempts have been reported [10, 11] that investigate the effect of cognitive styles of users on preference and performance issues in Web-based environments. We tried to balance the readability and security by adding contextual information in the form of natural conversation without reducing the distortion and noise.

II. VARIOUS TYPES OF CAPTCHA METHODS

A. Text-Based CAPTCHAs:

In this system computer generates a sequence of letters or digits after distorting them with a certain amount of noise render them on to the screen. The user is asked to identify the characters in order to pass the test. GIMPY [28] is a very reliable text CAPTCHA built by CMU in collaboration with Yahoo to protect chat rooms from spammers who were posting classified ads and writing scripts to generate free e-mail addresses (shown in fig. 1(a)). Another interesting character labeling based CAPTCHA is reCAPTHCA [3]. It is a free CAPTCHA service that helps to digitize books, newspapers (in fig. 1(b)). Microsoft uses a different CAPTCHA for services provided under MS Umbrella. They use eight characters (upper case) and digits. Microsoft's CAPTCHA [25] is used for services including Hotmail, MSN and Windows Live as shown in Fig. 1(c) foreground is dark blue, and background is grey. Warping is used to distort the characters, which makes computer recognition very difficult.



Fig. 1. Examples of text-based CAPTCHAs

Google [2] uses the CAPTCHA in many places. It is shown when URLs are added to Google and a new Gmail account is signed up (in fig. 1(d)). It is obvious that Google does not add background to the image.

Attacks on text-based systems mostly make use of OCR (**O**ptical **C**haracter **R**ecognition) algorithms. These algorithms first segment the images into small blocks containing only one letter, and use pattern recognition algorithms to classify the letters in each block [13, 14, 15]. In counter-attack to these segmentation algorithms, text-based CAPTCHA systems employ the following techniques to increase robustness [24, 17]:

1. Adding more noises in the form of scattered lines and dots to the background.
2. Characters are connected, overlapped or twisted to increase difficulty in character recognition.

However, all the above techniques make the task harder for humans too. Connecting characters together creates ambiguous characters such as “vv” can be similar to “w”, “cl” can be similar to “d”, “nn” can be “m”, “rn” can be “m” “rm” can be“nn” where users cannot be sure what they are.

Moreover hard to tell distorted O from 0, 6 from G and b, 5 from S/s, 2 from Z/z, 1 from l. In case of dictionary words non-English users get into trouble with predicting.

B. Image-Based CAPTCHAs:

In this, users is required to identify image. The advantage of image based CAPTCHA over text based is that pattern recognition is a hard AI Problem and therefore it is difficult to break this test using pattern recognition technique. The users of this CAPTCHA usually interact using a pointing device, e.g., mouse. In general, image-based CAPTCHAs require larger web page area, and need an image database maintained at the server. ESP-PIX[20] is a Captcha script that instead of asking you to type letters requires that you look at a set of pictures and then select the word that best describes all the images(in fig. 2(a)). It is available in English therefore end user must have a comprehensive English vocabulary. There are only 27% people in the world are English speaking [6].



Fig. 2. Examples of image-based CAPTCHAs

Scheme	No of Choices	English Dependency	Probability of entering a Bot	Average Solving Time(in seconds)	Interaction
Move & Select	4	Yes	1/40	6.02	Mouse
ESP-PIX	72	Yes	1/72	13	Mouse
Asirra	12	no	1/12	17	Mouse
MMC	4	yes	1/4	8	Mouse & Keyboard

TABLE I: Summarization of the image based CAPTCHAs

Multi Model CAPTCHA uses text and image based system together where end user is shown an image where four text labels associated with it. Text labels are attached in the image and the user is asked to select an appropriate text label [4]. A snapshot of Multi Model CAPTCHA is shown in fig. 2(b).

Move & Select [23] it's a 2 layer CAPTCHA, desired to improve security and reduce the solving time of human. In the proposed solution, we try to make use of human cognitive processing abilities into our CAPTCHA design (in fig. 2(d)). It is not suitable for visually impaired users. Also it may be challenging for users with learning disabilities. Table I summarizes the entire image based CAPTCHAs discussed above.

C. Audio and Video Based CAPTCHAs:

The program [18] picks a word or a sequence of numbers at random, renders the word or the numbers into a downloadable audio file and background noises are added to the sound clip using TTS software to make the test more robust against bots. It then presents the distorted sound clip to the user and asks users to enter its contents. These systems are highly dependent on the audio hardware and need to install essential software like Adobe Flash on their computers. These barriers lead to spend user's time more than standard response time which is typically about 5-15 seconds. Because of high level of distortion characters produce similar sound like “d “and “b” [7]. These English words are unfamiliar to non-English humans. It helps visually disabled users but the worst case is for people who have problem in both hearing and vision. YouTube which currently stores and indexes close to 150 million videos used as a video dataset in video based CAPTCHAs [14].However; video is also more complex and need more time and bandwidth to answer the challenge than other schemes.

III. MOTIVATION

A. Tradeoff between usability and robustness

There are two major issues involve in designing a strong CAPTCHA test:

1. Robustness (difficult to break)
2. Usability (human friendly)

In the system *robustness*, the characters must be more distorted, so that the malicious computer software (e.g., a robot program) could not recognize them. *Usability* is concerned with making CAPTCHA tests easy to learn, use, understand and interpret. It has become difficult for automated programs to pass tests by increasing distortion or noise. Consequently, it has also become difficult for humans too example given in Fig. 3(a). That is not a good design obviously. On the contrary, if the design is quite easy to be identified by the user, then the computer may also be able to easily identify and solve it as illustrated in Fig. 3(b). We therefore need to adopt even more advanced human cognitive processing abilities to enhance CAPTCHA to overcome this problem.

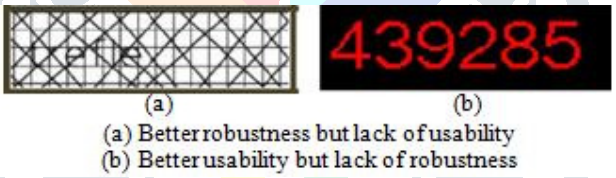


Fig. 3. Usability vs. Robustness

So there is a tradeoff between the usability and robustness in designing CAPTCHA test. Moreover, it is also important to note that answering CAPTCHAs added annoyance for potential users, who feel difficult to prove that they are human. To our best survey, there has been no such a CAPTCHA that considered human cognitive processing abilities into the CAPTCHA design.

IV. THE PROPOSED CAPTCHA TEST

In this paper, the proposed method has been developed to distinguish human users and computer programs from each other by the fact that human user have special cognitive processing abilities on the other hand it is nearly impossible for OCR programs to have that and it falls into hard AI problem.

Cognitive Psychology is the study of human perception, attention, memory and knowledge, and the ways in which these have been applied in the design of computer interfaces. Cognitive psychology[26] relates the use computer systems:

1. How humans perceive the world around them (e.g. Web Pages)
2. How they store and process information and solve Problems (e.g. CAPTCHA tests)
3. How they physically manipulate objects (e.g. clicking a link, button , type text etc)

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves[5].If a piece of information can be used to characterize the situation of a participant in an interaction, then that information is context.

In online we often engage ourselves in natural conversations by sending each other blog comments, chat messages, status updates and „tweets“. These messages are then indexed and made accessible through Internet search companies, instant messenger software or search index generators.

We know that there is a tradeoff between readability and security in solving CAPTCHA challenges. In our propose technique will try to balance the readability and security by adding contextual information in the form of natural conversation without reducing the distortion and noise.

In this paper we propose a human friendly CAPTCHA test, based on contextual conversation to secure web services as shown in Fig. 4. The test contains context information such as the knowledge in the form of natural language conversations.

In the server side, we need additional processing before presenting the CAPTCHA to the user. The required processing steps are discussed below.

1. Sample natural conversations (e.g. blog comments, chat messages, status updates and „tweets") are collected with the help of search index generators.
2. A meta conversation is collected from the sample.
3. The spell checking has been done using spell checker [27] to avoid spelling mistakes.
4. Using a parts of speech tagger [16] to identify the parts of speech in the meta dialogue and select the adjectives, adverbs and verbs as a candidate set for the distortion. This is necessary because exclusion of preposition and noun and pronouns are necessary, so we will only distort the adjectives, adverbs and verbs.
5. Then different degree of distortion (e.g. connected, overlapped, twisted and scaling etc) has been done to a randomly selected word from the candidate set.
6. Finally, fictitious characters are generated based on the meta conversation using only solid lines to avoid the bandwidth load on the network.

In the client side, users are presented with an image like Fig. 4. and they are asked to type only the distorted characters from the conversation. If the distorted characters are presented without this contextual conversation then the user will have some difficulty in interpreting the characters but now it is relatively easy because the human has special cognitive processing abilities. Therefore the context presented along with the CAPTCHA has decreased the solving time and thus increased the readability and usability.



Fig. 4. Proposed contextual conversation CAPTCHA

Other example of a natural conversation is the conversation between a father and his son. Here the context or the situation is "Sakir (son) has come to his father to tell his annual examination result". So the sample natural conversation between them is:

Son: "Papa, I have secured first position in the annual examination."**Father:** "*Congratulation*! My boy I am so proud of you."

If the word *Congratulation* is highly distorted in the actual Captcha test, it is possible for the user to recover the word by perceiving the contextual conversation. Therefore reducing the solving time thus improves the usability without hampering the robustness.

V. USER STUDIES AND RESULTS

A large-scale user study was conducted to investigate the actual user views related to perceptions, cognition, and user preferences related to CAPTCHA systems with real usage scenario. An invitation was announced on the web site of the University and on social networking sites, by sending invitation to known email addresses in order to recruit participants for the survey. The aim of this selection process was to recruit a representative sample of participants of varying profiles, intended to increase internal validity of the survey by involving expert, average and novice users with respect to CAPTCHA challenges. For this purpose, we first built a website which would present the users with sample challenges.

A. Demographics of Participants

A total of 110 people participated so far of age between 14 and 66 in the study between February and April 2012. 18 of them completed half of the test, due to internet disconnection or had double answers and were omitted from the test sample. The final sample included 92 valid participants which mean that they does not have any kind of vision problem that hampered their effort to identify colors, shapes, or patterns.

CAPTCHA test	Outcome		Average Submit Time (*in seconds*)	Average Success Time (*in seconds*)
	Success	*Failure*		
Contextual Conversation CAPTCHA	81/92 ≈88.04%	11/92 ≈11.96%	752/92 ≈ 8.17 s	752/81 ≈ 9.28 s
Google's re-CAPTCHA	85/92 ≈92.39%	7/92 ≈7.61%	1068/92 ≈ 11.61 s	1068/85 ≈ 12.56 s
Microsoft's CAPTCHA	77/92 ≈83.69%	15/92 ≈16.31%	1097/92 ≈ 11.93 s	1097/77 ≈ 14.24 s

Table II: Overview of the user result data



Fig. 5.Timing Distribution of each system for all users

B. *User Study Layout*

The participants were asked to visit a Web-page in order to take part in the study.

1. An initial questionnaire asking the users to enter the information (e.g. Age, Gender, Education, Years using the internet, Frequency of internet use, Native language, Familiarity with computers etc.)
2. A 1-page description of Contextual Conversation CAPTCHA, Microsoft's CAPTCHA and Google's re-CAPTCHA, showing users how to solve each challenge with pictorial description.
3. Then one challenge from each of the Contextual Conversation CAPTCHA, Microsoft's CAPTCHA and re-CAPTCHA.

A detailed pictorial description of each of the CAPTCHA"s was necessary for fair usage data and results.

There may be some users seeing a CAPTCHA for the first time on the other hand it is necessary to introduce the participants with our Contextual Conversation CAPTCHA. The study took an average of 5.6 minutes to complete for each participant.

C. *Usability Study*

Quoted from Jakob Nielsen [22], usability is defined by the following five quality components: Learnability, Efficiency, Memorability , Accuracy and Satisfaction.

Typically, the basic task that a CAPTCHA imposes to users is intuitive, easy to understand and easy to remember. Thus, CAPTCHA has a relatively good memorability.

1. ***Average solving time***

As shown in Table III, users complete text-based Contextual Conversation CAPTCHA challenges faster than that of Google"s re-CAPTCHA and Microsoft"s CAPTCHA. Each user takes an average of 3 seconds more to complete compare to Contextual Conversation CAPTCHA.

Average solving time in Contextual Conversation CAPTCHA is about 8.17 seconds from the distribution plots shown in fig 5(a) .On the other hand solving time is comparatively higher for Google"s re-CAPTCHA with most of the users taking around 11.61 seconds and for Microsoft"s CAPTCHA is 11.93 seconds (in fig 5(b), 5(c)) So it has better efficiency compare to others.

	CAPTCHA test		
	Contextual Conversation CAPTCHA	**Google's re-CAPTCHA**	**Microsoft's CAPTCHA**
Average solving time in seconds	8.17	11.61	11.93

Table III: Average Time taken per challenge for each of the systems (in seconds)

2. ***Accuracy***

Accuracy or the success rate is defined how successfully a participant can pass a CAPTCHA challenge .The total number of correct attempts of Contextual Conversation CAPTCHA (e.g. 88.04%) is higher than Microsoft"s CAPTCHA(e.g. 83.69%) as shown in Table II, which clearly indicates that users are able to solve more challenges of Contextual Conversation CAPTCHA correctly. Fig 6 shows a graphical representation of the difference in success rates among 3 tests used in our evaluation which signifies the proposed CAPTCHA has a higher accuracy or success rate.

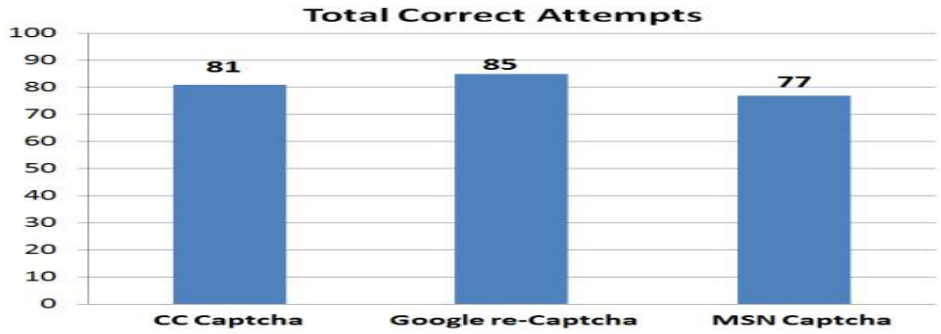


Fig. 6 .Total correct attempts out of 92 attempts

We notice that the success rate of Contextual Conversation CAPTCHA (e.g. 88.04%) is very close to re-CAPTCHA (e.g. 92.39%). This is actually surprising. We expected that solving a Contextual Conversation CAPTCHA challenge will be much harder than solving a re-CAPTCHA challenge because text-based CAPTCHAs have been widely deployed for a long time and users are quite familiar whereas users were experiencing our system for the very first time. This clearly suggests that our proposed design is pretty learnable.

VI. CONCLUSION AND FUTURE WORK

In this paper, we design a HIP named contextual conversation CAPTCHA just to balance the readability and security of the CAPTCHA challenge design. By adding distortion and noise we restrict BOTs to break the test, but it eventually also difficult to humans to solve. We overcame this problem by adding contextual information in the form of natural conversation without reducing the distortion and noise. We have validated our test through microscopic large-scale user study and find that our test is capable of being deployed in the internet. Moreover, the proposed approach is not suitable for visually impaired users and it may be challenging for users with learning disabilities. Our test may create problems for users who are not that much comfortable with English. So, our future work will undergo with the design of a native language CAPTCHA with the help of Google translator to localize the test, then that will be more usable to the users.

REFERENCES

[1] L. von Ahn, M. Blum, and J. Langford, "Telling humans and computers apart automatically," Communications of the ACM, vol. 47, no. 2, pp. 57-60, February 2004.

[2] Google Accounts, https://www.google.com/accounts/NewAccount.

[3] http://recaptcha.net/

[4] Almazayd, A.S. Ahmed , Y Kouchay, "Multi Modal CAPTCHA: A User Verification Scheme". Presented at Int Conf n Information Science and Application, Korea, 2011

[5] Rodden, T., Cheverst, K., Davies, K. Dix, A.. Exploiting Context in HCI Design for Mobile Systems. Workshop on Human Computer Interaction with Mobile Devices (1998)

[6] English Language taken from wikipedia and available at http://www.en.wikipedia.com/English_Language.

[7] Jeff Yan, Ahmad Salah El Ahamd. " Usability OF CAPTCHAs Or usability issues in CAPTCHA design"

[8] Fidas, C., Voyiatzis, A., and Avouris, N. On the Necessity of User-friendly CAPTCHA. In Proc. CHI 2011, ACM Press (2011), 2623-2626.

[9] Riding, R.J., and Cheema, I. Cognitive styles - An Overview and Integration. Educational Psychology 11, 3/4 (1991), 193-215.

[10] Brusilovsky, P., Kobsa, A., and Nejdl, W. The Adaptive Web: Methods and Strategies of Web Personalization. Springer, Berlin, Heidelberg, 2007.

[11] Brown, E., Brailsford, T., Fisher, T., Moore, A., and Ashman, H. Reappraising Cognitive Styles in Adaptive Web Applications. In Proc. WWW 2006, ACM Press (2006), 327-335.

[12] Van Oorschot, P. C. and Stubblebine, S. 2006. On Countering Online Dictionary Attacks with Login Histories and Humans-in-the –loop. ACM Transaction on Information and System Security. 9.3.August 2006:235-258.

[13] J. Yan and A. S. El Ahmad. A low-cost attack on a microsoft captcha. In Proceedings of the 15th ACM conference on Computer and communications security, CCS ¨08, pages 543–554, New York, NY, USA, 2008. ACM.

[14] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. ITube, You Tube, Everybody Tubes: Analyzing the World¨sLargest User Generated Content Video System. In Proc. IMC2007, ACM Press (2007), 1–14.

[15] K. Chellapilla and P. Simard. Using machine learning to break visual human interaction proofs (hips). In In Advances in Neural Information Processing Systems, pages 265–272, 2005.

[16] http://morphadorner.northwestern.edu/morphadorner/postagger/

[17] A. S. El Ahmad, J. Yan, and L. Marshall. The robustness of a new captcha. In Proceedings of the Third European Workshop on System Security, EUROSEC ¨10, pages 36–41, New York, NY, USA, 2010.ACM.

[18] T.Y Chan. "Using a Text-to-Speech Synthesizer to Generate a Reverse Turing Test", In Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, Sacramento, CA, USA, 2003, pp. 226-232.

[19] J. Elson, J. R. Doucerur, J. Howell, and J. Saul. Asirra: A Captcha that exploits interest aligned manual image categorization. In Proceedings of the 14th ACM conference on Computer and communications security, CCS ‟07, pages 366–374, New York, NY, USA, 2007. ACM.

[20] Esp-pix. http://server251.theory.cs.cmu.edu/cgi-bin/esp-pix/esp-pix.

[21] M. Chew and J. D. Tygar. Image recognition captchas. In Proceedings of the 7th International Information Security Conference (ISC), pages 268–279, 2004.

[22] Jakob Nielsen. Usability 101: Introduction to Usability, 2003. Available at http://www.useit.com/alertbox/20030825.html.

[23] Moin Mahmud Tanvee, Mir Tafseer Nayeem, Md. Mahmudul Hasan Rafee " Move & Select: 2-Layer CAPTCHA Based on Cognitive Psychology for Securing Web Services", International Journal of Video & Image Processing and Network Security IJVIPNS / IJENS vol. 11, issue 5 ,October,2011

[24] K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski. Designing human friendly human interaction proofs (hips). In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI‟05, pages 711–720, New York, NY, USA, 2005. ACM.

[25] Hotmail, https://signup.live.com/signup.aspx.

[26] http://en.wikipedia.org/wiki/Cognitive_psychology

[27] http://www.spellchecker.net/spellcheck/

[28] School of Computer Science. (2009, Dec.). Gimpy. Carnegie Mellon. [Online]. Available: http://www.captcha.net/captchas/gimpy/