PiCAPTion: Picture CAPTCHAs for Internet Authentication

Apoorva Rajendra Angre
Xavier Institute of Engineering
Mumbai, Maharashtra

Monal Dilip Kapadia
Xavier Institute of Engineering
Mumbai, Maharashtra

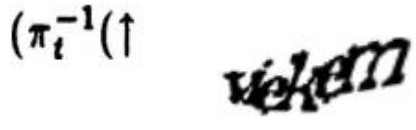
Meena Ugale
Asst. Prof of Information Technology
Xavier Institute of Engineering
Mumbai, Maharashtra

ABSTRACT
CAPTCHAs are tests that distinguish humans from software robots in an online environment. The most commonly encountered CAPTCHAs today take the form of a garbled string of words or characters. Unfortunately, existing text based CAPTCHAs suffer from several automated attacks. Thus, with the demonstration that character recognition CAPTCHAs are vulnerable to computer vision based attacks, this paper proposes alternatives to the traditional 'Text based CAPTCHAs', in the form of 'Image based CAPTCHAS', which require users to identify simple objects in the images presented, the argument is that object recognition is typically considered a more challenging problem than character recognition, due to the limited domain of characters and digits in the English alphabet.

Keywords
PiCAPTion, CAPTCHA, Bot, Security.

1. INTRODUCTION
CAPTCHA stands for "**C**ompletely **A**utomated **P**ublic Turing test to tell Computers and **H**umans **A**part". Since the dawn of the Internet, people have tried to abuse websites for both sport and profit. As the abuse became profitable, the scale of abuse grew using automated software (sometimes referred to as bots). In order to prevent bots from overrunning sites with spam, fraudulent registrations, fake sweepstakes entries, and other nefarious things, publishers responded by testing users to see if they were humans or not. CAPTCHA implementations can be found on more than 3.5 million sites globally, and human beings solve CAPTCHA implementations more than 300 million times a day.

The most common, old-fashioned way to make CAPTCHA "more secure" is to obfuscate the letters displayed with increasing severity. As bots have become better at bypassing CAPTCHA through OCR, most providers have responded by making their puzzles harder to read. This vicious cycle has, in turn, made it harder for humans (but more profitable for website owners).



[1]

What's worse is that some providers display unknown characters, foreign languages, and completely unintelligible "chicken scratch". [2]

[1]

Thus, there is a need for a better solution which not only verifies the user and grants access to desired content but also provides the user smooth access to the desired content. This paper will provide such a solution in the form of PiCAPTion (Picture CAPTCHAs for Internet Authentication),it will basically be an interactive image based CAPTCHA, which will prevent spam users from doing malpractices and also give users a CAPTCHA experience, which is less frustrating and more entertaining. It will use Graphics which engage human users, whilst differentiating human users from automated robots. PiCAPTion will save time on moderation of spam, improve content quality and add more usability to websites.

It will generate images for internet authentication. Thus, PiCAPTion will be an image-based CAPTCHA with the goal of making it easy for humans to read but harder for bots to crack. [3]

2. RELATED WORK
CAPTCHA is a verification code which Websites generally use to protect internet accounts against spam or other unauthorized account access. By entering the code**,** it is verified that the user is a human and not a spam-sending computer. [4]

Users may encounter a CAPTCHA code when:

- Sending an email containing HTML, links, embedded graphics, attachments
- Forwarding messages (chain letters, jokes, etc.)
- Performing too many failed sign in attempts.
- Creating a new web based account.

The most popular type of CAPTCHA currently used is text recognition; although they are widely used by many internet giants they have many drawbacks. (E.g. spammers could use software that would be able to recognize text embedded in the image and try all possible combinations to "break" the anti-spam mechanism), they are undoubtedly recognizable. [5]

This attack through Optical Character Recognition on text based CAPTCHAs can be best explained as explained below,

The normal text based CAPTCHA suffers from several weaknesses like fixed font face, fixed font size, no distortions, trivial background noise, and it's easy to segment. [6] In this experiment, a three step algorithm has been developed to break the text based CAPTCHA. The image is pre-processed

to remove noise using thresholding and a simple cleaning technique, and then segmented using vertical projections and candidate split positions. Four classification methods have been implemented: pixel counting, vertical projections, horizontal projections and template correlations. The system was trained on a sample of twenty text based CAPTCHAs to create thirty-six training templates (one for each character: 0-9 and A-Z). A separate sample of 100 text-based CAPTCHAs was used for testing. The following success rates have been achieved using the different classifiers: 8% pixel counting, vertical projections 97%, horizontal projections 100%, and template correlations 100%. [7]

Example:

Pre-process

Original: 

Grey Scale: 

Thresholding: 

Further Cleaning: 

Segment

Segmented:

Padded:

Classify

Pixel Counting: 8% Break Rate

Vertical Projections: 97% Break Rate

Horizontal Projections: 100% Break Rate

Template Correlations: 100% Break Rate[6]

Thus, this shows successful automated attacks have been developed against many existing text based CAPTCHA schemes. Algorithms have also been designed that can achieve character segmentation with a 90% success rate. Also text based CAPTCHAs may suffer from legibility issues.

Thus, basically three Image based CAPTCHAs have been implemented under PiCAPTion, the first one is a basic 'naming CAPTCHA' in which the user is given a set of images and he inputs his answer in the form of text and his input is compared with our large image database. Second CAPTCHA is basically an 'Anomaly CAPTCHA' in which again the user is given a set of images but here the user input is taken in the form of a single click and then compare with our large image database. Third CAPTCHA is 'Multi-Select' ,the user is given a question image and another grid containing images which are both similar and different to the image in the question, the user again has to click on the correct image , this is basically implemented by methods based on feature extraction. [8]

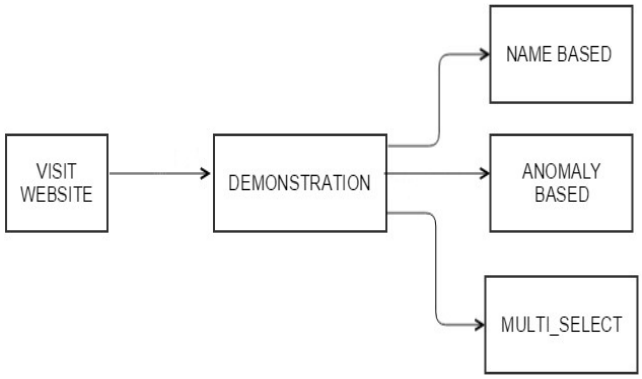


Fig 1: Types of captcha

3. ALGORITHM

- Images from the database are displayed randomly on the screen using random numbers.
- User's input is accepted.
- It is compared with a large image database using feature extraction and other methods.
- If the comparison matches, then the system outputs that the user is genuine and not a robot.
- Else a new CAPTCHA problem is displayed.
- User gets only 5 chances to solve the CAPTCHA, in each chance a new CAPTCHA problem is displayed before the user.

4. FLOWCHART

The System Flow as shown in Fig. 2 describes the overall flow of the system.

When any user visits a Website , he has to go through the registration process to access the resources of the Website. Any internet giant will want the user to solve a CAPTCHA when the user registers , to differentiate the human user from automated robots.

PiCAPTion will display the CAPTCHA after being processed from the backend.The backend of the system should go through the website's main Database to generate the CAPTCHA.

Fig 2:An example PiCAPTion

The user may then solve the CAPTCHA and input his answer.The answer that is inputted by the user may then be verified from the Database , if the entered answer is correct , the user proceeds to the next step of Registration.

But if the entered answer is incorrect, an entirely new CAPTCHA image is displayed before the user.The user gets only 5 chances to solve the CAPTCHA, in each chance a new CAPTCHA problem is displayed before the user. If the user fails to input the correct answer in the given attempts and asks for an attempt 6 , the system may reject the user as this

particular user might be an automated BOT. The main aim of giving only 5 chances is combating BRUTE FORCE attacks!

Also if the input is entered in less than 10seconds, then too there is space for speculation, whether the user is genuinely human or an automated BOT. Hence even in such a case, the system might reject the user, declaring it to be a BOT.

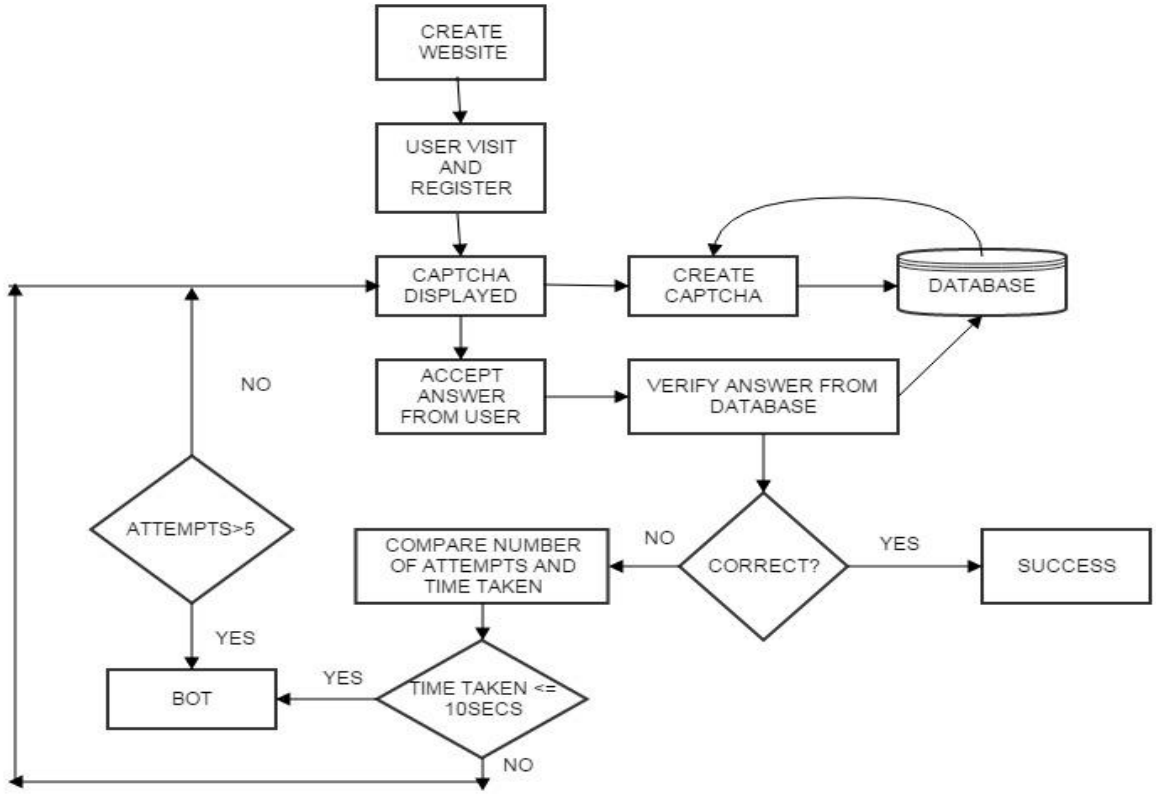


Fig 3: System flow

5. CONCLUSION

Thus, an improvement on current CAPTCHA systems was proposed by combining image processing techniques such as feature extraction with a knowledge-based component. The system asks users to identify certain images with a given set of questions and evaluates responses based on a certain level of tolerance to differentiate the user as human. In addition to this, the system was determined to be both interactive and user-friendly by allowing users the possibility of solving it with as little as two or three clicks with none of the usual strain. While the system described is a step in the right direction based on current offerings, it is still incomplete and thus a variety of steps need to be taken in the near future. As the system involves image generation, those same images can be used for advertising and marketing that is when a Company displays an image CAPTCHA for the user , the user gets a hint of the Company through the images in the CAPTCHA ,thus killing two birds in one stone, promoting its Company and authenticating the user through image generation.

6. ACKNOWLEDGMENTS

Monal Kapadia and Apoorva Angre thank our guide Asst. Professor Meena Ugale for her support and guidance.We also thank our dear friend Stalin Anudeep for his continuous encouragement.

7. REFERENCES

[1] http://solvemedia.com/security/captcha

[2] Ryan Doyle, "Image-based CAPTCHA with JACI", Centre for ICT 465 Elgar Road, Box Hill, Melbourne, Victoria Australia, November 2008.

[3] A.Krishnashanthi , Dr.K. Kuppusamy," Evolving New CAPTCHA using LCG Algorithm and Unpredictable Algorithm", International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 3, May-Jun 2012, pp.2258-2262.

[4] Ran Halprin," Dependent CAPTCHAs: Preventing the Relay Attack", Dec 16, 2007.

[5] Mayumi Takaya, Hikari Kato, Takeshi Komatsubara, Yusuke Watanabe, Akihiro Yamamura, ScienceDirect,The 9th International Conference on Cognitive Science.

[6] http://www.kloover.com/category/projects/

[7] Henry S. Baird and Terry Riopka, "ScatterType: a Reading CAPTCHA Resistant to Segmentation Attack ", Computer Science & Engineering Dept.

[8] Monica Chew and J. D. Tygar, "Image Recognition CAPTCHAs" In Proceedings of the 7th International Information Security Conference (ISC 2004), Springer, September 2004, pp. 268-279.

[9] Alessandro Basso,Stefano Sicco, "Preventing massive automated access to web resources", ScienceDirect.

[10] Bin B. Zhu, Jeff Yan, Guanbo Bao, Maowei Yang, and Ning Xu,"Captcha as Graphical Passwords?A New Security Primitive Based on Hard AI Problems",IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 9, NO. 6, JUNE 2014.

[11] Manar Mohamed ,Niharika Sachdeva, Michael Georgescu, Song Gao, Nitesh Saxena ,Chengcui Zhang, Ponnurangam Kumaraguru, Paul C. van ,Oorschot Wei-Bang Chen,"Three-Way Dissection of a Game-CAPTCHA: Automated Attacks, Relay Attacks, and Usability",Computer and Information Sciences, University of Alabama at Birmingham, USA Indraprastha Institute of Information Technology, India Computer Science, Carleton University, Canada Math and Computer Science, Virginia State University, USA.

[12] Abby A. Goodrum,"Image Information Retrieval:An overview of Current Research",Special Issue on Information Science Research Volume 3 No 2,2000

[13] Pranali Prakash Lokhande, P. A. Tijare,"Feature Extraction Approach for Content Based Image Retrieval",International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 2, February 2012 ISSN: 2277 128X