

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 4, Issue. 4, April 2015, pg.519 – 531

RESEARCH ARTICLE

SIMPLE TEXT BASED CAPTCHA FOR THE SECURITY IN WEB APPLICATIONS

Vidya P N, Dr. Shrinivasa Naika C.L

Computer Science and Engineering Department, U.B.D.T College of Engineering,
Davanagere, Karnataka, India

ABSTRACT

CAPTCHA is an acronym for Completely Automated Public Turing test to tell Computers and Humans apart. CAPTCHA systems are used as a security mechanism in web applications. Text-Based CAPTCHA is the most widely used CAPTCHA. Due to the wide popularity the text-based CAPTCHAs are more vulnerable to attacks. Hence there is an immense need to improve the strength of the text-based CAPTCHA to hinder many of such breaking attacks. This paper proposes a simple text based CAPTCHA approach that thwarts many of the breaking attacks and provides reliable and usable way to distinguish between humans and computers.

CHAPTER 1

INTRODUCTION

Presently, living has changed the dimension with the introduction of the Internet to mankind, ways people connect to each other, advertising, shopping, education, etc. Consequently, system security has become the most important issue for any websites since there

are many methods used to intrude the system over the internet. People have developed techniques, systems, programs and software systems that can replace a normal human being to do a job; such kinds of jobs include entering of data into systems, generate data automatically, handling events that occur on or within a system. As a matter of fact, web sites must ensure that the services are supplied to legitimate human users rather than bots to prevent service abuse. To thwart automated attacks, services often ask users to solve a puzzle before being given access to a service. Human Interactive Proofs (HIPs), focus on automation tests that virtually all humans can pass but current computer programs fail. Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) was an acronym that was coined in 2000. It is a type of challenge-response test that only a human completes successfully.

CAPTCHAs are designed to be simple problems that can be quickly solved by humans, but are difficult for computers to solve. Using Captchas, services can distinguish legitimate users from computer bots while requiring minimal effort by the human user . In the procedure, a computer or a program creates a test for its user, who is expected to be a human. The test is meant for the humans, that is, it is to be solvable only by humans and not any other machine, system or program. The user is required to provide a correct response to the test and then the user is permitted to access the work. When a correct response is received, it is presumed that the response arrived because of a human user.

CAPTCHA techniques have been classified into four categories:

- Text based Captcha.
- Audio based Captcha.
- Image based Captcha.
- Video based Captcha.

Each type is suitable to serve different group of users.

CHAPTER 2

LITERATURE SURVEY

Thomas V.A and Kaur K: proposed Cursor CAPTCHA–Implementing CAPTCHA Using Mouse Cursor, with the growing use of Internet and its services, a large number of organizations are making use of it to provide and seek information of the people using those services. This has raised the chances of attacks on such services by interrupting them sending multiple requests to the servers providing these services programmatically. CAPTCHA

(Completely Automated Public Turing test to tell Computers and Humans Apart) provides a way to differentiate users into humans and bots, or computer programs. This paper includes a new technique that utilizes image from custom mouse cursors and outperforms some most popular CAPTCHA techniques such as Text - based CAPTCHAs and previous Image - based CAPTCHAs.

Advantages:

CAPTCHA Using Mouse Cursor, with the growing use of Internet and its services, a large number of organizations are making use of it to provide and seek information.

Disadvantages:

This has raised the chances of attacks on such services by interrupting them sending multiple requests to the servers providing these services programmatically, so taking more time delay process.

Tamang T. and Bhattarakosol P: proposed Uncover impact factors of text-based CAPTCHA identification Presently, CAPTCHA is an important mechanism to gain access to the required system. However, there are some difficulties for users in typing CAPTCHA although they are authorized persons. Since the Text-based CAPTCHA is the most popular mechanism amongst all the CAPTCHA techniques, the difficulties of this text-based are studied and drawn out. The results of this study have shown that the presented character(s), genders of users, and their educational background are some of the important factors determining the correctness of CAPTCHA typing by its users. Therefore, generating a Text-based CAPTCHA must use the appropriate character, that also combining with the educational background and genders of the users.

Advantages:

Text-based CAPTCHA must use the appropriate character, that also combining with the educational background and genders of the user.

Disadvantages:

The CAPTCHA techniques, the difficulties of this text-based are studied and drawn out.

Sushma Yalamanchili, and M. Kameswara Rao: proposed a Framework for Devanagari Script-based Captcha, Human Interactive Proofs (HIPs) are automatic reverse Turing tests

designed to distinguish between various groups of users. Completely Automatic Public Turing test to tell Computers and Humans Apart (CAPTCHA) is a HIP system that distinguishes between humans and malicious computer programs. Many CAPTCHAs have been proposed in the literature that text-graphical based, audio-based, puzzle-based and mathematical questions-based. The design and implementation of CAPTCHAs fall in the realm of Artificial Intelligence. We aim to utilize CAPTCHAs as a tool to improve the security of Internet based applications. In this paper we present a framework for a text-based CAPTCHA based on Devanagari script which can exploit the difference in the reading proficiency between humans and computer programs. Our selection of Devanagari script-based CAPTCHA is based on the fact that it is used by a large number of Indian languages including Hindi which is the third most spoken language. There is potential for an exponential rise in the applications that are likely to be developed in that script thereby making it easy to secure Indian language based applications.

Advantages:

There is potential for an exponential rise in the applications that are likely to be developed in that script thereby making it easy to secure.

Disadvantages:

In this paper text-based process secure Indian language based applications not used other languages.

Baljit Singh Saini, and Anju Bala: proposed A Review of Bot Protection using CAPTCHA for Web Security, Today several daily activities such as communication, education, E-commerce, Entertainment and tasks are carried out by using the internet. To perform such web activities users have to register regarding the websites. In registering websites, some intruders write malicious programs that waste the website resources by making automatic false enrolments that are called as bots. These false enrolments may adversely affect the working of websites. So, it becomes necessary to differentiate between human users and Web bots (or computer programs) is known as CAPTCHA. CAPTCHA is based on identifying the distorted text, the color of image, object or the background. This paper examines CAPTCHAs and its working and literature Review. This paper also provides classification of CAPTCHAs, its application areas and guidelines for generating a captcha.

Advantages:

In this method some objects are chosen randomly and the pictures about these topics are searched and downloaded from the Internet.

Disadvantages:

The drawback is not prone to dictionary attacks because it does not include dictionary words

.Gossweiler R, Kamvar M, and Baluja S: proposed CAPTCHA Based on Image Orientation; 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. 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.

Advantages:

It is language-independent, lends itself to rapid implementation and has an almost limitless supply of images.

Disadvantages:

Does not require text-entry, and employs another domain for CAPTCHA generation beyond character obfuscation.

CHAPTER 3

THE PROPOSED SCHEME

During the design stage the main principles which play an important role for providing a more robust CAPTCHA. In our proposed scheme, multiple secure features which are extremely effective to obfuscate challenges for the breaking attack but easy to solve by users have been applied. Our processing will disallow replay of previously submitted challenge (never be reused), thus for every time the page is generated (or refreshed), the principle features will be changed as the following:

- CAPTCHA's code is a series of characters (uppercase and lowercase) and numbers.
- Multiple randomizing functions are used to generate a random code (stream of characters and numbers) in each challenge in order to make it not susceptible to a dictionary attack.
- The length of the code is varied (minimum length is 6 characters-numbers).
- Multiple font types are handled to prevent intrusion using image processing techniques when a consistent font is used.
- String/code are rotated at various angles.
- Lines are utilized to prevent segmentation. The numbers and the length of lines and their positions are varied each time in order to distort the text image randomly before being presented to the user.
- The text image is blurred using a specific technique in order to make CAPTCHA difficult for malicious software.
- Image dimensions are varied inconsistent with all the features above CAPTCHA's code and line colour are kept in grayscale colours but at a different level for each time.

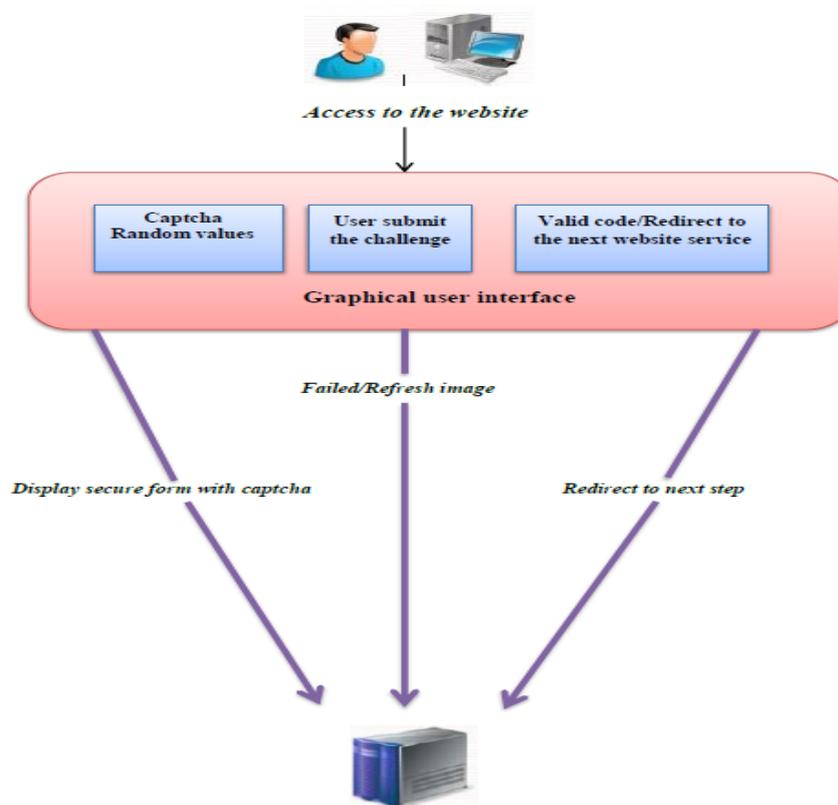


Fig Framework Of CAPTCHA

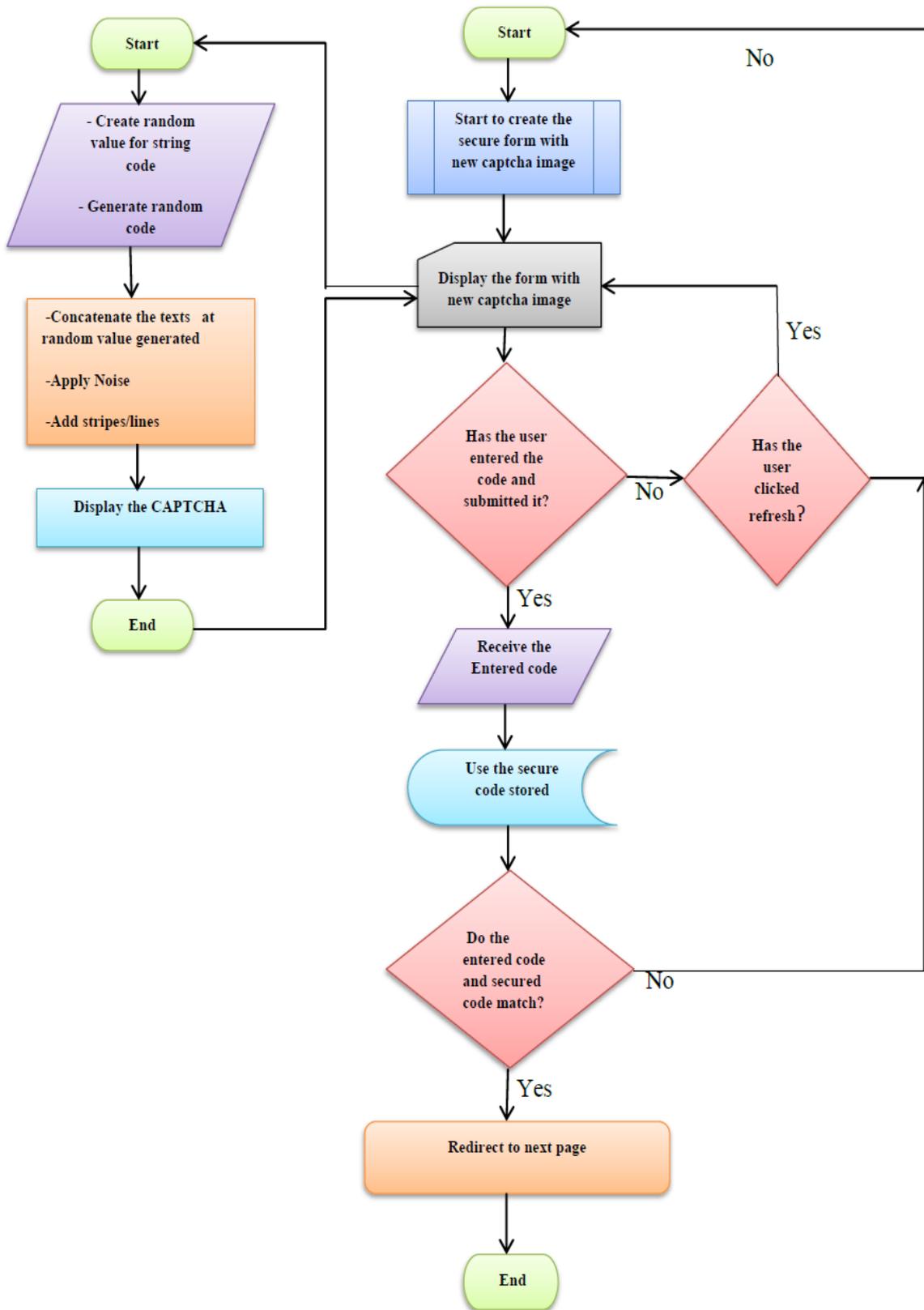


Fig Flow chart

CHAPTER 4

IMPLEMENTATION

Snapshots



Fig GUI

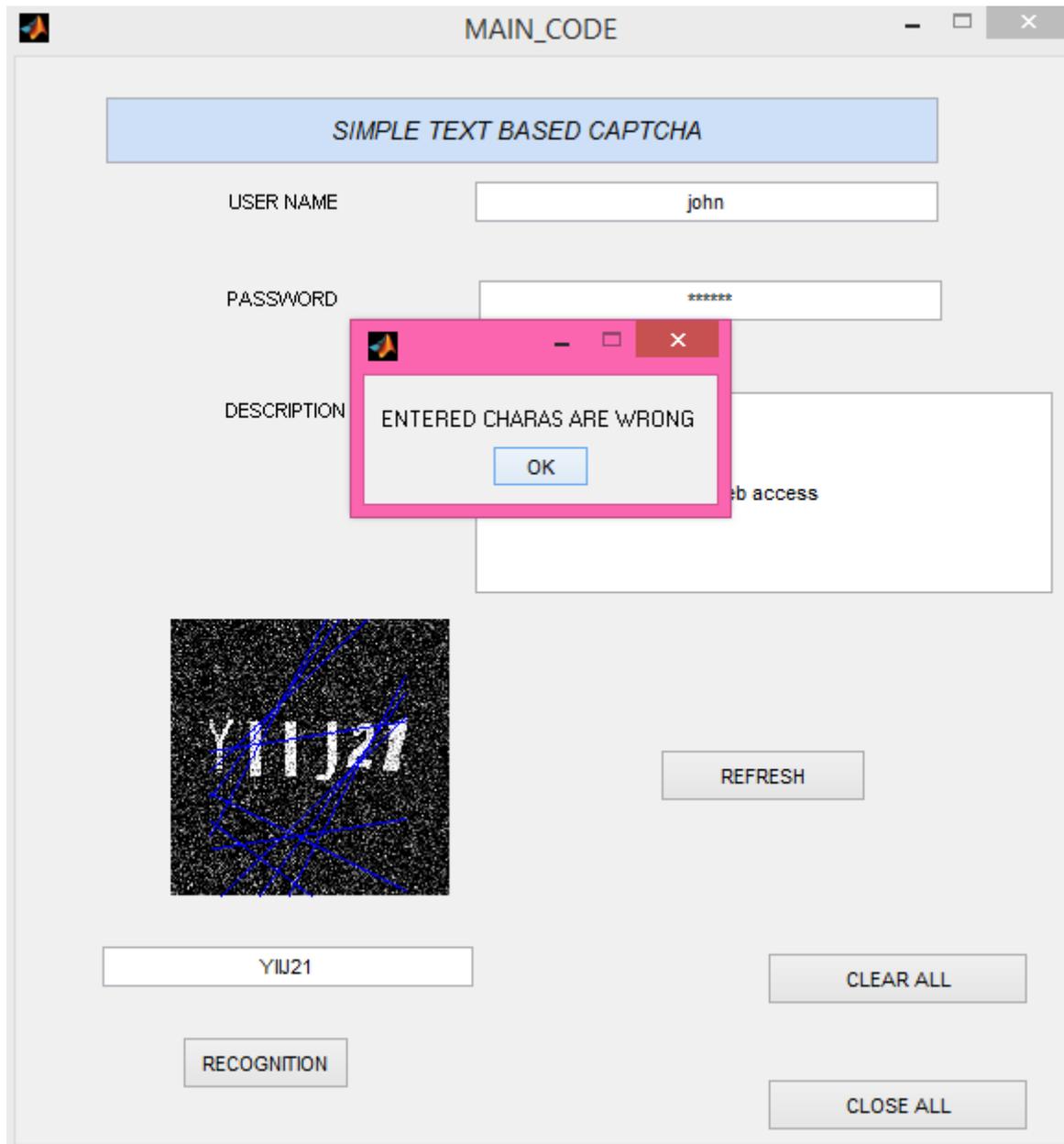


Fig Showing entered captcha is wrong

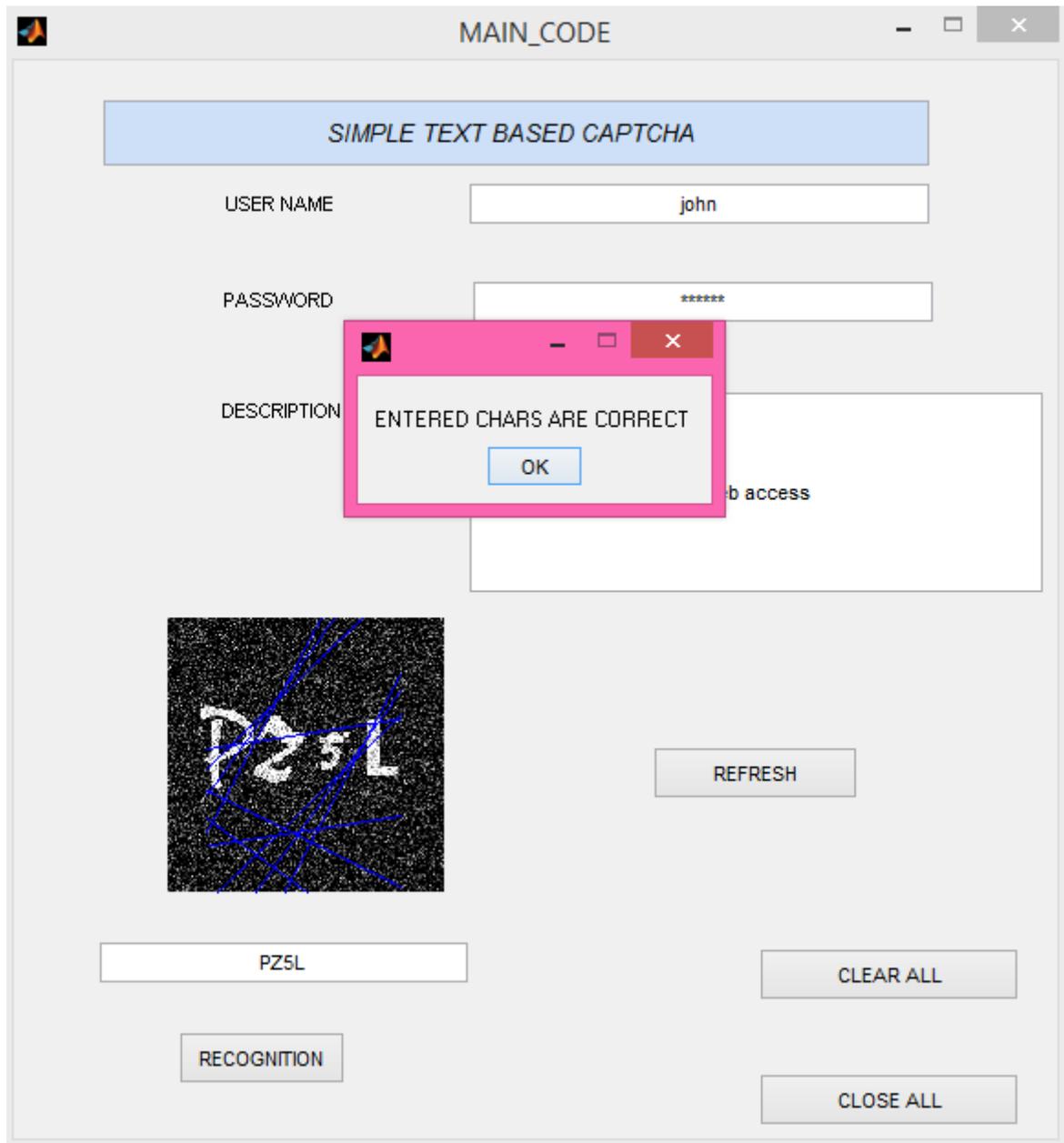


Fig. Showing entered captcha is correct

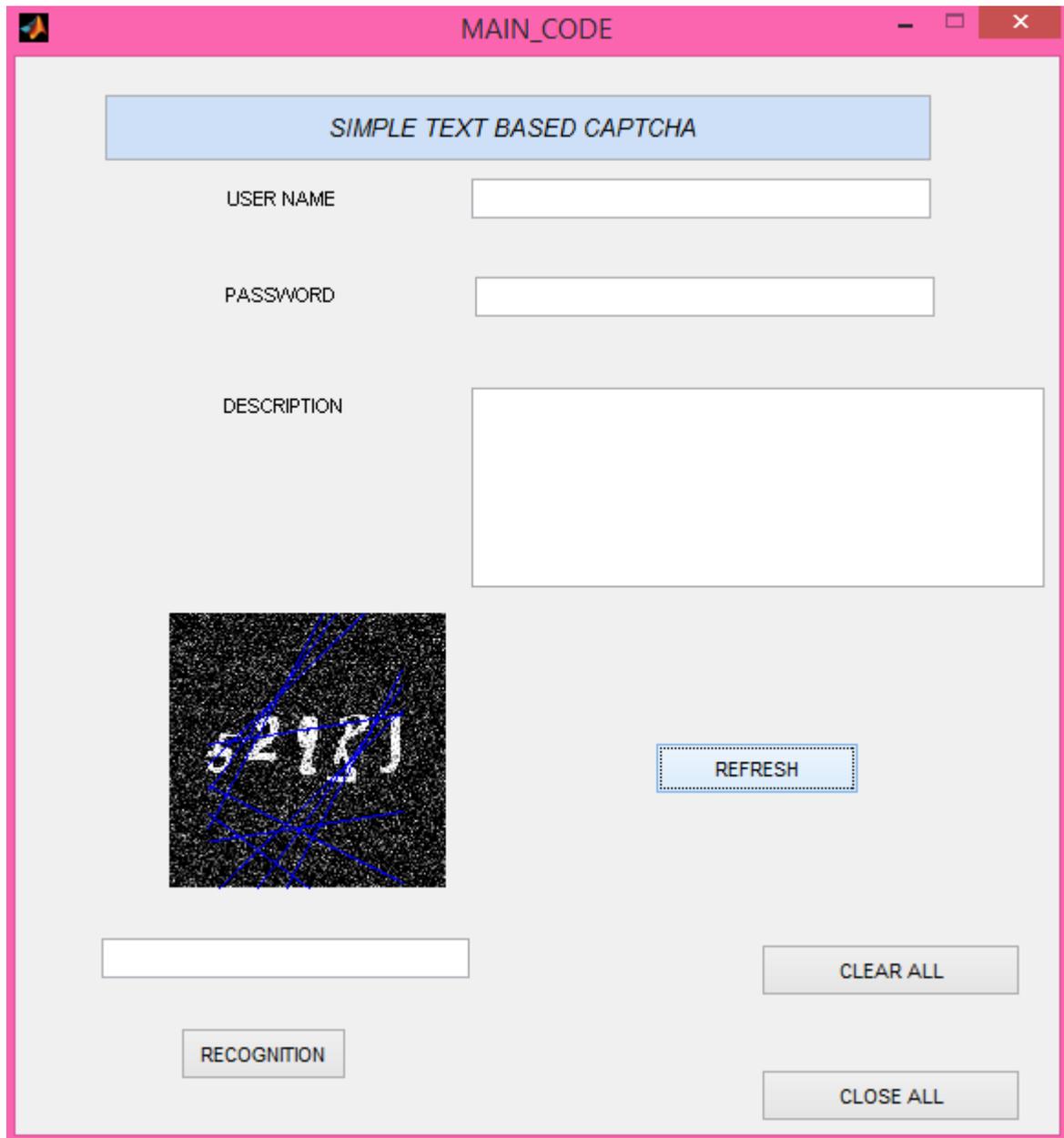


Fig. New CAPTCHA on refreshing

CONCLUSION

As a contribution toward improving the web security in the field of an automated challenge and response against attacks issued by automated programs, we proposed a simple text_based CAPTCHA. Two main goals have been considered to be achieved that is: simplicity of solving the technique for a human as well as the time that a human actually needs to find the solution.. To increase the difficulty for segmentation and recognition attacks on Captchas, we varied these significant features at each challenge in ranges potentially acceptable to human users. Our mechanism provides a solution to maximize the robustness and usability of text-based CAPTCHAs simultaneously.

REFERENCES

- [1] Mumtaz M. Ali AL-Mukhtar and Rana Riad K. AL-Taie : A More robust text based captcha for the security in web applications.
- [2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.
- [3] E. Bursztein and S. Bethard. Decaptcha: breaking 75% of eBay audio CAPTCHAs. In *Proceedings of the 3rd USENIX conference on Offensive technologies*, page 8. USENIX Association, 2009.
- [4] E. Bursztein, S. Bethard, Fabry C., Dan Jurafsky, and John C. Mitchell. Design parameters and human-solvability of text-based captchas. To appear.
- [5] Elie Bursztein, Steven Bethard, John C. Mitchell, Dan Jurafsky, and Celine Fabry. How good are humans at solving captchas? a large scale evaluation. In *Security and Privacy*, 2010.
- [6] J. Canny. A computational approach to edge detection. *Readings in computer vision: issues, problems, principles, and paradigms*, 184:87–116, 1987.
- [7] K. Chellapilla, K. Larson, P.Y. Simard, and M. Czerwinski. Computers beat humans at single character recognition in reading based human interaction proofs (hips). In *CEAS*, 2005.
- [8] K Chellapilla and P Simard. Using machine learning to break visual human interaction proofs. In MIT Press, editor, *Neural Information Processing Systems (NIPS)*, 2004.

[9] K. Chellapilla and P. Simard. Using machine learning to break visual human interaction proofs (HIPs).

Biographical Notes

VIDYA P N is presently pursuing M. Tech. final year in Computer Science and Engineering Department (Computer Science) from UBDTCE, Davangere, India.

Dr.SHRINIVASA NAIKA C.L is working as an Assistant Professor in Computer Science and Engineering Department, from UBDTCE, Davangere, India