Different Combinations of Color and Noise in Captcha Generation

[1]Kanika Singhal, [2]R S Chadha
[1,2]CDAC, Noida, UP, India

Abstract
Today, internet has become a global tool for accessing services whether it is education, entertainment or e-commerce. In order to register on these sites, a distorted image of pseudorandom letters and digits is to be entered at the end of the form in order to gain access to the service. That distorted image is called Captcha(Completely Automated Public Turing Test To Tell Computers And Humans Apart).In this paper we are presenting a technique of the combination of noise and color in Captcha which will make it more human friendly and difficult for automated programs to break the Captcha.

Keywords
Captcha, Noise, Automated Programs

I. Introduction

A lot of publicly available services on the Internet is a boon for the community. But unfortunately it has also invited new and novel abuses. Automated Programs (bots and spiders) are being created to steal services and to do fraudulent transactions. Free online accounts are being registered automatically many times and are being used to distribute copyrighted material. Recommendation systems are also vulnerable to artificial inflation or deflation of ranking. Automated programs enter these sites and sign up for the massive number of accounts that is used to send spam email. This leads to wastage of resources on sites .For this purpose Captcha is provided. It is used to verify whether the end user is human or not. Captcha methods were originally designed to prevent spammers from registering free accounts in the free email services. Another main application of
CAPTCHA methods is in bot detection. Bots are computer programs which simulate human behaviors. Captcha methods are used in different applications to identify bots. These methods identify bots to ban them from using services or to block them in automated attacks such as Denial of Service (DOS) attacks.

II. Related Work

There are several types of Captcha used today. Many OCR and non OCR methods have been proposed. CAPTCHA is now almost a standard security mechanism for defending against undesirable or malicious Internet bot programs, such as those that spread junk email and grab thousands of free email accounts.

A. Gimpy

Gimpy works by selecting several words from a dictionary and displays them, corrupted and distorted, in an image. Users must then enter the words in the image to gain entry to the service

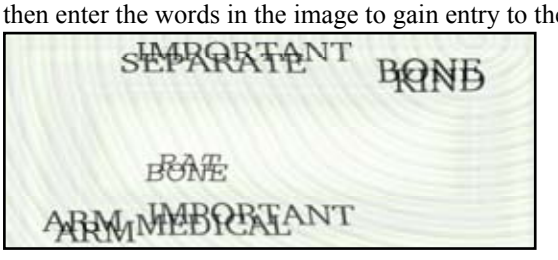

Fig. 1: Gimpy Captcha

B. Bongo

Bongo is based on a visual pattern recognition problem. A Bongo Captcha uses two sets of images; each set has some specific characteristic. One set might be boldface, while the other is not. The system then presents a single image to the user who then must specify the set to which the image belongs.

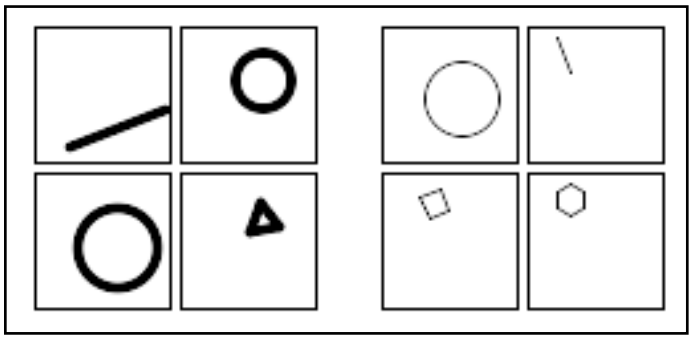

Fig. 2: Bongo Captcha

C. HVS

HVS masking characteristics Captcha is composed of English alphabets that are picked randomly and written with a combination of texture and edges with added noise such as to deceive the bots by randomly choosing the visibility of characters.

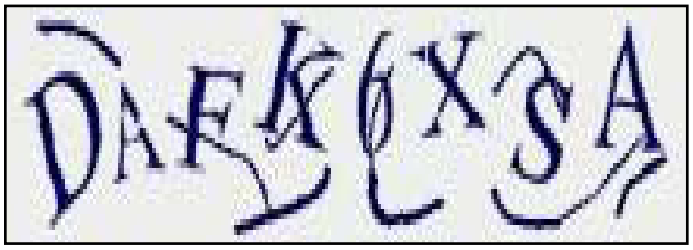

Fig. 3: Hotmail Captcha

D. Persian/Arabic

Persian/ Arabic Captcha uses Persian Arabic words. This method suggest the use of Persian words in which connected letters, presence of dots and right to left reading order makes program to fail. The major drawback is its limited domain users.

Question based Captcha is simple question based test contains mathematical problems. In this instead of some object name, image of the object is used.

Handwritten word Captcha contains database which is handwritten names of American cities selected from letters posted by people. But drawback is that due to poor quality of words to read for the human user

III. Proposed Approach

Different algorithms are used for different Captcha. The first step to create a CAPTCHA is to look at different ways humans and machines process information. Machines follow sets of instructions. If something falls outside the realm of those instructions, the machines aren't able to compensate. Similarly, it's unwise to build a CAPTCHA that doesn't distort letters and numbers in some way. An undistorted series of characters isn't very secure. Many computer programs can scan an image and recognize simple shapes like letters and numbers. Other CAPTCHA applications create random strings of letters and numbers. Using randomization eliminates the possibility of a brute-force attack — the odds of a bot entering the correct series of random letters are very low. The longer he string of characters, the less likely a bot will get lucky.

CAPTCHAs take different approaches to distorting words. Some stretch and bend letters in weird ways, as if looking at the word through melted glass. Others put the word behind a crosshatched pattern of bars to break up the shape of letters. A few use different colors or a field of dots to achieve the same effect. In the end, the goal is the same: to make it really hard for a computer to figure out what's in the CAPTCHA. Designers can also create puzzles or problems that are easy for humans to solve. Some CAPTCHAs rely on pattern recognition and extrapolation. For example, a CAPTCHA might include series of shapes and ask the user which shape among several choices would logically come next. The problem with this approach is that not all humans are good with these kinds of problems and the success rate for a human user can go below 80 percent.

In this paper, we have presented a new technique of combination of color and noise in the Captcha. Noise can be of different types. It can be in the form of intersecting lines, dots or zigzag lines. Similarly the shades of the colors can also be different. We can vary these noise and color with every Captcha image. This will be human friendly as well as difficult for automated programs to break the Captcha.

A. Background Interference Layer (Image, Noise)

Noise can be of different types.

1. Intersecting Lines

The noise can be in the form of intersecting lines, which interfere with the context letters and make it difficult to understand.

2. Dots

Noise can also be in the form of dots. These dots blurred the context letters and makes it difficult for machines to understand.

3. Circles

The noise can also be in the form of circles. This is human friendly but hard for machines to understand.

These noises are human friendly, humans can understand the characters, but it is difficult for machines to understand.

Similarly is the case with colors. Instead of using same color in each Captcha image we can use different color in each Captcha generated in the form. Human can easily adapt to these changing colors and noise, but it will be difficult for automated programs to break the Captcha when the color and noise will change with every Captcha image.

We can summarize this in the form of table

Table 1:

Captcha	Color	Noise
Captcha1	Purple	Intersecting lines
Captcha2	Light Purple	Dots
Captcha3	Dark purple	Circles

In this way, the color and noise of Captcha will change randomly with every changing image of Captcha. This will make almost impossible for automated programs to break the Captcha.

IV. Analysis of the Proposed Approach

In our approach, we have used different combination of color and image with every changing Captcha image. As shown in the table Captcha1 will have purple color and intersecting lines as noise. When the Captcha image will be changed after refresh, it will have different color and different form of noise as shown in the table. This will be easy for humans as humans can more easily adapt with changing colors and noise, so they can easily identify the distorted characters. But it will be difficult for automated programs to break the Captcha, as they have to apply different processes to scan and break the Captcha with every changing image. Different combination of color and noise will make the task of breaking the Captcha impossible for them. This will add more to the security and will give unbelievable results.

V. Conclusion

Security has been of great importance since last few decades. Automated programs have been designed to attack various services. The bots steal thousands of free e-mail accounts in a minute, and skew results. Over the past few years, an increasing number of public web services have attempted to prevent exploitation by bots and automated scripts, by requiring a user to solve a Turing-test challenge commonly known as a CAPTCHA before using the service. Most humans can easily pass the test but computers cannot. The most widely deployed CAPTCHAs are text-based schemes, which require a user to enter the text to gain access to the service. In this paper we have generated a Captcha which will have different combinations of noise and color. The noise and color will be changed with every Captcha refreshed .It will be

human friendly as they can easily adapt themselves with different combination of colors and noise, But it will be a real headache for automated programs to scan and apply processes to every changing image and break the Captcha.

References

[1] Jon Bentley, Collin mallows,"Analysing and improving challenge strings for Captcha DRAFT 0.27".

[2] Prem Shankar Yadava, Chandra Prakash Sahu, Sanjeev Kumar shukla,"Time variant Captcha :generating strong Captcha security by reducing time to automate computer programs", journal of emerging trends in computing and information sciences, Vol. 2, No. 12 December 2011.

[3] [Online] Availavle: http://www.golang.org/doc/codewalk/markov.go?h=os

[4] Kwan Woo Park,"Analysis of Captcha", Computer Science Department University of Southern California Los Angeles, CA 90089-0781 USA.

[5] Clark Pope, Khushpreet Kaur,"Is it human or computer: defending ecommerce with Captchas".

[6] Ahmad Salah El Ahmad, Jeff Yan,"The Robustness of a new captcha", school of Computing Science Newcastle University, UK.

[7] Yong Rui, Zicheng Liu,"ARTIFACIAL: Automated Reverse Turing test using FACIAL features", Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA.

[8] Ahmad S El Ahmad, Jeff Yan, Mohamad Tayara,"The Robustness of Google CAPTCHAs".

[9] Jeremy Elson, John R. Douceur, Jon Howell, Asirra: A CAPTCHA that Exploits Interest Aligned Manual Image Categorization", Microsoft Research.

Miss kanika singhal received his B.Tech in CSE from Apeejay college of engineering, India in 2011.Currently, she is doing M.Tech in CSE from C-DAC Noida(Affiliated to G.G.S.I.P.U New Delhi), India. She is working on the project "Captcha generation for secure web services". Her interest areas are Digital Image Processing, Operating Systems, and DBMS.

Mr. Ramneet Singh Chadha is an Assistant. Prof, MTech Head and currently working as Project Manager, in Health Informatics group. He has more than 12 years of domain expertise in Health care domain and has been closely involved in Design, Development and Implementation of e-Sushrut HMIS Software, since his joining C-DAC in 1997. He has interest in interoperability of hospitals for setting up NHIN using HL7 standards; cloud computing and telemedicine and other research area is health domain.