ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(42), DOI: 10.17485/ijst/2016/v9i42/97758, November 2016* ISSN (Online) : 0974-5645

Comparative Study of Algorithms used in CAPTCHA's and New Finding Set as LCG Algorithm

A. Clementeena[1*] and P. Sripriya[2]

[1]Vels University, Chennai – 600117, Tamil Nadu, India; teena8790@gmail.com
[2]School of computing sciences, Vels University, Chennai – 600117, Tamil Nadu, India; sripriya.phd@gmail.com

Abstract

Objectives: The completely automated public Turing tests are used generally to find the difference between the users whether he/she is a human or a bot. **Methods:** Once the randomly generated characters are held then it is converted into ASCII characters and this method will generate symbols, numbers and alphabets. As this very tedious one which is very difficult to be understood by bots so this method is very effective and robust**. Findings:** Number of authentication purpose is found out by misplacing the word's style and fonts. Very simple issues can be solved by this captcha's approach. The methods are very complicated for humans to understand but on the same side, it is easy for robots to keenly understand. So a new algorithm is used to find understandable captcha's into tougher one. Also, the LCG algorithm providesonly the designing principles to be followed in it by giving random function use. **Application/Improvements**: The main improvisation done here is that the application is built with more secure line and its usability is too high

Keywords: ASCII, CAPTCHA, LCG Algorithm

1. Introduction

The Internet is the great boon to this world and without internet usage, nothing can be done. For all-purpose we need internet and these purposes include on-line transaction, banking, education, communication and entertainment. The users need to register their observation in websites. Registration can be done by automated hacking software and thus, software relies on false acceptance which makes most of the resources occupied all time and thus reduces the performance and efficiency of the servers and even stops the entire web services. It is accepted that a good fine CAPTCHA's[1] must definitely address two major requirements: Robustness and usability. Robustness is the capability to resists through computer attacks and usability is the ease with which any human can pass its toughest challenges. CAPTCHA's are some challenge puzzles that is used to find whether the user is a human or a bot. CAPTCHA's are the programs that can generate and grade those tests that the human can pass the test but the current running computer programs cannot pass. The goal of this test is that the human can pass easily but bots cannot.

There are three types of CAPTCHA:

- Text-based Scheme[2] is used to check with a stream of text or data added with some disturbances or noise to confuse the user so that the user could find the puzzle easy or tough.
- The Sound-based Scheme is used to overcome the text-based scheme by intruding some audio or voices to find whether the user is human or a bot. Text-based scheme is not applicable to some people having impairment in their vision so its difficult to find out.
- Image-based Scheme[3] is used to check with some difficult images that have different colours and pixels but in that case, bots will easily find the other text-based scheme and audio-based scheme so to overcome this issue image-based scheme have been introduced.

The related work tells us about the CAPTCHA's improvisation in it and at first CAPTCHA's was introduced in 1997 which was devised by Andrei Broder. In the same year, Alta Vista website was using this CAPTCHA's method to distinguish between the computer programs or bots. Some distorted text will be displayed and that will be made to type into the box to check whether it is abot or

* *Author for correspondence*

human. Distortion is needed to make the OCR programs to find more difficult. CAPTCHA's can be classified into two ways as OCR based and NON-OCR based. After CAPTCHA was introduced several other methods was proposed side by side. Ibrahim Furkan Ince, Ilker Yengin, Yucel Batu Salman introduced a method or algorithm to splitting and rotating images against the OCR's machine methods.

2.1 "Handwritten CAPTCHA" Generation Algorithm

Handwritten text faces more challenges comparative to machine-printed text.More problems are faced in machine-printed text rather than handwritten text because handwritten can be distorted in any ways like given in Figure 1 to confuse the computer programs but in that case, machine-printed cannot satisfy. In this paper, we present an application called Human Interactive Proof's (HIP) which is used to defend online services to restrict the unwanted attacks of an intruder. This HIP uses set of protocols like Challenge-Response protocol that has series of automated tests relies on it which will make the human to pass the tests but, never have possibilities for computer programs or bots to pass the series of test. Some of the algorithmic approaches were introduced in this handwritten captchas such as Lexicon driver, lexicon free, parallel classifiers and combinations, pre and post processing routines, analytical and holistic methods. Their goal was to introduce the handwritten captcha's to check whether humans can pass or bots can do it.

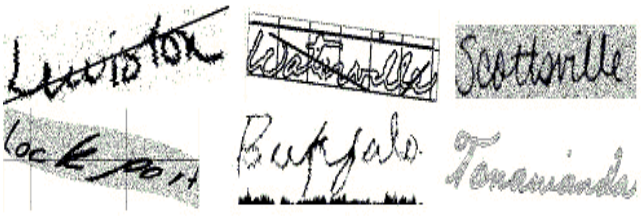

Figure 1. Examples given for handwritten CAPTCHA's.

In Figure 2 there are more word images which are very difficult to find out even not by computer recognition techniques and thus they are given with some small lists of words called lexicon. We use the handwritten word recognizers at the dispersal to conduct experiments and to come up with the arguments that are used to do automatic generation of word images and distortion of handwritten word images.

Image	Transformation	WMR	HMM
W. SENECA	Add noise	Recognizes	Fails
W. SENECA	Add noise, apply median filter	Recognizes	Fails
W. SENECA	Run chain code, empty letters	Fails	Fails
W. SENECA	Linear transformation	Recognizes	Recognizes
W. SENECA	Edge detector multiple times	Fails	Fails
W. SENECA	Affine transformation	Fails	Recognizes
W. SENECA	Empty letters, edge detector	Fails	Fails

Figure 2. Testing results of various word images.

We have created images in TIFF and HIPS format and also we have generated images with the help of two-word recognizers namely WMR and HMM. Both recognizers will match a word against its lexicon.

There were a sequence of tests held between word images of handwritten and word images with "nonsense" as given in Figure 3. what happen was the test with the handwritten was easy and legible so it was easy for them to find out but in nonsense word image case the letters were confused and they came out with many mistakes and for eg:'g'vs'q',e'vs'c',r'vs'n'.

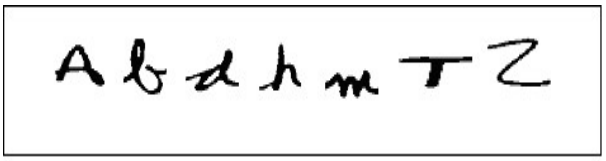

Figure 3. Word images with random nonsense letters.

2.2 Splitting and Rotating Algorithm

This method was introduced in 2011 to make the easier vision of bots and to distinguish between the human and computer programs. This algorithm works as splitting images[4] into several modules and each having their own random rotation values and drawing random lines for each of the random values in the formation of grid background. Thus the random lines have same colours and thus with distortion of images developed the Robustness feature of evolving new CAPTCHA's and also given more importance to the security of issues which was deployed

by Mega upload, a leading online storage block, and a delivery website. For example, e-mail service providers[5,6] like yahoo and g-mail activate their registration by this CAPTCHA as the last process to proceed. This CAPTCHA will surely help to stop the bots that activate hacking in the system and also helps in redundancy of user who distribute spam mails and messages. This paper intrudes out in splitting the image into several chunks in random height and width values and those split characters will be rotated in their random rotation angles that have particular place alone distorted with so as to confuse the user. It is very much difficult for the OCR's to find out where the image is being distorted and in which angle.

Carnegie Mellon designed Gimpy methods that do by selecting a particular word from the dictionary and ask the user who registers to login to type the word being seen in the image and make it reflect correctly what he has expected. In Figure 4, yahoo uses an easy way to render this method as EZ-Gimpy. In many cases, ahuman can read more than three images but bots cannot do this issue so it could easily track out.

Figure 4. Examples of CAPTCHA images for Yahoo, Gmail and Hotmail.

In this method characters are split into various blocks and so each has its own twist with rotation values and thus like it appears with certain angle domain interval as [-1, 1] , [-3, 3] [-5, 5]. Image parts are also split up and then they have been rotated with same rotation values and thus OCR programs find a tedious job in finding the start and end of the images. Rotation is included so as to confuse the user and so human can be differentiated from the computer programs or bots.

The text shown in in Figure 5 is actual 'W9XZq' but it is rotated with its rotation angle value so as to confuse the user. This CAPTCHA image is split up into 8 parts as of with (4x2) matrix and each split has its random rotation angle value between -3 and 3 degrees and also the background colour and the word image should be alike so as to confuse the user.

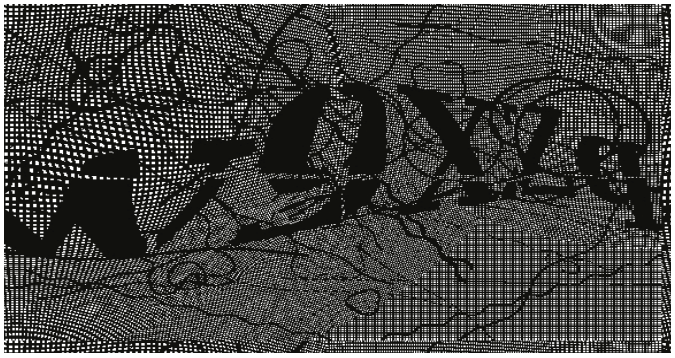

Figure 5. An CAPTCHA image with its rotation angle in it.

3. Linear Congruential Generators Algorithm

3.1 Algorithm General Procedures

Random NumbersàA series of numbers or a group of numbers which does not show relativity anywhere to each number within the series. The random numbers can be differentiated into two types as uniform and non-uniform random numbers. Generally

For all purpose, the uniform random number is referred and so it is based to retrieve some numbers and so one number is generated for usage of LCG.

D. H. Lehmer in 1948 proposed and paved way to generate numbers in

Random as a source and so he found this algorithm LCG. In LCG, each single number retrieved determines its own successor. The form or way the generator is given as,

Where m is called modulus, X_i, a and c are called as seed, multiplier, and increment as given.

Example: 1. Consider m = 31, a = 7, c = 0 and begin with .The next integers in the sequence are 9, 1, 7, 18, 2, 14, 5, 4, 28, 10, 8, 25, 20,16,19,9,1,7,18,2,14,5,4,28,10,8,25,20,16

And thus in the above examples random number generation is generated very bad and so it tend to repeat the same numbers which is being generated already

2.consider in second example with a change of

number as a=3 instead we would get the series of random numbers as 26, 16, 17, 20, 29, 25, 13, 8, 24, 10, 30, 28, 22, 4, 12, 5,15,14, 11, 2, 6, 18, 23, 7, 21, 1, 3, 9, 27 is of a good random number generator.

So from both the examples given we come to a conclusion that a good or bad random generator is based on the constant values such as a, c and m.

3.1.1 LCG Conditions

The period given is the smallest positive integer λ for which,

- **Condition 1:** The period cannot be greater than m value and so m is chosen as equal value or nearly equal one and so m is called as long period according to the computers.
- **Condition 2**: A full period generator is the one in which the normal period m and it is delivered as if:
- C is relatively prime to m value
- The multiple integers of q. For each prime factor q of m.
- Is a multiple of 4if m is given.
- **Condition 3:** If the increment, we can surely achieve a full period by such a follow-up.
- 1→very faster arithmetic operation.
- 2. Set b as high value as possible and for example b value is 63 in a 64-bit computer.
- 3. C value should be an odd-valued one.
- 4. set) as a multiple of 4.

3.2 Proposed Method

New unpredictable algorithms have been found out with the help of the LCG algorithm. This algorithm does procedure in following sequence as thus retrieving random numbers first and then converting it into ASCII character and this algorithm will help CAPTCHA's to generate alphabets, symbols, and numbers.as in procedure the first step is to take a constant value for LCG. Here we have used ASCII character of 256 in framing captchas .So that module m value is 256 as in condition (3.1) .considering c should be an odd-value as told before in condition (3.3).considering ,multiple of 4.

as in condition (3.4) and so we initialize X value as 1. Thus the LCG formulae can be applied with these constant values a. m and c to get the final random numbers.

92 119 82 237 72 99 62 217 52 79

2 42 197 32 59 22 177 12 39 2 157

3 248 19 238 137 228 255 218 117 208 235

But in the case of CAPTCHA delivering this output, it won't be that much effective. But their it contains only numbers which we are going to convert the numbers into ASCII values as follows.

\ w R í H c > Ù 4 O * Å

Space ; SYN ± FF ' STX ¥ ø DC3

Î ‰ ä ÿ Ú u Ð ë

And so from the above delivering it will very effective that bots or computer programs struggle to find out, but it's even more difficult for human to deliver because these symbols are not directly visible in the system and so to overcome this problem we need to consider only ASCII values from 33 to 126 and to neglect other values.so the above random numbers could be given as follows,

\ w R H c > 4 O * ; u

This text given comprises of five characters and each character has its own bending and size value

CAPTCHA 1: Displays first five characters

\ w R H c

CAPTCHA 2: Displays next five characters

> 4 O * ;

There is as many as examples given for captchas. The Random number[7,8] generator finds a very easy method with infinite set and endless loop hold to commence again and again in it. One major advantage in this generator algorithm is that it avoids redundancy of values. The main finding of this generation algorithm is that in

4. CAPTCHA Image Generation Process

The CAPTCHA Image Generation Process behaves differently for different applications and thus it works out very effective. There are some steps to be followed to make this process more effective

STEP 1: Create a CAPTCHA image in two-dimension as given in Figure 6 because it is so much enough for holding characters that come out as an output of LCG.

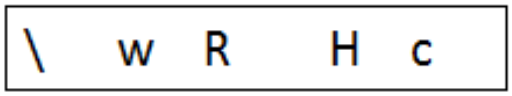

Figure 6. Step 1.

STEP 2: In Figure 7 Add some confusing techniques to uphold the characters and thus give some background colour as dark granite and letters colours in black so as to confuse the user while finding out. Thus, these techniques help the human to find out easier because the colours are alike the RGB spectrum colours which are difficult for the computer programs and bots to find out.

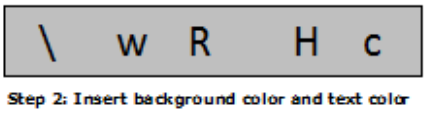

Figure 7. Step 2.

STEP 3: In step 2, we make the user confuse a lot as so to change the background and text colour as same in Figure 8 which will increase the rigidity of the user.

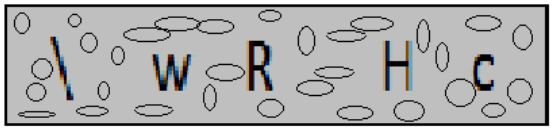

Figure 8. Step 3.

STEP 4: While designing CAPTCHA we need to give characters in small and also non-confusable characters as in Figure 9 and also there are 16 different font style which will confuse the user and thus eliminates the usage of bots.

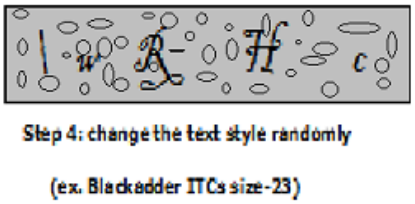

Figure 9. Step 4.

STEP 5: Then the word set is divided into several pieces as of in Figure 10 and each piece will have separate rotation with a random rotation value and the domain angle[-1, 1], [-3, 3], [-5, 5] in which it is rotated so as to confuse a lot the bots. The computer programs cannot find the word if it is dislocated at worst.

Figure 10. Step 4.

STEP 6: Atlast the worse more technique to confuse the computer program is to add random noise or disruption happened while entering the security tests.

Thus with all these enhancements in CAPTCHA[9,10] it becomes more difficult for the bots to find out. The old methods like handwritten, audio-based CAPTCHA was easily cracked out by the bots but this technique will be really a challenging one in future.

5. Result and Discussion

This paper gives out the entire map of the new unpredictable algorithm and thus how it works robust and effective[8,9]. The security and protection systems are still dynamically updated. As an advantage, it is very for the human readers to read than comparative to computer programs. This method is used in various websites and web services to feed input and to find out the easy way to find the user whether a human reader or a bot. Therefore this method is very easy to learn and follow up for the human readers. It can be easily learnt by any of the people even children can adapt it without any training. Thus if it is a failure for the OCR programs to find then, it is a success and thus a very effective method for the user.

5. Conclusion

The proposed new CAPTCHA method allows easy way of usage for the human than the bots and so it is very effective and more robust. LCG algorithm paves way for user friendliness and thus it assures high performance in security. There are three major advantages in this generation process. It is very much flexible and easy for

the human readers to solve and find. It is tougher for the bots than human readers and thus easy to generate and evaluate it. It paves a greater way for the future to resolve many issues created by computer programs.

6. References

1. Banday MT, Shah NA. A Study of CAPTCHAs for Securing Web Services. IJSDIA. 2009; 1(2):66- 74.
2. Bursztein E, Martin M, John C, Mitchell M. Text-based CAPTCHA Strengths and Weaknesses. ACM Computer and Communication Security. 2011. p. 1-14.
3. Datta R, Li JJZ, Wang W. Imagination by A Robust Image-based CAPTCHA Generation System. Proceedings of the 13th Annual ACM International Conference on Multimedia; 2005. p. 331-4.
4. Ince IF, Yengin I, Salman YB, Cho HG, Yang TC. Designing CAPTCHA algorithm: Splitting and rotating the images against OCRs. 3rd International Conference on Convergence and Hybrid Information Technology; Busan. 2008. p. 596-601.
5. Naveen K, Saravanan S, Lavanya M, Vaithayanathan V. Improved CAPTCHA based Authentication for E-mail ID. Indian Journal of Science and Technology. 2015; 8(35):1-4.
6. Kumar SP, Ramachandaran R, Saravanan A. Generation of Variant Random Order (VRO) in text graphics colour CAPTCHA for enhancing web security protection. Indian Journal of Science and Technology. 2016; 9(10):1-9.
7. Charles N, Zeeb Z, Patrick J. Burns by random number generator recommendation. 1997; 1-26.
8. Yan J, Ahmad SAE. Usability of CAPTCHAs or usability issues in CAPTCHA design. Proceedings of the 4th Symposium on Usable Privacy and Security; 2008. p. 44-52.
9. Chow R, Golle P, Jakobsson M, Wang L, Wang X. Making CAPTCHAs Clickable.
10. Soni R, Tiwari D. Improved CAPTCHA Method. International Journal of Computer Applications. 2010; 1(25):1234-5.