Performance Evaluation of CAPTCHA WORD RANKING Algorithm to Break Video CAPTCHA

[1]Maddela Shireesha
[1]ME (Computer Engineering)

[2]V.B.Gaikwad
[2] Associate Professor, Department of Computer Engineering

[1,2] Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India.

ABSTRACT

CAPTCHA – Completely Automated Public Turing test to tell Computer and Humans Apart, CAPTCHAs, on web employed as automated filters to distinguish humans and automated bots to prevent fraud and vandalistic activities such as spamming, automated bot-based account creation and corrupting websites by reducing the availability of site resources to legitimate users. At the moment, Captcha such as Text or image based, Math, and Audio have been broken accurately by bots and proven as less robust. Up recent, to address automated bot fraud activities and attacks a more robust CAPTCHA introduced called Video CAPTCHA. The Video Captcha is a captcha with moving letters and challenge will be given as to enter only few letters with specified color (for example enter only red color letters from moving letters) from moving letters. That means, breaking a captcha with motion along with specified color (red color) becomes difficult and tough job for bots. Moreover, this paper addresses a fact that like other traditional Captchas abovementioned, intelligent bot programs can able to break Video CAPTCHA too. Here, the paper described a novel schema to break Video CAPTCHA and procedure to sequentially break Video CAPTCHA. Yet, this research paper also covers how the implemented captcha breaking algorithm named CAPTCHA WORD RANKING algorithm breaks the Video CAPTCHA easily and effectively. Consequently, this experiment is conducted in an ethical manner and the implemented CAPTCHA WORD RANKING Algorithm only works with own created Video CAPTCHA samples where each sample ends with red color moving letters and captcha is moving from right to left in horizontal direction. Further, discussed the results of Static and Video Captcha samples with their breaking time, followed by performance evaluation of CAPTCHA WORD RANKING algorithm in terms of coverage, precision and accuracy. Subsequently, described the status (success/failure) of captcha samples in graphical representation. Finally, specified the experimental results, role of captcha application and backend database followed by implementation screen shots of some broken Video CAPTCHAs.

General Terms

CAPTCHA WORD RANKING Algorithm

Keywords

Automated, Bot, CAPTCHA, Spamming, Ranking

1. INTRODUCTION

These days, a known fact that, Internet has become essential part of human life to perform various activities such as commercial business, academics and education, entertainment and online shopping, administrative purpose and even in communication. Most of the websites today provides registration forms to be filled as a procedure to become legitimate member and to utilize the services and resource of the site. [1] [2] [3] [4].

At the same time, spamming, bot- based account creation and DOS attacks (Denial of Service attacks) also became very common to waste the websites resources. To restrict aforesaid hoax activities on Web and to avoid impersonation of automated bots to act as humans, a challenge and response test introduced called CAPTCHA. CAPTCHA-Completely Automated Public Turing Tests to Tell Computers and Humans Apart employed as an automated filter to perform challenge and response test on web forms to distinguish humans and bots or malicious script programs that perform the work of human [1] [3].

Due to rapid improvements in Information technology, bot programs have become intelligent and able to learn and break CAPTCHA challenge accurately and able to break in fewer attempts. In recent research, it is identified that using DeCaptcha tools most of existing CAPTCHA such as MATH, Text-based or Visual based CAPTCHA can be bypassed by bots. CAPTCHA used at well-known websites such as eBay, Digg and Wikipedia are compromised by DeCaptcha tool [5].

Recently, to improve the security and to make captcha breaking process difficult for bots, a new type of CAPTCHA emerged called Video CAPTCHA as most of traditional captchas aforementioned are broken accurately by bots. A Video CAPTCHA is a captcha with motion means here captcha letters are moving. NuCAPTCHA is a well known example of type Video CAPTCHA [6].

Moreover, research in this area already proven that conventional captcha in use from decades said above are weak, and this fact leads to doubt the behavior of Video CAPTCHA in terms of robustness and security. Here, the primary objective of this research is to prove that even Video CAPTCHA is not secured and robust as it claims to be by using a novel schema to break Video CAPTCH. For the explanation purpose, a procedure is given and applied on own Video CAPTCHA to break it sequentially. And, the same procedure is repeated with the rest of the Video CAPTCHA samples.

Next, most importantly, as part of proposed system, the paper describes the implemented algorithm named CAPTCHA WORD RANKING algorithm. Here, this paper, addresses how the implemented algorithm works to break own created Video CAPTCHA samples. Then, provided the status and breaking time of some Static and Video Captcha samples. Afterwards, discussed, about the performance of CAPTCHA WORD RANKING algorithm in terms of coverage, precision and accuracy. Further, explained the experimental results, function of captcha application and the significance of backend database followed by implementation screenshots.

2. BACKGROUND

In 1997, Chief Scientist Andrei Broder and his colleagues at AltaVista developed a primitive filter technique called CAPTCHA to abuse attacks from bots by adding URLS to their search engine. This technique is to generate an image of printed text randomly so that malicious automated programs or systems cannot read it but possible for humans.

Later, in 2000, von Ahn and Blum developed and publicized the basic idea of CAPTCHA that can include any kind of program that distinguishes humans from computers [1] [2] [3] [4] [7].
Greg Mori, Jitendra Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA," in their work specified a shape matching algorithm to break EZ-Gimpy CAPTCHA [8].
Elie Bursztein, "How we broke the NuCaptcha video scheme and what we propose to fix it" in his work explained how to break NuVideo CAPTCHA [9].
Y. Xu, G. Reynaga, S. Chiasson, J-M. Frahm, F. Monrose and P. van Oorschot, "Security and Usability Challenges of Moving-Object CAPTCHAs: Decoding Codewords in Motion," in their work given a high-level overview to break moving object CAPTCHAs such as NuVideo CAPTCHA [10].

3. NOVEL SCHEMA TO BREAK VIDEO CAPTCHA

To prove Video CAPTCHA is not secured and robust described a novel schema to break Video CAPTCHA (see Figure 1) [3]. To put the plan into action, developed own web application having Registration Form fields where users can register by submitting their personal details. The Registration Form contains fields which accept user's personal information and a Video CAPTCHA to distinguish users from bots. From Web Server, a Registration page or form is requested by the Bot program and as a response it receives all Registration form fields and even Video CAPTCHA.
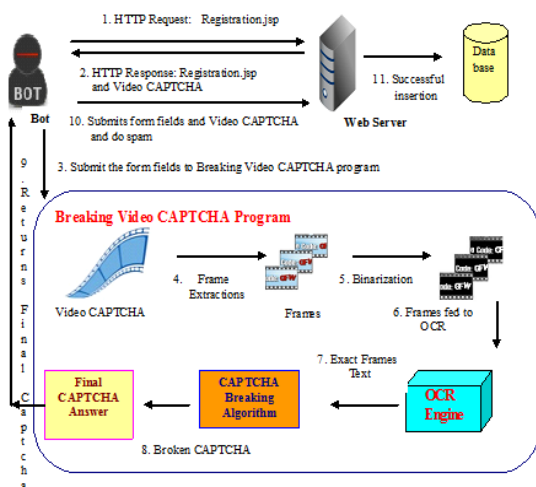

Fig 1: Novel Schema to Break Video CAPTCHA

Here, the implemented system contains 11 steps to show even Video CAPTCHA can be broken and is not so secured and robust.

1. Bot program requests Registration form from Web Server.
2. In response, Bot receives Registration form and Video CAPTCHA.
3. Bot submits Registration form fields to Breaking Video CAPTCHA Program.
4. In Breaking Video CAPTCHA Program, all the frames are extracted from Video CAPTCHA.
5. And extracted frames are binarized.
6. Then, binarized frames are fed to OCR Engine.
7. And with the help OCR Engine, extract Frames Text.
8. Extracted Frames Text given as input to the CAPTCHA WORD RANKING algorithm. This algorithm works on extracted frames text and gives the broken answer Captcha. The broken answer Captcha is the Final Captcha Answer.
9. The Final Captcha Answer is returned to Bot.
10. Bot submits all form fields including the Final Captcha Answer and do spam.
11. The successful record insertion in database shows that Bot has broken Video CAPTCHA successfully, and even done spam.

4. PROCEDURE TO SEQUENTIALLY BREAK VIDEO CAPTCHA

The following Figure 2 shows the procedure to break a sample Video CAPTCHA, **BUZZ 87M** [3].The same procedural steps are repeated with remaining Video CAPTCHA samples.
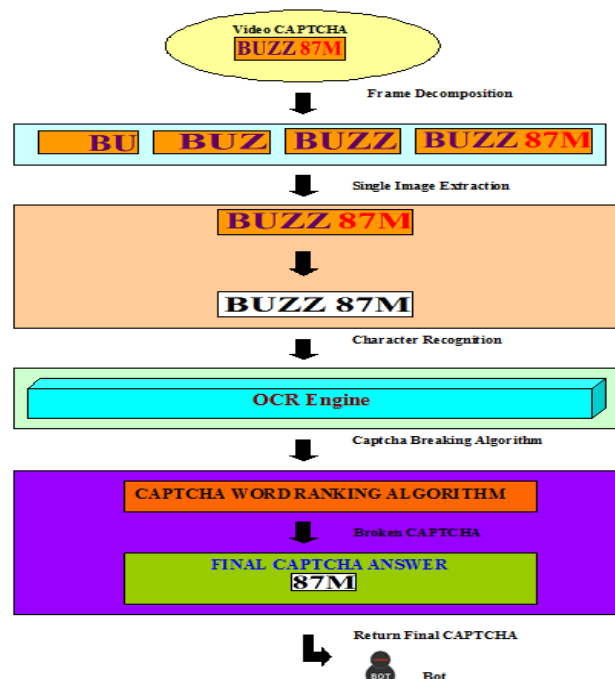
Fig 2: Procedure to Sequentially Break Video CAPTCHA

In this example, Bot requests Registration form and receives form fields and Video CAPTCHA, **BUZZ 87M**. Here, letters in Video CAPTCHA **BUZZ 87M**, are moving in relation to each other and from these captcha letters, the challenge is to enter only RED color letters to pass the test. And the bot must enter only RED color letters **87M**, to pass the challenge test as humans do.

Procedural steps to break Video CAPTCHA sequetially are given below:

1. Bot submits the received Video CAPTCHA, **BUZZ 87M** to Breaking Video CAPTCHA Program.
2. The above program decomposes Video CAPTCHA, **BUZZ 87M** into frames. Here, the Video CAPTCHA, **BUZZ 87M** decomposed into 241 frames. These decomposed 241 frames send to Pre-processing.
3. In Pre-processing, from these 241 frames, each frame is binarized and noised is removed.
4. Then these binarized 241 frames fed to OCR Engine for text recognition.
5. Once text recognition done by OCR Engine, these 241 frames text fed to Captcha Breaking Algorithm named CAPTCHA WORD RANKING Algorithm.
6. CAPTCHA WORD RANKING Algorithm computes all the 278 frames text and provides Final Captcha Answer **87M**.
7. Final Captcha Answer **87M**, retuned to Bot.

The above mentioned steps are applied to the remaining Video CAPTCHA samples and recorded their status in terms of success and failure.

5. CAPTCHA WORD RANKING ALGORITHM

In this section, briefly describe an implemented Captcha Breaking Algorithm named CAPTCHA WORD RANKING ALGORITHM shown in Figure 3.
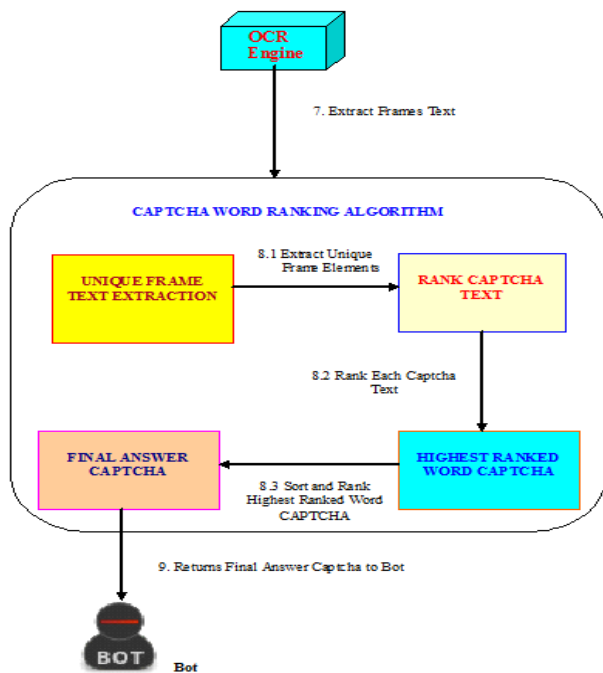

Fig 3: CAPTCHA WORD RANKING Algorithm steps to Break Video CAPTCHA

Internal working of CAPTCHA WORD RANKING Algorithm as follows:

1. Once all frames text of Video CAPTCHA recognized by OCR Engine, frames text is fed to CAPTCHA WORD RANKING ALGORITHM.

2. Now, CAPTCHA WORD RANKING ALGORITHM starts its work by extracting unique frame text elements from all frames.
3. After extracting unique frame text elements, it proceeds to rank each captcha text.
4. Then from each ranked captcha text, it sorts and ranks the highest ranked word captcha.
5. Finally, the highest ranked word captcha returned as final answer captcha to Bot.

6. PERFORMANCE EVALUATION

Performance Evaluation is based on evaluating effectiveness by measuring its accuracy [5]. Accuracy is the fraction of captchas that were answered correctly by the automated bot. The more effective way to evaluate the attack effectiveness is by metrics called Coverage and Precision. Coverage metric is the fraction of captchas that the automated bot attempts to answer. And the Precision metric is the fraction of captchas that were answered correctly.

Formulas:

$$N = \text{Number of Captcha Samples}$$
$$L = \text{Length of Captcha}$$
$$N_a = \text{Number of attempts to answer}$$
$$N_c = \text{Number of attempts to answered correctly}$$

$$\text{Coverage} = \left(N_a \times \frac{L}{N}\right) \times 100$$

$$\text{Precision} = \left(N_c \times \frac{L}{N}\right) \times 100$$

$$\text{Accuracy} = \text{Coverage} \times \text{Precision} \times L$$

In this experiment, 50 samples are tested and defined N=50 samples, Length of Captcha L=3, 5, 6, 7, 8, 9, 12, 13, 15 varies with respect to Captcha. Then number of attempts to answer kept always as 1 i.e., Na =1.

6.1 Static Captcha Samples

Static Captcha means CAPTCHA with no motion and covers single frame. The following Figure 4 shows the graphical representation about the status (success/failure) and breaking time of static captcha samples. From total 50 samples, 5 CAPTCHA samples are Static. And from these 5 Static CAPTCHA samples all are successfully broken associated with their breaking time. And the breaking time of these static captcha samples fall between 175 ms to 557 ms.
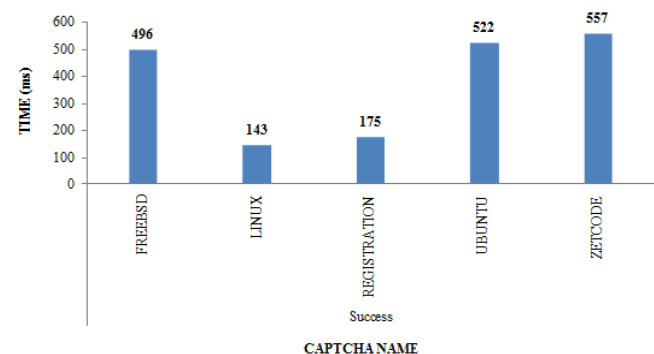

Fig 4: Status-Time of Static CAPTCHA Samples

6.2 Video Captcha Samples

Aforementioned, Video CAPTCHA is a CAPTCHA with motion and from this only last 3 red color letters need to be entered by bot to pass the test.

The following Figure 5 shows the graphical representation about the status(success/failure) and captcha breaking time in milliseconds (ms) of Video captcha samples with length L=6. Here, all the 7 Video CAPTCHA character length is same (i.e. Length L=6) but number of frames in each varies. Observed that, Video CAPTCHA with less frames broken with minimum breaking time as compared to other Video CAPTCHA samples with more frames. For example in Figure 5, BUG_25T Video CAPTCHA with 132 frames broken with minimum time 16037 ms and Video CAPTCHA with 271 frames broken with time 34266 ms. Means, as the number of frames increases, the breaking time of Video CAPTCHA increases.
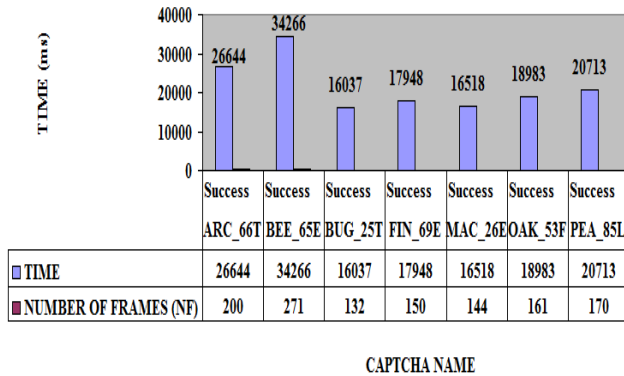
Fig 5: Status-Time of CAPTCHA Length (L):6

6.3   Status of Captcha Samples

Here, this section provides the Status (Success/Failure) of the CAPTCHA samples tested as part of the experiment.
The Figure 6 shows the graphical representation of Status of total 50 samples in terms of Success/Failure. Here, out of 50 samples, numbers of captchas answered correctly are 44 means Nc =44 and 6 are failure.
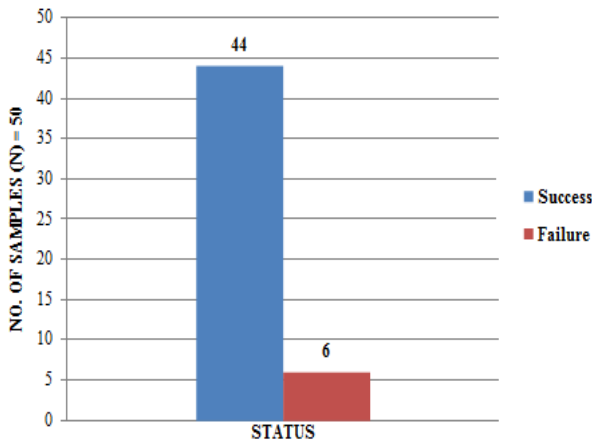

Fig 6: Status of CAPTCHA Samples

6.4   Performance of CAPTCHA WORD RANKING Algorithm

The performance of CAPTCHA WORD RANKING algorithm depends on number of captchas answered and answered correctly to measure the Coverage, Precision and Accuracy.
Figure 7, shows the Performance Evaluation in terms of Coverage, Precision and Accuracy of Sample size N=50. Using values of N, L, Na, and Nc, when evaluated the performance, CAPTCHA WORD RANKING algorithm shown Coverage=19.2%, Precision=55%, Accuracy=92.3%
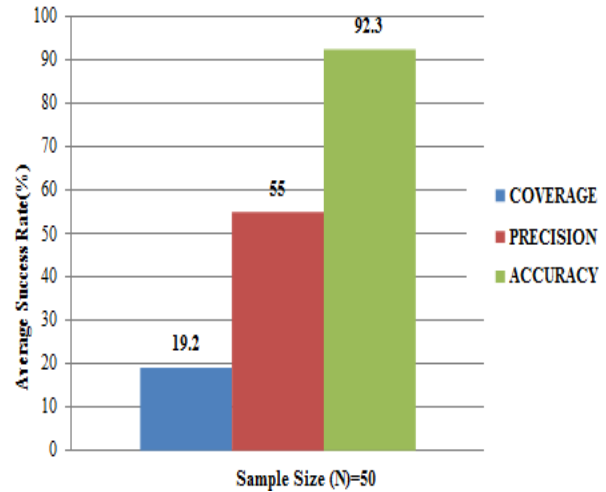

Fig 7: Performance Evaluation of CAPTCHA WORD RANKING Algorithm

7.   EXPERIMENTAL RESULTS

7.1   Captcha Samples

As mentioned earlier, the aim is prove that Video CAPTCHA (Captcha letters are moving in relation to each other) is not secured and robust for web applications, therefore,   in this research project,
created and used own Video CAPTCHA samples each ends with 3 RED color letters.
Here, in the following Figure 8, shows 28 Video CAPTCHA samples out of 50 samples used in the experiment.

Fig 8: CAPTCHA Samples

7.2   Captcha Application

To conduct this experiment in an ethical manner, created own captcha web application with registration page with form fields First Name, Last Name, Email, City, Zip, State, Country and Video CAPTCHA named **BUZZ 87M** shown in Figure 9.

Fig 9: Registration Page With Video CAPTCHA

Fig 10: StudentRegistration Table

And this, first the application is tested whether users can pass the challenge and response test or not before testing with bot. Each time, when user submits all the form fields and Video CAPTCHA correctly, an alert box shows the **correct captcha entered and success insertion**. And record is inserted into the StudentRegistraiton table. If user enters incorrect captcha, user is provided with pop up box with message **incorrect captcha entered, going back** and no recored is inserted into the database. In case of bot, a spam record is inserted into the Student Registration table when the bot breaks the Video CAPTCHA successfully and submits all the form fields including the broken Video CAPTCHA. When bot fails to break the Video CAPTCHA, no record is inserted to table.

7.3   Backend Database

However regarding database, here in this experiment, using Microsoft Access, Student Registration table is created in CaptchaDB database. And whenever user or bot successfully pass the challenge and response test, for each of them a record is entered with the details entered by them shown in Figure 10. For example, Once Video CAPTCHA is broken successfully by bot then, bot submits form fields that include following details:

FirstName = "Rock1373368907496" (unique timestamp)
LastName = "Test"
Email = Shireesha@abc.com
City= "Mumbai"
Zip= 400708
State = "MH"
Country = "India"
And the above record can be checked against the database in StudentRegistration table, if record found means spam is done successfully.
The last record inserted into the **Student Registration** table shows the personal information entered by user to distinguish user from bot. The record inserted by user after passing the challenge and response test is given below.
FirstName = "shireesha"
LastName = "y"
Email = siri@yahoo.com
City= "Airoli"
Zip= 400708
State = "MH"
Country = "India"

7.4   Implementation Screen Shots

Here, this section provides implementation screenshots of two broken Video CAPTCHA samples in Figure 11 and Figure 12.
The following Figure 11 shows the screenshot of Video CAPTCHA sample named **BUZZ 87M** of 241 frames successfully broken using CAPTCHA WORD RANKING algorithm and final captcha answer is **87M** with breaking time is 32563 ms.
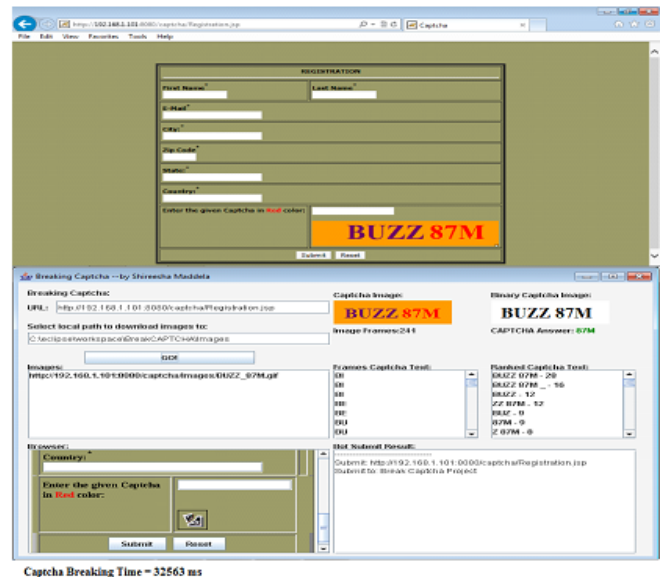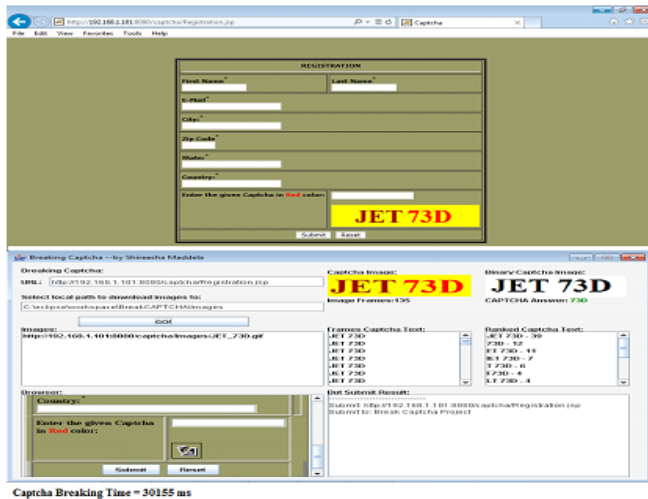

Fig 11: Broken Video CAPTCHA - BUZZ 87M

And Figure 12 shows another Video CAPTCHA named **JET 73D** of 135 frames successfully broken using CAPTCHA WORD RANKING algorithm and final captcha answer is **73D** with breaking time 30155 ms.

Fig 12: Broken Video CAPTCHA - JET 73D

8. CONCLUSION

This research project objective is to prove that Video CAPTCHA (Captcha as moving letter) is not that robust and secure as it appears to users.

Moreover, the limitation of this research project is that, the implemented CAPTCHA WORD RANKING algorithm and Bot program are applicable to own created Video Captcha samples.

Here, noticed that Static Captchas are broken successfully by the CAPTCHA WORD RANKING algorithm and the breaking time of this static captcha samples fall between 175 ms to 557 ms (i.e. less than 1000 ms).

Moreover, when algorithm applied to 7 Video CAPTCHA samples of 6 character length (L=6), each with different number of frames, all of them are broken successfully and status noted as success. And observed that, as the number of frames in Video CAPTCHA increases the breaking time increases.

Also noticed that, the breaking time of Video CAPTCHA is more than the breaking time of Static CAPTCHA because Static CAPTCHA represented in single frame where as Video CAPTCHA with more frames.

About status, out of 50 samples 44 are broken successfully and 6 are failure. And, performance of the CAPTCHA WORD RANKING algorithm depends on the number of captchas correctly answered by bot. While coming to performance evaluation, CAPTCHA WORD RANKING algorithm has shown coverage 19.2 %, precision 55% and accuracy 92.3 %.

Using status (Success/Failure) and performance information of the algorithm, it is clear that Video CAPTCHA can be broken accurately and is not that robust and secure as it seems to be.

9. REFERENCES

[1] A.A. Chandavale, Dr.A.M.Sapkal, Dr.R. M.Jalnekar, "Algorithm To Break Visual CAPTCHA," Second International Conference on Emerging Trends in Engineering and Technology (ICETET), pp. 258–262, 2009.

[2] Maddela Shireesha, "A Study on various possible ways to Attack CAPTCHA," In Proceedings of the Third National Conference on Nascent Trends in Information Technologies and Communication Technologies, equinox, pp.100-102, 2011.

[3] Maddela Shireesha, Prof. V.B. Gaikwad, "A Novel Schema to Break Video CAPTCHA," International Journal of Scientific Research in Computer Science Applications and Management Studies (IJSRCSAMS) ISSN 2319 – 1953, Volume 2, Issue 3, May 2013.

[4] Dr.K.Kuppusamy, A.Krishnashanthi, "SECURING WEB SERVICE USING 3D CAPTCHA," National Conference on Future Computing, Volume 1, pp. 56-60, 2012.

[5] Elie Bursztein, Matthieu Martin, John C. Mitchell, "Text based CAPTCHA Strengths and Weaknesses," ACM Computer and Communication security (CCS), pp. 1-14, 2011.

[6] "White Paper: NUCAPTCHA & TRADITIONAL CAPTCHA," Nu Captcha, pp. 1-6, 2011, [Online]. Available: http://www.nucaptcha.com/. [Accessed: Sept. 08, 2012].

[7] Anti-Spam Research Laboratory, "How CAPTCHA annoys genuine users," Digital Ecosystem and Business Intelligence Institute, Curtin University of Technology, pp. 1-7, [Online]. Available: http://asrl.debii.curtin.edu.au/. [Accessed: Sept. 07, 2012].

[8] Greg Mori, Jitendra Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA," Computer Science Division, University of California, Berkeley, CA 94720, pp. 1-8

[9] Elie Bursztein, "How we broke the NuCaptcha video scheme and what we propose to fix it" [Online]. Available: http://elie.im/blog/security/how-we-broke-the-nucaptcha-video-scheme-and-what-we-propose-to-fix-it/. [Accessed: Oct. 21, 2012].

[10] Y. Xu, G. Reynaga, S. Chiasson, J-M. Frahm, F. Monrose and P. van Oorschot, "Security and Usability Challenges of Moving-Object CAPTCHAs: Decoding Codewords in Motion," pp. 1-16, [Online]. Available: https://www.usenix.org/system/files/conference/.../sec12 final118.pdf, [Accessed: Oct. 13, 2012].