WEB SECURITY TECHNIQUES, ANALYSIS USING CAPTCHA AND ENHANCEMENT FOR FUTURE

Ravindra Meena[1] Devendra Bhavsar[2]
[1]Electrical Engineering Department , Career Point University Kota, India
[2]Computer Science Engineering Deptt , JK Lakshmipat University Jaipur, India
Email: [1]iitb.ravindra@gmail.com, [2]devendrebhavsar@jklu.edu.in

Abstract—There are many services in the internet including Email, search engine, social networking are provided with free of cost. To access these web services users have to register regarding the websites. During registration, some intruders or attackers write malicious programs that waste the website resources by making automatic false enrolments which are called as bots. Security researchers developed many techniques to prevent from accessing web resources by these bots. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a human authentication technique that generates tests to differentiate between human user and a malicious computer program. CAPTCHA techniques tries to maximize the difficulty for automated programs to pass tests .We can see many CAPTCHAs in web site with their own approach to provide security. In this article we take an overview of different and recently proposed CAPTCHA techniques. It proposes an approach to generate a robust CAPTCHA. The CAPTCHA is generated in two steps fist by randomizing alphanumeric characters and finally different queries are attached with it. The generated CAPTCHA is simple and can improve the usability and security with reduced solving time of human.

Keywords— Text based CAPTCHA, Web based security

I. INTRODUCTION

Internet has definitely made many aspects of modern life much more convenient by allowing users to access several online services. Security is an important issue for any website since there are many ways to intrude those services by programmed bots. CAPTCHA are well adopted solution for securing online services from programmed bots. This article proposes a new robust and two tier CAPTCHA to secure websites. First it studies history, importance of Internet and security along with different techniques to secure web services and applications. It also gives an overview of previous work and recently proposed CAPTCHA techniques. On the basis of their analysis we developed an algorithm and designed a CAPTCHA which can be generated easily with some randomness. It can also improve usability and ease for human user and maintain the robustness for programmed bots.

II. WAYS TO IMPEMENT WEB BASED SECURITY

A. Web security defence strategy

There are two techniques for finding excellent security. On one we would assign all of the resources required to maintain constant attentive to new security issues. we might ensure that all patches and updates are carried out at one time, have got all of the existing applications reviewed for correct security, be sure that only security knowledgeable programmers will deliver with your site and have absolutely their work checked carefully by security professionals. You would

also maintain a good firewall, antivirus protection and run IPS/IDS.

Our other option is to use an online scanning means to fix try out your existing equipment, applications and website code to discover in case a KNOWN vulnerability actually exists. While firewalls, antivirus and IPS/IDS are all worthwhile, it can be simple logic to also lock leading door. It's much more effective to correct half-dozen actual risks compared to leave them constantly in place and try to build higher and better walls around them. Network and site vulnerability scanning is regarded as the efficient security investment coming from all.

B. Web security using website security audit

Healthy defense against a attack on our website is always to regularly scan a competently setup domain that is certainly running current applications and whose website code was done well. Internet site testing, often known as web scanning or auditing, is really a hosted service furnished by Beyond Security called WSSA - Web page Security Audit. This particular repair requires no installing software or hardware and it is done without any interruption of web services. WSSA is usually run using regularly which means that your site are going to be tested against new vulnerabilities as they become known and offer you with solid data whether or not action is vital, needed or low priority. We will even be alerted if new code have been put into your website which is insecure, a different port continues to be opened that's unexpected, or May be a new service may be loaded and started that may present a chance to burglary. In complex, large systems it may be that daily web scanning would be the Best way to make sure that none of the many changes designed to site code or by using an application could possibly have opened an opening with your carefully established security perimeter. [1]

C. Network based security

Network-Based Security offers multiple, complementary services to help you protect our enterprise internally. These layers work in conjunction with endpoint computer antimalware software and internal company firewalls to intercept diverse threats.

D. CAPTCHA based web and appliction security

A CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a kind of challenge response test used to determine whether the user is human or not. Computers cannot decode the distorted words in a CAPTCHA easily, while humans can easily decipher the text. In the most common type of CAPTCHA user is provided with letters of a distorted image. Then the user is asked to solve the CAPTCHA by entering the correct characters. By definition CAPTCHAs are fully automated, it requires human maintenance very less. A CAPTCHA may have two characteristics such as usability and security. Security means the strength for preventing the variant attacks, while usability means the user friendliness of the CAPTCHA. The CAPTCHA is often a visual or audio challenge towards user in order to avoid bots and automated scripts from accessing the services protected because of it. It really is valuable for forums looking to prevent spambots/adbots from joining and protecting downloads from automated access by bots (that is not really a security risk in itself, but a bandwidth drain). A CAPTCHA will not provide some other form of security, it only provides defence against bots and the rate limiting that includes it. A CAPTCHA is really a program that protects websites against bots by generating and grading tests that humans can pass but current computer programs cannot. One example is, humans can read distorted text as the one shown below, but current computer programs can't.

E. Types of CAPTCHA

Text based CAPTCHA

These are generally an easy task to implement. The most convenient yet novel approach would be to present anyone with a few questions which a human user can solve. Gimpy: Created by Yahoo and CMU Accumulates 10 random words from dictionary and distorts, fills with noise User has to recognize at least 3 words If user is correct, he's admitted EZGimpy: A modified version of Gimpy Yahoo used this version in Messenger Has only 1 random string of characters Not just a dictionary word, so not at risk of dictionary attack A bad implementation, already broken by OCRs

Fig1 Gimpy & EZ-Gimpy Text CAPTCHA

Advantages of Text-based images

1) Text-based CAPTCHA is straightforward to implement so that it's mostly found in websites.
2) Battle Text-based CAPTCHA can be used to defeat dictionary attacks.
3) Re- CAPTCHA Text-based CAPTCHA uses new dictionary words that can't read using optical character recognition

Disadvantages of Text-based images

1) Therein type of CAPTCHAs, users has faced some problems to get in the best text or characters or letter. Following include the few reasons that confuse a person to identify the precise text. i. By using various lines.
ii. Utilization of various shapes. iii. Using multiple fonts. iv. Font size variation.
v. By using Blurred Letters.
2) Text-based CAPTCHAs can be simply broken by OCR techniques (example: Content based image retrieval).
3) The peoples which have low visibility power cannot easily pass the test.

Graphic based CAPTCHA

Graphic CAPTCHAs are challenges that entail pictures or objects who have getting some sort of similarity how the users ought to guess. They are visual puzzles, a lot like Mensa tests. Computer generates the puzzles and grades the answers, but is itself cannot solve it.

Fig 2 Graphic CAPTCHA

PIX is really a program which has a large database of labeled images. Most of these images are pictures of concrete objects (a horse, a table, a property, a flower). This program picks an object haphazardly, finds six images of these object by reviewing the database, presents these to the consumer and then asks the question "exactly what these pictures of?" Hence, writing a course that may answer the question "exactly what these pictures of?" is simple: search the database to the images presented and locate their label. Fortunately, this can be fixed. One way for PIX to become a CAPTCHA is usually to randomly distort the photographs before presenting these phones anyone, so that computer programs cannot easily search the database for that undistorted image.

Advantages of Image-based CAPTCHA

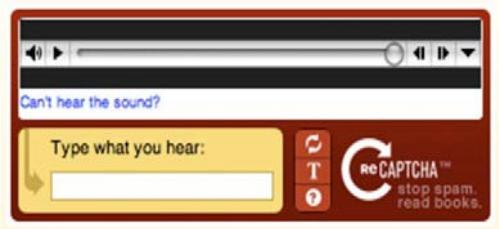
1) Within the text-based CAPTCHA zinc increases the safety.
2) Simple click based system so no necessity of typing.
3) Using Image-based CAPTCHA pattern recognition of image is tough AI program.

Disadvantages of Image-based CAPTCHA

1) Therein kind of CAPTCHAs, users have faced some problems to go in the precise text or characters or letter. Following are classified as the some reasons that confuse the users to recognize the correct text. i. Using various lines. ii. By using various shapes. iii. Usage of Multiple fonts. iv. Font size variation.
v. Usage of Blurred Letters
2) Text-based CAPTCHAs can be easily broken by OCR techniques (example: Content based image retrieval). 3) The peoples that contain low visibility power cannot easily pass quality.

Audio based CAPTCHA

The program picks a thing or even a sequence of numbers indiscriminately, renders the phrase or numbers in a sound clip and distorts the sound clip; it then is definitely the distorted sound clip towards the user and asks users to go into its contents. This CAPTCHA is founded on the real difference in ability between humans and computers in recognizing spoken communication. It is a crude strategy to filter humans in fact it is not so popular considering that the user must understand the language and also the accent the location where the sound clip is recorded.



Advantages of Audio-based CAPTCHA 1) It is employed for most people that have vision defect.

2) Friendly to peoples.

Disadvantages of Audio-based CAPTCHA

1) System available in the English so person needs to have a thorough English Vocabulary.

2) Similar sound characters.

3) Not working for dumb people or some people that have low listening power.

Video based CAPTCHA

Video CAPTCHA is a newer and fewer commonly seen CAPTCHA system. In video-based CAPTCHAs, three words (tags) are offered towards user which describes a movie. Anyone's tag must match into a set of automatically generated ground truth tags then exactly the test has been said to become passed. The phrase video CAPTCHA is used to any CAPTCHA which utilizes a relevant video since it's method for present information to your user Although video CAPTCHA is bound, both commercial and academic application do exist.

Advantages of Video-based CAPTCHA

1) It cannot break using Optical Character Recognition (OCR).

2) It cannot effect by laundry attacks.

3) In some cases it provides greater security than Text-based CAPTCHA and Image based CAPTCHA.

Disadvantages of Video-based CAPTCHA

1) Big is files are large, so problem face by users to download video and pass the CAPTCHA test.

2) Speed of video.

Puzzle based CAPTCHA

Usually in puzzle based CAPTCHA confirmed picture is divided to chunks. A user should combine these chunks so that you can make up the complete picture comparable to the first one. [2]

Advantages of Puzzle-based CAPTCHA

1) It looks like an exciting.

2) It helps the user to observe their brain.

3) It's just like a game so user can more communicate with this CAPTCHA system.

Disadvantages of Puzzle-based CAPTCHA 1) Time consuming.

2) User cannot identify the puzzle easily

III. LITRATURE REVIEW

This section studies and analyzes recently proposed CAPTCHAs and design principles. These would be useful to understand and develop robust CAPTCHA.

A. Question based CAPTCHA

Question based CAPTCHA was proposed in which on the basis of a series of pre-designed patterns of questions were prepared. In these patterns some of the elements of the problem were variable and changeable and they are chosen from some items randomly. For example question may be like "There are 5 cats, 3 apples, and 4 dogs on a table. How many pets are there on the table in total?" The answer is 7 (3 cats+4 dogs=7 pets). Here user only needs to enter a number. To increase the diversity and variety of the questions by designing different and various patterns and even make them more difficult and more sophisticated, too. In place of some words one can put an image of those words like in place of cat, apple, dogs and table one can put their images. Computer may recognize images but first it has to separate text from images which may be a difficult task. Probability of computers to successful answers of this type of question is very less because the computer requires the following abilities

1 Computer must recognize phrase shown in the image through OCR-based software.

2 Computer must recognize shapes shown in the image. Of course before that it must separate texts from the images which would be a difficult process in itself. 3 After recognition of texts and images the computer should be able to

understand the question.

4 At last and even if a computer does all the above-mentioned stages successfully it must be necessarily capable to answer the shown question.

Example of Question based CAPTCHA

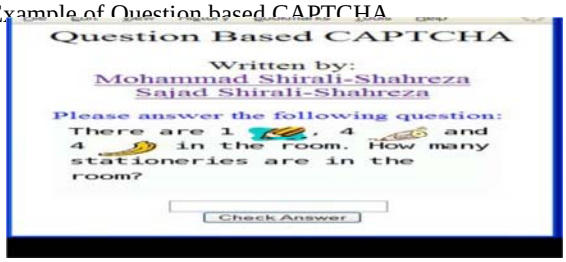


Fig 4 audio based CAPTCHA[3]

Advantages of Question based CAPTCHA

1. Unlike OCR-based CAPTCHA methods, this method requires only typing a number as the answer. So it is easy to use, saves users time and more comfortable for them.

2. In this method it is not necessary to have a keyboard and we have only to enter a number. Therefore we can use this method on devices which don't have a keyboard or on devices in which it is difficult to use a keyboard, such as mobile phones and Pocket PCs.

3. This method does not need any processing to be done by client and can be executed on small devices and on devices with limited resources. [3]

B. Word grouping CAPTCHA

The problem with current text based CAPTCHA (most popular CAPTCHA) schemes is that most of them have proven to be either not robust enough (easy to break them) or they are too complicated annoying to read even for humans.

Word grouping is a type of CAPTCHA in which user has to divide the given words in two subgroups. In word grouping CAPTCHA the user is presented with six words, and is asked to divide the group into two subsets, using any categorizing the user wishes. The words will be easier so that any user can do that.

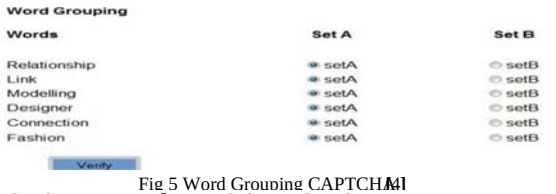


Fig 5 Word Grouping CAPTCHA[4]

C. A more robust CAPTCHA

A more robust CAPTCHA was proposed in [5] article withmultiple secure characteristics and very effective for breaking attack but easy to answer by user. It may avoid replay attack as after each page refresh principle characteristics ofCAPTCHA would be changed as:

CAPTCHA's code is a series of numbers and characters (uppercase and lowercase). To generate a random code (stream of characters and numbers) in each test in order to make it not susceptible to a dictionary attack, multiple randomizing functions are used.

The length of the code may be varied (minimum length is 6 characters-numbers). Multiple font types can be used to prevent intrusion using image processing techniques.

String/codes can be rotated at different angels. Lines can be used to prevent segmentation. The numbers and the length of lines and their positions can be changed each time in order to deform the text image randomly before presenting to the user. Using some specific technique the text image may be blurred in order to make CAPTCHA difficult for spiteful software. Image dimensions may be changed inconsistent with all other characteristics mentioned above. Each time CAPTCHA's code and line color were kept in gray scale colors at different levels.

D. Two-Tier CAPTCHA

In article [8] introduces a new CAPTCHA scheme called Two-Tier CAPTCHA. First a alphanumeric CAPTCHA code with image is generated. Second Query related to that

CAPTCHA code. E.g. enter only Digits .Rate of its toughness can be increase d in order to improve its resistance against the attacks by adding more and more query and combination in database. The algorithm of this scheme makes it hard for bot programs which mean that it is more secure.

The advantage of using Two-Tier CAPTCHA is it can solved by human users easily and difficult to solve by bots. This Two-Tier CAPTCHA methods use a same input method as used by many well-known web sites and services where users type some keywords or characters into an input box.

E. Hybrid Collage CAPTCHA

This paper [22] introduces a new type of

CAPTCHA called hybrid collage CAPTCHA. One of the CAPTCHA methods is Collage CAPTCHA. It is a procedure for distinction between human and computer programs through recognition and finding a picture of an object among some objects. This article improves the resistance of Collage CAPTCHA method by an improved method called Hybrid Collage CAPTCHA. This scheme displays images on left and right side of the screen.

On right side screen we have the corresponding images along with different texts in distorted form. Now the computer program asks user to choose the picture with correct texts. If the user select correctly, then user is allowed to enter the text of the image in the given text box. If entered name is correct, then we guess that user is human.

Fig 6 hybrid collage CAPTCHA [6]

F. Move & Select: 2-Layer CAPTCHA

This paper [23] proposes a technique differentiate human users and computer programs from each other by the fact that human user have exceptional cognitive processing abilities. A new CAPTCHA was proposed with some basic principles like Easy to solve for most people Difficult to solve for automated bots, Easy to create and estimate, User friendly

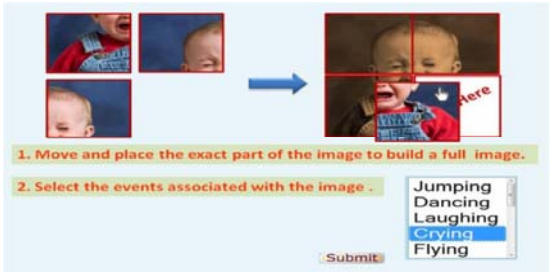


Fig7. 2-tier move & select CAPTCHA [7]

According to study and analysis in this article Move & Select CAPTCHA is more protected compare to other existing CAPTCHA techniques because of 2-layers. Here both layers may lead to hard AI problems for computer bots. Bots may try random guess attack based on probability.

B. Methods to design robust CAPTCHA

CAPTCHA designers may use methods to design a robust
CAPTCHA

1. Make it difficult to split the text from the background by using multiple colors for both foreground and background, leave no pattern that could help differentiate the foreground automatically, and contain some foreground colors into the background And vice versa.
2. Make it difficult to split each image by connecting characters with each other or add more cracks in each character.
3. Make it impossible to distinguish a character by counting its pixels by making all characters have the same pixel count all the time. Or make a character have very different pixel counts in different challenges (if the difference isn't large enough, an approximation method could almost certainly determine each character).

4.Random warping may provide good defense against the pixel count attack .for example, local warp can introduce "small ripples, waves, and elastic deformations along the pixels of the character" and global warp generates characterlevel, elastic deformations; both can make a character's pixel count less predictable[9].

IV. PROPOSED SCHEME

A. Existing text based CAPTCHA scheme

In this existing robust CAPTCHA scheme[8] algorithm random alphanumeric code of fixed length (size 6) is generated. This alphanumeric code is then converted into image with some noise. After this a random query is related to this code image in same session and user is asked to answer this query. Validation is done with user input and session value to allow user to proceed.

B. Proposed Scheme and Methodology

In this chapter, we propose a robust CAPTCHA technique which includes two stages: First is CAPTCHA generation in which a random series of characters was generated. Generated series of characters are combination of alphabets (upper or lower case) and numbers which may avoid the dictionary attack. In second stage a random query is associated with generated query. After that user is asked to pass the test, in which according to associated query user requires submitting few characters rather

than typing entire CAPTCHA. It saves time for user and associated query makes it hard to solve for program bots.

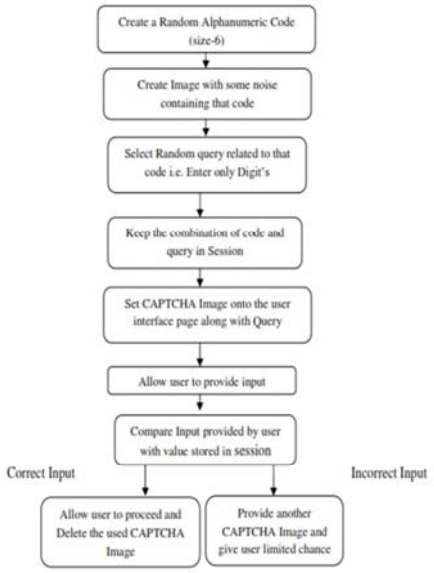


Fig 8 Algorithm of an existing robust CAPTCHA

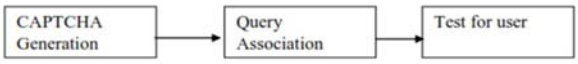


Fig 9 Block diagram of proposed scheme

C. Problem defination

In this work we analyzed previous works on CAPTCHA and found that so far many types of text based CAPTCHAS have been proposed and developed .All of them tried to make the test easier for human user and difficult enough for computer programs and bots. In their efforts to make a hard CAPTCHA many schemes used background confusion, blurring and tilting of texts which may lead to hard enough for human user to pass the test. More background confusion and tilting and twisting of test may cause recognition problem for human user also. Image based and group based CAPTCHA tests need large data base and may face usability problem. So the scheme should be simple and secured enough to avoid abuses from bots.

D. Objective for proposed scheme

1) To develop a robust text based CAPTCHA following the secure CAPTCHA design principles with random
character set made of alphabets (upper and lower case) and numbers.

2) To improve security and hardness length and font size of generated text is randomized every time.

3) Associate a random question to new generated CAPTCHA to improve usability and user response time.

E. Algorithm and flowchart of proposed scheme

1. Create random alphanumeric code

1.1. Vary the generated alphanumeric code by random length and font size.

2. Create Image with little background confusion

3. Randomly select a query related to that code.

4. Keeping the combination of generated code and query in session.

5. Set the CAPTCHA combination of generated code and query on user interface (login page).

6. Ask user to provide proper input.

7. Compare the input provided by user and session value.

7.1. If inputs provided by user and session value are same with correctness then user is allowed to access service page.

7.2. If inputs provided by user and session value are not same then user is asked to provide proper inputs again.

8. Stop

Fig 10. Flowchart of proposed scheme

F. Experiment table

User	Age	Response time (in sec.)	Input/solving time(in sec)	Experience of use
User-1	27	2	4.99	Good
User-2	24	3	6.10	Good
User-3	28	3	7.10	Good
User-4	25	4	6.80	Good
User-5	30	4	7.90	Good
User-6	23	3	5.50	Excellent
User-7	25	3	5.40	Good
User-8				

Table 1

Based on experiment it was found that a good experience of using the proposed CAPTCHA for every age group people which ensures better usability. It was also found that the average response time(3.14) and input/solving time(6.25) have been improved. It may be acceptable according to Bursztein's recommendations that the optimal recognition time for textbased CAPTCHAs is under 9 seconds [10].

G. Result

.

Fig 11 snapshot before user logged in

This snapshot shows the login page CAPTCHA code and assiciated query before login of the user.Based on query user should enter first, 4th and 5th character of the CAPTCHA code only.

.

Fig 12 snapshot when user logging in

This snapshot of result shows login page with CAPTCHA code and assciated queryAfter providing correct user name , password and CAPTCHA code according to query attached with it.

Snapshot of proposed scheme were shown above in which it can be obseved that here user needs to provide input according to query associated with the test. It saves time for user as it only provides few from entire code.Here type of query associated to CAPTCHA code is important and choosing query for query set should be done carefully. If query like" type how many numeric character? "," Type how many alphabets in the code?" are to be associated test may suffer with brute force or random guess attack. So we need to avoid query like this in which answer is single character or number.

We must consider queries whose answers will be easier for human user but difficult for bot programs.

V. CONCLUSION

In this article different recently and frequently used CAPTCHA were studied. According problem statement a robust CAPTCHA should not be only difficult to solve by computer programs, it should also be user friendly. Currently proposed scheme is designed in asp.net under visual studio platform with C# which is easy to implement. The solving process was kept simple so that human user can solve the test easily; they need to provide input according to associated query not entire CAPTCHA character set. Users require little time to provide input with accuracy. Principle design features were considered to provide better security and construct a generic text based CAPTCHA. To avoid several attacks like dictionary attack, segmentation attack and brute force attack (random guess attack) we have varied some considerable features and used different queries with randomness. It results improved usability, solving time and robustness of text based CAPTCHA. Using larger query set and choosing different kind of quires may improve the robustness of this type of CAPTCHA scheme in future.

REFERENCES

[1] Web Security, Your Site and Your Network beyond Security: Beyond security Inc.2015http://www.beyondsecurity.com/web-security-andweb-scanning.html -19/02/2015J.

[2] Ved Prakash Singh, Preet Pal" Survey of

Different Types of CAPTCHA" ,IJCSIT Vol. 5 (2) , 2014, 2242-2245.

[3] Mohammad Shirali-Shahreza, Sajad Shahreza,"Question Based CAPTCHA", IEEE International Conference on Computational Intelligence, 2007..

[4] Dayanand,"WORD GROUPING CAPTCHA-A NOVEL APPROACH FOR SECURING WEB SERVICES" International Journal of Electrical, Electronics and Data Communication Volume-1, Issue-, July-2013.

[5] Mumtaz M. Ali AL-Mukhtar and Rana Riad K. AL-Tai "A More Robust Text Based CAPTCHA For Security in Web Applications" IJETTCS vol.3 issue 2 march-april 2014.

[6] Divya Shanker, Prashant Gupta, Aditya Jaiswal "Hybrid Collage CAPTCHA" International Journal of Communication and Computer Technologies Volume 01 – No.31, Issue: 05 May.

[7] Moin Mahmud Tanvee, Mir Tafseer Nayeem, Md. Mahmudul Hasan Rafee "Move & Select: 2-Layer CAPTCHA Based on Cognitive Psychology for Securing Web Services" International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol: 11 No: 05

[8] Poonam Yadav and Sujata"Security Issues in Cloud Computing Solution of DDOS and Introducing Two-Tier CAPTCHA", IJCCSA Volume 3 no 3 june 2013.

[9] Jeff Yen, Ahmad Salah, "CAPTCHA security-Case Study", IEEE CS, 2009

[10] E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, andD. Jurafsky, "How good are humans at solving captchas? a large scale evaluation," in Proceedings of the 2010 IEEE Symposium on Security and Privacy, ser.

SP'10.