An Overview of the Application of Captcha In Mitigating Ddos Attack on Web Server

[1]AMADI E.C., [2]IBEH G. I., [3]EZIRIM C., [4]OPARA C. C.

[1-4]Department of Information Management Technology, Federal university of Technology, Owerri, Imo State, Nigeria

ec.amadi@gmail.com

ABSTRACT

In IT world today, Mitigation of DDOS attack on web servers and internet has been a major challenge for two decades or more. Introducing CAPTCHA application has proven a strong measure to secure web servers and internet activities. This paper reviews the application of CAPTCHA in mitigating DDOS attack on web servers. A detailed look at the various types of CAPTCHA, its working mechanism, and tuning test was also presented. The various drawbacks of CAPTHA was also presented with a clear road map for future areas of research to improve CAPTCHA technology.

Key words: DDOS, CAPTCHA, Attack, Tuning Test, Web Servers.

1.0 INTRODUCTION

Distributed Denial of Service (DDOS) attack is performed solely with the intention to deny the legitimate users to access services. DDOS attack is usually performed by means of bots, automated software. These bots send a large number of fake requests to the server which exceeds server buffer capacity which results in DDOS attack (Mehra, Agarwal, Pawar, & Shah, 2011).

Distributed Denial of Service attacks attempts to make a network resources and web servers become inaccessible, by interrupting services of host connected to network. Bots-affected computers are called as zombie and can be manipulated from the outside walls by hackers who seek to access web server resources or submit unwanted details into the servers.

In this paper presentation, we propose an idea to mitigate DDOS attack on web-sites which ask for user credentials before it allows them to access resources or submit any information. This approach is based on CAPTCHA application. CAPTCHAs are acronym for **C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part.

The term "CAPTCHA" was coined in 2000 by Luis Von Ahn, Manuel Blum, Nicholas J. Hopper (all of Carnegie Mellon University, and John Langford (then of IBM) (Jaiswal, 2010). They are challenge-response tests to validates that the users are indeed human not bots. The purpose of a CAPTCHA is to block form submissions from spam bots –automated scripts that harvest email addresses from publicly available web forms. A common kind of CAPTCHA used on most websites requires the users (human) to enter the string of characters that appear in a distorted form on the screen.

CAPTCHAs are used because of the fact that it is difficult for the computers to extract the text from such a distorted image (Civil, 2000), whereas it is relatively easy for a human to understand the text hidden behind the distortions. Therefore, the correct response to a CAPTCHA challenge is assumed to come from a human and the user is permitted into the website server (Guard, 2013).

Why would anyone need to create a test that can tell humans and computers apart? It's because of people trying to **game** the system -- they want to exploit weaknesses in the computers running the site (Holmes, n.d.). While these individuals probably make up a minority of all the people on the Internet, their actions can affect millions of users and Web sites. For example free e-mail service might find itself bombarded by account requests from an automated program. That automated program could be part of a larger attempt to send out spam mail to millions of people. The CAPTCHA test helps identify which users are real human beings and which ones are computer programs.

1.1 WHY USE CAPTCHAs IN SECURING WEB SERVERS

The proliferation of the publicly available services on the Web is a boon for the community at large. But unfortunately it has invited new and novel abuses. Programs (bots and spiders) are being created to steal services and to conduct fraudulent transactions(Jaiswal, 2010). Some examples:

a. Free online accounts are being registered automatically many times and are being used to distribute stolen or copyrighted material.

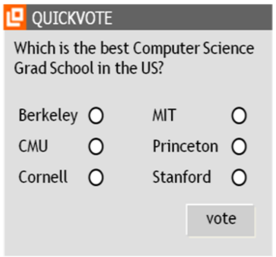
b. Recommendation systems are vulnerable to artificial inflation or deflation of rankings. For example, EBay, a famous auction website allows users to rate a product. Abusers can easily create bots that could increase or decrease the rating of a specific product, possibly changing people's perception towards the product.

c. Spammers register themselves with free email accounts such as those provided by Gmail or Hotmail and use their bots to send unsolicited mails to other users of that email service.

d. Online polls are attacked by bots and are susceptible to ballot stuffing. This gives unfair mileage to those that benefit from it.

In light of the above listed abuses and much more, a need was felt for a facility that checks users and allows access to services to only human users. It was in this direction that such a tool like CAPTCHA was created.

Well to completely understand its usage one can consider this story. Few years ago (November 99) ***www.Slashdot.org*** (a popular site in US) conducted following poll on internet.



QUICKVOTE

Which is the best Computer Science Grad School in the US?

Berkeley	○	MIT	○
CMU	○	Princeton	○
Cornell	○	Stanford	○

Figure1: Slashdot.org school pool

Now students at CMU and MIT instantly wrote a program which increased their vote counts using software and ultimately the poll had to be taken down because both MIT and CMU had millions of votes while others struggled to reach thousands.

There are situations like these where you need to distinguish whether user is a machine or a computer. This is where CAPTCHAs are need.

2.0 DEFINITIONS

CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart A.K.A. **Reverse Turing Test**

(Mahato, Saxena, & Mishra, 2015), Human Interaction Proof.

Turing Test: to conduct this test two people and a machine is needed here one person acts as an interrogator sitting in a separate room asking questions and receiving responses and goal of machine is to fool the interrogator.

HIP: human interaction proof is designed schemes that distinguish human and bots in a computer network.

Web server: A web server is a system that delivers web services to end users over the internet.

DDoS: Distributed Denial of Service attacks are attempts that are aimed at disrupting the normal function of a specific website from regular/legitimate users. There are three common DDoS attacks which include volumetric DDoS attacks, semantics DDoS attacks and Blended DDoS attacks(Bhandwalkar, Bhoi, Salve, Bantanur, & Pawar, 2015).

BOTS: Bots (robots) are networks of malware-infected machines that are commandeered by servers known as Command and Control (C&C) servers(Mahato et al., 2015). They are malwares that enables an attacker to get complete control within the affected computer. Computers that are contaminated with a 'bot' usually are known as 'zombies'. You can find literally hundreds of thousands of computers on the web which might be contaminated with some type of 'bot' and don't even realize it. Attackers will be able to access lists of 'zombie' PC's and activate these to help execute DDoS (distributed denial-of-service) attacks against Sites, host phishing attack Internet sites or distribute a large number of spam email messages.

2.1 BACKGROUND:

The need for CAPTCHAs rose to keep out the website/search engine abuse by bots. In 1997, **AltaVista** sought ways to block and discourage the automatic submissions of URLs into their search engines. Andrei Broder, Chief Scientist of AltaVista, and his colleagues developed a filter. Their method was to generate a printed text randomly that only humans could read and not machine readers. Their approach was so effective that in a year, "spam-add-ons'" were reduced by 95% and a patent was issued in 2001(Civil, 2000).

In 2000, **Yahoo**'s popular **Messenger** chat service was hit by bots which pointed advertising links to annoying human users of chat rooms. Yahoo, along with Carnegie Mellon University, developed a CAPTCHA called EZ-GIMPY, which chose a dictionary word randomly and distorted it with a wide variety of image occlusions and asked the user to input the distorted word(Civil, 2000).

In November 1999, s*lashdot.com* released a poll to vote for the best CS College in the US. Students from the Carnegie Mellon University and the Massachusetts Institute of Technology created bots that repeatedly voted for their respective colleges. This incident created the urge to use CAPTCHAs for such online polls to ensure that only human users are able to take part in the polls(Jaiswal, 2010).

2.2 CAPTCHAs AND THE TURING TEST:

CAPTCHA application has its foundation in an experiment called the **Turing Test**. Alan Turing, sometimes called the father of modern computing, proposed the test as a way to examine whether or not machines can think -- or appear to think -- like humans. The classic test is a game of imitation. In this game, an interrogator asks two participants a series of questions. One of the participants is a machine and the other is a human. The interrogator can't see or hear the participants and has no way of knowing which is which. If the interrogator is unable to figure out which participant is a machine based on the responses, the machine passes the Turing Test. (Mahato et al., 2015)

Of course, with a CAPTCHA, the goal is to create a test that humans can pass easily but machines can't. It's also important that the CAPTCHA application is able to present different CAPTCHAs to different users. If a visual CAPTCHA presented a static image that was the same for every user, it wouldn't take long before a spammer spotted the form, deciphered the letters, and programmed an application to type in the correct answer automatically.

2.3 TYPES OF CAPTCHAs

CAPTCHAs are classified based on what is distorted and presented as a challenge to the user. They are:

A. Text CAPTCHAs:

Text CAPTCHAs involves text distortion and the user is asked to identify the text hidden or the user is told to enter the provided text in a text area. There are two major types of text CAPTCHA (among others not commonly used) used in securing the web server, these includes;

1. Gimpy:

Gimpy is a very reliable text CAPTCHA built by CMU in collaboration with Yahoo for their Messenger service. Gimpy is based on the human ability to read extremely distorted text and the inability of computer programs to do the same. Gimpy works by choosing ten words randomly from a dictionary, and displaying them in a distorted and overlapped manner. Gimpy then asks the users to enter a subset of the words in the image. The human user is capable of identifying the words correctly, whereas a computer program cannot.

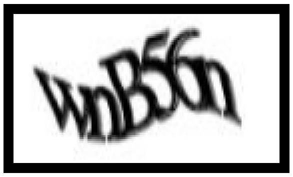


Figure 2: Sample of gimpy text CAPTCHA

2. Ez – Gimpy:

This is a simplified version of the Gimpy CAPTCHA, adopted by Yahoo in their signup page. Ez – Gimpy randomly picks a single word from a dictionary and applies distortion to the text. The user is then asked to identify the text correctly.

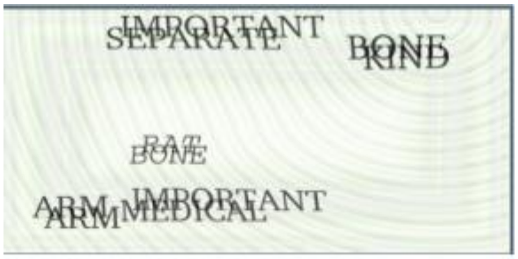


Figure 3: Sample of Ez-gimpy text CAPTCHA

B. Graphic CAPTCHAs:

Graphic CAPTCHAs are challenges that involve pictures or objects that have some sort of similarity that the users have to guess. Computer generates the puzzles and grades the answers, but is itself unable to solve it.

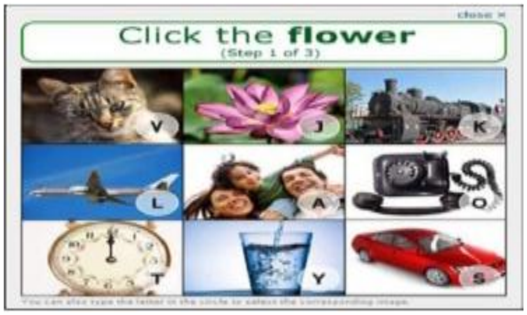

close ×
Click the flower
(Step 1 of 3)
V
J
K
L
A
O
T
Y
S
You can also type the letter in the circle to select the corresponding image.

Figure 4: Sample of graphic CAPTCHA

C. Audio CAPTCHAs:

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; the user is then asked to enter the exact words from the sound clip into a textarea.

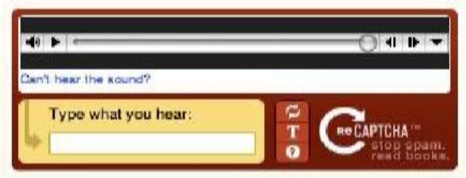
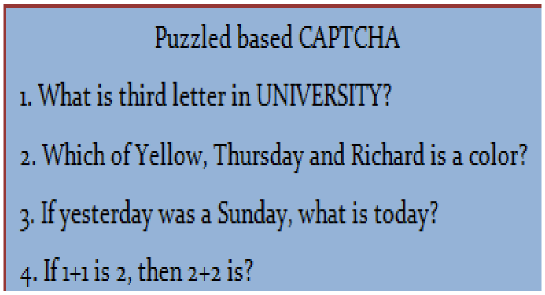


Figure 5: Sample of audio CAPTCHA

D. Puzzle based CAPTCHAs

Usually in puzzle based CAPTCHA; confirmed picture is divided into chunks. A user should combine these chunks so that you can make up the complete picture comparable to the first one. Again, in this type of CAPTCHA, you may be told to answer technical questions such as;



Puzzled based CAPTCHA

1. What is third letter in UNIVERSITY?

2. Which of Yellow, Thursday and Richard is a color?

3. If yesterday was a Sunday, what is today?

4. If 1+1 is 2, then 2+2 is?

Figure 6: Sample puzzle based CAPTCHA

3.0 HOW DOES CAPTCHA APPLICATION WORKS

A Web server may be holding both public and protected resources that may be in the form of web pages, data stored in a database or files or some other service intended to be used by human users on the client. User request for a resource is sent by the client computer to the server, which is granted to it if the resource is not protected. In case the resource is CAPTCHA protected, the access is granted to it only after passing CAPTCHA test as depicted in figure

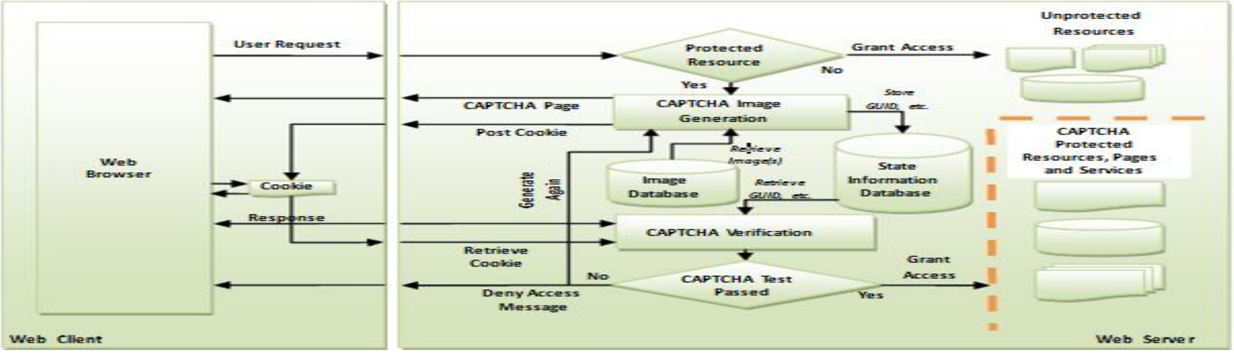


Figure 7: Working of CAPTCHA

The server uses some CAPTCHA image generation algorithm to generate a CAPTCHA image. Different CAPTCHA techniques use different algorithms for image generation which may employ use of images stored in an image database.

The state information along with global unique identifier (GUID) of the client and the CAPTCHA solution is stored in the state information database(SID) at the server. Storing GUID of the client ensures that only client that received CAPTCHA can produce a valid solution.

Instead of storing the CAPTCHA solution a nd other state information on server in SID, it be may stored in hashed or encrypted form in a cookie on the client. A web page containing the generated CAPTCHA image and the cookie is posted to the client which renders it in a web browser to the user.

A human operator responds to CAPTCHA test and the response is passed by t he client to the server. The server verifies the authenticity of CAPTCHA solution by comparing the stored GUID and the GUID of the client sending the solution.

The solution provided by the client is next compared with t he solution stored in SID or cookie and accordingly either access is granted or denied. In case access is denied, a message is posted to the client and the process starts afresh.

A CAPTCHA implementation may temporarily bl ock access for a client if it repeatedly fails to respond to a number of CAPTCHA tests. Further, for a particular session once a CAPTCHA a challenge has been passed by a client, subsequent accesses to protect resources on the server may be granted to it without putting it to further test.(Banday & Shah, 2011).

3.1 ISSUES WITH CAPTCHAs

There are many issues with CAPTCHAs, primarily because they distort text and images in such a way that, sometimes it gets difficult for even humans to read. Even the simplest, but effective CAPTCHA, like a mathematical equation “What is the sum of three and five?” can be a pain for cognitively disabled people.

Distortion is a big a problem when it is done in a very haphazard way. Some characters like ‘d’ can be confused for ‘cl’ or ‘m’ with ‘rn’ and ‘ɡ’or ‘q’ or ‘9’. Likewise in audio based CAPTCHAs, we may experience issues where the human can’t verify the sound of letters like “p”, “b”, and “e”. as a result of these issues, users tend to break text based CAPTCHAs with OCR “optical character recognition”(Jaiswal, 2010)

Audio CAPTCHAs can also cause compatibility issues. For example, many such schemes require JavaScript to be enabled. However, some users might prefer to disable JavaScript in their browsers. Some other schemes can be even worse. For example, we found that one audio scheme requires Adobe Flash support. With this scheme, vision-impaired users will not even notice that such a CAPTCHA challenge exist in the page, unless Flash is installed in their computers - apparently, no text alternative is attached to the speaker-like Flash object, either(Holmes, n.d.).

Apart from the above mentioned issues with CAPTCHAs, in the recent times, CAPTCHAs have been overridden by many ad-ons e.g. Rumola. Rumola is an add-ons which when installed on a browser and enabled, it automatically disables any CAPTCHA on the web page irrespective of the type of CAPTCHA and how strong it may be, if such occurs, the web servers are vulnerable to tremendous attack by bots. This has been the greatest pitfall of CAPTCHAs in the recent times.

Furthermore, expertise have been able to design many artificial intelligent viruses (bots) that can decipher CAPTCHAs faster than a human could comprehend, this involves mostly in an audio based CAPTCHAs where the human finds it difficult to understand letters such as “p” and “b” but the artificial intelligent virus gets it faster and more correct than the human may comprehend as well as solving puzzles.

As a result of these issues, many web service providers such as Google, Microsoft, facebook, whatsapp, e-banking system etc have switched off from CAPTCHA usage to direct access communication with human through sending tokens (authentication verification codes) as a text message of the user mobile number and sending links to user email.

4.0 FURTURE WORKS / RECOMMENDATIONS

Having seen the essence and pitfalls of CAPTCHA application for web server security in mitigation of DDoS, we hereby recommend further works on CAPTCHA application to provide better services;

1. CAPTCHAs (text, graphics, audio, puzzle) should be implemented with the consciousness of browser compatibility.
2. Stronger algorithms can be introduced for counter-measures such that OCR can't detect.
3. If there be a future work to be done on CAPTCHA application, it should be emphatically on the browser add-ons. Browsers should implement add-ons checks and policies before accepting any add-ons software.

4.1 CONCLUSION

Over the years, web servers have known so many threats from hackers by the use of bots and other automated forms, but the introduction and usage of CAPTCHAs verification application has proven a strong counter measure in mitigating DDoS attacks on web servers.

In this paper, we gave an overview on how to mitigate DDoS attacks on our web servers using CAPTCHAs, Though CAPTCHAs may have its pitfall in the recent days but it still serve as means to secure a system.

REFERENCES

Banday, M. T., & Shah, N. (2011). A Study of CAPTCHAs for Securing Web Services. *IJSDIA International Journal of Secure Digital Information Age*, *1*(2), 66–74. Retrieved from http://adsabs.harvard.edu/abs/2011arXiv1112.5605T

Bhandwalkar, D., Bhoi, A., Salve, K., Bantanur, C., & Pawar, M. V. (2015). Imagination, Detection and Mitigation of DDoS Attacks in Pdf File. *International Journal of Advanced Research in Computer Science and Software Engineering*, *5*(2), 804–810.

Civil, O. (2000). www.studymafia.org A Seminar report.

Guard, B. L. N. (2013). Universal DDoS Mitigation Bypass. *Black Hat USA 2013*.

Holmes, D. (n.d.). The F5 DDoS Playbook : Ten Steps for Combating DDoS in Real Time.

Jaiswal, U. (2010). *CAPTCHA and internet security*.

Mahato, S., Saxena, V. P., & Mishra, R. G. (2015). Securing Web Services and Applications using Captcha Security, *14*(April), 1–12.

Mehra, M., Agarwal, M., Pawar, R., & Shah, D. (2011). Mitigating denial of service attack using CAPTCHA mechanism. *Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET '11*. http://doi.org/10.1145/1980022.1980086