Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & the Africa AI Stakeholders Summit
Academic City University College, Accra Ghana
19th – 21st December, 2022
www.isteams.net/accrabespoke2022

Towards the Development of a Machine Learning Enhanceed Framework for Honeypot and CAPTCHA Intrusion Detection Systems

Muhammad, Hafsat, Longe, Olumide .B., Baale, Abimbola & Antai, U-O Ekpo
[1,3,4]Doctoral Programme in Cyber Security,
African Centre of Excellence for Technology Enhanced Learning (ACETEL)
National Open University of Nigeria, Abuja, Nigeria
[2]Faculty of Computational Sciences & Informatics, academic City University, Accra, Ghana
E-mail: olumide.longe@acity.edu.gh

ABSTRACT

With the continuous prevalence of cyber-attacks, information safety has become very important to governments and organizations all over the world. Individuals, organizations and government suffers from financial and reputational damages consequent to cyber-attacks. Hence development of good cyber security technique became very important in the literature. Generally, traditional IDS are passive in such a way that they detect and report attacks based on predefined rules. Traditional IDS focus on how to detect attacks base on a given rule, i.e. either assigned or abnormality (Muhammad 2010). Other traditional securityy detection strategies such as firewall and CAPTCHA intrusion detection systems (IDS) have been invented to protect the system's security, but there are still many critical issues which are reported every day (Salem et al 2008; Hauwa 2020). The situation worsens with the development of Internet technologies (Huang et al., 2019). Hence, cyber criminals continue to develop attack techniques against every cyber defense policy (Bukhari, S.et al,.2020). Therefore developed different cyber-attack methods such as phishing, image forgery, identity theft etc. On the other hand, academics and practitioners of cyber security are also developing cyber security measures. Hence, there is need for development of advanced security measures to curtail issues of cyber-attacks. This work proposes a framework that leverages on machine learning for same

Key words: Machine Learning, Enhanceed Framework, Honeypot, CAPTCHA Intrusion Detection

Proceedings Reference Format
Muhammad, Hafsat, Longe, Olumide .B., Baale, Abimbola & Antai, U-O Ekpo (2022): Towards the Development of a Machine Learning Enhanceed Framework for Honeypot and Captcha Intrusion Detection Systems. Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & the Africa AI Stakeholders Summit. Academic City University College, Accra Ghana, 2022. Pp 43-50. www.isteams.net/accrabespoke2022. dx.doi.org/10.22624/AIMS/ACCRABESPOKE2022/V34P4

1. INTRODUCTION

Modi et al (2010), stressed the development of soft techniques for Intrusion Detection and Prevention System. This led to the evolution of different "Completely Automated Public Turning Test to tell Computers and Humans Apart" (CAPTCHA) techniques for network security purpose (Huang 2019). There are too many malicious threats across the Internet which may compromise a system in the absence of any secure application which provides protection against such threats.

One such threat is the bot (Ling-zi. et al,. 2012) a malicious program which has the capability to run automated tasks over the network and thus creating problem in the network (Huang et al., 2019). Yet, even with CAPTCHA, Legitimate users of the internet can be attacked by web-bots in many ways (Abdullah et al., 2019) CAPTCHA and Honeypots are effective in introduction detection system. Yet, the lagged in detecting and stopping boats.

Andrew (2014) Classified methods used to exploit CAPTCHA into three: the Optical Character Recognition (OCR), Learning which used machine learning techniques to break CAPTCHAs and farming that exploit CAPTCHA by exposing it to humans to solve based on a certain reward, known or unknown to the solver and all were capable against many CAPTCHA defense systems. Hence a more robust technique is needed to curtail unwanted bots.

1.2 Problem Statement

With the continuous prevalence of cyber-attacks, information safety has become very important to governments and organizations all over the world. Individuals, organizations and government suffers from financial and reputational damages consequent to cyber-attacks. These attacks accounted for 6trillion dollars between the year 2019 and 2021 alone. Academics and security practitioners to the development of different cyber-defence techniques such as CAPTCHA. To an extent, CAPTCHA has been considered effective to certain intrusion detection and prevention system. On the other side, Attackers are discovering new techniques to break these security policies. Consequently, different bots, intelligent malware, and spywaresares capable of breaking through several CAPTCHA-trap IDS.

The White House's Office of Management and Budget revealed that of the 96 federal agencies it assessed, 74 percent were either "At-Risk" or "High Risk" for cyber-attacks, Which They needed immediate security improvements. Cyber-attacks can cause electrical blackouts, failure of military equipment, and breaches of national security secrets. They can result in the theft of valuable, sensitive data like medical records. They can disrupt phone and computer networks or paralyze systems, making data unavailable. It's not an exaggeration to say that cyber threats may affect the functioning of life as we know it. Hence, this research intend to use supervised learning on different CAPTCHA challenges data set on intrusions detection system using Random forest Algorithm.

1.3 Aim

The aim of this research is to create an intrusion detection system using Random Forest Algorithm reinforced CAPTCHA and honeypots.

1.4 Objectives

- Development of website to test IDS on malicious software for both intelligent and Non-Intelligent bots.
- Random forest will be used on training the Machine on different CAPTCHA challenges data set.
- Honeypots will be design and integrated into the website in other to lewd the malicious software
- A Reasonable amount of time for the period of six month to test the entire system for intrusion.

1.5 Motivation of the Research

This thesis intends to develop an intrusion detection system using Random Forest Algorithm to reinforce CAPTCHA technique against bots and other prevalent challenges associated with CAPTCHA. The research is motivated by:

- The motive is based on the work of Hauwa et al,. (2018) where their research was done on CAPTCHA based intrusion detection model an the work of abdullah et al,.(2020). enhenced intrusion detection system using honeypot and captcha techniques without the use of machine learning.
- Relative absence of research on use of machine learning to support CAPTCHA against bots.
- The need for new intrusion detection and prevention techniques to contain emerging network security challenges.

2. RELATED LITERATURE

This section summarizes some of the CAPTCHA techniques used for designing and developing IDS and the Honeypot and machine learning for reinforcing the CAPTCHA. According to Hauwa et al (2020), the world is advancing to global connected system many are getting connected to do business transactions and many aspects of life transforms. At the same time it also brings in lot of security risks to the business over the network. It is stated by Kumar et al (2013), with the growth of cyber-attacks, information safety has become an important issue all over the world. Different Intrusion detection techniques have been used to support the security of an organization against threats or attacks. On one side, attackers are discovering new techniques and ways to break these security policies (Abdullah et al 2019)

Generally, traditional IDS are passive in such a way that they detect and report attacks based on predefined rules . Traditional IDS focus on how to detect attacks base on a given rule, i.e. either assigned or abnormality (Muhammad 2010). That means a new attack that is not defined in the system will not be detected, also some interactions with genuine human may ambiguously be considered as a threat. Modi et al (2010) Reviewed most of the soft computing techniques used in IDS development. They comprise the biologically-inspired techniques like the genetic algorithm and some major machine learning tools like Fuzzy logic, support vector, artificial neural network and software defined networking. Honeypots and random forest algorithm are amassing recent attention of literature. Modi et al (2010), emphasized that soft techniques application in IDS and IPS will optimally improve system security.

Steven et al (2011) Developed 'DeCAPTCHA' software which enables the computer to render the CAPTCHA code legible by cleaning up the text. The software was successful in cleaning up 66% of Visa's Authorize.net CAPTCHAs, 70% of Blizzard's Entertainment and 73% of CAPTCHA.com's CAPTCHA. Gu et al (2006) Designed a generic method to break all text based CAPTCHAs, which is considered the best in generic attack, with the success rate of 5% to 77%, which can solve a puzzle in less than 15 seconds average speed in a standard desktop. Text-based CAPTCHAs are more of human friendly but vulnerable to attack. Andrew (2014) Classified methods used to exploit CAPTCHA into three: the Optical Character Recognition (OCR), Learning which used machine learning techniques to break CAPTCHAs and farming that exploit CAPTCHA by exposing it to humans to solve based on a certain reward, known or unknown to the solver.

Foley (2012), Reviewed different CAPTCHAs and categorized them into three i.e. the visual, non-visual and hybrid. They evaluate them and suggest some alternatives base on the given criteria that will be considered when prioritizing the selection and implementation of the CAPTCHAs. The criteria consist of the cost, efficiency and robustness on usability. They identified security and usability as the major barrier in CAPTCHA deployment.

They suggested some alternatives to CAPTCHA which were categorized into three that includes; Administrative, Interactive and cheating bots and tested by the same CAPCTHA criteria, with great improvement in both the usability and security in their deployment. Powell et al (2017), Designed fCAPTCHA, which consists of multiple image of human faces and non-face image with varying degree of distortion. Users get access by matching faces belonging to a single individual. After a comparative analysis of different CAPTCHAs and their alternative, Parita et al (2016) arrived at a conclusion with a suggestion, that honeypots and CAPTCHA have their respective weaknesses and drawbacks that makes them independently less effective, but by integrating and removing the weaknesses will form a viable defense for online system.

Powell et al (2017) designed a novel image-based CAPTCHA that uses object recognition inspired by negative selection algorithm of the biological immune system. It also has a two-phase filtering algorithm which ensures that the CAPTCHA is resilient to automated attack while remaining easy for human users to solve. The image CAPTCHA is not convenient for users, hence the need to completely eliminate CAPTCHA's should be considered (Josh, 2019). To eliminate traditional CAPTCHA, Google introduced reCAPTCHA that makes verification simple for users by only clicking on a checkbox while making it harder for bots.

The reCAPTCHA works using an advanced risk analysis that comprises of browsing history of the genuine user already tracked by google cookie just to determine the difficulty of challenge that is presented to the user, explores how aspect of the browser environment affects the risk analysis, canvasses rendering techniques to fingerprint users across machines and browsers, identifies how user-agent influence the user's reputation and the timing of movement and movement pattern of mouse to decide what type of challenges will be presented to the user (Suphannee et al 2019).

An advanced no CAPTCHA reCAPTCHA, which is invisible to human, was introduced in 2017 by Google due to the trouble with reCaptcha that drives users to the extreme edge of sanity.
The invisible CAPTCHA shows no challenge to user, instead it returns probability scores between 0.0 (100% bots) and 1.0 (100% human) (Abdullahi et al, 2019).

3. RESEARCH DIRECTION & METHODOLOGY

We present below our proposed framework.

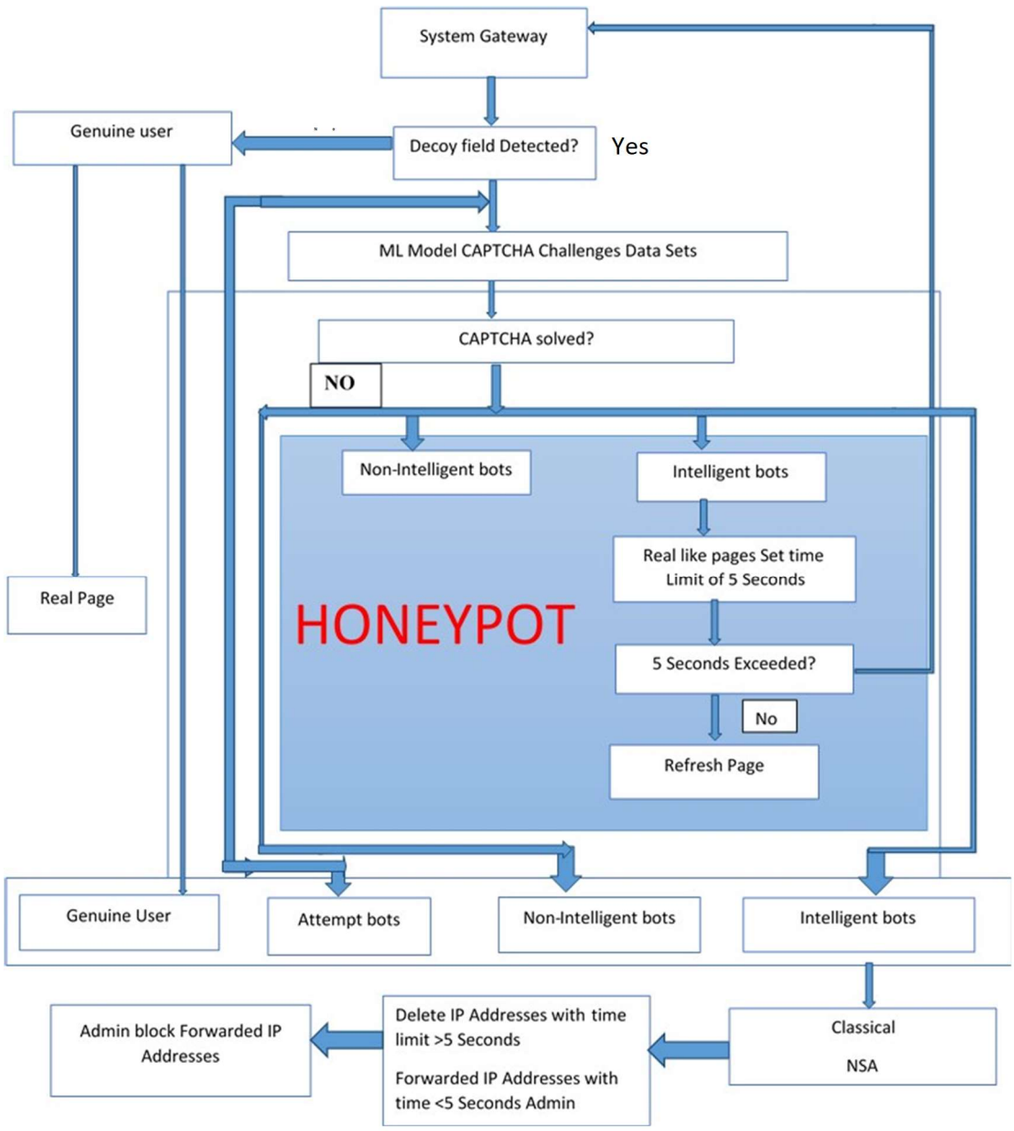

Fig 1: Proposed System Framework For MI Reinforced Captcha Model
Source: Researchers' Fieldwork

4. CONCLUDING REMARKS

With the prevalence of cyber-attacks, information safety has become very important to governments and organizations all over the world. Individuals, organizations and government suffers from financial and reputational damages consequent to cyber-attacks. To some extent, CAPTCHA has been considered effective to certain intrusion detection and prevention system. On the other hand, attackers are discovering new techniques to break these security policies. Consequently, different bots, intelligent malware, and spywaresares capable of breaking through several CAPTCHA-trap IDS. It has therefore become imperative to continue to evolve means of strengthening existing CAPTCHA Systems while also developing novel ideas that can enable the security community stay ahead of criminals in cyber space.

REFERENCES

1. I.M .Azhaghir, A.Rajesh and S.Kathick, (2016). "INTRUSION DETECTION AND PREVENTION SYSTEM: TECHNOLOGIES AND CHALLENGES". International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.87 © Research India Publications; http/www.ripublication.com/ijaer.htm
2. Joseph, C. (2018). Symantec Internet Security Threat Report 2018: The Top Takeaways. [Blog post]. Retrieved from: https://thycotic.com/company//blog/2018/04/17/symantec-internet-security-threat-report-2018/. It is a snapshot of the page as it appeared on 26 Aug 2021 10:22:50 GMT.
3. Abdullahi,M., Aliyu,S. and Junaidu,S.B(2019).''AN ENHENCED INTRUSION DETECTION SYSTEM USING HONEYPOT AND CAPTHA TECHNIQUES"*FUDMA Journal of Sciences (FJS)ISSN online: 2616-1370 ,ISSN print: 2645 – 2944Vol. 3 No. 3, September, 2019, pp 202- 209*
4. Singh,P.V. and Pal,P.(2014)."SURVEY OF DIFFERENT TYPES OF CAPTCHA"
5. Souley, B. and Abubakar, H.(2018)." A CAPTCHA –BASED INTRUSION DETECTION MODEL"International Journal of Software Engineering & Applications (IJSEA), Vol.9, No(1)DOI: 10.5121/ijsea.2018.9103 29
6. Ling-Zi,Z. and Yi-Chun Z.(2012) "A Case Study of Text-Based CAPTCHA Attacks," in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover.
7. Saini,S.B. and Bala,A.(2013)"A REVIEW OF BOT PROTECTION USNIG CAPTCHA FOR WEB SECURITY"IOSR Journal of Computer Engineering, pp. 36-42
8. Khalifa,W. and Hassan,A.(2016)"A SURVEY OF CURRENT RESEARCH ON CAPTCHA"AljabelAlGharbiUniversity,Faculty of Eng-Jadu ,EE dept, Gharyan, Libya.International Journal of Computer Science & Engineering Survey (IJCSES) (7) (3).
9. Diebold,P.Hess, and Schafer,G.(2005)."A Honey pot architecture for Detecting and Analysing Unknown NetworkAttacks"In Proc. Oh 14th KommunikationinVerteiltensystemen2005 (KiVS05), Kaiserslautern, Germany.

10. INTERNATIONAL JOURNAL FOR RESEARCH & DEVELOPMENT IN TECHNOLOGY Volume-7,Issue-2 (Feb-17) ISSN (O) :- 2349-3585 All rights reserved by www.ijrdt.org 62 HONEYPOTS FOR NETWORK SECURITY
11. Samuel AL. Some studies in machine learning using the game of checkers. IBM Journal of research and development. 1959 Jul; 3(3):210-29
12. Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2. ISBN 978-0-07-042807-2. Dept of MCA, Gnanamani college of Technolgy, Namakkal, INDIA
13. HAUWA A., BOUKARI S.,& ABDULSALAM Y. G. (2020) : AN IMPROVED CAPTCHA – BASED INTRUSION DETECTION SYSTEM BASED ON REDIRECTOR MODEL *Journal of Theoretical and Applied Information Technology*
14. Yesugade, K. D., Avinash, M. S., Satish, N. S., Sandeep, S. C., & Malav, S. Infracstructure Security Using IDS, IPS and Honeypot. *International Engineering Research Journal (IERJ), vol 2, issue3.*, (2016). pp. 851-855.
15. Agnaou, A., El Kalam, A. A., Ouahman, A. A., & De Montfort, M. Automated Technique to reduce Positive and Negative False from attacks collected through the deployment of distributed honeypot network. *International Journal of Computer Science and Information Security, vol 14 Issue 9.* (2016). *(IJCSIS).*
16. Ashwini, M. K., Pratiksha, G., Anuja, K., Varsharani, S., & Gayatri, S. Secure Network System using Honeypot. *International Journal of Advanced Research in Computer and Communication Engineering,(IJARCCE), vol 2*, (2017). pp.230-232.
17. Modi, U., & Jain, A. An Improved Method to detect Intrusion using Machine Learning Algorithms. *Informatics Engineering, an International Journal (IEIJ), vol 4 issue2.*, (2016). pp. 17-29
18. Kumar, B., Phani Raju, T. S., Ratnakar, M., Baba, S., & Sudhakar, N. Intrusion Detection System- Types and Prevention. *International Journal of Computer Science and Information Technologies (IJCSIT), Vol 4 issue 1, (2013). pp. 77- 82*
19. Abdullahi, M., Aliyu, S. and Junaidu, S. B. (2019) An Enhanced Intrusion Detection System Using Honeypot And Captcha Techniques
20. Joseph, C. (2018, April, 17). Symantec Internet Security Threat Report 2018: TheTop
21. Mohammad, M., &Mohammad Reza, K. (2014). CAPTCHA and its Alternatives: A Review.
22. *Security And Communication Networks Security Comm. Networks*, 8: 2135–2156.
23. Foley, A. (2012). Biometric Alternatives to CAPTCHA: Exploring Accessible Interface Options. *Dublin Institute of Technology.*
24. Mohammad A. F. & Syed S. H.(2010). Towards Cyber Defense: Research in Intrusion Detection and Intrusion Prevention Systems, *International Journal of Computer Science and Network Security* (IJCSNS), 10(7).
25. Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). Asurvey of intrusion detection techniques in cloud. *Journal of network and computer applications, 36*(1), 42-57.
26. Steven G., Mengjun X., Zhenyu Wu., &Haining W. (2011). Humans and Bots in Internet
27. Chat: Measurement, Analysis, and Automated Classification. *, ieee/acm transactions on networking*, 19(5).

28. Gu, G., Fogla, P., Dagon, D., Lee, W., Skori, C. (2006) Measuring intrusion detection capability: An information-theoretic approach. In: Proc. of the 2006 *ACM Symposium on Information, computer and communications security*, pp. 90–101.
29. Powell, B. M., Kalsy, E., Goswami, G., Vatsa, M., Singh, R., & Noore, A. (2017). *Attack-resistant aiCAPTCHA using a negative selection artificial immune system.* Paper presented at the 2017 IEEE Security and Privacy Workshops (SPW).
30. Andrew, D. (2014). ESCAPT: Easy Strategies for Computers to Avoid the Public Turing Test. *Mentor: Ming Chow Fall*.