Journal of Theoretical and Applied Information Technology
15th February 2020. Vol.98. No 03
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN IMPROVED CAPTCHA – BASED INTRUSION DETECTION SYSTEM BASED ON REDIRECTOR MODEL

[1] HAUWA ABUBAKAR,[2] BOUKARI SOULEY, [3] ABDULSALAM YA'U GITAL

[1] Department of Computer Science, Umar Suleiman College of Education Gashu'a, Yobe, Nigeria.
[2,3] Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria.
E-mail: [1] bbkrhauwa@gmail.com, [2]bsouley2001@yahoo.com, [3]asgital@yahoo.com

ABSTRACT

This study deals with an Improved **C**ompletely **A**utomated **T**urning test to tell **C**omputers and **H**umans **A**part (CAPTCHA)-Based Intrusion Detection System Based on Redirector Model. In advanced technological world today, internet usage is inevitable but there are certain Network security challenges such as intelligent spywares that are capable of breaking and bypassing the CAPTCHA. This study aims at designing CAPTCHA-based IDS so as to identify the intelligent spywares that are capable of breaking CAPTCHAs in Intrusion Prevention System (IPS). Therefore, the study tries to design IDS with intelligent redirector using CAPTCHA- Trap of detecting Intrusion by intelligent spywares. The work is set to design an email website which was integrated with the new CAPTCHA based IDS system and host it online. A dummy honeypot system was designed in which the IP addresses of the software intruders that intruded into the system were captured. The work also implemented and evaluated the performance metrics which includes Detection Rate (DR), Precision (PR), False Positive Rate (FPR), correctly and incorrectly classified instances of the system against the Existing system. The researchers use captured IP addresses as the datasets for the study. The work employed Waikato Environment for Knowledge Analysis (WEKA) and Python to analyze the experimental data. The findings of the study indicate a high percentage of Detection Rate and a very low False Positive rate as compared to the existing system. It is recommended that website users and webmasters should incorporate the use of the Improved CAPTCHA-based IDS, IPS and Honeypot so as to block intruders and collect information about them for necessary measures.

Keywords: *IDS, IPS, CAPTCHA, Detection Rate, Intelligent Spyware.*

1. INTRODUCTION

As 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 [1], with the growth of cyber-attacks, information safety has become an important issue all over the world. Different techniques are 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. Securing the network and detecting large number of attacks by intelligent spywares is one of the major concerns for network administrators. The drawback with use of the internet is the security challenges since intelligent spywares breaks and bypassed the Completely Automated Turning test to tell Computers and Humans Apart (CAPTCHA) in Intrusion Prevention System (IPS). Intrusion Detection system (IDS), IPS and Honeypot combined are some of the security measures applied on the network as a hybrid security model to detect those intelligent spywares. The objective of the hybrid security model is to detect, prevent and monitor the event of the intruders' intent on the network. This work designed and implemented Improved CAPTCHA-based IDS with a redirector to curb the security failure identified when CAPTCHA was used in IPS. The work integrated CAPTCHA in IDS to detect and identify spywares that bypass and break into the system. The system with the CAPTCHA- Trap IDS detects and redirects the detected attacker into the dummy Honeypot, the Honeypot lured the software and capture its IP address and block it from accessing into the real system. The work has a major significance in online transaction when network infrastructure requires regular security treat prevention, monitoring, detection and recovery. The need to evaluate the performance of the algorithms of attack detected through our CATPCHA- Based IDS based on redirector model to secure the system against the attack by the intelligent spywares.[2] Stated that detecting attacks is an essential need in networks. [3] Define Intrusion as effort to compromise Confidentiality, Integrity, and/or Availability (CIA) in a computer system or network. Intrusion is a process where software accesses web content that is protected with username and password. The software may use different infiltration methods to break the username and even the password. Intrusion occurs in webmail accounts, face book accounts, twitter, LinkedIn, or any login page. At a wider sense, intrusion may include both human and machine access to account having web content that is secured with username and password. Sometimes humans and software combined forces to achieve intrusion. To overcome this problem Network Security provides many techniques and one of the most important techniques is Intrusion Detection System (IDS), [4].

An Intrusion Detection system as defined by [5] is the process of intelligently monitoring the events occurring in a computer system or network, analyzing them for signs of violations of the security policy. The primary aim of Intrusion Detection Systems (IDS) is to protect the availability, confidentiality and integrity of critical networked information systems.

2. REVIEW OF RELATED WORKS

This section summarizes some of the techniques that used for designing and developing IDS. In [6], proposed the system which combines specific features and services of Intrusion -Detection-System (IDS), Intrusion-prevention-System (IPS) and Honeypot. Because various exploits are being used to compromise the network. These exploits are capable of breaking into any secured networks. In order to increase efficiency of network security, they introduce Honeypot. Honeypot detect attacks with the help of IDS; trap and deflect those packets sent by attackers. The result of their work indicates that the system handles multiple clients using the concept of Honeypot. They also pointed out that, Intrusion detection system (IDS) monitor whole network and looks for intrusion. When any intrusion occurs Honeypot will be activated. This activated Honeypot will divert the traffic to dummy/virtual servers & back track the source (IP address) or origin of that attack. The drawback of the system is that since it supports multiple clients including an attacker the system can easily be compromised.

In the work of [7] **,** which the system proposed combination of the features, functions and methodology of IDS, IPS and Honeypot and making Intrusion Detection System more effective, accurate and responsive. In their work it was evident that Honeypot, IDS and IPS are eventually deployed on the gateway for analyzing incoming network traffic. The main server was connected to Internet Service Provider (ISP) through external router. All incoming packets from external network will be first arriving on the mirror server i.e. Honeypot to a capture the logs. The result of their work shows that the proposed system is more stable and precise on operating system platform also on detection ratio. The systems has introduced a sophisticated and interactive user friendly interface to configure and monitor the software and also to analyze and log the behaviour of the intruder and intruding events. The only drawback of their proposed system the detection module did not include how the IDS should be capable of detecting intrusion by spywares since it is evident CAPTCHA on the IPS can be broken by spywares.

[3], proposed "an improved method to detect intrusion using machine learning algorithms" with KDDCUP 99 dataset, which is simulated on WEKA tool. Their proposed method detects individual attacks presents in KDDCUP 99 dataset fast and efficiently. They further pointed out that the algorithms can be modified according to environment of network to make better detection rate and time.

[8] proposed an automated algorithm that gives ability to detect attack before their occurrence; their approach reduces the positive and negative false rates. They also use of data related to malicious traffic captured using a network of Honeypot to recover potential threats sources. Their goal was to characterize the root cause of the attacks on the Honeypot.

[9], proposed the implementation of middle interaction production honeypot. Their main goal is to secure the server side using honeypot from the attackers. The result of the work indicates that Clients can communicate to the servers through the honeypot only. The clients have the fake IP address of the honeypot and not the server's. If the client is a genuine client then its request goes to honeypot. Honeypot changes its IP address and forwards the request to the original server. After that server gives response to honeypot. Again honeypot changes its IP address

and sends response to client. If the client is fake client, then attacker will be tracked , located, identified and saves information about attackers at the honeypot. Though it is an attacker it gives response to make them fool. In all these scenario, security is maintained. The limitation here is that if no any attacker comes in the honeypot system becomes useless the need for other established security tools such as IDS and IPS to be integrated with the honeypot.

[10], Their work develop a new hybrid model that is used to estimate the intrusion Scope threshold degree based on Network transaction data's optimal features. Detection time is reduced and accuracy rate also improved.

[11] In their approach they presented a statistical and complexity analysis of CIDDS-001 dataset. They further utilized supervised and unsupervised machine learning techniques to analyze the complexity of the dataset in eminent evaluation metrics. Their evaluation results shows that k-nearest neighbor, decision trees, random forests, naive bayes and deep learning based classifiers can be used to develop an efficient network intrusion detection system.

[12] In their work they conducted a set of experiments and evaluate machine learning techniques for detecting malware and their classification into respective families dynamically. they received a set of real malware samples and benign programs from Virus Total, and executed in a controlled & isolated environment which they recorded malware behavior for evaluation of machine learning techniques in terms of commonly used performance metrics which was extracted from malware and benign executable samples using a Cuckoo Sandbox and a Python based automated system to form a real malware dataset.

There are many researches that have been conducted in the field of Intrusion detection system. [13], described the log based IDS approach which they employed on the user request, the system was designed in such way that it will predict the multiple request and block that particular user and then analyze the auditing log. This was achieved by analyzing the user MAC and IP address and is then blocked.

2.1 Comparison of Different Intrusion Detection Systems used on a Network security

This section shows a comparison of most closely related work done by researchers using Intrusion detection security models in the Web and their limitations as displayed in table1 below.

Table 1: Comparison of Different Intrusion Detection Systems used on a network security

S/No	Author (s)	Research Title	Result provided	Issues addressed
1.	[14]	Honeypot Intrusion Detection System	In their study, they developed a Honeypot IDS system	Performance identified on the Network Security Monitor (NSM).
2.	[15]	A Survey on Potential Applications of Honeypot Technology in Intrusion Detection Systems.	The work combined Intrusion detection system with Honeypot that detect new attack pattern that do not exist in signature database.	Cost of Security on Corporate Network minimized, false positive alarm level also reduced.
3.	[16]	Simple Text Based CAPTCHA for the Security in Web Applications.	In their work, they implement a simple Text Based CAPTCHA which provides simplicity of solving the techniques for human as well as the time that a human actually needs to find the solution.	Solution to maximize robustness and usability of text based CAPTCHA simultaneously is provided.
4.	[17]	Optimize Machine Learning Based Intrusion Detection for Cloud Computing: Review Paper.	The research work presents an anomaly detection technique based on Genetic Algorithm (GA) and Support Vector Machine (SVM) performance valuation was improved.	Analysis of data set KDD99 the most popular data set on IDS provided.
5.	[7]	Infrastructure Security Using IDS, IPS and Honeypot	In their work, they Implement IDS, IPS and Honeypot making	Response time and efficiency of detection

			Intrusion Detection System more effective, accurate and responsive.	increased.
6.	[10]	Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model	The work develop a new hybrid model that is used to estimate the intrusion Scope threshold degree based on Network transaction data's optimal features.	Detection time is reduced and accuracy rate also improved.
7.	[18]	Honeypot Based Detection System with Snooping agents and Hash Tags	The work provides an Improvement on the Performance of Intrusion detection System using Honeypot with Snooping Agents and Hash Tags.	Detection rate improved and False positive reduced.
8.	[13]	Log Based Intrusion Detection Model.	The work Proposes the use of Intrusion detection model based on User Request and multiple request are predicted that belong to a particular user are blocked and his auditing logs also analyzed.	Unauthorized events are predicted and DDOs attacks are prevented by the log file.
9.	[19]	A two-stage flow-based Intrusion Detection Model for next Generation Networks	Their work propose a flow based Intrusion Detection System in two stage detection process, first stage uses one class SVM for detection all normal traffics are discarded and malicious traffic are forwarded to the second stage and only malicious flows are analyzed in details.	Grouping of Malicious flows in different attack clusters and accuracy in the separation in their separation is provided.

3. METHODOLOGY

The major concern of this work is to evaluate the performance of the security model in a network web based system proposed by [1] using CAPTCHA- based IDS as network security to secure users data in a web with a developed security model in the network using CAPTCHA-based IDS, IPS and Honeypot.

3. 1 Working Principle of Improved CAPTCHA-Based IDS based on Redirector Model with loosely Generated CAPTCHA

The working principle of the CAPTCHA – based IDS with a redirector as proposed in [1] is shown in Figure1 below.

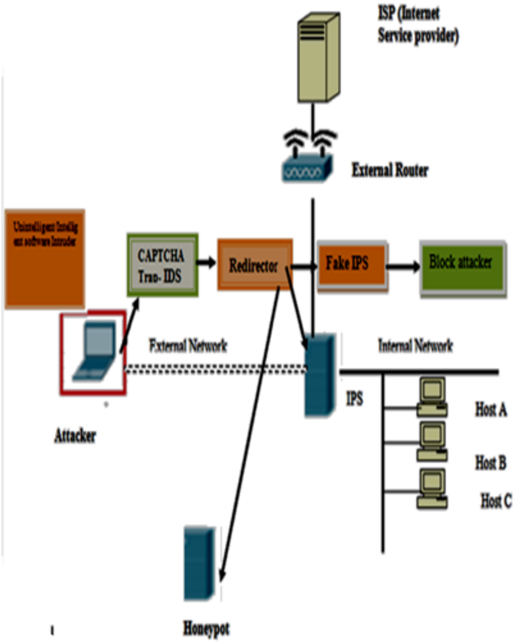


Figure 1. CAPTCHA- based Intrusion Detection System proposed in [1]

CAPTCHA is a tool commonly used in IPS, to prevent machine intruders (bots) from intruding into a system, however, in this research work, CAPTCHA will also be used as IDS. The technique to be used is cognizance of the fact that there are software intruders that can read

CAPTCHA and attempt to infiltrate it and intrude into system. CAPTCHAs with weak design pattern and fixed length with varying colours on text will be employed for use in web-based system acting like IPS while in real sense it is an IDS that will attempt to lewd software intruders using machine learning-based attack to successfully read the text-based CAPTCHA and infiltrates the system. Likewise software intruders using unwitting human labour can easily read the CAPTCHAs and infiltrate the system as well. The CAPTCHA character is not only to be read and re-type back to the system, it is to be read, understand and abide by. For instance one of such CAPTCHA may displayed "**DO NOT TYPE ANYTHING IN THE TEXTBOX**" a wise human will understand that the textbox should be left empty, while a software intruder that successfully read the words will just rush and type "**DO NOT TYPE ANYTHING IN THE TEXTBOX**" in the textbox or something similar to that sentence. As soon as anything is typed in the provided textbox, the system quickly detects that an intrusion has taken place and the intruder is quickly redirected to the Honeypot model for post intrusion activities. Figure 2 illustrates the IDS using CAPTCHA as trap.

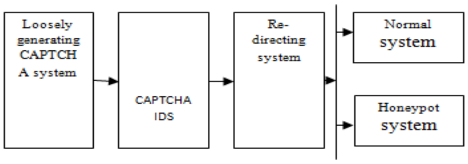


Figure 2: IDS using CAPTCHA

The presence of CAPTCHA in this system naturally deters software intruders, since CAPTCHAs are generally seen as IPS and therefore, its presence in the system will naturally interpreted as an IPS system. However, this system faking software intruders to believe it is IPS may not be very efficient as some software intruders may attempt login ignoring the CAPTCHA and the textbox provided for it, this will re-direct the intruder to the login authentication and somehow this may have been an intrusion bypassing the fake IPS and the CAPTCHA-trap IDS. This is an instance where the system may theoretically be bypassed and subsequent researches on similar are recommended studying this gap.

3. 2 Implementation process

The system was developed as a real world system which is an Email web server and a CAPTCHA Trap IDS incorporated on the site using the following technologies as: **Back End** includes PHP (Hypertext Pre-processor) and MySQL. For the **Front End we use** JavaScript, HTML (Hyper Text Mark-up Language) and CSS (Cascading Style Sheet). The following algorithm best illustrate our CAPTCHA-based IDS model as follows;

Algorithm for the proposed model

```
Step 1: Start
Step 2: Enter Login details
Step 3: Solve CAPTCHA challenge
Step 4: IF (login details correct) {
Step 5: IF (CAPTCHA Text = 1) {
    Intelligent spyware Get IP address
    Send User IP address to Honeypot and Database
Step 6 :}
        Else {
Step 7: Redirect user to Mailbox
Step 9 :}
Step 10: Else {
Step 11: Repeat Step 2
Step 12 :}
Step 13: Stop
```

3. 3 Working Interfaces and structure of the proposed system

Figure 3 below shows the system interface for the Login into the Honeypot mail of the system by the Administrator. It contains Home, Sign Up and Login.

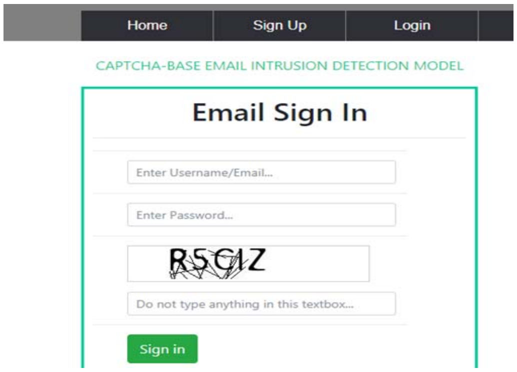
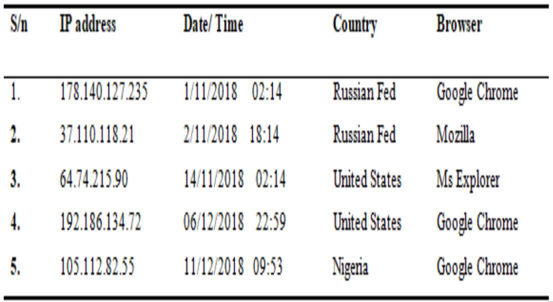

Home
Sign Up
Login
CAPTCHA-BASE EMAIL INTRUSION DETECTION MODEL
Email Sign In
Enter Username/Email...
Enter Password...
R5GIZ
Do not type anything in this textbox...
Sign in

Figure 3: CAPTCHA based email intrusion detection login Sign Up and login page

Table 2 shows report from event recording by the proposed system. Successful intrusions are stored in MySQL table for record-keeping purposes. Items captured and save on the database are: IP address of machine the software intruder is coming from, Date and time of intrusion. The browser used and country.

Table 2: Report Of Intrusion Detected On The System



S/n	IP address	Date/ Time	Country	Browser
1.	178.140.127.235	1/11/2018 02:14	Russian Fed	Google Chrome
2.	37.110.118.21	2/11/2018 18:14	Russian Fed	Mozilla
3.	64.74.215.90	14/11/2018 02:14	United States	Ms Explorer
4.	192.186.134.72	06/12/2018 22:59	United States	Google Chrome
5.	105.112.82.55	11/12/2018 09:53	Nigeria	Google Chrome

3.4 Experimental data

The system was experimented using the extracted data from the database. The entire system was fully implemented.

The system was experimented using different data sets as follows;

1. No of Months
2. No of IP address captured
3. Class (human/ Bots)

3.5 Performance Metrics

An IDS is evaluated by the measure of Accuracy, detection rate and F- Measure. It should have a very low false alarm. [20].

The performance of the existing and the proposed systems was evaluated using the following performance metrics:

The analysis was performed using the WEKA 3.8.0 (Waikato Environment for knowledge analysis).WEKA which is an open source machine learning scripting tool which was developed by Waikato University) New Zealand. The existing and proposed system was evaluated using the following perimeters:

Precision: - It express as = $\frac{TP}{TP+FP}$ … (1)

Precision: - is a measure of what fraction of test data detected as attack are actually from the attack class.

Detection Rate; which is the number of suspected attacks over all the attacks that exist in the dataset and the formula for computing the Detection Rate is as follows;

Detection Rate = $\left(\frac{TP}{TP+FN}\right) * 100$ … (2)

TP expressed as True Positive
TN expressed as True Negative
FP expressed as False Positive and
FN expressed as False Negative As adopted by [21]

False Positive Alarm Rate /Ratio this can be calculated through the use of confusion matrix parameters which are represented by the following values TP, FN, FP and TN

1. **True Positive (TP)** - Indicates the instances which are predicted as normal
2. **False Negative (FN)** – This illustrate that incorrectly classification process occurs. Where attacks are classified as normal packets.
3. **False positive (FP)** – This value represents incorrect instance where normal packets are classified as attacks.
4. **True negative (TN)** – This represent the correct instance normal packets as being normal.

Here the False Positive Rate is defined as FPR = $\left(\frac{TP}{TP+TN}\right) * 100$ … (3)

As adopted by [22] and [23].

We also obtained the detection accuracy for both systems using the Naïve Bayes classifier algorithm in Python language, which we obtained the percentage accuracy of detection for both the Existing and the proposed system. The naïve Bayes is a classification method based on Bayes theorem which is used to predict class labels. [24].

4. RESULTS AND DISCUSSIONS

The Improved CAPTCHA – based IDS detection technique was developed on a Windows operating system 500 GB Hard drive, 4GB RAM and a Core i5 processor of 2.20GHZ. A modem for internet access is needed a domain from ISP also used. The software requirement includes: Xampp local server for running the application on local host. Latest internet browser Google chrome, Firefox, Mozilla or Opera etc. and a strong Internet connectivity was required. The analysis was performed using the WEKA 3.8.0 (Waikato Environment for knowledge analysis).WEKA which is an open source machine learning scripting tool which was developed by Waikato University) New Zealand. To perform the experiments, we use all the 96 captured IP address normal instances and the 33 attack instances that make up the data Training set for the head and tail. Our system performs binary classification; the classes on each dataset are reduced to only two: normal and attack, where the attack class consists of all Captured IP address of the host intruders that attack the system within period of time. We also obtained the detection accuracy for both systems using the Naïve Bayes classifier in Python. Naive Bayes is a supervised learning classifier that based on Bayes' Theorem with the "naive" notion of independence among variables of a problem. This notion means that the presence of one variable in a problem does not have any effect on the presence of another variable. Naive Bayes uses the conditional probability. It classifies the problem by combining previous calculated likelihood and probabilities to make the next probability using Bayes rule. It is used in text classification, spam detection, recommendation system etc. [25].

4.1 Result

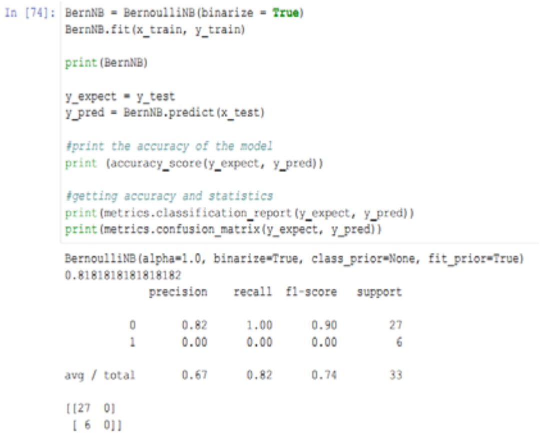
The results for the experimental analysis of the Improved CAPTCHA-base IDS are described in this section.

4.2 System Testing

The system was tested on a real world; an email web site was designed and in-cooperated on web server having a login and signing up details for account creation.

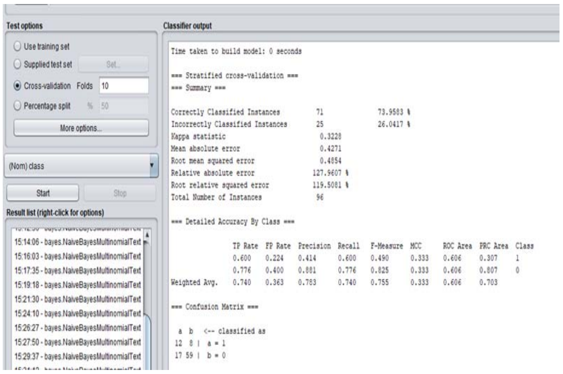
4. 3 Analysis of the experimental results

Figure 2 below illustrate result analysis obtain from the experiment conducted using Python



```
In [74]: BernNB = BernoulliNB(binarize = True)
         BernNB.fit(x_train, y_train)

         print(BernNB)

         y_expect = y_test
         y_pred = BernNB.predict(x_test)

         #print the accuracy of the model
         print (accuracy_score(y_expect, y_pred))

         #getting accuracy and statistics
         print(metrics.classification_report(y_expect, y_pred))
         print(metrics.confusion_matrix(y_expect, y_pred))

         BernoulliNB(alpha=1.0, binarize=True, class_prior=None, fit_prior=True)
         0.8181818181818182
                      precision    recall  f1-score   support

                   0       0.82      1.00      0.90        27
                   1       0.00      0.00      0.00         6

         avg / total       0.67      0.82      0.74        33

         [[27  0]
          [ 6  0]]
```

Figure 2. Screen captured of Run analysis of the Improved CAPTCHA based IDS in python



Test options
○ Use training set
○ Supplied test set
Set...
● Cross-validation Folds 10
○ Percentage split % 50
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
15:14:06 - bayes.NaiveBayesMultinomialText
15:16:03 - bayes.NaiveBayesMultinomialText
15:17:35 - bayes.NaiveBayesMultinomialText
15:19:18 - bayes.NaiveBayesMultinomialText
15:21:30 - bayes.NaiveBayesMultinomialText
15:24:10 - bayes.NaiveBayesMultinomialText
15:26:27 - bayes.NaiveBayesMultinomialText
15:27:50 - bayes.NaiveBayesMultinomialText
15:29:37 - bayes.NaiveBayesMultinomialText
Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          71               73.9583 %
Incorrectly Classified Instances        25               26.0417 %
Kappa statistic                          0.3228
Mean absolute error                      0.4271
Root mean squared error                  0.4854
Relative absolute error                127.9607 %
Root relative squared error            119.5081 %
Total Number of Instances               96

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.600    0.224    0.414      0.600    0.490      0.333    0.606     0.307     1
               0.776    0.400    0.881      0.776    0.825      0.333    0.606     0.807     0
Weighted Avg.  0.740    0.363    0.783      0.740    0.755      0.333    0.606     0.703

=== Confusion Matrix ===

  a  b   <-- classified as
 12  8 |  a = 1
 17 59 |  b = 0
```

Figure 3. Screen capture from WEKA analysis tool showing detail accuracy by class for the CAPTCHA -based IDS

4. 3.1 Comparison Analysis of Results obtained from the Malicious Email IDS and the Improved CAPTCHA- based IDS

Table 3 below shows the comparison analysis of the result obtained from both the existing the Improved CAPTCHA based IDS based on redirector model using Precision and detection Ratio of captured attacks.

Table 3: Percentage Of Weighted Average Of Naïve Bayes Classifier With Python Obtained From The Existing And Proposed System

Evaluation Metrics	Existing System (%)	Proposed System (%)	Performance differences (%)
Accuracy	75.0	82.0	7.0
Precision	57.0	67.0	10.0
Recall	76.0	82.0	6.0
F1-score	65.0	74.0	9.0

Table 4: Comparison Of Experimental Result Obtained From The Naïve Bayes Multinomial Classifier In WEKA

Evaluation Metrics	Existing System	Proposed System	Performance differences
TP Rate	0.57	0.74	0.17
FP Rate	0.58	0.36	0.22
Precision	0.53	0.78	0.25
Detection Rate	0.57	0.74	0.17
F- Measure	0.54	0.75	0.21

Table 5: Correctly And Incorrectly Classified Instances Obtained From The Naïve Bayes Multinomial Classifier.

Systems	Total Attacks	Correctly Classified Instances	Incorrectly Classified Instances
Existing System	49	57.0	42.9
Proposed System	96	74.0	26.0
Performance differences	47	17	16.9

4.4 Discussion of results

The precaution to protect the network is the duty of network administrator as pointed out by [26]. Network security infrastructure is one of the major concerns for network administrators, and detecting large number of attacks by intelligent spywares. Developing IDS model that can perform the task of intelligently monitoring network systems for an attack by spyware is important aspect of today's technological advancement, therefore the need for efficient and effective Intrusion detection system that can handle the detection and reporting of frequent attacks by not only humans but intelligent spywares. Thus, this section presents the discussion on the Detection Ratio and Precision by both the existing the proposed systems. Furthermore discussion on the differences obtained from detection Rate and correctly and incorrectly classified instances is another aspect that needs to be considered in the work.

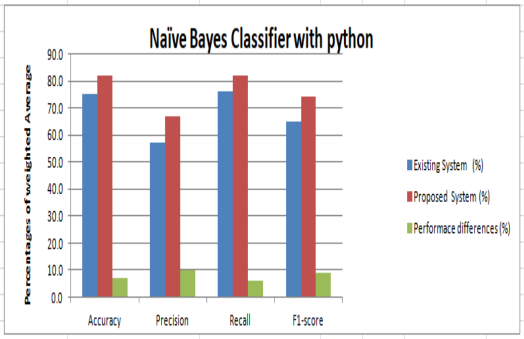


Figure 4. Graphical Representation Of Percentages Of Weighted Average Of The Evaluation Metrics

4. 4. 1 Detection Rate achieved on the Number of detected Attacks

As shown on the graph figure 4 above ,the confusion matrix parameter are used to help in detecting attacks that are indicated as attacks and are presented as attacks (TP), also the (FP) .This value represents incorrect instance where normal packets are classified as attacks. From the table 3 above indicates that the proposed system has a higher number of detected attacks up to 96 instances with high detection Rate of 0.74 (74%). In comparison to the existing system the proposed system has better and efficient detection rate than the existing system which is having total of 49 attacks detected with 0.57 (57%) detection rate and with a precision of 0.78(78%) for the proposed system and 0.53 (53%) precision for the existing system. Precision this is a percentage of attacks that are properly detected which is also term as Accuracy.

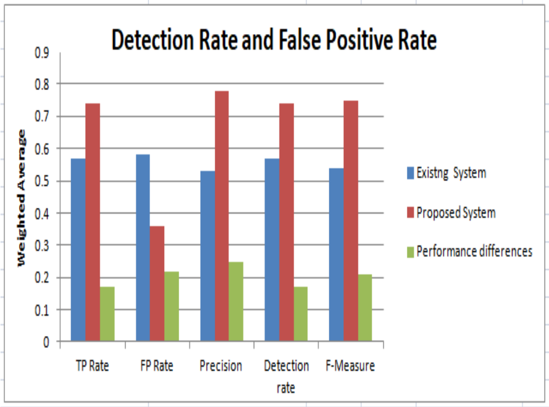


Figure 5: Graphical Representation Of Detection Rate And Precision Of The Two Systems.

Generally, we can say our approach achieved good result when the performance of the system was evaluated using the Naïve Bayes classifier of the machine learning analysis technique which is an important performance metric for measuring effectiveness of Intrusion detection system. As shown in the graph figure 4 above the proposed system after analysis indicates a better Accuracy of (82%) with a high correctly classified instance of 74% classification compared to the existing system which has Accuracy of (75%). This confirms that our work has the ability to detect intrusion by intelligent spyware effectively and efficiently.

4. 4. 2 Correctly and Incorrectly classified instances

For correctly classified instances means attacks that are detected as attack and are classified correctly as attack. While incorrectly classified instances signifies in inappropriate classification of attacks detected. This accuracy of detection is better achieved from the confusion matrix of the WEKA analysis tool. In comparison to the existing system the proposed system has better and efficiency of correctly classified instance of 74% and only 26% incorrectly classified instance , whereas the existing system has worse correctly classified instance of

57% and 43% incorrectly classified instances. This indicates that from the experimental result the Improved CAPTCHA based IDS which is based on redirector model performs better in terms of detecting actual attacks than the Existing system. Figure 6 below illustrates the graphical representation of the correctly and incorrectly classified instances obtained from both the existing and the Improved CAPTCHA- based IDS system.

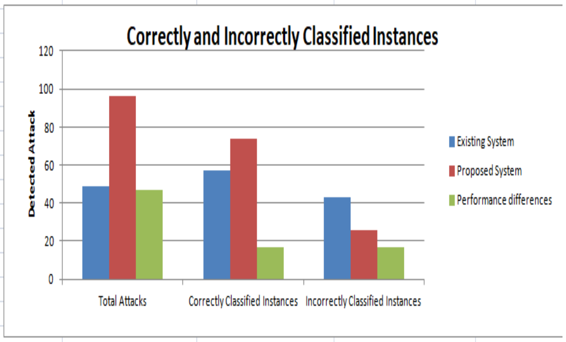


Figure 6.Graphical Representation Of Correctly And Incorrectly Classified Instances.

4. 4. 3 Remarkable differences obtained from Accuracy, Precision and recall

Table 3 presents the results of the analysis of the systems Accuracy, precision and recall also called Detection Rate. The result indicates that there are remarkable differences obtained from the Experimental result carried out on the attack detected which the Accuracy, precision recall values were obtained. The proposed system proved to have a better performance with higher differences between the Accuracy obtained. There is also a difference of 7% Accuracy and 10% difference Precision. This is because the existing system considers attacks coming into the system generally without considering the intelligence of the attack, whereas in the proposed system it consider the attack on the spyware which are classified as been unintelligent and intelligent capable of breaking and bypassing CAPTCHA. The percentage of the bots penetrating the existing system is more than that of the proposed system, which is achieved as a result of the redirecting that takes place for the intelligent spywares.

4. 4. 4 Remarkable differences obtained from Detection Rate and False Positive rate

Table 4 presents the results of the analysis of the systems Detection Rate and FP rate. The result indicates that there are remarkable differences obtained from the Experimental result carried out on the attack detected which the Detection Rate and FP rate values were obtained. The proposed system proved to have a better performance with higher differences between the detection rate and the Precision; in which difference of 0.17% of detection rate and 0.25% difference Precision. Furthermore, the proposed system indicates better performance than the previous system in terms of FP rate obtained by the two systems, the previous system has higher FP rate of 0.58 compared to the proposed system which has 0.36. Another consideration is the percentage of correctly classified instance obtained by the proposed system as compared to the previous system indicates the proposed system performs more efficiently than the previous in which from the analysis performed the proposed system works well for having 74% correctly classified instances while only having 26% incorrectly classified instances. Unlike the previous system which obtained 57% correctly classified instance and having larger value of 43% incorrectly classified instances.

4.5 Findings

Having analysed the data and discussed the result, the following ideas have been identified and presented as the findings of the study:

i. The study indicates a high percentage of Detection Rate and a very low False Positive rate as compared against the existing system.
ii. The study also obtained a remarkable deference which the proposed system proves to have a better performance on the accuracy obtained against the existing system.
iii. Furthermore, result of the analysis provides an evidence of CAPTCHA break claim by intelligent spywares as revealed by [27].

4.6 Remarkable difference of the existing system and the Proposed system

In the existing system IPS is included because it is a security tool that is used for preventing spywares from getting intrusion into a system. And one of the techniques it used is CAPTCHA, the CAPTCHA is distinguishing between human and spywares, human beings can read CAPTCHA and pass but spyware cannot read CAPTCHA and pass. With advance development in spyware, new sophisticated spywares are now designed in such a way that they can break CAPTCHA under IPS, in which they use two methods; (i) The Unwitting human labour and (ii) machine –based attack, here they look at the design at the background and use pattern recognition to read the CAPTCHA. Therefore the existing system with the use of CAPTCHA under IPS did not indicate how it can find out about analysis of intruder intent and the category of spyware. In comparison with the proposed system, the Improved CAPTCHA-based system outperforms the existing system whereby the proposed system developed a CAPTCHA BASED IDS, which CAPTCHA was integrated into IDS instead of an IPS. The system work in a way that an active CAPTCHA is presented in the gateway with a placeholder tag “DO NOT TYPE ANYTHING IN THIS TEXTBOX”. Human users were tasked to read, comprehend and abide by the instruction. While an intelligent bots which targets the system will solve the CAPTCHA and pass in with the zeal of bypassing the security, but unfortunately will be redirected to a dummy page where his information will be tracked and blocked instantly. Their aim also includes identifying those intelligent bots that attack our systems, study their behaviours for further research and production.

5. CONCLUSION

In the paper, an improved Intrusion Detection System named CAPTCHA-based IDS based on redirector model as a means of detecting spywares was designed. The system was an improvement on the CAPTCHA-based IDS purposely designed in order to solve the problem of web-based security. It was achieved by designing an email web based, a CAPTCHA –Trap IDS which was integrated in to the website and host online with a dumpy honeypot. The system was designed in order to strengthen the security of web-based application. As it was evident that intelligent spywares uses machine learning and unwitting human labour to break and bypasses CAPTCHA when used in IPS, detected attacks are redirected to honeypot and their IP addresses were captured and other information about them are collected this can be useful in strengthen our security techniques. The proposed system combines, IDS, CAPTCHA, IPS and honeypot to form a hybrid online security application. This research will go a long way in curbing the menace of sophisticated software that have the capability of infiltrating CAPTCHA when used in IPS and thereby saving many organisations and institutions from cyber-attacks of all forms through login intrusion. The research will also open room for subsequent researches on the use of CAPTCHA in IDS apart from its traditional use in IPS.

In the paper, Accuracy has been measured on each datasets. We evaluate the performance in terms of the percentage of weighted average with the Naïve Bayes classifier in python. Evaluation metrics such as Accuracy, Precision, Detection rate (Recall), and F-1 score were illustrated and discussed, we further evaluate the performance of the system with the Naïve Bayes Multinomial text classifier in WEKA and we successful obtained classification accuracy of the performance such as TP rate, FP rate, F-measure Precision ROC Area, with correctly and incorrectly classified instances of the detected attacks on the system.

The outcomes of our result have led to the following recommendations;

i. The system makes emphasis on detecting intelligent spywares capable of breaking through CAPTCHA with less consideration on the unintelligent spywares, hence it is recommended for subsequent research to consider studying this part.

ii. Further categorization of detected attacks based on their activities on the system can also improve the detection rate for the system.

iii. It was observed from our studies that CAPTCHA is obsolete, people don’t want it anymore, and hence a lot of improvement can still be done by the research community to effectively handle spywares that break into the web-based application without CAPTCHA so as to minimised false positive rate which means from the study some bots were found that neglect the CAPTCHA and are considered been genuine on the system producing a high number of false positive rate.

iv. However, this system faking software intruders to believe it is IPS may not be very efficient as some software intruders may attempt login ignoring the CAPTCHA and the textbox provided for it, this will re-direct the intruder to the login authentication and somehow this may have been an intrusion bypassing the fake IPS and the CAPTCHA-trap IDS. This is an instance where the system may theoretically be bypassed and subsequent researches on similar are recommended studying this gap.

6. FUTURE WORK

In future, it is important to make new researches to eliminate the deficiencies with CAPTCHA based IDS because it was observed by our model that, a genuine user can mistakenly fill in the CAPTCHA box, hence will be regarded as bot. furthermore, bots neglect the CAPTCHA and can be consider been genuine, that is False positive. When this happened the entire system will be penetrated and compromise. Therefore, the CAPTCHA can be made hidden from the user and more detailed analysis of the existing and upcoming problems with IDS can be performed with other suitable performance evaluation metrics.

REFERENCES:

[1] Boukari, S., & Abubakar, H. A CAPTCHA - Based Intrusion detection Model. *International Journal of Software Engineering & Applications (IJSEA), vol 9 Issue 1.*, (2018). pp. 29-40.

[2] Kansra, M., & Chadha, P. D. Cluster Based detection of Attack IDS using Data Mining. *International Journal of Engineering Development and Research (IJEDR), vol 4issue 3*, (2016). pp. 1082-1087.

[3] 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.

[4] 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.*

[5] Igbe, O., Darwish, I., & Saadawi, T. Distributed Network Intrusion Detection System: An Artificial Immune System Approach. *IEEE First Conference on Connected Health: Applications, Systems and Engineering Technologies.* (2016). IEEE Computer Society.

[6] Malav, S., Avinash, M. S., Satish, N. S., & Sandeep, S. c. Network Security Using IDS, IPS and Honeypot. *International Journal of Recent Research in Mathematics Computer Science and Information Technology. Vol 2issu 2.*, (2016). pp. 27-30.

[7] 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.

[8] 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).*

[9] 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.

[10] Aljawarneh, S., Aldwairi, M., & Yasin, M. B. Anomaly-based Intrusion Detection System through feature Selection Analysis and building hybrid efficient model. *Journal of Computational Science*, (2017). pp. 1-10.

[11] Verma, A., & Ranga, V. On Evaluation of Network Intrusion Detection Systems:Statistical Analysis of CIDDS-001 Dataset Using Machine Learning Techniques. *Pertanika Journal of Science & Technology. Vol 26 issue 3, Universiti Putra Malaysia Press.*, (2018). pp.1307 - 1332.

[12] Zhao, H., Li, M., Wu, T., & Yang, F. Evaluation of Supervised Machine Learning Techniques for Dynamic Malware Detection. *International Journal of Computational Intelligence Systems. vol. 11. Atlantis Press*, (2018). pp.1153- 1169.

[13] Kumar Reddy, K. M., Kumar, Y. S., Ranjit, N., & Sammeta, M. N. Log based Intrusion Detection Model. *International Journal of Engineering Technology Science and Research. Vol 5 issue 3,* (2018). *(IJETSR).*, pp. 1362-1365.

[14] Ogweno, K. L., Oteyo, O. E., & Henry, D. O. Honey Pot Intrusion Detection System. *International Journal of Engineering Inventions, vol 4 issue 5*, (2014). PP. 28-41.

[15] Baykara, M., & Daş, R. A Survey on Potential Applications of Honeypot Technology in Intrusion Detection Systems. *International Journal of Computer Networks and Applications (IJCNA), vol 2 issue 5,* (2015). *EverScience Publications*, pp. 203-211.

[16] Vidya, P. N., & Naika, S. C. (2015). Simple Text based CAPTCHA for the security of Web Applications. *International Journal of Computer Science and Mobile Computing, vol 4 issue 4 :*, 519-531.

[17] Ali, H. M., Zolkipli, M. F., & Mohammed, M. A. (2016). Optimize Machine Learning Based Intrusion Detection for Cloud Computing: Review Paper. *Journal of*

Engineering and Applied Sciences, vol 11 No 2 Medwel Journal, pp. 3254- 3264.

[18] Joshi, V., & Kakkar, P. (2017). Honeypot Based Intrusion Detection System with Snooping agents and Hash Tags. *International Journal of Computer Science and Information Technologies, (IJCSIT), vol 8 issue 2*, 237-242.

[19] Umer, M. F., Sher, M., & Bi, Y. (2018). A two- stage- flow-based Intrusion Detection Model for Next Generation Networks, vol 13 issue 1. *PLoS ONE*, pp. 1-20.

[20] Ugochukwu, C. J., & Bennett, O. E. (2018). An Intrusion Detection System using Machine Learning Algorithm. *International Journal of Computer Science and Mechanical theory, vol 4 issue 1 IIARD*, pp. 39-47.

[21] Hussein, S. M. (2016). Performance Evaluation of Intrusion Detection System using Anomaly and Signature based algorithms to Reduce False Alarm Rate and Detect Unknown Attacks. *International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1064-1069). IEEE, Confernce Publishing Services CPS.

[22] Shewale, V. R., & Patil, H. D. Performance Evaluation of Attack Detection Algorithms using Improved Hybrid IDS with online Captured Data. *International Journal of Computer Applications, vol 146 issue 8*, (2016). pp. 35 - 40.

[23] Milan, Sardana, H., & Singh, K. (2018). Reducing False positive Alarms in Intrusion Detection Systems - A Survey. *International Research Journal of Engineering and Technology (IRJET); Impact factor value : 6.171*, pp. 9- 12.

[24] Georgina, O. N., Isah, A., & Alhassan, J. Analytical Study of some selected Classification Algorithms in WEKA using Real Crime Data. *International Journal of Advanced Research in Artficial Intelligence, (IJARAI). Vol 4 issue 2*, (2015).pp. 44-48.

[25] Ashraf, N., Ahmad, W., & Ashraf, R. A Comparative Study of Data Mining Algorithms for High Detection Rate in Intrusion Detection System. *Annals of Emerging Technologies in Computing (AETiC). Vol 2 No1*, (2018). pp. 49-57.

[26] Iskandar, A., Virma, E., & Ahmar, A. S. Implementing DMZ in Improving Network Security of Web Testing in STMIK AKBA. *International Journal of Engineering & Technology. Vol 7 issue 2.*, (2018). pp. 99-104.

[27] Sano, S., Otusko, T., Itoyama, K., & Okuno, H. G. HMM- based Attacks on Google's ReCAPTCHA with Continous Visual and audio symbol. *International Journal of Imformation Processing, vol 23 No 6 (2015)*, Pp. 814-826.