Volume 4, Issue II, February, 2025, No. 59, pp. 755-767
Submitted 2/2/2025; Final peer review 30/2/2025
Online Publication 3/3/2025
Available Online at http://www.ijortacs.com

SMART CYBER THREAT DETECTION AND MITIGATION USING DEEP PACKET INSPECTION AND DECEPTION BASED MACHINE LEARNING TECHNIQUE

By

OgbonnayaIrukwu Joe[1], Kufre M. Udofia[2], Akaninyene Bernard Obot[3].

Department of Electrical and Electronic Engineering, University of Uyo, Uyo, Akwaibom State[1,2,3]

Corresponding Author's Email and Tel: irukwuogbonnaya@gmail.com[1], [1]08036766845

Abstract

The increasing complexity of cyber threats has made traditional network defence mechanisms inadequate; especially in smart environments handling high-volume data traffic. Modern-day attackers utilize advanced tactics, exploiting vulnerabilities within networks and systems, often bypassing conventional defences. The aim of this study is modelling of smart cyber threat detection and mitigation using deep packet inspection and deception based machine learning technique. To achieve this, data was collected from Silexsecure limited, Alibaba and Kaggle repository considering six attack classes which are brute force, benign, distributed denial of service, Structured Queried Language (SQL) injection attack, and normal packet. Three machine learning algorithms which are Support Vector Machine (SVM), Decision Tree (DT) and Artificial Neural Network (ANN), were selected and trained to generate three Deep Packet Inspection (DPI) models, using Matlab programming language. Comparative analysis was performed on the models with recorded accuracy of 91.8% for DT, 89.9% for ANN and 81% for SVM. Upon selection of the DT based DPI model as the best, a honeypot-based deception security model was selected and then integrated with the DPI as a smart deception security model using Python programming language. Several simulation experiments were performed to demonstrate the effectiveness of the model and results showed its reliability in security network infrastructures against selected online threats. The model was implemented to secure an online network infrastructure employed by users for E-commerce activities using Javascript and Python programming language. The results when tested with legitimate packet, successfully allow the user access to the main server, however when tested with SQL injection attack allowed the user access to a decoy facility where the threat information were collected at the back-end and for threat intelligence analysis.

Keywords: Cyber Threat, Deep Packet Inspection, Threat Detection, Deception-Based Security Model, Threat Mitigation, Machine Learning

1. INTRODUCTION

Generally, reliability of a network infrastructure involved collective components of integrity, availability and confidentiality [1,2,3]. Over the years, several Deep Packet Inspection (DPI) models have been developed considering both machine and deep learning models as some of the most recent techniques in literature applied and proposed to solve network security problem [4,5,6]. Another innovative approach is the application of deception technique which is aimed at diverting the attacker's attention from the

main network facility [7,8,9]. However, both techniques have their limitations when deployed as standalone security solution, but integration of the two concepts presents a strong force for network security solution.

Accordingly, this study presents a hybrid approach which merged the DPI and deception solution to provide an enhanced security solution against several threats on network facility [10,11]. This was achieving through the adoption of a suitable data model that characterized network threats and then after process was applied to train three selected machine learning algorithms to present three DPI models which were comparatively analyzed to identify the most suitable for the proposed hybrid solution. More so, a deception solution of honey-web was adopted and then integrated with the ML based DPI model to presented a comprehensive network defence system using Python and JavaScript [11]. Several experiments were carried out to assess the model performance considering several threats scenarios and also legitimate users. The solution performance is evaluated to demonstrate the ability of the security solution to offer reliable network security based on a case study infrastructure.

2. Data collection, data description and data pre-processing

The study utilized both primary and secondary data. The primary dataset was collected from Silexsecure limited, Alibaba and Kaggle repository considering six attack classes which are brute force, benign, Distributive Denial of Service (DDoS), Structured Queried Language (SQL) injection attack, and normal packet. Three machine learning algorithms which are Support Vector Machine (SVM), Decision Tree (DT) and Artificial Neural Network (ANN), were selected and trained to generate three Deep Packet Inspection (DPI) models, using Matlab programming language. The secondary is the Software Defined Network (SDN) intrusion dataset from the Kaggle repository which is an online network infrastructure. Other data used for the research were collected from interdisciplinary domain experts in cyber security, data science, telecommunication engineering, and data analysis. The data was processed using imputation technique which by fixing duplicate and missing values automatically based on mean imputation, then feature extraction was performed using Analysis of Variance (ANOVA).

The SDN intrusion dataset is made up of 79 attributes of SDN threat and grouped into four different attack classes of DDoS, Cross-Site Scripting (XSS) intrusions, brute force, SQL injections, normal packet and benign traffic. The sample size of the data is 1,188,333 rows of observations related to network intrusions and white-listed traffic. The breakdown of observations for each type of traffic is presented in Table 1

Table 1: Attack class and number of observations

Attack class	Attack feature size
Benign Traffic	798,322
Web Attack XSS Traffic	1,962
Web Attack SQL Injection Traffic	60
Web Attack Brute Force	4,550
DDoS Traffic	338,139
Normal packet	45,345

2.2 The Deep Packet Inspection and Deception Solution

The focus in this study is on cyber threat detection and mitigation on an online infrastructure using a combination of machine learning and software-defined network-based honeyweb (SDN-honeyweb) approach. The process flow diagram for the cyber threat detection and mitigation solution is given in Figure 1.

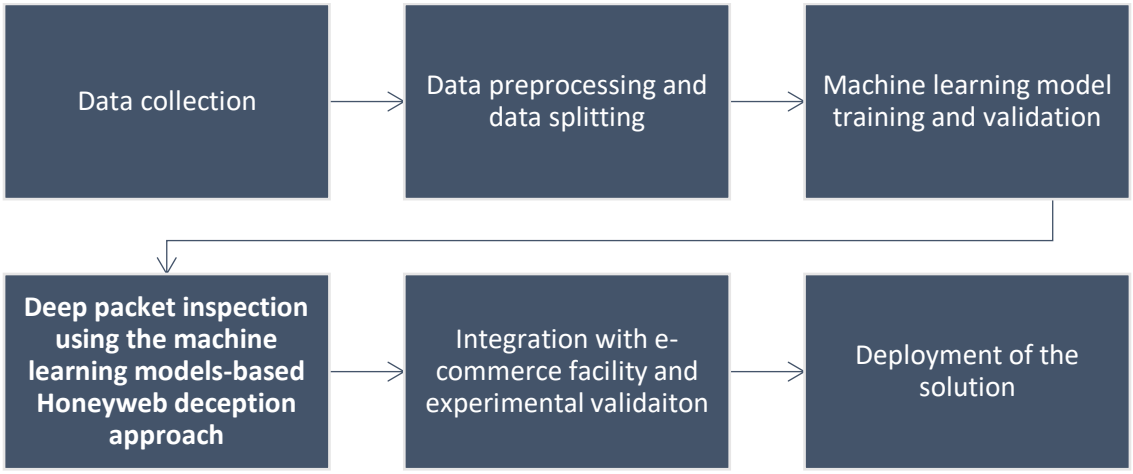

Figure 1: The process flow diagram for the cyber threat detection and mitigation solution

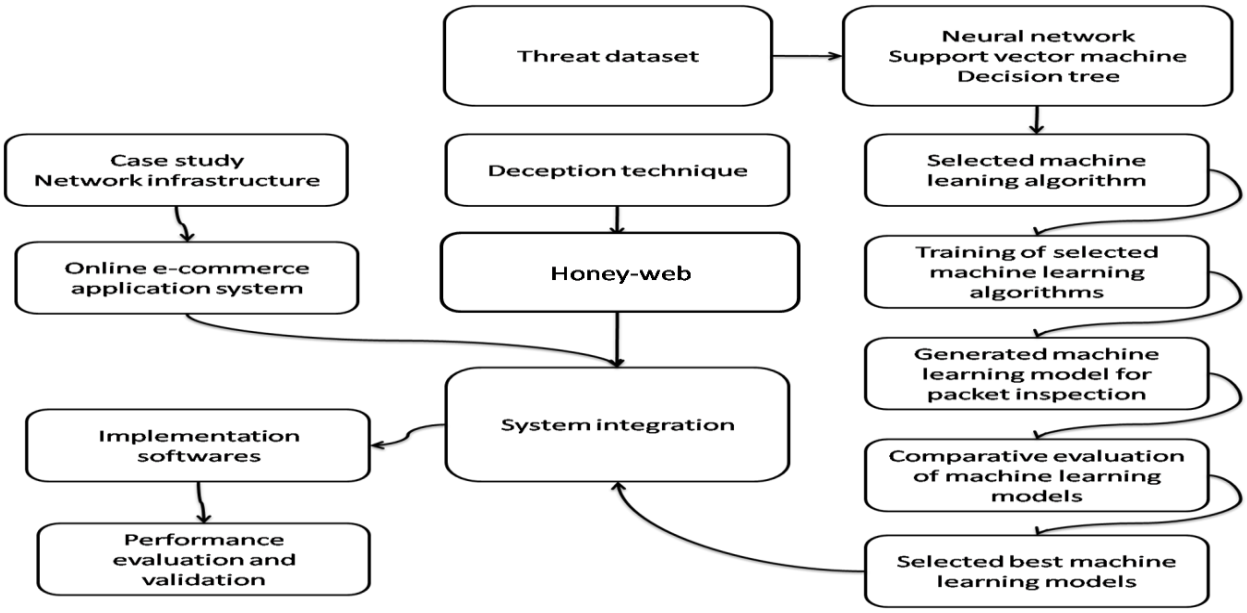

Figure 2: Mind mapping diagram of the solution procedure

According to Figure 1, the cyber threat detection and mitigation solution began with data collection of network threat information to curate the problem dataset. Next, is the data pre-processing and data splitting followed by the training of the three selected machine learning (ML) algorithms which included Support Vector Machine

(SVM), Decision Tree (DT) and Artificial Neural Network (ANN). The cyber threat detection ability of the ML models are compared using well-defined success metrics and the best model is selected for integration into the packet inspection model and merged with a deception facility that is based on the Honey-web model. This is applied as a smart deception framework for protection of online network infrastructure. Real-time controlled tests, taking into account the several threats such as SDN injection attack are used to validate the intelligent threat detection system. The smart deception solution is then integrated with e-commerce facility and tested under threats condition to validate the work practically. The mind mapping diagram of the solution procedure is presented in Figure 2.

2.3 The Deep Packet Inspection Models (DPIM) and their Performance Evaluation Metrics

In this work, three DPIMs were generated considering an artificial feed-forward neural network, a support vector machine, and a decision tree. The DPIMs are developed with the aim of conducting a comparative assessment of the models to identify the best model which is then used for the network security applications. In the study, the DPIM is used as the monitoring mechanism, which inspects the features of packets penetrating towards the online network infrastructure. The DPIM extracts the features of the packet and then classifies the feature patterns with the trained model of the SDN attack. When threats or malicious features are detected, the packet is flagged as a threat, and the user is identified as an intruder. The parameters utilized for the DPI model assessment are positive predictive value, false discovery rate, accuracy, true positive rate and area under curve.

i). Positive Predictive Value (PPV): The PPV quantifies true positive predictions and it is computed as:

$$PPV = TP / (TP + FP) \quad (1)$$

Where: TP: True Positives; FP: False Positives

ii). False Discovery Rate (FDR): FDR quantifies false positive predictions and it is computed as:

$$FDR = FP / (TP + FP) \quad (2)$$

Where: FP: False Positives

iii). Accuracy: Accuracy measures the overall correct predictions and it is computed as

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

Where: TP: True Positives; TN: True Negatives; FP: False Positives; FN: False Negatives

iv). True Positive Rate (TPR): TPR calculates the proportion of actual positives correctly predicted and it is computed as:

$$TPR = TP / (TP + FN) \quad (4)$$

v). False Negative Rate (FNR): FNR represents the proportion of actual positives

that were incorrectly predicted as negativesand it is computed as:

$$FNR = FN / (TP + FN) \quad (5)$$

2.4 System Integration of the security model on the network infrastructure

The architecture of the system integration is presented in Figure 3. Incoming packet generated by the user are transferred to the network through the SDN, at this time, the deep packet inspection model monitors for the packet features through real time classification process, then the decision algorithm allows access to the normal server for classified normal users and then the decoy server for the classified threat suspects. The use case architecture of the integrated system operation is presented in Figure 4which shows the use case demonstration of threat from attacker and then the use case of the normal packet transfer from normal user.

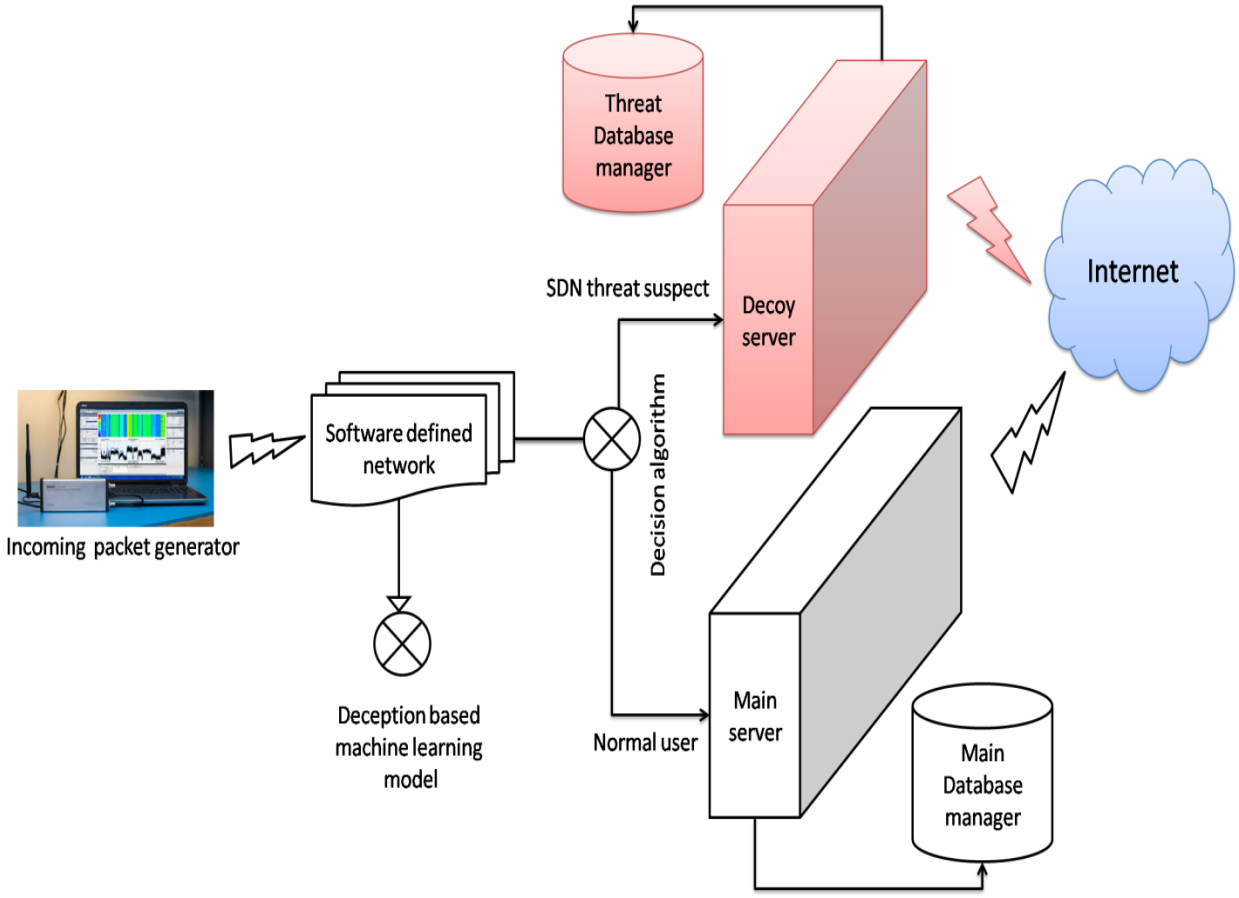

Figure 3: Architecture of the integrated security model

Specifically, in Figure 4, the architecture of the secured network with deep packet inspection and SDN honeyweb was tested with SDN threat from an attacker and then normal packet from another user. The incoming packets generated were processed through feature identification and packet inspection to classify the status of the

packet, and when malicious features were detected due to the SDN attack penetration from the attacker use case, the decision boundary allowed the user access to the decoy server, while if the packets were classified as normal, the user was allowed access to the main server for the e-commerce activity.

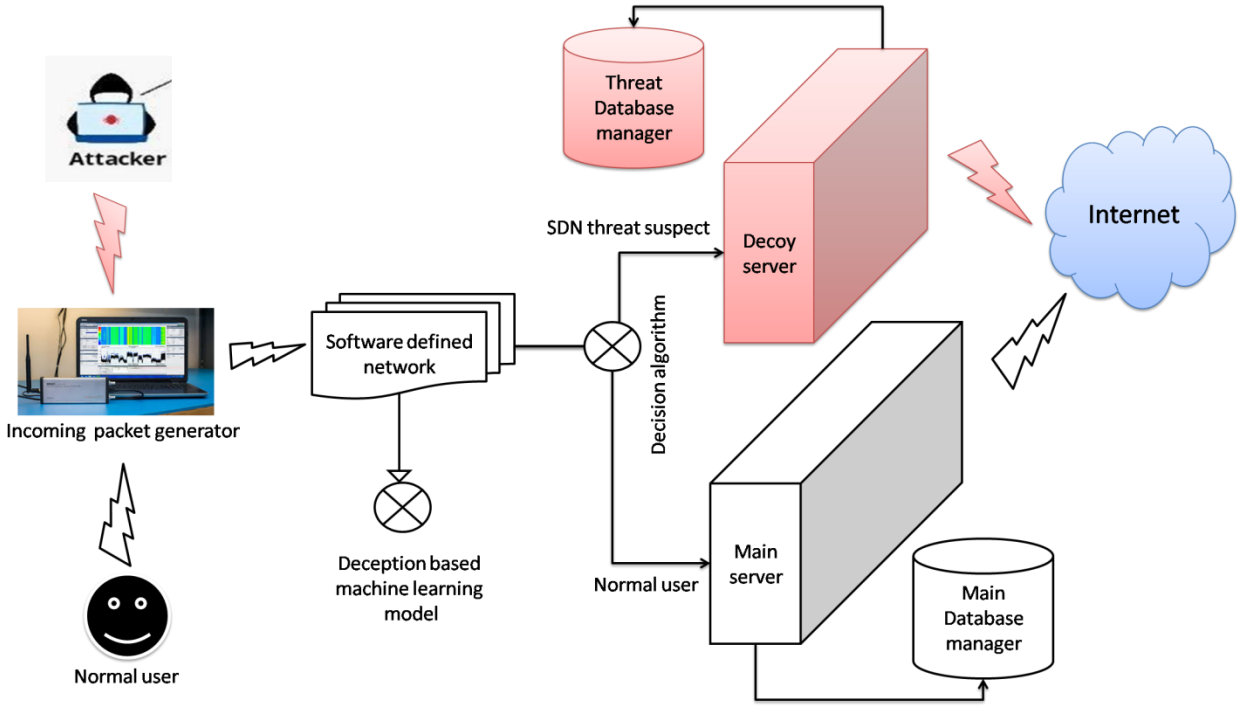

Figure 4: Use case architecture of the integrated system operation

2.5 System implementation

The implementation of this work was done with four software namely; Matlab, Python MySQL and Javascript. Matlab was applied to implement the training of the three machine learning algorithms as a deep packet inspection models. This was done using classification learner application software. The trained and validated models were used as building blocks to develop the smart deception solution which is DPI-based-SDN-Honey-web solution for network security using Python programming language and MySQL.

The python software was used in creating a honeypot deception model that simulates the diversion of attackers and legitimate users between a decoy server and the main (genuine) server. Using randomly generated IP addresses, the model assigns attackers different attack types such as SQL Injection, DDoS, XSS, Benign, legitimate, and Brute force, each with varying probabilities of being diverted to a decoy server. Legitimate users also face a small chance of false diversion. The implementation includes real-time detection times and visualizes the diversion process through dynamic bar plots, allowing for an analysis of how efficiently attackers are redirected away from the main

system while legitimate users access it. This approach, executed in Google Colab, demonstrated the model's workability and adaptability in handling diverse cyber attack scenarios.The smart deception model was then integrated to protect an online network infrastructure used for E-commerce activities using Javascript and MySQL programming language. The configuration parameters used for this work are reported in Table 2.

Parameter	Value
Time	2024-09-22 12:00:00
Main Server IP	192.168.1.10
Decoy Server IP	192.168.1.20
Attacker Type 1 IP	203.0.113.1
Attacker Type 1 Method	SQL Injection
Number of Attacker Type 1 Users	5
Attacker Type 2 IP	203.0.113.2
Attacker Type 2 Method	DDoS
Number of Attacker Type 2 Users	3
Attacker Type 3 IP	203.0.113.3
Attacker Type 3 Method	Brute force
Number of Attacker Type 3 Users	5
Attacker Type 4 Method	Benign
Number of Attacker Type 4 Users	4
Attacker Type 5 Method	XSS attack
Number of Attacker Type 5 Users	4
Legitimate User 1 IP	198.51.100.1
Legitimate User 1 Activity	Accessing Server
Legitimate User 2 IP	198.51.100.2
Legitimate User 2 Activity	Accessing Server
Legitimate User 3 IP	198.51.100.3
Legitimate User 3 Activity	Accessing Server
Number of Normal Users	10
Honeypot Trigger Time	2024-09-22 12:01:00
Attack Detection Time	2024-09-22 12:02:00
Redirection to Decoy Time	2024-09-22 12:03:00
Incident Response Time	2024-09-22 12:05:00

Table 2: Simulation parameters

3. RESULTS AND DISCUSSIONS

The results of the performance of the DPI model-based SDN-honeyweb on legitimate users are shown in Figure 5. The bar chart in Figure 5 showed the connection of legitimate users with their IP address to the main server to perform online activities stating the number of attempts the user tries to connect to the main server. The results showed that the honeypot was able to connect the users to the main server at every connection attempt.

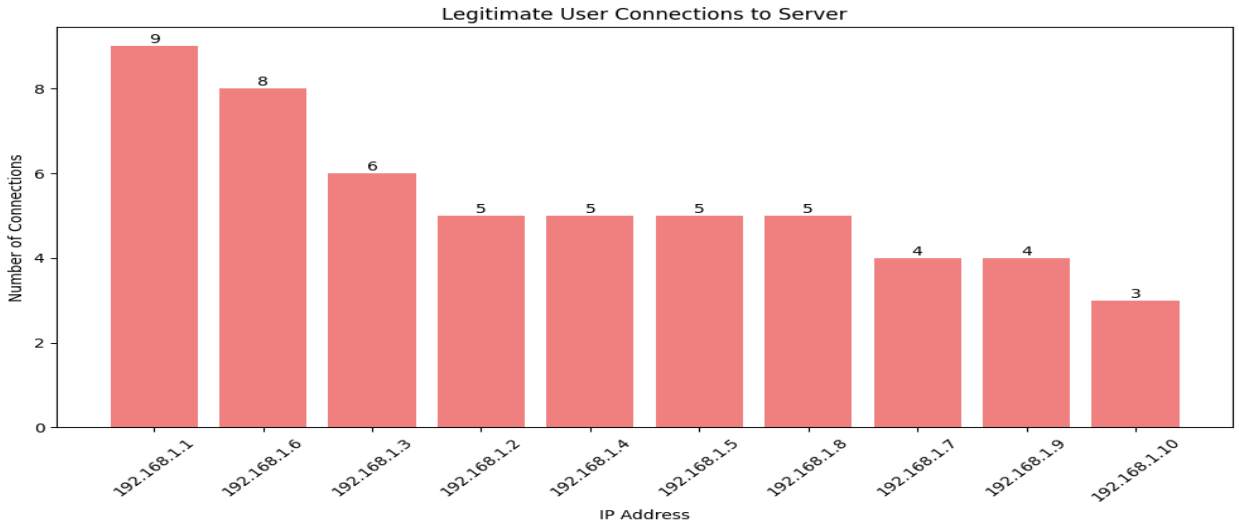

Figure 5: Legitimate user connection to server

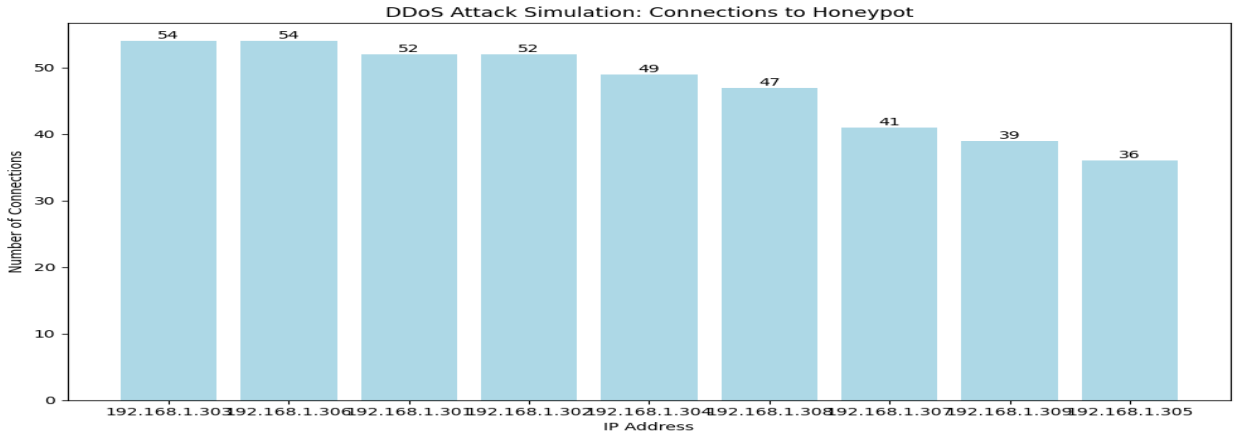

Figure 6: DPI based SDN-honeyweb against DDoS attack

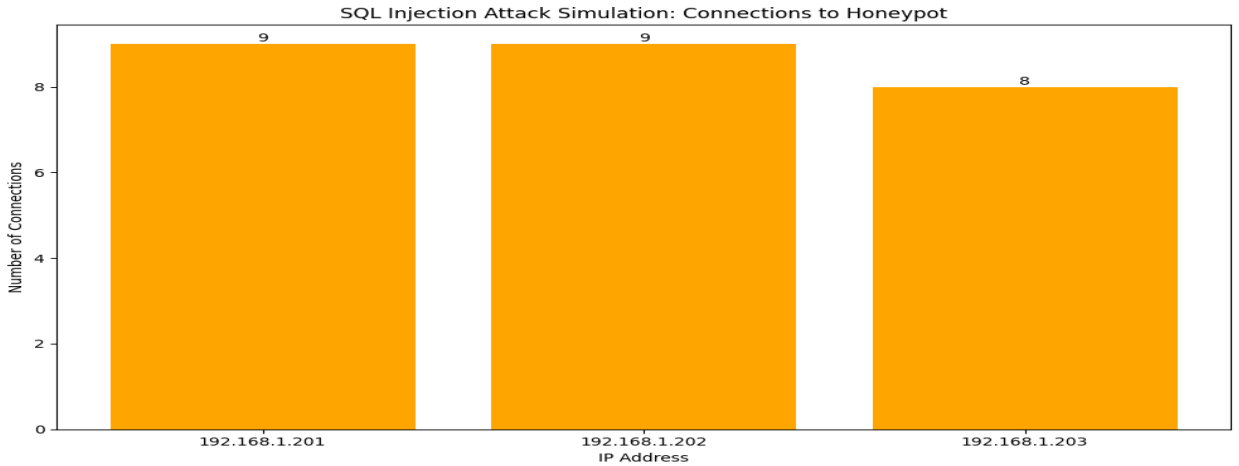

Figure 7: DPI-SDN-honeyweb with SQL injection attack

The result for users with DDoS attack are shown in Figure 6. The bar chart in Figure 6 showsthe performance of the SDN-honeyweb with themachine learning-based DPI models while considering DDoS attack. From the result, it was observed that 9 individual attackers were used to try DDoD attack with each flooding the server with over 35 flood IP threats respectively. The results showed that the honeypot was able to correctly detection them and connect to honeypot server which is the deception facility to divert and monitor threat information.

The Cross-Site Scripting (XSS) attack was evaluated with the SDN-honeyweb and DPI technique and theresult are shown in Figure 8. The DPI based SDN-honeyweb was able to connect the XSS attackers to the honeypot server where their information will be

The SQL injection attack was evaluated with the SDN-honeyweb and DPI technique and theresult are shown in Figure 7. From the results, it was observed that the SDN-honeyweb with the ML DPI model was able to correctly classify three users sending SQL attack to the network. The results showed that each of the user injected over 7 SQL attacks to the network and the honeypot was able to detect and allow the attacker access to the connection to honeypot server collected and used to improve the main network facility. Again, the Brute force attack was evaluated with the SDN-honeyweb and DPI technique and theresult are shown in Figure 9.

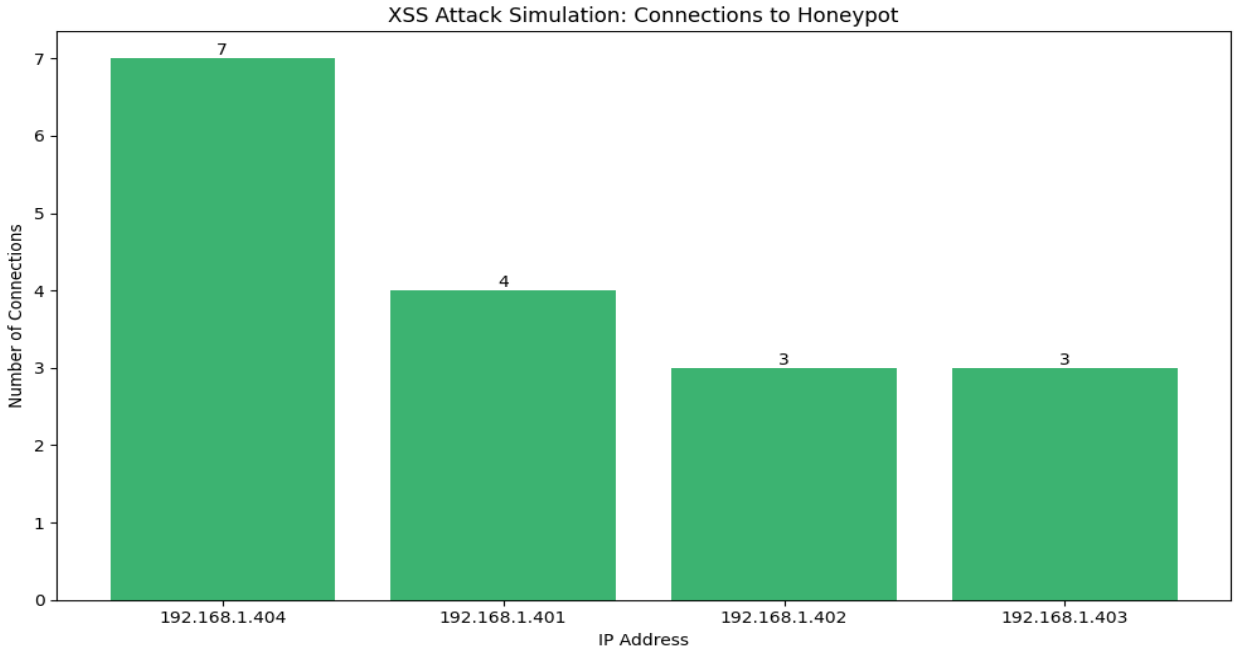

Figure 8: DPI based SDN-honeyweb with XSS attack

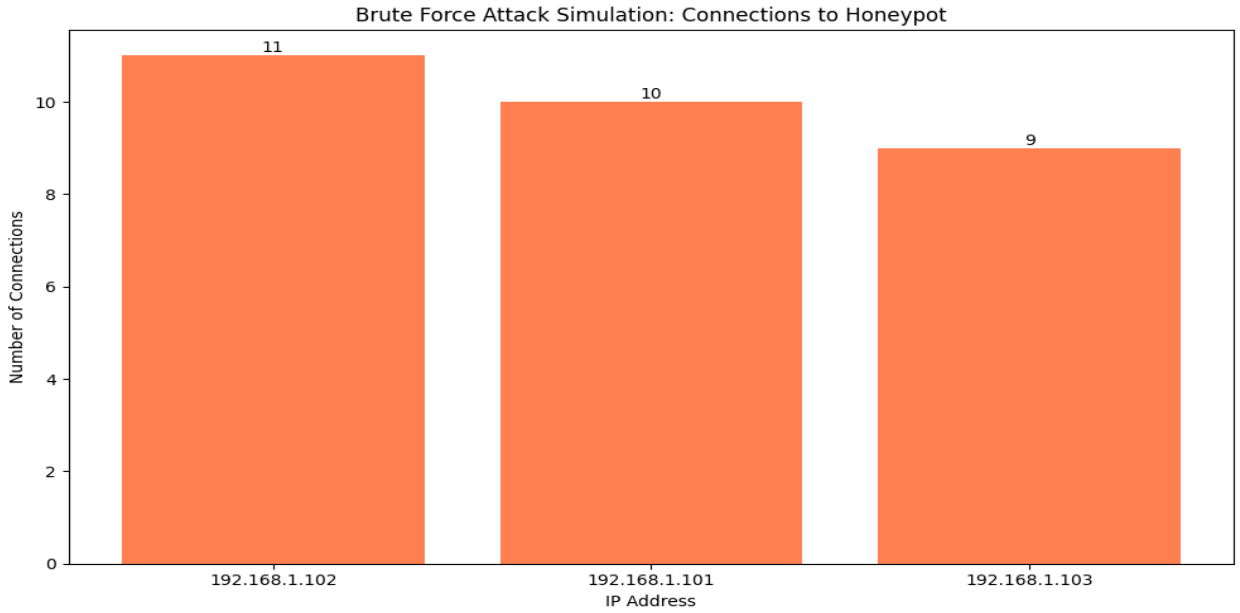

Figure 9: DPI-SDN-honeyweb with Brute force attack

The results presented in Figure 10 is a comprehensive reported of the honeypot server information showing number of connection attempts by illegitimate users trying to connect to the network and how the SDN-honeyweb diverted the attackers to the honeypot server. The result also showed the number of connections attempts from each IP address of attackers, and overall they were successfully connected to the honey-pot server. The results in Figure 11 reported the diverted legitimate user to the main server.

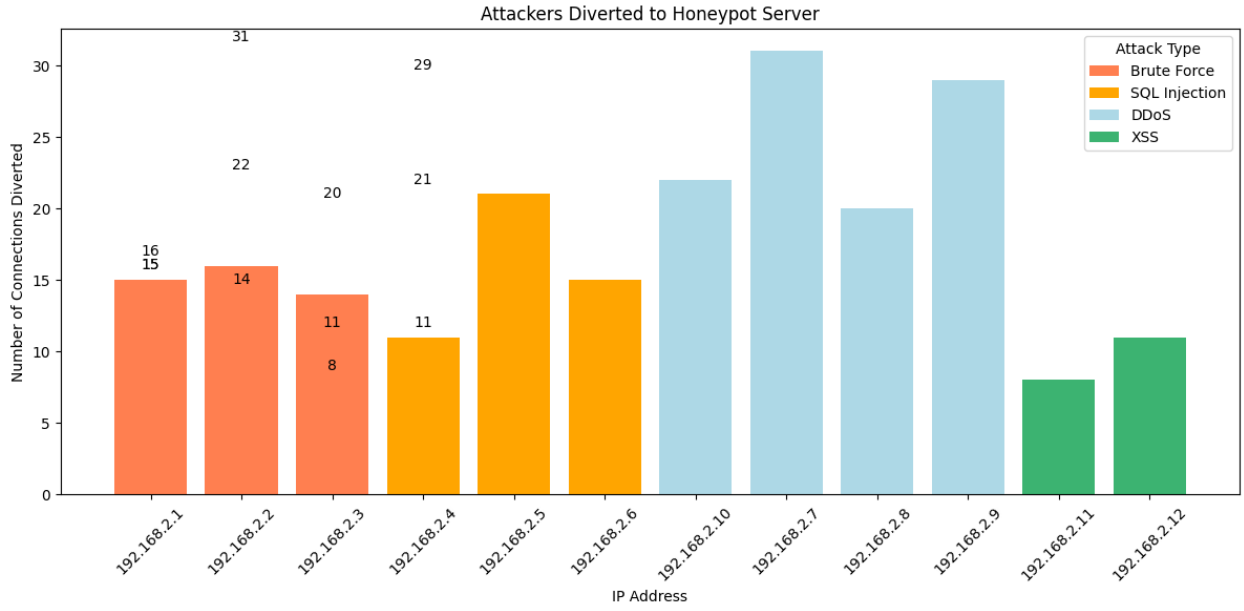

Figure 10: Attackers diverted to the honeypot server with DPI- Honeyweb

```
[*] Listening on 127.0.0.1:8889
Legitimate Client from 127.0.0.1 trying to connect...
[!] Connection from ('127.0.0.1', 55324)
[!] Legitimate User ('127.0.0.1', 55324) connected to the real server.
Client received: Welcome to the REAL server!

Legitimate Client from 127.0.0.1 trying to connect...
[!] Connection from ('127.0.0.1', 44550)
[!] Legitimate User ('127.0.0.1', 44550) connected to the real server.
Client received: Welcome to the REAL server!

Legitimate Client from 127.0.0.1 trying to connect...
[!] Connection from ('127.0.0.1', 44566)
[!] Legitimate User ('127.0.0.1', 44566) connected to the real server.
Client received: Welcome to the REAL server!
```

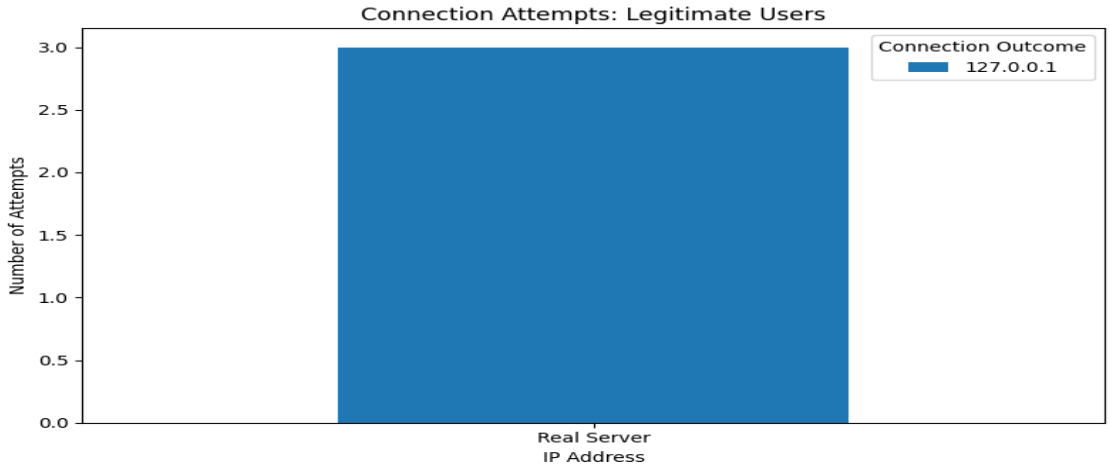

Figure 11: Classification of normal user and diversion to main server

The comparative analysis of the models and their performance against benign and DDoS attack are presented in Table 3.It can be seen the DT model has the best accuracy of 91.8 % for both the benign and DDoS attack while SVM has the worst accuracy with a value of 89.9 %. Similar trends are observed in the accuracy of the DT model in the other attack categories; the DT maintained highest accuracy in the detection of attacks and benign users. As such, the DT model based DPI model is adopted.

Table 3: Comparison of the results of the machine learning-based DPI performance on Benign and DDoS Attack

Metrics	**ANN-Benign**	**ANN-DDoS**	**DT-Benign**	**DT-DDoS**	**SVM-Benign**	**SVM-DDoS**
Positive Predictive Value (PPV):	96.7	96.1	99.9	97.9	85.7	80.7
False Discovery Rate (FDR):	3.3	4.9	0.1	2.1	14.3	19.3
True Positive Rate (TPR)	97.2	93.5	98.6	99.9	88.9	64.8
False Negative Rate (FNR)	2.8	6.5	1.4	0.1	11.2	36.2
Accuracy (ACC)	89.9	89.9	91.8	91.8	89.9	89.9

4. CONCLUSION

An approach for cyber threat detection and mitigation on an online infrastructure using a combination of machine learning and software-defined network-based honeyweb (SDN-honeyweb) is presented. The solution was tailored for an e-commerce platform. The SDN-honeyweb has machine learning-based mechanism to

detect threat and benign traffic and hence divert the attention of the attacker from the main online network infrastructure to a fake network online facility using honeypots. Once the attacker is detected and diverted to the fake online platform, the system then collect threat intelligence from both the front and back ends of the online infrastructure. On the front end, the malicious threat data is collected and applied to divert the attacker to a fake server; on the back end, the threat information is collected from the virtual server and used for threat intelligence.

The three selected machine learning (ML) algorithms includes Support Vector Machine (SVM), Decision Tree (DT) and Artificial Neural Network (ANN). The three models were trained using data that was collected from Silexscure limited, Alibaba and Kaggle repository. In the dataset, six attack classes were considered and they include brute force, benign, distributed denial of service, Structured Queried Language (SQL) injection attack, and normal packet. In all, the DT model performed best among the three ML models in detecting and clarifying the traffic as benign or threat.

REFERENCES

1. Adelani, F. A., Okafor, E. S., Jacks, B. S., &Ajala, O. A. (2024). Theoretical insights into securing remote monitoring systems in water distribution networks: lessons learned from Africa-US projects. *Engineering Science & Technology Journal*, *5*(3), 995-1007.
2. Jha, R. K. (2023). Cybersecurity and confidentiality in smart grid for enhancing sustainability and reliability. *Recent Research Reviews Journal*, *2*(2), 215-241.
3. Desai, B., Patil, K., Mehta, I., & Patil, A. (2024). A Secure Communication Framework for Smart City Infrastructure Leveraging Encryption, Intrusion Detection, and Blockchain Technology. *Advances in Computer Sciences*, *7*(1).
4. Deri, L., & Fusco, F. (2021, July). Using deep packet inspection in cybertraffic analysis. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)* (pp. 89-94). IEEE.
5. Song, W., Beshley, M., Przystupa, K., Beshley, H., Kochan, O., Pryslupskyi, A., ... & Su, J. (2020). A software deep packet inspection system for network traffic analysis and anomaly detection. *Sensors*, *20*(6), 1637.
6. Shhadih, M. A. (2023). *Cyber Deception Techniques and an Adversary Engagement Platform for Cybersecurity Enhancement* (Doctoral dissertation, The George Washington University).
7. Al Amin, M. A. R., Shetty, S., Njilla, L., Tosh, D. K., &Kamhoua, C. (2021). Hidden markov model and cyber deception for the prevention of adversarial lateral movement. *IEEE Access*, *9*, 49662-49682.
8. Mohan, P. V., Dixit, S., Gyaneshwar, A., Chadha, U., Srinivasan, K., &Seo, J. T. (2022). Leveraging computational intelligence techniques for defensive deception: a

review, recent advances, open problems and future directions. *Sensors*, *22*(6), 2194.

9. Sowndeswari, S., Kavitha, E., & Krishnamoorthy, R. Enhancing security in wireless sensor networks: A fusion of deep learning and energy-efficient routing. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-16.
10. Mutambik, I. (2024). Enhancing IoT Security Using GA-HDLAD: A Hybrid Deep Learning Approach for Anomaly Detection. *Applied Sciences*, *14*(21), 9848.
11. Shahid, W. B., Aslam, B., Abbas, H., Afzal, H., & Khalid, S. B. (2022). A deep learning assisted personalized deception system for countering web application attacks. *Journal of Information Security and Applications*, *67*, 103169.