IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

KNN based Identification of Denial-of-Service Attack in Software Defined Network

Bhartesh Goyal[1], Gursewak Singh[2]

Department of Computer Engineering, Baba Farid College of Engineering and Technology, Bathinda, Punjab, India

ABSTRACT: Utilizing their capacity to identify patterns and abnormalities in real-time data, machine learning algorithms significantly contribute to the suppression of attacks. Machine learning offers a more dynamic and responsive defence mechanism than static techniques in intrusion detection because of its ability to adjust to shifting threat landscapes. The incorporation of machine learning technology is anticipated to facilitate the development of increasingly proficient and versatile intrusion detection systems as the field progresses. Given the recent spike in cybersecurity concerns and the necessity of defence against a variety of assaults, keeping an extensive and up-to-date knowledge base is imperative. The main source of these worries is the exponential expansion of computer networks and the pervasive use of related applications by both individuals and organisations, particularly with the growing uptake of the Internet of Things (IoT). The results of the experiments demonstrate the good outcomes for three distinct methods under various variations for three distinct datasets.

KEYWORDS: KNN, DDoS, SDN

1. INTRODUCTION

In various domains, machine learning plays a crucial role, particularly in applications such as fraud detection, computer-aided diagnosis, bioinformatics, medical diagnosis, and network attack detection. Machine learning is extensively employed in the realm of network attack detection, where the primary objective is to identify potential attacks early on to safeguard network resources. Some attacks demand significant management attention due to their suitability and severity.

The detection of network attacks has emerged as a major contributor to the erosion of privacy by malicious actors. Consequently, it is imperative to proactively predict and monitor these attacks at their nascent stages to mitigate potential harm.

1.1 Cyber Attacks

1.1.2 DoS and DDoS attacks

A denial-of-service (DoS) attack aims to overwhelm a system's resources, rendering it incapable of responding to legitimate service requests. Similarly, Distributed Denial of Service (DDoS) attacks seek to exhaust system resources, but they are distinct as they originate from multiple host computers infected with attacker-controlled malware. These attacks are aptly named "denial of service" attacks since the targeted website becomes unable to cater to users attempting to access it. In a DoS attack, the targeted page is inundated with malformed requests, and each response consumes resources as the site must address each request. This hinders the website from functioning normally and often leads to a complete shutdown.

1.1.3 MITM Attack

A man-in-the-middle (MITM) cyberattack is a type of cybersecurity breach where an attacker intercepts and eavesdrops on the data exchanged between two parties, networks, or computers. The term "man-in-the-middle" stems from the attacker being positioned between the two communicating parties. During a MITM attack, the two parties involved appear to be communicating normally, unaware that the transmitted messages have been tampered with or accessed by the attacker before reaching their intended destination.

1.1.4 Phishing attack

A phishing attack occurs when a malicious actor sends an email that mimics the appearance of a message from a trusted and legitimate source, aiming to deceive the recipient into divulging sensitive information. This type of attack

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

combines elements of social engineering and technology. The term "phishing" is derived from the idea that the attacker is "fishing" for sensitive information by posing as a trustworthy entity.

II. LITERATURE REVIEW

Different measures can be considered for the early stage of Cyber Attack prediction S. Shanthi the referenced section [1] provides a review of existing research focused on the detection and classification of malware and malicious code, utilizing text analysis and data mining techniques. Notably, Data Mining technology proves versatile in the context of Malware detection. Suh-Lee et al. (2016) contributed to the field by detecting security threats through a combination of data mining, text classification, natural language processing, and machine learning. Their approach involved extracting relevant information from unstructured log messages to enhance security measures. Kakavand et al. (2015) proposed an anomaly detection model, known as Text Mining-based Anomaly Detection (TMAD). This model specifically targets the detection of HTTP attacks on network traffic. TMAD utilizes n-gram text classification and the Term Frequency-Inverse Document Frequency (TF-IDF) method to achieve its objectives. Norouzi et al. (2016) introduced various classification methods for the detection of malware programs. Their approach focuses on classifying malware based on the functionality and behavior exhibited by each program. Fan et al. (2015) explored the application of hook technology to track dynamic signatures attempted by malware programs. In their classification processes, they employed machine learning algorithms such as Naïve Bayesian, J48 (decision trees), and support vector machines to enhance the accuracy of detection. These studies collectively contribute to the evolving landscape of malware detection, showcasing diverse methodologies and technologies aimed at enhancing the capabilities of security systems.

2.1 RESEARCH GAP

In recent times the use of different types of data analytical technique in different fields like health, traffic, population, etc. Different researchers are emphasis on different techniques on different types of datasets in different fields. Health sector is one such sectors where abundance amount of work has already been undertaken. Different techniques are applied with different variants on to different datasets. This is creating confusion that which technique with which variant is better way for prediction compared to other on different datasets.

III. METHODOLOGY

The proposed research methodology is based on various sequential steps.

Step1 Different research papers are taken for study purpose.

Step2 Identify the techniques whose results are in higher band for cyber-attack related datasets.

Step3 List all the techniques and their different variants that are to be implemented for comparative study. Different performance parameters are considered for comparison for the outcome.

Step4 Select the datasets in Cyber-attack for which these techniques are to be implemented.

Step5 Apply the techniques using python onto the selected datasets.

Step6 Extract the results on the basis of selected parameters.

Step7 Show the results using different types of graphical means for better representation.

Step8 Write the research paper to publish the results for the selected techniques.

Step9 Complete the dissertation.

IV. ALGORITHM

Step1 Dataset D will be collected.

Step2 N (D) = N, the normalized dataset will be generated from the real dataset.

Step3 Subdivide the dataset D=> Train (D), Test (D)

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

Step4: Apply KNN (Train (D))

Step5: Apply KNN (Test (D))

Step6: Confusion Matrix (KNN(Test(D))

V. DATASETS

There are three datasets that are considered while experimenting with different techniques under different variants.

S. No.	Dataset Name
1	Cyber_attack_2019
2	Cyber_attack_2021
3	Cyber_attack_2022

Table 1 Dataset

Comparison of KNN with Manhattan distances for cyber-attack dataset with different sizes training and testing set.

VI. RESULTS

6.1 Results with KNN

	KNN with Euclidean Distance			KNN with Chi Square Distance			KNN with Manhattan Distance		
Parameters	70-30	30-70	50-50	70-30	30-70	50-50	70-30	30-70	50-50
Accuracy	0.6	0.7	0.71	0.58	0.65	0.67	0.59	0.67	0.68
precision	0.81	0.8	0.85	0.79	0.778135	0.797297	0.8	0.87	0.8
sensitivity	0.68	0.7	0.72	0.65	0.683616	0.670455	0.66	0.73	0.7
specificity	0.4	0.55	0.53	0.37	0.448	0.423077	0.43	0.56	0.5
FPR	0.65	0.5	0.45	0.59	0.282486	0.378788	0.69	0.67	0.4

Table 2 Results with KNN

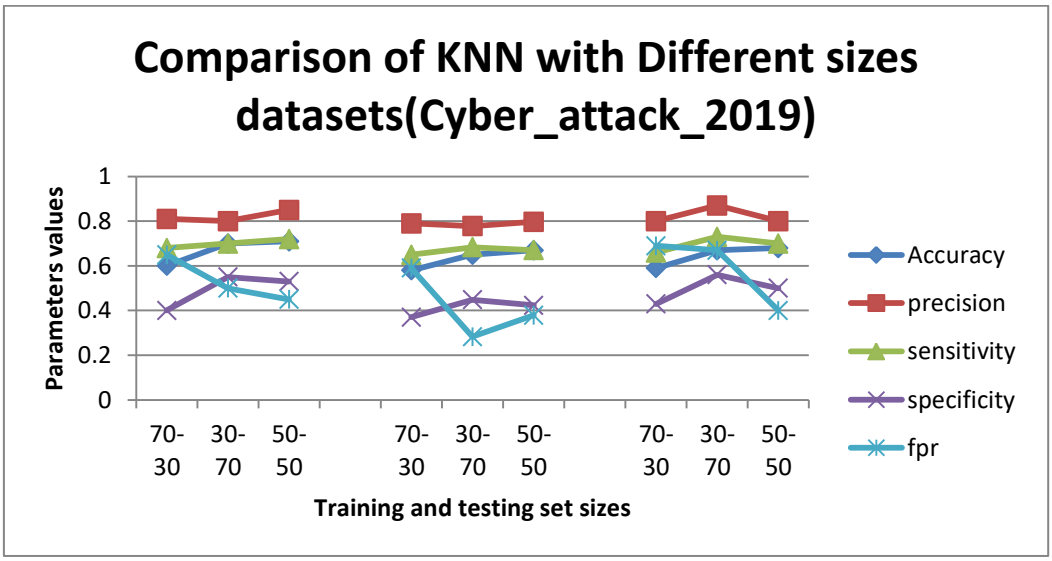

Fig. 1 Comparison with KNN

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

6.1.1 Comparison for SVM with different kernels on Cyber_attack_2022 dataset

	SVM with Kernel rbf			SVM with Kernel poly		
	70-30	30-70	50-50	70-30	30-70	50-50
Accuracy	0.6341	0.6618	0.64035	0.634146	0.6618	0.6404
precision	1	1	1	1	1	1
sensitivity	0.6341	0.6618	0.64035	0.634146	0.6618	0.6404
specificity	0	0	0	0	0	0
FPR	0.4878	0.2088	0.2924	0.487805	0.2088	0.2924

Table 3 SVM results

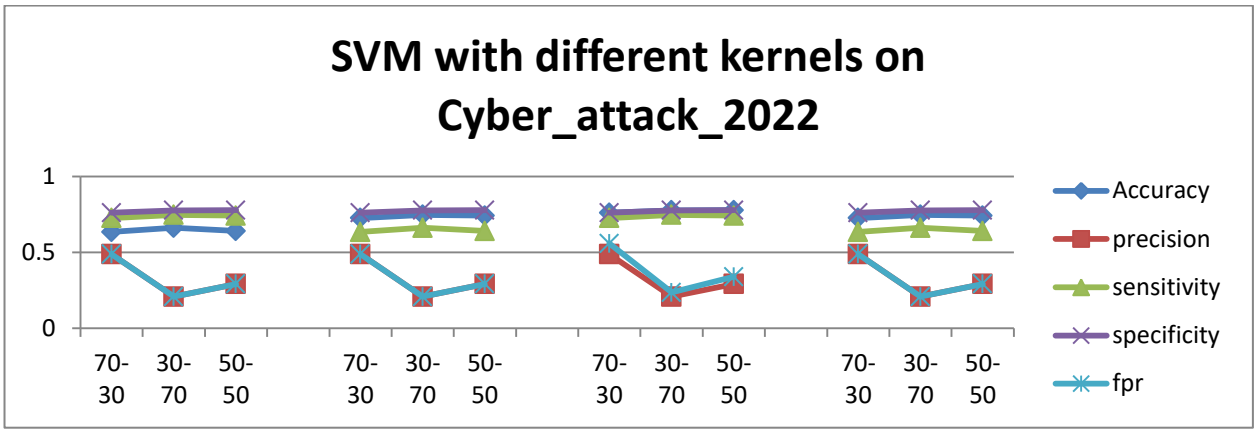

Fig. 2 Graph with SVM

6.1.2 Regression with dataset

	Simple Linear Regression			Multiple Linear Regression		
	70-30	30-70	50-50	70-30	30-70	50-50
Accuracy	76.693	73.305	75.814	0.78	0.79	0.81
RMSE	0.54	0.63	0.65	0.15	0.62	0.63

Table 4 Regression with Cyber_attack_2022 dataset

	Logistic Regression		
	70-30	30-70	50-50
Accuracy	0.78	0.67	0.77
precision	0.63	0.67	0.78
sensitivity	0.6257	0.65	0.68
specificity	0.56	0.56	0.54
FPR	0.5848	0.45	0.34

Table 5 Regression with Cyber_attack_2022 with training set sizes

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

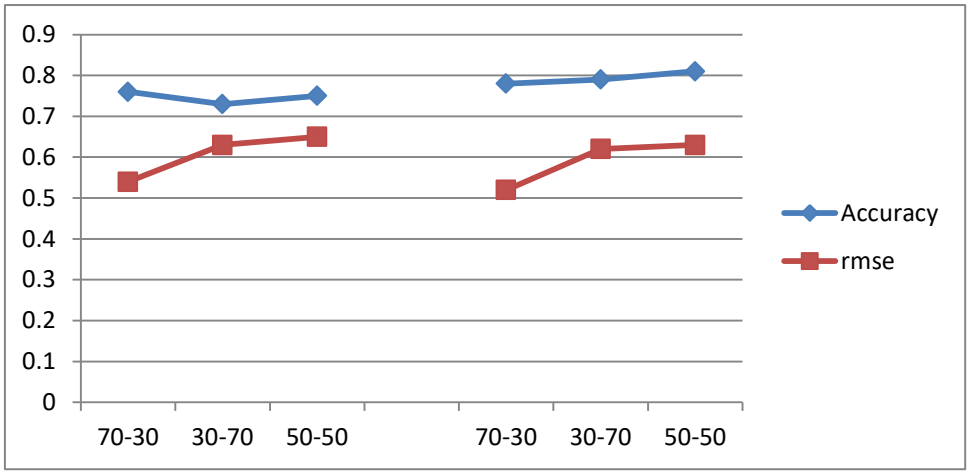

Fig. 3 Comparison of simple and multiple linear regression

Fig. 5.9 shows the comparison of the single and multiple linear regression is shown in the diagram. The performance of the multiple linear regression is better compared to the single linear regression.

VII. DISCUSSIONS

The experimentation results show the satisfactory results for three different techniques under different variants for three different datasets. The result shows that the performance for the KNN with the Manhattan distance is having better performance on all the datasets, SVM sigmoid kernel performance if better compared to the other kernels on all the dataset, multiple variable regression in having better performance compared to the other types of regressions.

VIII. CONCLUSION AND FUTURE WORK

8.1 CONCLUSION

The experimentation of the different techniques for different datasets with different types of variants performance has been interpreted. The results show that same technique performance on all different dataset is remaining high. In current scenario KNN with the Manhattan distance performance is better in all the cases. The performance of the multiple variable regressions is better compared to the other types of regressions. The performance of the SVM with the sigmoid kernel is having better performance. On the overall the performance of multiple variable regressions is comparatively higher than the other two types of techniques.

8.2 FUTURE WORK

The comparison of the different techniques with different variants with different datasets has been done. This work can be further enhanced by including various other techniques available as machine learning techniques for prediction purpose.

REFERENCES

1. Tavallaee, M.; Stakhanova, N.; Ghorbani, A.A. Toward credible evaluation of anomaly-based intrusion-detection methods. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2010, 40, 516–524. [Google Scholar] [CrossRef]
2. Aviv, A.J.; Haeberlen, A. Challenges in Experimenting with Botnet Detection Systems. In Proceedings of the 4th Workshop on Cyber Security Experimentation and Test (CSET 11), San Francisco, CA, USA, 8 August 2011. [Google Scholar]
3. Velan, P.; Čermák, M.; Čeleda, P.; Drašar, M. A survey of methods for encrypted traffic classification and analysis. Int. J. Netw. Manag. 2015, 25, 355–374. [Google Scholar] [CrossRef]
4. De Lucia, M.J.; Cotton, C. Identifying and detecting applications within TLS traffic. In Proceedings of the Cyber Sensing 2018, Orlando, FL, USA, 15–19 April 2018; Volume 10630. [Google Scholar] [CrossRef]
5. Kaur, S.; Singh, M. Automatic attack signature generation systems: A review. IEEE Secur. Priv. 2013, 11, 54–61. [Google Scholar] [CrossRef]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205221 |

6. Ahmed, M.; Naser Mahmood, A.; Hu, J. A survey of network anomaly detection techniques. J. Netw. Comput. Appl. 2016, 60, 19–31. [Google Scholar] [CrossRef]
7. Zeek IDS. 2021. Available online: https://zeek.org (accessed on 10 May 2021).
8. Shiravi, A.; Shiravi, H.; Tavallaee, M.; Ghorbani, A.A. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Comput. Secur. 2012, 31, 357–374. [Google Scholar] [CrossRef]
9. Moustafa, N.; Slay, J. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Inf. Secur. J. Glob. Perspect. 2016, 25, 18–31. [Google Scholar] [CrossRef]
10. Maciá-Fernández, G.; Camacho, J.; Magán-Carrión, R.; García-Teodoro, P.; Therón, R. UGR '16: A new dataset for the evaluation of cyclostationarity-based network IDSs. Comput. Secur. 2018, 73, 411–424. [Google Scholar] [CrossRef][Green Version]
11. Lippmann, R.P.; Fried, D.J.; Graf, I.; Haines, J.W.; Kendall, K.R.; McClung, D.; Weber, D.; Webster, S.E.; Wyschogrod, D.; Cunningham, R.K.; et al. Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation. In Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX'00), Hilton Head, SC, USA, 25–27 January 2000; Volume 2, pp. 12–26. [Google Scholar] [CrossRef][Green Version]
12. Siddique, K.; Akhtar, Z.; Khan, F.A.; Kim, Y. KDD Cup 99 data sets: A perspective on the role of data sets in network intrusion detection research. Computer 2019, 52, 41–51. [Google Scholar] [CrossRef]

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details