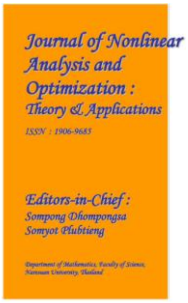
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.10 : 2024
ISSN : **1906-9685**



Journal of Nonlinear
Analysis and
Optimization :
Theory & Applications
ISSN : 1906-9685

Editors-in-Chief :
Sompong Dhompongsa
Somyot Plubtieng

Department of Mathematics, Faculty of Science,
Naresuan University, Thailand

Enhanced DDoS Detection Using Semi-Supervised Machine Learning Techniques

[1] KAVALI LAHARI, [2] V. NARAHARI, [3] D. LAKSHMI NARAYANA REDDY

[1] *PG Scholar, Department of Computer Science and Engineering, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, Andhra Pradesh*

[2,3] *Assistant Professor, Department of Computer Science and Engineering, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, Andhra Pradesh*

Abstract: The proliferation of malicious applications poses a significant threat to the Android platform, exploiting various network interfaces to the pilfer users' personal data and execute attack operations. In this study, we propose an efficient and automated method for detecting malware using the semantic analysis of network traffic text. Our approach threats each HTTP flow generated by mobile applications as a textual document, amenable to natural language processing (NLP) for feature extraction. Leveraging the network traffic data, we construct a robust malware detection model. Initially, we employ the N-gram method from NLP to analyze the traffic flow headers, extracting meaningful features. Subsequently, we introduce an automatic feature selection algorithm based on the chi-square test to identify significant associations between variables. We present a novel approach to malware detection using NLP techniques, treating mobile traffic as textual documents. Through the application of an automatic feature selection algorithm based on N-gram sequences, we derive insightful features from the semantics of traffic flows. Our methods have uncovered malware instances that elude detection by traditional antivirus scanners. Furthermore, we develop a detection system capable of monitoring traffic across institutional enterprise networks, home networks, and 3G/4G mobile networks. This system integrates with computers to detect suspicious network behaviors.

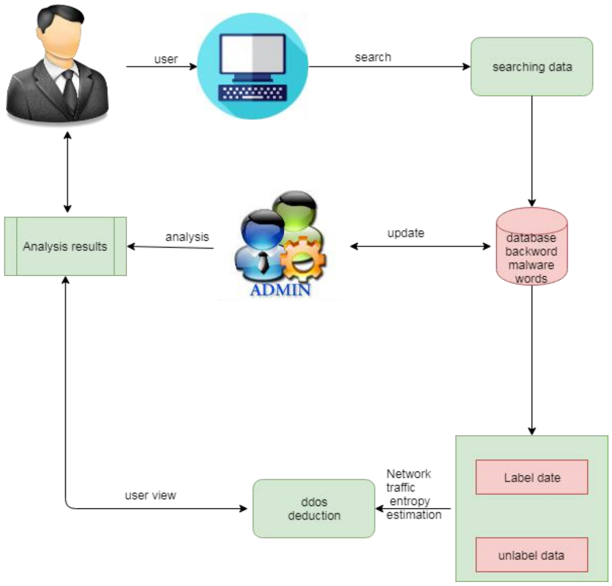
Keywords – Malware detection, HTTP flow analysis, text semantics, machine learning.

I. INTRODUCTION

Over the last two decades, data mining techniques have played a crucial role inn developing sophisticated intrusion detection systems. Techniques such as Artificial Intelligence (AI), Machine Learning (ML), Pattern recognition, Statistics, and information Theory are widely employed for intrusion detection. As the internet becomes more integral to daily operations, ensuring data availability has become a critical challenge. Data availability is essential for a network system's security. Distributed Denial of Service (DDoS) attacks pose a significant threat by intentionally disrupting or degrading the quality of a network or service. These attacks often involve numerous compromised online devices, known ass botnets, which facilitate massive attacks by leveraging the collective power of many devices. Botnets also obscure the attack's origin, complicating mitigation effort. Additionally, distinguishing between legitimate high-traffic events (flash crowds) and DDoS attacks remains challenging.

There are two primary methods for launching DDoS attacks. The first involves sending malformed packets to the victim to confuse a protocol or application (a vulnerability attack). The second, more common method involves overwhelming the target with excessive traffic. This paper outlines the life cycle of DDoS detection, which includes several phases:

1. **Prevention:** This phase focuses on protecting a system against attacks by implementing appropriate security measures at various points. It aims to safeguard server resources and ensure that online services are available to legitimate clients.
2. **Mitigation:** Applied when an attack occurs, this phase involves executing suitable security countermeasures to handle or slow down the attack, effectively stopping the malicious activity.
3. **Detection:** This phase involves analyzing the system to identify malicious traffic indicative of a DDoS attack. Detection techniques include pattern matching, clustering, statistical methods, deviation analysis, associations, and correlation, which help identify abnormal GET request traffic against a web server.
4. **Monitoring:** Essential for real-time detection, monitoring involves collecting necessary information about a host or network using tools like network monitoring software. This phase becomes more complex when attackers use botnets distributed globally to launch low-rate DDoS attacks.
5.



II LITERATURE SURVEY

[1] An empirical evaluation of information metrics for low-rate and high-rate ddos attack detection: Distributed Denial of Service (DDoS) attacks represent a major threat to uninterrupted and efficient Internet service. In this paper, we empirically evaluate several major information metrics, namely, Hartley entropy, Shannon entropy, Renyi's entropy, generalized entropy, Kullback–Leibler divergence and generalized information distance measure in their ability to detect both low-rate and high-rate DDoS attacks. These metrics can be used to describe characteristics of network traffic data and an appropriate metric facilitates building an effective model to detect both low-rate and high-rate DDoS attacks. We use MIT Lincoln Laboratory, CAIDA and TUIDS DDoS datasets to illustrate the efficiency and effectiveness of each metric for DDoS detection.

[2] Constructing detection knowledge for ddos intrusion tolerance: Intrusion tolerance is the ability of a system to continue providing (possibly degraded but) adequate services after a penetration. With the rapid development of network technology, distributed denial of service (DDoS) attacks become one of the most important issues today. In this paper, we propose a DDoS ontology to provide a common terminology for describing the DDoS models consisting of the Profile model (the representation of the behaviors of system and users) and the Defense model (the descriptions of Detection and Filter methodologies). Also, the Evaluation strategy based upon current statuses of users' behaviors is used to evaluate the degree of the intrusion tolerance of the proposed models during DDoS attacks. Based upon the ontology, four KCs (Profile model, Evaluation strategy, Detection methodology, and Filter methodology Knowledge Classes) and their relationships are then proposed, where each KC may contain a set of sub-KCs or knowledge represented as a natural rule format. For an arbitrarily given network environment, the default knowledge in the Profile KC and the Evaluation KC, the appropriate detection features in the Detection KC, and the suitable access control list policies in the Filter KC can be easily extracted and adopted by our proposed integrated knowledge acquisition framework. We are now implementing a NORM-based DDoS intrusion tolerance system for DDoS attacks to evaluate the proposed models.

[3] Defending against flooding-based distributed denial-of-service attacks: Flooding-based distributed denial-of- service (DDoS) attack presents a very serious threat to the stability of the Internet. In a typical DDoS attack, a large number of compromised hosts are amassed to send useless packets to jam a victim or its Internet connection, or both. In the last two years, it is discovered that DDoS attack methods and tools are becoming more sophisticated, effective, and also more difficult to trace to the real attackers. On the defense side, cur-rent technologies are still unable to with stand large-scale attacks. The main purpose of this article is therefore twofold. The first one is to describe various DDoS attack methods, and to present a systematic review and evaluation of the existing defense mechanisms. The second is to discuss a longer-term solution, dubbed the Inter-net-firewall approach, that attempts to intercept attack packets in the Internet core, well before reaching the victim

[4] Distributed denial of service attack and defense: This brief provides readers a complete and self-contained resource for information about DDoS attacks and how to defend against them. It presents the latest developments in this increasingly crucial field along with background context and survey material. The book also supplies an overview of DDoS attack issues, DDoS attack detection methods, DDoS attack source trace back, and details on how hackers organize DDoS attacks. The author concludes with future directions of the field, including the impact of DDoS attacks on cloud computing and cloud technology. The concise yet comprehensive nature of this brief makes it an ideal reference for researchers and professionals studying DDoS attacks. It is also a useful resource for graduate students interested in cyberterrorism and networking

[5] Four decades of data mining in network and systems management. How has the interdisciplinary data mining field been practiced in Network and Systems Management (NSM)? In Science and Technology, there is a wide use of data mining in areas like bioinformatics, genetics, Web, and, more recently, astro informatics. However, the application in NSM has been limited and inconsiderable. In this article, we provide an account of how data mining has been applied in managing networks and systems for the past four decades, presumably since its birth. We look into the field’s applications in the key NSM activities—discovery, monitoring, analysis, reporting, and domain knowledge acquisition. In the end, we discuss our perspective on the issues that are considered critical for the effective application of data mining in the modern systems which are characterized by heterogeneity and high dynamism.

III SYSTEM ANALYSIS

EXISTING SYSTEM:

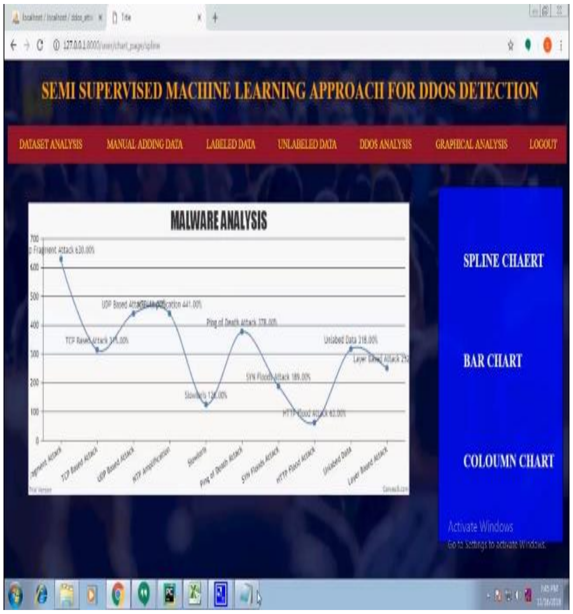
The first phase of their approach consists of dividing the incoming network traffic into three type of protocols TCP, UDP or Other. Then classifying it into normal or anomaly traffic. In the second stage a multi-class algorithm classifies the anomaly detected in the first phase to identify the attacks class in order to choose the appropriate intervention. Two public datasets are used for experiments in this paper namely the UNSW-NB15 and the NSL-KDD Several approaches have been proposed for detecting DDoS attack. Information theory and machine learning are the performances of network intrusion detection approaches, in general, rely on the distribution characteristics of the underlaying network traffic data used for assessment. The DDoS detection approaches in the literature are under two main categories unsupervised approaches and supervised approaches. Depending on the benchmark datasets used, unsupervised approaches often suffer from high false positive rate and supervised approach cannot handle large amount of network traffic data and their performances are often limited by noisy and irrelevant network data. Therefore, the need of combining both, supervised and unsupervised approaches arise to overcome DDoS detection issues.

DISADVANTAGES:

- The datasets above are split into train subsets and test subsets using a configuration of 60% and 40% respectively. The train subsets are used to fit the Extra-Trees ensemble classifiers and the test subsets are used to test the entire proposed approach. Before fitting the classifiers, the train subsets are normalized using the *Min, Max* method
- This section presents the details of the proposed approach and the methodology followed for detecting the DDoS attack. The proposed approach consists of five major steps: Datasets preprocessing, estimation of network trafficEntropy, online co-clustering, information gain ratio.
- The aim of splitting the anomalous network traffic is to reduce the amount of data to be classified by excluding the normal cluster for the classification. For DDoS detection normal traffic records are irrelevant and noisy as the normal behaviors continue to evolve. Most of the time the new unseen normal traffic instances cause the increase of the false positive rate and the decrease of the classification accuracy. Hence, excluding some noisy normal instances of the network traffic data for classification is beneficial in terms of low false positive rates and classification accuracy. Assuming that after the network traffic clustering one cluster contains only normal traffic, a second one contains only DDoS traffic and a third one contains both DDoS and normal traffic.

PROPOSED SYSTEM:

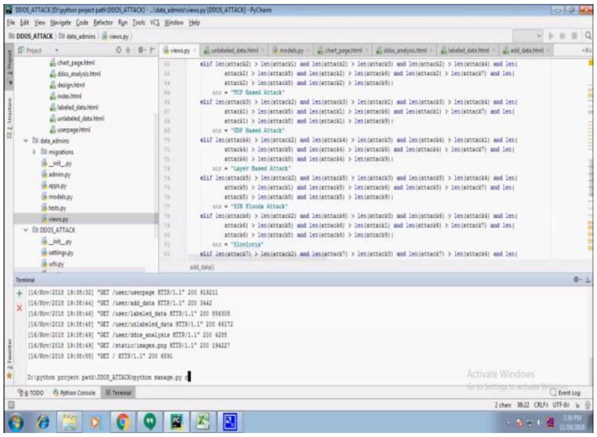
This section introduces our methodology to detect the DDoS attack. The five-fold steps application process of data mining techniques in network systems discussed in characterizes the followed methodology. The main aim of combining algorithms used in the proposed approach is to reduces noisy and irrelevant network traffic data before preprocessing and classification stages for DDoS detection while maintaining high performance in terms of accuracy, false positive rate and running time, and low resources usage. Our approach starts with estimating the entropy of the FSD features over a time-based sliding window. When the average entropy of a time window exceeds its lower or upper thresholds the co-clustering algorithm split the received network traffic into three clusters. Entropy estimation over time sliding windows allows to detect abrupt changes in the incoming network traffic distribution which are often caused by DDoS attacks. Incoming network traffic within the time windows having abnormal entropy values is suspected to contain DDoS traffic. The focus only on the suspected time windows allows to filter important amount of network traffic data, therefore only relevant data is selected for the remaining steps of the proposed approach. Also, important resources are saved when no abnormal entropy occurs. In order to determine the normal cluster, we estimate the information gain ratio based on the average entropy of the FSD features between the received network traffic data during the current time window and each one of the obtained clusters. As discussed in the previous section during a DDoS period the generated amount of attack traffic is largely bigger than the normal traffic. Hence, estimating the information gain ratio based on the FSD features allows to identify the two cluster that preserve more information about the DDoS attack and the cluster that contains only normal traffic. Therefore, the cluster that produce lower information gain ratio is considered as normal and the remaining clusters are considered as anomalous. The information gain ratio is computed for each cluster as follows:



ADVANTAGES:

- Where *subset* represents the received subset of network data during the time window *w*, *Ci (i = 1, 2, 3)* are the obtained clusters from *subset* and |*Ci*| is the size of the*ith*cluster. *Avg H(subset)* is the average entropy of the FSD features of the input *subset* and |*subset* | represents the size
- The clustering of the incoming network traffic data allows to reduce important amount of normal and noisy data before the preprocessing and classification steps. More than6% of a whole traffic dataset can be filtered.

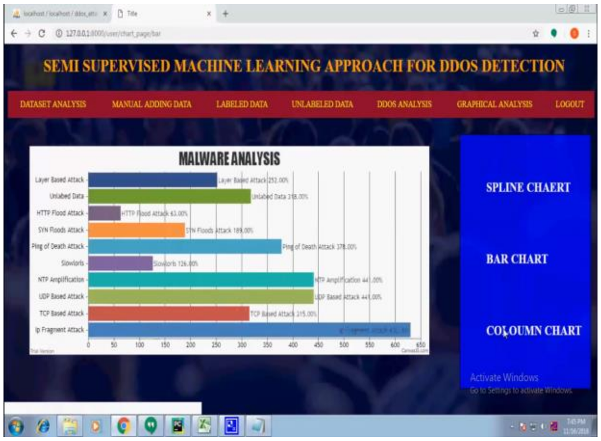
IV RESULTS

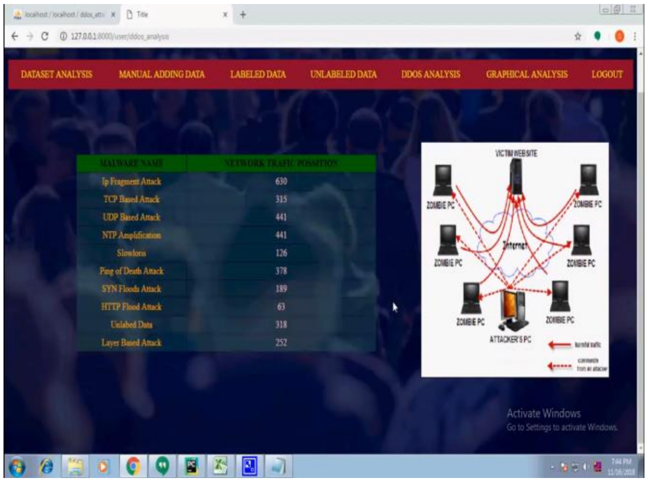


```
    elif len(attack2) > len(attack1) and len(attack2) > len(attack3) and len(attack2) > len(attack4) and len(
            attack2) > len(attack5) and len(attack2) > len(attack6) and len(attack2) > len(attack7) and len(
            attack2) > len(attack8) and len(attack2) > len(attack9):
        str = "TCP Based Attack"
    elif len(attack3) > len(attack2) and len(attack3) > len(attack1) and len(attack3) > len(attack4) and len(
            attack1) > len(attack5) and len(attack1) > len(attack6) and len(attack1) > len(attack7) and len(
            attack1) > len(attack8) and len(attack1) > len(attack9):
        str = "UDP Based Attack"
    elif len(attack4) > len(attack2) and len(attack4) > len(attack3) and len(attack4) > len(attack1) and len(
            attack4) > len(attack5) and len(attack4) > len(attack6) and len(attack4) > len(attack7) and len(
            attack4) > len(attack8) and len(attack4) > len(attack9):
        str = "Layer Based Attack"
    elif len(attack5) > len(attack2) and len(attack5) > len(attack3) and len(attack5) > len(attack4) and len(
            attack5) > len(attack1) and len(attack5) > len(attack6) and len(attack5) > len(attack7) and len(
            attack5) > len(attack8) and len(attack5) > len(attack9):
        str = "SYN Floods Attack"
    elif len(attack6) > len(attack2) and len(attack6) > len(attack3) and len(attack6) > len(attack4) and len(
            attack6) > len(attack5) and len(attack6) > len(attack1) and len(attack6) > len(attack7) and len(
            attack6) > len(attack8) and len(attack6) > len(attack9):
        str = "Slowloris"
    elif len(attack7) > len(attack2) and len(attack7) > len(attack3) and len(attack7) > len(attack4) and len(
```

```
[16/Nov/2018 19:38:32] "GET /user/userpage HTTP/1.1" 200 919211
[16/Nov/2018 19:38:44] "GET /user/add_data HTTP/1.1" 200 3442
[16/Nov/2018 19:38:46] "GET /user/labeled_data HTTP/1.1" 200 858308
[16/Nov/2018 19:38:48] "GET /user/unlabeled_data HTTP/1.1" 200 66172
[16/Nov/2018 19:38:49] "GET /user/ddos_analysis HTTP/1.1" 200 4288
[16/Nov/2018 19:38:49] "GET /static/images.png HTTP/1.1" 200 194227
[16/Nov/2018 19:38:55] "GET / HTTP/1.1" 200 6891

D:\python project path\DDOS_ATTACK>python manage.py r
```





MALWARE NAME	NETWORK TRAFIC POSSITION
Ip Fragment Attack	630
TCP Based Attack	315
UDP Based Attack	441
NTP Amplification	441
Slowloris	126
Ping of Death Attack	378
SYN Floods Attack	189
HTTP Flood Attack	63
Unlabed Data	318
Layer Based Attack	252

V CONCLUSION

Android represents a rapidly growing threat in the realm of malware. Currently, many research methods and antivirus scanners are inadequate against the expanding size and diversity of mobile malware. To address this, we introduce a novel solution for mobile malware detection based on network traffic flows. This approach treats each HTTP flow as a document and applies NLP string analysis to the HTTP flow requests. Using N-Gram line generation, a feature selection algorithm, and the SVM algorithm, we develop an effective malware detection model. Our evaluation demonstrates the efficiency of this solution, significantly improving upon existing methods and successfully identifying malicious leaks, albeit with some false positives. The detection rate for harmful traffic is 99.15%, with a false positive rate of 0.45%. Further validation with newly discovered malware confirms the performance of our proposed system. In real-world environments, our model detects 54.81% of harmful applications, outperforming other popular antivirus scanners. Our tests reveal that malware models can be detected by our system, which does not interfere with the detection capabilities of other virus scanners. Additionally, it is possible to obtain new malicious models from Virus Total detection reports. Once new samples are added, we will re-train, refresh, and update our model to incorporate the latest malware.

REFERENCES

[1] P. Naresh, P. Srinath, K. Akshit, M. S. S. Raju and P. VenkataTeja, "Decoding Network Anomalies using Supervised Machine Learning and Deep Learning Approaches," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1598-1603, doi: 10.1109/ICACRS58579.2023.10404866.

[2] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205. [3] Ahmed, E., Naser Mahmood, A. K., & Hu, J. (2016). DDoS attack detection method based on semi-supervised SVM. Procedia Computer Science, 96, 835-842.

[4] Jaiswal, A., & Kumar, A. (2020). Semi-supervised machine learning approach for DDoS attack detection. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[5] Saeed, A., & Zeadally, S. (2019). A survey of DDoS detection and defense mechanisms in cloud computing. Journal of Network and Computer Applications, 133, 42-60.

[6] Tang, H., & Yu, F. (2017). Semi-supervised learning for classifying network attacks. In 2017 13th International Conference on Computational Intelligence and Security (CIS) (pp. 93-97). IEEE.

[7] Pham, T. V., & Nguyen, N. D. (2018). Semi-supervised deep learning for DDoS detection in SDN-based IoT systems. In 2018 10th International Conference on Knowledge and Systems Engineering (KSE) (pp. 1-6). IEEE.

[8] Peng, Y., & Leckie, C. (2016). Distributed semi-supervised learning for intrusion detection in networked systems. IEEE Transactions on Dependable and Secure Computing, 15(2), 233-246.

[9] P. Naresh, K. Pavan kumar, and D. K. Shareef, ‘Implementation of Secure Ranked Keyword Search by Using RSSE,’ International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 2 Issue 3, March – 2013.

[10] Kumar, N., & Goyal, V. (2017). A survey of ensemble learning algorithms for semi-supervised learning. In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (ICCDE) (pp. 1250-1255). IEEE.

[11] Alshammari, R. A., & Alshammari, G. (2020). Deep Learning Techniques for DDoS Attack Detection and Defense Mechanisms: A Review. Computers, Materials & Continua, 65(3), 1603-1627.

[12] Yu, Y., & Yao, Y. (2019). A DDoS attack detection model based on improved semi-supervised deep learning. Mathematical Problems in Engineering, 2019.

[13] Sunder Reddy, K. S. ., Lakshmi, P. R. ., Kumar, D. M. ., Naresh, P. ., Gholap, Y. N. ., & Gupta, K. G. . (2024). A Method for Unsupervised Ensemble Clustering to Examine Student Behavioral Patterns. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 417–429. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4854..

[14] Hussan, M.I. & Reddy, G. & Anitha, P. & Kanagaraj, A. & Pannangi, Naresh. (2023). DDoS attack detection in IoT environment using optimized Elman recurrent neural networks based on chaotic bacterial colony optimization. Cluster Computing. 1-22. 10.1007/s10586-023-04187-4.

[15] Ahmed, E., & Mahmood, A. K. N. (2017). An adaptive semi-supervised SVM-based DDoS attack detection system in cloud computing. In 2017 International Conference on Engineering & MIS (ICEMIS) (pp. 1-6). IEEE.