Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 1 (2025) 69 - 76 e-ISSN: 2580-0760

Enhancing Network Security: Evaluating SDN-Enabled Firewall Solutions and Clustering Analysis Using K-Means through Data-Driven Insights

Ahmad Turmudi Zy[1]*, Isarianto[2], Anggi Muhammad Rifai[3], Agung Nugroho[4], Abdul Ghofir[5]
[1,2,3,4] Department of Informatics Engineering, Faculty of Engineering, Pelita Bangsa University, Bekasi, Indonesia
[5] Department of Informatics Engineering, Faculty of Engineering, President University, Bekasi, Indonesia
[1]turmudi@pelitabangsa.ac.id, [2]isarianto@pelitabangsa.ac.id, [3]anggimuhammad@pelitabangsa.ac.id,
[4]agung@pelitabangsa.ac.id, [5]geoff@president.ac.id

Abstract

In the face of escalating and increasingly complex cyber threats, enhancing network security has become a critical challenge. This study addresses this issue by investigating the optimization of SDN-enabled firewall solutions using a data-driven approach. The research employs K-Means clustering to analyze attack patterns, aiming to identify and understand distinct patterns for improved firewall effectiveness. Through the clustering process, attack data was classified into three clusters: Cluster 0, indicating concentrated attack sources likely tied to high-activity regions or networks; Cluster 1, representing a dispersed *distribution of attacks, pointing to diverse origins; and Cluster 2, linked to specific geographic regions or unique attack behaviors. The clustering efficacy was evaluated using the Silhouette Score (0.606) and the Davies-Bouldin Index (0.614), indicating meaningful and reliable clustering outcomes. These findings provide actionable insights into network threat patterns, enabling the refinement and enhancement of SDN-enabled firewalls. The study contributes to the field by demonstrating the potential of clustering techniques in uncovering patterns overlooked by traditional methods and paving the way for further research into alternative clustering algorithms and broader applications in network security.*

Keywords: attack patterns; data-driven analysis; K-Means clustering; network security; SDN-enabled firewalls

How to Cite: Ahmad Turmudi Zy, Isarianto, A. M. Rifa'i, A. Nugroho, and A. Ghofir, "Enhancing Network Security: Evaluating SDN-Enabled Firewall Solutions and Clustering Analysis Using K-Means through Data-Driven Insights", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 1, pp. 69 - 76, Jan. 2025.
DOI: https://doi.org/10.29207/resti.v9i1.6056

1. Introduction

In the constantly evolving digital landscape, organizations face an ever-growing range of cyber threats that are both more sophisticated and frequent [1]. These threats target network infrastructures, aiming to exploit vulnerabilities in systems, potentially leading to significant security breaches and loss of sensitive data [2]. Attack methods vary widely, from simple intrusions to highly complex and sustained campaigns known as advanced persistent threats (APTs) [3]. The diversity and complexity of these attacks require more than just traditional defensive strategies to effectively protect network integrity [4].

The rise in cyberattacks has been driven by technological advancements that have made it easier for malicious actors to bypass outdated security measures [5]. As organizations adopt more interconnected and dynamic systems, vulnerabilities in networks become harder to manage [6]. This expanding attack surface makes traditional security solutions such as static firewalls and signature-based detection systems inadequate in the face of increasingly sophisticated threats [7]. These conventional systems often struggle to detect new or evolving forms of attack, leaving networks exposed to considerable risk [8].

Furthermore, traditional security measures typically rely on predefined rules or known signatures to detect threats, making them inherently reactive and slow to adapt to new challenges [9]. With cyberattacks becoming more adaptive, there is a pressing need for security solutions that can offer real-time protection [10]. As attackers become more creative and persistent, organizations require advanced solutions that can not only detect but also mitigate threats before they cause serious damage [2].

Received: 18-09-2024 | Accepted: 23-12-2024 | Published Online: 25-01-2024

Recent studies have highlighted the limitations of conventional security measures, sparking interest in more adaptive technologies such as Software-Defined Networking (SDN) [11]. SDN allows for a separation of the control plane from the data plane, offering a centralized and flexible approach to managing network traffic [12]. This architecture has opened the door to more advanced and dynamic security systems, where rules and policies can be updated automatically in response to emerging threats. Many researchers have identified SDN as a key technology for transforming network security by offering greater control over data flows [13].

Among the innovations driven by SDN, SDN-enabled firewalls have garnered particular attention for their ability to provide programmable, real-time responses to threats [14]. Unlike traditional firewalls, which require manual configuration and are slower to adapt, SDN-enabled firewalls can be programmed to adjust dynamically based on live network data [15]. Studies show that these firewalls not only enhance security but also improve overall network efficiency by minimizing disruptions caused by manual security updates [16].

However, while SDN-enabled firewalls have shown promise, their deployment faces certain challenges [17]. The sheer volume of data that these systems must analyze can overwhelm even the most advanced algorithms. Several research papers have emphasized the importance of integrating data analytics techniques to enhance SDN firewalls' ability to process and act upon this data efficiently [18]. This need for real-time analysis has led to the exploration of various machine learning techniques, particularly clustering methods like K-Means, to provide insights into attack patterns that could improve the adaptability and performance of SDN-enabled firewalls [19], [20].

Despite the advances in SDN-enabled firewalls, optimizing their effectiveness remains a significant challenge due to the dynamic and complex nature of network threats [21]. As these firewalls are programmed to respond to real-time network conditions, understanding and categorizing attack data becomes crucial for fine-tuning their performance [22]. Current research suggests that without a clear method to analyze this vast amount of data, SDN-enabled firewalls cannot fully leverage their potential [14]. The need for advanced analytical tools to process security data has become more urgent, as traditional methods fall short in categorizing and predicting new attack vectors [23].

Data-driven approaches such as K-Means clustering have been proposed as a solution to this problem [24]. K-Means clustering, a machine learning technique, can process large datasets to identify and group similar attack patterns [25]. This ability to categorize and analyze security data provides critical insights into how threats evolve and behave, allowing for more targeted and effective firewall configurations [26]. By understanding the distinct attack patterns, network administrators can make more informed decisions on how to deploy their firewall resources to protect against both current and future threats [27].

The integration of K-Means clustering with SDN technology represents an innovative approach to addressing the growing complexity of cyberattacks [28]. By using this clustering method, it is possible to move from a reactive to a proactive security posture. The combination of real-time threat analysis with adaptive firewall responses ensures that organizations can better safeguard their networks against emerging threats, offering a more robust defence system overall [25].

Recent studies by [29] have introduced advanced approaches to enhance security in Software-Defined Networks (SDNs), with a focus on addressing Distributed Denial of Service (DDoS) attacks. One such approach is the Whale Optimization Algorithm-based DDoS detection (WOA-DD), which utilizes a metaheuristic clustering technique to identify and mitigate DDoS threats. This algorithm has been evaluated in various conditions and has demonstrated significant robustness, stability, and efficiency. WOA-DD has been found to outperform several traditional solutions, proving its effectiveness in securing SDN environments against DDoS attacks.

This study is designed to evaluate the effectiveness of SDN-enabled firewalls by integrating K-Means clustering to analyze and categorize network attack patterns. The research aims to address the critical challenge of optimizing firewall configurations by providing a more detailed understanding of the threats that networks face [30]. By applying K-Means clustering, this study seeks to identify distinct attack clusters that can inform more adaptive and dynamic firewall policies, enhancing overall network protection against sophisticated and evolving cyber threats [31].

This research discusses imlementation of K-Means clustering to enhance the performance of SDN-enabled firewalls in identifying and mitigating cyber threats in real-time. The study focuses on clustering attack data to uncover patterns and insights that traditional analysis methods might overlook. By analyzing large datasets of attack patterns and classifying them into distinct clusters, the research aims to improve real-time threat detection and response capabilities within SDN-enabled networks [32].

The primary objective of this study is to evaluate the combined impact of K-Means clustering and SDN technology on network security. Addressing this objective fills a critical gap in the literature, as most existing studies explore the benefits of SDN and clustering techniques in isolation. The findings demonstrate that integrating advanced data analysis techniques enhances the efficacy of SDN-enabled firewalls, providing organizations with actionable

insights to bolster their security measures in the face of increasingly complex cyber threats.

The structure of this research is as follow figure 1: first, input the dataset from AWS Honeyspot; then pre-processing to find missing data fields; next step, add feature selection and split the dataset for training 80% and testing 20%; our clustering method is K-means clustering, and for measures prediction, we use the silhouette score and Davies-Bouldin index.

2. Research Methods

This research aims to evaluate the performance of SDN-enabled firewall solutions by integrating clustering analysis using the K-Means algorithm. The approach begins with the collection and preprocessing of network traffic data, which will serve as the foundation for identifying network anomalies. A data-driven methodology is employed, where key features are extracted and analyzed to determine patterns and potential security threats. The clustering process groups traffic flows based on similar characteristics, allowing for enhanced detection of abnormal behavior. The firewall solution is then tested under various network conditions to assess its efficiency in mitigating potential threats. The performance metrics of the solution are evaluated using accuracy, detection rate, and false-positive rate, providing insights into its overall effectiveness in enhancing network security and research purpose shown in Figure 1.

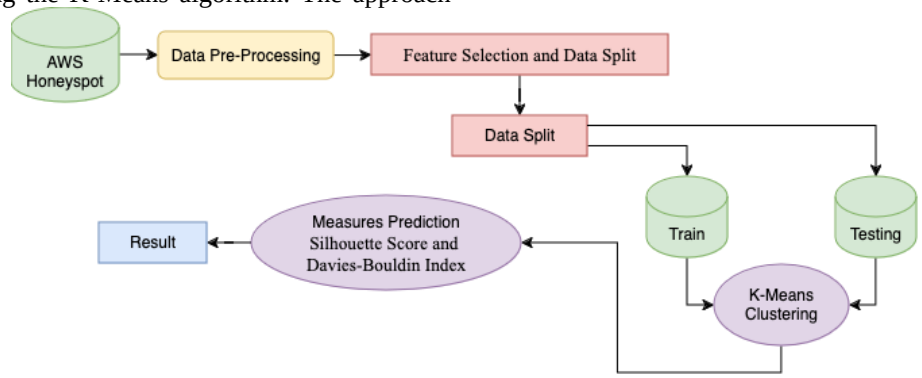

Figure 1. Research Purpose

Data was gathered from the AWS Honeypot [33], which collects detailed information on cyberattacks. The dataset includes attack details such as source IP addresses, ports, protocols, geographic information, and timestamps. The dataset spans various attack types, including TCP, UDP, and ICMP protocols.

The collected data required extensive preprocessing due to the presence of missing and irrelevant fields. First, fields such as 'type' and 'Unnamed: 15' were removed due to their high rate of missing values. Further, rows with missing latitude and longitude values were excluded to maintain geographic accuracy for the clustering analysis. Additionally, any outliers in latitude values exceeding 90 degrees were also filtered out. The dataset was resampled into hourly intervals to aggregate attack counts and prepare the time series data for further analysis. To normalize the attack distribution, a logarithmic transformation was applied to the count of attacks, mitigating the impact of extreme values on the clustering and prediction models shown in Table 1.

Table 1. Sample AWS Honeypot Dataset

No.	datetime	host	src	proto	spt	dpt	srcstr	CC	country	locale	latitude	longitude
0.	2013-03-03 21:53:00	groucho-oregon	1032051418	TCP	6000.0	1433.0	61.131.218.218	CN	China	Jiangxi Sheng	28.5500	115.9333
1.	2013-03-03 21:57:00	groucho-oregon	1347834426	UDP	5270.0	5060.0	80.86.82.58	DE	Germany	NaN	51.0000	9.0000
2.	2013-03-03 21:58:00	groucho-oregon	2947856490	TCP	2489.0	1080.0	5.180.184.106	TW	Taiwan	Taipei	25.0392	121.5250
3.	2013-03-03 21:58:00	groucho-us-east	841842716	UDP	43235.0	1900.0	50.45.128.28	US	United States	Oregon	45.5848	-122.9117
4.	2013-03-03 21:58:00	groucho-singapore	3587648279	TCP	56577.0	80.0	213.215.43.23	FR	France	NaN	48.8600	2.3500
....	....	....	....	....	....	....	....	....	....	....	....	....
4515805.	2013-09-08 05:55:00	groucho-tokyo	28142724	TCP	3555.0	445.0	1.173.108.132	TW	Taiwan	Taipei	25.0392	121.525

For feature selection, key attributes such as protocol type (TCP, UDP, ICMP), source country, and geographic location (latitude and longitude) were considered. These features provided essential information for analyzing attack vectors and clustering them based on geographic or protocol behavior. The dataset was split into training and testing sets to evaluate the predictive performance of the clustering and firewall solution models.

K-Means clustering was applied to analyze patterns in the attacks, grouping them based on geographical distribution and protocol types [34]. The algorithm was tuned to find the optimal number of clusters using the elbow method, which allowed us to capture the most distinct clusters of attack behavior. Each cluster represented a grouping of attacks with similar characteristics, which is vital for understanding how SDN-enabled firewalls could dynamically adapt to network threats based on geographic and protocol-based attack clusters [35]. By identifying distinct groups, we could gain insights into the frequency and type of attacks in specific regions, allowing for more targeted firewall rule applications.

K-Means is an unsupervised learning algorithm commonly used to solve clustering problems [36]. Its goal is to partition data into distinct groups or clusters based on specific features, where each cluster is represented by a centroid that acts as the center of all the data points within the cluster [37]. The algorithm begins by initializing a predetermined number of clusters, K, and randomly selecting K data points from the dataset as initial centroids. These centroids serve as the initial cluster centers [38].

In the next step, the algorithm assigns each data point to the nearest centroid. The distance between a data point and a centroid is typically calculated using the Euclidean Distance formula, although other distance measures can be used. The Euclidean Distance between two points, x and y, in an n-dimensional space is given by Formula 1 [39].

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

d(x,y) represents the distance between the points, and x_i and y_i are the coordinates of the points in the iii-th dimension. Once all data points are assigned to the closest centroid, the algorithm updates the centroid positions by calculating the mean of all data points assigned to each cluster. The new centroid for a cluster is the average of the data points in that cluster, computed using Formula 2 [40].

$$C_i = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i \quad (2)$$

C_j represents the new centroid of cluster j, n_j is the number of data points in the cluster, and x_i are the data points in that cluster.

This process of assigning points to the nearest centroid and updating centroids is repeated iteratively until the centroids no longer change significantly, or a predefined number of iterations is reached. The objective of K-Means is to minimize the Sum of Squared Errors (SSE), which is the total squared distance between each data point and its nearest centroid. The SSE is calculated as Formula 3 [39].

$$SSE = \sum_{j=1}^{K}\sum_{x_i \in C_j} ||x_i - C_j||^2 \quad (3)$$

K is the number of clusters, x_i represents the data points, and C_j is the centroid of cluster j. By minimizing SSE, K-Means ensures that the data points within each cluster are as close to their respective centroid as possible, leading to tighter and more cohesive clusters.

To evaluate the effectiveness of the SDN-enabled firewall, the model's predictive power in classifying network traffic was assessed using common measures such as the Silhouette Score and Davies-Bouldin Index [41]. These measures were calculated based on how well the clustering algorithm distinguished between different types of attacks. Additionally, the number of false positives and negatives was examined to measure the firewall's sensitivity and ensure it does not block legitimate traffic while efficiently detecting threats.

The analysis showed that K-Means clustering could successfully identify attack patterns based on geographical regions and protocol types. By integrating these insights with SDN-enabled firewall configurations, we could create adaptive firewall rules that automatically respond to incoming threats. The firewall's dynamic adjustments based on the clustering model led to improved detection and response times, especially in regions with high attack volumes like China and the United States. This approach not only enhances network security but also provides a scalable solution for real-time threat detection in dynamic network environments.

3. Results and Discussions

3.1 Data Analysis

The graph illustrates in Figure 2, the number of attacks per hour between March 2013 and September 2013, revealing significant fluctuations in attack frequency over time.

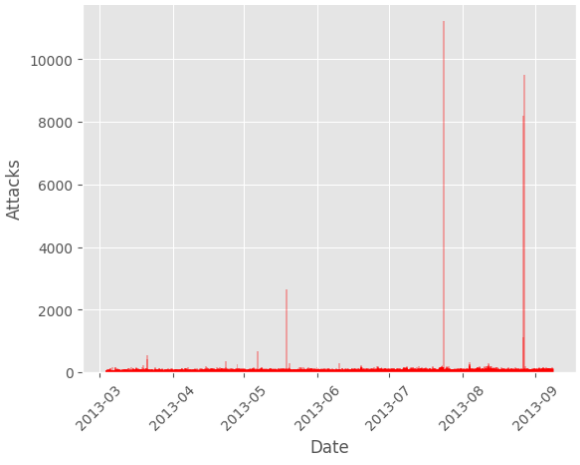

Figure 2. Attack by Date

For most of the period, the number of attacks remains consistently low, generally below 2,000 attacks per hour, indicating a relatively stable environment with minimal attack activity. However, based on the analysis, a notable increase in attacks was observed in May, with over 2,000 incidents recorded. This was

followed by a decline in June and July. In August, there was a dramatic spike, where the number of attacks surged to 11,216 per hour, marking the most significant event in the analyzed period. After this peak, the number of attacks decreased slightly but remained high in September, with over 9,000 attacks recorded. These intense spikes suggest major attack incidents, likely large-scale Distributed Denial of Service (DDoS) attacks or similar coordinated efforts, significantly impacting the network. The overall pattern highlights the fluctuating nature of cyberattacks, with critical periods of heightened vulnerability in May and August, contrasting sharply with the generally lower attack frequency throughout the rest of the period.

Based on the analysis of Figure 3, it is evident that the average values for Break-In attempts remain relatively stable, demonstrating a consistent fluctuation pattern over time. However, significant spikes in activity were observed on specific dates, including May 18th, July 24th, and August 26th. These sudden increases suggest heightened security vulnerabilities or targeted attack efforts on those dates, which may require further investigation to understand the underlying causes and implications. Such fluctuations are critical for identifying periods of elevated risk and for refining security strategies in response to these potential threats.

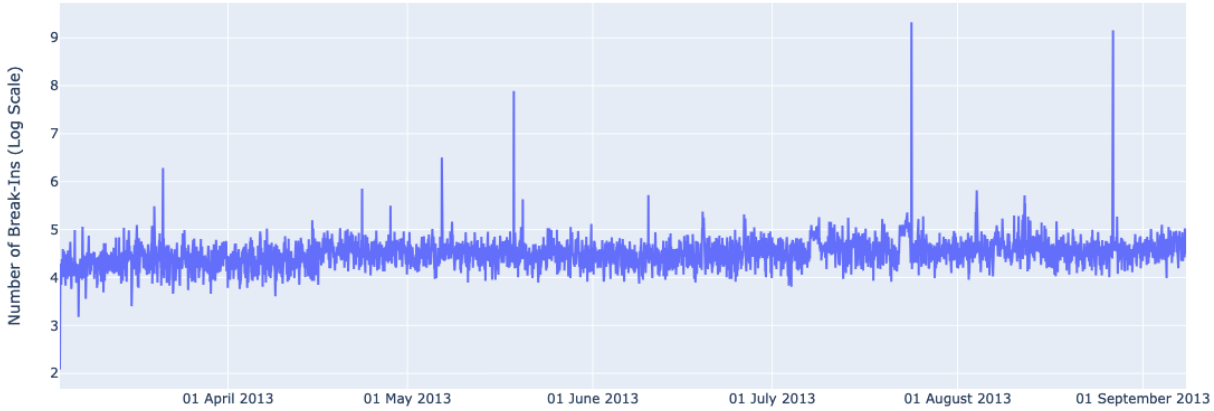

Figure 3. Time Series of Break-In Attempts

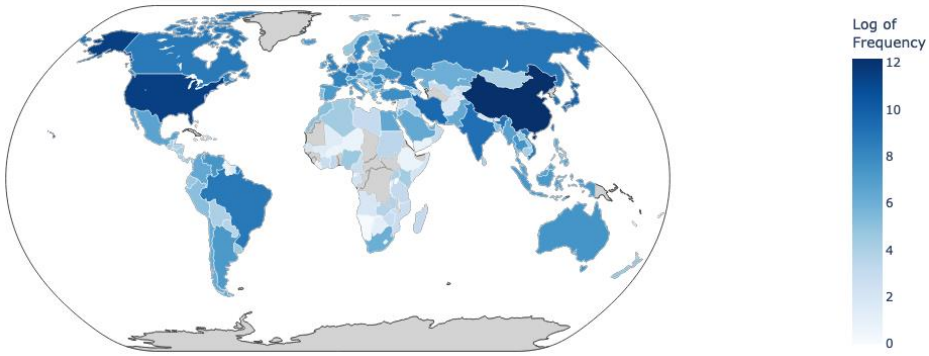

Figure 4. Attack location by Country

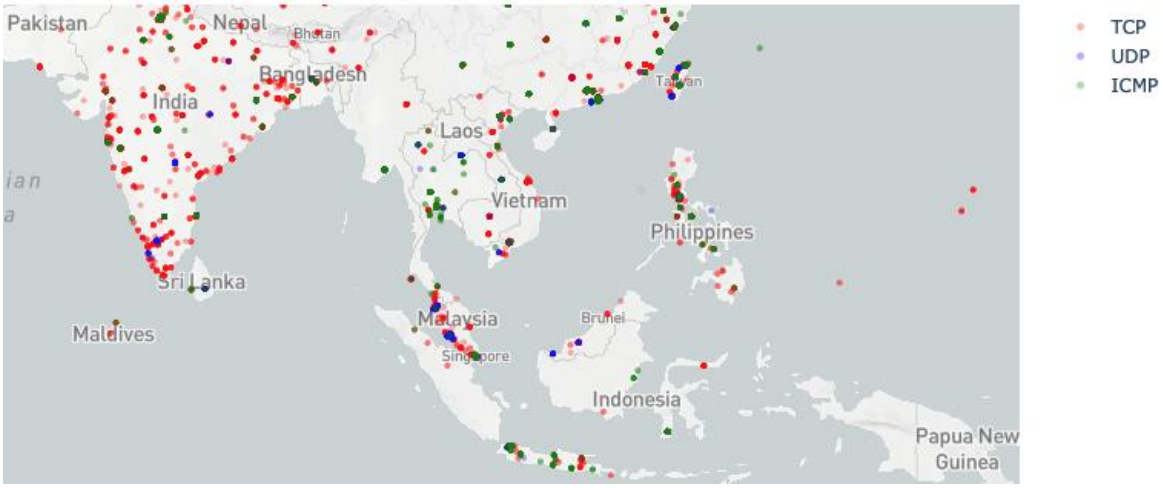

Figure 5. Detail Type by Location

The analysis of attack patterns show in Figure 4 attack reveals that China is the predominant source of cyberattacks, contributing a total of 191,394 incidents. This is followed by the United States, which accounts for 89,941 attacks. Japan ranks third with 17,204 attacks, while Iran and Taiwan reported 13,042 and 12,148 attacks, respectively. Other countries contributing significantly include the Netherlands (10,739), India (9,418), South Korea (9,316), Vietnam (7,826), and Russia (7,211). The attack traffic observed comprises three main types: TCP, UDP, and ICMP, each contributing to different dimensions of the threat landscape show in Figure 4 and Figure 5 for detail type each location. Conversely, countries such as Andorra, Bermuda, Barbados, Namibia, and Cape Verde have reported the lowest attack incidences, each contributing fewer than five cases. This stark contrast between the highest and lowest attack sources provides insights into the global distribution of cyber threats, underlining the necessity for targeted security strategies in high-risk regions while acknowledging the global nature of cybersecurity challenges.

3.2 Result Prediction

The K-Means clustering analysis yielded three distinct clusters, each representing different characteristics of the observed attacks, as shown in Figure 6. The Green Cluster highlights a concentrated area of attack origins, indicating high activity from a specific region or network, likely due to vulnerabilities or coordinated efforts. This localized nature makes it essential to develop targeted mitigation strategies. In contrast, the Purple Cluster shows a more decentralized attack pattern, suggesting multiple regions or networks as sources, requiring broader security responses. The Yellow Cluster, while also region-specific, reveals unique attack behaviors that may reflect distinct techniques or motives, necessitating specialized defenses. The strong clustering performance, supported by a Silhouette Score of 0.606 and a Davies-Bouldin Index of 0.614, confirms that these distinct patterns offer valuable insights for adaptive security measures, particularly in environments using SDN-enabled firewalls.

However, several key challenges must be addressed to fully leverage these insights. Potential biases in the dataset—such as geographic or attack type bias from uneven honeypot deployments—could skew the analysis. Scalability issues arise with increasing data volumes, affecting storage, processing, and real-time response capabilities. Additionally, real-time implementation poses difficulties, including the need for low-latency data streaming, efficient threat detection, and timely alerts. To overcome these challenges, the data pipeline must be optimized, algorithms improved for efficiency, and dynamic anomaly detection models integrated to ensure scalability and responsiveness in real-time applications.

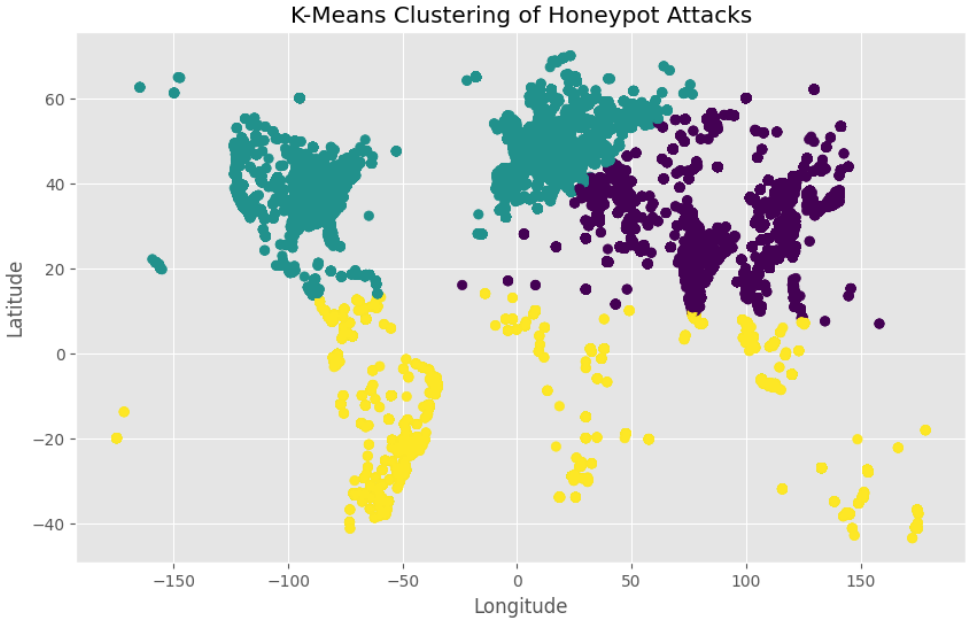

Figure 6 Result K-Means Clustering

4. Conclusions

In conclusion, the K-Means clustering analysis successfully identified distinct attack patterns across three key clusters, each representing unique characteristics in terms of geographical origin and attack behavior. The Green Cluster, which displayed a high concentration of attacks, pointed to a specific region or network that is particularly vulnerable or subjected to coordinated malicious activities. This finding suggests the presence of concentrated cyber threats, potentially due to exploitable vulnerabilities or orchestrated efforts by malicious actors. Targeting such a region with focused security measures and localized threat mitigation strategies would be crucial in reducing the attack surface within this hotspot. The Purple

Cluster, by contrast, showed a more dispersed attack pattern, indicating a wider range of attack origins spanning multiple regions or networks. This decentralized nature of attacks poses a different challenge, requiring broader and more adaptable security responses, as these threats do not emanate from a single source. Such a diverse origin profile could indicate more sophisticated, distributed attack campaigns or botnet-based threats, which necessitate global monitoring and scalable defenses. The Yellow Cluster exhibited a distinct set of characteristics, revealing attacks originating from a specific geographic region but distinguished by unique behavioral patterns or techniques. This suggests the attackers in this cluster may have different motives or methods compared to those in the other clusters. Understanding these unique behaviors allows for the development of specialized defenses that cater to the particular tactics employed by this group. The robustness of the clustering was confirmed through evaluation metrics, with a Silhouette Score of 0.606 and a Davies-Bouldin Index of 0.614, both indicating good clustering performance. These metrics suggest that the clustering model effectively separated distinct groups while maintaining cohesion within each cluster. The insights gained from this analysis are particularly valuable for designing SDN-enabled firewall systems, which can dynamically adjust to the geographic and protocol-based attack clusters. This allows for a more nuanced and responsive approach to network security, tailored to the specific characteristics of the threats identified. The implications of the identified clusters for network defense strategies are crucial. The Green Cluster, with its high concentration of attacks, necessitates proactive measures such as localized security enhancements and vulnerability assessments to counter coordinated threats. Conversely, the Purple Cluster's dispersed attack patterns require a scalable defense approach, utilizing global threat intelligence and adaptive security architectures like Software-Defined Networking (SDN) to effectively manage threats from multiple sources. Lastly, the Yellow Cluster indicates the need for tailored defenses, employing advanced analytics to understand and respond to the unique tactics of attackers in this group. By implementing these strategies, organizations can enhance their resilience against the diverse threats highlighted by the clustering analysis.

References

[1] P. Sharma and H. Gupta, “Emerging Cyber Security Threats and Security Applications in Digital Era,” in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida: IEEE, May 2024.

[2] Ö. Aslan, S. S. Aktuğ, M. Ozkan-Okay, A. A. Yilmaz, and E. Akin, “A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions ,” *Electronics (Basel)*, vol. 12, no. 6, Mar. 2023.

[3] A. Sharma, B. B. Gupta, A. Kumar Singh, and S. V. K., “Advanced Persistent Threats (APT): evolution, anatomy, attribution and countermeasures,” *J Ambient Intell Humaniz Comput*, vol. 14, pp. 9355–9381, May 2023.

[4] M. Shokry a, A. Ismail Awad b c d e, M. Khaled Abd-Ellah f, and A. A.M. Khalaf, “Systematic survey of advanced metering infrastructure security: Vulnerabilities, attacks, countermeasures, and future vision,” *Future Generation Computer Systems*, vol. 136, pp. 358–377, Nov. 2022.

[5] Y. Perwej, S. Q. Abbas, J. Pratap Dixit, N. Akhtar, and A. K. Jaiswal, “A Systematic Literature Review on the Cyber Security,” *International Journal of Scientific Research and Management*, vol. 9, no. 12, pp. 669–710, 2021.

[6] A. Annarelli, F. Nonino, and G. Palombi, “Understanding the management of cyber resilient systems,” *Comput Ind Eng*, vol. 149, Nov. 2020.

[7] R. Mazzolin and A. Madni Samueli, “A Survey of Contemporary Cyber Security Vulnerabilities and Potential Approaches to Automated Defence,” in *2020 IEEE International Systems Conference (SysCon)*, Montreal: IEEE, Dec. 2020.

[8] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, “A Survey on Machine Learning Techniques for Cyber Security in the Last Decade,” *IEEE Access*, pp. 222310–222354, Nov. 2020.

[9] U. Ikechukwu Okoli, O. Chimezie Obi, A. Okechukwu Adewusi, and T. Oluwaseun Abrahams, “Machine learning in cybersecurity: A review of threat detection and defense mechanisms,” *World Journal of Advanced Research and Reviews*, vol. 21, no. 1, pp. 2286–2295, 2024.

[10] F. Akowuah and F. Kong, “Real-Time Adaptive Sensor Attack Detection in Autonomous Cyber-Physical Systems,” in *2021 IEEE 27th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, Nashville: IEEE, Jul. 2021.

[11] R. Sultana, J. Grover, and M. Tripathi, “Security of SDN-based vehicular ad hoc networks: State-of-the-art and challenges,” *Vehicular Communications*, vol. 27, p. 100284, Jan. 2021, doi: 10.1016/j.vehcom.2020.100284.

[12] S. Ahmad and A. H. Mir, “Scalability, Consistency, Reliability and Security in SDN Controllers: A Survey of Diverse SDN Controllers,” *Journal of Network and Systems Management*, vol. 29, no. 1, p. 9, Jan. 2021, doi: 10.1007/s10922-020-09575-4.

[13] R. Deb and S. Roy, “A comprehensive survey of vulnerability and information security in SDN,” *Computer Networks*, vol. 206, p. 108802, Apr. 2022, doi: 10.1016/j.comnet.2022.108802.

[14] O. Yurekten and M. Demirci, “SDN-based cyber defense: A survey,” *Future Generation Computer Systems*, vol. 115, pp. 126–149, Feb. 2021, doi: 10.1016/j.future.2020.09.006.

[15] A. Shaghaghi, M. A. Kaafar, R. Buyya, and S. Jha, “Software-Defined Network (SDN) Data Plane Security: Issues, Solutions, and Future Directions,” in *Handbook of Computer Networks and Cyber Security*, Cham: Springer International Publishing, 2020, pp. 341–387. doi: 10.1007/978-3-030-22277-2_14.

[16] A.-D. Tudosi, A. Graur, D. G. Balan, and A. D. Potorac, “Research on Security Weakness Using Penetration Testing in a Distributed Firewall,” *Sensors*, vol. 23, no. 5, p. 2683, Mar. 2023, doi: 10.3390/s23052683.

[17] M. Rahouti, K. Xiong, Y. Xin, S. K. Jagatheesaperumal, M. Ayyash, and M. Shaheed, “SDN Security Review: Threat Taxonomy, Implications, and Open Challenges,” *IEEE Access*, vol. 10, pp. 45820–45854, 2022, doi: 10.1109/ACCESS.2022.3168972.

[18] J. C. Correa Chica, J. C. Imbachi, and J. F. Botero Vega, “Security in SDN: A comprehensive survey,” *Journal of Network and Computer Applications*, vol. 159, p. 102595, Jun. 2020, doi: 10.1016/j.jnca.2020.102595.

[19] J. Cunha *et al.*, “Enhancing Network Slicing Security: Machine Learning, Software-Defined Networking, and Network Functions Virtualization-Driven Strategies,” *Future Internet*, vol. 16, no. 7, p. 226, Jun. 2024, doi: 10.3390/fi16070226.

[20] A. M. Rifai, S. Raharjo, E. Utami, and D. Ariatmanto, “Analysis for diagnosis of pneumonia symptoms using chest X-ray based on MobileNetV2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (CLAHE),” *Biomed Signal*

Process Control, vol. 90, p. 105857, Apr. 2024, doi: 10.1016/j.bspc.2023.105857.

[21] T. Han *et al.*, "A comprehensive survey of security threats and their mitigation techniques for next-generation SDN controllers," *Concurr Comput*, vol. 32, no. 16, Aug. 2020, doi: 10.1002/cpe.5300.

[22] K. Kallepalli and U. B. Chaudhry, "Intelligent Security: Applying Artificial Intelligence to Detect Advanced Cyber Attacks," 2021, pp. 287–320. doi: 10.1007/978-3-030-87166-6_11.

[23] R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, "Machine Learning Models for Secure Data Analytics: A taxonomy and threat model," *Comput Commun*, vol. 153, pp. 406–440, Mar. 2020, doi: 10.1016/j.comcom.2020.02.008.

[24] Md. Zubair, MD. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Annals of Data Science*, Jun. 2022, doi: 10.1007/s40745-022-00428-2.

[25] O. I. Al-Sanjary, M. A. Bin Roslan, R. A. A. Helmi, and A. A. Ahmed, "Comparison and Detection Analysis of Network Traffic Datasets Using K-Means Clustering Algorithm," *Journal of Information & Knowledge Management*, vol. 19, no. 03, p. 2050026, Sep. 2020, doi: 10.1142/S0219649220500264.

[26] A. Parizad and C. J. Hatziadoniu, "Cyber-Attack Detection Using Principal Component Analysis and Noisy Clustering Algorithms: A Collaborative Machine Learning-Based Framework," *IEEE Trans Smart Grid*, vol. 13, no. 6, pp. 4848–4861, Nov. 2022, doi: 10.1109/TSG.2022.3176311.

[27] B. Shreeve, C. Gralha, A. Rashid, J. Araújo, and M. Goulão, "Making Sense of the Unknown: How Managers Make Cyber Security Decisions," *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 4, pp. 1–33, Jul. 2023, doi: 10.1145/3548682.

[28] M. Arunkumar and K. Ashok Kumar, "Malicious attack detection approach in cloud computing using machine learning techniques," *Soft comput*, vol. 26, no. 23, pp. 13097–13107, Dec. 2022, doi: 10.1007/s00500-021-06679-0.

[29] M. Shakil, A. Fuad Yousif Mohammed, R. Arul, A. K. Bashir, and J. K. Choi, "A novel dynamic framework to detect DDoS in SDN using metaheuristic clustering," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 3, Mar. 2022, doi: 10.1002/ett.3622.

[30] D. Bringhenti, G. Marchetto, R. Sisto, F. Valenza, and J. Yusupov, "Automated Firewall Configuration in Virtual Networks," *IEEE Trans Dependable Secure Comput*, vol. 20, no. 2, pp. 1559–1576, Mar. 2023, doi: 10.1109/TDSC.2022.3160293.

[31] M. Landauer, F. Skopik, M. Wurzenberger, and A. Rauber, "System log clustering approaches for cyber security applications: A survey," *Comput Secur*, vol. 92, p. 101739, May 2020, doi: 10.1016/j.cose.2020.101739.

[32] M. Landauer, F. Skopik, M. Wurzenberger, and A. Rauber, "System log clustering approaches for cyber security applications: A survey," *Comput Secur*, vol. 92, p. 101739, May 2020, doi: 10.1016/j.cose.2020.101739.

[33] J. Jacobs and B. Rudis, "DDS Dataset Collection Honeypots," Mar. 2014.

[34] A. A. J. Al-Abadi, M. B. Mohamed, and A. Fakhfakh, "Enhanced Random Forest Classifier with K-Means Clustering (ERF-KMC) for Detecting and Preventing Distributed-Denial-of-Service and Man-in-the-Middle Attacks in Internet-of-Medical-Things Networks," *Computers*, vol. 12, no. 12, p. 262, Dec. 2023, doi: 10.3390/computers12120262.

[35] M. S. El Sayed, N.-A. Le-Khac, M. A. Azer, and A. D. Jurcut, "A Flow-Based Anomaly Detection Approach With Feature Selection Method Against DDoS Attacks in SDNs," *IEEE Trans Cogn Commun Netw*, vol. 8, no. 4, pp. 1862–1880, Dec. 2022, doi: 10.1109/TCCN.2022.3186331.

[36] A. O. Salau and M. M. Beyene, "Software defined networking based network traffic classification using machine learning techniques," *Sci Rep*, vol. 14, no. 1, p. 20060, Aug. 2024, doi: 10.1038/s41598-024-70983-6.

[37] L. Wang, J. Yang, X. Xu, and P.-J. Wan, "Mining Network Traffic with the *k* -Means Clustering Algorithm for Stepping-Stone Intrusion Detection," *Wirel Commun Mob Comput*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6632671.

[38] A. Fahim, "K and starting means for k-means algorithm," *J Comput Sci*, vol. 55, p. 101445, Oct. 2021, doi: 10.1016/j.jocs.2021.101445.

[39] H. Kim, H. K. Kim, and S. Cho, "Improving spherical k-means for document clustering: Fast initialization, sparse centroid projection, and efficient cluster labeling," *Expert Syst Appl*, vol. 150, p. 113288, Jul. 2020, doi: 10.1016/j.eswa.2020.113288.

[40] Y. Liu, S. Ma, and X. Du, "A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Initial Cluster Centroids of K-means," *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/ACCESS.2020.3044069.

[41] R. Kusumastuti, E. Bayunanda, A. M. Rifa'i, M. R. G. Asgar, F. I. Ilmawati, and K. Kusrini, "Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC)," *CogITo Smart Journal*, vol. 8, no. 2, pp. 501–513, Dec. 2022, doi: 10.31154/cogito.v8i2.438.501-513.

