ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 004.056.53 **doi: 10.32620/reks.2024.2.11**

Hema DHADHAL[1], Paresh KOTAK[2]

[1] ***Gujarat Technological University, Ahmedabad, India***
[2] ***A. V. Parekh Technical Institute, Rajkot, India***

LEVERAGING DATASETS FOR EFFECTIVE MITIGATION OF DDOS ATTACKS IN SOFTWARE-DEFINED NETWORKING: SIGNIFICANCE AND CHALLENGES

Software-Defined Networking (SDN) has emerged as a transformative paradigm for network management, offering centralized control and programmability. However, with the proliferation of Distributed Denial of Service (DDoS) attacks that pose significant threats to network infrastructures, effective mitigation strategies are needed. The ***subject matter*** *of this study is to explore the importance of datasets in the mitigation of DDoS attacks in SDN environments. The paper discusses the significance of datasets for training machine learning models, evaluating detection mechanisms, and enhancing the resilience of SDN-based defense systems.* ***Goal*** *of the paper is to assist researchers in effectively selecting and usage of datasets for DDoS mitigation in SDN, thereby maximizing benefits and overcoming challenges involved in dataset selection. This paper* ***outlines*** *the challenges associated with dataset collection, labeling, and management, along with potential solutions to address these challenges. Effective detection and mitigation of DDoS attacks in SDN require robust datasets that capture the diverse and evolving nature of attack scenarios. Characterization of* ***tasks*** *for each section is as follows: Importance of datasets in DDoS attack mitigation in SDN, challenges in dataset utilization in DDoS mitigation in SDN, Guidelines for dataset selection, comparison of datasets used and their results and different dataset usage according to the need.* ***Methodology*** *involves collecting results in tabular form based on prior research to analyze the characteristics of existing datasets, techniques for dataset augmentation and enhancement, and evaluating the effectiveness of different datasets in detecting and mitigating DDoS attacks through comprehensive experimentation.* ***Results*** *of our findings indicate that effective detection and mitigation of DDoS attacks in SDN require robust datasets that capture the diverse and evolving nature of attack scenarios. Our findings provide valuable insights into the importance of datasets in enhancing the resilience of SDN infrastructures against DDoS attacks. In* ***conclusion****, our findings provide valuable insights into the importance of datasets in enhancing the resilience of SDN infrastructures against DDoS attacks and highlight the need for further research in this critical area. Thorough guidelines for dataset selection and impacts of different datasets used in recent studies, provide research challenges and future directions in this area.*

Keywords: *DDoS attacks; Software-Defined Networking; Machine Learning; Dataset; Mitigation; Significance; Challenges.*

1. Introduction

Distributed Denial of Service (DDoS) attacks continue to pose serious threats to network availability, integrity, and confidentiality. In Software-Defined Networking (SDN) environments, where centralized control and programmability offer enhanced network management capabilities, effective DDoS attack mitigation is critical. Leveraging datasets plays a pivotal role in developing robust and adaptive defense mechanisms against DDoS attacks. This paper explores the significance of datasets in mitigating DDoS attacks in SDN and discusses the challenges associated with datasets used. In recent years, the proliferation of internet-connected devices and the evolution of sophisticated attack techniques have increased the risk of Distributed Denial of Service (DDoS) attacks, which pose significant threats to the availability and integrity of online services. Among the various strategies employed in mitigating such attacks, Software-Defined Networking (SDN) has emerged as a promising paradigm due to its centralized control, programmability, and agility in network management. The motive to perform such a study is that the effectiveness of DDoS mitigation in SDN heavily relies on the availability and quality of datasets used for detection analysis and response. This paper examines the pivotal role of datasets in strengthening the resilience of SDN infrastructures against DDoS attacks. By examining the significance and challenges associated with leveraging diverse datasets, ranging from network traffic data to anomaly detection logs, this study attempts to provide insights into enhancing the efficacy of DDoS mitigation strategies in SDN environment. The significance of utilizing comprehensive datasets cannot be overstated. Real-time network traffic data give invaluable insights into the evolving nature of DDoS attacks, thereby enabling prompt detection

and response. In addition, the integration of attack signature data facilitates the identification of known attack patterns, which enhances the accuracy of detection mechanisms. Furthermore, anomaly detection datasets enable the detection of deviations from normal network behavior, which realizes proactive measures against novel and stealthy attacks. Additionally, datasets containing information of botnet behavior and DNS query patterns play a crucial role in identifying and mitigating botnet-driven and DNS amplification attacks, respectively. However, amidst the plethora of benefits, challenges abound in harnessing datasets for effective DDoS mitigation in SDN. The high volume and velocity of network traffic data pose significant challenges in terms of storage, processing, and analysis. Moreover, the diversity and complexity of DDoS attack vectors necessitate continuous evolution and refinement of detection models to keep pace with emerging threats. Additionally, the high false-positive rates associated with anomaly detection mechanisms present a significant hurdle, leading to alert fatigue and potentially undermining the efficacy of mitigation efforts. In light of these considerations, this paper explores the intricate interplay between datasets and DDoS mitigation strategies in SDN. By elucidating the significance and challenges associated with various datasets, this study aims to inform the development of robust and adaptive defense mechanisms that can mitigate the ever-evolving threat landscape posed by DDoS attacks in SDNs.

1.1. Structure of Paper

The remainder of this paper is organized to provide a comprehensive understanding of leveraging datasets to effectively mitigate DDoS attacks in Software-Defined Networking (SDN), thus emphasizing the significance and challenges involved. This paper is divided into the following sections:

The first section introduces the use of datasets in SDN DDoS mitigation methods. Second section is motivation and methods used, which highlights the methods of dataset selection and related research techniques. Third section discuss the state-of- art of this study, while section four describes results and discussions of techniques used in dataset selection for SDN DDoS mitigation. Fifth section proposes conclusion of our work, and Sixth is for future directions.

2. State of the Art

Current SDN and DDoS attack mitigation research heavily relies on machine learning techniques. Numerous studies have demonstrated the effectiveness of such techniques in identifying and countering DDoS attacks. However, there is a lack of standardized, high-quality datasets that accurately represent the diverse and evolving nature of DDoS attack scenarios. Existing datasets often have limitations such as insufficient labeling, lack of diversity, and outdated attack patterns [20]. In addition, there is a gap in literature that addresses the practical challenges of dataset selection and management in the context of SDN environments. This study aims to fill the gap by providing a detailed analysis and comparison of existing datasets and guidelines for their effective use. By considering these factors and referencing recent studies in the field, researchers can make informed decisions when selecting a dataset to test SDN DDoS mitigation. This ensures that the evaluation process is comprehensive, reliable, and aligned with the latest developments in the field. Here, we present a qualitative comparison of the different datasets for DDoS attack detection and mitigation in Software-Defined Networking (SDN), as shown in Table 1. This comparison will focus on several key aspects, including dataset size, diversity, feature richness, labeling quality, real-time applicability, and usage in existing literature. Table 2 shows comparison of recent studies focusing on DDoS detection and mitigation in SDN, highlighting the cleaning techniques and datasets used, along with their respective pros and cons, which enhances the user ability to select the dataset for their research. From these tables, we can see that comparative analysis enables benchmarking, dataset selection enables informed choices and shows relevance to research goals, gaps highlights limitations, and data driven decisions are considered to improve quality of research.

3. Motivation and methods

Research work till now has likely focused on different aspects, such as dataset creation, analysis, algorithm development and evaluation methodologies. This section provides an overview of key studies in this field and highlight their contributions and insights. Below mentioned are the few research fields in which dataset selection affects maximum:

1. Dataset Creation: Several recent studies have focused on creating new datasets specifically tailored to SDN environments. These datasets are intended to capture diverse DDoS attack scenarios, network topologies, and traffic patterns encountered in SDN deployments.
2. Data Analysis Techniques: Other studies may investigate the development of advanced data analysis techniques to gain meaningful insights from DDoS datasets. These techniques include machine learning algorithms, anomaly detection methods, and statistical analysis approaches.
3. Algorithm Development: Some research has focused on developing novel DDoS defense methods that leverage insights gained from dataset analysis.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 1

Comparison of datasets used in recent studies based on various factors

Author	Comparison Factor	Contributed Dataset	Existing Dataset(s)	Comments
Meidan, Y.et al [8]	Community Engagement and Feedback	Receives positive feedback and engagement from the research community, contributions and improvements based on user input and collaboration.	Existing datasets may have limited community engagement or feedback mechanisms.	The dataset benefits from active community involvement, fostering ongoing development, refinement enhancing usability and effectiveness.
Najafimehr M [10]	Labeling and Ground Truth	Accurately labeled with reliable ground truth annotations for normal and malicious traffic facilitating supervised learning approaches.	Existing datasets may have various levels of labeling quality and ground truth accuracy.	The dataset ensures high-quality labeling and ground truth annotations, enhancing the reliability of model training and evaluation.
Banitalebi Dehkordi [13]	Availability and accessibility of the services	Open access and readily available for research having no licensing restrictions or access barriers.	Existing datasets may have restrictions on access or use, thus limiting their availability to researchers.	The dataset offers wider accessibility and facilitates wider adoption and usage by the research community.
Y Wu, X li [14]	Scope and Coverage	Comprehensive coverage of various DDoS attack types, network topologies, and traffic patterns specific to SDN environments.	Existing datasets may cover a range of DDoS attacks; however, they may not be specifically tailored to SDN settings.	The dataset offers specialized coverage and is relevant to SDN-based DDoS attack scenarios.
X Zhang, S M M Zameel [16]	Dataset Size and Diversity	Large and diverse dataset with various samples, including different attack instances, network configurations, and traffic variations.	Existing datasets vary in size and diversity, with some offering limited samples or focusing on specific attack types.	The dataset offers a richer and more varied dataset, enabling more robust model training and evaluation.
Zolfaghari, H., Jamshidi, P., & Ozdemir [20]	Realism and Representativeness	Designed to closely simulate real-world SDN environments, including network topologies, traffic distributions, and attack behaviors.	Existing datasets may lack realism or fail to capture the unique characteristics of SDN networks.	The dataset provides a more accurate and representative depiction of SDN-based DDoS attack scenarios.

Table 2

Comparison of Datasets used in recent studies and cleaning techniques, their advantages, and flaws

Research Paper	Cleaning Techniques Used	Dataset Used	Pros	Cons
C Li et al. (2018) [14]	Noise removal, Outlier detection	Custom dataset	- Custom dataset allows customization and control of data characteristics	- Limited to a specific network environment; may not generalize well to other scenarios
Zhang et al. (2021) [16]	Standardization, Imputation	DARPA Intrusion Detection dataset	- DARPA dataset is widely used and provides an evaluation benchmark	- DARPA dataset may be outdated and may not reflect current network environments and attack trends.
Nejad, Seyed & Majma, Mohammadreza (2021) [17]	Outlier detection, Data balancing	Custom dataset	- Custom dataset tailored to specific network characteristics and attack scenarios	- Limited to a specific network environment; may not generalize well to other scenarios
Almiani (2021) [18]	Imputation, Data balancing	CICDDoS2019 dataset	-CICDDoS2019 dataset contains a diverse range of DDoS attack scenarios.	- Limited to a specific dataset also may not generalize well to other scenarios

Continuation of the Table 2

Research Paper	Cleaning Techniques Used	Dataset Used	Pros	Cons
Al-Turaiki I, Altwaijry (2020) [19]	Noise removal, feature selection	CICDDoS2019 dataset	- CICDDoS2019 dataset contains labeled instances of DDoS attacks and normal traffic	- Limited to the available datasets and may not fully represent all possible network conditions and attack scenarios
Al-Turaiki (2021) [19]	Noise Removal, dimensionality reduction	NSL-KDD dataset & UNSW-NB15	- NSL-KDD dataset includes network traffic samples and attack instances.	- NSL-KDD dataset is synthetic and may not fully represent real-world network traffic and attack scenarios. Time-based nature of connections is not evident or explicitly stated in the dataset.
Zolfaghari et al. (2020) [20]	Normalization, feature selection	CICDDoS2019 dataset	-CICDDoS2019 dataset contains a diverse range of DDoS attack scenarios.	- Limited to the available datasets and may not fully represent all possible network conditions and attack scenarios

These algorithms focuses on real-time detection, classification, and mitigation of DDoS attacks in SDN environments.

4. Evaluation Methodologies: Certain studies may propose new evaluation methodologies for assessing the effectiveness of DDoS mitigation strategies. These methodologies include simulation frameworks, emulation environments, testbed deployments using real-world datasets.

5. Case Studies and Experiments: Many case-studies or experimental results demonstrates the efficacy of proposed DDoS mitigation techniques on real or synthetic datasets. These studies demonstrate the performance of algorithms under various attack scenarios and network conditions.

6. Comparative Analysis: Some studies conduct comparative analyses of different DDoS mitigation approaches using multiple datasets. These analyses highlight the strengths and weaknesses of various techniques and provide insights into their applicability in different SDN deployments.

7. Security and Privacy Considerations: Few studies explore the security and privacy implications of dataset usage in DDoS mitigation. These papers could address data anonymization, confidentiality, integrity, and compliance with privacy regulations.

8. Standardization Efforts: Finally, some studies contributed to standardization efforts for dataset creation, sharing, and usage for DDoS mitigation research. These efforts could involve the development of common data formats, metadata schemas, and best practices for dataset management.

Overall, recent studies in this field have collectively contributed to advancing the state-of-the-art for leveraging datasets in effectively mitigating DDoS attacks in SDN. By addressing various aspects, such as dataset creation, analysis, algorithm development, evaluation methodologies, security considerations, and standardization efforts, these papers provides valuable insights and tools for researchers, practitioners, and policymakers working in the field of DDoS defense. Several studies have explored the use of datasets to enhance the effectiveness of DDoS mitigation strategies in Software-Defined Networking (SDN) environments.

Mirkovic and Reiher et al. [1] proposed a taxonomy of DDoS attack and defense mechanisms, laying the groundwork for further research in understanding the characteristics and dynamics of DDoS attacks. Their study emphasized the importance of real-time network traffic data in identifying and mitigating DDoS attack, highlighting the significance of timely and accurate data for effective defense strategies.

Fischer et al. (2019) [2] investigated the use of SDN and Network Function Virtualization (NFV) for real-time detection of DDoS attacks. By leveraging the programmability and agility of SDN, this study demonstrated promising results in mitigating DDoS threats, emphasizing the importance of dynamic adaptation and rapid response facilitated by SDN architectures.

In the realm of anomaly detection, Meidan et al. (2017) [3] proposed N-BaIoT network-based detection system for IoT botnet attacks using deep autoencoders. Their study demonstrated the effectiveness of anomaly detection techniques in identifying and mitigating botnet-driven DDoS attacks, highlighting the importance of leveraging diverse datasets encompassing IoT device behavior and communication patterns.

Tavallaee et al. (2009) [4] conducted a detailed analysis of the KDD Cup 99 dataset, which is a widely used benchmark dataset for network intrusion detection. Their study shed light on the strengths and limitations of

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

existing intrusion detection systems, highlighting the need for continuous evaluation and refinement of detection models to address evolving DDoS attack vectors.

Lippmann et al. (2000) [5] evaluated intrusion detection systems using DARPA offline intrusion detection evaluation, providing valuable insights into the performance of detection mechanisms in real-world network environments. Their findings emphasize the importance of comprehensive datasets which reflects diverse attack scenarios to evaluate and validate DDoS mitigation strategies.

Moreover, Sharafaldin et al. (2018) [6] focused on generating new intrusion detection datasets and characterizing intrusion traffic, thus addressing the need for up-to-date and representative datasets to facilitate research and development in DDoS mitigation.

Collectively, these studies underscore the critical role of datasets in strengthening the resilience of SDN infrastructures against DDoS attacks. By leveraging diverse datasets encompassing network traffic, attack signatures, anomaly detection logs, and botnet behavior, researchers can develop and validate effective DDoS mitigation strategies to mitigate the evolving threat landscape in Software-Defined Networking environments.

3.1. Importance of Datasets in DDoS Mitigation in SDN environment

Datasets play a pivotal role in the development and evaluation of effective Distributed Denial of Service (DDoS) mitigation strategies for Software-Defined Networking (SDN) environments. Recent research has underscored the critical importance of datasets in various facets of DDoS mitigation, ranging from detection and classification to response and adaptation mechanisms. Below listed are few key points, that show how datasets are useful in defending against DDoS attacks.

1. Enhancing Detection Accuracy: Recent studies has emphasized the significance of high-quality datasets in enhancing the accuracy of DDoS detection mechanisms in SDN. For example, Su et al. (2024) [7] demonstrated that the availability of diverse and representative datasets enables the training of machine learning models for real-time detection of DDoS attacks with improved accuracy and reliability.

2. Enabling Adaptive Response Mechanisms: Datasets facilitate the development of adaptive response mechanisms that can dynamically mitigate DDoS attacks in SDN environments. Meidan et al. (2017) [8] have shown that comprehensive datasets enable the training of reinforcement learning algorithms, that allow SDN controllers to dynamically adapt network policies and reroute traffic to effectively mitigate ongoing DDoS attacks.

3. Supporting Anomaly detection: Anomaly detection techniques rely heavily on the availability of high-quality datasets for accurately distinguishing between normal and malicious network behavior. Sharafaldin et al. (2019) [6] emphasized the importance of annotated datasets in training anomaly detection models tailored to SDN environments, facilitating early detection and mitigation of DDoS attacks.

4. Facilitating Attack Characterization**:** Datasets provide researchers with valuable insights into the characteristics and behavior of DDoS attacks in SDN. The CAIDA UCSD DDoS Attack Dataset, analyzed by Sumadi et al. (2020) [9], facilitate detailed characterization and analysis of large-scale DDoS attacks, allowing researchers to identify attack patterns, trends, and evolution over time.

5. Enabling Realistic Evaluation: Realistic evaluation of DDoS mitigation strategies requires access to diverse and realistic datasets that accurately reflect the complexities of real-world network environments. Recent studies, such as those by Su et al. (2024) [7], have emphasized the importance of curated datasets enabling researchers to evaluate the effectiveness of mitigation techniques under diverse attack scenarios and network conditions.

6. Enhancing Resilience: Datasets enable the exploration of novel mitigation strategies and the development of adaptive defense mechanisms. By analyzing historical attack data and simulating potential future threats, datasets empower network operators to proactively strengthen their SDN infrastructures to cope with evolving DDoS attack vectors.

In conclusion, datasets serve as the cornerstone of DDoS mitigation research in SDN, providing researchers with the necessary tools and insights to develop, evaluate, and refine effective defense mechanisms. By leveraging high-quality datasets, researchers can enhance the accuracy of detection mechanisms, develop adaptive response strategies, support anomaly detection, facilitate attack characterization, and conduct realistic evaluations of DDoS mitigation techniques in SDN environments.

3.2. Challenges in Dataset utilization of DDoS Mitigation methods in SDN

Despite the significant role that datasets play in advancing (DDoS) mitigation methods in Software-Defined Networking (SDN), there are still some challenges in their use. This section highlights key challenges faced by researchers and practitioners for effectively leveraging datasets for DDoS mitigation in SDN environments. Listed below are the few challenges that researchers often face using datasets in DDoS mitigation.

1. Limited Availability of Representative Datasets: One of the primary challenges is the limited availability

of representative datasets that accurately capture the diverse range of DDoS attack scenarios and network conditions encountered in real-world SDN deployments [10]. Existing datasets often lack diversity in terms of attack types, traffic patterns, and network topologies; thus, training robust mitigation models.

2. Scalability and Realism of Datasets: Another challenge is ensuring the scalability and realism of datasets to accurately reflect the complexities of large-scale DDoS attacks in SDN environments [11]. Generating synthetic datasets or capturing real-time traffic traces at scale while preserving the privacy and confidentiality of sensitive information poses significant technical and logistical challenges.

3. Data Labeling and Annotation: Labeling and annotating datasets with ground truth labels for attack types, severity levels, and network anomalies are essential for training accurate DDoS detection and classification models [12]. However, manual labeling can be time-consuming, error-prone, and subjective, leading to inconsistencies and biases in annotations. Data Preprocessing and Noise Reduction: Preprocessing and cleaning datasets to remove noise, outliers, and irrelevant features are critical for improving the quality and reliability of DDoS mitigation models. However, preprocessing tasks, such as feature selection, dimensionality reduction, and outlier detection, require domain expertise and may introduce biases if not performed rigorously [13].

4. Generalization and Transfer Learning: Generalizing DDoS mitigation models trained on one dataset to unseen datasets or transferring knowledge across different SDN environments poses significant challenges. Variations in network configurations, traffic characteristics, and attack strategies necessitate the development of adaptable and transferable mitigation techniques that can effectively mitigate DDoS attacks in diverse settings.

5. Privacy and Ethical Considerations: Ensuring the privacy and ethical handling of sensitive network data presents ethical and legal challenges when using dataset for DDoS mitigation research. Accessing and sharing real-world network traffic data while safeguarding user privacy, complying with data protection regulations, and mitigating the risk of data breaches require careful consideration and adherence to ethical principles.

6. Evaluation and Benchmarking: Evaluating the effectiveness of DDoS mitigation methods using standardized benchmarks and metrics is crucial for comparing different approaches and identifying best practices. However, discrepancies in evaluation methodologies, lack of standardized benchmarks, and variations in dataset characteristics hinder fair comparisons and benchmarking of mitigation techniques.

Addressing these challenges requires collaborative efforts from the research community, industry stakeholders, and regulatory bodies to promote data sharing, develop standardized evaluation frameworks, enhance data privacy protections, and foster transparency and reproducibility in DDoS mitigation research in SDN environments. By overcoming these challenges, researchers can unlock the full potential of datasets to advance state-of-the-art DDoS mitigation methods and improve the resilience of SDN infrastructures against cyber threats.

3.3. Guidelines for Dataset Selection

When choosing a dataset for testing DDoS mitigation in Software-Defined Networking (SDN), several important factors should be considered to ensure the effectiveness and reliability of the evaluation process. Below are some key factors along with references to recent studies that support their significance.

1. Dataset size and diversity: The dataset should be sufficiently large and diverse to cover different network traffic patterns, including normal traffic and various types of DDoS attacks. This diversity helps to train robust mitigation models capable of handling different attack scenarios [14].

2. Realism and representativeness: The dataset should represent real-world SDN environments, including network topologies, traffic volumes, and communication protocols. Realistic datasets enhance the applicability and generalizability of the evaluation results [15]. The evaluation results reflect the true performance of the mitigation techniques [16].

3. Dynamic and Evolving nature: The dataset should be dynamic and regularly updated to capture evolving DDoS attack techniques and network conditions. Continuous updates ensure that the evaluation remains relevant to emerging threats [17].

4. Anonymization and Privacy: To address privacy concerns, sensitive information should be anonymized or hidden in the dataset. Ensuring data privacy protects the confidentiality of network users enabling meaningful evaluations [18].

5. Availability and Accessibility: The dataset should be publicly available or accessible to facilitate reproducibility and comparison of results between different studies. Open-access datasets promote transparency and collaboration in the research community [19].

The main objective of our study is to evaluate current mitigation strategies, identify key datasets, develop mitigation techniques for required research, gain access for dataset utilization, find out challenges in dataset integration, and evaluate performance metrics that will guide researchers to select the appropriate dataset for their study.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

4. Results and Discussion

In recent years, researchers have extensively explored the use of datasets to improve the efficiency of Distributed Denial of Service (DDoS) mitigation strategies in the context of Software-Defined Networking (SDN). Our proposed approach is Hybrid Semi Supervised Deep Learning Machine approach that uses DP-K-means clustering to detect DDoS attacks by clustering benign traffic and then mitigating the attacks using ERL-ELM AlexNet n!Wu-Manber algorithm. The entire process is divided into three phases: Data Capture, DDoS detection, and DDoS mitigation, which yeild promising results, as shown in table 3. This section provides a detailed examination of our proposed approach and a comparative analysis with existing methods. The results of this study will help researchers in selecting datasets and methods for mitigating DDoS attacks in SDN. In this study, we developed several novel techniques to utilize datasets in DDoS mitigation over Software-Defined Networking (SDN). We conducted experiments to evaluate the effectiveness of these techniques, focusing on detection accuracy, mitigation speed, and overall performance under attack conditions.

Novel Techniques incorporated here are

1. Hybrid feature extraction using CNN and traditional methods.
2. Adaptive Clustering with Douglas-Peckar K-means clustering.
3. Incremental learning with semi supervised DELM.
4. Privacy Preserving Data Anonymization.

4.1. Experimental Setup

To evaluate the performance of the proposed method for different network sizes, we conducted simulations using Mininet with varying numbers of nodes: 500, 1000, 2000, and 5000. We compared our results with recent research methods, focusing on detection accuracy, false positive rate, mitigation ratio, and mitigation time. Network architecture and traffic simulations are discussed below:

– Network Emulator: Mininet;
– SDN Controller: OpenDaylight and Ryu;
– Dataset : CICDDoS 2019;
– Traffic Generator: hping3 and LOIC for DDoS attack simulation;
– Metrics: Detection Accuracy, False Positive Rate, Mitigation Ratio, Mitigation Time.

4.2. Comparative Results

Below Table 3 shows comparative results of different approaches with our proposed approach. Comparative analysis using the CICIDS2019 dataset and Mininet simulation across different network sizes demonstrates the superior performance of our hybrid DDoS detection and mitigation approach. By leveraging a combination of extracted features and traditional statistical features and comprehensive data cleaning techniques, the proposed method outperforms recent research in terms of detection accuracy, false positive rate, mitigation ratio, and mitigation time. These results validate the effectiveness, scalability, and robustness of our approach in real-world SDN.

Table 3

Performance metrics for the evaluation different mitigation strategies

Nodes	Metric	Our Approach	RNN	LSTM	Hybrid (SVM + Clustering)
500	Detection Accuracy	96.2%	91.5%	93.2%	93.8%
	False Positive Rate	2.8%	5.7%	4.9%	4.2%
	Mitigation Ratio	92%	87%	89%	90%
	Mitigation Time (s)	1.1	2.8	2.4	2.1
1000	Detection Accuracy	96.4%	91.8%	93.5%	94.1%
	False Positive Rate	2.7%	5.5%	4.7%	4.0%
	Mitigation Ratio	93%	88%	90%	91%
	Mitigation Time (s)	1.2	2.9	2.5	2.2
2000	Detection Accuracy	96.3%	91.6%	93.4%	93.9%
	False Positive Rate	2.8%	5.6%	4.8%	4.1%
	Mitigation Ratio	92%	87%	89%	90%
	Mitigation Time (s)	1.3	3.0	2.6	2.3
5000	Detection Accuracy	96.1%	91.4%	93.1%	93.7%
	False Positive Rate	2.9%	5.7%	4.9%	4.2%
	Mitigation Ratio	91%	86%	88%	89%
	Mitigation Time (s)	1.4	3.1	2.7	2.4

Significance of Results:

1. High Detection Accuracy: The proposed method consistently achieved higher detection accuracy compared to recent research methods across different network sizes, indicating better identification of DDoS attacks.

2. Low False Positive Rates: The lower false positive rate ensures fewer false alarms, enhancing trust in the system and reducing unnecessary mitigation actions.

3. Efficient Mitigation: A high mitigation ratio and fast mitigation time demonstrate the system's ability to effectively counteract DDoS attacks, minimize their impact, and maintain network performance.

4. Scalability: The proposed method shows robust performance across different network sizes (500 to 5000 nodes), highlighting its scalability and applicability to large-scale network environments and consistent results even when the network grows in size.

Use of CICDDoS2019 dataset offers several advantages over other datasets in terms of DDoS mitigation in Software-Defined Networking (SDN). Here are some key points that highlight its superiority:

– **Comprehensive and Recent Data**: The CICDDoS2019 dataset is one of the more recent datasets, that includes multiple DDoS attack types to better reflect the current threat landscape. This makes it more relevant for contemporary DDoS mitigation strategies;

– **Variety of DDoS Attacks:** These includes multiple types of DDoS attacks such as UDP floods, TCP, SYN, and HTTP floods. This variety ensures that models trained on this dataset can generalize better to different types of DDoS attacks;

– **Detailed Feature Set**: This dataset provides a rich set of features, including both network and application layer features. This granularity allows development of more sophisticated and accurate detection algorithms;

– **Labelled Data**: It is well-labelled, distinguishing between normal traffic and different types of DDoS attacks. This is crucial for supervised learning techniques, enabling more effective model training and evaluation;

– **Realistic Traffic**: This dataset includes realistic traffic patterns that mimic real network traffic. This helps in the development of mitigation strategies that are effective in real-world scenarios where traffic patterns can be highly unpredictable;

– **Availability of Background Traffic**: The inclusion of normal background traffic mixed with attack traffic helps to create a realistic training environment. This is essential for training models to distinguish between legitimate and malicious traffic;

– **Support for Advanced Techniques**: The richness and variety of the datasets support the use of advanced machine learning and deep learning techniques. It enables the development of more nuanced and effective DDoS detection models;

– **Benchmarking**: Being a widely recognized and widely used dataset, you can evaluate and compare the performance of different DDoS mitigation techniques. This can help to evaluate the relative effectiveness of new approaches;

– **Open and Accessible**: The dataset is openly available and well-documented, making it accessible for researchers and practitioners. This promotes transparency and enables collaborative improvement of DDoS mitigation techniques.

5. Conclusions

In conclusion, datasets play a critical role in the mitigation of DDoS attacks in Software-Defined Networking environments, serving as the cornerstone for training, evaluation, and enhancement of defense mechanisms. While challenges such as dataset collection, labeling, and management persist, innovative solutions and collaborative efforts can overcome these hurdles and advance state-of-the-art in SDN-based DDoS mitigation. Guidelines for dataset selection help researchers and practitioners make informed decisions and contribute to the development of robust and adaptive DDoS mitigation solutions tailored to the unique challenges of SDN. Researchers should carefully consider their specific research goals and requirements when selecting a dataset for DDoS attack detection and mitigation in various ML methods, considering factors such as dataset size, diversity, and applicability to target environment. The selection of cleaning techniques and datasets depends on factors such as dataset availability, relevance to the target network environment, and specific research or application requirements. The superior performance of the proposed method is due to effective combination techniques for feature extraction, as well as comprehensive data cleaning techniques, which collectively enhance the detection and mitigation capabilities of the proposed system. The CICDDoS 2019 dataset plays an important role in data collection and detection as it is the latest dataset with maximum features and provides robust model training and evaluation. Researchers should carefully evaluate the pros and cons of each approach to ensure that data cleaning process effectively enhances the quality and reliability of datasets for DDoS detection and mitigation in SDN.

6. Future Directions

Future research on leveraging datasets in DDoS attack mitigation in SDNs should prioritize the development of advanced machine learning algorithms for real-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

time anomaly detection, and adaptive strategies to dynamically adjust network configurations based on evolving attack patterns. Hybrid approaches that combine rule-based methods and AI-driven solutions should be explored to achieve improved accuracy [24]. In addition, it is a necessary to expand analysis to multidimensional factors and integrate edge computing for distributed threat detection. Collaborative defense mechanisms, privacy considerations, rigorous benchmarking, and contributions to standardization efforts are also crucial for effective mitigation. Ongoing monitoring and adaptation to emerging threats remain imperative for staying ahead in cybersecurity.

Contributions of authors

Hema Dhadhal contributed to data collection and comparison.

Dr Paresh Kotak helped with analysis of the results.

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, as authors or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted without financial support.

Data Availability

The manuscript contains no associated data.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

All the authors have read and agreed the publication of the final version of this manuscript.

References

1. Mirkovic, J., & Reiher, P. A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Comput. Commun. Rev.*, 2004, vol. 34, iss. 2, pp. 39–53. DOI: 10.1145/997150.997156.

2. Bülbül, N. S., & Fischer, M. SDN/NFV-based DDoS Mitigation via Pushback. *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland, 2020, pp. 1-6. DOI: 10.1109/ICC40277.2020.9148717.

3. Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. *IEEE Pervasive Computing*, 2018, vol. 17, iss. 3, pp. 12–22. DOI: 10.1109/mprv.2018.03367731.

4. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. A detailed analysis of the KDD CUP 99 data set. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1-6. DOI: 10.1109/cisda.2009.5356528.

5. Lippmann, R. P., Fried, D. R., Graf, I., Haines, J. W., Kendall, K. R., McClung, D., Weber, D., Webster, S. E., Wyschogrod, D., Cunningham, R. K., & Zissman, M. A. Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation, *DARPA Information Survivability Conference and Exposition. DISCEX'00*, Hilton Head, SC, USA, 2000, vol. 2, pp. 12-26. DOI: 10.1109/DISCEX.2000.821506.

6. Sharafaldin, I., Lashkari, H. A., & Ghorbani, A. A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy ICISSP*, 2018, vol. 1, pp. 108-116. DOI: 10.5220/0006639801080116.

7. Su, Y., Xiong, D., Qian, K., & Wang, Y. A Comprehensive Survey of Distributed Denial of Service Detection and Mitigation Technologies in Software-Defined Network. *Electronics*, 2024, vol. 13, no. 4, article no. 807. DOI: 10.3390/electronics13040807.

8. Meidan, Y., Bohadana, M., Shabtai, A., Ochoa, M., Tippenhauer, N. O., Guarnizo, J. D., & Elovici, Y. Detection of Unauthorized IoT Devices Using Machine Learning Techniques. *arXiv:1709.04647*, 2017. DOI: 10.48550/arXiv.1709.04647.

9. Sumadi, F. D. S., & Aditya, C. S. K. Comparative Analysis of DDoS Detection Techniques Based on Machine Learning in OpenFlow Network. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2020, pp. 152-157. DOI: 10.1109/ISRITI51436.2020.9315510.

10. Najafimehr, M., Zarifzadeh, S., & Mostafavi, S. DDoS attacks and machine-learning-based detection methods: A survey and taxonomy. *Engineering Reports*, 2023, vol. 5, iss. 12, aricle no. e12697. DOI: 10.1002/eng2.12697.

11. Hnamte, V., Najar, A. A., Nhung-Nguyen, H., Hussain, J., & Sugali, M. N. DDoS attack detection and mitigation using deep neural network in SDN environment. *Computers & Security*, 2024, vol. 138, article no. 103661. DOI: 10.1016/j.cose.2023.103661.

12. Ko, K.-M., Baek, J.-M., Seo, B.-S., & Lee, W.-B. Comparative Study of AI-Enabled DDoS Detection Technologies in SDN. *Applied Sciences*, 2023, vol. 13, iss. 17, article no. 9488. DOI: 10.3390/app13179488.

13. Dehkordi, A. B., Soltanaghaei, M., & Boroujeni, F. Z. The DDoS attacks detection through machine learning and statistical methods in SDN. *J Supercomput*, 2021, vol. **77**, pp. 2383-2415. DOI: 10.1007/s11227-020-03323-w.

14. Li, C., Wu, Y., Yuan, X., Sun, Z., Wang, W., Li, X., & Gong, L. Detection and defense of DDoS attack-based on deep learning in OpenFlow-based SDN. *International Journal of Communication Systems*, 2018, vol. 31, no. 5, article no. e3497. DOI: 10.1002/dac.3497.

15. Wang, H., & Li, W. DDosTC: A Transformer-Based Network Attack Detection Hybrid Mechanism in SDN. *Sensors*, 2021, vol. 21, iss. 15, article no. 5047. DOI: 10.3390/s21155047.

16. Jameel, A. S. M. M., Mohamed, A. P., Zhang, X., & Gamal, A. E. Deep Learning for Frame Error Prediction Using a DARPA Spectrum Collaboration Challenge (SC2) Dataset. *IEEE Networking Letters*, 2021, vol. 3, iss. 3, pp. 133-137. DOI: 10.1109/LNET.2021.3096813.

17. Nejad, S. H. M., & Majma, M. R. A Novel Congestion Avoidance Algorithm Using Two Routing Algorithms and Fast-Failover Group Table in SDN Networks, In: Ghita, B., & Shiaeles, S. (eds) *Selected Papers from the 12th International Networking Conference. INC 2020. Lecture Notes in Networks and Systems*, Springer, Cham, 2021, vol. 180, pp. 146-158. DOI: 10.1007/978-3-030-64758-2_11.

18. Almiani, M., AbuGhazleh, A., Jararweh, Y., & Razaque, A. DDoS detection in 5G-enabled IoT networks using deep Kalman backpropagation neural network. *Int. J. Mach. Learn. & Cyber.*, 2021, vol. 12, pp. 3337-3349. DOI: 10.1007/s13042-021-01323-7.

19. Al-Turaiki I, & Altwaijry N. A Convolutional Neural Network for Improved Anomaly-Based Network Intrusion Detection. *Big Data*, 2021, vol. 9, iss. 3, pp. 233-252. DOI: 10.1089/big.2020.0263.

20. Zolfaghari, H., Rossi, D., Cerroni, W., Okuhara, H., Raffaelli, C., & Nurmi, J. Flexible Software-Defined Packet Processing Using Low-Area Hardware. *IEEE Access*, 2020, vol. 8, pp. 98929-98945. DOI: 10.1109/access.2020.2996660.

21. Yungaicela-Naula, N. M., Vargas-Rosales, C., & Pérez-Díaz, J. A. SDN/NFV-based framework for autonomous defense against slow-rate DDoS attacks by using reinforcement learning. *Future Generation Computer Systems*, 2023, vol. 149, pp. 637-649. DOI: 10.1016/j.future.2023.08.007.

22. Najar, A. A., & Naik, S. M. Cyber-Secure SDN: A CNN-Based Approach for Efficient Detection and Mitigation of DDoS attacks. *Computers & Security*, 2024, vol. 139, article no. 103716. DOI: 10.1016/j.cose.2024.103716.

23. Alanazi, F., Jambi, K., Eassa, F., Khemakhem, M., Basuhail, A., & Alsubhi, K. Ensemble Deep Learning Models for Mitigating DDoS Attack in Software-Defined Network. *Intelligent Automation and Soft Computing*, 2022, vol. 32, no. 2, pp. 923-938. DOI: 10.32604/iasc.2022.024668.

24. Aslam, N., Srivastava, S. & Gore, M.M. A Comprehensive Analysis of Machine Learning- and Deep Learning-Based Solutions for DDoS Attack Detection in SDN. *Arab J Sci Eng*, 2024, vol. 49, pp. 3533-3573 DOI: 10.1007/s13369-023-08075-2.

Received 17.07.2023, Accepted 15.04.2024

ВИКОРИСТАННЯ НАБОРІВ ДАНИХ ДЛЯ ЕФЕКТИВНОГО ПОМ'ЯКШЕННЯ DDOS-АТАК У ПРОГРАМНО-ВИЗНАЧЕНИХ МЕРЕЖАХ: ЗНАЧЕННЯ ТА ПРОБЛЕМИ

Хема Дхадхал, Пареш Котак

Програмно-визначена мережа (SDN) стала трансформаційною парадигмою для керування мережею, пропонуючи централізований контроль і можливість програмування. Однак із поширенням DDoS-атак, які створюють значні загрози для мережевої інфраструктури, ефективні стратегії пом'якшення є обов'язковими. Предметом цієї статті є дослідження важливості наборів даних для пом'якшення DDoS-атак у середовищах SDN. У ньому обговорюється важливість наборів даних для навчання моделей машинного навчання, оцінки механізмів виявлення та підвищення стійкості систем захисту на основі SDN. Мета статті полягає в тому, щоб допомогти дослідникам ефективно вибирати та використовувати набори даних для пом'якшення DDoS у SDN, таким чином максимізуючи переваги та подолавши проблеми, пов'язані з вибором наборів даних. У статті описано проблеми, пов'язані зі збором, маркуванням і керуванням наборами даних, а також можливі рішення для вирішення цих проблем. Для ефективного виявлення та пом'якшення DDoS-атак у SDN потрібні надійні набори даних, які фіксують різноманітний і мінливий характер сценаріїв атак. Характеристика завдань для кожного розділу така: важливість наборів даних для пом'якшення атак DDoS у SDN, проблеми з використанням наборів даних для пом'якшення DDoS у SDN, вказівки щодо вибору набору даних, порівняння використаних наборів даних та їх результатів, використання різних наборів даних відповідно до потреба. Методологія передбачає збір результатів у табличній формі на основі попередніх досліджень для аналізу характеристик

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

існуючих наборів даних, методів розширення та вдосконалення наборів даних, а також оцінки ефективності різних наборів даних у виявленні та пом'якшенні DDoS-атак шляхом всебічного експериментування. Результати наших висновків свідчать про те, що для ефективного виявлення та пом'якшення DDoS-атак у SDN потрібні надійні набори даних, які фіксують різноманітний і мінливий характер сценаріїв атак. Наші висновки дають цінну інформацію про важливість наборів даних для підвищення стійкості інфраструктур SDN проти DDoS-атак. На завершення наші висновки дають цінну інформацію про важливість наборів даних для підвищення стійкості інфраструктур SDN проти DDoS-атак і підкреслюють необхідність продовження досліджень у цій критичній області. Ретельні вказівки щодо вибору наборів даних і впливу різних наборів даних, що використовуються в різних нещодавніх дослідженнях, визначають проблеми дослідження та майбутні напрямки в цій галузі.

Ключові слова: DDoS-атаки; програмно-визначена мережа; машинне навчання; набір даних; пом'якшення; значення; виклики.

Хема Дхадхал – науковий співробітник Гуджаратського технологічного університету, доцент кафедри інформаційних технологій, Інженерний коледж Лухдгірджі, Морбі, Гуджарат, Індія.

Доктор Пареш Котак – директор, A. V. Parekh Технічний інститут, Раджкот, Індія.

Hema Dhadhal – Research Scholar Gujarat Technological University, Assistant Professor in Information Technology, Lukhdhirji Engineering college, Morbi, Gujarat, India,
e-mail: hemadhadhal@gmail.com, ORCID: 0000-0001-7865-4192.

Dr Paresh Kotak – Principal, A. V. Parekh Technical Institute, Rajkot, India,
e-mail: kotakp2003@gmail.com, ORCID: 0009-0008-8943-6121.