**CHAPTER 2**

**LITERATURE REVIEW**

Bushra Alhijawi et al [1] included two main categories as presented a new classification of DoS attacks and mitigation techniques which is related to SDN. The classification that included mitigation methods to cope with DoS attacks on SDN and SDN-based solutions for mitigating DoS attacks. The first category indicates detection, mitigation, and prevention methods for DoS attacks against the SDN. The second category indicates the solutions that utilized SDN characteristics to handle DoS attacks in other networks such as cloud computing, legacy networks, and data centers. This paper analyzed the methods to cope with DoS attacks in SDN and classified them into six types, they are Table-Entry-based, Scheduling-based, Architectural-based, Flow statistics-based, Machine Learning-based, and Hybrid solutions. In future research, there are several ways to identify this detection as an establishment in a good progress in Machine Learning-based detection for DoS/DDoS, then more contributions can be developed using KNN, clustering methods and graphs. Second, a promising future direction would be to develop additional SDN-based solutions for DoS/DDoS mitigation against other environments such as IoT, cloud computing paradigms, and ISP networks. The development of optimization-based methods for detecting DoS/DDoS against SDN and in SDN-based solutions to minimize the false positive and maximize the true positive. It is an additional potential to increase an area of research. The different options for selecting baselines, datasets, and evaluation methodology, and their selections were led to consistent. Many parameter evaluations wanted to ensure the reliable results and facilitate the comparison between many approaches.

Raj Kumar Batchu et al [2] represented fast and efficient detection way for identifying real-world DDOS attacks in the networking. They had used the CICDDoS-2019 traffic dataset for this detection that are closer to real-time traffic activities. The

traffic activity contains noisy data, class imbalances, and irrelevant attributes. They had proposed a method that addresses all type of issues that are begin with class imbalance problem which are dealing to avoid bias towards majority class instances. Furthermore research, the memory optimizations are directed to enhance the processing speed of massive traffic and conserve multiple resources. The Hybrid Attribute selection technique with parameter tunings are implemented to extracting the essential attributes. These attributes are retrieved to the ELM and obtained the best parameters that discriminate DDOS attacks from regular traffic. These results have shown good generalization performance with six features in less training time. As a result, the suggested model can be utilized as a detection system to discover DDOS attack patterns in any network. This work is focused on binary classification. In future, this work can be extended for multi-attack classification, identifying IOT network attacks, and Software-defined networks attack

Juan Fernando Balarezo et al [3] analyzed about the existing DOS/ DDOS attacks mathematical models, firstly categorized the types of attacks which kind of network that are targeting either traditional, SDN or virtual networks. They had found that an amount of mathematical models in the literature that are much less compared to an amount of proposed mitigation solutions. These Attack models can provide a deeper understanding of DOS/DDOS attacks that would lead us to develop more efficient solutions and measures to mitigate their impact. They had classified the attack models which are found in the literature according to the type of network they were developed for the parameters considered and the way the authors analysed about the network environment and the attack patterns. Most of the attack models were developed for traditional networks, which makes an important to adapt or develop the models to the novel SDN paradigm, DDOS attacks are one of the biggest threats that software defined networks can face. It is possible to apply these models to software defined networks because SDN networking devices can also be exhausted in similar ways, the main difference relies on how traffic is processed between the different planes in SDN, so it is required to adapt the models. As future work they had planned to improve the mentioned models to the SDN in order to study with additional parameters that are required to be simulated the models to verify how much close with the model that behaves with the real test bed.

R. Amrish et al [4] proposed the machine learning algorithm for detecting DDOS attacks. This research study is evaluated using four distinct algorithm, they are ANN, KNN, Random Forest, and Decision Tree that are used to determine the classification model performs best in detecting malicious IP addresses. It is noted that the ANN model performs other classifier models with an accuracy of 99.95%. Furthermore results, these works can be expanded to develop and deploy in a block chain system to store and blacklist the IP Address of the node, if the traffic is classified as malicious by the machine learning algorithm. The use of the block chain provides an extra layer of security measure so that the data cannot be tampered.

James Halladay et al [5] analyzed the performance of eight machine learning algorithms and one deep learning model by using the dataset of CICDDoS2019. They had used these datasets for different two scenarios. Scenario A classified as benign versus DDOS traffic and Scenario B categorized as DDOS attack types. In these both scenarios, the models were initially trained with the control dataset consisting of 70 features. These models were trained on the proposed subset of 25 time based features. The top five models had accuracies of over 99% on the control features and over 98% on the time-based features in Scenario A. These results are as accurate as top DDOS classifiers seen in existing literature. Overall, the nine models had a median accuracy decrease of a mere 1.62% and a substantial median training time reduction of 36.28% when using only the proposed time-based features. Scenario B classified specific DDOS traffic as MSSQL, SSDP, SYN Flood, PORTMAP, DNS, LDAP, NETBIOS, SNMP, TFTP, NTP, UDP Flood, or UDP-Lag. The top performing model, XGB, had an accuracy of 74.08% on the control features and 69.05% on the time-based features. The nine classifiers had a median decrease of 7.43% in accuracy with a median training time reduction of 25.29% using only time-based features. Our top classifier’s performance metrics matched or exceeded the accuracies of previously discussed multiclass works, especially considering the uniquely large number of attack types in our experiments. Overall, XGB, RF, LD, LGBM, and KNN were the top-performing models. They had established the effectiveness of the time-based feature set in detecting and classifying Tor traffic. Likewise, our work demonstrated time-based features viability in the domain of DDOS. The results indicate that time-based features warrant further experimentation in traffic based research where a smaller dataset may prove beneficial to prevent over fitting and potentially to improve accuracy.

Furqan Rustam et al [6] exploited the DOS/DDOS attacks and the number of attacks have step increased over the past few years. Due to advanced and sophisticated attack detection approaches, the network security remains challenging problem in day to day activities. This research study proposed a machine learning-based framework to detecting DDOS/DOS attacks at the application layer by using multi-features. For multi-features, the best features from PCA and SVD are extracted to obtain superior performance. Extensive experiments are carried out using LR, RF, GB, ETC, as well as, CNN, LSTM, CNN-LSTM, and RNN models using different dataset sizes and different features. Experimental results indicate that PCA features tend to show better performance as compared to SVD features. Despite the results being better with using all features, multi-features help optimize the models’ performance. The RF model outperforms all other models by obtaining 100% accuracy when used with multi-features. On average, the tree-based ensemble models show better performance than linear models. This study shows excellent results, yet has several limitations. First, the approach is tested on a single dataset and requires further experiments regarding attacks on other layers. Second, the impact of dataset size is not investigated. Increasing the size might produce better results for deep learning models. Third, the influence of feature set size is also not analyzed, which we intend to perform in future work.

Francesco Musumeci et al [7] evaluated ML assisted DDOS attack detection frameworks for application in SDN environment considering Standalone and Correlated DAD architectures. Leveraging the potential of data-plane programmability enabled by P4 language, They had evaluated the detection latency is reduced when performing features extraction at P4 switches. To do so, they had compared different ML classifiers in terms of accuracy and computational time, and deployed the algorithms in a real-time scenario in which the P4 switch provides different types of data to the ML classifiers, namely, packet mirroring, header mirroring, and P4-metadata extraction. Numerical results show that attack detection can be performed with classification accuracy, precision, recall and F1-score higher than 98% in most cases, and with drastic time reduction in case of P4 is used for features extraction. As a future work, they had planned to investigate attack-type identification by developing multiclass ML classifiers, and implementing attack detection exploiting ML algorithms which leverage historical data, such as Recurrent Neural Networks.

Ashfaq Ahmad Najar et al [8] had performed several classiﬁcation models are used in our paper for detecting whether a packet is normal or an attack packet. We used the Random Forest model for the binary classiﬁcation which showed an accuracy of 99.13% on train data 99.13% accuracy on validation data and 97% on test data with an f1 score of 0.9661% which is very effective. We used another model which is a Multi-Layer Perceptron model to detect which type of attack packet it is, the model showed an accuracy of 97.96% on train data and 98.53% on validation data and 74% on the test data. These models can be used to detect malicious activity on the network. Having shown great accuracy these models will be very beneﬁcial for employing a defence mechanism on a network for detection of malicious activity.

Ismail et al [9] proposed a complete systematic approach for detection of this DDOS attack. They had applied the supervised machine learning techniques. This model generated the classification results from the supervised algorithm. They had used Random Forest and XG Boost classification algorithms. In their first classification, they had observed that both Random Forest Precision and Recall are getting approximately 89% accurate. For their second classification, they had noted that both XG Boost Precision (PR) and Recall (RE) are approximately 90% accurate. By comparing these classification result proposal to existing research works. Looking into the future works, for functional applications, it is important to provide a more user-friendly, faster alternative to deep learning calculations, and produce better results with a shorter burning time. It is important to work on unsupervised learning toward supervised learning for unlabelled and labelled datasets. Moreover, we will investigate how non-supervised learning algorithms will affect the DDOS attack detection.

Andrés Chartuni et al [10] presented the various transformation techniques were applied to the dataset, during the pre-processing stage. They had improved an accuracy metrics and a training time of the proposed model. During the proposed pre-processing stage, it was possible to perform a cleaning of Na N-type and null values. Similarly, a data normalization and quantile transformation were performed on the dataset. To obtain a balanced dataset in the context of classes, the SMOTE technique was applied only to the sample used for model training. The model was tuned based on a hyper parameter configuration. From the results of model training and validation, the tuning was applied mainly to the number of hidden layers, the rate used in the Dropout layer, and the learning rate. The model was trained and evaluated in three different scenarios. The results, metrics, and confusion matrix were presented for each scenario. The results show the third scenario as the best. This analysis was based on the performance metrics of the scenarios and number of labels used in the training phase. Scenario 2 yielded 94.57% accuracy, greater than Scenario 3 by 0.31%. However, it was not trained with the 13 labels available in the dataset, unlike the third scenario, which was trained with all labels. The added value of the proposed model consists in identifying the type of attack evidenced in a flow, as compared with the binary classification of recent works. As future work, they had proposed to deploy and evaluate the model in a computer network that has a flow not previously seen by the model. This deployment will consist of a model evaluation phase. Depending on the results, we propose to conduct a model training phase with the captured data. Once the model has been trained and validated, it could be instantiated as a network traffic manager whose objective is to reject or accept network flows based on the evaluation performed.

Kishore Babu Dasari et al [11] presented a comparison of the performance of six machine learning classification algorithms on eleven individual different DDOS attacks datasets. Unfortunately, the most common effective DDOS attack detection method for all DDOS attacks has yet to be identified. Some DDOS attacks have common effective methods and some attacks have different effective methods. Decision tree and random forest 467 algorithms gave poorer results than others. Logistic regression, Ada Boost, KNN, and NB show good results. In this paper, classification algorithms applied to different individual DDOS attack datasets get the best scores in all metrics with Google COLAB. TPU processor which is a powerful hardware accelerator and 12GB RAM. This configuration is more expensive. All datasets are big data size. The idea of next research would be to use feature selection to reduce data [22] and detect DDOS attacks using low-cost hardware

Mona Alduailij et al [12] analyzed the DDOS attack detection is a common problem in a distributed environment. This type of attack causes the unavailability of cloud service, which makes it essential to detect this attack. A machine learning model can be used to identify this type of attack. The research objective of this work is to detect a DDOS attack, with improved performance. This experiment was performed on the CICIDS 2017 and CICDDOS 2019 datasets. Different files related to DDOS attack were included in experiments, from both datasets. They had selected the most relevant features, by applying the MI and the RFFI methods. The selected features are fed to machine learning algorithms including (RF, GB, WVE, KNN, LR). The overall prediction accuracy of RF with 16 features, is 0.99993, and with 19 features, is 0.999977, which is better, Symmetry 2022, 14, 1095 13 of 15 compared to other methods. They had concluded that RF, GB, WVE, KNN, and LR are achieving good results, by using MI and RFFI as feature selection techniques. In the future, they may use wrapper feature selection methods, such as sequential feature selection, with neural networks, for DDOS and other attack detection.

Tariq Emad Ali et al [13] represented difference between DDOS attacks with various rates and patterns and normal traffic. Over the years, many effective ML/DL methods for DDOS attack detection have been suggested by different researchers. Sadly, however, the applicability of these techniques is severely constrained due to attackers constantly changing their attack tactics. To find involving the SLR protocol are evaluated and drawn from in this review in order to assess the state-of-the-art DDOS assault detection systems based on ML/DL approaches. The literature has been summarized in Section 4 in accordance with the suggested taxonomy for DDOS attack detection using ML/DL techniques, with each study’s respective advantages and disadvantages listed. The accuracy rate reported in much of the literature is over 99%. Because the majority of these studies assessed their models using offline data analysis for evaluation and comparison, certain metrics for performance may vary in a real-world or production settings. In particularly, they had noted that the existing papers have generally not employed the same DS or assessment techniques, making comparisons between their results difficult.

Li Xinlong et al [14] improved detection of DDOS attacks in terms of their methodology, which is among the most essential and complicated concerns in information security. Dealing with these systems calls for highly developed computer programs that can use time sequences and other generally advanced intelligence qualities to conquer complex challenges. In this spirit, an innovative hybrid model of the HTM system is provided in this study. The system’s architecture is modified by the addition of an LSTM cell so that the system can encode time sequences that comprise inflow data. Experiments showed that the proposed methodology successfully resolved the issue of accurately detecting DDOS attacks. An essential realization is that the proposed system can be a function of mapping the input data on the cells of the system’s final area, incorporating any spatial and temporal information discovered between the data. Different input data, that is, different values of the characteristics of the original input vector, and at different time frames within each sequence will lead to the activation of other cells. Therefore, it is easy to see that by attempting to map the relationship between cells and input data, and it is highly likely that we will be able to interpret the decisions of the proposed system, as is the case with simpler models such as decision trees because it is easy to see that by attempting to map the relationship between cells and input data, it is highly likely that we will be able to map the relationship between cells and input data. However, this method demands work and a significant amount of research, both of which we might address in a later line of investigation.

Meenakshi Mittal et al [15] discriminated the DDOS attacks with different rates and pattern from benign traffic is a very challenging issue. Many efficient DL approaches have been proposed by fellow researchers for DDOS attack detection over the years. But unfortunately, the scope of these methods is very limited as the attackers are continuously updating their attack strategies and skills very rapidly to launch unknown or zero-day DDOS attacks with unique traffic patterns every time. In this paper, we have used the SLR protocol to review the DDOS attacks detection system based on DL approaches and results of the SLR protocol are analyzed.

Deepak Kumar et al [16] proposed that a Deep Learning model is used to classify DDOS attacks on the network, which is likely to be more effective than the machine learning model. The LSTM model was chosen as a viable model for this investigation as it encompasses both feature selection and extraction in its model, which makes it superior to shallow machine learning methods. In this research work, the LSTM model has been used for the classification of benign and threats on the CICDDOS 2019 dataset, the LSTM model, which is employed as a deep learning model, has approximately 98.6% for DDOS attacks classification which is large as compared to KNN and ANN model. Furthermore, using the CICDDoS2019 dataset with LSTM to detect DDOS attacks provides direction for other DDOS intrusion detection research. Because of its great accuracy in attack detection, it appears that incorporating the LSTM model into the software-based networks is a good option. For future work, to capturing network traffic incorporates incremental learning. So the machine can update with a new type of attack.