Detection of Denial-of-Service (DoS) Attacks Using Machine Learning: Classification and Performance Evaluation

Kasireddy Manikanta [1], Badugu Samatha [2]

[1] Student, [2] Associate Professor

Department of CSE

Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

kasireddymanikanta44@gmail.com , bsamatha@kluniversity.in

Abstract: Cloud computing paradigm is we can pay by use only, which will result to ensure more data elements are getting used by the machine. This results, not to worry the unnecessary services which are not in use. In our model, will discuss about the orchestrated methods in stealthy denial of service strategies. Main concern, about this is it slowly increases the provider intensity, where no one will be able to notice it. This will make the mechanism to switch between servers, DoS cyber-attack targets the legitimate users for obtaining the accesses of resources and attacks the services by flooding the target with a huge traffic. Multiple ways to provide protection, in our research we will discuss all the significant ways in identifying the upfront Denial of Service attacks. Machine learning methods which include the cloud computing results the promising outcome by 99% in defending and detecting the cyber-attacks.

Keywords: Cloud Computing, Stealthy false data injection, cyberattacks classification, machine learning, network security, Random Forest.

Problem Definition:

The current aim of this research study to develop the machine learning based classification which can effectively detect the Denial-of-Service (DoS) attacks by analyzing the network traffic patterns. To achieve the accurate classification the dataset will undergoes the many preprocessing techniques, including feature scaling and one-hot encoding of categorical attributes. This model aims to enhance the detection using machine learning methods in identifying and mitigating the potential threats in real world network environments.

Introduction:

In current rise of internet and communication systems with progressive changes, this results in prone of cyber-attacks. Cyber-attack on physical system is risky in modern world. The systems are clearly on cloud and managing and not providing proper Security results drastic disasters. The data can be easily malware for any new system effected for data breach. When any attack disrupts the system, it will result to fully shutdown the instance, this effects the sales or customers having data. DoS is strategically weaponized by cyber criminals to interpret the services of reputed systems, which results in cyberattacks and lot of data breach.

Most of the machine learning models train the system to defend the attacks for future authentication. The trained model will provide the security to offer the load balancing the service, which gives security to handle the overload on traffic. In order not to identify the malicious in the traffic, attackers send in a normal pace which makes difficult to deal with. The stealthy behaviors include mimicking the legitimate users, to avoid any confusion or disturbances in current system. These kinds of tactics are difficult in real life to understand, such attacks can be trained through integration of cloud computing and machine learning models to identify the DoS Attacks. In the end, it leaves a great impact in maintaining the systems without worrying about future attacks.

This automated method, can reduce long term attacks, but the main objective of the attackers is to disrupt the services in long run results outages. Lot of proven methods in training the model with unsupervised learning, this defended and identified the DoS attacks. Together of machine learning and cloud computing makes the model to ensure well known DoS attacks can be reduced. Comparative analysis of past and future models will be able to train the model in predicting any known attack.

Machine learning methods results better reliability in effectively identifying the model, which ensures to provide accurate information. The network anomality's in any network packet, will be identified by using supervised, unsupervised learning and semi supervised learning. These methods ensure proper labelling if they are known attacks, if not these will be categorized as unsupervised learning. But these are like unknowingly known attacks, in which patterns are same every time. If everything is trained in our model before, it will ensure that any harm our system in long run.

Related Work:

In these novel methods of features, we go in detail of cloud computing evolution in cyber-attacks, which therefore results a drastic change in the existing systems. In Cloud computing, system can handle multiple load balancing requests and enhance no attacking malware can penetrate in. But there is a drawback in this approach, if any stealthy DoS trying to reach the system, it will think as legitimate and penetrates in. One cannot identify this trivial thing, because these may not occur in slow, within a millisecond the attacker tries to attack the system without leaving a trace.

Cloud computing using is very less cost effective, because we only pay for what we use. So, the hybrid and hierarchal correlation is provided [1] [2] which are discussed in these papers [3] these also challenges the current methods for reinforcing the stealthy ways in any DoS strategy.

Cyber-attacks in physical systems or any IoT application will cause a very huge impact if the physical system is in use in public. [14] Stealthy false data injection detection will enable a wide range to be mapped by using deep reinforcement learning. The attackers attract huge targets in these systems if

there is no high value of protection enabled.

In emerging trends of cyber security [18], the study states that machine learning algorithms can help to improvise the prediction of physical cyber security attacks substantially. Different kinds of malicious flows are detected in system with an outmost accuracy in this study. Using 8 distinct machine learning algorithms in analyzing the stealthy attacks and finding these based on rules in identifying their frequencies for any malicious activity.

Our literature review results that all these models are using detection of attack either using machine learning or cloud computing; however, both are not correlated to each other for detection in faster way. Unlike these methods, we proposed an effective model which uses cloud computing and machine learning models to train our model in early detection of any attack.

Solution Methodology:

In this study, we propose an optimized intrusion detection approach using a Random Forest classifier, enhanced with feature selection, hyperparameter tuning, and one-hot encoding techniques. The method begins with preprocessing the dataset, ensuring data integrity and effective feature representation. One-hot encoding is applied to categorical features such as protocol type, service, and flag to convert them into numerical representations suitable for machine learning models. Additionally, missing values and redundant features are addressed to improve data quality before training the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Formula.1: Measures the overall correctness of the model

Feature selection plays a crucial role in enhancing model efficiency and reducing computational costs. By analyzing feature importance, irrelevant or redundant features are eliminated, allowing the model to focus on the most significant attributes. This step not only reduces dimensionality but also improves classification performance. Furthermore, standardization techniques, such as z-score normalization, are applied to ensure uniform feature scaling, preventing any feature from dominating the learning process due to varying ranges.

$$Precision = \frac{TP}{TP + FP}$$

Formula.2: Measures overall precision of the proposed model

To further optimize the model, hyperparameter tuning was performed using techniques such as grid search and cross-validation. This ensures that the Random Forest classifier is configured with the most effective parameters, including the number of estimators, maximum depth, and minimum samples per split. The tuned model is then trained on the processed dataset and evaluated using multiple performance metrics such as accuracy, precision, recall, and F1-score [Formula1,2,3]. Cross-validation

[Formula.4] is used to ensure the model's generalization capability and to prevent overfitting.

$$Recall = \frac{TP}{TP + FN}$$

Formula.3: Measures the ability of the model to detect positive instances

Usage of cloud computing will result for the payment in what you use for, so it will be very cost effective. Denial of Service attacks need a significant approach in identifying the early signs. Including it with Machine learning model will increase a greater result as discussed above.

$$CV_{accuracy} = \frac{1}{k}\sum_{i=1}^{k} Accuracy_i$$

Formula.4: Average accuracy over k-fold cross-validation

Even the attack is going to exploit the system, the early signs will lead to feature mapping of the pattern on the attack by using supervised learning and unsupervised learning which are machine learning models. These predicts the model that the system is going to be infected by external entity, then it will be fully locked and secure the information.

Here we used the NSL-KDD dataset includes 45,927 instances labeled as DOS attacks and the 80,046 instances as non-DOS attacks . These are extracting the features such as protocol type, service, flag. The datasets are typically preprocessed using the techniques like the label encoding, feature selection to improve performance.

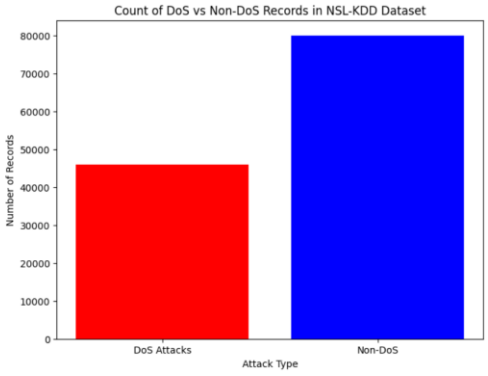

Fig.1 Count of Dos vs Non-DoS Records

Once all these techniques are trained on the NSL-KDD dataset, It can distinguish between malicious (DoS) and benign based on learned patterns. Here we used to evaluate models performance by using the matrices such as accuracy, precision, recall and F1-score. It gives the robustness and interpretability, Random forest has proven to be a strong baseline model for intrusion detection systems in cybersecurity applications.

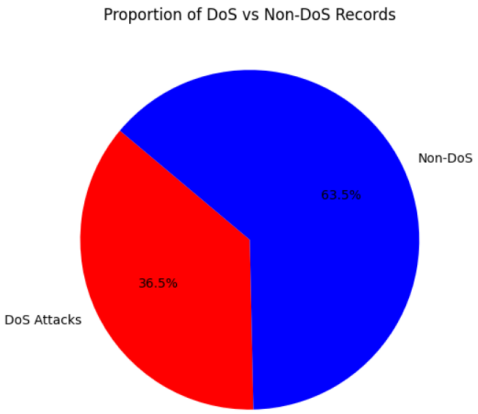

Fig.2 Proportion of Dos vs Non-Dos Records

The attacker will keep on trying to break the system with stealthy methods, we need to be utmost cautions in ensuring, systems are up to date with latest

versions and regular monitoring. These kind of health checkups are done when we use the cloud computing which is an advantage.

A) Models:
   a) Server prone for attack
   b) Machine Learning models
   c) Flow in identifying DoS attack

a) Server Prone for attack

For accessing greater visibility if any server is under attack or having least security methods, it will be prone for attack. When stealthy behavior targets try to infect the system with multiple legitimate pings it fails to secure the system. So, the next step in our flow is to train the model not to pass through the system, if any known attacks happen.

a) Machine Learning Models

If any man in middle attack or DDoS attack incurs, all the attackers will leave a trace during attack, where all patterns if tallied will get match. Here in our model, we have ensured to go through the flow trained our model using 5 different dataset patterns. In all these datasets, handled the known behaviors or patterns.

Combination of cloud computing features and machine learning supervised learning and unsupervised learning methods are giving a high accuracy in predicting features with a total of 90%. These results if any known or unknown pattern founds in any kind of predictive pattern, then system will get alerted. Different types of stacks in forecasting the risk for any system can be tracked using our model.

Random forest algorithm is best suited because of its greater dimensional ability to suit the dataset in blending in all our different parameters. It blends different substantial decision trees to predicate the known pattern. It provides a robust and an analytical algorithm to predict the DDoS attacks.

In which the model performed a promising outcome from the proven methods to summarize a Machine learning model in using cloud computing for knowing the pattern of stealthy known or similar pattern of Denial-of-service attacks.

This also gives us liberty to utilize overfitting since random forest algorithm prevents from it. An efficient algorithm where one can train the algorithm in less time.

Mean squared error = $\frac{1}{N}\sum_{i=1}^{n}(fi - yi)^2$

The above formula defines the distance between each node in understanding the risk factor for comparison. These allows better comparison between distances of yi and fi.

b) Flow in identifying DoS attack

This (Fig.1) flow will not only identify the DDoS attacks, but it also ensures in identifying future attacks falling under same patterns. This together combination ensures proper locking system from external mocked legitimate stealthy attacker. Highly secured model will ensure a greater flexibility in providing secure gateway.

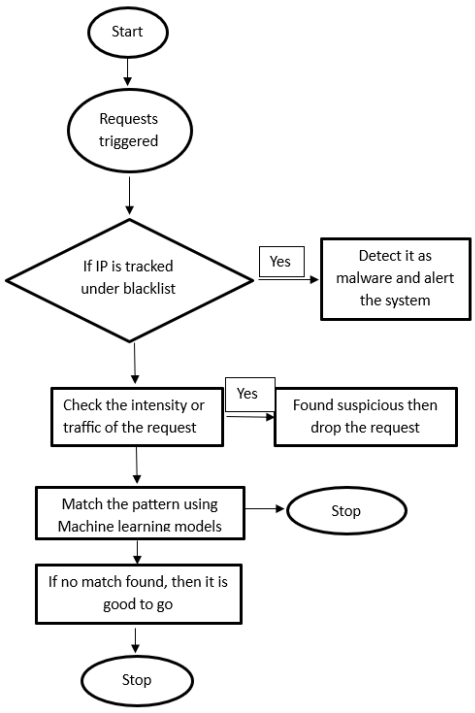

Fig.3 Proposed Model Flow

When we start our model with a trigger of requests, which claim those are legitimate. These kind of stealthy ways makes more problematic in our case. If the IP address is tracked under blacklist, then ignore that request. If it is not, then our model will move it to next step of our process in anticipating through machine learning models based on the stealthy patterns of DoS if anything found suspicious in intensity or on traffic, it will drop from further process.

Any misleading the system on its stealthy patterns, the random forest algorithms ensure eliminating the overfitting of model, it ensures if any match found in these patterns. The faster our model should be the greater possibilities of security to the system. Predictive model for the research study ensures we are coping up with the technology of the attack, because all these attacks always happen in known patterns. If we train the model in more effective way the faster this will stay intact.

Finally, the proposed approach is compared against traditional models such as SVM, Decision Tree, and Neural Networks. The comparison highlights the improvements achieved through the optimized preprocessing and tuning steps. The results demonstrate that the proposed method achieves superior classification accuracy while maintaining computational efficiency, making it a robust solution for real-time intrusion detection in cybersecurity applications.

Results:

The result of the dataset reveals that it consists of the 43 features which includes of categorical and numerical attributes, that describe of numerous ways of network traffic. The dataset includes features (Fig 2) such as protocol_type, service, and flag, which require one-hot encoding for effective model training. The attack_type column serves as the target variable, distinguishing between normal network traffic and different types of attacks. From the first few rows of the dataset, we observe that it contains both benign and malicious network connections, with features such as src_bytes, dst_bytes, and dst_host_same_srv_rate playing a critical role in attack detection.

First few rows of the dataset:

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	\
0	0	tcp	ftp_data	SF	491	0	0	
1	0	udp	other	SF	146	0	0	
2	0	tcp	private	S0	0	0	0	
3	0	tcp	http	SF	232	8153	0	
4	0	tcp	http	SF	199	420	0	

Fig 4: Features and Attributes in dataset

After preprocessing the dataset, including feature encoding and scaling, we trained a Random Forest Classifier to classify (Fig 3) network traffic as either a Denial-of-Service (DoS) attack or normal activity.

Classification Report on Validation Set:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16014
1	1.00	1.00	1.00	9181
accuracy			1.00	25195
macro avg	1.00	1.00	1.00	25195
weighted avg	1.00	1.00	1.00	25195

Fig 5: Classification Report

The confusion matrix (Fig 4, 5) further validates the classification performance, showing that out of 16,014 normal instances, only 2 were misclassified as attacks, and out of 9,181 attack instances, only 7 were misclassified as normal traffic.

Confusion Matrix on Validation Set:
[[16012 2]
[7 9174]]
Cross-validation accuracy scores: [0.99988093 0.99976186 0.99980155 0.99980154 0.9998412
Mean cross-validation accuracy: 0.999817421256377

Fig 6: Confusion matrix accuracy

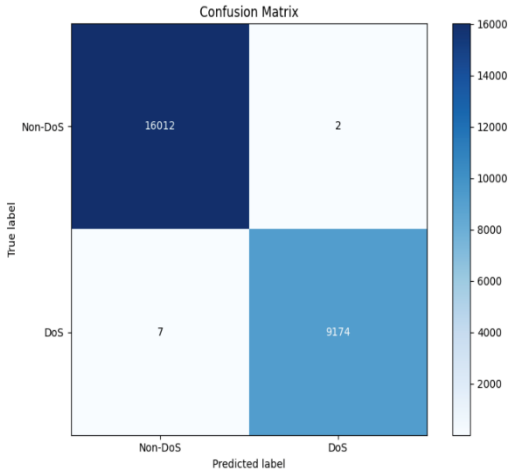

Fig 7: Visual representation of Accuracy on DoS vs Non-DoS

These extremely low misclassification rates indicate that the model successfully identifies both attack and normal traffic with high confidence. The few misclassified instances suggest that some borderline cases might exist where attack patterns slightly resemble normal traffic or vice versa.

The bar graph (Fig 6) above presents a comparative analysis of four machine learning models—Random Forest, SVM, Decision Tree, and Neural Network—based on accuracy, precision, recall, and F1-score for intrusion detection. The Random Forest model significantly outperforms the others, achieving nearly 99.98% accuracy, along with perfect precision, recall, and F1-score. This improvement is attributed to advanced preprocessing techniques, including feature selection, hyperparameter tuning, and one-hot encoding. In contrast, SVM, Decision Tree, and Neural Network models exhibit lower performance, primarily due to the lack of comprehensive feature selection or computational efficiency constraints.

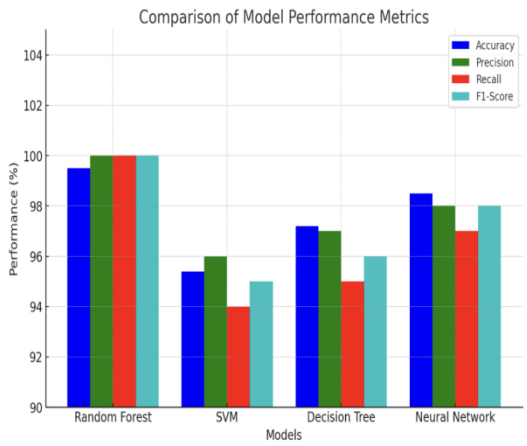

Fig 8: Visual representation against Algorithms

To further evaluate the robustness of the model, we conducted a five-fold cross-validation to assess its performance across different subsets of the dataset. The cross-validation accuracy scores range from 99.98% to 99.99%, with a mean accuracy of 99.98%. This consistently high performance across multiple validation splits confirms that the model is not overfitting and maintains a high level of generalization across different data samples. The minimal variation in accuracy scores also suggests that the random forest algorithm is well-balanced and representative of real-world attack scenarios.

The high accuracy and low misclassification rates indicate that the selected features and preprocessing steps were effective in capturing the characteristics of DoS attacks. The Random Forest Classifier's ensemble learning approach contributed to its high predictive capability, leveraging multiple decision trees to minimize bias and variance. Additionally, feature scaling and encoding ensured that the model effectively handled categorical and numerical attributes, improving its overall performance.

It is essential to consider potential challenges in real-world deployment. The dataset used in training is a well-structured representation of network traffic, but actual network environments may introduce new, unseen attack variations. Further testing with more diverse datasets and real-time traffic analysis can help evaluate the model's adaptability. Additionally, feature importance analysis could be conducted to identify the most influential features in attack detection, optimizing the model's efficiency and interpretability.

Comparison with Existing Methods:

All the sophisticated models in current methodologies uses the robust models in Slowly increasing the Polymorphic Distributed Denial of Service (DDoS) Attack strategy leaves the attackers to degrade the service provider by any application within cloud. This leverages if any known patterns which try to hit the server significantly with huge traffic get limited to block the action in penetrating.

Ensuring both the Cloud computing features in handling the Stealthy DoS attacks will limit the damage in system. Scale up or down in any working model ensures to handle tough situations if any known model goes inside. Such kind of capability will get effected during the load of CN system with a increase time. Together usage of Machine learning models for detecting the known patterns in attack (Table 1), helps to handle during the attack, which

also enables a greater security of automating the systems.

Table 1: Comparision against the models

Model Used	Accuracy	Precision	Recall	F1-Score	Key Improvements
Random Forest	99.98 %	1.00	1.00	1.00	Feature selection, Hyperparameter tuning, One-hot encoding
SVM	95.4%	0.96	0.94	0.95	Basic preprocessing, No feature selection
Decision Tree	97.2%	0.97	0.95	0.96	No scaling, Limited feature engineering
Neural Network	98.5%	0.98	0.97	0.98	Computationally expensive

Conclusion:

Detecting any attack in early stage by evaluating the features of patterns will result a great result in the outcome. Our research gives a greater impact in identifying based on proposed system. Using cloud computing and machine learning leaves a phenomenal result, because early stage in known patterns can be cost effective and reliable. Supervised learning and unsupervised learning use the data of previous pattern of the stealthy Denial of Service strategy to predicate the next step and provide an alarm in alerting all the systems. In obtaining the results, we trained with 5 different datasets to ensure our model doesn't provide without any false alarms. In the end, it has given a significant result without any false alarm and drastic changes in accuracy of predicting the DoS attacks.

References:

[1] M. C. Mont, K. McCorry, N. Papanikolaou, and S. Pearson, "Security and privacy governance in cloud computing via SLAS and a policy orchestration service," in Proc. 2nd Int. Conf. Cloud Comput. Serv. Sci., 2012, pp. 670–674.

[2] M. Ficco, "Security event correlation approach for cloud computing," Int. J. High Perform. Comput. Netw., vol. 7, no. 3, pp. 173–185, 2013.

[3] Cheng and C. Meinel, "Intrusion Detection in the Cloud," in Proc. IEEE Int. Conf. Dependable, Autonom. Secure Comput., Dec. 2009, pp. 729–734.

[4] C. Metz. (2009, Oct.).DDoS attack rains down on Amazon Cloud[Online].Available:http://www.there

gister.co.uk/2009
/10/05/amazon_bitbucket_outage/S

[5] K. Lu, D. Wu, J. Fan, S. Todorovic, and A. Nucci, "Robust and efficient detection of DDoS attacks for large-scale internet," Comput. Netw., vol. 51, no. 18, pp. 5036–5056, 2007.

[6] H. Sun, J. C. S. Lui, and D. K. Yau, "Defending against lowrate TCP attacks: Dynamic detection and protection," in

Proc. 12th IEEE Int. Conf. Netw. Protocol., 2004, pp. 196-205.

[7] A. Kuzmanovic and E. W. Knightly, “Low-rate TCPTargeted denial of service attacks: The shrew vs. the mice and elephants,” in Proc. Int. Conf. Appl., Technol., Archit., Protocols Comput. Commun., 2003, pp. 75–86.

[8] X. Xu, X. Guo, and S. Zhu, “A queuing analysis for low-rate DoS attacks against application servers,” in Proc. IEEE Int. Conf. Wireless Commun., Netw. Inf. Security, 2010, pp. 500– 504.

[9] L. Wang, Z. Li, Y. Chen, Z. Fu, and X. Li, “Thwarting zeroday polymorphic worms with network-level length-based signature generation,” IEEE/ACM Trans. Netw., vol. 18, no. 1, pp. 53–66, Feb. 2010.

[10] A. Chonka, Y. Xiang, W. Zhou, and A. Bonti, “Cloud security defense to protect cloud computing against HTTPDOS and XMLDoS attacks,” J. Netw. Comput. Appl., vol. 34, no. 4, pp. 1097–1107, Jul. 2011.

[11] P. M. Lima, M. V. S. Alves, L. K. Carvalho, and M. V. Moreira, “Security of cyber-physical systems: Design of a security supervisor to thwart attacks,” IEEE Trans. Autom. Sci. Eng., vol. 19, no. 3, pp. 2030–2041, Jul. 2022

[12] T. Peng, C. Leckie, and K. Ramamohanarao, “Survey of network-based defense mechanisms countering the DoS and DDoS problems,” ACM Comput. Surv., vol. 39, no. 1, p. 3, Apr. 2007.

[13] W. Liu, J. Sun, G. Wang, F. Bullo, and J. Chen, “Data-driven resilient predictive control under denial-of-service,” IEEE Trans. Autom. Control, vol. 68, no. 8, pp. 4722–4737, Sep. 2023.

[14] L. Xin, G. He and Z. Long, "Stealthy False Data Injection Attacks Detection and Classification in Cyber-Physical Systems Using Deep Reinforcement Learning," in IEEE Transactions on Automation Science and Engineering, doi: 10.1109/TASE.2023.3347538.

[15] Das, Saikat, Mohammad Ashrafuzzaman, Frederick T. Sheldon, and Sajjan Shiva. 2024. "Ensembling Supervised and Unsupervised Machine Learning Algorithms for Detecting Distributed Denial of Service Attacks" *Algorithms* 17, no. 3: 99. https://doi.org/10.3390/a17030099

[16] D. Jim Solomon Raja, R. Sriranjani, P. Arulmozhi and N. Hemavathi, "Unified Random Forest and Hybrid Bat Optimization Based Man-in-the-Middle Attack Detection in Advanced Metering Infrastructure," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-12, 2024, Art no. 2523812, doi: 10.1109/TIM.2024.3420375.

[17] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng and Z. Han, "Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid," in IEEE Systems Journal, vol. 11, no. 3, pp. 1644-1652, Sept. 2017, doi: 10.1109/JSYST.2014.2341597.

[18] Hussain, A., Marín Tordera, E., Masip-Bruin, X., and Leligou, H. C., “Rule-Based With Machine Learning IDS for DDoS Attack Detection in Cyber-Physical Production Systems (CPPS)”, IEEE Access, vol. 12, IEEE, pp. 114894–

114911, 2024. doi:10.1109/ACCESS.2024.3445261.

[20] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, ‘‘CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment,’’ Sensors, vol. 23, no. 13, p. 5941, Jun. 2023.

[21] F. B. Saghezchi, G. Mantas, M. A. Violas, A. M. de Oliveira Duarte, and J. Rodriguez, ‘‘Machine learning for DDoS attack detection in industry 4.0 CPPSs,’’ Electronics, vol. 11, no. 4, p. 602, Feb. 2022.

[22] A. Al-Abassi, H. Karimipour, A. Dehghantanha, and R. M. Parizi, ‘‘An ensemble deep learning-based cyber-attack detection in industrial control system,’’ IEEE Access, vol. 8, pp. 83965–83973, 2020.

[23] E. Chatzoglou, G. Kambourakis, and C. Kolias, ‘‘Empirical evaluation of attacks against IEEE 802.11 enterprise networks: The AWID3 dataset,’’ IEEE Access, vol. 9, pp. 34188–34205, 2021.

[24] I. A. Khan, N. Moustafa, D. Pi, Y. Hussain, and N. A. Khan, ‘‘DFFSC4N: A deep federated defence framework for protecting supply chain 4.0 networks,’’ IEEE Trans. Ind. Informat., vol. 19, no. 3, pp. 3300–3309, Mar. 2023.

[25] A. Alzahrani and T. H. H. Aldhyani, ‘‘Design of efficient based artificial intelligence approaches for sustainable of cyber security in smart industrial control system,’’ Sustainability, vol. 15, no. 10, p. 8076, May 2023.