Enhancing DDoS Attack Detection: A Hybrid Machine Learning Approach

*

1st Marcus Lundgren
Blekinge Institute of Technology
Master of Science: ML and AI
Karlskrona, Sweden
mamt19@student.bth.se

2nd Fredrik Johansson
Blekinge Institute of Technology
Master of Science: ML and AI
Karlskrona, Sweden
frjh19@student.bth.se

***Index Terms*—DDoS, machine learning, deep learning, hybrid model, network security, anomaly detection, feature selection, intrusion detection systems, real-time detection, scalability**

I. Introduction

Distributed Denial of Service (DDoS) attacks pose a significant risk to network security by flooding target networks with harmful traffic [1]. A notable instance is the October 2016 Dyn attack [2], which underscored the susceptibility of online service platforms to such threats. This attack utilized a botnet—comprising compromised Internet of Things (IoT) devices, referred to as bots or zombies [3]—to interrupt the operations of leading companies like Netflix and Amazon.

Since that landmark incident in 2016, strides have been made in identifying DDoS hazards. Nevertheless, the challenge persists, especially with the number of IoT devices online hitting 10.07 billion in 2021 and anticipated to soar to 24.1 billion by 2030 [4]. The burgeoning proliferation of IoT gadgets, along with other ubiquitous devices like smartphones and PCs, offers a chance to enhance the precision and accuracy of models aimed at detecting these threats. Current endeavors in the research of DDoS threat identification are experiencing significant advancements through the incorporation of machine learning and deep learning methodologies. These techniques are increasingly vital for elevating the detection and formulation of countermeasures against diverse DDoS onslaughts [3] [4] [9] [10] [11].

The CIC-DDoS2019 Dataset: Developed by the Canadian Institute for Cybersecurity, the CIC-DDoS2019 dataset comprises extensive data collection on DDoS (Distributed Denial of Service) attacks and benign traffic [5]. It is tailored for real-time DDoS attack detection research, emphasizing minimal computational overhead. This dataset stands out for its capacity to train and evaluate machine learning models, thanks to its realistic depiction of network traffic encompassing a range of contemporary DDoS attack vectors. A notable aspect of DDoS attacks is their execution under various protocols. In the CIC-DDoS2019 dataset, several files are exclusively dedicated to one protocol each. It is important to discuss the significance of these protocols to understand the differences between DDoS attacks.

1) **DNS (Domain Name System):** This system translates domain names to IP addresses, allowing users to access websites with human-readable names. In DDoS attacks, attackers can exploit DNS servers to amplify traffic, known as DNS amplification attacks, overwhelming the target with a massive amount of data [6].
2) **LDAP (Lightweight Directory Access Protocol)** is a protocol used for accessing and maintaining distributed directory information services over an Internet Protocol (IP) network. LDAP can be exploited in DDoS attacks by making a high volume of complex requests to the server, leading to resource exhaustion [7].
3) **MSSQL (Microsoft SQL Server)** is a relational database management system by Microsoft. In the context of DDoS, attackers might target MSSQL servers with SQL injection attacks to disrupt services or execute commands that overwhelm the system [8].
4) **NetBIOS (Network Basic Input/Output System):** Provides services related to the session layer of the OSI model, allowing applications on different computers to communicate over a local area network. NetBIOS can be exploited in DDoS attacks to flood a network with traffic, causing disruption [8].
5) **NTP (Network Time Protocol)** is used to synchronize the clocks of computers over a network. However, NTP servers can be misused in reflection DDoS attacks, where attackers send requests with forged return addresses, causing servers to respond to the targeted victim, amplifying the volume of traffic [8].
6) **SNMP (Simple Network Management Protocol)** manages devices on IP networks. It can be vulnerable to DDoS attacks when attackers flood a network with SNMP requests with spoofed IP addresses, overwhelming the network or the SNMP servers with excessive traffic.

7) **SSDP (Simple Service Discovery Protocol):** This protocol is used within UPnP (Universal Plug and Play) to discover or advertise network services. It can be exploited for DDoS attacks by sending requests to devices with a forged return address, causing them to respond to a target, significantly amplifying the traffic [8].
8) **UDP (User Datagram Protocol)** is a simple, connectionless Internet protocol that offers limited service when messages are exchanged between computers in a network. UDP flood attacks involve sending a large number of UDP packets to random ports on a remote host, causing the host to check for the application listening at that port repeatedly and (when no application is found) reply with an Internet Control Message Protocol (ICMP) Destination Unreachable packet, thus saturating the network [8].
9) **SYN (Synchronization)** is a part of the Transmission Control Protocol (TCP) three-way handshake used to establish a connection between a client and server. SYN flood attacks involve sending a rapid succession of SYN requests to overwhelm a target's system, making it unable to handle legitimate traffic [8].
10) **TFTP (Trivial File Transfer Protocol)** is a simple, lockstep file transfer protocol with no authentication. It can be exploited in DDoS attacks by overwhelming a server with requests or by distributing malicious files through unsecured TFTP servers [8].

The proposed project introduces a mixed-model approach, combining a Support Vector Machine (SVM) with a Decision Tree (DT), as suggested by S. Sumathi and N. Karthikeyan. [9]. This innovative SVM-DT hybrid aims to harness the SVM's proficiency in managing complex, high-dimensional datasets alongside the DT's straightforward interpretability and user-friendliness. Such a blend is designed to identify diverse attack patterns effectively. In addition, to provide benchmarks for this hybrid construct, it will be evaluated against other established models in the field of DDoS detection, namely DT, SVM, Extreme Gradient Boosting (XGBoost), and Random Forest (RF), which are recognized as standard tools by the DDoS detection community [4] [10] [11]. The objective is to achieve a solution with accuracy, efficiency, and scalability. The CIC-DDoS2019 dataset will serve as the foundation for training and validation, ensuring the model is thoroughly tested in scenarios mimicking real-world attacks. This process emphasizes reducing false positives while enhancing the accuracy of threat detection.

II. Background and Related Works

In their latest report, Cloudflare revealed a significant surge in DDoS threats, observing a 117% year-over-year increase in network-layer DDoS attacks during Q4. This period notably saw heightened DDoS activity aimed at retail, shipment, and public relations sectors, particularly around peak shopping times like Black Friday and the holiday season. This trend underscores the evolving landscape of cyber threats and the importance of robust cyber defenses for businesses, especially during critical sales [12].

Chuyu She et al. developed a method to differentiate between genuine users and botnets that launch DDoS attacks at the application level. This approach hinged on analyzing seven distinct features extracted from user sessions. Applying a one-class SVM algorithm to their collected data demonstrated their model's effectiveness in identifying DDoS threats targeting the application layer [10].

S. Sumathi and N. Karthikeyan conducted a comprehensive analysis of various traditional and hybrid machine learning algorithms by testing them on the KDDcup99 and DARPF datasets. Their findings highlighted that DT and Fuzzy C-Means algorithms outshone others in performance, with the Fuzzy C-Mean algorithm achieving an impressive 98.7% accuracy in detecting DDoS traffic, alongside a swift detection time of 0.15 seconds. Inspired by their work, which underscores the efficacy of combining different machine learning approaches, we explored the potential of a hybrid model SVM with DT [9].

A. Seifousadati et al. applied several machine learning algorithms, including Naive Bayes, SVM, XGBoost, AdaBoost, KNN, and RF, to the CICDDoS2019 dataset for binary classification. Among these, their developed XGBoost model was notably successful, achieving a remarkable accuracy of 100% and an F1-score of 1, showcasing its exceptional capability in identifying binary classifications accurately [4].

III. Problem Statement and Objectives

A. Problem Statement

Despite advances in cybersecurity, Distributed Denial of Service (DDoS) attacks remain a formidable threat to the stability of online platforms and services. As the Internet of Things (IoT) continues to expand, the volume and sophistication of these attacks have escalated, outpacing the capabilities of many existing detection systems. Conventional DDoS detection methods struggle to adapt to evolving attack patterns and often incur high false-positive rates. This study addresses the pressing need for an adaptive, accurate, and scalable DDoS attack detection system that can keep pace with the dynamic nature of cyber threats.

B. Objectives

The primary objectives of this research are to:

- Develop a hybrid machine learning model that combines the strengths of Support Vector Machines (SVM) and Decision Trees (DT) to improve the detection rates of DDoS attacks.
- Evaluate the performance of the proposed hybrid model against established machine learning algorithms on the CIC-DDoS2019 dataset to benchmark its effectiveness.
- Analyze the model's ability to generalize from seen to unseen DDoS attack protocols, thereby assessing its practical applicability in diverse network environments.
- Implement preprocessing techniques to address the class imbalance in the dataset, thereby enhancing the model's predictive accuracy and reducing the likelihood of false positives and negatives.

- Enhance the transparency and interpretability of the model, facilitating trust and understanding in its decision-making processes for end-users.

IV. Method

We embraced the Cross Industry Standard Process for Data Mining (CRISP-DM) to tackle detecting DDoS attacks within network traffic. The following results were gained when executing CRISP-DM.

*1) **Business Understanding**:* Our mission is to pinpoint DDoS attacks precisely, bolster cybersecurity measures, ensure fairness, and adhere to ethical guidelines.

*2) **Data Understanding**:* In the exploratory data analysis (EDA) phase, we conducted a feature importance assessment to identify which features contribute most significantly to the predictive model's decision-making process before data cleaning and processing. 1 illustrates the top 10 features ranked by their importance as determined by the Decision Tree algorithm. The feature 'Inbound' demonstrates a predominant influence, with 'Source Port' and 'Destination Port' also showing substantial importance. This analysis indicates the raw, unprocessed dataset has inherent patterns that the Decision Tree model considers highly relevant for classifying the traffic data.

This preliminary feature importance assessment provides a clear indication that certain features, like 'Inbound', 'Source Port', and 'Destination Port', are key indicators for classifying traffic data, despite the dataset not being cleaned or processed. However, as seen in Table I, there exists a significant imbalance in the data. The 'BENIGN' category represents an extremely small fraction compared to the 'Attack Type' categories across all datasets. For example, 'BENIGN' data only constitutes 1.18% of the 'NTP' dataset and an even smaller 0.025% of the 'Syn' dataset.

This imbalance is visually corroborated by 2, which depicts the class distribution, illustrating the stark contrast between the counts of 'BENIGN' and various types of 'DDoS' attacks, and in Table I, which compares the benign instances to attack instances, highlighting an overwhelming predominance of attack data with benign instances constituting a mere fraction—often less than 1%—of the total dataset in several categories. Such disproportionality can potentially lead to a biased model that overfits the 'Attack Type' data while underrepresenting the 'BENIGN' class. It is crucial, therefore, to address this imbalance as part of the data preprocessing to ensure that the predictive model does not simply learn to predict the majority class but can discern and correctly classify the underrepresented 'BENIGN' traffic. The steps we take to balance this dataset will directly impact the model's ability to generalize and accurately detect DDoS activities in a real-world scenario where 'BENIGN' interactions would naturally occur more frequently than DDoS attacks.

*3) **Data Preprocessing**:* The dataset underwent several preparatory steps, including cleaning, normalization, and transformation, to ensure its readiness for analysis. Only a few missing values (NaN) were detected during the cleaning

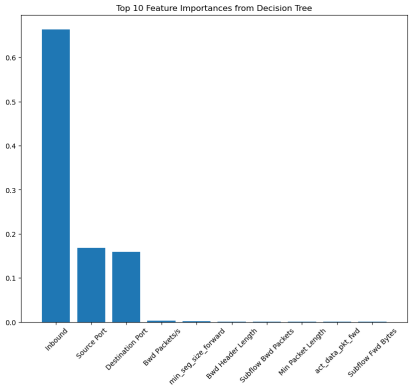

Fig. 1. Feauture importance before data processing.

TABLE I
Summary of Datasets and Counts

Dataset	Attack Type	Count	Ratio
NTP	NTP	1,202,642	98.81%
	BENIGN	14,365	1.18%
Syn	Syn	1,582,289	99.98%
	BENIGN	392	0.025%
DNS	DNS	5,071,011	99.93%
	BENIGN	3,402	0.067%
TFTP	TFTP	20,082,580	99.87%
	BENIGN	25,247	0.126%
UDPLag	UDP-lag	366,461	98.88%
	BENIGN	3,705	1.00%
	WebDDoS	439	0.12%
SSDP	SSDP	2,610,611	99.97%
	BENIGN	763	0.029%
NetBIOS	NetBIOS	4,093,279	99.96%
	BENIGN	1,707	0.042%
MSSQL	MSSQL	4,522,492	99.96%
	BENIGN	2,006	0.044%
UDP	UDP	3,134,645	99.93%
	BENIGN	2,157	0.069%
LDAP	LDAP	2,179,930	99.93%
	BENIGN	1,612	0.074%
SNMP	SNMP	5,159,870	99.97%
	BENIGN	1,507	0.029%

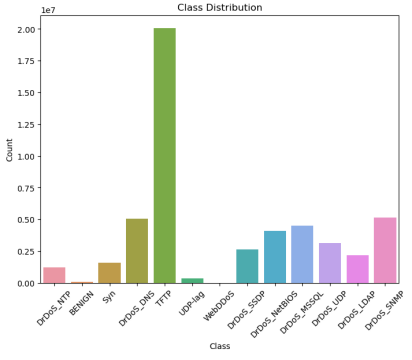

Fig. 2. Class Distribution.

phase and subsequently removed. A more significant issue was the presence of duplicate entries, which were eliminated to reduce noise and improve the dataset's integrity. Since specific algorithms, particularly SVMs, are sensitive to data scale, feature scaling was deemed essential. Finally, to enhance the dataset's efficiency and compatibility with various models, data types were optimized by downcasting, and object-type features were converted to categorical types, preparing the dataset for effective training.

Delving deeper into our rationale for feature exclusion, particularly concerning variance, we implemented a rigorous approach. Features with zero variance were eliminated outright, as they contribute unnecessary noise rather than valuable information to the model. Furthermore, we applied a variance threshold of 0.01—a relatively low figure—signifying our intent to retain features that exhibit a minimal linear relationship with the target variable. Features failing to demonstrate linear association were consequently removed from consideration. Following this meticulous selection process, we identified 18 features of significance, see Figure 3. The subsequent section assesses and details their importance to the model's performance.

4) ***Data Preperation****:* Upon thoroughly reviewing the dataset and consultation with relevant academic articles, we identified several features that could be omitted from our analysis. These include a range of metadata and attributes displaying zero or near-zero variance. Specifically, the metadata features we have decided to exclude are "Flow ID," "Fwd Header Length.1," "Source IP," "Src IP," "Source Port," "Src Port," "Destination IP," "Dst IP," "Destination Port," "Dst Port," "Timestamp," "Unnamed: 0," "Inbound," and "SimillarHTTP."

Metadata is data about other data [13]. It provides descriptive information about a dataset's content, facilitating easier identification, use, or management of the data it describes. In the context of our dataset, metadata includes details like source and destination ports, which indicate the communication endpoints in network transactions.

The rationale for excluding certain metadata features, particularly source and destination ports, is elaborated in the study by L. D'hooge at al. [14]. This research underscores how incorporating metadata into the model introduces powerful predictive shortcuts. Such shortcuts lead to a model's overreliance on specific features for prediction, potentially causing overfitting. This overfitting manifests as high accuracy on training data but poor performance on unseen data, primarily because the model does not adequately learn the more intricate relationships within the data.

These predictive shortcuts enable the model to make rapid classifications based solely on metadata, neglecting a more thorough examination of the complex patterns or interactions that may be present. This compromises the model's comprehensive understanding of the dataset and increases its vulnerability to manipulation. By altering metadata, attackers could evade detection, exploiting the model's dependency on these features.

In light of these considerations, our decision to exclude features such as "Source Port" and "Destination Port" aims to strengthen the model's capacity for learning and generalization. This strategy seeks to ensure the model's robustness and effectiveness in detecting a wide array of intrusion attempts, enhancing its reliability as a tool for network security. By studying Figure 2, we can see that inbound, destination port, and source port have high feature importance.

Given the significant imbalance in the target label class of the CICDDoS2019 dataset, it's crucial to ensure a balanced representation of classes for effective model training. While considering different methods to address this challenge, we evaluated two common strategies: over-sampling the minority class and under-sampling the majority class. Although effective in equalizing class distribution, over-sampling carries the risk of overfitting due to the duplication of minority class data. Conversely, under-sampling might lead to the loss of potentially valuable data from the majority class, compromising the model's ability to learn effectively.

After carefully considering various approaches and their associated risks, we determined that a targeted sampling method would be the most suitable for our purposes. Unlike stratified sampling, which aims to maintain the original distribution of classes across different subsets, targeted sampling allows us to intentionally balance the classes within our dataset. This method involves separately sampling an equal number of instances from minority and majority classes to achieve a 50/50 distribution. This approach directly addresses the challenge of class imbalance without over-sampling, which risks overfitting due to the duplication of minority class data, or under-sampling, which may result in the loss of potentially valuable data from the majority class. By employing targeted sampling, we create a balanced and reflective dataset of the diversity we aim to model, thus optimizing our chances of developing a robust and generalizable model for the highly imbalanced CICDDoS2019 dataset. This method ensures that our training data accurately represents both classes, enhancing model performance and fairness.

The dataset was categorized into seen and unseen protocols to assess the model's performance on familiar and novel data. Protocols for training were selected randomly, with seen protocols including NTP, SYN, DNS, TFTP, and UDPLag. The unseen protocols, tested to evaluate the model's generalization ability, comprised SSDP, NetBIOS, MSSQL, UDP, LDAP, and SNMP. This strategy aimed to rigorously test the model's predictive capabilities across different types of network traffic.

5) ***Modeling****:* We deploy a comprehensive range of machine-learning algorithms to classify network traffic efficiently. This ensemble includes DT, RT, SVM, XGBoost, and a hybrid SVM+DT model. Each of these algorithms has earned recognition for their effectiveness within the DDoS detection domain [4] [9] [10] [11]. This strategic selection aims to leverage the unique strengths of each model to enhance the accuracy and reliability of network traffic classification, ensuring a robust defense mechanism against DDoS attacks.

6) ***Evaluation****:* In step 5 of the process, we evaluate the models using key performance metrics, namely accuracy and the F1 score, to align with our business objectives. This

assessment ensures that the chosen models not only demonstrate high precision in predictions but also maintain a balance between sensitivity and specificity, reflecting their effectiveness in meeting our organization's strategic goals.

- Accuracy scores calculate the proportion of correctly predicted labels to the overall number of labels. Although initially, the data might be imbalanced, making accuracy a less reliable metric, applying techniques to balance the dataset justifies its use as a valid evaluation measure.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- The F1-Score represents the harmonic mean of Precision and Recall, making it an effective metric for evaluation alongside the Accuracy Score. It is precious because it accounts for False Positives and False Negatives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Finally, within the hybrid SVM + DT model, we employ a majority voting mechanism to consolidate the predictions from both the SVM and DT components. This approach allows the model to leverage the strengths of both algorithms, with the final prediction determined by the outcome that receives the majority vote from both SVM and DT. If both models agree on a classification, that classification is adopted.

7) ***Deployment****:* In the final step of the CRISP-DM process, we present our findings and the developed models to stakeholders, outlining a detailed deployment strategy for practical application or academic investigation. This is accomplished through a comprehensive report. This phase is crucial for ensuring that the insights gained and the methodologies applied are communicated, facilitating informed decision-making and enabling effective integration into real-world scenarios or further scholarly research.

V. Result

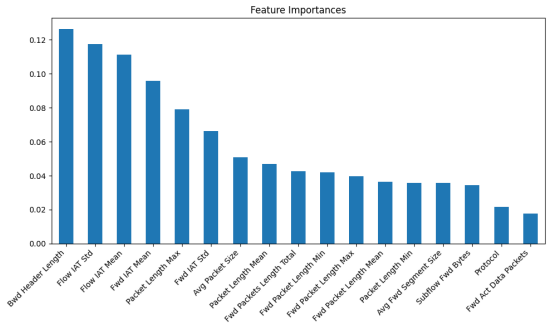

Fig. 3. Feature importance.

TABLE II
Training and Evaluation on seen protocols

Seen protocols	Training and Evaluation		
	Mean CV acc score (5)	*F1-score*	*Final pred acc*
RF	0.9980	0.9985	0.9985
SVM	0.7675	0.7660	0.7777
XGB	1.00	0.9985	0.9985
DT	0.9970	0.9982	0.9982
SVM + DT	X	0.9982	0.9982

TABLE III
Testing on unseen protocols

Unseen protocols	Testing	
	F1-score	*Final pred acc*
RF	0.8410	0.9964
SVM	0.7973	0.9975
XGB	0.8313	0.9961
DT	0.6412	0.9814
SVM + DT	0.9906	0.9814

VI. Analysis

Reviewing the data, Table 2 showcases the outcomes for five distinct models: RF, SVM, XGBoost, DT, and a hybrid model of SVM and DT. The findings indicate that RF, XGBoost, and DT each display notable accuracy and F1-scores when dealing with previously encountered protocols. On the other hand, SVM could perform better in these scenarios.

Turning our attention to Table 3, which assesses the models' efficacy with unseen protocols, we observe that both RF and XGBoost maintain their accuracy levels to a degree but encounter a significant decline in F1-scores. This drop hints at these models' challenges in accurately classifying new data, marked by increased false positive rates. DT's performance declines more drastically in terms of its F1-score, even more so than RF and XGBoost, highlighting a notable disparity. Interestingly, the SVM model exhibits an improved handling of unseen protocols, potentially due to its capacity for better generalization and reliance on robust features that are not overly specific to the previously seen protocols. This improvement suggests that the unseen protocols may present more distinct or straightforward patterns that align more closely with SVM's classification boundaries, unlike the possibly more complex patterns found in the seen protocols.

Despite the individual limitations observed in the SVM and DT models, the hybrid approach stands out for its stellar performance. This highlights the potential of model fusion to capitalize on each model's unique strengths, thereby enhancing overall model adaptability and resilience, particularly in varied or unpredictable settings. The synergy achieved by integrating SVM and DT into a single hybrid model appears to bolster their individual decision-making qualities, offering a more balanced approach to tackling the challenges posed by unseen protocols.

VII. Discussion

The outcomes of this study are intricately linked to the specific data preparation methods employed, particularly the

selection of protocols for training versus those designated for prediction. This choice significantly influences the results, contrasting with numerous studies that opt to train on all available protocols. Given the critical role these rates play in detecting DDoS attacks, this investigation aimed to delve into the dynamics of false-positive rates when differentiating between seen and unseen protocols.

Moreover, the choice of models in this study brings to the forefront considerations regarding explainable AI (XAI) and interpretability. DTs and RFs are noted for their higher interpretability, with DTs offering transparent decision paths from root to leaf, clearly delineating how predictions are made. RFs, composed of multiple DTs, provide valuable insights through feature importance scores. On the other hand, the SVM and XGBoost models, despite their solid predictive capabilities, present challenges in terms of interpretability due to their complex decision mechanisms. The introduction of a hybrid model, blending SVM and DT, potentially complicates interpretability further by integrating the complex decision boundaries of SVM. It is crucial to acknowledge the balance between performance and interpretability. This report has tried to enhance transparency in these complex models by applying feature-importance techniques to illuminate the decision-making process.

Additionally, the report considers training time, which becomes particularly pertinent when dealing with large datasets. The training duration for the SVM model notably exceeds that of other models, posing a challenge for maintaining the timeliness of the hybrid model, especially in adapting to new attacks. This consideration underscores the need for a balanced approach in model selection, weighing the benefits of predictive accuracy against the practical constraints of model training and maintenance.

VIII. Limitations

The project faces limitations, including a significant imbalance in the dataset between attack and non-attack instances, necessitating target sampling and leading to data underutilization. This approach might hinder model optimization. It focuses on familiar protocols, excluding unfamiliar ones, though the report evaluates model performance to assess adaptability. The reliance on majority voting for the hybrid model is a constraint; future work could explore diverse integration methods for enhanced outcomes, including a broader application of hybrid models beyond the current scope.

IX. Conclusion

Our research introduced a combined SVM and DT model for binary classification tasks. This model showcased remarkable effectiveness across both familiar and unfamiliar protocols. The model's strength lies in its interpretability and transparency. Decision Trees offer clear insights into decision-making processes, and although SVMs are typically less transparent, incorporating feature importance metrics significantly improves their interpretability. Surpassing our initial projections, this hybrid model achieved an impressive approximate 98% in both F1-scores and accuracy rates for classifying both seen and unseen protocols, demonstrating its robustness and adaptability.

References

[1] I. Sharafaldin, A. Habibi Lashkari, S. Hakak, and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," in *Proc. IEEE 53rd International Carnahan Conference on Security Technology*, Chennai, India, 2019.

[2] "Top 5 Most Famous DDoS Attacks," Microsoft, Feb. 17, 2023. [Online]. Available: https://www.microsoft.com/en-us/microsoft-365-life-hacks/privacy-and-safety/top-5-most-famous-ddos-attacks. [Accessed: Feb. 13, 2024].

[3] I. Ko, D. Chambers, and E. Barrett, "Unsupervised learning with hierarchical feature selection for DDoS mitigation within the ISP domain," ETRI, 2019.

[4] A. Seifousadati, S. Ghasemshirazi, and M. Fathian, "A Machine Learning Approach for DDoS Detection on IoT Devices," ETRI, 2019.

[5] "DDoS evaluation dataset (CIC-DDoS2019)," Canadian Institute for Cybersecurity, [Online]. Available: https://www.unb.ca/cic/datasets/ddos-2019.html. [Accessed: March. 13, 2024].

[6] Cloudflare, 'What is a DNS amplification DDoS attack?' [Online]. Available: https://www.cloudflare.com/learning/ddos/dns-amplification-ddos-attack/ [Accessed: March. 15, 2024].

[7] L. Segal, "Old Protocols, New Exploits: LDAP Unwittingly Serves DDoS Amplification Attacks," F5 Labs, Nov. 15, 2016. [Online]. Available: https://www.f5.com/labs/articles/threat-intelligence/old-protocols-new-exploits-ldap-unwittingly-serves-ddos-amplification-attacks-22609. [Accessed: March. 15, 2024].

[8] DDoS-Guard, 'DDoS attack types,' DDoS-Guard.net. [Online]. Available: https://ddos-guard.net/en/terms/ddos-attack-types. [Accessed: March. 12, 2024]

[9] S. Sumathi and N. Karthikeyan, "Search for Effective Data Mining Algorithm for Network-Based Intrusion Detection (NIDS)-DDOS Attacks," in 2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), 2018, pp. 41-45.

[10] C. She, W. Wen, Z. Lin, and K. Zheng, "Application-layer DDoS detection based on a one-class support vector machine," in International Journal of Network Security & Its Applications (IJNSA), vol. 9, no. 1, pp. 13-24, 2017.

[11] E. Navruzov and A. Kabulov, "Detection and analysis types of DDoS attack," in 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2022, pp. 1-7, doi: 10.1109/IEMTRONICS55184.2022.9795729.

[12] Cloudflare, "DDoS Threat Report 2023 Q4," 2023. [Online]. Available: https://blog.cloudflare.com/ddos-threat-report-2023-q4. [Accessed: March. 9, 2024].

[13] Wikipedia contributors, 'Metadata,' Wikipedia, The Free Encyclopedia. [Online]. Available: https://en.wikipedia.org/wiki/Metadata. [Accessed: March. 20, 2024]."

[14] L. D'hooge, M. Verkerken, B. Volckaert, T. Wauters, and F. De Turck, "Establishing the Contaminating Effect of Metadata Feature Inclusion in Machine-Learned Network Intrusion Detection Models," in Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2022), Lecture Notes in Computer Science, vol. 13358, L. Cavallaro, D. Gruss, G. Pellegrino, and G. Giacinto, Eds. Cham: Springer, 2022, pp. [page numbers]. https://doi.org/10.1007/978-3-031-09484-2_2