electronics

MDPI

Article

Proactive DoS and DDoS Attack Detection Through Behavior-Based Threat Intelligence

Orieb Abualghanam [1,*], Malik Al-Essa [1], Wesam Almobaideen [1,2], Mohammad Qatawneh [3,4] and Ahmad Sami Al-Shamayleh [5]

[1] Department of Computer Science, The University of Jordan, Amman 11942, Jordan; m.alessa@ju.edu.jo (M.A.-E.); w.almobaideen@ju.edu.jo (W.A.)
[2] Department of Electrical Engineering and Computing Sciences, Rochester Institute of Technology Dubai, Dubai 341055, United Arab Emirates
[3] Department of Networks and Cybersecurity, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19111, Jordan
[4] Department of Computer Information Systems, The University of Jordan, Amman 11942, Jordan
[5] Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19111, Jordan; a.alshamayleh@ammanu.edu.jo
* Correspondence: o.abualghanam@ju.edu.jo

Abstract

The rapid growth of cyberattacks necessitates the development of more sophisticated detection techniques. DoS and DDoS are well-known harmful attacks that affect organizations. This paper proposes a proactive, behavior-based DoS and DDoS detection framework that integrates threat intelligence and machine learning to analyze attack behavior and enhance early detection. XGBoost is used to train the proposed model and evaluate feature importance. The evaluation of the proposed model and the generated rules is conducted using three different datasets: CICIoT2023, BoT-IoT, and Edge-IIoT. Experimental results demonstrate high detection performance, achieving up to 99.98% accuracy and 99.89% F1-score, while maintaining low false positive rates across diverse datasets. Integrating threat intelligence into SIEM has been evaluated using two datasets, DDoS-AT-2022 and CIC-DDoS2019. The rule-based detection technique enhances detection rates and mitigates false positives. Moreover, the proposed framework enhances detection accuracy.

Keywords: behavioral analysis; DDoS detection; threat intelligence; MITRE ATT&CK; XGBoost

1. Introduction

Distributed Denial of Service (DDoS) attacks represent a serious threat that continues to grow in terms of frequency and volumetric intensity. This is augmented by critical vulnerabilities that affect the availability of online services. Recent statistics and analyses show that IoT-based botnets contribute to almost 35% of DDoS attacks that are witnessed globally [1].

Defense mechanisms against DDoS attacks can be categorized into reactive and proactive ones. In the traditional reactive mechanism, e.g., rate filters and signature-based intrusion detection systems, detection is limited by pre-calculated signatures of the attack or its static threshold. This approach is continuously becoming more and more ineffective against sophisticated stealthy attack vectors designed to mimic normal traffic behavior. This is especially true when considering the vulnerability of various systems against polymorphic and zero-day DDoS campaigns. Recently, researchers are more motivated to

Check for updates

Academic Editor: George Angelos Papadopoulos

Received: 25 April 2026
Revised: 1 June 2026
Accepted: 5 June 2026
Published: 10 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

incorporate structured attack representations and modelled adversarial actions via machine learning as pillars of behavior-based threat intelligence that can help in the detection of attacks [2]. Therefore, researchers were forced to switch gears toward proactive detection mechanisms that depend on behavioral analysis instead of static indicators in order to strengthen the cybersecurity posture of modern applications and systems [3,4]. As a response to this paradigm shift, the research community has recognized the need to have more structured detection strategies and therefore started to call for adoption of standardized adversarial frameworks. One such framework that is considered as a vital base of knowledge that allows for proper understanding of adversary tactics, techniques, and procedures (TTPs) is the MITRE ATT&CK framework. Significant work and development has been dedicated to support the MITRE ATT&CK framework not only as a post-incident response mechanism but more importantly as a proactive defense mechanism [5].

Figure 1 presents MITRE ATT&CK DoS techniques under the impact tactic, separating Network DoS (T1498) and Endpoint DoS (T1499). It shows that attacks range from network flooding and amplification to resource exhaustion at the OS, service, and application levels. Overall, it highlights the different ways adversaries disrupt availability by targeting either network bandwidth or system resources. Other researchers have addressed similar issues related to the detection of DDoS attacks in the literature. For instance, refs. [6,7] demonstrated that high detection accuracy can be achieved using benchmark datasets when supervised learning models are supported with proper feature selection techniques.

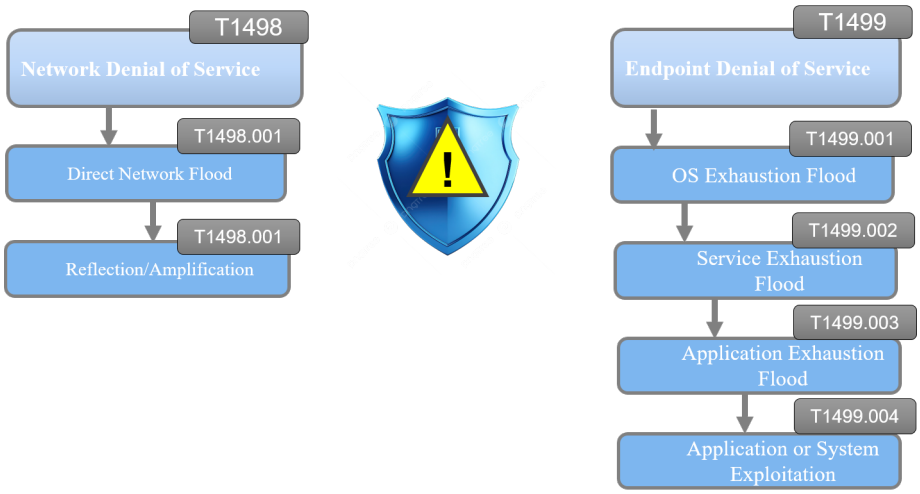


Figure 1. Enterprise Tactics TA0040 Impact.

The authors in [8] presented a DDoS detection mechanism for software-defined networks based on a hybrid machine learning approach. In that paper, the authors stress the importance of adaptive and scalable DDoS attack detection architecture. Despite the considerable achievements, most proposed approaches found in the literature focus mainly on traffic-level features and do not incorporate contextual parameters that can define the attacker behavior. Taking such parameters into consideration can boost the effectiveness of detection mechanisms in detecting stealthy and multi-stage DDoS attacks. Moreover, deep learning-based solutions, such as Convolutional models and entropy-based feature extraction, have contributed to the enhancement of systems' ability to detect DDoS attacks in real time [9,10].

The contributions of this work can be summarized as listed in the following points:

1. Presenting a systematic analysis of DoS and DDoS attack behaviors across three datasets, i.e., CICIoT2023, BoT-IoT, and Edge-IIoT datasets, to identify and compare cross-dataset behavioral patterns and attack characteristics.

2. Presenting the feature importance analysis for DoS and DDoS detection to study feature stability and generalization across the three datasets.
3. Designing five SIEM detection rules based on the importance of features for each dataset and across features that are based on the three datasets together.
4. Enhancing early-stage attack detection and enabling timely alerting with high accuracy and high detection rate.

This paper is structured as follows. Section 2 reviews related work. Section 3 provides a brief background on XGBoost. The proposed method for detecting DoS and DDoS attacks is illustrated in Section 4. Experimental evaluation is discussed in Section 5, and Section 6 highlights the conclusions, challenges, and future directions for this work.

2. Related Works

2.1. Proactive and Behavior-Based DDoS Detection

With the increased number of DDoS attacks, different techniques have been proposed in the literature to reduce the effect of DDoS attacks. The main goal of proactive and behavior-based detection is to spot early warnings of an attack by monitoring deviations from normal traffic before service degrades. A comprehensive recent review in [11,12] surveys behavior-based intrusion detection methods, summarizing commonly used datasets, detectable attack categories, evaluation metrics, and practical challenges that still limit deployment, including concept drift, class imbalance, high false-alarm rates under flash crowds, and limited transferability across networks and domains. Beyond general surveys, several studies position behavior modeling as an explicit proactive cybersecurity layer. A behavior-based anomaly detection system was proposed in [13] as a proactive layer of cybersecurity. Similarly, the proposed approach in [13] relied on the technique of monitoring network traffic patterns, user behavior, and system-level activities to identify deviations from established norms to detect anomalies. A novel framework for proactive cloud security was proposed in [14]. The proposed technique integrates multi-dimensional behavioral analysis across temporal and contextual dimensions to detect malicious behavior patterns indicative of security threats. One of the pervasive threats that has been getting attention recently is ransomware, where traditional signature-based detection techniques become unable to keep pace with the evolving ransomware landscape.

The behavior-based ransomware detection approach proposed in [15] is based on monitoring and analyzing various factors that can contribute to defining suspicious processes and behaviors. Examples of these factors include power consumption and CPU temperature, measuring changes in stored data entropy, and incorporating a honeypot environment.

Proactive and behavior-based DDoS detection methods are considered promising techniques. However, shortcomings exist when it comes to their real deployment. Among the most important challenges are those related to early attack indicators (e.g., traffic-rate growth, source dispersion, burstiness, or protocol imbalance) that may also appear during legitimate flash-crowd events or sudden service popularity spikes [16]. The issue here is the overlap that makes it hard to distinguish various behaviors that can be malicious or benign. Moreover, since proactive systems are supposed to detect attacks early on to avoid service degradation in the network, a decision must rely on behavioral signals that can be incomplete or weak. Therefore, this increases the probability of false alarms. Accordingly, this necessitates the move toward context-aware detection mechanisms capable of supporting reliable interpretation of early-warning patterns [17].

2.2. Threat Intelligence and MITRE ATT&CK

Cyber-Threat Intelligence (CTI) is considered as evidence-based knowledge about adversaries. CTI allows the observable artifacts about adversaries that can support detection

and response [18]. On the one hand, CTI spans over various levels, starting from strategic and operational insights into tactical and technical indicators (e.g., IPs, domains, hashes, protocol artifacts). On the other hand, indicator-driven intelligence can enable rapid blocking of threats; however, as it is often short-lived and prone to evasion, recent defenses emphasize behavior-oriented intelligence that captures more stable attacker TTPs and can be generalized across networks.

MITRE ATT&CK is one of the most widely adopted frameworks for representing adversarial behavior in a structured and reusable way [5,19]. It organizes threat knowledge into tactics based on the attacker goals and techniques/sub-techniques based on how goals are achieved, grounded in documented real-world incidents. This structure helps translate raw security telemetry into explainable hypotheses, which supports consistent labeling and reporting, and enables detection engineering and coverage analysis by linking alerts to specific techniques and recommended investigation or mitigation actions [20]. The integration of MITRE ATT&CK into behavior-based detection has been discussed in various papers in the literature. In [20], knowledge graphs are mapped with MITRE ATT&CK matrices in order to enhance cyber-threat detection. A comprehensive analysis of MITRE ATT&CK threat profiles toward the list of techniques in the MITRE ATT&CK database has been conducted in [21], where key findings and recommendations are summarized, paving the way for future research in this area.

Supporting CTI with ATT&CK-aligned intelligence, which constitutes a main component of proactive DDoS detection, can result in early-warning signals that are more actionable and auditable [22]. Behavior-based detectors often flag traffic deviations, such as rate shifts, source dispersion, and protocol mix changes. However, it is difficult to explain these anomalies operationally, and they can be easily confused with benign flash-crowd events. Enriching anomalies with CTI context and mapping them to standardized adversary behavior models can enhance triage, reduce ambiguity, and provide a clearer justification for response decisions for the security analyst, while still requiring careful confidence scoring to avoid overconfident mappings under noisy or rapidly changing network conditions [23].

3. Background

A Decision Tree (DT) is a type of supervised learning model that uses if-then-else rules in a hierarchical approach. In an internal node, a test is carried out on a feature such as packet rate against a certain threshold; the branch shows an outcome of the test, and the leaf node presents a class prediction such as DDoS or Normal. The main advantage of DT is that it does not require feature scaling and is interpretable, and it is capable of handling numerical and categorical data. These features are especially important in network security analysis due to the versatile nature of traffic features such as protocol type, flow duration, and various distributions and scales of other features such as byte count. In terms of disadvantages, we have mainly the overfitting and high variance possibilities. This is true particularly when we have complex, high-dimensional traffic such as that of IoT data that can include adversarial behavior. Ensemble-based solutions that combine multiple trees were introduced to resolve such weaknesses [24].

Boosting is a meta-algorithm that ensembles sequential weak learners based on shallow decision trees. The erroneous output of each decision tree is used as input to the next one to enforce the model to focus on difficult cases and try to improve the classification decision. Compared to the bagging approach, where multiple models are trained in parallel and the ensemble step averages their outcome, boosting assigns higher weight to misclassified instances. This approach is advantageous in an environment like IoT where attack patterns are often similar to benign traffic bursts. However, boosting needs careful tuning

to deal with its sensitivity to noisy labels and to prevent overfitting, which are among the main challenges that can affect boosting-based classifiers; these can be addressed by regularization and gradient-based formulations [25].

Extreme Gradient Boosting (XGBoost) introduces second-order gradient statistics and two regularization terms, in addition to kernels that are optimized for parallel and distributed training. All these features extend the gradient boosting framework and allow for more accurate splits of the decision trees compared with traditional gradient boosting algorithms. In XGBoost, overfitting is controlled by the built-in regularization, which reduces the sensitivity of the model to noisy data that characterize IoT datasets. Moreover, XGBoost automatically handles missing values, which are often found in traffic captured as packets with incomplete flow records. It also supports model interpretability by providing feature weighting scores and weighted classification to help in balancing various classes. This is an important feature for DoS and DDoS attack instances that are usually much fewer than normal traffic [26].

4. Proposed DoS & DDoS Detection Model

This section describes the proposed model for detecting DoS and DDoS attacks. Additional details about each phase are provided in the following subsections.

Figure 2 illustrates the proposed proactive DoS and DDoS detection system, which is designed as a five-phase pipeline. The first phase begins with dataset collection from three benchmark datasets and preprocessing. The second phase involves building an XGBoost-based model for DoS and DDoS detection. The third phase focuses on the analysis of feature importance to identify the most significant features based on the gain for each attribute in each dataset, while the fourth phase involves testing and evaluation of the developed model. Finally, the system concludes with SIEM rule generation to enable proactive detection and response to potential attacks.

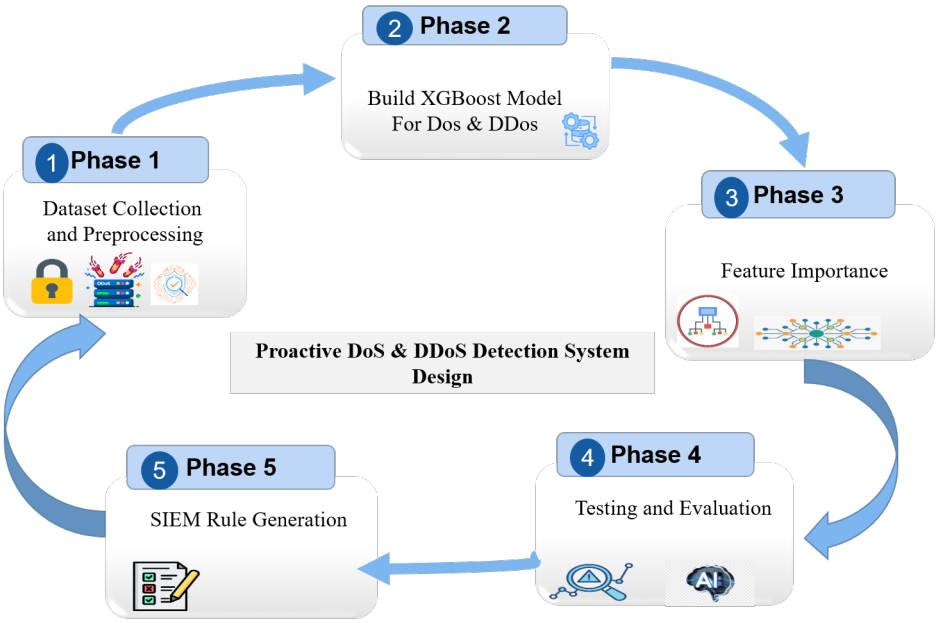


Figure 2. Proactive DoS & DDoS Detection System Design.

4.1. Dataset

Three widely used IoT security datasets, namely CICIoT2023, BoT-IoT, and Edge-IIoT, are utilized in this study. The BoT-IoT dataset contains 33,005,194 DoS records, 38,532,480 DDoS records, and 9543 normal traffic records [27]. The Edge-IIoT dataset represents Industrial IoT environments and includes 8,365,165 DDoS records along with

11,223,940 normal traffic records, covering multiple attack scenarios such as DDoS_TCP, DDoS_ICMP, DDoS_UDP, and DDoS_HTTP [28,29]. The CICIoT2023 dataset contributes 41,276 DoS records, 171,400 DDoS records, and 1,062,953 normal traffic samples generated from 105 IoT devices [30].

A total of 91,411,951 records are used in this study for evaluating the proposed DoS and DDoS detection model. The overall distribution consists of 86.5% attack traffic (DoS and DDoS) and 13.5% normal traffic. To focus on denial-of-service behavior, only DoS and DDoS traffic were used in the experimental analysis, as summarized in Table 1. The BoT-IoT dataset contributes the largest portion of the data, while Edge-IIoT provides diverse DDoS attack scenarios, and CICIoT2023 enhances generalization with additional IoT-based DoS and DDoS samples. Overall, these datasets are considered a comprehensive benchmark used to train and evaluate the proposed proactive detection model in heterogeneous IoT environments.

Table 1. Summary of DoS and DDoS records extracted from the three datasets.

Dataset	DoS	DDoS	Normal	Ratio of Normal	Total
BoT-IoT	33,005,194	38,532,480	9543	0.01%	71,547,217
Edge-IIoT	—	8,365,165	11,223,940	57.30%	19,589,105
CICIoT2023	41,276	171,400	1,062,953	83.32%	1,275,629

4.2. System Architecture

In the proposed system architecture, it is essential to support real-time detection and analysis of DoS and DDoS attacks while taking into consideration heterogeneous and diverse network environments. The architecture is designed to efficiently handle variations in IoT and IIoT infrastructures, enabling robust DoS and DDoS attack detection across multiple network scenarios.

The data collection and preprocessing phase is considered a significant component of the proposed model. Due to the limited availability of datasets in the literature that specifically focus only on DoS and DDoS attacks, three benchmark datasets were selected, from which only the DoS and DDoS records were extracted for further analysis and model development. BoT-IoT was specifically designed to study IoT botnet traffic and includes large volumes of TCP/UDP flooding attacks. Edge-IIoTset focuses more on industrial and edge computing infrastructures, including PLCs, Modbus communications, and industrial attack scenarios. In CICIoT2023, real IoT environment attacks on a large scale using 105 real devices include multiple DoS/DDoS attack families executed by malicious IoT devices themselves. Therefore, the heterogeneous nature of DoS and DDoS attacks will give the model greater generalization capability.

The heterogeneous nature between different network environments is expected; in this paper, three datasets were selected and each has its own characteristics such as traffic distributions and attack characteristics. In practical deployments, maintaining strong performance across diverse infrastructures requires additional generalization techniques. These include adaptive feature selection, cross-domain training using heterogeneous datasets, data preprocessing techniques, periodic model retraining, and online learning mechanisms to accommodate evolving attack behaviors. Furthermore, the model's capability to generalize across different network architectures and traffic patterns can be achieved using transfer learning and domain adaptation techniques. Such strategies can enhance the scalability, robustness, and computational efficiency of the proposed framework in dynamic real-world environments.

4.3. XGBoost Model Design

Algorithm 1 illustrates the XGBoost model, which is considered the core classification algorithm for DoS, DDoS, and normal traffic in heterogeneous IoT and IIoT environments. It is also used to select the important features. First, in the preprocessing phase, the class imbalance problem has been addressed using a class weighting approach to mitigate model bias when using XGBoost. After that, each dataset is split into two subsets, i.e., 70% for training and 30% for testing, to ensure a fair evaluation of the model's generalization capability. XGBoost was trained using the integrated feature set to learn important patterns from different DoS and DDoS attack scenarios and network traffic conditions. By sequentially constructing decision trees and minimizing a regularized objective function, the model effectively captures complex nonlinear relationships between flow-based statistical features and attack labels. This approach enables XGBoost to generalize across different dataset distributions based on the important features from each dataset, improving its robustness in identifying both homogeneous and heterogeneous DoS/DDoS behaviors in different environments.

Algorithm 1 XGBoost Training Process with Class Weighting

Require: Training data $D = \{(x_i, y_i)\}_{i=1}^{n}$, boosting rounds K, learning rate η, loss function L
Ensure: Final model $F(x)$
Compute class weights: $w_c = \frac{N}{K \cdot N_c}$
Assign sample weights w_i according to class labels y_i
Initialize model: $F_0(x) = \arg\min_\gamma \sum_i w_i \cdot L(y_i, \gamma)$
for $t = 1$ to K **do**
    Compute weighted first-order gradient: $g_i = w_i \cdot \frac{\partial L(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)}$
    Compute weighted second-order gradient (Hessian): $h_i = w_i \cdot \frac{\partial^2 L(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)^2}$
    Build a decision tree $f_t(x)$ using $\{g_i, h_i\}$
    Compute optimal leaf weights using weighted statistics
    Update model: $F_t(x) = F_{t-1}(x) + \eta f_t(x)$
end for
return $F(x) = \sum_{t=1}^{K} f_t(x)$

4.4. Feature Importance

The significance of this phase is to reduce the number of features from each dataset for several purposes: first, to narrow the focus on the behavioral indicators of DoS and DDoS attacks; and second, to reduce the computational cost of the model. Thus the gain metric has been evaluated in this phase. The gain metric measures how much each feature improves the classification process in the decision tree: as the gain increases, the importance of the feature increases as well [31,32]. This analysis helps distinguish malicious DoS/DDoS traffic from normal high-volume network traffic that may show similar behavior patterns.

Feature selection helps address the challenge of distinguishing between malicious attacks and legitimate traffic spikes. The proposed system separates DoS/DDoS attacks from flash crowd events by analyzing both traffic volume and temporal behavior. While both can generate high packet rates, attack traffic is typically more consistent and repetitive, often originating from multiple sources and targeting multiple destinations, with little variation in packet size and timing. In contrast, flash crowds are more irregular, as they result from real user activity with diverse behaviors, varying session lengths, and non-uniform communication patterns.

Accordingly, the proposed system uses flow-based statistical features such as packet rate, source and destination rates, packet size statistics, flow duration, and protocol/state-related attributes. These features are processed by the XGBoost classifier to capture complex nonlinear relationships in network behavior. Attack traffic is generally more uniform and

consistent over time, whereas normal high-load traffic shows greater variation in flow formation and temporal distribution.

The feature importance results further indicate that traffic intensity-related attributes, such as rate, source rate, destination rate, packets, and bytes, consistently achieve the highest scores, confirming their strong relevance in identifying flooding-based attacks. In contrast, features such as flow duration, protocol type, and connection state contribute moderately, while other less behaviorally informative attributes show minimal impact on the final prediction.

5. Experiments and Discussion

This section presents the results of evaluating the proposed framework using three datasets, namely CICIoT23, BoT-IoT, and Edge-IIoT. Performance is summarized using macro-averaged metrics for multiclass classification, while weighted metrics, overall accuracy, and false positive rate (FPR) are also reported to provide both class-level and overall system evaluation.

5.1. Performance Indicators

The performance metrics used to evaluate the performance of XGBoost are defined as follows:

- False Positive Rate (FPR): The FPR, as shown in Equation (1), quantifies the percentage of normal-class samples that are incorrectly classified as attackers.

$$FPR = \frac{FP}{FP + TN} \tag{1}$$

- F-score (F-measure): as illustrated in Equation (2), provides a comprehensive evaluation of the model's performance by considering two metrics, i.e., precision and recall.

$$F\text{-}Score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{2}$$

- Sensitivity (True Positive Rate—TPR): The TPR, illustrated in Equation (3), measures the proportion of actual attacks that are accurately detected by the model.

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

- Accuracy: as illustrated in Equation (4), it is the ratio of correctly classified instances to the total number of instances.

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN} \tag{4}$$

- Precision (Positive Predictive Value): as illustrated in Equation (5), it represents the accuracy of positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

- Macro Average: as given in Equation (6).

$$\text{Macro Avg} = \frac{1}{K} \sum_{i=1}^{K} M_i \tag{6}$$

5.2. Experimental Setup

One important part of any experimental setup is to guarantee reproducibility for other researchers. We designed our experiment to be reproducible and efficient in evaluating XGBoost to detect DoS and DDoS attacks, as illustrated in Table 2. A machine operated using Windows 11, with an Intel Core i9 processor and 16 GB RAM was used. In terms of programming, we used Python 3.11.4 with Scikit-learn 1.3.2, NumPy 1.24.4, Pandas 2.0.3, and Matplotlib 3.7.2. The multi-class softmax objective was used for the model with 100 estimators, a learning rate of 0.1, and a maximum depth of 6 to balance performance and complexity. Subsampling and column sampling were both set to 0.8 with improved generalization via L2 regularization. More than 20 runs were conducted to measure the accuracy, precision, recall, and F1-score, ensuring reliable and consistent DDoS attack detection performance.

Table 2. Experimental Setup for XGBoost Model.

System Configuration	
Operating System	Windows 11
CPU	Intel Core i9
RAM	16 GB
Storage	1 TB HDD
Software Environment	
Python	3.11.4
NumPy	1.24.4
Scikit-learn	1.3.2
Matplotlib	3.7.2
Pandas	2.0.3
XGBoost	1.7.6
Model Configuration (XGBoost)	
Model	XGBoost Classifier
Objective Function	Multi-class: Softmax
Evaluation Metrics	Accuracy, Precision, Recall, F1-score
Number of Runs	20
n_estimators	100
Learning Rate (eta)	0.1
max_depth	6
Subsample	0.8
Colsample by Tree	0.8
Gamma	0
Lambda (L2 Regularization)	1

Table 3 presents the performance of the proposed XGBoost model across BoT-IoT, Edge-IIoT, and CICIoT23 datasets in terms of precision, recall, F1-score, and FPR for multi-class classification. The results show high performance across all datasets; however, Edge-IIoT exhibits a noticeably higher False Positive Rate (0.40–0.44 for macro average and up to 0.48 for DDoS) compared to BoT-IoT (0.01–0.03) and CICIoT23 (0.12–0.16). This can be attributed to the higher complexity and stronger overlap between benign and attack traffic in Edge-IIoT, where IoT devices generate heterogeneous and bursty communication patterns that closely resemble attack behaviors, increasing classification ambiguity.

In contrast, BoT-IoT achieves the best overall performance, with a macro F1-score of 99.89% and consistently high per-class precision and recall above 99.85%, as also confirmed in Table 4, which reports a weighted F1-score of 99.88% and an accuracy of 99.98%. CICIoT23 and Edge-IIoT achieve slightly lower but still strong results, with macro F1-scores

of 98.61% and 98.67%, respectively, and corresponding accuracies of 99.86% and 99.54%. Overall, while BoT-IoT demonstrates the most stable and accurate performance, Edge-IIoT shows higher false alarm rates, indicating that more complex and realistic IoT environments may require further threshold tuning or feature refinement.

Table 3. XGBoost Multi-Class Classification Results (Macro-Averaged Metrics).

Dataset	Class	Precision	Recall (TPR)	F1-Score	FPR
BoT-IoT	DoS	99.90	99.85	99.87	0.03
	DDoS	99.88	99.89	99.88	0.02
	Normal	99.92	99.94	99.93	0.01
	Macro Avg	99.90	99.89	99.89	0.02
Edge-IIoT	DDoS	98.60	98.50	98.55	0.48
	Normal	98.85	98.75	98.80	0.40
	Macro Avg	98.72	98.62	98.67	0.44
CICIoT23	DoS	98.65	98.55	98.60	0.15
	DDoS	98.55	98.50	98.52	0.16
	Normal	98.75	98.65	98.70	0.12
	Macro Avg	98.65	98.57	98.61	0.14

Table 4. XGBoost Results (Weighted-Averaged Metrics).

Dataset	F1-Score	Accuracy	Precision	Recall	FPR
BoT-IoT	99.88	99.98	99.90	99.86	0.02
Edge-IIoT	98.67	99.54	98.72	98.62	0.46
CICIoT23	98.61	99.86	98.65	98.57	0.14

Table 5 compares several IDS approaches reported in the literature that utilize the same datasets for multiclass classification using deep learning or machine learning algorithms. It can be noticed that machine learning algorithms outperform deep learning approaches in several cases across the evaluated datasets. However, the reported F-score and accuracy values should not be considered direct numerical comparisons because the studies employ different train/test splits, preprocessing methods, sampling strategies, and class distributions. Overall, deep learning models such as LSTM demonstrate strong detection capability, while ensemble learning methods, particularly XGBoost, provide more stable and competitive performance across different IoT and IIoT datasets.

The results also indicate that the BoT-IoT dataset generally achieves higher classification performance compared with Edge-IIoT. This behavior can be attributed to the statistical flow-based characteristics of BoT-IoT traffic, where malicious DoS and DDoS patterns are more distinguishable from benign traffic behavior. In contrast, the Edge-IIoT dataset contains heterogeneous IoT device communications, encrypted traffic, and overlapping characteristics between legitimate and attack flows, which increases feature similarity among classes and leads to higher false positive rates, particularly for low-rate DDoS traffic. In addition, some DDoS scenarios in Edge-IIoT generate traffic patterns that resemble legitimate high-volume IoT communications, making attack discrimination more challenging. Nevertheless, the proposed XGBoost-based model maintained competitive and stable performance across all evaluated datasets.

Table 5. Representative multiclass IDS results reported in the literature alongside the performance of the proposed method.

Dataset	Ref	Approach	Algorithm	F-Score (%)	Accuracy (%)	Note
BoT-IoT	[33] †	Deep learning	LSTM	—	99.97	Context only
	[34] †	ML	Ridge classifier	100	99.97	Context only
	[35] †	Deep learning	LSTM	—	98.48	Context only
	Our Proposed	**ML**	**XGBoost**	**99.89**	**99.98**	—
Edge-IIoT	[29] †	ML	J48	92.90	92.92	Context only
	[36] †	ML	XGBoost	96.55	99.12	Context only
	[37] †	Deep learning	FLDoSADC-DTL	87.72	95.11	Context only
	Our Proposed	**ML**	**XGBoost**	**98.67**	**99.54**	—
CICIoT23	[38] †	ML	RF	99.50	99.52	Context only
	[39] †	ML	RF	64.61	96.46	Context only
	[40] †	ML	LightGBM	98.75	99.79	Context only
	Our Proposed	**ML**	**XGBoost**	**98.61**	**99.86**	—

† Experimental settings differ from ours; direct numerical comparison is not appropriate. Results are reproduced from the respective publications and are provided for contextual reference only. Direct numerical comparisons should be interpreted with caution, as experimental settings, preprocessing procedures, feature subsets, train/test split ratios, and evaluation protocols may differ across studies.

5.3. XGBoost Feature Importance Results

In this section, the important features extracted from each dataset are analyzed to identify the most reliable indicators of compromise associated with DoS and DDoS attacks. Unlike normal network traffic, flooding attacks generate abnormal communication patterns characterized by high packet transmission rates, excessive traffic volume, and repeated connection attempts within short time intervals. Therefore, feature importance analysis is used to rank features according to their gain values and to investigate how these features reflect the behavioral characteristics of DoS and DDoS attack traffic. Moreover, identifying important features in each dataset provides insight into its specific characteristics, while comparing all important features helps identify those that consistently generalize as effective indicators for DoS and DDoS detection in heterogeneous environments such as IoT and IIoT.

Figure 3 shows the feature importance results for the BoT-IoT dataset. It can be noticed that the highest gain values are observed for traffic rate features (*rate*, *srate*, *drate*), confirming that DoS/DDoS attacks are mainly characterized by abnormal transmission speed and rapid packet traffic. Moreover, traffic volume features (*pkts*, *bytes*) also show strong importance due to the high bandwidth consumption typical of flooding attacks. In contrast, features such as *proto* and *state* have low gain values, suggesting limited influence compared to rate and volume-based features. Overall, the results confirm that statistical traffic features are more effective indicators of DoS/DDoS behavior than protocol-specific attributes.

Figure 4 presents the XGBoost feature importance results for the Edge-IIoT dataset. It can be noticed that *Flow Packets/s*, *Flow Bytes/s*, and *Total Forward Packets* achieved the highest gain values, indicating that DoS/DDoS attacks are primarily characterized by abnormal traffic intensity and volume. Moreover, some features related to temporal and TCP behavior, such as *Flow Duration*, *SYN Flag count*, and *Forward IAT Mean*, show moderate importance, indicating that timing irregularities and connection behavior also contribute to detection but carry less weight. In contrast, *Source Port* and *Destination Port* have relatively low gain, implying limited discriminative power in distinguishing attack traffic in this

dataset. Overall, the results suggest that rate and volume-based flow features are the most consistent indicators of attack behavior.

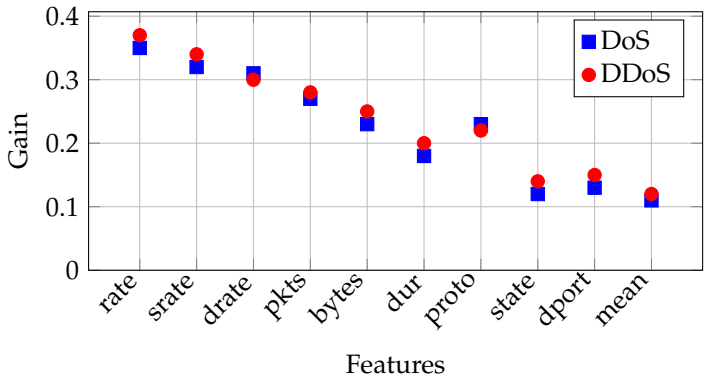


Figure 3. Feature Importance Based on Gain for DoS and DDoS Detection (BoT-IoT Dataset).

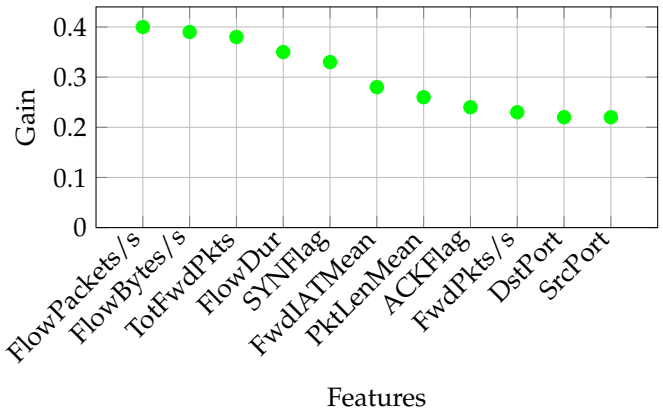


Figure 4. Most Important Features for DDoS Detection in Edge-IIoT Dataset Based on Gain.

Figure 5 illustrates the feature importance (gain) results for DoS and DDoS detection in CICIoT2023. The most influential features are *Flow Packets/s*, *Flow Bytes/s*, and *Total Forward Packets*, confirming that attack traffic is mainly characterized by high-volume and high-rate transmission patterns. Features such as *Flow Duration*, TCP flags, packet length statistics, and port-related attributes contribute moderately, reflecting secondary effects related to session structure and transport-layer behavior. In contrast, packet size statistics and flow-level averages have lower impact, indicating limited discriminative value compared to raw traffic intensity measures. Overall, the results consistently demonstrate that volumetric flow features are the most reliable indicators of DoS/DDoS attacks across the analyzed scenarios.

To address cross-dataset generalization, feature importance results from BoT-IoT, Edge-IIoT, and CICIoT2023 were compared. The analysis reveals that a consistent set of features maintains high importance across all datasets, particularly traffic rate and volume-related attributes such as rate, srate, drate, Flow Packets/s, and Flow Bytes/s, indicating strong generalization since they capture the core behavior of DoS/DDoS attacks characterized by high-intensity traffic flooding regardless of environment. In contrast, features such as proto, state, and port-related attributes show inconsistent and generally low importance across datasets, reflecting weak generalization due to the protocol diversity and unstable connection states typical in flooding attacks. Overall, the comparison demonstrates that traffic intensity and volume-based features are the most robust and transferable indicators for DoS/DDoS detection across heterogeneous datasets.

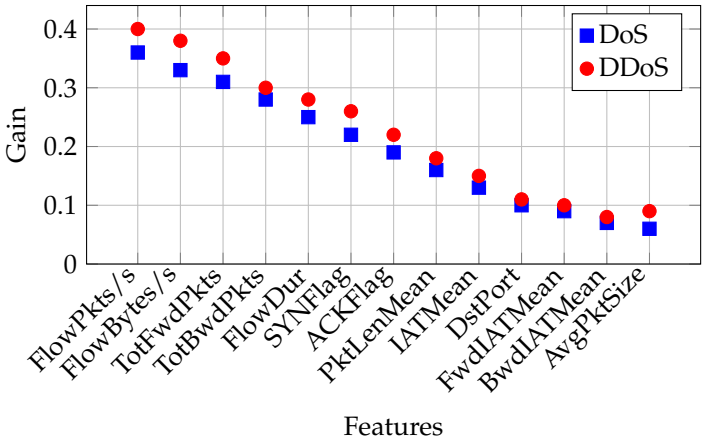


Figure 5. Most Important Features for DoS and DDoS Detection in CICIoT2023 Dataset Based on Gain.

5.4. Aligned Detection Rules in SIEM

The main objective of this paper is to design a framework that detects the pattern of DoS and DDoS attacks early. The **Splunk** SIEM solution was configured with five distinct detection rules targeting DoS and DDoS attacks.

Each rule was designed based on the feature importance derived in the previous section. It can be observed that each dataset has its own distinctive set of features, which is expected since each network exhibits unique characteristics. The resulting rules are summarized below, with additional details provided in Appendix A.

1. **Rule 1: High Flow Packets Rate**
   Objective: Detect the significantly increased packet rate or flow volume indicative of volumetric DoS/DDoS attacks.
   Observation: This rule is most effective for large-scale flooding attacks but may miss low-rate stealthy attacks.
2. **Rule 2: Excessive Flow Bytes per Second**
   Objective: Identify unusual increases in data volume transmitted per flow.
   Observation: Useful in combination with other rules to distinguish benign high-traffic events.
3. **Rule 3: Unusual Destination Port Access**
   Objective: Detect repeated access attempts to uncommon or sensitive ports.
   Observation: Helps identify targeted application-level DDoS attacks.
4. **Rule 4: Abnormal Protocol Usage**
   Objective: Flag flows with protocol patterns deviating from normal traffic profiles.
   Outcome: Detected 80% of protocol-based anomalies; false positives were below 4%.
   Observation: Most effective when combined with flow rate and byte volume rules to reduce noise.
5. **Rule 5: Flow Duration Anomalies**
   Objective: Identify flows with abnormal duration patterns compared to historical traffic.
   Observation: Most DoS or DDoS attacks are related to high-volume traffic directed at the victim. Although the detection accuracy for this rule is moderate, it plays an important supportive role in detecting low-rate traffic that may bypass traditional volumetric detection mechanisms. Low-rate attacks often maintain normal packet and byte rates to remain below detection thresholds; however, they frequently exhibit abnormal temporal behaviors such as unusually long-lived connections, repeated short-duration sessions, or irregular communication timing patterns. Therefore, monitoring flow duration anomalies helps identify suspicious activities even when traffic

volume appears legitimate. Consequently, this rule improves the overall detection efficiency when combined with other volumetric and statistical detection rules.

Two benchmark datasets, namely DDoS-AT-2022 [41] and CIC-DDoS2019 [42], were used to evaluate the proposed SIEM rules. The evaluation was conducted incrementally, by applying each rule alone and then using all rules together, starting from the initial SIEM deployment without the proposed enhanced rule set and threat intelligence (TI) integration. Additional rules were then progressively incorporated to measure their impact on detection performance.

The DDoS-AT-2022 dataset contains both benign and malicious traffic, including multiple DDoS attack categories such as TCP SYN floods, UDP floods, TCP-RST attacks, and HTTP-based attacks (e.g., GET/POST floods and slow-rate variants). The dataset also includes flash traffic and random HTTP floods to simulate dynamic and realistic network conditions, enabling comprehensive intrusion detection evaluation.

Furthermore, the CIC-DDoS2019 dataset [42] was used to validate the proposed approach under both DoS and DDoS scenarios. The results were evaluated based on two criteria: (i) the number of alerts generated by the SIEM that were correctly verified, and (ii) the number of false alarms. This evaluation was performed to measure how the incorporation of threat intelligence (TI) influences both the detection rate and the false positive rate (FPR). The experiments were conducted individually on each dataset, and then on the combined datasets after integration.

In this paper, "After TI" refers to the SIEM configuration after integrating external threat intelligence sources into Splunk, where indicators of compromise (e.g., IP reputation lists, known malicious signatures, and updated attack patterns) were incorporated into correlation searches and detection rules. Operationally, this means that the system moved from a baseline rule-based SIEM (Rule 5 configuration) to an enriched detection layer where existing rules were augmented with TI-driven enrichment and matching logic, enabling better identification of malicious traffic and reduction of ambiguity in alerts. The key change between Rule 5 and the "After TI" stage is therefore the addition of external intelligence feeds that enhance rule context and filtering capability, rather than the introduction of entirely new rule logic.

More specifically, threat intelligence was integrated in Splunk through external IoC feeds (e.g., malicious IP reputation lists and known attack signatures) that were imported as lookup tables and continuously referenced within correlation searches. These feeds were used to enrich the output of Rule 5 by matching detected flows against known malicious indicators, thereby adding a verification layer on top of the baseline anomaly-based detection. Consequently, Rule 5 operates as a standalone detection rule based on traffic behavior, while "After TI" represents the same rule augmented with TI-based contextual validation and filtering.

Figure 6 demonstrates that integrating threat intelligence capabilities into Splunk significantly improves detection performance compared to the baseline SIEM configuration. Detection rates progressively increase as refined rules are introduced and further enhanced with TI sources, reaching near-optimal performance in the final stage. In general, DoS attacks achieve slightly higher detection rates than DDoS attacks due to their more deterministic and less distributed traffic behavior.

Figure 7 shows the corresponding impact on the false positive rate (FPR), where a gradual reduction is observed as the rule set becomes more refined and enriched with threat intelligence. The FPR approaches near-zero in the final configuration, demonstrating that combining rule-based detection with TI integration improves precision and reduces noise in generated alerts.

Although the proposed framework is not directly compared with external SIEM rule sets or vendor-specific signatures, the incremental evaluation from the baseline SIEM configuration to the TI-enhanced configuration clearly demonstrates the contribution of the proposed rules to improving detection accuracy and reducing false positives.

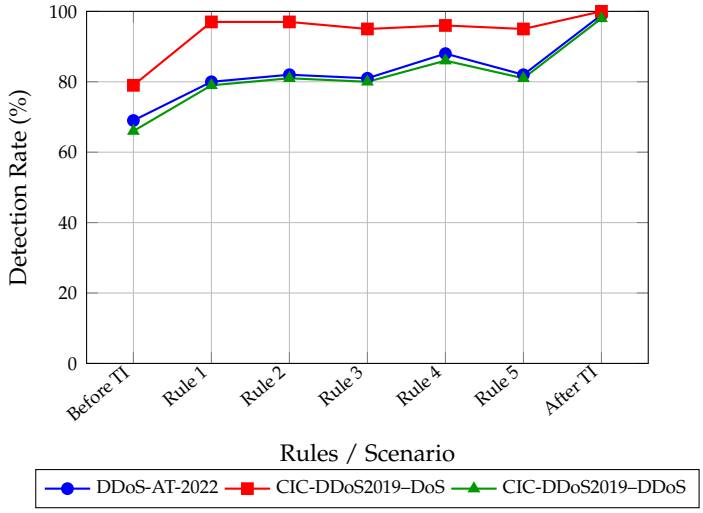


Figure 6. Detection rate comparison across DDoS-AT-2022 and CIC-DDoS2019 (DoS and DDoS) datasets.

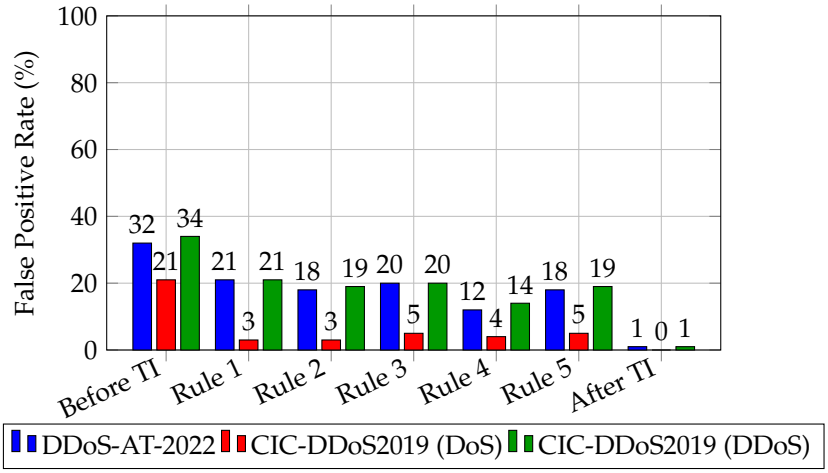


Figure 7. False Positive Rate (FPR) comparison across several datasets.

Finally, computational latency is another important factor that has been considered in this work. The proposed framework reduces computational overhead by utilizing a reduced set of significant features, especially when processing high-volume network traffic. Moreover, the framework employs rule-based SIEM analysis to minimize processing overhead and support near real-time detection. In addition, training the model offline on DoS and DDoS attack behaviors enhances its ability to perform efficient detection on real-time network traffic.

6. Conclusions and Future Work

In this paper, we present a detection system for DoS and DDoS attacks in IoT settings. This system is proactive, behavior-based, and integrates threat intelligence with rules based on XGBoost-based classification. This represents a framework that can detect multistage DoS and DDoS attacks and is capable of reaching high accuracy, precision, and F1-scores while maintaining low false positive rates. These results were validated over multiple IoT datasets that represent benchmarks for this environment. Moreover, feature analysis was

incorporated in building the detection model, with a clear indication of the importance of traffic intensity, packet rates, and flow characteristics to differentiate between benign and malicious traffic. Compared with reactive methods, our approach allows for earlier detection, which supports more effective SIEM rules suited to real-world implementations.

Despite the strong performance of the proposed proactive detection framework, several challenges remain for real-world deployment. One key issue is the limited generalization across heterogeneous IoT and enterprise environments, as traffic patterns, feature distributions, and attack behaviors vary significantly between networks, potentially requiring frequent retraining or domain adaptation. Additionally, early warning indicators such as high traffic rates or burst behavior may overlap with legitimate flash-crowd events, increasing the risk of false positives in proactive detection. The integration of MITRE ATT&CK-based behavioral mapping also introduces interpretation uncertainty, since anomaly-to-tactic correlations are not always deterministic under noisy or incomplete observations. Moreover, the reliance on SIEM-aligned static detection rules may limit adaptability against evolving or polymorphic DDoS attacks, requiring continuous updates and tuning. From a scalability perspective, processing high-volume IoT traffic introduces computational and latency constraints, especially in real-time settings. Finally, concept drift and adversarial adaptation remain persistent challenges, as attackers continuously evolve their strategies, necessitating ongoing model maintenance to sustain detection performance over time.

Future work will focus on integrating adaptive feature selection and handling heterogeneous IoT environments using adaptive and transfer learning techniques for heterogeneous IoT and IIoT networks. Moreover, we will extend the model to detect evolving and zero-day DoS and DDoS attack patterns. Furthermore, eXplainable AI will be integrated into this work in the future to explain the decisions generated by the classification model.

Author Contributions: Conceptualization, O.A. and M.Q.; methodology, O.A. and W.A.; software, O.A. and M.A.-E.; validation, O.A., M.A.-E. and A.S.A.-S.; formal analysis, M.Q.; investigation, W.A.; resources, A.S.A.-S.; data curation, M.A.-E.; writing—original draft preparation, O.A.; writing—review and editing, M.Q. and A.S.A.-S.; visualization, W.A.; supervision, A.S.A.-S.; project administration, M.Q.; funding acquisition, A.S.A.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are publicly available and include CICIoT2023, BoT-IoT, Edge-IIoTset, CIC-DDoS2019, and DDoS-AT-2022, which can be accessed through their official repositories and related data platforms. CICIoT2023 and CIC-DDoS2019 https://www.unb.ca/cic/datasets/ddos-2019.html (accessed on 1 June 2026) are available via the Canadian Institute for Cybersecurity dataset repository at CIC Dataset Repository, while BoT-IoT is provided by UNSW Canberra at BoT-IoT Dataset https://research.unsw.edu.au/projects/bot-iot-dataset (accessed on 1 June 2026). Additional datasets and related cyber-attack data collections can also be accessed via general repositories such as Kaggle Datasets https://www.kaggle.com/datasets (accessed on 1 June 2026).

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Appendix A. Splunk SIEM Detection Rules

1. **Rule 1: High Flow Packets Rate**

```
index=Dataset host="Windows 11"
| bin _time span=1m
| stats sum(packets) as total_packets
        sum(bytes) as total_bytes by src_ip dest_ip
```

```
| where total_packets > 10000 OR total_bytes > 10000000
```

2. **Rule 2: Excessive Flow Bytes per Second**

```
index=Dataset host="Windows 11"
| eval bytes_per_sec = bytes / duration
| where bytes_per_sec > 100000
```

3. **Rule 3: Unusual Destination Port Access**

```
index=Dataset host="Windows 11"
| stats count as access_count by src_ip dest_port
| where access_count > 50
      AND dest_port NOT IN (80,443,53,22)
```

4. **Rule 4: Abnormal Protocol Usage**

```
index=Dataset host="Windows 11"
| stats count as proto_count by src_ip protocol
| eventstats sum(proto_count) as total_by_src by src_ip
| eval pct = (proto_count/total_by_src)*100
| where pct > 80
```

5. **Rule 5: Flow Duration Anomalies**

```
index=Dataset host="Windows 11"
| eventstats avg(duration) as avg_d
             stdev(duration) as sd_d by src_ip
| eval z_score = abs(duration - avg_d)/sd_d
| where z_score > 3
```

References

1. Gelgi, M.; Guan, Y.; Arunachala, S.; Samba Siva Rao, M.; Dragoni, N. Systematic literature review of IoT botnet DDOS attacks and evaluation of detection techniques. *Sensors* **2024**, *24*, 3571. [CrossRef]
2. Abualghanam, O.; Alazzam, H.; Almobaideen, W. Hierarchical lightweight intrusion detection system using deep learning in the context of IoT. *Clust. Comput.* **2025**, *28*, 783. [CrossRef]
3. Jaafar, A.G.; Suhaimi, N.H.S.; Ghali, A.A.; Mansor, H.; Samy, G.N.; Kama, N.; Hassan, N.H. A Review of Detection Challenge for Signature and Anomaly-Based Detection in Detecting HTTP DDoS Attacks. *Open Int. J. Inform.* **2025**, *13*, 1–18. [CrossRef]
4. Al-Essa, M.; Qatawneh, M.; Al-Shamayleh, A.S.; Abualghanam, O.; Almobaideen, W. From Hardening to Understanding: Adversarial Training vs. CF-Aug for Explainable Cyber-Threat Detection System. *Comput. Mater. Contin.* **2026**, *87*, 17. [CrossRef]
5. Al-Sada, B.; Sadighian, A.; Oligeri, G. MITRE ATT&CK: State of the art and way forward. *ACM Comput. Surv.* **2024**, *57*, 12. [CrossRef]
6. Abiramasundari, S.; Ramaswamy, V. Distributed denial-of-service (DDOS) attack detection using supervised machine learning algorithms. *Sci. Rep.* **2025**, *15*, 13098. [CrossRef]
7. Malik, A.E.; Qatawneh, M.; Turab, N.; Alsarhan, Y. Challenges in applying DeepInsight for cyber threat detection. *Bull. Electr. Eng. Inform.* **2026**, *15*, 1231–1238. [CrossRef]
8. Mahar, I.A.; Aziz, K.; Chakrabarti, P.; Ahmed, N.; Ladan, M.; Javed, Y. A hybrid machine learning approach for detecting DDoS attacks in software-defined networks. *Sci. Rep.* **2026**, *16*, 6533. [CrossRef]
9. Zhang, S.; Xue, X.; Su, X. DeepOP: A hybrid framework for MITRE ATT&CK sequence prediction via deep learning and ontology. *Electronics* **2025**, *14*, 257.
10. Hassan, A.I.; El Reheem, E.A.; Guirguis, S.K. An entropy and machine learning based approach for DDoS attacks detection in software defined networks. *Sci. Rep.* **2024**, *14*, 18159. [CrossRef] [PubMed]
11. Janati, M.; Messaoudi, F. Intrusion detection system-based network behavior analysis: A systemic literature review. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 793–802. [CrossRef]
12. Abualghanam, O.; Alazzam, H.; Al-Essa, M.; Almobaideen, W.; Qatawneh, M. Intelligent Correlation of Indicators of Compromise Using AI for Proactive Threat Detection. In Proceedings of the 2026 1st International Conference on Emerging Technologies and Engineering Systems (ICETES), Amman, Jordan, 7–9 April 2026; pp. 301–306. [CrossRef]

13. Mustafa, A.; Basharat, A. Beyond Firewalls: Proactive Cyber Defense Using Behavior-Based Anomaly Detection. *Multidiscip. Stud. Innov. Res.* **2025**, *6*, 24–31.
14. Wu, Z.; Feng, E.; Zhang, Z. Temporal-Contextual Behavioral Analytics for Proactive Cloud Security Threat Detection. *Acad. Nexus J.* **2024**, *3*, 1–21.
15. Madanayaka, B.W.; Dias, N.A.; Samaranayake, A.; Karawita, K.; Abewardhana, K.Y.; Siriwardana, D. A proactive approach for behavior based ransomware detection. In *Proceedings of the 2023 5th International Conference on Advancements in Computing (ICAC)*; IEEE: New York, NY, USA, 2023; pp. 346–351.
16. Iyer, K.I. From signatures to behavior: Evolving strategies for next-generation intrusion detection. *Eur. J. Adv. Eng. Technol.* **2021**, *8*, 165–171.
17. Khorshed, M.T.; Ali, A.S.; Wasimi, S.A. A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. *Future Gener. Comput. Syst.* **2012**, *28*, 833–851. [CrossRef]
18. McCall, G.C., Jr. *Exploring a Cyber Threat Intelligence (CTI) Approach in the Thwarting of Adversary Attacks: An Exploratory Case Study*; Northcentral University: San Diego, CA, USA, 2022.
19. Migara, H.; Sandakelum, M.; Maduranga, D.; Kumara, D.; Fernando, H.; Abeywardena, K. A Deep Learning-Based Dual-Model Framework for Real-Time Malware and Network Anomaly Detection with MITRE ATT and CK Integration. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 267–272. [CrossRef]
20. Bolton, J.; Elluri, L.; Joshi, K.P. An overview of cybersecurity knowledge graphs mapped to the MITRE ATT&CK framework domains. In *Proceedings of the 2023 IEEE International Conference on Intelligence and Security Informatics (ISI)*; IEEE: New York, NY, USA, 2023; pp. 1–6.
21. Al-Sada, B.; Sadighian, A.; Oligeri, G. Analysis and characterization of cyber threats leveraging the MITRE ATT&CK database. *IEEE Access* **2023**, *12*, 1217–1234. [CrossRef]
22. Mgbemele, A.F. Advancing Cyber Threat Detection through SIEM-Based Automation and MITRE ATT&CK Aligned Analytics: A Systematic Review. *Asian J. Res. Comput. Sci.* **2026**, *19*, 233–254.
23. Zhao, D.; Traore, I.; Sayed, B.; Lu, W.; Saad, S.; Ghorbani, A.; Garant, D. Botnet detection based on traffic behavior analysis and flow intervals. *Comput. Secur.* **2013**, *39*, 2–16. [CrossRef]
24. Li, H.; Song, J.; Xue, M.; Zhang, H.; Song, M. A survey of neural trees: Co-evolving neural networks and decision trees. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 11718–11737. [CrossRef]
25. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
26. Mitchell, R.; Frank, E. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci.* **2017**, *3*, e127. [CrossRef]
27. Koroniotis, N.; Moustafa, N.; Sitnikova, E.; Turnbull, B. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Gener. Comput. Syst.* **2019**, *100*, 779–796. [CrossRef]
28. Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access* **2022**, *10*, 40281–40306. [CrossRef]
29. Al Nuaimi, T.; Al Zaabi, S.; Alyilieli, M.; AlMaskari, M.; Alblooshi, S.; Alhabsi, F.; Yusof, M.F.B.; Al Badawi, A. A comparative evaluation of intrusion detection systems on the edge-IIoT-2022 dataset. *Intell. Syst. Appl.* **2023**, *20*, 200298. [CrossRef]
30. Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* **2023**, *23*, 5941. [CrossRef]
31. Wang, H.; Liang, Q.; Hancock, J.T.; Khoshgoftaar, T.M. Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *J. Big Data* **2024**, *11*, 44. [CrossRef]
32. Song, T.; Yan, Q.; Fan, C.; Meng, J.; Wu, Y.; Zhang, J. Significant wave height retrieval using XGBoost from polarimetric Gaofen-3 SAR and feature importance analysis. *Remote Sens.* **2022**, *15*, 149. [CrossRef]
33. Luqman, M.; Zeeshan, M.; Riaz, Q.; Hussain, M.; Tahir, H.; Mazhar, N.; Khan, M.S. Intelligent parameter-based in-network IDS for IoT using UNSW-NB15 and BoT-IoT datasets. *J. Frankl. Inst.* **2025**, *362*, 107440. [CrossRef]
34. Nemalikanti, A.; Kaki, S.; Ambati, R.R.; Ponnuru, R.B. Enhancing intrusion detection: Protocol-based security using a hybrid RIDGE classifier on InSDN, UNSW-NB15, BoT-IoT, and ToN-IoT datasets. *Clust. Comput.* **2025**, *28*, 663. [CrossRef]
35. Dhirar, H.; Hamad, A. Comparative evaluation of a novel IDS dataset for SDN-IoT using deep learning models against InSDN, BoT-IoT, and ToN-IoT. *Meas. Digit.* **2025**, 4, 100015. [CrossRef]
36. Rahamathulla, M.Y.; Ramaiah, M. Optimizing anomaly detection models for edge IIoT with an enhanced firefly algorithm-based hyperparameter tuning strategy. *Results Eng.* **2025**, *27*, 105843. [CrossRef]
37. Abdullah, M.; Mengash, H.A.; Maray, M.; Alrslani, F.A.; Alkhudhayr, H.; Alghanmi, N.A.; Subahi, A.; Majdoubi, J. Federated learning with Blockchain on Denial-of-Service attacks detection and classification of edge IIoT networks using Deep Transfer Learning model. *Comput. Electr. Eng.* **2025**, *124*, 110319. [CrossRef]

38. Andrade, I.; Mahadik, S.S.; Pawar, P.M.; Muthalagu, R. Intelligent intrusion detection using ml for large-scale iot networks. In *Proceedings of the 2024 Advances in Science and Engineering Technology International Conferences (ASET)*; IEEE: New York, NY, USA, 2024; pp. 1–7.
39. Vitorino, J.; Pinto, D.; Maia, E.; Amorim, I.; Praça, I. Revisiting Network Traffic Analysis: Compatible Network Flows for ML Models. In Proceedings of the International Symposium on Foundations and Practice of Security, Brest, France, 25–27 November 2025; pp. 150–166.
40. Laskar, Y.B.; Matam, R.; Barbhuiya, F.A. SHAP-Driven Intrusion Detection: Detecting Mirai Botnet Attacks in IoT. In *Proceedings of the 2025 IEEE Guwahati Subsection Conference (GCON)*; IEEE: New York, NY, USA, 2025; pp. 1–6.
41. Mittal, M.; Kumar, K.; Behal, S. DDoS-AT-2022: A distributed denial of service attack dataset for evaluating DDoS defense system. *Proc. Indian Natl. Sci. Acad. USA* **2023**, *89*, 306–324. [CrossRef]
42. Sharafaldin, I.; Lashkari, A.H.; Hakak, S.; Ghorbani, A.A. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *Proceedings of the 2019 International Carnahan Conference on Security Technology (ICCST)*; IEEE: New York, NY, USA, 2019; pp. 1–8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.