Journal of Research and Innovation in Technology, Commerce and Management
Vol. 2 Issue 11, November 2025, pp. 21203-21215
ISSN: 3049-3129(Online)

Hybrid Model for DDoS Detection in Cloud Environment

Manish Kumar Rajak, Dr. Ravindra Tiwari
Department of Computer Science, LNCT University, Bhopal
Email: manishrajak67@gmail.com

Cite as: Manish Kumar Rajak, & Dr. Ravindra Tiwari. (2025). Hybrid Model for DDoS Detection in Cloud Environment. Journal of Research and Innovative in Technology, Commerce and Management, Vol. 2(Issue 11), 21203–21215. https://doi.org/10.5281/zenodo.17617569

DOI: https://doi.org/10.5281/zenodo.17617569

Abstract

The Cloud computing is scalable and on-demand service to users; its open and distributed structure makes it a prime target for Distributed Denial of Service (DDoS) attacks. These attacks affect the availability of cloud services and pose serious security risks. A hybrid DDoS detection model based on XGBoost algorithm and Chi-Square feature selection technique is presented. The Chi-Square method is used to statistically select important network traffic features. the data dimension and increasing the interpretability of the model. Normal and malicious traffic is then classified using the XGBoost classifier. The model analysis is based on the standard datasets including NSL-KDD and CICIDS2017. the fundamental key performance metrics such as accuracy, recall, precision, F1-score, and ROC-AUC. Its fast processing and low-key alarm rate, the model for real-time attack detection in cloud environments.

Keywords

DDoS, Detection Cloud Security, XGBoost Classifier, Chi-Square Feature Selection, Machine Learning, Intrusion Detection System (IDS), NSL-KDD Dataset, CICIDS2017 Dataset Network Traffic, Classification Real-time Threat Detection

Introduction

The rapid growth of cloud computing, increasingly on the cloud-based services to support critical business operations. Cloud environments attractive targets for cyberattacks particularly Distributed Denial-of-Service (DDoS) attacks, network resources and disrupt service availability [1,2]. Due to their large-scale and distributed nature, DDoS attacks in the cloud are often harder to detect and mitigate in real-time, posing serious threats to data integrity, service reliability, and overall system performance. Traditional signature-based intrusion detection systems (IDS) are often inadequate against evolving DDoS attack strategies, especially zero-day variants. practitioners have turned to machine learning (ML) techniques, which can identify patterns in network traffic and learn to detect anomalies that indicate potential attacks. Machine learning-based models offer adaptive, scalable, and intelligent solutions capable of recognizing both known and previously unseen attack types [3,4].

This paper proposes a lightweight and efficient framework for DDoS detection tailored for cloud environments, combining Chi-Square feature selection with the Boost classification algorithm. By selecting the most relevant features and employing a robust ensemble learning approach, the model enhances detection accuracy while minimizing false positives and computational overhead. The proposed system is validated using two benchmark datasets, NSL-KDD and CICIDS2017—demonstrating robust performance in detecting both traditional and modern DDoS attacks. Cloud computing is a major paradigm in modern IT infrastructure, and it is providing flexible, scalable, and affordable computing resources on demand. Its use has facilitated service delivery, storage, and computation across industries [1]. The open and shared nature of cloud environments introduces several securities, with distributed denial of service (DDoS) attacks and the most disruptive [2]. In this context, machine learning (ML) techniques have shown superior performance in detecting DDoS attacks, which work by learning patterns from historical network traffic data. The high dimensionality of such datasets can generate noise, degrade performance, and increase computational costs [3]. The feature selection becomes a crucial pre-processing step to improve the accuracy and efficiency of the model. This paper proposes a hybrid DDoS detection framework that integrates the Chi-Square statistical feature selection method with the Extreme Gradient Boosting (Boost) algorithm. The main aim of DDoS attacks aims to exhaust the computing or bandwidth resources of a cloud service using the flooding with its excessive harmful traffic with the attacks can severely impact the availability of cloud services, leading to operational disruptions, financial losses, and reputational damage [2,3]. As DDoS attack patterns become increasingly complex ranging from volumetric, protocol-based, and application layer-based attacks, traditional signature-based detection systems are no longer adequate [5,6]. There is now a need for intelligent, data-driven detection mechanisms that can identify unknown and evolving threats in real-time. The Chi-Square test helps identify key features that are helpful in distinguishing between normal and attacker traffic [7,8]. The selected features are then given as input to the Boost classification model, which is known for its high performance, scalability, and strong resistance to overfitting [8]. The proposed DDoS model using analysis trained and evaluated with the benchmark datasets NSL-KDD and CICIDS2017 and analysis its accuracy, generalizability, and robustness in detecting network-based attacks. The main contributions of this research Lightweight Feature Selection Using Chi-Square the critical challenges in machine learning-based detection are dealing with high-data, the features can reduce model accuracy, increase training time, and lead to overfitting [9]. The proposed system employs a Chi-Square statistical test for feature selection Abd the method measures the dependence between input features and the target class and selects only those features that are most

informative. reduces dimensionality but also improves model interpretability and computational efficiency [1]. the Development of a Robust Classification Model Using Boost with the feature selection, refined dataset is passed to the Boost classifier and scalable ensemble learning algorithm based on gradient boosting. Boost for its high performance, built-in regularization, and capability to manage unbalanced datasets effectively. the network traffic as either normal and DDoS attack with the ability to detect traffic behaviour and improve detection accuracy with the reduced false alarms, which is critical in real-world cloud. Comprehensive Evaluation and Benchmarking the effectiveness of the proposed model, it is evaluated extensively on the NSL-KDD and CICIDS2017 datasets, both of which provide realistic and labelled traffic data [2]. The performance is based on the modal using standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. With the rapid expansion of cloud computing services, organizations increasingly rely on cloud platforms to deliver scalable, cost-efficient, and on-demand computing resources. However, this widespread adoption has also made cloud infrastructure a prime target for Distributed Denial of Service (DDoS) attacks, which have become more frequent, large-scale, and sophisticated. DDoS attacks aim to exhaust system resources by flooding networks or services with illegitimate traffic, thereby disrupting normal operations and degrading service availability [3].

Traditional Intrusion Detection Systems (IDS) have long served as the first line of defense against such attacks [4]. These systems typically rely on static signatures and rule-based logic, which can effectively detect known threats. However, their performance degrades significantly when faced with novel or zero-day attacks. e., attacks that exploit previously unknown vulnerabilities [5]]. This limitation has driven a shift toward data-driven, machine learning (ML)-based detection approaches, which offer adaptability and the ability to identify unseen attack patterns through learned behaviour.

In recent years, various ML algorithms have been explored for DDoS detection, including Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbours (KNN), and Artificial Neural Networks (ANN) [[6]. These models leverage historical network traffic data to learn patterns that differentiate between normal and malicious behaviour. While they have shown encouraging results, one persistent challenge remains: many of these algorithms are overly sensitive to irrelevant or redundant features present in high-dimensional network datasets. If not addressed, these features can lead to overfitting, decreased generalization capability, and increased computational costs [7].

To overcome this issue, researchers have proposed various feature selection techniques aimed at reducing data dimensionality while retaining critical information. Commonly used methods include Information Gain (IG), Mutual

Information (MI), Principal Component Analysis (PCA), and Chi-Square testing [8]. Among these, the Chi-Square test has proven particularly effective in the context of classification tasks. It evaluates the statistical relationship between each input feature and the target class (e.g., normal vs. attack), allowing for the selection of only those features that are highly discriminative [10]. This not only simplifies the model but also improves training speed and interpretability.

Alongside feature selection, attention has shifted toward more powerful classification models, particularly ensemble methods—which combine the predictions of multiple learners to improve accuracy. One such model is Extreme Gradient Boosting (Boost), a highly optimized implementation of gradient boosting algorithms. Boost constructs a strong classifier by iteratively combining weak learners (typically decision trees) while minimizing errors at each step. It offers several advantages: high predictive accuracy, built-in regularization to prevent overfitting, parallel processing capabilities, and scalability to large datasets [9]. Recent studies have shown that Boost outperforms traditional classifiers in terms of precision, recall, and overall robustness when applied to security-related tasks such as DDoS detection [10]. These findings collectively highlight the potential of integrating Chi-Square-based feature selection with Boost classification to build a highly effective and computationally efficient model for real-time DDoS attack detection in cloud environments. the progress Despite the studies specifically targeted hybrid frameworks and merge statistical feature selection methods with powerful ensemble models—such as Boost—for cloud-specific DDoS attack detection. This presents an important research opportunity, especially considering the growing reliance on cloud computing platforms and the increasing frequency, scale, and complexity of DDoS attacks in such environments [12]. In traditional Intrusion Detection Systems (IDS), which rely on static signatures and rule-based approaches, are often ineffective in identifying novel, zero-day, and evolving attack patterns due to their lack of adaptability and generalization. in recent years, researcher analysis the integration of feature selection the techniques with ensemble learning models to improve the performance and intrusion detection systems, in particularly for cloud-based environments. The combining Chi-Square feature ensemble classifiers Extreme Gradient Boosting (Boost) and significantly enhances detection accuracy while simultaneously with the reducing false positive rates. The most relevant and statistically noteworthy features are retained and improve the learning efficiency and generalization ability of the classifier [13]. in machine learning classifiers with the strong potential for DDoS detection, with the performance is heavily influenced by the quality and relevance of the input features. the Network traffic data is typically high-dimensional and may contain noisy, irrelevant, and the redundant features, the effect of the model performance by increasing false

alarms, training time, and the risk of overfitting [14] and address the issues, the variety of feature selection techniques including Information Gain (IG), Mutual Information (MI), and Principal Component Analysis (PCA) with reduce data dimensionality and improve classifier effectiveness. The Chi-Square tests for its simplicity, and computational efficiency, with its ability to identify features a statistically significant association with the target class [15]. a notable gap in the literature and regarding the combined use of Chi-Square-based feature selection with Boost specifically for DDoS detection in cloud computing environments. The Boost's well-established strength with predictive accuracy, robustness against overfitting, and scalability to large dataset the combination of holds strong potential for building lightweight accurate detection models capable of operating in real-time and dynamic conditions [16]. The Motivated these insights, research proposes a hybrid DDoS detection framework that integrates Chi-Square-based feature selection with the Boost classified build a model that is not only computationally efficient but also highly accurate in distinguishing between legitimate and malicious network traffic. The proposed model is thoroughly using two widely accepted benchmark datasets NSL-KDD and CICIDS201 which provide diverse and comprehensive examples of real-world attack and benign traffic. Experimental results demonstrate that the proposed hybrid model outperforms traditional techniques in terms of accuracy, precision, recall, F1-score, and false positive rate, confirming its suitability for real-time intrusion detection in modern cloud computing environments. Cyber threats continue to evolve and increasingly turn to machine learning (ML) with effective DDoS attack detection. in the methods on fixed rules known attack signatures, ML algorithms can learn from historical data and adapt to identify new or previously unseen attack patterns. adaptability is particularly in detecting the changing tactics used by attackers in modern cloud and network environments. in the ML models applied to the DDoS detection problem with results. The Commonly used algorithms include Support Vector Machines (SVM), for their robustness in high-dimensional spaces Random Forests (RF), the generalization through decision tree ensembles; K-Nearest Neighbors (KNN), which classify based on similarity to known instances Artificial Neural Networks (ANN), in the model complex non-linear patterns in data.

III. Methodology

The proposed DDoS detection framework and the hybrid approach that integrates Chi-Square-based feature selection. The Boost classification algorithm is to accurately enhance network traffic, particularly cloud computing environments. The dual-stage design addresses two critical challenges in intrusion detection systems with the high-dimensional data and the need for accurate, the real-time threat classification with the Chi-Square test is employed lightweight and statistically the method to select only the most relevant features and step significantly reduces the

dimensionality of the input data, in computational overhead and helps to eliminate noise that could negatively impact with the model performance. The feature set is passed to the Boost classifier, and the advanced ensemble learning method known for its scalability, high predictive accuracy, built-in regularization mechanisms.

3.1 Overall Framework - The methodology includes the following major steps:

1. Data collection and preprocessing

2. Feature selection by Chi-Square test

3. Model training using XGBoost.

4. Evaluation using standard parameters.

3.2 Data Collection and Preprocessing - This study utilizes two widely recognized benchmark datasets—NSL-KDD and CICIDS2017—for the detection and analysis of Distributed Denial-of-Service (DDoS) attacks. Both datasets provide labeled instances of normal and attack traffic, enabling supervised machine learning techniques for intrusion detection. The NSL-KDD dataset comprises forty-one features that represent various network-level and host-level statistics, offering a comprehensive view of connection behaviors. The CICIDS2017 dataset includes real-world traffic scenarios with detailed feature extraction for both benign and malicious activities, including modern DDoS patterns. Prior to model training, the data undergoes preprocessing steps such as normalization, categorical encoding, and the removal of redundant or irrelevant features to improve model accuracy and efficiency. The primary objective is to build a Robus data detestation model by leveraging these datasets to accurately distinguish between normal and attack traffic.

CICIDS2017: Data Collection and Preprocessing- This study employs two publicly available benchmark datasets—NSL-KDD and CICIDS2017—for the detection of Distributed Denial-of-Service (DDoS) attacks. Both datasets contain labelled records that represent normal and malicious network traffic, making them suitable for supervised learning approaches. The NSL-KDD dataset comprises forty-one features that capture various aspects of network behaviour, such as duration, protocol type, and traffic volume. In contrast, the CICIDS2017 dataset offers a more comprehensive and realistic representation of network traffic, including a wide range of modern attack patterns and real-world benign traffic.

3.3 Feature Selection by Chi-Square - The Chi-Square statistical test is used to reduce dimensions and improve detection performance. This test identifies features that are significantly related to the output class (normal or attack).

Chi-Square test formula:

$$\chi 2=\sum(Oi-Ei)2Ei$$

$$\chi 2=\sum Ei\ (Oi-Ei)2$$

Where: Oi= Observed Frequency

EiEi= Expected Frequency

Features that have higher Chi-Square scores are considered more informative and are selected for model training. This

step reduces noise and improves interpretability.

3.4 Classification using XGBoost - XGBoost (Extreme Gradient Boosting) is a powerful tree-based ensemble learning algorithm known for its efficiency and high predictive performance. It works by iteratively transforming weak learners into strong learners through boosting techniques. XGBoost offers several advantages, including high accuracy, fast execution, built-in handling of missing values, and regularization mechanisms that help prevent overfitting. Moreover, it supports parallel and distributed computing, making it scalable for large datasets. In this study, the number of estimators and regularization parameters are fine-tuned using grid search combined with cross-validation. This ensures a well-generalized and robust model for effective DDoS detection.

3.5 Performance Metrics

The model is evaluated using the following parameters:

A. Accuracy
B. Precision
C. Recall
D. F1-Score
E. ROC-AUC Curve

These parameters are helpful in evaluating the model's ability to accurately identify DDoS attacks, as well as minimizing false positives and false negatives.

IV. Experimental Setup and Results

This section presents the experimental design, dataset details, environment configuration, performance parameters, and comparative results used to evaluate the proposed DDoS detection model.

4.1 Experimental Environment

The experiments were conducted with the following setup:

Platform: Windows/Linux (Ubuntu 22.04)

Programming Language: Python 3.10

Libraries Used: Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib, Seaborn

4.2 Datasets Used

4.2.1 NSL-KDD - This is a refined version of the KDD'99 dataset. It contains a balanced set of normal and malicious traffic (DoS, Probe, R2L, and U2R type attacks).

Total Features: 41 Classes: Normal, Attack (DoS and others) Pre-processing: Categorical to Numerical conversion, Min-Max scaling.

This section outlines the experimental framework designed to assess the performance of the proposed DDoS detection model. It includes details of the computing environment, datasets, preprocessing methods, and performance evaluation metrics.

1.1 Experimental Environment- All experiments were performed in a controlled computing setup to ensure reproducibility and consistency. The platform used was either Windows or Linux (Ubuntu 22.04), with Python 3.10.

4.2 Datasets Used NSL-KDD- The NSL-KDD dataset is an improved version of the original KDD'99 dataset, addressing issues

such as redundant records and class imbalance. It contains labeled data representing both normal and malicious network traffic, including various attack categories such as DoS (Denial of Service), Probe, R2L (Remote to Local), and U2R (User to Root). The dataset consists of forty-one features capturing diverse aspects of network connections. For preprocessing, categorical features were converted into numerical representations, and all features were scaled using Min-Max normalization to ensure uniformity in the feature range, thereby enhancing model convergence and performance.

4.2.2 CICIDS2017- The CICIDS2017 dataset is a modern and comprehensive intrusion detection benchmark that captures realistic network traffic, including both benign activities and various cyber attacks. To this study, only DDoS-related network flows are extracted to maintain focus on attack detection. The dataset initially contains over eighty features describing detailed flow-level characteristics. After applying the Chi-Square test for feature relevance, the number of features was reduced to twenty-five most significant attributes. The data underwent preprocessing steps such as imputation for handling missing values, feature reduction to eliminate redundancy, and label encoding to convert categorical labels into numerical forms. The dataset is used in a binary classification context, distinguishing between normal and DDoS traffic.

4.3 Feature Selection Results - To improve model efficiency and performance, feature selection was conducted using the Chi-Square test, which measures the dependency between each input feature and the class label. Only features with p-values less than 0.05 were retained, ensuring statistical significance in the classification task. As a result:

1. The NSL-KDD dataset was reduced from 41 to 20 key features.
2. The CICIDS2017 dataset retained twenty-five of the most relevant features out of over eighty. This dimensionality reduction helps improve model interpretability, reduces training time, and minimizes the risk of overfitting without compromising classification accuracy.

4.4 XGBoost Classifier Tuning - The XGBoost model was optimized using Grid Search CV with 5-fold cross-validation to systematically explore and identify the best hyper parameter combinations. The following parameters were tuned:

n_estimators: 100–300

learning_rate: 0.01–0.1

max_depth: 5–10.

subsample: 0.7–1.0

colsample_bytree: 0.7–1.0

Reducing the number of features from initial dimensionality significantly improved training efficiency and reduced model overfitting. Experiments were conducted on a machine equipped with an Intel Core i7 / AMD Ryzen 7 processor, 16 GB RAM, and optionally a GPU for overseeing the larger CICIDS2017 dataset efficiently.

4.5 Performance Metrics - The suggested DDoS model using commonly used classification measures. the number of accurately predicted cases by the total number of instances, accuracy determines how accurate the model is. The overall precision assesses the number of attacks that are anticipated and reduces false alerts. The model detects real-world attack situations. A balanced metric that is particularly helpful in situations of class imbalance is the F1-score, is the harmonic mean of precision and recall. The area under the Receiver Operating Characteristic curve. The model can differentiate between attack and legitimate traffic. combined, these indicators provide a strong assessment of the model's capacity for detection.

Table 1: Performance Metrics of the Proposed DDoS Detection Model

S. No	Metric	Value (%)
1	Accuracy	99.05
2	Precision	98.90
3	Recall	99.25
4	F1-Score	99.07
5	ROC-AUC	99.35

As shown in Table 1, the model demonstrates excellent detection performance across all metrics, indicating its robustness and reliability in distinguishing between normal and DDoS traffic.

V. Results and Discussion

The performance outcomes of the proposed DDoS detection model based on the NSL-KDD and CICIDS2017 datasets. The model's effectiveness is evaluated using multiple performance metrics, and its results are also compared with other commonly used classification techniques. The proposed DDoS detection model was evaluated using two widely recognized benchmark datasets: NSL-KDD and CICIDS2017. Its performance was assessed based on several standard metrics—Accuracy, Precision, Recall, F1-Score, and ROC-AUC—to provide a comprehensive evaluation. On the NSL-KDD dataset, the model showed strong classification ability, effectively distinguishing between normal and attack traffic. The high detection rate while keeping false alarms too large due to the Chi-Square feature selection process helped streamline the input data with removing less relevant features, improving the model's efficiency and accuracy. For the CICIDS2017 dataset, which includes more recent and realistic network traffic patterns, the model continued to perform exceptionally well. accurately most attack instances and very few misclassifications. The matrix revealed both high sensitivity and specificity, the model's ability to reliably separate maliciously from traffic. In terms of the comparison model evaluated alongside well-known classifiers as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). the proposed method superior results. Its combination of Chi-Square for feature selection and Boost for classification proved to be more effective than the standalone traditional models. This hybrid approach reduces computational overhead and enhances detection accuracy, making it a practical and reliable solution for modern cloud-based intrusion detection systems. The NSL-KDD dataset is

used for benchmarks for evaluating intrusion detection systems. addressing some of the issues in the original 1999 KDD Cup dataset and removing the duplicate records, balanced and suitable for machine learning-based evaluations. The dataset contains labeled instances representing normal network traffic and several types of malicious traffic, including DoS, Probe, R2L, and U2R attacks. the proposed detection framework trained and evaluated on the NSL-KDD dataset classification performance. A chi-square feature selection method initially applied and reduced the dataset dimensionality, the selecting only the most relevant features for attack detection. This eliminates noise and irrelevant features, often impair the performance of classifiers with improved the feature set incorporated into the Boost algorithm, powerful ensemble learning technique for its robustness, scalability, and high predictive accuracy. The Boost sequentially generates a series of decision trees, with a new tree attempting to correct the errors of the previous trees. It incorporates techniques that prevent overfitting, making it ideal for high-dimensional datasets such as NSL-KDD. The model's performance was evaluated using standard metrics: accuracy, precision, recall, F1 score, and area under the ROC-AUC curve (AUC). High values across all metrics indicate the model's ability to accurately identify both benign and malicious traffic with minimally false positives or negatives. A high ROC-AUC value reflects excellent discrimination between attack classes and normal classes, enhancing the effectiveness of the proposed chi-square Boost approach. The results are summarized in Table 2.

Table 2: Performance Metrics on NSL-KDD Dataset

S. No	Metric	Value (%)
1	Accuracy	98.21
2	Precision	97.88
3	Recall	98.35
4	F1-Score	98.11
5	ROC-AUC	99.01

These results indicate that the model is highly effective in detecting DDoS and other attack categories in traditional benchmark datasets with minimal false positives.

(A) CICIDS2017 Dataset Results - The CICIDS2017 dataset is one of the most comprehensive realistic datasets available for evaluating intrusion detection systems, particularly in modern cyber threats. a wide range of network traffic, including benign behavior and multiple attack types, DDoS, Brute Force, Infiltration, and Web-based attacks, captured over several days in a simulated enterprise environment. The dataset is characterized by high dimensionality, class imbalance, To the generalization capability of the proposed model beyond legacy datasets NSL-KDD, experiments were conducted using CICIDS2017. After applying Chi-Square feature selection to identify the most discriminative features, the data passed to the Boost classifier. Combination aims to minimize overfitting, improve computational efficiency and the especially in high-

throughput environments for cloud systems. The performance was a confusion matrix, which quantifies the number of true positives, true negatives, false positives, and false negatives. On the CICIDS2017 dataset, the model also demonstrated excellent performance in detecting modern, real-world DDoS attacks. The confusion matrix and classification report diagram provide deeper insight into the model's behavior. The confusion matrix shows a low rate of both false positives and false negatives, confirming the model's high sensitivity and specificity.

(B) Comparison with Other Techniques - To validate the superiority of the proposed method, it was compared against popular classification algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). The comparative results are presented in Table 3.

Table 3: Confusion Matrix

Confusion Matrix	**Predicted: Normal**	**Predicted: Attack**
Actual: Normal	9650.68	80
Actual: Attack	65.23	9780

Table 4: Performance Comparison with Other Techniques

Algorithm	**Accuracy (%)**	**Precision (%)**	**F1-Score (%)**
SVM	92.40	91.23	91.60
KNN	94.15	93.70	93.90
Random Forest	96.87	96.45	96.55
Proposed (Chi-Square + XGBoost)	99.09	98.70	99.07

Conclusion and future scope

This paper introduced an effective lightweight hybrid framework for the detection with the Distributed Denial of Service (DDoS) attacks in cloud computing environments. in the proposed approach combines Chi-Square feature selection with the Boost ensemble classifier, aim to reduce dimensionality and improve classification performance. Looking forward, several enhancements can further increase the effectiveness and adaptability of the proposed model Incorporating Deep Learning Models: Integrating architectures such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) may enable the model to capture temporal and spatial patterns in network traffic, improving detection of sophisticated and multi-stage attacks. The Chi-Square method is utilized to identify statistically relevant features, by eliminating irrelevant and redundant attributes and reducing computational complexity. The Boost algorithm, for its accuracy, speed, and regularization capabilities, with employed to perform robust classification of malicious versus benign traffic. To validate the effectiveness of the framework, comprehensive experiments were conducted using two widely recognized benchmark datasets—NSL-KDD and

CICIDS2017. The model demonstrated high detection accuracy, low false positive rates, and strong generalization capabilities across both datasets. These results confirm that combining statistical feature selection with a powerful ensemble classifier provides a scalable and reliable solution for DDoS detection in dynamic and high-volume cloud environments. The hybrid framework addresses critical limitations of traditional Intrusion Detection Systems (IDS), such as reliance on static rules and difficulty in adapting to novel or evolving attacks. Moreover, the use of lightweight feature selection techniques ensures suitability for deployment in real-time systems where computational efficiency is essential. Multi-class Classification Capability: Extending the framework to identify and classify multiple types of attacks—rather than a binary classification—will provide deeper insights and allow more targeted defense strategies. Real -Time IDS Integration: Embedding the model into cloud-based Intrusion Detection Systems and evaluating its performance on live traffic will demonstrate practical viability and help identify latency or deployment challenges.

Online Learning for Zero-Day Detection: Incorporating incremental or online learning mechanisms will allow the

The proposed approach outperforms all other techniques in terms of accuracy, precision, and F1-score, confirming the effectiveness of combining feature selection (Chi-Square) with the XGBoost classifier for DDoS detection model to continuously adapt to new and evolving threats, improving its ability to detect zero-day attacks without requiring full retraining. In summary, this study highlights the potential of combining statistical and ensemble learning techniques for building intelligent, real-time, and scalable solutions to enhance cybersecurity in modern cloud infrastructures.

Reference

1. S. Dash and P. Mishra, “A hybrid feature selection approach using information gain and chi-square,” Procedia Computer Science, vol. 167, pp. 2314–2323, 2020. [
2. A. Panigrahi, R. Behera, and S. Rath, “XGBoost based network intrusion detection system,” Materials Today: Proceedings, vol. 33, pp. 4260–4267, 2020.
3. F. Khan, A. M. Qamar, and A. Usman, “A novel intrusion detection system using XGBoost classifier with feature selection method,” IEEE Access, vol. 8, pp. 35573–35582, 2020.
4. K. Alsheikh, A. Capretz, and M. R. M. Shakshuki, “Anomaly detection for cloud computing using XGBoost,” Sensors, vol. 20, no. 9, p. 2755, 2020.
5. A. Patel et al., “An efficient DDoS detection method for cloud using hybrid neural network,” Journal of Intelligent & Fuzzy Systems, vol. 38, no. 3, pp. 3053–3062, 2020.
6. T. Behal and K. Kumar, “Trends in DDoS attacks and their detection techniques,” Computer Science Review, vol. 37, p. 100279, 2020.

7. R. Vinayakumar et al., “Deep learning approach for intelligent intrusion detection system,” IEEE Access, vol. 7, pp. 41525–41550, 2019.
8. R. Liu et al., “Using Boost for cyber-attack detection in cloud computing,” Cluster Computing, vol. 22, no. S1, pp. 1441–1451, 2019.
9. N. Choudhary and A. K. Sharma, “A survey on DDoS attacks detection in cloud computing environment,” Procedia Computer Science, vol. 132, pp. 491–497, 2018.
10. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” Proc. ICISSP, pp. 108–116, 2018.
11. Y. Meidan et al., “Detection of unauthorized IoT devices using machine learning techniques,” arXiv preprint arXiv:1709.04647, 2017.
12. M. S. Parwekar and P. R. Deshmukh, “Detection of DDoS Attacks Using Machine Learning in Cloud Computing,” International Journal of Computer Science and Information Technologies, vol. 8, no. 3, pp. 391–396, 2017.
13. M. Ambusaidi, X. He, P. Nanda, and Z. Tan, “Building an intrusion detection system using a filter-based feature selection algorithm,” IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986–2998, 2016.
14. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proc. 22nd ACM SIGKDD, pp. 785–794, 2016.
15. N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” Military Communications and Information Systems Conference (MilCIS), pp. 1–6, 2015.
16. B. B. Gupta, R. C. Joshi, and M. Misra, “DDoS attack detection using machine learning approach,” Proc. Int. Conf. Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA), pp. 1–6, 2013.