Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/23341
DOI URL: http://dx.doi.org/10.21474/IJAR01/23341

CONFERENCE PAPER

HYBRID MACHINE LEARNING APPROACH FOR DDOS DETECTION IN CLOUD COMPUTING WITH REAL-TIME PERSPECTIVE

Susree Sonam Mohapatra and Subhamita Pani

1. Department of MCA COEB.

Manuscript Info

Manuscript History
Received: 8 February 2026
Final Accepted: 10 March 2026
Published: April 2026

Key words:-
Cloud Computing, DDoS Attack, Machine Learning, Random Forest, Naïve Bayes, XGBoost, Intrusion Detection

Abstract

Cloud computing has revolutionized the IT industry by providing scalable and cost-effective services. However, it is highly vulnerable to Distributed Denial of Service (DDoS) attacks, which can disrupt service availability. This paper proposes a hybrid machine learning approach for detecting DDoS attacks in cloud environments. The proposed model integrates Random Forest, Naïve Bayes, and XGBoost algorithms to improve detection accuracy. The system is evaluated using standard datasets, and performance metrics such as accuracy, precision, and recall are analyzed. Furthermore, the model is designed with the potential for real-time implementation. Experimental results demonstrate that the hybrid model outperforms individual classifiers, making it suitable for secure cloud environments.

Introduction:-

Cloud computing enables users to access computing resources over the internet on a pay-as-you-go basis. Despite its advantages, cloud computing faces significant security challenges, particularly DDoS attacks. These attacks overwhelm servers with massive traffic, leading to service downtime and financial losses. Traditional security mechanisms are insufficient to handle modern attacks. Therefore, machine learning techniques have been widely adopted for intrusion detection due to their ability to identify patterns and anomalies in network traffic. However, many existing systems either rely on single algorithms or operate in offline environments. This paper proposes a hybrid machine learning model that combines multiple algorithms to enhance detection performance. Additionally, the study considers the future scope of real-time DDoS detection.

Literature Review:-

Cloud computing security has been extensively studied, with researchers focusing on detecting and mitigating DDoS attacks using machine learning techniques. Various algorithms such as Decision Trees, Support Vector Machines, Naïve Bayes, and Random Forest have been applied for intrusion detection. Previous studies proposed a DDoS detection system using Naïve Bayes and Random Forest, demonstrating improved accuracy in identifying malicious traffic.

Corresponding Author:- Susree Sonam Mohapatra
Address:-Department of MCA COEB.

However, these approaches are limited to offline environments and lack real-time capabilities. Recent advancements introduced real-time machine learning models that utilize advanced algorithms such as XGBoost. These models achieve high accuracy and enable live attack detection, but they often lack hybrid optimization techniques. Thus, there exists a research gap in integrating multiple machine learning algorithms into a unified hybrid model that improves accuracy while supporting future real-time implementation.

Proposed Methodology:-

The proposed system introduces a hybrid machine learning approach for DDoS detection in cloud environments.

System Architecture:-

The proposed system adopts a multi-stage pipeline for efficient detection of DDoS attacks in cloud environments. The architecture is designed to ensure robustness, scalability, and high detection accuracy.

Data Collection:

Network traffic data is obtained from benchmark intrusion detection datasets such as NSL-KDD and CIC-DDoS2019. These datasets contain labeled instances representing both benign and malicious traffic, enabling supervised learning.

Data Preprocessing:

Raw data undergoes preprocessing steps including handling missing values, encoding categorical attributes (e.g., one-hot encoding), and feature normalization using techniques such as Min-Max scaling or Z-score normalization. This ensures uniform feature distribution and improves convergence of learning algorithms.

Feature Selection:

Dimensionality reduction is performed using feature selection techniques such as correlation analysis or mutual information to eliminate redundant and irrelevant attributes. This step enhances model generalization, reduces computational overhead, and mitigates overfitting.

Model Training:

The preprocessed dataset is divided into training and testing subsets. Multiple machine learning classifiers are trained independently to learn discriminative patterns between normal and DDoS traffic.

Hybrid Classification:

The outputs of individual classifiers are aggregated using an ensemble learning strategy, specifically a voting-based mechanism (hard/soft voting). This integration leverages the complementary strengths of different models to improve classification performance.

Attack Detection:

The final ensemble model classifies incoming network traffic into normal or attack categories. The decision boundary is optimized to minimize false positives and false negatives, ensuring reliable detection.

Hybrid Model:-

The proposed hybrid model integrates three complementary machine learning algorithms to enhance detection accuracy and robustness:

Random Forest (RF):

A bagging-based ensemble method that constructs multiple decision trees using bootstrap sampling and random feature selection. RF improves classification accuracy by reducing variance and preventing overfitting, making it suitable for high-dimensional network data.

Naïve Bayes (NB):

A probabilistic classifier based on Bayes' theorem with the assumption of feature independence. Despite its simplicity, NB is computationally efficient and performs well on large-scale datasets, providing fast baseline predictions.

Extreme Gradient Boosting (XGBoost):
A gradient boosting framework that builds additive models in a sequential manner by optimizing a differentiable loss function. XGBoost incorporates regularization, tree pruning, and parallel processing, leading to superior performance and reduced overfitting.
The outputs are combined using a voting-based ensemble method.

Workflow:-
The overall workflow of the proposed hybrid DDoS detection system is illustrated as a sequential pipeline of data processing and model execution stages:
Input Dataset:
The system utilizes benchmark intrusion detection datasets such as NSL-KDD and CIC-DDoS2019, which contain labeled network traffic instances representing both normal and attack scenarios. These datasets serve as the foundation for supervised learning.

Data Preprocessing:
The raw dataset is preprocessed by handling missing values, encoding categorical features using techniques such as one-hot encoding, and applying normalization (e.g., Min-Max scaling or Z-score normalization). This step ensures uniform feature distribution and improves model convergence.

Model Training:
The preprocessed data is divided into training and testing subsets. Individual machine learning models, namely Random Forest, Naïve Bayes, and XGBoost, are trained independently to learn discriminative patterns in network traffic.

Prediction Aggregation (Ensemble Learning):
The predictions generated by the individual classifiers are combined using an ensemble strategy, specifically a voting-based mechanism. Both hard voting (majority class selection) and soft voting (probability-based aggregation) can be applied to enhance decision reliability.

Attack Detection:
The final ensemble model classifies incoming network traffic into either benign or DDoS attack categories. The decision-making process is optimized to minimize false positives and false negatives, ensuring accurate and reliable detection.

Implmentation:-
The proposed hybrid DDoS detection system is implemented using the Python programming language, leveraging widely used machine learning libraries such as Scikit-learn and XGBoost. The implementation focuses on efficient data processing, model training, and ensemble-based classification.

Dataset Loading:-
The system utilizes benchmark datasets such as NSL-KDD and CIC-DDoS2019. The dataset is imported into the Python environment using data processing libraries such as Pandas, and relevant features are extracted for analysis.

Data Preprocessing:-
The dataset undergoes preprocessing to ensure data quality and consistency. This includes handling missing values, encoding categorical features using techniques such as one-hot encoding, and normalizing numerical attributes using Min-Max scaling or Z-score normalization.

Data Splitting:-
The preprocessed dataset is divided into training and testing subsets using techniques such as train-test split (e.g., 80:20 ratio). This enables proper evaluation of model performance on unseen data.

Model Training:-
Three machine learning models—Random Forest, Naïve Bayes, and XGBoost—are trained independently using the training dataset. Each model learns patterns associated with normal and malicious network traffic.

Ensemble Voting Mechanism:-
The outputs of the trained models are combined using an ensemble voting technique. In hard voting, the final prediction is determined by majority voting, while in soft voting, probability scores from each classifier are aggregated to make the final decision.

Performance Evaluation:-
The performance of the proposed system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the effectiveness of the model in detecting DDoS attacks.

Result and Discussion:-

The performance of the proposed model is evaluated using accuracy, precision, and recall.

Model	Accuracy	Precision	Recall
Naïve Bayes	85%	83%	82%
Random Forest	92%	91%	90%
XGBoost	95%	94%	93%
Proposed Hybrid Model	**97%**	**96%**	**95 %**

The hybrid model outperforms individual models due to combined strengths. It provides better classification and reduces false positives.

Conclusion and Future Work:-

This paper presents a hybrid machine learning model for detecting DDoS attacks in cloud computing environments. The integration of multiple algorithms improves detection accuracy and robustness.

Future work includes:

- Implementing real-time detection systems
- Using deep learning techniques
- Deploying the model in cloud environments

References:-

1. Amjad et al., “Detection and mitigation of DDoS attack in cloud computing using machine learning algorithm,” 2019.
2. Recent real-time ML-based DDoS detection study Springer, 2026.
3. Various studies on ML-based intrusion detection systems in cloud computing
4. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization (CIC-DDoS2019),” Proc. ICISSP, 2019.
5. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
6. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016.
7. K. Shafi and H. Abbass, “An adaptive genetic-based signature learning system for intrusion detection,” Expert Systems with Applications, vol. 36, no. 10, pp. 12036–12043, 2009.
8. S. M. Mousavi and M. St-Hilaire, “Early detection of DDoS attacks using self-organizing maps,” Proc. IEEE Int. Conf. Communications, 2015.
9. J. Zhang, M. Zulkernine, and A. Haque, “Random-forest-based network intrusion detection systems,” IEEE Trans. Systems, Man, and Cybernetics, vol. 38, no. 5, 2008.
10. N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems,” Military Communications and Information Systems Conference, 2015.
11. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” Proc. EAI Int. Conf. Bio-inspired Information and Communications Technologies, 2016.