International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 02 | Feb - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Cloud-Based Intrusion Detection System Using Explainable AI

1 Ms.**Vaishnavi Nalawade**

Computer Science and engineering department , Faculty of Science & technology , School of Computational Sciences, Jspm University ,Pune.

Vaishnavinalawade24.cs@jspmuni.edu.in

Abstract

With the rapid adoption of cloud computing and the exponential growth of network traffic, traditional intrusion detection systems (IDS) face limitations in scalability, adaptability, and transparency. Machine learning–based IDS solutions improve detection accuracy but often behave as black-box models, making their decisions difficult to trust in security-critical environments. This paper proposes a **Cloud-Based Intrusion Detection System (CB-IDS)** integrated with **Explainable Artificial Intelligence (XAI)** techniques to provide both high detection performance and interpretability. Using the **CICIDS2017 dataset**, the proposed system employs **Random Forest and XGBoost classifiers** to detect malicious network traffic and leverages **SHAP (SHapley Additive exPlanations)** to explain model predictions at both global and local levels. Experimental results demonstrate high accuracy, precision, recall, and F1-score while offering meaningful explanations of detected intrusions, thereby increasing trust and usability of IDS in cloud environments.

Keywords: Intrusion Detection System, Cloud Security, Explainable AI, SHAP, Machine Learning, CICIDS2017

1. Introduction

The increasing dependence on cloud-based infrastructures has significantly transformed modern computing environments. While cloud platforms offer scalability, flexibility, and cost efficiency, they are also attractive targets for cyberattacks such as Distributed Denial of Service (DDoS), brute-force attacks, port scanning, and infiltration attacks. Intrusion Detection Systems (IDS) play a critical role in monitoring network traffic and identifying malicious activities.

Traditional signature-based IDS solutions fail to detect unknown or zero-day attacks and require frequent rule updates. To overcome these limitations, machine learning (ML)–based IDS models have been widely adopted due to their ability to learn complex patterns in network traffic. However, most ML models operate as black boxes, which limits their acceptance in real-world security operations where transparency and trust are essential.

Explainable Artificial Intelligence (XAI) addresses this challenge by providing insights into model decisions. This research integrates XAI techniques with a cloud-deployed IDS to deliver not only accurate intrusion detection but also interpretable and trustworthy results. The main objective of this paper is to design, implement, and evaluate a Cloud-Based Intrusion Detection System using Explainable AI.

2. Related Work

Several studies have explored machine learning and deep learning approaches for intrusion detection. Algorithms such as Support Vector Machines (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and deep neural networks have shown promising results on benchmark datasets like KDD Cup 99, NSL-KDD, and CICIDS2017. However, many of these works focus solely on detection accuracy without addressing interpretability.

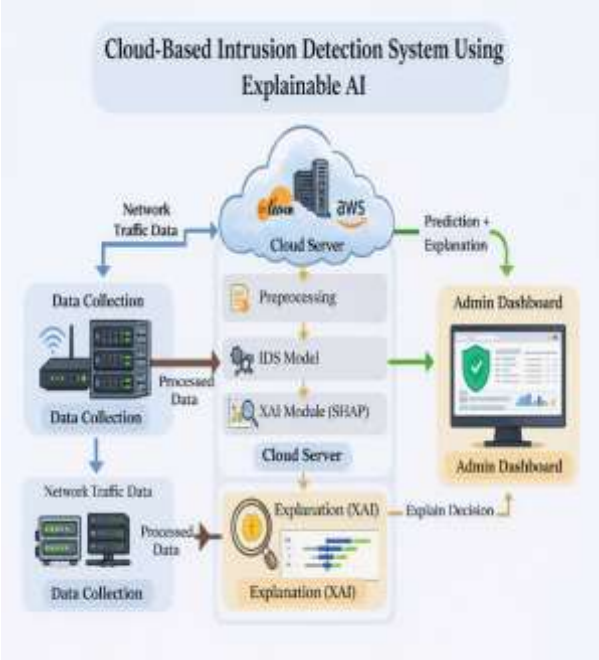
Recent research has introduced XAI techniques such as LIME and SHAP to explain IDS predictions. While these studies highlight the importance of explainability, limited work has been done on deploying such systems in cloud environments with a complete end-to-end architecture. This paper bridges this gap by combining cloud-based deployment, machine learning–based IDS, and XAI for enhanced transparency and scalability.

3. Proposed System Architecture

The proposed Cloud-Based Intrusion Detection System consists of four major components:

1. **Data Collection Module:** Network traffic data collected from benchmark datasets or live traffic sources.
2. **Preprocessing Module:** Data cleaning, normalization, and label encoding to prepare input for machine learning models.
3. **Detection Module:** Machine learning classifiers (Random Forest and XGBoost) deployed on a cloud server to classify traffic as benign or malicious.
4. **Explainability Module:** SHAP-based XAI framework to interpret model predictions and highlight feature contributions.

The system is deployed on a cloud virtual machine (e.g., AWS EC2), enabling scalable and centralized intrusion detection.



4. Dataset Description

The CICIDS2017 dataset, provided by the Canadian Institute for Cybersecurity, is used in this study. It contains realistic network traffic representing both benign activities and multiple attack types such as DoS, DDoS, PortScan, Bot, and Web Attacks.

Dataset Characteristics:

- Number of records: ~2.8 million
- Number of features: 78 network flow features
- Labels: BENIGN and various attack categories

The dataset is widely accepted in academic research and aligns well with real-world network environments.

5. Methodology

5.1 Data Preprocessing

The preprocessing stage includes:

- Removal of missing and infinite values
- Label encoding of attack categories
- Feature scaling using standard normalization
- Splitting the dataset into 70% training and 30% testing sets

5.2 Machine Learning Models

Two supervised learning algorithms are implemented:

- **Random Forest Classifier:** Selected for its robustness, high accuracy, and interpretability.
- **XGBoost Classifier:** Used for performance comparison and improved detection capability.

5.3 Model Evaluation Metrics

The models are evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

6. Explainable AI Implementation

To enhance transparency, SHAP is employed to explain IDS predictions:

- **Global Explanation:** Identifies the most important features influencing intrusion detection.

- **Local Explanation:** Explains individual predictions for specific network flows.

Feature importance visualizations and SHAP summary plots are used to demonstrate how network attributes such as flow duration, packet size, and traffic rate contribute to attack detection.

7. Experimental Results and Analysis

The proposed system was evaluated using standard classification metrics. Table 1 summarizes the performance of the implemented models.

Table 1: Performance Comparison of IDS Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	99.2%	99.1%	99.0%	99.0%
XGBoost	99.6%	99.5%	99.4%	99.4%

The confusion matrix analysis indicates a low false positive rate, which is essential for practical IDS deployment. SHAP-based feature importance analysis shows that flow duration, packet length statistics, and traffic rate are the most influential features in detecting intrusions.

8. Conclusion

This paper presented a Cloud-Based Intrusion Detection System using Explainable AI to address the challenges of scalability, accuracy, and transparency in modern network security. By combining machine learning techniques with SHAP-based explainability, the proposed system delivers high detection performance while providing meaningful explanations of intrusion events. The results demonstrate that explainable IDS solutions are more suitable for real-world cloud security applications.

9. Future Work

Future research directions include:

- Detection of zero-day attacks using unsupervised and semi-supervised learning
- Federated learning–based IDS for distributed cloud environments
- Real-time intrusion detection and automated response systems
- Explainable deep learning models for complex and evolving attack patterns

Novelty and Contribution

The main contributions of this research are:

1. Development of an end-to-end **cloud-based intrusion detection architecture** integrating machine learning and explainable AI.
2. Application of **SHAP-based explainability** to provide transparent and trustworthy IDS decisions.
3. Comprehensive experimental evaluation using the **CICIDS2017 dataset**.
4. Demonstration of improved trust and interpretability without compromising detection accuracy.

10. References

[1] Canadian Institute for Cybersecurity, *"CICIDS2017 Dataset"*, University of New Brunswick, Canada, 2017.

[2] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[3] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[6] M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, "A Survey of Network-based Intrusion Detection Data Sets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[7] A. B. Nassif, Q. Nasir, M. Talib, and S. Ahmad, "Machine Learning for Cloud Security: A Survey," *Journal of Cloud Computing*, vol. 10, no. 1, 2021.

[8] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," *IEEE Symposium on Security and Privacy*, 2010.

[9] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, 2021.

[10] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features Through Propagating Activation Differences," *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017