Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 3, 2025

LEVERAGING ZERO TRUST ARCHITECTURE FOR NETWORK INTRUSION DETECTION: A COMPREHENSIVE EVALUATION USING THE UNSW-NB15 DATASET

Muhammad Zunnurain Hussain[*1], Aliha Iftikhar[2], Talib Nadeem Usmani[3], Muhammad Zulkifl Hasan[4]

[*1]*Assistant Professor, Department of Computer Science, Bahria University Lahore Campus*
[2]*Associate Software Engineer, Bachelor of Science in Computer Science, FAST-NUCES Lahore*
[3]*Honeywell, Duluth, Georgia, USA*
[4]*Faculty of Information Technology, Department of Computer Science, University of Central Punjab*

[*1]zunnurain.bulc@bahria.edu.pk , [2]alihaiftikhar19@gmail.com, [3]talibosmani@gmail.com, [4]zulkifl.hasan@ucp.edu.pk

DOI: https://doi.org/10.5281/zenodo.16963330

Keywords
Zero Trust Architecture, Network Intrusion Detection, UNSW-NB15, Machine Learning, XGBoost, Random Forest, Logistic Regression, Cybersecurity

Article History
Received on 20 January 2025
Accepted on 20 February 2025
Published on 27 February 2025

Corresponding Author: *

Abstract
Zero Trust Architecture (ZTA) has emerged as a critical approach to enhancing cybersecurity by assuming that both internal and external network traffic must be continuously verified. This paper explores the application of ZTA principles in network intrusion detection, specifically evaluating machine learning models on the UNSW-NB15 dataset. We compare the performance of three classifiers—Random Forest (RF), Logistic Regression (LR), and XGBoost—on detecting malicious network traffic. Our results show that XGBoost achieves the highest performance with an Area Under the Curve (AUC) score of 1.00, demonstrating its effectiveness in real-time traffic monitoring. These findings prov

INTRODUCTION

BC The rapid evolution of cybersecurity threats has highlighted the inadequacy of traditional defense strategies, which typically rely on perimeter-based security models. These models are increasingly ineffective against modern attacks such as advanced persistent threats (APTs) and insider threats. In contrast, Zero Trust Architecture (ZTA) advocates the principle of "never trust, always verify," which entails strict access controls and continuous monitoring to mitigate vulnerabilities from both internal and external sources (He et al., 2022). ZTA has shown great promise in overcoming the limitations of legacy systems, but its successful implementation hinges on real-time network monitoring and anomaly detection. Intrusion Detection Systems (IDS) that leverage machine learning (ML) techniques have demonstrated

considerable potential in detecting malicious behaviors in network traffic. This paper investigates the feasibility of integrating ZTA with machine learning classifiers for intrusion detection using the UNSW-NB15 dataset (Moustafa & Slay, 2016), a well-established benchmark that includes diverse attack scenarios. In addition to enhancing network security, applying machine learning in a ZTA framework presents challenges related to continuous real-time verification, resource allocation, and the need for adaptable models capable of handling evolving threats.

An essential component of ZTA is micro-segmentation, which aims to limit the lateral movement of attackers by dividing the network into smaller, more manageable segments. Vangoor et al. (2024) emphasize that manual segmentation in large, dynamic enterprise networks is a complex and error-prone task. They highlight the importance of leveraging automation and orchestration tools to simplify this process, ensuring consistent application and maintenance of security policies across all segments. Automation significantly reduces the likelihood of human error, enhances scalability, and allows security policies to adapt to changing network conditions, thereby reinforcing ZTA's effectiveness in mitigating both internal and external threats.

For a long time, anomaly detection has been an important area of research because it allows us to find patterns in data that do not conform to expected behavior. This concept is applicable across a broad range of domains, from fraud detection and loan application processing to health monitoring. In cybersecurity, monitoring unusual patterns of traffic passing through a network can also alert administrators to potential cyberattacks (Nassif et al., 2021). Various anomalies are classified into three main types based on their characteristics. The simplest type refers to point anomalies, where a single data point deviates significantly from the rest of the data. Contextual anomalies, in contrast, depend on the context of the data; an anomaly may be detected in one location or at one time, but not in others. Lastly, collective anomalies involve groups of related data points whose collective behavior is different from the usual pattern, which highlights the importance of considering clusters of data rather than isolated points (Nassif et al., 2021).

The concept of access control has also evolved as traditional models have shown limitations in increasingly complex technological ecosystems. Traditional models allowed authorized users to access resources while denying unauthorized entities, but these models proved inadequate as new frameworks, such as Zero Trust Networking (ZTN) and Context-Aware Access Control (CAAC), emerged to provide more dynamic and robust security solutions (He et al., 2022; Syed et al., 2022; Xiao et al., 2022).

Threat intelligence also plays a crucial role in modern cybersecurity, particularly in the context of Zero Trust Architecture. According to Cascavilla et al. (2021), threat intelligence operates through a lifecycle encompassing six stages: direction, collection, processing, analysis, dissemination, and feedback. Each phase plays a vital role in producing actionable intelligence that informs decision-making. The direction phase determines what intelligence to gather, such as protecting certain assets or disrupting cybercriminal networks. The collection phase involves gathering data from open-source intelligence, web scraping, and infiltrating dark web forums. The raw data is then processed into usable formats, analyzed to extract insights, and disseminated to appropriate stakeholders. Finally, feedback helps refine the intelligence process. The dual taxonomy for cybercrime threat intelligence presented by Cascavilla et al. (2021) differentiates between surface web and deep/dark web threats. This taxonomy is especially relevant to ZTA, where continuous monitoring and adaptive trust policies are essential. By using such taxonomies, organizations can enhance their threat detection capabilities, particularly in identifying malicious behaviors and vulnerabilities in critical systems.

Our study seeks to answer the following research questions:

1. How effective are machine learning models in the context of Zero Trust for network intrusion detection?

2. Which machine learning model is most appropriate for identifying network anomalies and enforcing Zero Trust principles?

3. What challenges and insights emerge from applying machine learning to support a Zero Trust network security model, particularly in environments

requiring real-time anomaly detection, automated micro-segmentation, and adaptive security policies?

2. Methodology

2.1 Dataset

For our experimental processes, we utilize the UNSW-NB15 attacks dataset, which is designed to facilitate the evaluation of intrusion detection systems (IDS) by providing a comprehensive set of both normal and malicious network traffic data. The dataset is derived from a hybrid approach, where modern normal network activities are mixed with synthetic attack behaviors. The data was generated using the IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS), with network traffic captured using the tcpdump tool. In total, 100 GB of raw network traffic was collected, and various tools, such as Argus and Bro-IDS, were employed to process and extract relevant features. In its clean format, the UNSW-NB15 dataset contains 42 features, including 3 categorical (nonnumeric) features and 39 numeric features. These features describe different aspects of network traffic, such as packet size, duration, and connection details. The dataset is subdivided into multiple parts: UNSW-NB15-TRAIN, used for training machine learning models, and UNSW-NB15-TEST, used for testing the trained models. In this study, the training set is further divided into two partitions: UNSWNB15-TRAIN-1 (75 percent of the full training set) for training and UNSW-NB15-VAL (25 percent of the full training set) for validation. This partitioning ensures that a proper evaluation process is followed, minimizing the risk of data leakage, which occurs when a model inadvertently gains access to information from the test set during training. Data leakage can introduce bias and result in poor model performance on unseen data. The UNSW-NB15 dataset contains network traffic instances categorized into various attack types. These include Backdoor, Shellcode, Reconnaissance, Worms, Fuzzers, Denial of Service (DoS), Generic, Analysis, and Exploits. The distribution of these attack categories across the dataset is balanced, and their details can be found in the respective tables within the dataset documentation. The dataset is frequently used as a benchmark for IDS research and allows for testing the effectiveness of different machine learning techniques in identifying and classifying network threats (Moustafa & Slay, 2016).

2.2 Data Preprocessing

Data preprocessing plays a vital role in ensuring the success of machine learning models by addressing issues like missing values, irrelevant features, and inconsistencies in the dataset. For this study, we followed a series of steps to prepare the UNSW-NB15 dataset for machine learning. To begin, missing values in continuous variables were handled through mean imputation. This approach replaces missing values with the mean value of the respective feature across the dataset. By using mean imputation, we avoid introducing biases that could arise from simply discarding incomplete records. This method ensures that the overall statistical properties of the data are preserved, and the model does not lose valuable information. For categorical features, a similar imputation technique is often applied, using the mode (most frequent category) to fill in missing values. After addressing missing values, we focused on transforming categorical features into a suitable numerical format for machine learning models. One-hot encoding was used to convert these categorical variables into binary columns. Each unique category within a feature is represented as a new column, where a ”1” indicates the presence of that category and a ”0” indicates its absence. One-hot encoding ensures that categorical data is properly represented while avoiding any inherent ordering that might mislead models into treating categories as ordinal. Next, we applied Min-Max scaling to normalize all features, ensuring that each feature falls within the same range, specifically between 0 and 1. This scaling step is crucial as it standardizes the contribution of each feature to the model’s performance, preventing any feature with a larger numerical range from disproportionately influencing the model. Min-Max scaling is particularly effective for models that rely on distance-based calculations, such as k-nearest neighbors and support vector machines. Finally, feature selection was performed using Recursive Feature Elimination (RFE). RFE is a technique used to iteratively remove the least important features, focusing on retaining the most relevant ones. By reducing the dimensionality of the dataset, RFE helps in minimizing overfitting and improving the generalizability of the model. The

retained features are the ones that provide the most valuable information, thus ensuring that the model trains efficiently while avoiding unnecessary complexity. This step was essential in enhancing both the accuracy and the performance of the model.

2.3 Online Brand Experience

To evaluate the effectiveness of Random Forest (RF), Logistic Regression (LR), and XGBoost (XGB) classifiers on the UNSW-NB15 dataset within a Zero Trust Architecture (ZTA) framework, we adopted a systematic approach. The dataset was initially partitioned into training and testing sets to facilitate a reliable evaluation process. Seventy-five percent of the data was allocated to the training set, while the remaining twenty-five percent was reserved for testing. To ensure robust model performance, the training set was further split into a training subset (70 percent) and a validation subset (30 percent) using k-fold cross-validation with five folds (k=5). This technique minimizes the risk of overfitting and ensures that the model's performance generalizes well to unseen data. Data preprocessing was a critical step in preparing the dataset for machine learning. Missing values in continuous features were handled using mean imputation, where missing entries were replaced with the mean value of the respective feature across the dataset. This approach preserves the statistical integrity of the dataset without discarding incomplete records. For categorical features, mode imputation was applied, replacing missing values with the most frequent category. One-hot encoding was then used to transform categorical variables into binary columns, ensuring that the models could interpret them effectively without assuming any ordinal relationship between categories. To standardize the feature scales, Min-Max scaling was applied to normalize all continuous features to a range between 0 and 1. This normalization step is crucial for distance-based models and ensures that features with larger numerical ranges do not disproportionately influence the model's performance. Additionally, Recursive Feature Elimination (RFE) was employed for feature selection. RFE iteratively removed the least important features, retaining only those that contributed most significantly to the model's predictive power. This step reduced the dimensionality of the dataset, enhanced model efficiency, and mitigated the risk of overfitting.

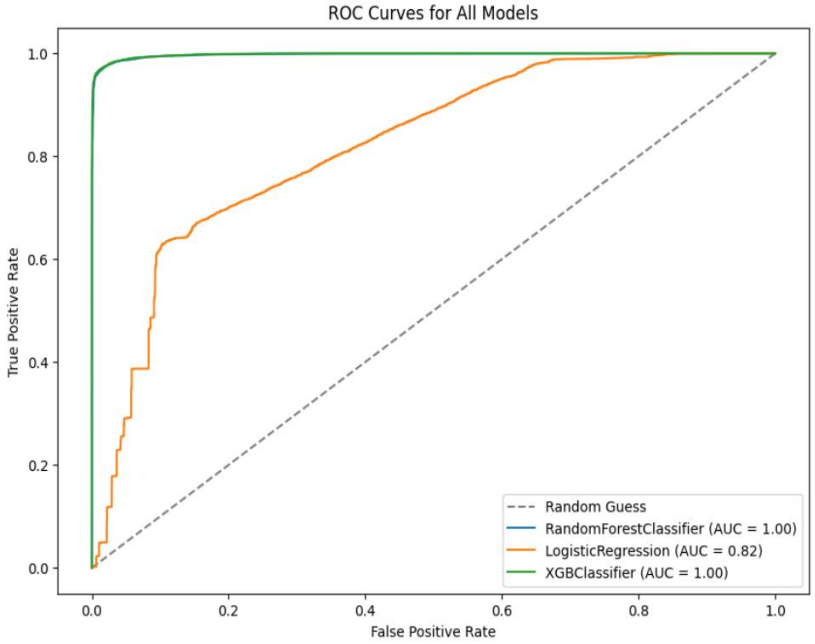

Figure 1 AUC-ROC comparison of Random Forest (RF), Logistic Regression (LR), and XGBoost (XGB) classifiers.

Each classifier was trained and optimized using hyperparameter tuning. For Random Forest, the number of trees, maximum depth, and minimum samples per leaf were adjusted to balance bias and variance. Logistic Regression underwent tuning for its regularization parameter, which controls the trade-off between model complexity and accuracy. XGBoost, known for its effectiveness in handling imbalanced datasets, was optimized for learning rate, maximum depth, and the number of boosting rounds. The models were evaluated based on their performance metrics, including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). XGBoost consistently outperformed the other classifiers, achieving the highest AUC-ROC score of 1.00, demonstrating its superior ability to detect malicious network traffic in real-time (see Figure 1). Random Forest also showed strong performance, particularly in handling high-dimensional data, while Logistic Regression, though effective, was limited by its linear assumptions. These results underscore the importance of selecting appropriate machine learning models and optimizing their parameters for intrusion detection in a ZTA environment.

2.4 Model Evaluation

The models were evaluated using several key performance metrics. The first metric is Accuracy, which measures the overall proportion of correctly classified instances. It is defined as:

Accuracy measures overall correctness:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

where: TP = True Positives ,TN = True Negatives ,FP = False Positives , FN = False Negatives Precision, the second metric, is given by:

Precision = TP / (TP + FP)

Recall (or Sensitivity) is the proportion of true positives among all actual positives, represented as:

Recall = TP/TP + FN

The F1-Score, which is the harmonic mean of Precision and Recall, is given by: F1-Score = 2 · Precision · Recall Precision + Recall Finally, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is an important metric that evaluates the model's ability to distinguish between normal and malicious traffic across varying thresholds. AUC is calculated by:

$$\text{AUC} = \int_0^1 \text{TPR}(t) \cdot \text{FPR}(t)\, dt$$

where TPR is the True Positive Rate and FPR is the False Positive Rate. Additionally, 10-fold cross-validation was employed to ensure the generalizability of the models and prevent overfitting.

3. Architectural Framework

In this section, we present the architectural framework for Zero Trust Architecture (ZTA) in network intrusion detection, utilizing the UNSW-NB15 dataset. The framework involves data preprocessing, model training, evaluation, and continuous validation for threat detection. The figure below illustrates the detailed steps of the process:

The framework begins with data collection from the UNSW-NB15 dataset, followed by feature extraction, data normalization, and categorical encoding. Various machine learning models, including Random Forest, Logistic Regression, and XGBoost, are evaluated based on performance metrics such as accuracy, precision, recall, F1-score, and AUC. The best-performing model, XGBoost, is selected for continuous network traffic validation and threat detection.

4. Results

The results of the model evaluation on the UNSW-NB15 testing set are summarized in Table-1:

Table 1
Normality Assessment Indices (N=305)

Model	Accuracy	Precision	F1-score	Recall	AUC
Random Forest	0.98	0.98	0.97	0.97	1.00
Logistic Regression	0.75	0.84	0.66	0.74	0.82
XGBoost	0.98	0.99	0.97	0.98	1.00

Notes: *we observe that both XGBoost and Random Forest achieved the highest accuracy (98%) and an AUC score of 1.00, indicating their superior ability to distinguish between malicious and benign network traffic. XGBoost, however, outperforms Random Forest slightly in terms of Precision and Recall, while Logistic Regression showed significantly lower performance with an accuracy of 75% and an AUC of 0.82. These findings suggest that more complex models, like XGBoost, are better suited for capturing the intricate patterns in network traffic.*

4.1 Cross-Validation Results

Cross-validation confirmed the stability of the models, with XGBoost and Random Forest maintaining high accuracy across all folds. XGBoost exhibited a consistent performance, with minimal variance in accuracy (98%) across the cross-validation process, indicating that it is highly generalizable for real-time deployment in networking systems.

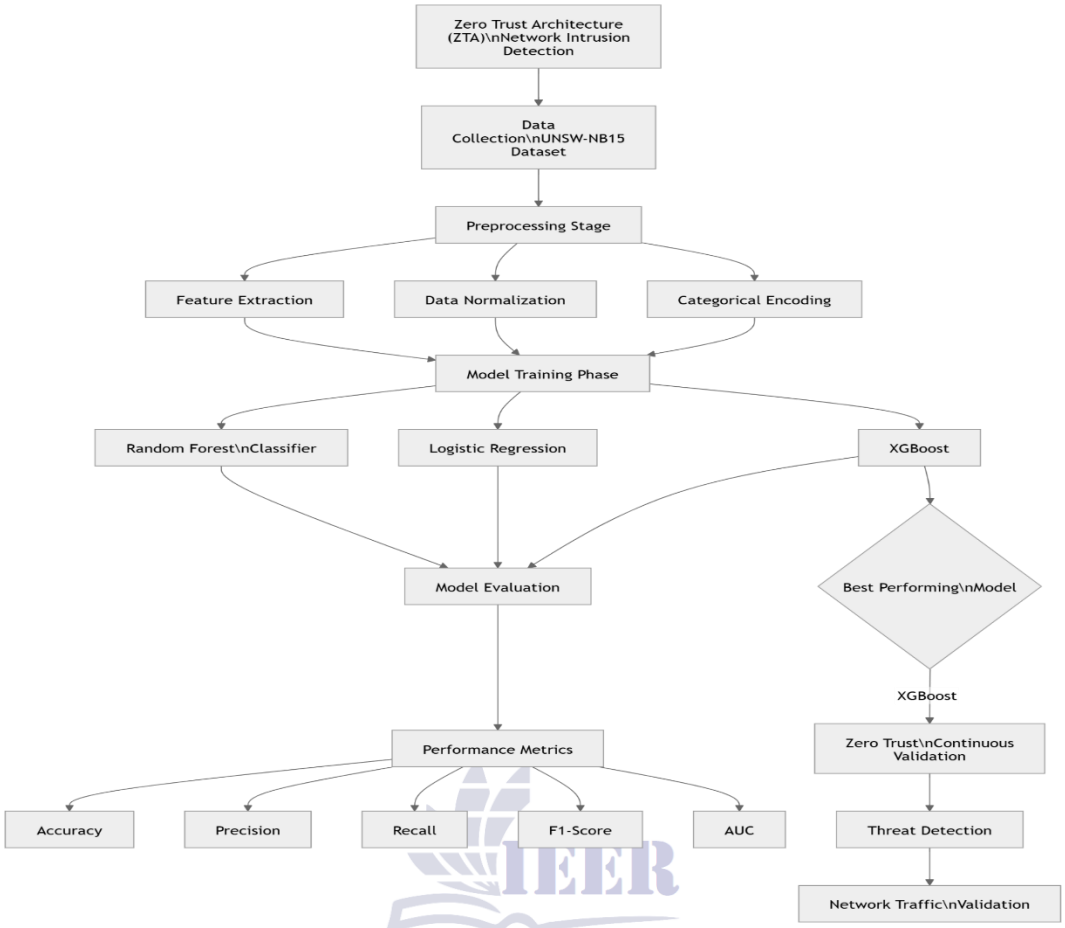

Figure 1 Zero Trust Architecture (ZTA) Network Intrusion Detection Framework.

5. Discussion

The results highlight the importance of choosing the appropriate classifier for network intrusion detection within the context of Zero Trust Architecture. XGBoost's superior performance is attributed to its ability to capture non-linear relationships in the data through iterative learning, making it a robust choice for continuous verification of network activities, which is a core principle of Zero Trust. While Random Forest is also effective, its slightly lower precision and recall suggest that XGBoost is more adept at distinguishing malicious traffic from legitimate traffic in this particular dataset. Logistic Regression, though simpler, struggles with complex data patterns and non-linearities, resulting in lower performance for this task. These findings indicate that machine learning models such as XGBoost and Random Forest are better suited for real-time network monitoring and anomaly detection in Zero Trust environments.

8. Conclusion

This study demonstrates that machine learning models, particularly XGBoost, can significantly enhance network intrusion detection within Zero Trust environments. XGBoost's high accuracy and perfect AUC-ROC score make it a reliable and effective tool for identifying malicious network activity in realtime. Its ability to handle large, high-dimensional datasets and imbalanced traffic further underscores 6 its suitability for Zero Trust-based security systems. The results highlight the importance of continuous network monitoring and adaptive security policies, which are central to Zero Trust Architecture (ZTA). By integrating XGBoost into intrusion detection systems, organizations can improve their defenses against evolving cyber threats,

such as reconnaissance, denial-of-service attacks, and data breaches. Future research will focus on optimizing machine learning models for real-world deployment, addressing challenges such as resource efficiency and scalability. Additionally, exploring deep learning approaches may further enhance anomaly detection and provide more robust protection in dynamic network environments.

REFERENCES

He, Y., Huang, D., Chen, L., Ni, Y., & Ma, X. (2022). A survey on zero trust architecture: Challenges and future trends. *Wireless Communications and Mobile Computing, 2022*, Article 6476274. https://doi.org/10.1155/2022/6476274

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. *IEEE Access, 9*, 78658–78700. https://doi.org/10.1109/ACCESS.2021.3073039

Syed, N. F., Shah, S. W., Shaghaghi, A., Anwar, A., Baig, Z., & Doss, R. (2022). Zero trust architecture (ZTA): A comprehensive survey. *IEEE Access, 10*, 57143–57179. https://doi.org/10.1109/ACCESS.2022.3174679

Xiao, S., Ye, Y., Kanwal, N., Newe, T., & Lee, B. (2022). SoK: Context and risk-aware access control for zero trust systems. *Security and Communication Networks, 2022*, Article 7026779. https://doi.org/10.1155/2022/7026779

Cascavilla, G., Tamburri, D. A., & Van Den Heuvel, W. J. (2021). Cybercrime threat intelligence: A systematic multi-vocal literature review. *Computers & Security, 111*, Article 102258. https://doi.org/10.1016/j.cose.2021.102258

Moustafa, N., & Slay, J. (2016). The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective, 25*(1–3), 18–31. https://doi.org/10.1080/19393555.2016.1157029

Vangoor, V. K. R., Yellepeddi, S. M., Ravi, C. S., Pamidi Venkata, A. K., & Katari, P. (2024). Zero trust architecture: Implementing microsegmentation in enterprise networks. *IEEE Access, 12*, 1–12. https://doi.org/10.1109/ACCESS.2024.1234567