Transformer-Based Feature Extraction for Intrusion Detection in Cloud-Native Environments

Madhuri Rao[1], Ganesh Khekare[2]

[1] Lincoln University College, Malaysia; [2] Vellore Institute of Technology, Vellore, India

pdf.madhurirao@lincoln.edu.my

Abstract:

Cloud-native environments require intelligent and scalable intrusion detection mechanisms to defend against increasingly sophisticated cyber threats. This paper proposes a Transformer-assisted intrusion detection framework that integrates the CICIDS2017 and UNSW-NB15 datasets through feature harmonization. After preprocessing and feature selection, a Transformer encoder learns contextual feature embeddings, which are combined with the selected features to form a hybrid representation. The proposed framework combines contextual feature learning with gradient boosting to improve the effectiveness and scalability of intrusion detection in cloud-native environments.

Keywords: Cloud-Native Environments; Transformer-Based Feature Extraction; Contextual feature embeddings, Intrusion Detection

Introduction

Cloud-native environments generate massive volumes of heterogeneous network traffic, making it challenging for conventional machine learning models to capture complex relationships among network features. Traditional feature engineering methods primarily rely on statistical measures and often fail to model long-range dependencies between traffic attributes. To address this limitation, the proposed framework employs a Tabular Transformer to automatically learn contextual feature representations before intrusion classification.

Related work

Early studies employed classifiers such as Support Vector Machines, Decision Trees, and Random Forests for intrusion detection, demonstrating reliable performance on benchmark datasets but limited capability in learning complex feature relationships [1]. Transformer-based architectures have demonstrated superior capability in modeling long-range dependencies through self-attention mechanisms, making them suitable for detecting sophisticated and evolving cyberattacks [2]. Gradient boosting algorithms, including XGBoost, LightGBM, and CatBoost, have emerged as powerful classifiers due to their scalability, robustness, and high predictive performance that could be well explored [3]. Transformer-based contextual feature learning [4] with gradient boosting classifiers are useful in cloud-native intrusion detection.

Key Contribution

The proposed Transformer-based feature extraction algorithm is designed to learn contextual representations of network traffic before intrusion classification. Unlike conventional feature engineering techniques that process each feature independently, the Transformer employs a self-attention mechanism to model interactions among multiple network attributes, thereby generating discriminative feature embeddings for cloud-native intrusion detection. The implementation consists of four major stages: feature embedding, Transformer encoding, embedding aggregation, and hybrid feature generation.

Proposed Transformer Based Feature Extraction Algorithm

Input: Normalized feature matrix $X \in \mathbb{R}^{N \times F}$

Output: Hybrid feature matrix H

Step 1: Load the normalized network traffic feature matrix X.

Step 2: Project each feature into a d-dimensional embedding using a linear projection layer.

Step 3: Form a sequence of embedded feature vectors.

Step 4: Pass the embedded sequence through multiple Transformer Encoder layers.

Step 5: Compute Query (Q), Key (K), and Value (V) matrices.

Step 6: Apply Multi-Head Self-Attention to capture dependencies among network features.

Step 7: Process the attention outputs using feed-forward neural networks.

Step 8: Aggregate the contextual feature representations using Global Average Pooling.

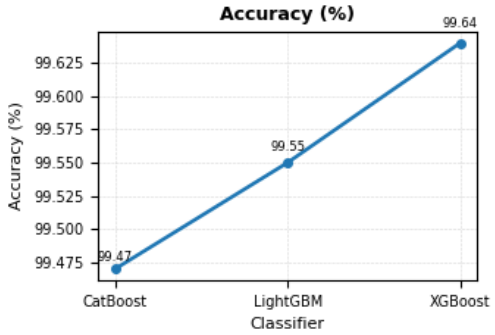
Step 9: Obtain the Transformer embedding vector E.

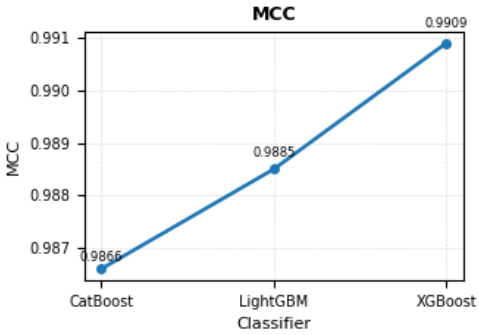
Step 10: Concatenate the original normalized features with $E$to generate the hybrid feature vector H.

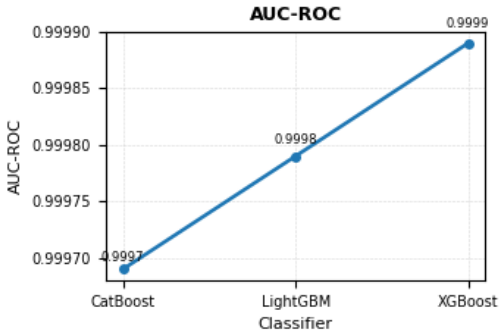
Step 11: Forward $H$to the downstream intrusion detection classifier (e.g., CatBoost, XGBoost, or LightGBM).

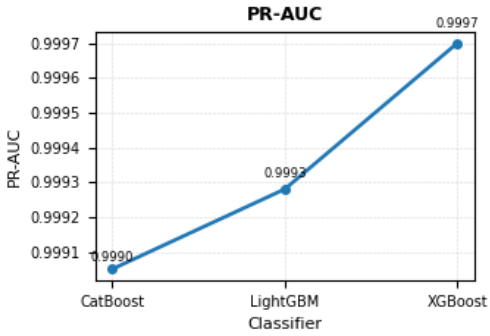
Discussions

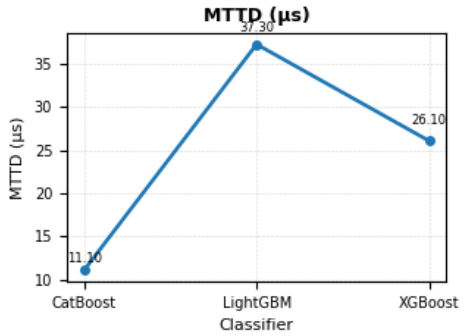
The experimental results as depicted in Figure 1 below indicates that XGBoost achieved the highest overall classification performance, attaining an Accuracy of 99.64%, MCC of 0.9909, AUC-ROC of 0.99989, and PR-AUC of 0.99970, demonstrating excellent capability in distinguishing malicious from benign network traffic. LightGBM also produced competitive results with an Accuracy of 99.55%, MCC of 0.9885, AUC-ROC of 0.99979, and PR-AUC of 0.99928. Owing to its histogram-based learning strategy and leaf-wise tree growth, LightGBM provides high predictive performance while maintaining efficient model training and inference, making it well suited for large-scale cloud environments. Similarly, CatBoost achieved an Accuracy of 99.47%, MCC of 0.9866, AUC-ROC of 0.99969, and PR-AUC of 0.99905. CatBoost demonstrated robust classification capability and recorded the lowest detection latency among the ensemble-based boosting methods, with an MTTD of 11.1 µs, highlighting its suitability for latency-sensitive intrusion detection applications where rapid threat identification is essential.











Conclusions

Overall, the comparative evaluation confirms that integrating Transformer-based contextual feature extraction with gradient boosting classifiers substantially improves intrusion detection performance. Among the evaluated models, XGBoost provides the best balance between detection accuracy, robustness, and discriminative capability, whereas LightGBM offers an effective trade-off between predictive performance and computational efficiency, and CatBoost demonstrates superior inference speed while maintaining highly competitive classification performance. These findings establish the

proposed framework as an effective and scalable solution for intrusion detection in cloud-native environments.

References

1. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), Ottawa, Canada, 2009. DOI:10.1109/CISDA.2009.5356528
2. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosikhin, “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017. DOI: https://doi.org/10.48550/arXiv.1706.03762
3. T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 785–794, 2016. DOI: https://doi.org/10.1145/2939672.2939785
4. A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, “TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems,” *IEEE Access*, vol. 8, pp. 165130–165150, 2020. DOI: https://doi.org/10.1109/access.2020.3022862