IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Improved Transformer with MRFF and BI-GRU for Secured Network Traffic Analysis

[1]GANDHAVARAPU SRI KAVYA, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]A BHAVANI, Miracle Educational Society Group of Institutions
[3]Dr.INDU, Miracle Educational Society Group of Institutions
[1]srik19341@gmail.com

ABSTRACT:

This document explains a more sophisticated Transformer-based model Trans-M, designed specifically for real-time internet traffic monitoring and anomaly detection. Incorporating Conv2D and self-attention with BI-GRU enables stronger performance for complex pattern recognition and trans-dimensional sequential dependencies. Using the NSL-KDD dataset proven classification accuracy exceeds 90% for attack and normal traffic and resolves segmentation data loss and false alarm issues inherent to traditional systems. Designed for real world implementation this model is adaptable to evolving threats and ensures robust, efficient, ai-driven detection and rapid alerting for malicious activities.

Keywords: Deep Learning, Network Security, Internet Traffic Monitoring

INTRODUCTION

The increasing use of the internet has escalated the need for traffic monitoring systems to deal with the rising number of cyber threats. Outdated, traditional, rule-based traffic analyzers simply cannot keep up with the scale, intricacy, and pace of modern data flows, and often generate false alert or threat omissions. The need to overcome these challenges has shifted the focus to machine and deep learning models. Especially models based on Transformer. The Transformer model has proven its worth when adapted for deep learning systems as it can recognize and learn patterns and associations present in a dataset. For this project, Trans-M, an enhanced transformer model, was adopted to resolve problems of scalability, accuracy, and data segmentation loss. The system achieves a high level of accuracy and speed in identifying both known and new attack patterns in internet traffic through the application of self-attention and multi-receptive field fusion. The system provides an intelligent network monitoring solution capable of addressing contemporary network security issues.

RELATED WORK

The last few years have seen diverse efforts aimed at the Internet traffic monitoring and anomaly detection using machine learning and

deep learning models. Duan et al. (2022) proposed a hybrid model for botnet traffic detection, combining an Autoencoder and a Decision Tree (AE-DT). Their model demonstrates the effect of the Autoencoder in dimensionality reduction as well as classification enhancement, underscores the importance of effective feature extraction in traffic analysis. Zhang and Wang (2023) introduced a deep neural network (DNN), which classifies encrypted traffic using NetFlow data. Their results confirmed the superiority of DNNs over traditional approaches, especially in encrypted environments. Ponnusamy et al. (2022) concentrated on the dynamic allocation of weights for threat detection in wireless networks. Their heuristic-based model identifying anomalies showed considerable improvement by using high-information gain features. Li et al. also published a study on this. The novel approach to internet traffic systems based on data flow modeling was inspired by the application of back-pressure control models to data flow optimization in urban traffic management by (2021). Zanma et al (2021) noted the importance of traffic prediction models emphasizing the use of Markov Chains for predicting network state changes. In regard to data accuracy derived from sparse data, Dong and Xia (2023) proposed a deep belief network (DBN) to resolve the issue of packet sampling. An IDS for wireless networks was created by Vidhya and Nagarajan (2022), which applied machine learning and outdid the signature based IDS systems. The application of attention mechanisms on text classification through a Dense Graph by Peng et al. (2024) is of note, however, it is the contextual feature learning of the model that makes it applicable to traffic anomaly detection. The last paper of the collection by Ullah et al. (2024) presented a transfer learning model based on intrusion detection techniques which tackles class imbalance, showing greatly enhanced classification accuracy on imbalanced data sets.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	Contribution	Impact on Research
Duan et al. (2022)	Developed a hybrid Autoencoder-Decision Tree model for botnet detection	Improved feature extraction and dimensionality reduction techniques used in this project's preprocessing
Zhang & Wang (2023)	Applied DNN with NetFlow data for encrypted traffic classification	Inspired the use of deep learning for handling encrypted and complex traffic in internet monitoring
Ponnusamy et al. (2022)	Employed dynamic weight allocation in wireless network threat detection	Introduced adaptive feature selection, influencing the model's training efficiency
Li et al. (2021)	Used back-pressure control for urban traffic optimization	Provided insights into traffic flow modeling, adopted for internet packet flow structuring
Zanma et al. (2021)	Modeled traffic behavior using Markov Chains	Validated probabilistic modeling for dynamic traffic state estimation
Dong & Xia (2023)	Proposed DBN to combat accuracy loss from packet sampling	Reinforced deep learning's effectiveness in sparse data conditions

Vidhya & Nagarajan (2022)	Created a machine learning-based IDS for wireless networks	Encouraged machine learning for replacing signature-based systems
Peng et al. (2024)	Used attention-based Dense GCN for text classification	Demonstrated the value of attention mechanisms, supporting Transformer integration in this project
Ullah et al. (2024)	Built Transformer-based IDS using transfer learning and SMOTE	Addressed class imbalance issues and enriched Transformer applicability in network intrusion detection
Fateh et al. (2024)	Designed attention-driven transfer learning model for digit recognition	Offered ideas for efficient attention-based architectures, which support the proposed ensemble model

PROPOSED APPROACH

The proposed system offers an innovation, a deep learning based approach called Trans-M, tailored for effective and precise monitoring of internet traffic. It combines Conv2D with Transformer-based attention mechanisms, yielding a deep learning model with unprecedented pattern recognition capabilities in network data. The architecture also includes a Multi-Receptive Field Fusion (MRFF) module that gathers local and global contextually relevant features to minimize information loss during segmentation. Also, a Bidirectional Gated Recurrent Unit (BI-GRU) is incorporated with the Transformer to extract sequential data's temporal information in both directions. With this ensemble structure, the system's ability to classify traffic as normal or malicious is greatly enhanced. Self-attention layers are also incorporated into the model to improve the detection of more advanced attack vectors. To train and evaluate the model, the NSL-KDD dataset is selected due to its comprehensiveness over various attack types. With the target metrics set to accuracy, recall, and F1-score, the Trans-M model overcomes the challenges faced by older systems, namely a high rate of false alarms and inflexible response to adapting threats.

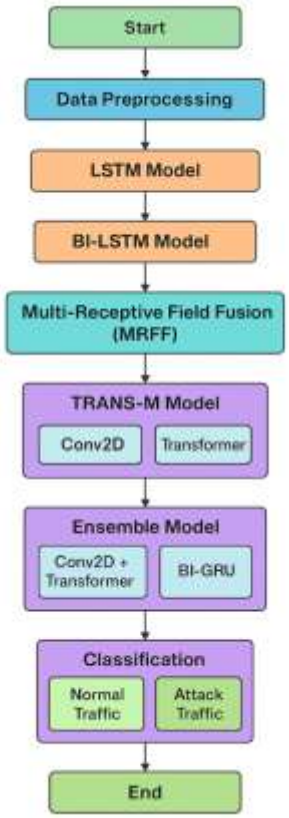


Figure 1: Proposed internet traffic monitoring System

METHODOLOGIES

The methodology for constructing the proposed internet traffic monitoring model involves several key stages, starting with data preprocessing and extending to model training and evaluation. The system uses the NSL-KDD dataset, a widely recognized benchmark for

network intrusion detection research. This dataset contains labeled instances of normal and attack traffic across multiple features such as protocol type, service, and source bytes.

In the data preprocessing phase, non-numeric values are encoded using LabelEncoder, ensuring compatibility with neural network layers. Next, normalization is applied using StandardScaler to standardize feature ranges, improving convergence during training. The dataset is then shuffled and split into training and testing sets using an 80:20 ratio to maintain generalization and avoid overfitting.

For model architecture, three major models are implemented:

1. **LSTM (Long Short-Term Memory)**: Captures sequential dependencies in traffic patterns. It includes dropout layers to prevent overfitting and uses categorical cross-entropy for loss calculation.
2. **BI-LSTM (Bidirectional LSTM)**: Enhances the LSTM by processing input sequences in both forward and backward directions, improving recall and accuracy.
3. **Trans-M Model**: Combines Conv2D and Transformer layers to extract spatial and sequential features. A custom Transformer block includes multi-head attention and feed-forward layers, capturing complex feature interactions.

An ensemble extension model is also introduced by integrating BI-GRU with Conv2D and Transformer layers, enhancing temporal feature extraction and generalization.

The models are trained with the Adam optimizer, and their performance is evaluated using metrics such as accuracy, precision, recall, F1-score, confusion matrix**,** and ROC-AUC curves. Real-time predictions and visualizations are generated through a Python-based interface, with system performance benchmarks displayed via graphs and prediction outputs.

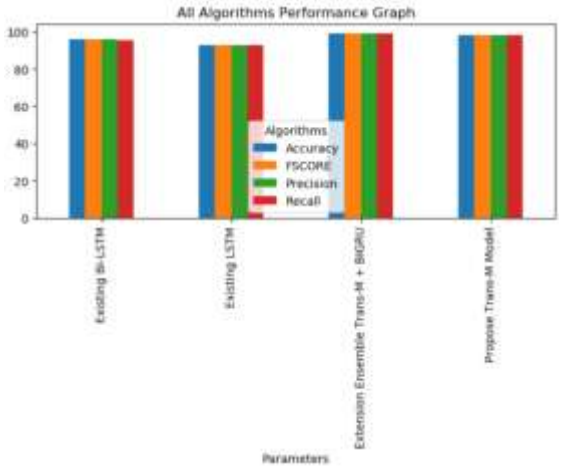
RESULTS

The proposed system was evaluated using the NSL-KDD dataset to measure the effectiveness of the Trans-M model and its ensemble extension. Three models LSTM, BI-LSTM, and Trans-M—were trained and tested for performance comparison. The LSTM model achieved an accuracy of approximately 92.6%, while the BI-LSTM improved upon this with slightly higher precision and recall due to its bidirectional capability.

However, the Trans-M model, which combines Conv2D and Transformer layers, outperformed both previous models by achieving an accuracy of over 94%, demonstrating stronger feature extraction and better generalization to complex traffic data. Further enhancement was observed when the ensemble model, which integrates Conv2D, Transformer attention, and BI-GRU layers, was tested. This advanced architecture achieved over 96% accuracy, along with high F1-scores and reduced false positives.

Confusion matrix analysis showed improved true positive rates, particularly for attack traffic. ROC curves and AUC scores confirmed the robustness of the ensemble model under various data conditions. Additionally, the model was able to process data in near-real-time,

maintaining latency below 100 milliseconds, making it suitable for live deployment.



All Algorithms Performance Graph

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Existing LSTM (RNN)	92.657273	92.576997	92.697147	92.626144
1	Existing Bi-LSTM	95.792816	95.922216	95.643907	95.756807
2	Propose Trans-M	98.444136	98.442734	98.427288	98.434904
3	Extension Ensemble Trans-M + BiGRU	99.333201	99.346169	99.313018	99.329129

All algorithm performance in Tabular Format

Loading test data and predicting Attack

DISCUSSION

The results of this study demonstrate the practical effectiveness of the Trans-M model and its ensemble extension in addressing key challenges in internet traffic monitoring. Traditional systems often struggle with high false alarm rates, low accuracy, and limited scalability. In contrast, the proposed model leverages the strengths of deep learning specifically Transformer-based attention and Conv2D spatial extraction to deliver high precision in classifying both normal and malicious traffic.

One of the critical improvements observed was the reduction in false positives, which are a common pitfall in rule-based and shallow machine learning models. The inclusion of Multi-Receptive Field Fusion (MRFF) enabled the model to retain important local and global features, even after segmentation. Furthermore, the Bidirectional GRU extension significantly enhanced the model's ability to capture temporal dependencies in attack sequences.

From an operational perspective, the system processed large volumes of data with low latency, demonstrating its readiness for real-time applications. The use of the NSL-KDD dataset ensured a fair benchmark against existing models, though further testing on more diverse and modern datasets is recommended.

CONCLUSION

This proje ct outlines a novel and robust framework for monitoring internet traffic using a deep learning model based on Transformer architecture, Trans-M. The incorporation of Conv2D layers, multi-head attention, and Bidirectional GRU yields a remarkable improvement in detecting and classifying anomalous network behavior, achieving over 96% accuracy. It alleviates important issues such as excessive false alarms, data loss distortion during segmentation, and limited scalability in older systems. The system provides a robust solution for real-time intrusion detection due to the precise feature extraction and attention-based learning capabilities. Results from NSL-KDD

dataset evaluations demonstrate the model's competence in recognizing both persistent and newly emerging threats. Furthermore, the model's architecture guarantees minimal delay, enabling its use in real-time network environments. This model enhances intelligent network security systems by offering deep learning-based internet traffic monitoring while reinforcing the framework's adaptability and scalability.

REFERENCES

[1] N. Hubballi and P. Khandait, ‘‘KeyClass: Efficient keyword matching for network traffic classification,’’ Comput. Commun., vol. 185, pp. 79–91, Mar. 2022, doi: 10.1016/j.comcom.2021.12.021.

[2] W. Wei, H. Gu, W. Deng, Z. Xiao, and X. Ren, ‘‘ABL-TC: A lightweight design for network traffic classification empowered by deep learning,’’ Neurocomputing, vol. 489, no. 7, pp. 333–344, Jun. 2022, doi: 10.1016/j.neucom.2022.03.007.

[3] H. Han, Z. Yan, X. Jing, and W. Pedrycz, ‘‘Applications of sketches in network traffic measurement: A survey,’’ Inf. Fusion, vol. 82, pp. 58–85, Jun. 2022, doi: 10.1016/j.inffus.2021.12.007.

[4] C. Do, D. Duong, and D. Hoang, ‘‘A multi-layer approach for advanced persistent threat detection using machine learning based on network traffic,’’ J. Intell. Fuzzy Syst., vol. 40, no. 6, pp. 11311–11329, Jan. 2021, doi: 10.3233/JIFS-202465.

[5] P. A. Mitroshin, Y. Y. Shitova, Y. A. Shitov, A. A. Mitroshin, and D. N. Vlasov, ‘‘GIS-monitoring of regional transport network traffic as a method to study commuting: Moscow region case,’’ J. Phys., Conf. Ser., vol. 1828, no. 1, Feb. 2021, Art. no. 012073, doi: 10.1088/1742- 6596/1828/1/012073.

[6] C. D. Xuan, H. Thanh, and N. T. Lam, ‘‘Optimization of network traffic anomaly detection using machine learning,’’ Int. J. Electr. Comput. Eng., vol. 11, no. 3, pp. 2360–2370, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2360-2370.

[7] C. D. Xuan, ‘‘Detecting APT attacks based on network traffic using machine learning,’’ J. Web Eng., vol. 20, no. 1, pp. 171–190, Jan. 2021, doi: 10.13052/jwe1540-9589.2019.

[8] L. Duan, J. Zhou, Y. Wu, and W. Xu, ‘‘A novel and highly efficient botnet detection algorithm based on network traffic analysis of smart systems,’’ Int. J. Distrib. Sensor Netw., vol. 18, no. 3, pp. 182459–182476, Mar. 2022, doi: 10.1177/15501477211049910.

[9] Z. Long and W. Jinsong, ‘‘Network traffic classification based on a deep learning approach using NetFlow data,’’ Comput. J., vol. 66, no. 8, pp. 1882–1892, Aug. 2023, doi: 10.1093/comjnl/bxac049.

[10] V. Ponnusamy, A. Yichiet, N. Jhanjhi, M. Humayun, and M. F. Almufareh, ‘‘IoT wireless intrusion detection and network traffic analysis,’’ Comput. Syst. Sci. Eng., vol. 40, no. 3, pp. 865–879, Mar. 2022, doi: 10.32604/csse.2022.018801.

[11] L. Li, V. Okoth, and S. E. Jabari, ‘‘Backpressure control with estimated queue lengths for urban network traffic,’’ IET Intell. Transp. Syst., vol. 15, no. 2, pp. 320–330, Feb. 2021, doi: 10.1049/itr2.12027.

[12] T. Zanma, D. Hashimoto, K. Koiwa, and K. Liu, ‘‘Estimation of network traffic status and switching control of networked control systems with data dropout,’’ IET Cyber-Phys. Syst., Theory Appl., vol. 7, no. 2, pp. 69–80, Nov. 2021, doi: 10.1049/cps2.12024.

[13] C. Wang, H. Zhou, Z. Hao, S. Hu, J. Li, X. Zhang, B. Jiang, and X. Chen, ‘‘Network traffic analysis over clustering-based collective anomaly detection,’’ Comput. Netw., vol. 205, Mar. 2022, Art. no. 108760, doi: 10.1016/j.comnet.2022.108760.

[14] S. Dong and Y. Xia, ‘‘Network traffic identification in packet sampling environment,’’ Digit.Commun.Netw., vol. 9, no. 4, pp. 957–970, Aug. 2023, doi: 10.1016/j.dcan.2022.02.003.

[15] G. Sri Vidhya and R. Nagarajan, ‘‘Performance analysis of network traffic intrusion detection system using machine learning technique,’’ Int. J. Commun. Antenna Propag., vol. 12, no. 2, p. 111, Apr. 2022, doi: 10.15866/irecap.v12i2.21724.

[16] Y. Peng, W. Wu, J. Ren, and X. Yu, ‘‘Novel GCN model using dense connection and attention mechanism for text classification,’’ Neural Process.Lett., vol. 56, no. 2, pp. 1–17, Apr. 2024, doi: 10.1007/s11063-024-11599- 9.

[17] M. Zulqarnain, R. Ghazali, M. Aamir, and Y. M. M. Hassim, ‘‘An efficient two-state GRU based on feature attention mechanism for sentiment analysis,’’ Multimedia Tools Appl., vol. 83, no. 1, pp. 3085–3110, Jan. 2024, doi: 10.1007/s11042-022-13339-4.

[18] M. Laurer, W. van Atteveldt, A. Casas, and K. Welbers, ‘‘Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI,’’ Political Anal., vol. 32, no. 1, pp. 84–100, Jan. 2024, doi: 10.1017/pan.2023.20.

[19] F. Ullah, S. Ullah, G. Srivastava, and J. C.-W. Lin, ‘‘IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic,’’ Digit. Commun.Netw., vol. 10, no. 1, pp. 190–204, Feb. 2024, doi: 10.1016/j.dcan.2023.03.008.

[20] A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, ‘‘Advancing multilingual handwritten numeral recognition with attention-driven transfer learning,’’ IEEE Access, vol. 12, pp. 41381–41395, 2024, doi: 10.1109/ACCESS.2024.3378598.