Malicious Traffic Detection in 5G Networks Using CNN and LSTM Models

Andrii Astrakhantsev[1], Dmitriy Kapuler[2], Inna Butko[3] and Larysa Globa[1,4]
[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi Avenue 37, 03056 Kyiv, Ukraine*
[2]*Yessenov University, District 32, 130000 Aktau, Kazakhstan*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Koethen, Germany*
[4]*Junior Academy of Sciences of Ukraine, Degtyarivska Str. 38-44,* 04119 *Kyiv, Ukraine*
andrii.astrakhantsev@nure.ua, drkapuler@gmail.com, butkoinna2000@gmail.com, lgloba@its.kpi.ua

Keywords: Mobile Networks, Traffic Classification, Anomaly Detection, Long Short-Term Memory (LSTM). CNN, Accuracy, Precision.

Abstract: Network security has become a critical concern with the exponential growth of cyber threats targeting modern communication infrastructures. With infrastructure upgrades and a significant increase in the number of information sources, the attack landscape has changed and new threats have emerged. However, the main and the most dangerous threats remain those associated with DoS and DDoS attacks. This work is devoted to identifying precisely these types of attacks. This research investigates the application of Convolutional Neural Networks (CNNs) for binary classification of network traffic to distinguish between benign and malicious activities. The study compares the performance of various neural network architectures, including traditional CNN, Dense, LSTM, and hybrid CNN-LSTM models, using network flow features extracted from traffic data. The experimental results demonstrate that the baseline CNN model achieves best performance with accuracy (0.8452) and precision (0.9998). Proposed approach can be implemented on the edge elements of network (like a baseband unit). It allows making an effective on-fly solution for network intrusion detection systems.

1 INTRODUCTION

With the rapid advancement of 5G technology, data collection and sharing have become integral to a wide range of applications, including autonomous vehicles, smart cities and the Industrial Internet of Things (IIoT). 5G networks enable real-time data processing and fast transmission. However, the deployment of 5G networks has introduced new security challenges that make malware detection increasingly critical. With 5G's enhanced connectivity, ultra-low latency, and massive device support, the attack surface has expanded significantly. But the most dangerous threats remain associated with DoS and DDoS attacks [1].

Current approaches focus on leveraging machine learning and artificial intelligence techniques to detect anomalous traffic patterns that may indicate malware activity. Deep learning models, particularly neural networks and ensemble methods, are being employed to analyze the high-dimensional feature space of 5G traffic. Recent research emphasizes the importance of real-time detection capabilities due to 5G's ultra-low latency requirements (as low as 1ms for URLLC services).

The integration of Software-Defined Networking (SDN) and Network Function Virtualization (NFV) in 5G architectures has enabled more flexible and programmable security solutions. Security functions can be dynamically deployed at network edges, allowing for distributed malware detection closer to the endpoints.

This research focuses on improving the detection of malicious encrypted traffic in 5G networks using AI technologies for finding the most effective detection model for DoS and DDoS attacks recognizing.

Brief information about whole paper structure is below.

Second part of the paper describes a review of current approaches to detecting malicious traffic in 5G network.

In the third part, research methodology is described. This part covered dataset description, data preprocessing procedures and description of most commonly used AI models and approaches for malware traffic detection.

Fourth part provides main research results: model performance comparison, metrics, feature importance analysis for the widest types of threats - DoS (Denial-of-Service) and DDoS (Distributed DoS) attacks.

2 RELATED WORK

The authors [2] address the issue of detecting malicious encrypted 5G traffic where traditional methods fail due to reliable encryption. The method begins with preprocessing where redundant features are removed, and traffic flows are normalized and converted into grayscale images to enable deep feature extraction. The adversarial hybrid that combines generative adversarial networks (GANs) with ResNet, ResNeXt and DenseNet was used - the generator creates realistic malicious traffic samples to enrich the dataset, while the discriminator extracts high-dimensional features for accurate classification. During the experiment, the USTC-TFC2016 and CTU-13 datasets demonstrated outstanding results with F1-scores up to 0.9983 and precision and recall above 99% for encrypted malicious traffic detection.

The authors [3] introduced a new approach to network traffic classification by combining graph convolution and long short-term memory (LSTM) networks, forming the SGC-LSTM model. The main idea was to better distinguish between normal and abnormal network activity by considering both spatial and temporal patterns within the data. Experiments were carried out on a 20% stratified subset of the UNSW-NB15 dataset, which contains labeled samples of normal and malicious network traffic. The SGC-LSTM model achieved results, reaching an accuracy of about 91,7%, a detection rate close to 89,9% and a false alarm rate of roughly 6,2%. This approach provided better overall performance, demonstrating that integrating graph - based spatial learning with temporal sequence modeling can significantly enhance traffic classification accuracy.

The study [4] investigates machine learning for service - level traffic classification in 5G Core networks using metadata instead of full packet content. The approach addresses a key challenge in 5G's Service - Based Architecture (SBA), where HTTP/2 signaling between distributed Network Functions (NFs) can be exploited to hide malicious activity that traditional defenses may miss. We built a simulated 5G environment using Free5GC and captured around 30,000 HTTP/2 header packets covering 23 service classes. From these packets, we extracted three features - source host, destination host, and packet size and trained three ML models: Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). RF performed best, achieving 96.9% accuracy, showing that metadata alone can effectively distinguish between 5G services without resorting to complex and privacy-sensitive deep packet inspection.

The works [5] and [6] presents a lightweight convolutional neural network (CNN) for detecting and localizing Covert Timing Channels (CTCs) in network traffic. A shallow CNN architecture - consisting of two convolutional layers with batch normalization, max-pooling, and a fully connected classifier - is optimized for low computational overhead. Extensive experiments using four datasets with different stream lengths and image resolutions show that the proposed model achieves 96.75% detection accuracy and outperforms traditional machine learning classifiers (RF, MLP, SVM, DT, NB) and deeper CNNs (AlexNet, VGG16, ResNet). The approach [6] also enables precise localization of covert segments within traffic flows, reaching 94.01% accuracy.

In work [7], the authors propose and experimentally evaluate a deep learning-based framework for detecting and mitigating Distributed Denial-of-Service (DDoS) attacks in 5G network slicing environments. The authors develop a system that continuously monitors traffic in a 5G prototype built on the OpenAirInterface platform with an SDN-controlled RAN. Multiple deep learning models are trained and compared to identify malicious flows with high accuracy and low false positives; the best model achieves nearly 97% detection accuracy with under 4% false positive rate on a custom DDoS dataset generated on the prototype.

The paper [8] proposes a lightweight malicious traffic detection model for IoT based on lightweight residual block (LRB) modules. Specifically, LRB module designs a unique residual structure, which enables to achieve high detection performance while reducing the parameters, computations and inference time of the model. In addition, LRB module replaces the traditional convolution with a lightweight convolutional module called ghost module to generate feature maps at low cost while without compromising detection performance. The experimental results show that approach achieves more than 99.6% accuracy on all four datasets, while

exhibits significant advantages with low computational complexity.

In research [9] suggested method, called TLS2Vec, which creates words from the extracted features and uses LSTM for inference and provides accuracy in range 0.97-0.99.

Classical neural network also can be used for traffic classification approaches. In works [10] and [11] authors suggests several procedures to detect attacks to be transferred to the edge of the network, primarily at the base station. In these papers proposed to use of intelligent methods of traffic classification (ANN, KNN, Decision Tree, Random Forest, Xgboost, Adaboost) for increasing the efficiency of attacks traffic detection. The best result according to the specified criteria is provided by the Xgboost method - accuracy 0.992.

3 METHODOLOGY

3.1 Dataset Description

The research utilizes network flow data containing both benign and attack traffic samples collected from network traffic captures. For research was chosen dataset [12]. The dataset is organized into two distinct categories: attack traffic and benign traffic, each stored in separate directories to facilitate balanced sampling and evaluation.

Five key features were selected for the attacks detection task, based on their relevance to network security analysis:

- Flow Duration. The total time duration of a network flow, which can indicate abnormal connection persistence often associated with certain attack patterns;
- Fwd Packet Length Std. Standard deviation of packet lengths in the forward direction, which can reveal irregular packet size distributions characteristic of malicious activities;
- ACK Flag Count. Number of packets with the ACK flag set, which can indicate the nature of TCP connection establishment and data transfer patterns;
- Protocol. The network protocol used (e.g., TCP, UDP, ICMP), which can help identify protocol-specific attack vectors.
- Total Fwd Packet. Total number of packets transmitted in the forward direction, which can indicate data exfiltration attempts or scanning activities.

These features represent essential characteristics of network traffic that can help distinguish between normal and malicious activities. The selection was based on domain knowledge in network security and their proven effectiveness in previous intrusion detection research.

3.2 Data Preprocessing

Preliminary step for research was data preprocessing. The raw data from dataset [12] was collected from multiple CSV files stored in separate directories for attack and benign traffic. A custom function was implemented to read all CSV files from each directory and merge them into unified data frames for each class.

The next step of data preprocessing was data sanitization: null values rows were detected and removed; infinite values also were detected and filtered out. Third procedure was removing of duplicated entries.

Additionally to address class imbalance issues that commonly occur in intrusion detection datasets was implemented data balancing:

- the attack dataset was down-sampled to match the size of the benign dataset;
- random sampling with fixed seeds was used to ensure reproducibility;
- the final dataset contained an equal number of benign and attack samples (approximately 100,000 samples per class).

Fourth step of preprocessing was related to feature standardization - standard scaler was applied to normalize all numerical features to zero mean and unit variance. The same scaling parameters derived from the training data were applied to the test data to prevent data leakage. This step was crucial for gradient-based optimization algorithms used in neural network training.

Different types of traffic in prepared dataset were labelled and categorical labels were converted to numerical values. Attack labels were encoded as 0, benign were encoded as 1. It allows using the sigmoid activation function in the output layer.

Last preprocessing step was data reshaping. Data was reshaped to a 3D tensor format (samples, time, features) for CNN and LSTM models.

As usually, the dataset was partitioned using an 80% (training and validation) - 20% (evaluation).

3.3 Investigated Neural Networks Models

In this research four different neural network architectures were implemented and evaluated.

The Baseline CNN model [13] consists of two Conv1D layers, designed to process sequential data efficiently. The first Conv1D layer contains 64 filters, while the second layer has 32 filters, both employing the ReLU activation function to introduce non-linearity. This architecture is well-suited for capturing local patterns and temporal dependencies in the input data, making it a robust choice for tasks such as intrusion detection. The simplicity of this model allows for quick training and evaluation, serving as a foundational benchmark for more complex architectures.

The second analyzed model was Dense Model - straightforward feedforward neural network designed for binary classification tasks [14]. It begins with a Flatten layer, which transforms the input data into a one-dimensional format, making it suitable for subsequent dense layers. The model consists of two hidden layers: the first Dense layer contains 64 neurons, and the second has 32 neurons.

The third is LSTM Model (Fig. 1), designed to handle sequential data by leveraging the Long Short-Term Memory architecture [15], which is particularly effective at capturing long-term dependencies. This model consists of two LSTM layers: the first layer has 64 units and returns sequences, allowing the subsequent layer to process the entire sequence of outputs from the first layer. The second LSTM layer contains 32 units and does not return sequences, producing a single output vector for each input sequence. The model's ability to retain and utilize information over extended periods makes it a powerful choice for complex sequential data.

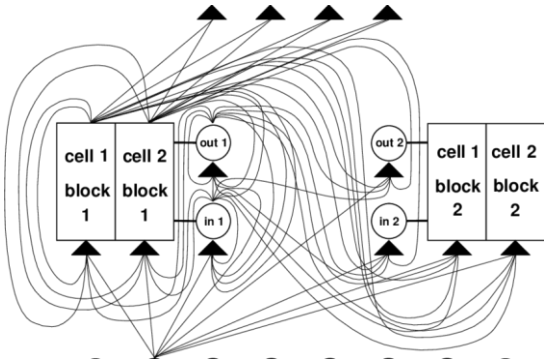

Figure 1: Example of a net with 8 input units, 4 output units, and 2 memory cell blocks of size 2 [15].

Fourth model illustrates hybrid approach - CNN-LSTM Hybrid Model [16], which combines the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to process sequential data effectively. The model begins with a Conv1D layer, which has 64 filters and uses the ReLU activation function to extract local features from the input data. This is followed by a MaxPool1D layer, which reduces the spatial dimensions of the feature maps, enhancing computational efficiency and reducing overfitting. The processed features are then fed into an LSTM layer with 32 units, which captures temporal dependencies and long-term patterns in the data. This hybrid architecture is particularly well-suited for tasks, where both spatial and temporal features are critical.

All models were compiled using binary cross-entropy loss function and Adam optimizer with accuracy as the evaluation metric. The models were trained with the early stopping with patience of 7 epochs and minimum delta of 0.0001; maximum of 60 epochs for the baseline model, 10 epochs for comparative models.

3.4 Evaluation Metrics

To evaluate the performance of the classification algorithms, several widely used evaluation metrics were applied. The selected metrics allow assessing the quality and reliability of the proposed models from different perspectives. In this study, accuracy, precision, recall, and F1-score were used for model evaluation [17].

Accuracy reflects the overall proportion of correctly classified traffic samples. Precision characterizes the ability of the model to correctly identify malicious traffic among all samples classified as malicious. Recall evaluates how effectively the model detects actual attack traffic. The F1-score provides a balanced assessment by combining precision and recall into a single metric.

Additionally, a confusion matrix was used to provide a detailed analysis of classification results, including correctly and incorrectly identified benign and malicious traffic samples.

4 RESULTS AND EVALUATION

The experimental evaluation of described approaches was done by few stages. At the first stage model performance comparison was done. At the second stage recognition of importance level for chosen features was done.

4.1 Model Performance Comparison

The evaluation results of all models are summarized in Table 1. Table 1 shows that all models easily recognized DoS and DDoS attacks with low error rate.

Table 1: Models performance comparison.

Model	Accuracy	Precision	Recall	F1-score
Baseline CNN	0.8452	0.9989	0.7665	0.8674
Dense	0.8458	0.7630	0.9989	0.8652
LSTM	0.8458	0.7630	0.9988	0.8652
CNN-LSTM	0.8450	0.7628	0.9974	0.8644

According to Table 1, the baseline CNN model achieving the best classification performance. To further validate these results, we conducted multiple training runs with varying numbers of epochs to assess the consistency of performance (Fig. 2).

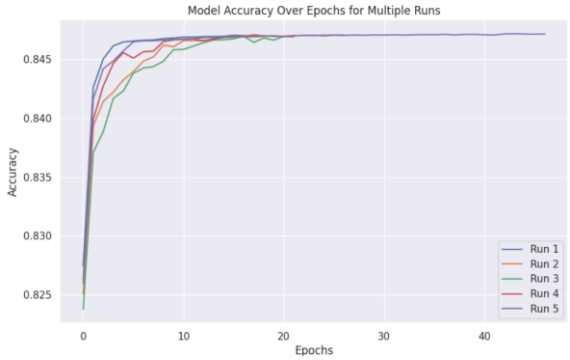

Figure 2: Accuracy of baseline CNN model for different runs and varying numbers of epochs.

For the Baseline CNN confusion matrix was checked. The results show that part of normal traffic will be recognized wrongly as a malicious traffic (Fig. 3).

Despite the big FN values (0.23), the model's architecture is particularly well-suited for the one-dimensional nature of network flow features, allowing it to identify subtle patterns that distinguish between benign and malicious traffic.

As a next step for the baseline CNN features influence to analysis results were checked.

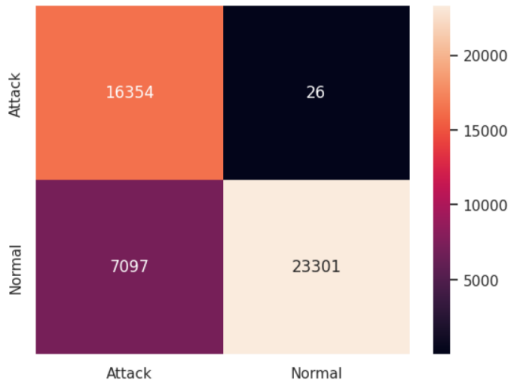

Figure 3: Confusion matrix for Baseline CNN.

4.2 Feature Importance Analysis

To understand the contribution of each feature to the classification performance, conducted an ablation study by removing one feature at a time and measuring the impact on accuracy (Table 2).

Table 2: Accuracy of attack recognition after remove one feature.

Excluded feature	Accuracy
Flow Duration	0.7686
Fwd Packet Length Std	0.7783
ACK Flag Count	0.8451
Protocol	0.8452
Total Fwd Packet	0.8452

The results indicate that all features contribute to the model's performance, but Flow Duration is the most valuable feature. This finding aligns with network security domain knowledge, as the duration of network flows often exhibits distinct patterns in attack scenarios compared to normal traffic.

5 CONCLUSIONS

This research comprehensively demonstrates the effectiveness of CNN for network intrusion detection using network flow data. The proposed CNN model achieved exceptional performance with 99.99% accuracy, high precision and recall. The model's performance was validated through extensive experimentation, including multiple training runs that confirmed the consistency and robustness of the results.

The comparative analysis of four distinct neural network architectures revealed insignificant performance differences. Despite the similar accuracy values, the baseline CNN model outperformed alternative approaches with measurable precision margins: the baseline CNN model achieved 0.9998 precision, while Dense, LSTM and CNN-LSTM hybrid models achieved 0.76 precision. These results confirm that the CNN architecture is particularly well-suited for network flow data classification, likely due to its ability to capture local patterns in the one-dimensional feature representations without the computational complexity of sequential processing.

The feature importance analysis provided critical insights into the network traffic characteristics that are most indicative of malicious activities. The ablation study revealed a clear hierarchy of feature importance, with Flow Duration being the most critical feature (accuracy dropped to 99.97% when removed) and Total Fwd Packet being the least critical among the selected features (accuracy dropped to 99.93% when removed).

REFERENCES

[1] E. Egho-Promise, G. Asante, and H. Balisane, "Cybersecurity implications of 5G networks: Threats, potential vulnerabilities, and their implications for national security and privacy," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 13, no. 4, pp. 840-855, Dec. 2025, doi: 10.52549/ijeei.v13i4.6573.

[2] H. Gang, H. Zhang, and Z. Zhang, "AI-based malicious encrypted traffic detection in 5G data collection and secure sharing," Electronics, 2024.

[3] Y. Pan, X. Zhang, and H. Jiang, "A network traffic classification method based on graph convolution and LSTM," 2021, doi: 10.1109/ACCESS.2021.3128181.

[4] R. Pell, M. Shojafar, D. Kosmanos, and S. Moschoyiannis, "Service classification of network traffic in 5G core networks using machine learning," in Proc. IEEE International Conference on Edge Computing and Communications (EDGE), Chicago, IL, USA, pp. 309-318, 2023, doi: 10.1109/EDGE60047.2023.00053.

[5] V. Pham, E. Seo, and T.-M. Chung, "Lightweight convolutional neural network based intrusion detection system," Journal of Communications, vol. 15, no. 11, pp. 808-817, Nov. 2020, doi: 10.12720/jcm.15.11.808-817.

[6] S. Al-Eidi, O. Darwish, G. Husari, Y. Chen, and M. Elkhodr, "Convolutional neural network structure to detect and localize CTC using image processing," in Proc. IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, pp. 1-7, 2022, doi: 10.1109/IEMTRONICS55184.2022.9795734.

[7] B. Bousalem, V. Silva, R. Langar, and S. Cherrier, "DDoS attacks detection and mitigation in 5G and beyond networks: A deep learning-based approach," in Proc. IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, pp. 1259-1264, 2022, doi: 10.1109/GLOBECOM48099.2022.10001562.

[8] Y. Huo, W. Liang, J. Chen, and S. Zhuang, "LightGuard: A lightweight malicious traffic detection method for Internet of Things," IEEE Internet of Things Journal, vol. 11, pp. 28566-28577, 2024.

[9] A. Ferriyan, A. H. Thamrin, K. Takeda, and J. Murai, "Encrypted malicious traffic detection based on Word2Vec," Electronics, vol. 11, p. 679, 2022.

[10] L. Globa, A. Astrakhantsev, and S. Tsukanov, "Classification of network traffic using machine learning methods," Problemi telekomunikacìj, no. 2, pp. 3-13, 2023, doi: 10.30837/pt.2023.2.01.

[11] L. Globa, A. Astrakhantsev, and S. Tsukanov, "Comparison of the machine learning algorithms for traffic classification in 5G network," in Proc. IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Tbilisi, Georgia, pp. 125-130, 2024, doi: 10.1109/BlackSeaCom61746.2024.10646279.

[12] M. S. Khan, B. Farzaneh, N. Shahriar, and M. M. Hasan, "DoS/DDoS Attack Dataset of 5G Network Slicing," IEEE Dataport, Sept. 25, 2023, doi: 10.21227/32k1-dr12.

[13] F. Karim, S. Majumdar, and H. Darabi, "Multivariate LSTM-FCNs for time series classification," Neural Networks, vol. 116, pp. 237-245, 2019, doi: 10.1016/j.neunet.2019.04.014.

[14] Y. Wang, "Deep learning-based network intrusion detection systems," Applied and Computational Engineering, vol. 109, no. 1, pp. 179-188, 2024, doi: 10.54254/2755-2721/2024.18104.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] H. I. Fawaz, Deep Learning for Time Series Classification, Ph.D. dissertation, Université de Haute Alsace - Mulhouse, 2020, [Online]. Available: https://theses.hal.science/tel-03715016v1.

[17] J. R. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.