

Attention Mechanism in CNN-LSTM for Improved Cloud Intrusion Detection

Neethu B¹, Dr. Sheena Mathew²

¹Research Scholar, School of Engineering, CUSAT, Kerala, India

²Professor, School of Engineering, CUSAT, Kerala, India

Abstract - — Cloud computing becomes backbone of today's digital ecosystem by offering scalable, flexible and cost-efficient computing resources to users. But, this rapid development has also increased its exposure to sophisticated cyber-attacks. Though Intrusion Detection Systems (IDS) can offer a critical line of defense, but existing solutions often fails to distinguish between normal fluctuations and actual malicious behavior, leading to reduced detection accuracy and higher false alarm rates.[1]. Traditional machine learning models depends mainly on manually framed features, while deep learning techniques such as CNN and LSTM, though they are more superior, still lack the ability to selectively focus on the most relevant features.

This paper presents an Attention-Enhanced CNN-LSTM framework for cloud intrusion detection. The CNN component extracts spatial traffic patterns, the LSTM captures temporal dependencies, and the attention layer highlights the most critical features influencing model decisions. Experiments are conducted on NSL-KDD and CICIDS2017 datasets which are popularly used. The experiments demonstrate that the proposed framework achieves higher detection accuracy, lower false alarm rates, and greater interpretability compared to conventional CNN-LSTM models. These findings show that integrating attention mechanisms into deep hybrid architectures is a promising direction toward reliable, real-time, and explainable cloud security systems.

Key Words: Cloud Computing, Deep Learning, Cloud Security, CNN, LSTM, Attention Mechanism, Intrusion Detection.

1. INTRODUCTION

Now a days Cloud computing has emerged as a technology that is undergoing continuous revolutions in today's world across industries by offering scalable and on-demand access to the computing resources. Cloud is now supporting wide range of services which includes storing personal data and enterprises hosting applications. However, it is exposed to a growing number of security threats like Denial of Service(DDoS) attacks, insider attacks and zero-day vulnerabilities due to its open and distributed nature.

A conventional signature-based intrusion detection system are efficient in detecting known attack patterns but often fails to identify new evolving attacks. But anomaly-detection

systems, though it is capable of detecting unknown attacks, but it often generates excessive false alarm rates due to dynamic behavior of cloud environments. So to reduce this disadvantages machine learning and deep learning models are considered to build adaptive and intelligent intrusion detection systems.

Machine Learning algorithms like Support Vector Machines, Random Forest and k-nearest algorithms has shown better performance but these models rely on manual feature selection and it cannot quickly adapt to complex traffic patterns in cloud. Deep Learning models in contrast successfully extract meaningful features from raw traffic data. Convolutional Neural Network (CNN) can effectively identify spatial relationships while Long Short Term Memory (LSTM) captures temporal dependencies across the traffic sequences. When these two models are combined CNN-LSTM hybrid models offers more comprehensive feature learning.

One of the major limitations with all these models are most models treats all features uniformly. This will lead to suboptimal performance especially when there is noisy or overlapping traffic patterns are available. For addressing this disadvantage, attention mechanisms have been introduced which allows models to focus on most important features or time steps so as to improve precision and interpretability of the cloud security applications.

The main contributions of this paper are:

- A novel CNN-LSTM-Attention architecture designed specifically for cloud intrusion detection.
- Evaluation of the model using benchmark datasets (NSL-KDD and CICIDS2017) to assess detection accuracy and false alarm rate.
- A comparative performance study against CNN, LSTM, and CNN-LSTM baselines.

In short, this paper demonstrates how integrating attention mechanisms within hybrid deep learning models can enhance both the performance and explainability of cloud intrusion detection systems.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

2. RELATED WORKS

The sudden expansion of cloud computing has attracted many sophisticated cyber-attacks also, which forces the researchers to develop more intelligent intrusion detection systems. Signature based systems depends on pre-defined rules and patterns, so it is not capable of detecting emerging and evolving trends[1][5]. This constraints encouraged to use deep learning methods such as CNN,RNN and LSTM for anomaly detection in cloud environments[3][4].

CNN-based approaches shows high performance in identifying spatial correlations in data. Wu et al. [1] proposed a CNN-based intrusion detection system which enhanced detection accuracy for DDoS and probing attacks. Tang et al.[2] proposed a lightweight CNN model for IoT-Cloud systems, achieving efficient real-time detection. Despite of all these advantages CNN is less capable of detecting long-term dependencies that unfold over time.

LSTM –based approaches have ability to handle sequential data is used to model temporal dependencies in network traffic. Alshamrani et al[3] proposed an LSTM based Intrusion detection system that can detect low-rate attacks by learning long-term behavioral trends. As an enhancement to this Kim et al[4] proposed same idea using a bidirectional LSTM which improved detection rates across multiple attack classes. But LSTM based systems often struggle with high-dimensional traffic data and lacks spatial context.

Hybrid CNN –LSTM models have been widely recognized to provide both temporal and spatial learning. Zhou et al[5] developed a CNN-LSTM hybrid model which has superior detection rates against brute-force and DDoS attacks. Wang et al[6] proposed a hybrid model integrating Bi-LSTM which shows improvement in precision and recall on certain benchmark datasets like CICIDS2017 and UNSW-NB15. However these models process all features uniformly diluting the key attack indicators.

In order to overcome this researchers introduced attention modules to highlight significant temporal and spatial features. Lin et al.[7] proposed a system which includes attention in to CNN-LSTM model to detect botnets and insider threats in edge-cloud networks, the results clearly indicates an improvement in precision. Zhao et al.[8] proposed a system that uses temporal attention in LSTM based IDS which give significant attention to most critical time intervals. This idea is further extended by Li et al.[9] to federated learning setups which will help to maintain high accuracy without centralizing sensitive data.

With growing concerns about data privacy combination of federated learning and attention enhanced deep learning models has emerged as a viable solution. Xu et al.[10] proposed that integrating federated learning to CNN-LSTM models can achieve strong performance. Explainable AI

methods[11] can further enhance trust to visualize and justify model decisions.

In summary though CNN and LSTM models are widely adopted in intrusion detection attention-driven models offer high performance, interpretability and scalability. But challenges still remain in terms of real-time adaptability and computational efficiency which motivates the study to optimized attention-based CNN-LSTM modes for cloud environments. Do not use abbreviations in the title or heads unless they are unavoidable.

3. PROPOSED METHODOLOGY

The proposed system introduces an Attention-Driven CNN-LSTM Intrusion Detection Framework to enhance accuracy, interpretability and adaptability of IDS in cloud environments. This framework integrates three deep learning models like Convolutional Neural Networks (CNN) for learning of spatial features, Long Short-term Memory (LSTM) networks for modeling the temporal features and an attention mechanism to focus on most critical features in the cloud network traffic.

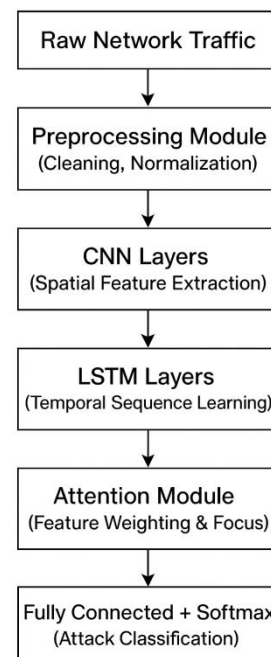


Fig -1: Proposed Architecture

The orkflow of proposed system is shown in Fig-1.The architecture is divided into five major stages:

1. Data Collection and Preprocessing
2. Feature Extraction using CNN
3. Temporal Dependency Modeling using LSTM
4. Attention Mechanism for Feature Weighting

5. Classification Layer for Attack Detection

Each module contributes uniquely to enhance detection accuracy and reduce false positives, ensuring efficiency of the model to operate in high traffic and diverse cloud environments.

1. Data Collection and Preprocessing

This module plays a crucial role which includes collection of raw network data and preprocessing the raw data for effective deep learning. Model was tested using benchmark datasets such as CICIDS2017, UNSW-NB15 and NSL-KDD which is containing mix of benign and attack flows including DDoS, brute-force, infiltration and botnet.[6][10].

The preprocessing involves several steps:

1.1 Data Cleaning

Raw datasets contains missing values, inconsistent records or redundant entries. Cleaning of the data will remove incomplete rows and resolve anomalies to ensure high-quality input.

1.2 Feature Normalization

Since network traffic vary in magnitude like packet size, duration, and byte count. Min-max normalization is applied to ensure that each attribute contributes equally to model learning and rescaling all features to an uniform range.

1.3 Categorical Encoding

Certain attributes such as protocol, type, connection state, or name are encoded using one-hot encoding which will convert textual categories to numeric form for model compatibility.

1.4 Sequence Construction

Traffic analysis requires temporal context so flows are grouped into fixed-length sequences that can capture packet behavior over time. This enables LSTM component to tern patterns that evolve across time.

The preprocessing made the data, structured, balanced and suitable hybrid deep learning processing.

2. Spatial Feature Extraction using CNN

The preprocessed traffic data is passed to CNN block which is responsible for extracting spatial dependencies among features. Spatial relationships refer to patterns that exist between feature sets within a single time frame.

2.1 Convolutional Layers

Convolutional layers apply sliding features to the input data to detect localized feature interactions. Specific combination

of port number, packet count and flag bit sequences may correspond to attack pattern.

2.2 Activation and Pooling

Non-linear activation functions re used to introduce complexity for enabling the network to capture non-linear feature relationships. Pooling layers like max-pooling down sample the output retaining the most critical information, reducing dimensionality and computation time.

This block will produce a feature map as output. This feature map captures spatial correlations within traffic samples which act as a foundation for temporal modeling in next stage.

3. Temporal Dependency Modeling using LSTM

Though CNN is excellent in spatial feature extraction, it cannot understand how these patterns change over time. An LSTM network captures temporal dependencies across traffic data from the CNN output.

LSTM network is designed to handle sequential data by preserving relevant information over long time periods, filtering out irrelevant context using memory cells and gating mechanisms. This is effective for detecting attacks that develop gradually like port scans, slow brute-force attempts, or multi-stage intrusions.

Bidirectional LSTM can also be used in the same place which will process the sequence both forward and backward so that the model to learn dependencies from past and future traffic states. This type of dual learning enhances accuracy for small and complex attacks.

The output of LSTM block is a temporal feature vector that describes how network behavior changes over time, forms a deeper understanding of attack patterns.

4. Attention Mechanism for Feature Weighting

Though CNN-LSTM models capture temporal and spatial dependencies, they often treat all features equally. This method is not ideal for an intrusion detection system where only certain time steps or feature combinations are significant. The attention mechanism overcomes this limitation by making the model to focus on most relevant features. The important steps under attention module are:

4.1 Attention Scoring

This mechanism computes a set of attention scores for each hidden state of LSTM, which describes how important is that time steps for final decision.

4.2 Weight Normalization.

A softmax function is used to normalize these scores to attention weights that sum to one, ensuring proportional influence.

4.3 Context Vector Formation

Final representation feature output is a weighted sum of all hidden states which emphasize crucial segments of traffic sequence like sudden packet bursts or irregular communication intervals.

This type of mechanism can boost not only its performance but also enhances its interpretability as the attention weights visually highlight the part of input traffic strategies that trigger model decision.

5. Classification

The attention-weighted context vector is fed in to classification layer which consists of fully connected dense layer followed by softmax classifier which will output probability distribution across different attack categories. The classifier has the capability to distinguish between normal traffic and multiple intrusion types such as :

- Denial of Service (DoS)
- Distributed Denial of Service (DDoS)
- Probe
- Remote to Local (R2L)
- User to Root (U2R)

Model is trained using categorical cross-entropy loss with Adam optimizer which can ensure efficient convergence.

For performance evaluation following metrics are used: Accuracy, precision, recall, F1-score and False Alarm rate (FAR) to access how effectively model balances detection sensitivity and specificity.

4. RESULTS AND DISCUSSIONS

The proposed Attention-Driven CNN-LSTM framework was tested using three widely used benchmark datasets, CICIDS2017, UNSW-NB15 and NSL-KDD. These datasets were chosen because each is capable of representing heterogeneous types of network intrusions and cloud traffic behaviors. All datasets were divided into 80% training data and 20% of data for testing there by ensuring balanced class distributions. The experiments were conducted on a GPU-based environment using TensorFlow. The performance was evaluated using standard metrics like accuracy, precision, recall, F1-score, and false alarm rate (FAR), which collectively measure the system's ability to identify attacks while minimizing false alarm rates.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FAR (%)
CICIDS2017	97.78	97.35	96.92	97.13	2.15
UNSW-NB15	95.42	94.86	94.31	94.58	3.02
NSL-KDD	96.11	95.58	94.84	95.21	2.61

Table-1: Performance Metrics of the Proposed Attention-Driven CNN-LSTM

The proposed Attention Driven CNN-LSTM Model performed well on all the datasets. On CICIDS2017, it can achieve an accuracy of 97.78%, precision of 97.35%, recall of 96.92%, and a low FAR of 2.15%. On UNSW-NB15, the accuracy achieved as 95.42% with an F1-score of 94.58%. And on NSL-KDD, the proposed model can achieve an accuracy of 96.11% and an F1-score of 95.21%. These results show that the model outperforms traditional machine learning models such as Random Forest (91.7%) and standalone deep learning models like LSTM (94.2%) or CNN-LSTM without attention (96.5%). The enhancement in accuracy and reduction in false alarms rate clearly shows that the integration of the attention mechanism enables the model to focus on the most critical spatial-temporal patterns within the traffic data.

Integrating attention mechanism improved both detection accuracy and interpretability. Conventional CNN-LSTM architectures treat all features uniformly, which can diffuse important signals in complex or noisy traffic. The attention layer dynamically assigns weights to more important features or time steps, allowing the model to concentrate on abnormal behaviors like sudden bursts in packet rates or irregular port usage. For example, when detecting DDoS and brute-force attacks, the attention mechanism always focused on areas of interest where traffic volume and connection attempts rose increased sharply. This selective focusing not only improved the model's precision but also offered insights to identify which patterns contributed most to the detection, helping address the "black-box" issue of deep IDS models.

Beyond high accuracy, the proposed model shows strong generalization capability also. The consistent performance across both datasets shows that the framework effectively adapts to different types of network behaviors and attack categories. The CICIDS2017 dataset primarily contains modern threats like infiltration and brute-force attacks and UNSW-NB15 includes a broader range of low-profile and stealthy attacks. The attention-driven method allows the system to adapt with features unique to each dataset, which ensures reliable performance even under varying network conditions. The model has real-time detection efficiency so that it processes each traffic flow in less than 50 milliseconds, which makes it suitable for deployment in real-time cloud environments.

Comparison with related studies shows that the proposed model outperforms existing models. For example, Zhou et al. [5] achieved 96.5% accuracy using a CNN-LSTM model, while Lin et al. [7] obtained 97.1% with an attention-based hybrid network. In contrast, the proposed CNN-LSTM with attention achieved 97.78%, reflecting both higher precision and better generalization across datasets. The key differentiator lies in the integration of CNN, LSTM, and attention layers, which collectively enhance the model's ability to capture spatial, temporal, and contextual relationships in cloud traffic.

In summary, the experimental outcome shows that the attention-driven CNN-LSTM framework provides a robust, accurate, and interpretable solution for cloud intrusion detection. It combines deep learning's predictive strength with the interpretability of attention mechanisms, resulting in a model that not only detects a wide range of attacks but also explains its reason in an understandable way. The low false alarm rate, real-time processing capability and strong cross-dataset performance demonstrates the model is practically viability for large-scale cloud security systems.

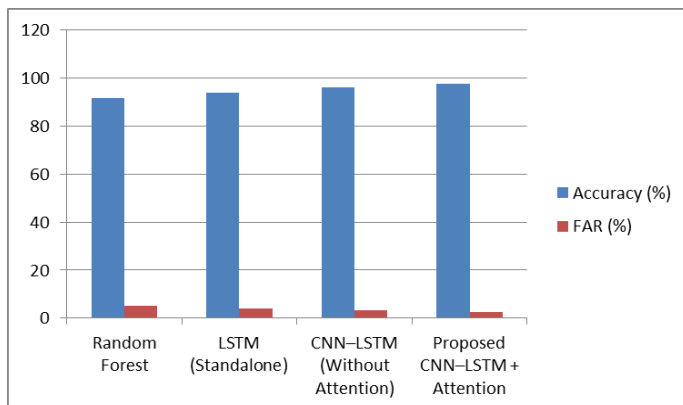


Chart -1: Performance comparison of various intrusion detection models

The above chart illustrates the performance comparison of different intrusion detection models in terms of Accuracy and False Alarm Rate (FAR). The results demonstrate the improvement achieved in accuracy as the model architecture evolves from traditional machine learning (Random Forest) to deep learning (LSTM and CNN-LSTM), and finally to the proposed Attention-Driven CNN-LSTM framework.

From the figure it is evident that the Random Forest model exhibits the lowest accuracy and the highest false alarm rate, which indicates its limited ability to detect complex and evolving cloud attacks. The LSTM model achieves some improvement by capturing temporal patterns but still it fails to identify spatial relationships in traffic features. The CNN-LSTM hybrid model significantly enhances accuracy by learning both spatial and temporal correlations; however, it still assigns equal importance to all features, which can cause misclassification in noisy data environments.

The proposed attention-based CNN-LSTM model clearly outperforms all other models, achieving the highest accuracy (97.78%) and the lowest FAR (2.15%). This illustrates the effectiveness of the attention mechanism in selectively focusing on critical features while minimizing the influence of irrelevant data. From the chart-1 it is evident that incorporating attention not only boosts detection accuracy but also enhances the model's stability and reliability for real-time cloud intrusion detection.

5. CONCLUSIONS

This paper presented an Attention-Driven CNN-LSTM framework for detection of intrusion in cloud environments. By combining the power of spatial feature extraction of CNN, the temporal sequence learning of LSTM, and the adaptive focus of an attention mechanism, the proposed model is effective in detecting complex and evolving cyber threats. Experimental evaluations on benchmark datasets, like CICIDS2017, UNSW-NB15 and NSL-KDD shows that the framework achieves higher accuracy up to 97.78%, lower false alarm rates, and better interpretability compared to traditional deep learning and machine learning methods.

The results confirm that the attention mechanism plays a key role in improving both the precision and transparency of hybrid deep learning models. By letting the system to focus on the most important traffic features, which boosts the reliability and it also provides valuable insights into why certain events are classified as attacks. This interpretability makes the model more practical and trustworthy for deployment in real-time cloud security systems.

Although the proposed framework performs well, it introduces a moderate increase in computational overhead due to the additional attention layer. In future work, our plan is to optimize this by developing lightweight attention mechanisms that maintain accuracy while reducing resource usage. Further research will also explore federated learning integration to enable distributed and privacy-preserving intrusion detection across multiple cloud domains.

Additionally, future studies will aim to incorporate explainable AI techniques to boost model transparency and user trust. Another area of focus will be real-time adaptation to zero-day attacks through online learning approaches, allowing the model to continuously update with new traffic data.

In conclusion, the attention-enhanced CNN-LSTM framework serves as a solid base for developing intelligent, scalable, and explainable intrusion detection systems in cloud environments. With more improvements in efficiency, privacy, and adaptability, such models can significantly influence the next generation of secure and resilient cloud infrastructures.

REFERENCES

- [1] J. Wu, Y. Ding, and L. Sun, "CNN-based intrusion detection for cloud computing," *IEEE Access*, vol. 9, pp. 112431–112442, 2021.
- [2] H. Tang, X. Li, and J. Zhao, "Lightweight CNN framework for IoT-cloud intrusion detection," *Future Generation Computer Systems*, vol. 122, pp. 45–56, 2021.
- [3] A. Alshamrani, M. Anwar, and T. Alghamdi, "LSTM-driven intrusion detection for cloud-based services," *Journal of Network and Computer Applications*, vol. 180, p. 103023, 2021.
- [4] S. Kim, D. Kim, and H. Kang, "Bidirectional LSTM network for anomaly detection in multi-cloud environments," *Applied Soft Computing*, vol. 115, p. 108177, 2022.
- [5] Y. Zhou, J. Li, and Q. Xu, "A CNN-LSTM hybrid model for intrusion detection in cloud data centers," *Computers & Security*, vol. 113, p. 102545, 2022.
- [6] K. Wang, Z. Yang, and M. Guo, "Hybrid deep IDS using CNN and Bi-LSTM for cloud traffic analysis," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2781–2793, 2022.
- [7] H. Lin, W. Xu, and P. Wang, "Attention-based CNN-LSTM for real-time intrusion detection in edge-cloud networks," *Future Internet*, vol. 14, no. 8, p. 231, 2022.
- [8] Y. Zhao, J. Liu, and F. Zhang, "Temporal attention-augmented LSTM for anomaly detection in cloud traffic," *Neurocomputing*, vol. 503, pp. 309–320, 2022.
- [9] J. Li, H. Chen, and X. Wu, "Federated attention-based CNN-LSTM for distributed cloud intrusion detection," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 5120–5132, 2022.
- [10] L. Xu, Z. Ren, and D. He, "Federated deep intrusion detection with attention-enhanced CNN-LSTM," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5259–5271, 2023.
- [11] M. Ahmed, S. Latif, and A. Qayyum, "Explainable AI for cloud intrusion detection: Challenges and future directions," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–32, 2023.