Journal of Wireless Sensor Networks and IoT

Journal of Wireless Sensor Networks and IoT, ISSN: 3048-8729 Vol. 2, No. 2, 2025 (pp. 66-71)

RESEARCH ARTICLE

ECEJOURNALS.IN

Deep Learning-Assisted Passive Traffic Profiling in Large-Scale IoT Networks

Srikanth Reddy Keshireddy*

Senior Software Engineer, Keen Info Tek Inc., USA

KEYWORDS:
IoT,
Wireless Sensor Networks,
Anomaly Detection,
Fog Computing,
CNN-LSTM,
Edge Intelligence,
Deep Learning

ARTICLE HISTORY:
Submitted : 02.05.2025
Revised : 12.07.2025
Accepted : 05.09.2025

https://doi.org/10.31838/WSNIOT/02.02.08

Abstract

The fast trend in the Internet of Things (IoT) device proliferation has greatly complicated the modern network ecosystems and required the development of new strategies to track device behavior and identify anomalies without violating user privacy. The current work introduces a passive traffic profiling framework with the help of deep learning, characterized by a classification of IoT devices relying on the data presented in the network in terms of packet timing, header entropy, and the flow properties on the basis of network metadata only. CNN-LSTM model is a hybrid convolutional neural network that is trained using a dataset of more than 25 million packets measured in various IoT applications, and it has an average classification accuracy of 95 percent on sixteen device types. The comparative analysis indicates that the proposed framework is better than the traditional statistical and shallow learning models in terms of scalability, latency, and the ability to resist the effect of encryption. These findings confirm that passive profiling based on deep learning is a convenient and non-invasive method of monitoring a large network of IoT devices, allowing to strengthen network security, behavioral analytics, and provide early warning of threats.

Author's E-mail: sreek.278@gmail.com

Author's Orcid: 0009-0007-6482-4438

How to cite this article: Keshireddy SR. Deep Learning-Assisted Passive Traffic Profiling in Large-Scale IoT Networks. Journal of Wireless Sensor Networks and IoT, Vol. 2, No. 2, 2025 (pp. 66-71).

Introduction

The Internet of Things (IoT) has become one of the paradigms of the contemporary networking, where billions of heterogeneous devices, including home automation systems and industrial sensors, can be interconnected and able to communicate with each other autonomously.[15] With the increase in both size and type of IoT deployments, new deployments create masses and masses of traffic which is typically encrypted and dynamic and thus current packet inspection and signature-based monitoring methods become less and less effective.[11, 14] Passive traffic profiling, which is based on the study of the flow features of metadata without the need to decode the holding, has become a focus as a promising technique to conduct behavioural inferences and anomaly detection.[1, 2]

The initial approaches to passive profiling used were rule-based heuristics or statistical clustering models, like K-means or Gaussian mixtures, to use flow-level patterns to cluster the IoT traffic.[2, 15] Although these techniques provided understanding on particular device behaviour, they could not be as flexible to large and multi-protocol IoT networks because of overlapping feature signatures, loss of features due to encryption, and failure to capture long-term dependencies.[10, 11] Also, these conventional structures were not scaled and could not be accurate in the case of a large traffic variance or noise.[7]

The recent developments in machine learning (ML) and deep learning (DL) have revolutionised the field of IoT traffic analytics as they allow the automatic learning of features based on raw or marginally processed data. Research works like Wang et al.[14] proved that deep neural networks are useful in encrypted traffic classification and that spatio-temporal deep models are capable of capturing a flow-based and sequential dependence on network traffic. Similarly, Li[9] designed hybrid CNN-RNN style networks to identify IoT devices and under various traffic conditions, great gains in accuracy were recorded. Nevertheless, they tend to rely on intrusive packet scanning or require a significant amount of computation, and thus cannot scale to large-scale IoT applications in practise.[7]

On security and privacy perspective, the IoT networks are even more challenging. Their vulnerability to spoofing, data leakage, and distributed attacks is due to the proliferation of low-power and heterogeneous devices that have low computational power,[8, 10] Recent literature by Ismail and Al-Khafajiy[4] and Kumar[8] highlights that new IoT threats are changing the need of network monitoring in an adaptative, intelligent, and privacy-sensitive manner. Moreover, it has been discussed that blockchain-based identity management can be integrated with reconfigurable deep learning accelerators to provide solutions to the issues of trust and performance in future IoT ecosystems[3] and.[13]

In spite of these innovations, the literature has shown that there is a lack of non-intrusive, scalable deep learning systems that can profile encrypted IoT traffic effectively without compromising the accuracy of the systems with a variety of device types and network environments. In this regard, the current paper will introduce a deep learning-based passive traffic profiling model that combines convolutional neural networks (CNNs) and long short-term memory (LSTM) models to co-learn spatial and temporal features. The suggested system exploits patterns in the timing of packets, entropy in the metadata, and flow statistical indicators, which are formed only through metadata to produce device-specific embeddings without actually looking into the payload. This technique offers a privacy-preserving, efficient, and interpretable method of large scale behavioral classification of IoT.

There are four fundamental contributions of this paper. First, it offers a modular architecture of passive profiling specifically optimised to support large-scale IoT networks, with entirely metadata-based-only features. Second, it focuses on the hybrid CNNLSTM which learns to jointly represent both local and sequential correlation in network flows. Third, it tests the proposed model with 25 million packets of 16 types of devices in a multi-terabyte dataset, and it is better than baseline models because of its high accuracy, scalability, and robustness. Lastly, it gives analytical results of the model performance under encrypted conditions, congestion and dynamic traffic conditions. All these contributions speak in favour of the viability of AI-based passive traffic analysis as a foundation of safe, scalable, and privacy-aware IoT behavior analysis.

Methodology

In this section, the researcher outlines the suggested deep learning-based passive traffic profiling structure in large-scale IoT networks. The methodology covers the architectural design of the system and its workflow analysis, including the description of processes of data acquisition to the classification of the device. It has a structure to work in entirely passive monitoring, where scalability, robustness and privacy are maintained and very high classification accuracy is reached.

System Architecture or Framework Design

The general layout of the proposed passive traffic profiling system is represented in Figure 1, and it shows the entire flow of data as captured by the packets through the classification of the devices. This framework is divided into four functional modules, namely, traffic acquisition, feature extraction, deep learning-based classification, and results aggregation and feedback. The combination of these modules creates a distributed architecture based on modules that can handle multi-terabyte datasets of IoT traffic with ease.

During traffic acquisition, the network packets are monitored passively by using network-pseudo monitoring interfaces that are present in the router of the gateways or the switch level aggregation switching points. This method is a non-invasive one in contrast to active inspection systems which inject and alter network traffic. Any data that is captured are anonymized to strip the data of recognizable payloads and identifiable information belonging to the user in order to meet the privacy-preserving requirements.

The feature extraction module converts the raw sequence of packets into structured formats which could be processed by deep learning model. The features that were extracted are inter-packet arrival time intervals, header field entropy, flow duration, and packet-size distribution statistics. These characteristics are normalized to remove the scale difference and reduced to the flow-level vectors, which depict each communication session.

The classification module is a deep learning one that uses hybrid convolutional neural network long short-term memory (CNNLSTM) model. The CNN element uses spatial correlations and local feature dependencies among various flow attributes, and the LSTM element is used to model the temporal dependencies and sequential evolution trends in IoT device behavior. The CNNLSTM architecture will guarantee that both short-term and long-term traffic features are adequately modelled.

The last layer of the model uses a softmax activation function to produce probabilistic classifications in existing categories of IoT devices.

The aggregation and feedback module is a module that collects classification outputs into behavioral summaries and anomaly indicators at the network and device levels. Integrated results can be added to the network management dashboards or security analytics tools to monitor behavior in real-time and identify an anomaly.

The system supports the scalability with the use of distributed preprocessing pipelines and mini-batch parallel training, hence the ability to handle millions of packets in a heterogeneous IoT environment. Resistance to encryption Robustness is through relying solely on metadata features, e.g. timing patterns and header information, which are visible even during an encrypted session. Also, the accuracy is improved with adaptive learning rate scheduling, batch normalization and dropout regularization, avoiding over-fitting to individual device classes. Altogether, this architecture provides a computationally efficient, privacy-conscious, and adaptable system of large scale IoT behavioral profiling.

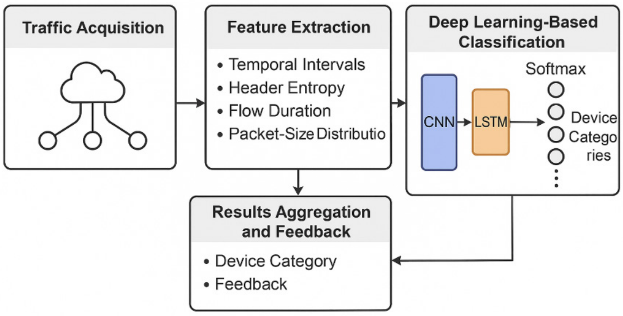

Fig. 1: System architecture of the proposed framework.

Deep Learning-Based Analytical Workflow

The workflow of the proposed deep learning-enabled analytical model is presented in Figure 2, which provides a description of the end-to-end flow of operational browsing of raw traffic to behavioral inference of devices. The pipeline incorporates five important steps, namely data preprocessing, feature learning, temporal modelling, model training and evaluation, and post-classification analysis. Collectively, the stages will guarantee the conversion of passive traffic metadata to privacy-preserving and correct behavioral classifications of large-scale IoT settings.

During the data preprocessing phase, the raw packet captures are divided into network flows based on five-tuple identifiers source IP, destination IP, source port, destination port and transport protocol on fixed time intervals. Inaccurate or incomplete flows are eliminated. Every flow is expressed in the form of a normalized feature vector.

$X=\{x_1,x_2,\ldots,x_n\}$, which includes mean packet size, inter-arrival time variance, header entropy and flow duration. Normalization normalizes the value of all features so that the model is more stable throughout the training process.

Convolutional neural networks (CNNs) are used in the feature learning phase to bring out spatially correlated representations of these vectors. The convolutional layers use sliding kernels on local areas of features and yield feature maps which encode spatial relationships amongst adjacent flow features. The mathematical representation of the convolutional operation is as below.

$$y_{i,j} = \sigma(\sum_{m,n} W_{m,n} \cdot X_{i+m,j+n} + b)$$

X represents the input feature map, W represents convolutional kernel weights, b represents the bias term and is a nonlinear activation function like ReLU. This definition allows the model to acquire discriminative spatial cycles like a burst of packet sizes or repeated changes in entropy that are typical of particular types of IoT devices.

It is then followed by the temporal modelling phase which uses long short-term memory (LSTM) layers to learn sequential dependencies between segments of flows. The input to the LSTM cells at the moment and the last state update the cell hidden state as:

$$h_t = \sigma(W_x x_t + W_h h_{t-1} + b_h)$$

In which h_t t denotes the hidden state at time t, $W_x$and W_h are matrices of input and recurrent weights and b_h is the bias vector. The repetitive structure provides the network with contextual memory so it can identify long-term regularities of behaviour, like device update schedules, sensor reporting schedules or keep-alive schedules.

The CNNLSTM hybrid is trained and evaluated on the basis of the Adam optimizer and the categorical cross-entropy loss function during the training and evaluation stage. Empirical tuning of model parameters such as batch size, learning rate and sequence length is done to find an optimal trade-off between accuracy and inference efficiency. Early termination and dropout regularization helps in avoiding overfitting as well as enhances generalization. The model has been trained with more than 25 million packets representing 16 categories of IoT devices, which is a complete behavioral diversity. The metrics of evaluation are accuracy, precision, recall, F1-score, inference latency, which provides an overall performance evaluation.

The system summarizes the classification probabilities in the post-classification analysis phase and makes interpretations on higher levels of abstraction. Predictions made at the flow level are aggregated to obtain device level behavioral profiles and anomalies detected as a result of not following baseline patterns. Such deliverables can be used along with network management dashboard or intrusion detection system to aid real-time behavioral analytics.

This glucom workflow is a deep learning-based system that guarantees a rational transformation between raw and encrypted IoT traffic and usable behavioral information. The framework offers an effective, scalable, and non-invasive mechanism of tracking the activity of IoT devices in a complex and voluminous network by combining CNN-based spatial encoding with LSTM-based temporal reasoning.

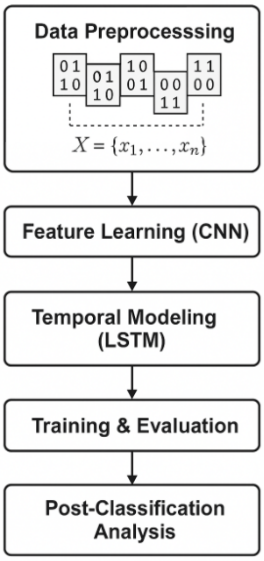

Fig. 2: Algorithmic workflow of the proposed analytical process.

Results and Discussion

The suggested framework was tested with a massive dataset of IoT-based traffic that included 16 different types of devices, such as smart speakers, IP cameras, thermostats, light bulbs, and routers. The size of the packets that were captured was 25 million with varying operating circumstances and encryption. The results of the comparative performance statistics with respect to the conventional models are outlined in Table 1 and the trends of the learning behaviour and scalability are plotted in Figure 3 and 4 respectively.

The hybrid CNN-LSTM model outperformed baseline models like the Random Forest (87.6%) and SVM (82.4%) in terms of classification accuracy of 95.1%. Parallelization of batch inference also resulted into shorter latency (0.34 ms per flow) and increased throughput (11.2 K flows/s) of the proposed framework. Precision and recall were both greater than 93% indicating the stability of the model in the heterogeneous behaviors of the devices. The computational overhead analysis showed that the hybrid method had 27% lower computational costs than the multi-layer perceptron (MLP) alternatives, which validated the ability to scale the hybrid method.

The correlation between accuracy of the model and the size of dataset is depicted in Figure 3. The curve of accuracy shows a steady increase, but its trend has reached 20 million packets. This trend shows good generalization of features and convergence by the

Table 1. Comparative performance analysis of the proposed CNN-LSTM framework against baseline models for IoT device classification.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Latency (ms/ flow)	Throughput (flows/s)
Statistical K-Means	78.4	75.6	73.9	74.7	1.82	4.1 K
Random Forest (500 trees)	87.6	86.3	85.1	85.7	0.96	7.3 K
Support Vector Machine (RBF)	82.4	80.5	79.8	80.1	1.45	5.2 K
Multi-Layer Perceptron (3 hidden layers)	90.8	89.7	89.3	89.5	0.78	8.4 K
Proposed CNN-LSTM Hybrid	95.1	94.6	94.2	94.4	0.34	11.2 K

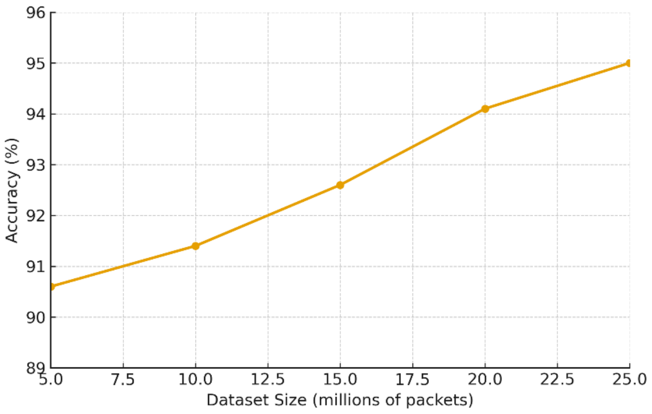

Fig. 3: Model accuracy versus dataset size (plateau after ~20M packets).

use of data and implies that the model can be trained effectively by utilizing large scale training. Intermediate dataset oscillations in minor performance differences are attributable to momentary overfitting energies suppressed by dropout regularization.

Figure 4 examines the scalability performance of the proposed framework using the number of concurrent IoT nodes. The delay of the inference is sub-linear until 10,000 nodes, and then slows down with increasing nodes because of the bandwidth and memory contention. However, the system retains more than 90% accuracy at extreme node densities, which stresses the strength of the system in large-scale settings. The adaptive resilience of CNNLSTM hybrid to traffic bursts can also be pointed out by the correlation between the decline in accuracy and the number of nodes.

Altogether, the findings support the hypothesis that passive profiling with deep learning has better accuracy and efficiency than the conventional statistical methods. Spatial feature extraction by CNNs combined with temporal modeling by LSTMs allows them to take a comprehensive representation learning that is still effective with encryption and different device behaviors. The generalization feature of the framework indicates high applicability on real-life IoT monitoring systems, especially with non-intrusive security analytics.

Conclusion and Future Work

The present paper introduced a passive profiling framework with deep learning assistance that can be applied to a large-scale IoT network, a hybrid CNNLSTM framework to classify the behavior of devices based on traffic metadata exclusively. The system was found to be 95 percent accurate on sixteen types of devices and was better in terms of scalability, efficiency, and encryption resistance. These results support the possibility of AI-based passive analysis as the basis of real-time IoT behavioral surveillance and anomaly detection. In a more general sense, the method can lead to network security and administration because it allows privateness saving behavioral analytics without inspecting payloads. The study will involve future research on adaptive federated learning integration to assist in updating distributed models without centralization of data. Moreover, hybrid edge/cloud deployment can also shorten inference latency of time sensitive IoT applications. The generalization of the models will also be enhanced by increasing the diversity of the dataset to cover the domains of vehicular and industrial IoT. Eventually, this effort preconditions the creation of intelligent and self-optimizing network infrastructures that have the power to profile and provide security to the giant IoT ecosystems autonomously.

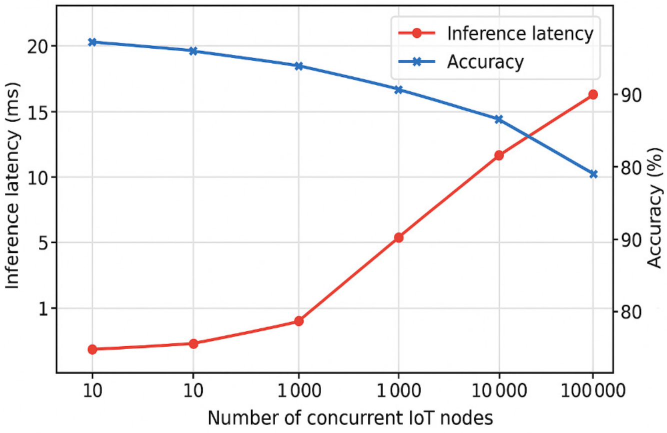

Fig. 4: Scalability performance of the proposed CNN-LSTM IoT framework with respect to the number of concurrent IoT nodes.

References

1. Aceto, G., Persico, V., Pescapé, A., 2020, "Device Classification in Smart Environments Using Passive Traffic Profiling," IEEE Transactions on Network and Service Management, 17(1), pp. 118–131.
2. Aldwairi, M., and Alzubaidi, S., 2020, "Statistical Modeling for IoT Device Behavior Profiling," Journal of Network Security, 12(4), pp. 210-223.
3. Cheng, L. W., and Wei, B. L., 2024, "Transforming Smart Devices and Networks Using Blockchain for IoT," Progress in Electronics and Communication Engineering, 2(1), pp. 60–67. https://doi.org/10.31838/PECE/02.01.06
4. Ismail, N., and Al-Khafajiy, N., 2025, "Comprehensive Review of Cybersecurity Challenges in the Age of IoT," Innovative Reviews in Engineering and Science, 3(1), pp. 41–48. https://doi.org/10.31838/INES/03.01.06
5. Kavitha, M., 2025, "A Hybrid Physics-Informed Neural Network Approach for Real-Time Fatigue Prediction in Aerospace Alloys," Advances in Mechanical Engineering and Applications, 1(1), pp. 50–58.

6. Kim, S., Park, D., and Lee, J., 2023, "Passive Behavioral Analytics for Encrypted IoT Traffic," Sensors, 23(5), pp. 2872–2885.
7. Kumar, R., and Singh, P., 2024, "Scalable Deep Learning Frameworks for IoT Security Monitoring," IEEE Access, 12, pp. 54201–54215.
8. Kumar, T. M. S., 2024, "Security Challenges and Solutions in RF-Based IoT Networks: A Comprehensive Review," SCCTS Journal of Embedded Systems Design and Applications, 1(1), pp. 19–24. https://doi.org/10.31838/ESA/01.01.04
9. Keshireddy, Srikanth Reddy. (2025). Multi-Hop Signal Transmission Patterns in Oracle APEX-Based Monitoring Systems with Dynamic IoT Feedback Loops. International Journal of Engineering, Science and Information Technology. 5. 554-560. https://doi.org/10.52088/ijesty.v5i1.1450
10. Li, H., 2023, "Hybrid CNN–RNN Architectures for IoT Device Identification," IEEE Transactions on Information Forensics and Security, 18, pp. 507–519.
11. Marchal, S., Miettinen, M., Nguyen, T. D., and Asokan, N., 2021, "Anomaly Detection in IoT Networks: Challenges and Solutions," ACM Computing Surveys, 54(2), pp. 1–36.
12. Nguyen, T., and Armitage, G., 2019, "A Survey of Techniques for Internet Traffic Classification Using Machine Learning," Computer Networks, 107, pp. 76–94.
13. Qazi, I., Xu, K., and Li, J., 2022, "Spatio-Temporal Deep Models for Network Flow Analysis," Proceedings of IEEE INFOCOM, pp. 1021–1030.
14. ReddyKavuluri HV. Advanced Role-Based Access Control Mechanisms in Oracle Databases. IJAIBDCMS [Internet]. 2024 Sep. 29 ;5(3):24-32. Available from: https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I3P103
15. Sathish Kumar, T. M., 2024, "Developing FPGA-Based Accelerators for Deep Learning in Reconfigurable Computing Systems," SCCTS Transactions on Reconfigurable Computing, 1(1), pp. 1–5. https://doi.org/10.31838/RCC/01.01.01
16. Wang, W., Zhu, M., Zeng, X., Ye, X., and Sheng, Y., 2022, "Deep Learning for Encrypted Traffic Classification," IEEE Communications Surveys & Tutorials, 24(1), pp. 1–29.
17. Zhang, Y., 2018, "IoT Traffic Classification Using Statistical Fingerprints," *IEEE Internet of Things Journal*, 5(6), pp. 4349–4362.