IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/46ICT/2025/4636325/
Vol. 2025 (35), Article ID 4636325

IoT Device Classification Methodes Based on Network Traffic*

Piotr KONTOWICZ, Mariusz GŁĄBOWSKI, Marek FECHNER, Jagoda PIECHOCKA and Michał WEISSENBERG

Poznan University of Technology, Faculty of Computing and Telecommunications,

Piotrowo 3, 60-965 Poznań, Poland

Correspondence should be addressed to: Piotr KONTOWICZ, piotr.kontowicz@put.poznan.pl

* Presented at the 46th IBIMA International Conference, 26-27 November 2025, Ronda, Spain

Abstract

The rapid growth of Internet of Things (IoT) devices in smart cities produces substantial and heterogeneous network traffic, creating significant challenges for security and management. Accurate device classification is crucial for protecting networks against threats, such as malicious devices, and for facilitating efficient resource allocation and management. Although various classification methods have been proposed, a comprehensive review evaluating their effectiveness in addressing the specific challenges of smart cities, such as large-scale device heterogeneity and extensive encrypted traffic, remains lacking.
This paper provides a systematic review of the literature, analyzing and comparing IoT device classification methods that utilize network traffic data. The methodology emphasizes two primary aspects: feature extraction techniques (including packet-level, flow-level, and automated approaches) and the classification algorithms employed (supervised, unsupervised, and deep learning methods).
The analysis shows that packet-based methods achieve high precision but face limitations with encrypted traffic and scalability. In contrast, flow-based and deep learning approaches exhibit greater adaptability. Machine learning (ML) and deep learning (DL) algorithms consistently demonstrate strong performance, with reported accuracy rates reaching up to 99%. The review identifies several critical future research directions for smart city applications, including experimental validation, integration with edge computing, and the development of hybrid classification models.

Keywords: Smart City, IoT devices classification, network traffic features extraction methods, network

Traffic Characteristics

Introduction

For the purpose of this paper, the definition of a smart city proposed by [1] was adopted, which defines a smart city as an environment that uses information and communication technologies (ICT) and related technologies to improve the quality of life of its residents. It highlights the use of Internet of Things (IoT) device networks, such as smart sensors, actuators, and cameras, as means to collect and transmit data to smart systems and services.
Many types of threats, vulnerabilities, and attacks targeting IoT device networks within smart cities have been identified [2]. One category mentioned includes attacks involving the insertion of a malicious device into the network, data manipulation, information duplication, or packet dropping [3].

Review Methodology

This article is a review paper. The methodology involved an analysis of articles concerning IoT device classification methods. The aim was to identify, compare, and evaluate their potential advantages and limitations. The analysis was divided into main parts:

Cite this Article as: Piotr KONTOWICZ, Mariusz GŁĄBOWSKI, Marek FECHNER, Jagoda PIECHOCKA and Michał WEISSENBERG, Vol. 2025 (35) "IoT Device Classification Methodes Based on Network Traffic" Communications of International Proceedings, Vol. 2025 (35), Article ID 4636325,
https://doi.org/10.5171/2025.4636325

1. Analysis of feature extraction methods: examining which network traffic features are used for device classification. This included extraction methods based on packet-level features, flow-level features, and automatic extraction methods.
2. Analysis of classification algorithms: an analysis of the classes of algorithms presented in the literature for the automatic classification of network devices. This covered methods based on MAC/IP addresses, statistical approaches, behavioral profiling techniques, and a broad spectrum of machine learning methods, including supervised, unsupervised, and deep learning algorithms.

The discussed elements were evaluated in the context of the unique characteristics and challenges specific to smart cities. These challenges include the large number and diversity of devices, and high network traffic volume, which poses a performance challenge for the classification algorithms employed.

Comparative Analysis

In accordance with the adopted methodology, this section presents the results of the comparative analysis of IoT device classification methods. The results are presented in the following sections.

Feature extraction

The first stage of the classification process is feature extraction from network traffic. Three main categories are distinguished:

1. Packet-based: These methods rely on the analysis of features derived from individual packets. They include port-based analysis, deep packet inspection (DPI), stochastic packet inspection (SPI), and approaches utilizing machine learning (ML). These methods are well-researched but may struggle with encrypted traffic. These methods are summarized in Table 1.
2. Flow-based: This approach focuses on analyzing a set of packets grouped into a 'flow,' which can be defined based on various criteria, such as a fixed number of packets or a time window. These methods are simple to implement and often rely on statistical methods, correlations, or analysis. These methods are summarized in Table 2.
3. Automatic extraction: These methods utilize deep learning techniques. They eliminate the need for manual feature selection, as features are automatically selected from the collected data. This approach performs well in high-traffic environments.

The choice between packet-level and flow-level features is critical. Packet-level features offer high granularity, but their analysis can be computationally expensive. Flow-level features generally offer a better balance between granularity and computational efficiency.

Table1: Packet based methods

Approach	Reference	Description	Potential advantages	Potential weaknesses
Port based	4, 5, 6, 7, 8	Analysis based on predefined ports used by devices and services	Well-researched, easy to implement, requires access to the header only	The effectiveness of analyzing traffic from devices using non-standard ports may be limited
Deep packet inspection (DPI)	4, 6, 7, 8	Packet contents are analyzed for signatures typical of a specific device or service	Well-researched, can be used for devices using non-standard ports	Difficult to implement for encrypted data
Stochastic Packet Inspection (SPI)	8, 9	Statistical information from packet payloads is analyzed for classes typical of a specific device or service	Can be used for encrypted traffic	Computationally expensive, may require class updates for new devices
Packet-based approaches	4, 5, 6, 7, 8	Selected data from the packet header	Depending on the selected features, it	May require high computational costs

utilizing Machine Learning (ML)		or payload is analyzed using ML models	may be effective for encrypted traffic	depending on the features and algorithms used

Table 2. Flow-based extraction methods

Approach	Reference	Description	Potential advantages	Potential weaknesses
Based on statistical properties	4, 19, 6, 7, 10, 8	Based on the statistical properties of the flow	Potentially useful for distinguishing device types, can handle encrypted traffic	Rely solely on statistics, not always easily extendable
Based on flows correlations	8	Based on flow correlations	Allow handling of encrypted traffic	Computationally demanding
Based on behavior	7, 8	Based on network traffic received by hosts	Not computationally demanding, effectively handle encrypted traffic	Classification results may be ambiguous

The choice of method depends on the requirements and constraints of the classification task. Packet-based methods offer simplicity but may struggle with encrypted traffic. Flow-based techniques provide a balance between effectiveness and performance.

In the context of a smart city environment, which is characterized by a large number of devices and high volumes of network traffic, packet-level feature extraction can be difficult to implement due to computational complexity. Relying solely on basic packet features may be insufficient because of the risk of spoofing by malicious nodes. Payload analysis is often not feasible due to encryption.

Classification Algorithms

Extracted features are used by various classes of algorithms. The simplest approach is to classify devices based on MAC or IP addresses, but this is limited due to the potential for spoofing and the lack of standardization. Another approach involves statistical methods, although these are often restricted to specific applications or protocols.

With the advancement of machine learning and deep learning techniques, these methods are gaining popularity. They include both traditional supervised and unsupervised algorithms.

Table 3. Methods by Algorithm Used

Approach	Reference	Algorithms	Accuracy
Supervised ML	11	Naive Bayes Multinomial (stage 1), Random Forest (RF)	99%
	12	Random Forest (RF), k-Nearest Neighbors (k-NN), Gaussian i Bernoulli Naive Bayes.	70.55%
	13	RF, k-NN, MLP, Naive Bayes (NB), logistic regression, SVM	99%
Unsupervised ML	14	RF, XGB, Decision Tree (DT), NB, SVM (stage 1). XGB, RF, DT	99%
	15	K-Means, thresholding, k-NN	89%

	16	K-Means clustering	99%
Deep Learning	17	Convolutional Neural Network (CNN)	99%
	18	Graph Contrastive Neural Network (GCNN)	87% - 99% (depend on the model)
	19	CNN	98%

Machine learning–based approaches offer greater capabilities. The results presented in Table 3 indicate their high effectiveness:

1. Supervised methods: The main challenge is the need for high-quality labeled training data, which is time-consuming to prepare. In addition, models need to be periodically updated.
2. Unsupervised methods: These reduce reliance on labeled data, but selecting the right features and achieving sufficient classification granularity remain challenges.
3. Deep learning methods appear particularly promising, as they are capable of automatic feature extraction and are well suited to the complexity and dynamic nature of network traffic.

Discussion

The choice of an appropriate classification approach largely depends on the characteristics of the environment. Implementing these methods in a real-world smart city environment must take into account the general assumptions about the network environment:

1. a large number of diverse IoT devices
2. high volume of network traffic
3. complex traffic patterns
4. presence of malicious nodes

The main challenge lies in the data. Developing a deployable solution requires setting up a test environment that includes a sufficiently large and diverse set of IoT devices.

Conclusions

This article reviewed selected IoT device classification methods based on network traffic characteristics. The analysis focused on the traffic features used and the classes of algorithms applied.

In terms of feature selection, the study discussed packet-based and flow-based approaches, further categorizing them into manual and automatic feature extraction methods.

Based on the challenges identified in this work, several directions for future research are proposed:

1. Experimental validation: conducting experimental comparisons of the discussed methods to evaluate their performance under realistic conditions.
2. Integration with edge computing: exploring the use of edge computing mechanisms to reduce latency and improve the scalability of real-time classification methods.
3. Alternative classification mechanisms: investigating approaches to device classification beyond network traffic analysis, potentially incorporating hardware-level features.

These research directions aim to address gaps in current methodologies and contribute to the development of robust, scalable, and efficient IoT device classification systems suitable for deployment in smart cities.

Acknowledgements

This research is funded by the Ministry of Education and Science, Grant 0313/SBAD/1314.

Bibliography

- Silva, B.N., Khan, M., Han, K.: Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. Sustainable Cities and Society 38, 697–713 (2018). DOI 10.1016/j.scs.2018.01.053

- Fei, W., Ohno, H., Sampalli, S.: A Systematic Review of IoT Security: Research Potential, Challenges, and Future Directions. ACM Computing Surveys 56(5), 1–40 (2024). DOI 10.1145/3625094
- Li, B., Lin, Y., Khan, I.A.: Self-Supervised Learning IoT Device Features With Graph Contrastive Neural Network for Device Classification in Social Internet of Things. IEEE Transactions on Network and Service Management 20(4), 4255–4267 (2023). DOI 10.1109/TNSM.2023.3252806
- Abbasi, M., Shahraki, A., Taherkordi, A.: Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. Computer Communications 170, 19–41 (2021). DOI 10.1016/j.comcom.2021.01.021
- Jmila, H., Blanc, G., Shahid, M.R., Lazrag, M.: A Survey of Smart Home IoT Device Classification Using Machine Learning-Based Network Traffic Analysis. IEEE Access 10, 97117–97141 (2022). DOI 10.1109/ACCESS.2022.3205023
- Shafiq, M., Tian, Z., Bashir, A.K., Jolfaei, A., Yu, X.: Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. Sustainable Cities and Society 60, 102177 (2020). DOI 10.1016/j.scs.2020.102177
- Tahaei, H., Afifi, F., Asemi, A., Zaki, F., Anuar, N.B.: The rise of traffic classification in IoT networks: A survey. Journal of Network and Computer Applications 154, 102538 (2020). DOI 10.1016/j.jnca.2020.102538
- Zhao, J., Jing, X., Yan, Z., Pedrycz, W.: Network traffic classification for data fusion: A survey. Information Fusion 72, 22–47 (2021). DOI 10.1016/j.inffus.2021.02.009
- Mantia, G.L., Rossi, D., Finamore, A., Mellia, M., Meo, 10.1109/ICC.2010.5502280
- Zhang, J., Chen, X., Xiang, Y., Zhou, W., Wu, J.: Robust network traffic classification. IEEE/ACM Transactions on Networking 23, 1257–1270 (2015). DOI 10.1109/TNET.2014.2320577
- Sivanathan, A., Gharakheili, H.H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., Sivaraman, V.: Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. IEEE Transactions on Mobile Computing 18(8), 1745–1759 (2019). DOI 10.1109/TMC.2018.2866249
- Thangavelu, V., Divakaran, D.M., Sairam, R., Bhunia, S.S., Gurusamy, M.: DEFT: A Distributed IoT Fingerprinting Technique. IEEE Internet of Things Journal 6(1), 940–952 (2019). DOI 10.1109/JIOT.2018.2865604
- Ganesan, E., Hwang, I.S., Liem, A.T., Ab-Rahman, M.S.: SDN-Enabled FiWi-IoT Smart Environment Network Traffic Classification Using Supervised ML Models. Photonics 8(6), 201 (2021). DOI 10.3390/photonics8060201. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute
- Zeng, T., Ye, K., Lou, F., Chang, Y., Yin, M., Hu, T.: Dual-IoTID: A Session-Based Dual IoT Device Identification Model. Applied Sciences 14(11), 4741 (2024). DOI 10.3390/app14114741. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute
- Koball, C., Rimal, B.P., Wang, Y., Salmen, T., Ford, C.: IoT Device Identification Using Unsupervised Machine Learning. Information 14(6), 320 (2023). DOI 10.3390/info14060320. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute
- Wang, H.: A method of classifying IoT devices based on attack sensitivity. Journal of Information Security and Applications (2024)
- Liu, X., Han, Y., Du, Y.: IoT Device Identification Using Directional Packet Length Sequences and 1D-CNN. Sensors 22(21), 8337 (2022). DOI 10.3390/s22218337. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute
- Li, B., Lin, Y., Khan, I.A.: Self-Supervised Learning IoT Device Features With Graph Contrastive Neural Network for Device Classification in Social Internet of Things. IEEE Transactions on Network and Service Management 20(4), 4255–4267 (2023). DOI 10.1109/TNSM.2023.3252806
- Yin, S., Zhang, W., Feng, Y., Xiang, Y., Liu, Y.: Automatic IoT device identification: a deep learning based approach using graphic traffic characteristics. Telecommunication Systems 83(2), 101–114 (2023). DOI 10.1007/s11235-023-01009-1