Machine Learning Based Classification and Identification of IoT devices

PARAM BHARATBHAI PATEL (234101062) AND PRATIK MAHENDRA KAMBLE(234101039)
UNDER THE SUPERVISION OF DR. TAMARAPALLI VENKATESH

INTRODUCTION

- The Internet of Things (IoT) represents a transformative network of physical objects or devices, ranging from household appliances like smart bulbs, thermostats, and security cameras to industrial tools, medical monitors, and environmental sensors, all embedded with sensors, software, and other technologies to connect and exchange data.
- The classification and identification of IoT devices involve categorizing them based on their characteristics, behavior, or functionality, while uniquely recognizing and distinguishing individual devices within a network to enhance network management and security.
- IoT device identification ensures security and efficient management by detecting rogue entities, supporting anomaly detection, and enabling timely interventions with lightweight methods, crucial for diverse and heterogeneous networks.
- Challenges in IoT device identification includes Device Heterogeneity , Dynamic Network Environments, Scalibility.

MOTIVATION

- The integration of IoT devices has enhanced efficiency but introduced critical security risks, making accurate identification essential for mitigating threats.
- Device identification isolates compromised devices, prevents spoofing, and detects anomalies to reduce security breaches and unauthorized activities.
- It supports network traffic analysis to identify suspicious patterns like DDoS attacks or data transfers, ensuring secure and trustworthy IoT ecosystems.

OBJECTIVE

- To develop or refine a Machine Learning-based approach for the classification and identification of IoT devices, emphasizing the creation of a generalizable method that performs reliably across diverse environments and network contexts.

LITERATURE SURVEY

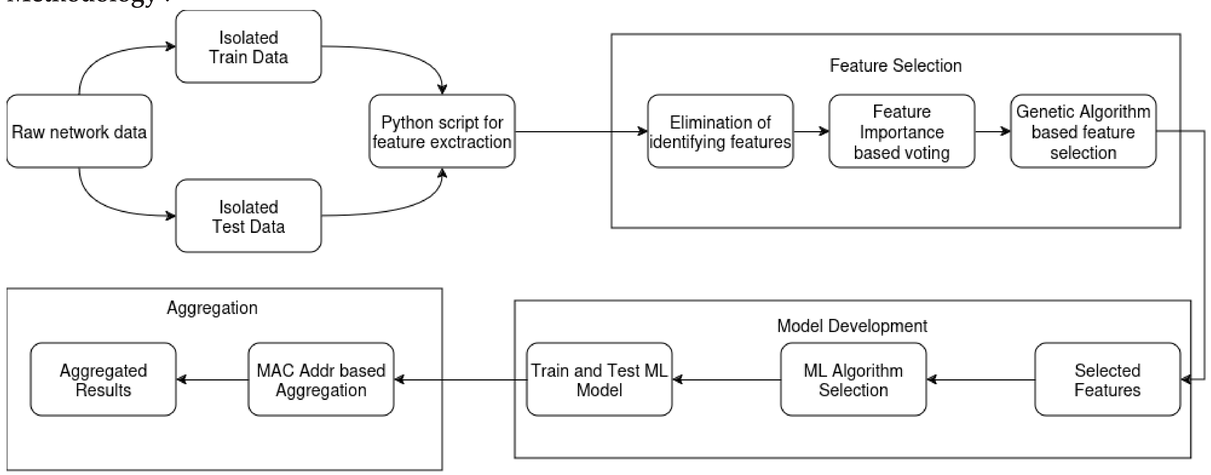
- Traditional IoT identification methods[2],[3],[4] relying on static identifiers (e.g., MAC addresses, DNS queries) and rule-based systems were effective in controlled environments but struggled with spoofing, evolving behaviors, and non-IP protocols. Machine learning introduced behavior-based approaches, offering adaptability, scalability, and accuracy, addressing these limitations in dynamic IoT ecosystems.
- IoT Sentinel[3] introduced protocol-based features (e.g., DNS, port numbers) for fingerprinting devices during setup, using Random Forest for classification. Effective in controlled environments but struggled with spoofing and non-IP protocols.
- Sivanathan et al[5] collected long-term traffic data from 28 IoT devices, extracting statistical features like flow volume and DNS queries. Used multi-stage machine learning models but faced challenges with environment-specific dependencies.
- SysID[4] simplified feature extraction using single-packet headers (e.g., TCP sequence numbers, port numbers) for lightweight classification. Optimized for resource-constrained environments but limited in handling dynamic behaviors.
- IoTDevID[1] proposed a behavior-based fingerprinting method, leveraging packet-level features and Genetic Algorithms for feature selection. Achieved high generalizability and accuracy across datasets with robust aggregation techniques.

IMPLEMENTATION AND VALIDATION

The implementation focuses on implementation of [1]'s framework on Aalto and UNSW dataset and then validation of the same on CIC IoT 2022 dataset.

Datasets used :
Aalto dataset : 27 device classes with IP devices
UNSW dataset : 32 device classes with both IP/Non-IP devices.
CIC 2022 dataset : 40 device classes with IP/Non-IP under normal and attack states.

Methodology :



Evaluation Results :
For Aalto dataset F1 score of **0.8551** was noted.(0.7282 for individual and 0.8099 for aggregated)
For UNSW dataset F1 score of **0.9316** was noted.(0.8281 for individual and 0.9205 for aggregated)
For CIC dataset F1 score of **0.925** for IP-only devices and **0.788** for (IP+Non-IP) devices was noted.

CONCLUSIONS

- While achieving high accuracy across diverse datasets, it highlighted the robustness of protocol-independent features and mixed classification methods in addressing challenges across varied network environments.
- However data imbalance ,non-IP protocols,devices with similar hardware or software which exhibit similar behavior continue to pose challenges,affecting the model's ability to generalize and accurately classify devices.

FUTURE WORK

- Employ meta learning techniques like stacking with base models to address the challenges and enhance the performance of the model by combining the strength of base models.
- Future work will involve experimenting with different datasets to assess the model's generalization across various IoT environments and network traffic patterns.
- Exploring methods like Recursive Feature Elimination, clustering-based selection, and PCA to identify relevant features and reduce model complexity.

REFERENCES

[1] Kahraman Kostas, Mike Just, and Michael A Lones. Iotdevid: A behavior-based device identification method for the iot. *IEEE Internet of Things Journal*, 9(23):23741–23749, 2022.

[2] Arunan Sivanathan, Daniel Sherratt, Hassan Habibi Gharakheili, Adam Radford, Chamith Wijenayake, Arun Vishwanath, and Vijay Sivaraman. Characterizing and classifying iot traffic in smart cities and campuses. In *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 559–564. IEEE, 2017.

[3] Markus Miettinen, Samuel Marchal, Ibbad Hafeez, Nadarajah Asokan, Ahmad-Reza Sadeghi, and Sasu Tarkoma. Iot sentinel: Automated device-type identification for security enforcement in iot. In *2017 IEEE 37th international conference on distributed computing systems (ICDCS)*, pages 2177–2184. IEEE, 2017.

[4] Ahmet Aksoy and Mehmet Hadi Gunes. Automated iot device identification using network traffic. In *ICC 2019-2019 IEEE international conference on communications (ICC)*, pages 1–7. IEEE, 2019.

[5] Arunan Sivanathan, Hassan Habibi Gharakheili, Franco Loi, Adam Radford, Chamith Wijenayake, Arun Vishwanath, and Vijay Sivaraman. Classifying iot devices in smart environments using network traffic characteristics. *IEEE Transactions on Mobile Computing*, 18(8):1745–1759, 2018.