Available online: www.ijsrmes.com

International Journal for Scientific Research in Modern Engineering and Science

Published in: Vol. 1, Issue 1, pp. 49-57, March-April 2026

REVIEW OF MACHINE LEARNING METHODS FOR DETECTING AND CLASSIFYING IOT AND NON-IOT DEVICES

Divya Shukla [1]*, Mohan Rao Mamdikar [2]

[12]PDepartment of Department of Computer Science & Engineering, Vishawavidyalaya Engineering College, Ambikapur

Abstract

One of the most important technologies in the field of IoT is the abstraction of IoT devices. The Internet of Things (IOT) refers to a global network of interconnected devices. It combines widespread communications, pervasive computing, and ambient intelligence. In this paper, we present a systematic survey of ML technologies for identifying IoT devices and detecting compromised ones. In this report, we have mentioned the proximity-based literature review along with the problem identification of various IoT devices. Later in this paper we have discussed briefly about the device type identification, i.e. whether the device is IoT or Non-IoT (NoT). The results show that IoT and non-IoT devices can be distinguished with greater accuracy, and IoT devices can be classified into the appropriate classes with the required accuracy.

Keywords: *Internet of Things, Security, Machine Learning, Device Type Identification, Machine Classifier, . Device type identification*

* *Corresponding author*

1. INTRODUCTION

IoT, or Internet of Things, refers to the ability to access and manage everyday devices and equipment via the Internet. The Internet of Things (IoT) refers to a network of interconnected computing devices, mechanical and digital systems, objects, animals, or people, each equipped with unique identifiers and the capability to exchange data over a networkwithout needing direct human involvement or human-computer interaction. An increasingly dynamic domain, the Internet of Things (IoT) entails the linking and communication of intelligent devices, that is, IoT devices equipped with embedded sensors, onboard data processing, and communication features enable automated services that would otherwise be unattainable [1].

By around 2020, it was anticipated that trillions of IoT devices connected to networks would be part of the global infrastructure [2]. The IoT is increasingly woven into daily life, supporting a wide range of new services and applications in urban and community settingsparticularly in healthcare. Transportation, energy/utilities, and other sectors. This research is currently centered on the threats that IoT devices present

to major corporate organizations. Enterprise IoT security depends not only on the organization's practices but also on the actions of its employees. IoT devices that organizations deploy on their own can support a range of enterprise applications. For example, smart cameras and smoke detectors boost security, while smart thermostats, light bulbs, and outlets help cut down on energy useand similar devices follow the same pattern. Therefore, precautions should be taken to ensure that these Web-connected devices do not inadvertently widen the organization's exposure to cyberattacks. Smart TVs commonly found in conference rooms serve as a solid example.

IoT devices, or Internet of Things devices, are non-standard gadgets that connect wirelessly to a network, enabling communication with each other and the exchange of data. IoT devices are expanding internet connectivity to go beyond traditional devices like smartphones, laptops, tablets, and desktops. Integrating these devices with technology allows us to communicate and interact through networks, and they can also be monitored and controlled remotely. A wide range of IoT devices is available, compliant with the IEEE 802.15.4 standard. These devices include wireless motes, attachable sensor boards, and interface boards, all valuable tools for researchers and developers. IoT devices encompass computers, software, wireless sensors, and actuators. These IoT devices are linked via the internet, allowing data to be exchanged between objects or people automaticallywithout any human involvement.

2. LITERATURE REVIEW

In their recent study, the authors developed a system that uses machine learning to automatically classify IoT devices and restricts communication from vulnerable ones to reduce potential harm to the network [3]. However, the protocol depends on MAC addresses to identify new devices attempting to authenticate with the network, a method that can be faked. Additionally, the work fails to account for scenarios in which a previously authenticated device is compromised or requires re-authentication each time it reconnects to the network.

Bremler et al. Prioritize differentiating between IoT and non-IoT NoT) devices using machine learning classifiers to apply appropriate security policies. to the device [4]. Nevertheless, among several shortcomings, the most significant limitation lies in their classification method, which can only determine whether a device is IoT or non-IoT. Failing to resolve this issue could enable an attacker to exploit a vulnerable device to carry out actions reserved for another type of devicefor instance, using a compromised smart camera to unlock a door or open a garage.

In a separate study, the authors seek to automatically identify potentially malicious IoT devices within a network using the Random Forest classifier and whitelist those deemed trustworthy [5]. A major flaw in this method is that if a previously whitelisted vulnerable device is compromised by an adversary, the attacker could gain full access to the networkbecause the system lacks the ability to detect changes in the device's behavior.

In their other work, the authors suggest IoT security solutions leveraging machine learning methodssuch as reinforcement learning, unsupervised learning, and supervised learningto enhance resistance against spoofing, improve detection capabilities, and authenticate devices to safeguard data privacy [6]. Their work aims to detect

attacks using various machine learning methods; yet, each attack is flagged by a separate model, potentially consuming significant resources like memory and processing time. Moreover, in our view, if models are trained to spot attacks by recognizing specific patterns, real-world attacks may deviate from those patterns, potentially allowing some attacks to go unnoticed. A survey on machine learning-based classification methods for detecting and identifying legitimate and unauthorized IoT devices, offering security solutions where traditional cryptographic protocols are not feasible [7]. Nevertheless, the paper has limitations, as it does not present a complete and formal authentication protocol capable of integrating the ML techniques described to safeguard the network against specific real-world attacks.

3. PROXIMITY BASED LITERATURE REVIEW

The proximity-based solution presented in [8] relies on multiple physical actions carried out by the user. While it offers a meaningful advancement in IoT device authentication, it demands considerable effort from the user. Moreover, the authors focus solely on initial authentication and fail to cover the necessary steps required when a previously authenticated but vulnerable device is compromised.

In their other work, the authors suggest a user authentication method based on proximity for voice-enabled IoT devices [9]. The work introduces a voice-based method for estimating distance to authenticate IoT devices, leveraging multiple advanced technologies; however, its main limitation is that it applies exclusively to voice-enabled IoT devices.

Shafagh and Hithnawi suggest an alternative proximity-based method for authenticating IoT devices, relying exclusively on the wireless communication interface [10]. However, their approach has drawbacks, as it fails to consider the risk of a nearby device being compromised due to its own security vulnerabilities. vulnerabilities that could enable an adversary to carry out attacks such as actuation, network poisoning, and interception of network traffic.

In additional work, the authors suggest a method for device identification that relies on fingerprinting the wireless device's chipset [11]. However, their solution cannot identify a compromised legitimate device.

4. PROBLEM IDENTIFICATION

Machine Learning Based Problem Identification

Authors	Focus on	Techniques /System Used	Problem Identified
[3] M.Miettinen, S.Marchal et al.	Automatic Identification of Iot Device	ML and MAC Address	If a previously authenticated device is compromised, or if the device must be reauthenticated each time it reconnects to the network.

[4]. Bremler et al.	Distinguish between Iot and Non-Iot Devices	ML classifiers	Can only classify a device as IoT or non-IoT, but cannot determine the specific type of IoT device.
[5]Y. Meidan, M. Bohadana et al.	Automatic Detection of Suspicious IoT Devices	Random Forest classifier and white List devices that are classified as trust-worthy	A compromised device that has been whitelisted will allow the attacker unauthorized access to the network, as the system cannot identify changes in its behavior.
[6] L. Xiao et al.	To Identify attacks	ML techniques, including reinforcement learning, unsupervised learning, and supervised learning	These models are trained to identify attacks by recognizing specific pat-terns; however, real-world attacks may differ from those patterns, po-tentially allowing some attacks to go undetected.
[7] Y. Liu, et al.	IoT Device type Classifi-cation	Multiple Classifier	Do not provide a formal authentica-tion protocol.

Proximity Based Problem Identification

Authors	Focus on	Problem Identified
[8] J. Zhang et al.	Based on Physical Activity performed by user	Avoid discussing the measures required if an already au-thenticated vulnerable device is compromised.
[9] N. Z. Gong et al.	Use of voice-powered IoT devices	Applies exclusively to voice-enabled IoT devices.
[10] H. Shafagh et al.	utilizing the wireless communication interface	which might prompt an adversary to carry out attacks like actuation, poisoning the network, or intercepting network traffic.
[11] P. Robyns	device identification based on fingerprint recogni-tion	unable to identify a compromised legitimate device.

5. THREAT MODE OF ROGUE DEVICES IN IOT

This section offers a concise overview of the threat vectors posed by rogue devices in IoT, alongside correspond-ing defenses. We break down the attack sequence and pinpoint the core needs for detecting and identifying IoT devices: confirming the authenticity of legitimate devices, spotting unfamiliar or spoofed devices, and identifying compromised devices exhibiting anomalous behavior. The IoT's cyber infrastructure enables devices of varying capabilities and security weaknesses to exchange information and work together. This scheme promotes a broad,

open system with minimal barriers to entry. Conversely, adversaries can carry out unauthorized activities with significant ease [12]. Typically, adversaries in IoT employ two main types of attack modes: passive attacks and proactive attacks. During a passive attack, adversaries refrain from causing harm or slowing down performance over an extended period. Nevertheless, they passively monitor devices' communication and usage patterns, offering attackers blueprints for future preemptive strikes. If we view passive attackers as spies quietly and peacefully collecting intelligence, proactive attackers go to great lengths to disrupt performance or exploit systems to carry out harmful actions. In real-world attacks, both proactive and passive methods are often used together.

We divided the whole attack chain into five stages, as follows:

1. Penetration
2. Spying:
3. Data analytics:
4. Planning
5. Attack:

In the penetration phase, both passive and proactive attacks are integrated. From the viewpoint of network operators or cybersecurity surveillance agents, if we can stop adversaries from successfully impersonating legitimate devices during the initial stagepenetrationor if we can detect compromised devices in the subsequent stagespying. Network operators and surveillance agents have the capability to dismantle the entire attack chain.

6. LEARNING-ENABLED DEVICE IDENTIFICATION IN IOT

This section examines techniques for identifying devices and their types in IoT systems, with most relying on analyzing network traffic and wireless signal patterns. We begin by examining methods for identifying device types, commonly employed to recognize commercial IoT devices.

A. Device type identification

Although device types aren't directly tied to a device's identity, they remain crucial for effective network management and risk control. Figure 5 shows a brief overview of common IoT devices, while their Physical Layer, Data Link Layer, and overall data transmission characteristics are compared in references [13], [14], and [15], respectively. As shown in Figure 1, WiFi is widely used in smart homes, whereas smart cities tend to favor dependable cellular networks. Device type identification is commonly carried out across network, transport, and application layers and is often implemented within Software Defined Network (SDN) controllers or software routers [16][17].

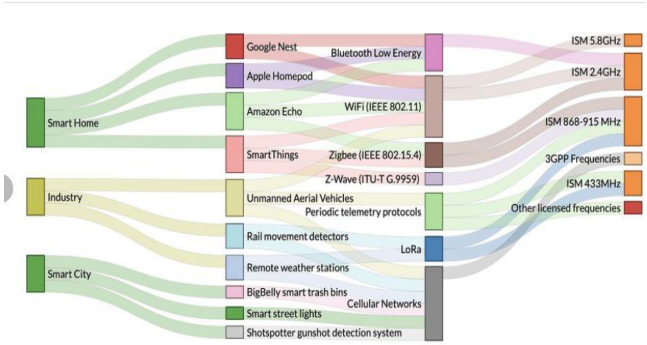

Figure 1. Typical IoT devices and protocols.

The types of devices indicate their capabilities and usage patterns. Figure 2 presents a taxonomy of features used to identify device types.

As shown in Figure 2, remote service remains a widely exploited attack vector for revealing the device typeor even its identity. This is because IoT devices interact with remote service providers via the REST API [18]. Even with encrypted sensitive data, certain distinctive strings in web requests can still be exploited to deduce device types. In [19], the authors demonstrate that a Naive Bayesian classifier can achieve high accuracy in categorizing 28 commercial IoT devices when relying solely on port numbers, domain names, and cipher suites.

While modeling remote service requests from devices shows promise for identifying device types, such approaches might fail when interacting with anonymous service providers. For alleviation, their activity and data flow patterns can be put to use. In [20], the authors suggest that their Random Forest classifier achieves a high accuracy of 95% in distinguishing among 20 IoT devices when combining activity features, network data flows, and remote service requests. In [21], device types are determined according to the regularity of their activities. Initially, the authors apply the Discrete Fourier Transform (DFT) and discrete autocorrelation to identify the dominant periods in protocol-specific activities. They subsequently employ statistical and stability metrics to model device behaviors.

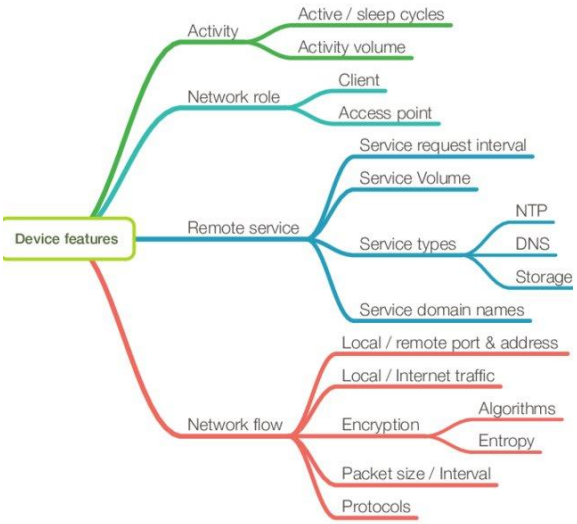

Figure 2. Features for device type identification.

7. MACHINEL EARNING TECHNIQUES FOR IOT DEVICE SECURITY ENHANCEMENT

Machine learning methodssuch as supervised, unsupervised, and reinforcement learninghave been extensively used to enhance network security, including tasks like authentication, access control, anti-jamming offloading, and malware detection.

Supervised learning methodsincluding support vector machines (SVM), naive Bayes, K-nearest neighbor (K-NN), neural networks, deep neural networks (DNN), and random forestscan be applied to classify or regress IoT device network traffic or application traces, enabling the construction of predictive models [22]. For instance, IoT devices can employ SVM to identify network intrusions [22] and spoofing attacks [23], use K-NN for detecting network intrusions [13] and malware [24], and leverage neural networks to spot network intrusions [25] and DoS attacks [26]. IoT devices can employ Naive Bayes for intrusion detection, while random forest classifiers are suitable for malware detection. Devices with adequate computational and memory capacity can leverage DNNs to identify spoofing attacks [27].

8. CONCLUSION

This survey offers a broader perspective on current advanced technologies designed to detect and identify IoT devices. As the Internet of Things (IoT) grows, an increasing number of IoT devices are being deployed across diverse environments, such as businesses, homes, warehouses, and roadways. Ensuring appropriate security for IoT devices is essential, as different IoT devices possess varying characteristics. The literature review uncovered a substantial volume of cited works focusing on IoT device identification and classification. Despite this, most work on small enterprise networks currently relies on static data like port details, MAC addresses, and device model test results. Looking ahead, we aim to explore a wider variety of IoT and non-IoT devices to foster smarter environments, examine emerging communication technologies and advanced deep learning methods, test datasets from IoT devices infected with spyware or targeted by cyber espionage, and identify unauthorized devices for enhanced security.

References

[1] S. Jeschke, C. Brecher, H. Song, and D. Rawat, Industrial Internet of Things: Cyber manufacturing Systems. Cham, Switzerland: Springer, 2017.

[2] J. Sun, “An open-access book about decoding mode-s and ads-b data,” https://mode-s.org/, May 2017.

[3] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma. IoT sentinel: Automated device-type identification for security enforcement in IoT. In Proc. of International Conference on Distributed Computing Systems (ICDCS), pages 2177–2184. IEEE, 2017.

[4] A. Bremler-Barr, H. Levy, and Z. Yakhini. IoT or not: Identifying IoT devices in a short time scale. In Proc. of Network Operations and Management Symposium, pages 1–9. IEEE, 2020.

[5] Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N. O. Tippenhauer, J. D.Guarnizo, and Y. Elovici. Detection of unauthorized IoT devices using machine learning techniques. arXiv preprint arXiv:1709.04647, 2017.

[6] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu. IoT security techniques based on machine learning: How do IoT devices use ai to enhance security? IEEE Signal Processing Magazine, 35(5):41–49, 2018.

[7] Y. Liu, J. Wang, J. Li, S. Niu, and H. Song. Machine learning for the detection and identification of internet of things devices: A survey. IEEE Internet of Things Journal, 9(1):298–320, 2021.

[8] J. Zhang, Z. Wang, Z. Yang, and Q. Zhang. Proximity based IoT device authentication. In Proc. of IEEE INFOCOM, pages 1–9. IEEE, 2017.

[9] N. Z. Gong, A. Ozen, Y. Wu, X. Cao, R. Shin, D. Song, H. Jin, and X. Bao. Piano: Proximity-based user authentication on voice-powered internet-of-things.

[10] H. Shafagh and A. Hithnawi. Poster: come closer: proximity-based authentication for the internet of things. In Proc. of Annual International Conference on Mobile Computing and Networking, pages 421–424, 2014.

[11] P. Robyns, E. Marin, W. Lamotte, P. Quax, D. Singel´ee, and B. Preneel. Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning. In Proc. of ACM Conference on Security and Privacy in Wireless and Mobile Networks, pages 58–63, 2017.

[12] K.-H. Wang, C.-M. Chen, W. Fang, and T.-Y. Wu, "On the security of a new ultra-lightweight authentication protocol in iot environment for rfid tags," The Journal of Supercomputing, vol. 74, no. 1, pp. 65–70, 2018.

[13] S. Al-Sarawi, M. Anbar, K. Alieyan, and M. Alzubaidi, "Internet of things (iot) communication protocols," in 2017 8th International conference on information technology (ICIT). IEEE, 2017, pp. 685–690.

[14] I. Jawhar, N. Mohamed, and J. Al-Jaroodi, "Networking architectures and protocols for smart city systems," Journal of Internet Services and Applications, vol. 9, no. 1, p. 26, 2018.

[15] F. Metzger, T. Hoßfeld, A. Bauer, S. Kounev, and P. E. Heegaard,"Modeling of aggregated iot traffic and its application to an iot cloud," Proceedings of the IEEE, vol. 107, no. 4, pp. 679–694, 2019

[16] S. Han, K. Jang, K. Park, and S. Moon, "Packetshader: a gpuaccelerated software router," ACM SIGCOMM Computer Communication Review, vol. 40, no. 4, pp. 195–206, 2010.

[17] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and openflow: From concept to implementation," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2181–2206, 2014.

[18] L. Gao, C. Zhang, and L. Sun, "Restful web of things api in sharing sensor data," in 2011 International Conference on Internet Technology and Applications. IEEE, 2011, pp. 1–4.

[19] A. Sivanathan, "Iot behavioral monitoring via network traffic analysis," arXiv preprint arXiv:2001.10632, 2020.

[20] A. Sivanathan, D. Sherratt, H. H. Gharakheili, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Characterizing and classifying iot traffic in smart cities and campuses," in 2017 IEEE

[21] S. Marchal, M. Miettinen, T. D. Nguyen, A.-R. Sadeghi, and N. Asokan, "Audi: Toward autonomous iot device-type identificationusing periodic communication," IEEE Journal on Selected Areas in Communications, vol. 37, no. 6, pp. 1402–1412, 2019.

[22] M. Abu Alsheikh, S. Lin, D. Niyato, and H. P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," IEEE Commun. Surveys and Tutorials, vol. 16, no. 4, pp. 1996–2018, Apr. 2014.

[23] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," IEEE Trans. Neural Networks and Learning Systems, vol. 27, no. 8, pp. 1773–1786, Mar. 2015.

[24] J. W. Branch, C. Giannella, B. Szymanski, R. Wolff, and H. Kargupta, "In-network outlier detection in wireless sensor networks," Knowledge and Information Systems, vol. 34, no. 1, pp. 23–54, Jan. 2013.

[25] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," Soft Computing, vol. 20, no. 1, pp. 343–357, Jan. 2016.18

[26] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Commun. Surveys and Tutorials, vol. 18, no. 2, pp. 1153–1176, Oct. 2015. Networks, pp. 3437–3444, Atlanta, GA, Jun. 2009.

[27] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in Proc. ACM Int Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), pp. 1–10, Chennai, India, Jul. 2017

[28] L. Xiao, Q. Yan, W. Lou, G. Chen, and Y. T. Hou, "Proximity-based security techniques for mobile users in wireless networks," IEEE Trans. Information Forensics and Security, vol. 8, no. 12, pp. 2089–2100, Oct. 2013..

[29] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A system for Denial-of-Service attack detection basedon multivariate correlation analysis," IEEE Trans. Parallel and Distributed Systems, vol. 25, no. 2, pp. 447–456, May 2013.

[30] X. He, H. Dai, and P. Ning, "Improving learning and adaptation in security games by exploiting information asymmetry," in IEEE Conf. Computer Commun. (INFOCOM), pp. 1787–1795, Hongkong, China, May 2015.

[31] V. Mnih, K. Kavukcuoglu, D. Silver, et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Jan. 2015.

[32] Y. Gwon, S. Dastangoo, C. Fossa, and H. Kung, "Competing mobile network game: Embracing anti-jamming and jamming strategies with reinforcement learning," in Proc. IEEE Conf. Commun. and Network Security (CNS), pp. 28–36, National Harbor, MD, Oct. 2013.