Available online: www.ijsrmes.com

International Journal for Scientific Research in Modern Engineering and Science

International Journal for Scientific Research in Modern Engineering and Science, 3(4):58-66 (2022)

MACHINE LEARNING APPROACHES FOR DETECTION AND IDENTIFICATION OF IOT AND NON - IOT DEVICES: - A REVIEW

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 stands for Internet of Things, which means accessing and controlling daily usable equipments and devices using Internet. The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. a rapidly evolving field, the Internet of Things (IoT) involves the interconnection and interaction of smart objects, i.e., IoT devices with embedded sensors, on board data processing capabilities, and means of communication, to provide automated services that would otherwise not be possible [1].

Trillions of network-connected IoT devices are expected to emerge in the global network around 2020 [2].The IoT is becoming pervasive parts of everyday life, enabling a variety of emerging services and applications in cities and communities including in health. Transportation energy/ utilities, and other areas. The current research focuses on the risks IoT devices pose to large corporate organizations. IoT security in

enterprises is associated with the behavior of the organization itself, as well as its employees. Self-deployed IoT devices may support a variety of enterprise applications. For instance, smart cameras and smoke detectors enhance security; smart thermostats, smart light bulbs and sockets facilitate power savings; and so forth. Given this, care should be taken to make sure that such Web-enabled devices do not contribute to an expansion of the cyber attack surface within the organization. The smart TVs typically installed in conference rooms are a good example.

IoT Devices Internet of Things Devices is non-standard devices that connect wirelessly to a network with each other and able to transfer the data. IoT devices are enlarging the internet connectivity beyond standard devices such as smart phones, laptops, tablets, and desktops. Embedding these devices with technology enable us to communicate and interact over the networks and they can be remotely monitored and controlled There are large varieties of IoT devices available based on IEEE 802.15.4 standard. These devices range from wireless motes, attachable sensor-boards to interface-board which are useful for researchers and developers. IoT devices include computer devices, software, wireless sensors, and actuators. These IoT devices are connected over the internet and enabling the data transfer among objects or people automatically without human intervention.

2. LITERATURE REVIEW

In recent work, the authors have provided a system capable of automatically identifying the type of IoT device through the use of machine learning (ML) and limiting the communication of vulnerable devices to minimize damage inflicted on the network [3]. However, the protocol relies on MAC addresses to identify a new device trying to authenticate with the network, which can be spoofed. Furthermore, the work does not consider the case where a previously authenticated device is compromised or a re-authentication process for every time the device is reintroduced into the network.

Bremler et al. focus on distinguishing between IoT a Non-IoT (NoT) devices through the use of ML classifiers to assign relevant security policies . to the device [4]. However, among several drawbacks, the biggest limitation is demonstrated by their classification technique since their classification model is only capable of classifying a device as IoT or not-IoT. Not addressing this issue would allow the adversary to compromise a vulnerable device to actuate activities that should only be performed by a different kind of device, for example, using a vulnerable smart camera to open a lock or a garage door.

In another work, the authors aim to automatically detect suspicious IoT devices in a network through the Random Forest classifier and white list devices that are classified as trustworthy [5]. The drawback of this approach is that if a vulnerable device that has already been white listed is compromised by an adversary, it would give the adversary unrestricted access to the network, since the network is not capable of identifying the change in device behavior.

In other work, the authors propose IoT security solutions based on ML techniques, including reinforcement learning, unsupervised learning, and supervised learning to improve IoT systems spoofing resistance and detection

and to authenticate a device to protect data privacy [6]. Their work attempts to detect an attack through several machine learning techniques; however, each attack is identified through a different machine learning model, which can end up utilizing a large number of resources such as memory and time. Furthermore, in our opinion, if the models are trained to detect an attack by observing a certain pattern, the attack in real-life might differ from what the models are trained to identify which could potentially leave attacks undetected. A survey related to ML-based classification techniques to detect and identify legitimate and rogue IoT devices to provide security where conventional approaches that use cryptographic protocols cannot be applied [7]. However, the paper demonstrates limitations as it does not present a complete and formal authentication protocol that can incorporate the ML techniques presented in order to protect the network from certain attacks in real life.

3. PROXIMITY BASED LITERATURE REVIEW

Proximity-based solution proposed in [8] is based on several physical activities performed by a user, The work provides a notable contribution in the field of IoT device authentication however, it requires a significant amount of work to be performed by the user. Furthermore, the authors only address initial authentication and do not discuss the measures that must be taken if an already authenticated vulnerable device is compromised.

In other work, the authors propose a proximity-based user authentication solution for voice-powered IoT devices [9]. The work presents a voice-based distance estimation

technique to authenticate IoT devices using various advanced technologies the biggest constraint for the proposed method is that it is only applicable to voice-powered IoT devices.

Shafagh and Hithnawi propose another proximity-based solution for IoT device authentication by solely utilizing the wireless communication interface [10] However, their solution presents limitations, since it does not account for a device in close proximity being compromised due to its security. vulnerabilities, which could lead an adversary to perform attacks such as actuation, poisoning the network, and capturing network traffic.

Finally in other work, the authors propose a device identification based on fingerprint recognition of the wireless device chipset [11]. However, their solution is not capable of 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	Previously authenticated device is compromised or a reauthentication process for every time the device is reintroduced into the network

[4]. Bremler et al.	Distinguish between Iot and Non-Iot Devices	ML classifiers	only capable of classifying a device as IoT or not-IoT. Not able to classify the 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 trustworthy	Compromised device that has been whitelisted will give the adversary unauthorized access to the network .since it cannot detect changes in device behaviors
[6] L. Xiao et al.	To Identify attacks	ML techniques, including reinforcement learning, unsupervise learning, and supervised learning	this models are trained to detect an attack by observing a certain pattern, the attack in real-life might differ from what the models are trained to identify which could potentially leave attacks undetected.can not detect real life attacks.
[7] Y. Liu, et al.	IoT Device type Classification	Multiple Classifier	Donot Provide Formal Authentication protocol

Proximity Based Problem Identification

Authors	Focus on	Problem Identified
[8] J. Zhang et al.	Based on Physical Activity performed by user	do not discuss the measures that must be taken if an already authenticated vulnerable device is compromised.
[9] N. Z. Gong et al.	Use of voice-powered IoT devices	Only Applicable for voice-powered IoT devices
[10] H. Shafagh et al.	utilizing the wireless communication interface	which could lead an adversary to perform attacks such as actuation, poisoning the network, and capturing network traffic.
[11] P. Robyns	device identification based on fingerprint recognition	not capable of identify a compromised legitimate device.

5. THREAT MODE OF ROGUE DEVICES IN IOT

This section briefly reviews the threat modes of rogue devices along with countermeasures in IoT. We analyze the attack chain and identify the essential requirements of IoT device detection and identification: verifying legitimate devices' identity, detecting unknown or falsified devices, and detecting compromised (hijacked) devices

with abnormal behaviors. The cyber infrastructure of IoT allows sharing information and collaborating among devices with different capacities and vulnerabilities. On the one hand, this scheme cultivates a large open system with low entry restrictions. On the other hand, adversaries can conduct rogue activities with great convenience [12].

Generally, the attack modes of adversaries

in IoT are in two folds: passive attack and proactive attacks. In a passive attack, adversaries do not cause damage or performance degradation for a long time. Still, they passively analyze devices' communication and activity patterns, providing road maps for proactive attacks in the future. If we regard passive attackers as spies secretly and peacefully gathering intelligence, the proactive attackers do whatever possible to degrade performances or exploit devices to conduct malicious activities. In practical attacks, proactive and passive attacks are combined

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

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

Among these stages, passive and proactive attacks are combined in the penetration stage. From the perspective of network operators or cyber security surveillance agents, if we can prevent the adversaries from successfully impersonating legitimate devices in the first stage (penetration) or can identify hijacked devices in the second stage (spying). Network operators and surveillance agents can destroy the whole attack chain.

6. LEARNING-ENABLED DEVICE IDENTIFICATION IN IOT

This section reviews methods to recognize devices' identities sand types in IoT. Most of them are based on network traffic and wireless signal pattern recognition. We first review device type identification methods, which are widely used in identifying commercial IoT devices.

A. Device type identification

Even though device types are not directly related to devices' identities, they still provide essential information for network management and risk control. A brief diagram of typical IoT devices is in Figure 5, and comparisons of their Physical Layer, Data Link Layer as well as aggregated data transmission characteristics

are presented in [13], [14] and [15], respectively. As in Figure 1, WiFi is pervasively utilized in smart homes while smart cities prefer reliable cellular networks. Device type identifications are frequently performed on network, transportation, and application layers and implemented in Software Defined Network (SDN) controllers or Software Routers [16]–[17].

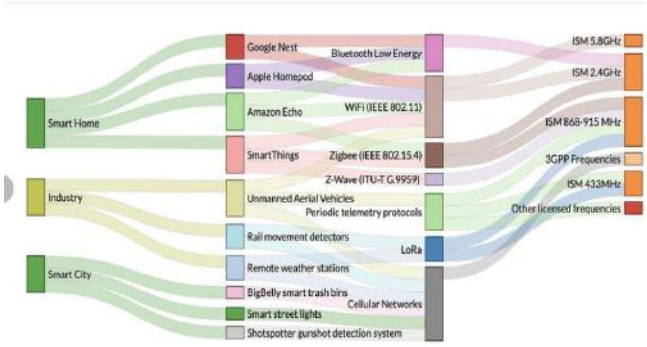

Figure 1. Typical IoT devices and protocols.

Device types reveal functionalities and activity profiles. A taxonomy of features for device type identification is presented in Figure 2.

As in Figure 2, remote service is a popular attack surface to disclose the device type or even identity. The reason is that the IoT devices communicate with remote service providers through the REST API [18]. Even though sensitive data are encrypted, some unique strings in their Web requests can still be exploited to infer device types. Authors in [19] present that using only port numbers, domain names, and cipher suites, a Naive Bayesian classifier can reach high accuracy in classifying 28 commercial IoT devices.

Even though modeling devices' remote service requests provides promising results in device type identification, these solutions may not work if they interact with anonymous service providers. For alleviation, their activity and data flow patterns can be utilized. Authors in [20] propose that their Random Forest classifier reaches a high accuracy of 95%in identifying 20 IoT devices when features of activities, network data flows, and remote service requests are utilized simultaneously. In [21], devices' types are identified based on the periodicity of activities. The authors first use the Discrete Fourier Transform (DFT) and discrete autocorrelation to find the dominant periods in protocol-specific activities .They then use statistical and stability metrics to model devices' behaviors.

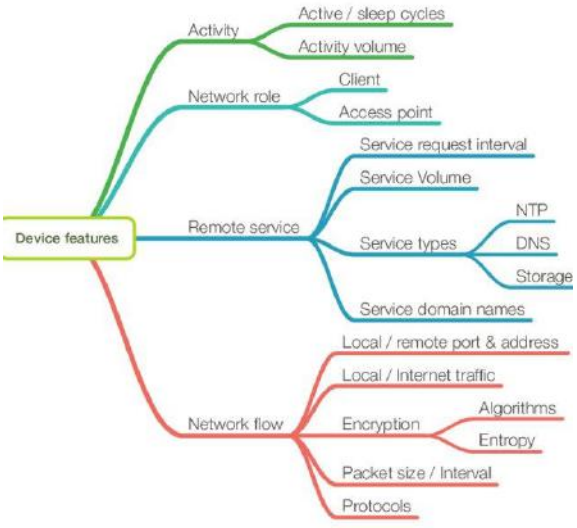

Figure 2. Features for device type identification.

7. MACHINEL EARNING TECHNIQUES FOR IOT DEVICE SECURITY ENHANCEMENT

Machine learning techniques including supervised learning, unsupervised learning, and reinforcement learning (RL) have been widely applied to improve network security, such as authentication, access control, anti-jamming offloading and malware detections.

Supervised learning techniques such as support vector machine (SVM), naive Bayes, Knearest neighbor (K-NN), neural network, deep neural network (DNN) and random forest can be used to label the network traffic or app traces of IoT devices to build the classification or regression model [22]. For example, IoT devices can use SVM to detect network intrusion[22] and spoofing attacks [23], apply K-NN in the network intrusion [13] and malware [24] detections, and utilize neural network to detect network intrusion [25] and DoS attacks [26]. Naive Bayes can be applied by IoT devices in the intrusion detection and random forest classifier can be used to detect malwares .IoT devices with sufficient computation and memory resources can utilize DNN to detect spoofing attacks [27].

Unsupervised learning does not require labeled data in the supervised learning and investigates the similarity between the unlabeled data to cluster them into different groups [24]. For example, IoT devices can use multi-variate correlation analysis to detect DoS attacks [29] and apply IGMM in the PHY-layer authentication with privacy protection [28].

Reinforcement learning techniques such as Q-learning, Dyna-Q, post-decision state (PDS)[30] and deep Q-network (DQN) [31] enable an IoT device to choose the security protocols as well as the key parameters against various attacks via trial-and-error For example, Q-learning as a model free RL technique has been used to improve the performance of the authentication [6], anti -jamming offloading [6], [32] and malware detections [6], IoT devices can apply Dyna-Q in the authentication and malware detections , use
PDS to detect malwares [and DQN in the anti-jamming transmission

8. CONCLUSION

This survey provides a more comprehensive overview of existing advanced technology for detecting and identi-fying IoT devices. With the rise of the Internet of Things (IoT), a large number of IoT devices are being built in a variety of situations, including businesses, homes, warehouses, and highways. The appropriate security of IoT devices is critical because the state of different IoT devices has different properties. The literature review revealed a large number of cited works on IoT device identification and classification. Still, the majority of the work done on small enterprise networks is based on static information such as port information, MAC addresses, and device model train test .In the future, we hope to investigate a broader range of IoT device types and Non-IoT devices for creating an intelligent environment, investigate new communication technologies and deep learning techniques, experiment with data from IoT devices compromised with spyware and cyber espionage, and detect un authorized devices for security purposes.

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.