A Report on the Security Measures and Cost of Privacy for Devices in IoT

Monika Balamurugan

Technische Universität Dortmund
monika.balamurugan@tu-dortmund.de

Abstract. As our lives migrate to the digital realm, our online identity has evolved to become an increasingly robust collection of data about every aspect of our online and offline lives. This data is extremely appealing to companies who wish to use it for a variety of analytics. In this report, we create awareness for the consumers around the world who have only a vague understanding of how much of their data is being tracked, where, when, and by which companies.

Keywords: Data Privacy, IoT Analytics, Consumer Data Collection.

1 Introduction

Devices powering the Internet of Things (IoT) are everywhere [21]. Every connected device with the ability to send data through a network autonomously without any human interaction qualifies [11]. This includes modern passenger and commercial fleet vehicles, industrial robotics, battery-powered sensors, consumer devices, and several other smart machines. IoT is no longer a new technology that people wish to experience in the future. IoT is actively deployed and growing rapidly.

As more devices come to market, research forecasts for IoT solutions also grow exponentially [19, 27]. A 2019 study from Business Intelligence predicted more than 64 billion IoT devices by 2025. The growth is directly attributed to advantages that IoT introduces to businesses, health care organizations, and the industrial system (Industrial IoT or IIoT) [9]. Moreover, the introduction of 5G networking will serve developers with new opportunities to create low-power, high-speed communications devices with almost zero transmission delays.

Yet, the most problematic concern about this technology is its security. IoT devices are known to be highly vulnerable to cyber attacks such as DDoS, spoofing, malware, and privacy issues [28]. Regulators, manufacturers, and enterprise users are all equally responsible for the security of this technology.

At the same time, penetration testing (often referred to as pentesting) is still one of the available solutions that guarantee the strength of IoT security. Pentesting is the process of hacking into computer systems, networks or web applications in search of finding vulnerabilities that lead to cyber attacks. Pentesting remains a manual process carried out by ethical hackers. Hence, we are here to give an

overview of how pentesting, with all its pros and cons, is used to increase IoT security.

Also companies are collecting data about our online activities, browsing the Internet, posting on social media, etc. Together, such collected data make up a digital identity for each person. There is no official list of what data is collected or leaves the consumer devices, they are also confused about where they are being tracked. Across the world, people are uncomfortable with how their data is being used by companies, even when it comes to worthy causes like COVID-19 tracking. We report that consumers often have to sacrifice their privacy to access their desired services, thus leading to decreasing levels of trust. There are *3 Ts* namely Tracking, Trust, and Time that reveal the trade-offs faced by consumers when using online services [3].

2 Benefits of Pentesting an IoT Environment

For enterprises, the usefulness of IoT only comes with its safety. Therefore, conducting comprehensive pentesting on all the elements of the IoT ecosystem will bring various advantages including; managing risks, detecting security threats, empowering devices security, and ensuring business continuity.

Plus, securing the IoT ecosystem will help enterprises evade any data breaches and thus violating data protection laws such as GDPR. More, the final result of a pentesting process will assist stakeholders and executives to make business decisions in the future. Further, deploying tests on IoT devices could lead to discovering new attack vectors and approaches, and consequently fostering IoT security.

3 Awareness on Social Media Tracking

As consumers scroll through their feeds, they need to have a general awareness that companies are tracking their data [7]. Many believe that social media companies track at least one element of their personal data. People from the Netherlands likely expect social media companies to actively collect data on their posts and only 40% do not think their data is safe. In France, 59% of French people are not aware that their posts are being tracked by social media companies, and 45% of Australians agree that their data is being tracked.

There is a popular expression that *if you're not paying for the product, you are the product*. The Internet provides a wealth of information and free services to consumers. But people often forget that the free service they receive from an Internet search, and social media account are being funded, and advertisers are buying access to them. From a recent report from Cambridge Analytica, the My Health Record in Australia has suffered potential data breaches, shrinking levels of privacy. Yet, many consumers are not aware of the routine private data tracking and data harvesting that is common today.

4 Tracking a Global Pandemic like COVID-19

Recently people have been immersed with COVID-19 news, which provides rapid updates about case counts, reopening plans, etc. Hence, just 4% of Australians are unaware of the smartphone data collection efforts that aim to track the spread of COVID-19. In other countries like Dutch, 14% of respondents are unaware, 35% of Americans are unaware, and 17% of Germans are unaware of data tracking efforts to contain the novel coronavirus. In Australia, 19% of people are opposed to data tracking even for the COVID-19 purpose. Hence COVID-19 is closing the data tracking education gap. Whether or not people understand which data is being tracked, they are overall unhappy about companies tracking their data, regardless of its purpose of collection. Around 91% of people are uncomfortable with at least one kind of company tracking their data.

5 Distrust of Social Media Providers

Almost half of the people have doubts that their conversations and biometric data are being tracked by companies, although there is consensus as to which type of services/companies are collecting data. People from the Netherlands show the highest rate of discomfort since 90% of people are saying they are upset with personal conversation tracking, and 88% are uncomfortable with biometric data gathering. The majority of people from various age groups, for example, 43% of Australians are totally not comfortable with any form of personal data collected for any purpose. The most common data that people think companies are tracking is location data. 57% are uncomfortable with services tracking their location. Recently in [8], Google was storing location data on smartphones even if the consumer had turned off location access.

The companies that provide us free service own our data that is generated when using their services. To give consumers control over their data being provided to companies, recently, some studies [10] have proposed the idea that the services that collect personal data should pay the data owner (source of data generation). Although there is a provision for people to get compensated for their data, still 37% say no, and 27% are unsure if a one-time payment is worth the data. But for specific types of data (critical ones), 76% of people are unwilling to sell at least one type of their data at any price. Hence, consumers value their data more than earning money.

6 Steps Necessary for Successful IoT Pentesting

First, the IoT ecosystem demands three components to operate suitably, which are:

The things. Devices such as self-driving cars, cameras, sensors, and all the devices [23, 25] that reside on the edge of the network.

The gateways. Those are the materials that function as a bridge between the IoT devices and the data aggregation-spot. It can be a router or any device that connects two or more elements on the network.

Cloud data centers. This could be either private or public clouds and it's where data is stored and analyzed. This is the place where all the magic happens.

Second, pentesters should carry out a reconnaissance process on five levels, which are:

Hardware-level. Both edge devices and gateways hardware, chips, storage, and sensor should be investigated via reverse engineering and disassembling to identify any subversion vulnerabilities on them.

Network-level. This includes evaluating wireless protocols [13] such as Wi-Fi, Bluetooth, ZigBee, and narrowband (NB) 5G; Encryption protocols, and end-to-end authentication and authorization for any potential weaknesses.

Firmware-level. Diverse types of operating systems should be analyzed to search for possible vulnerabilities, such as privilege escalation, Buffer Overflow, and zero-day exploits. This is done by examining the updating process, checking cryptographic primitives, and password storing mechanisms.

Web Application-level. targeting the APIs to look for any SQL injection, XSS, and Broken Authentication and Session Management that could lead to unauthorized access to the devices.

Cloud-level. Conducting a test on the operating systems and network infrastructure of the data aggregation point is mandatory to spot any issues that could threaten data privacy. If it's a public cloud, then both parties, vendors and end-users, are responsible for its security.

After completing the recon process and gathering all the essential information, pentesters need to start attacking all the components using the appropriate tools. For example, pentesters should run a "man-in-the-middle" attack on the network-level to check if the encryption algorithms are working accurately [4].

Another scenario that the pentester should undertake is to interrogate the user-interface with brute-force attacks and see if the passwords used are sufficiently strong. Be aware that most IoT devices come with default passwords established by the manufacturer, and this is one of the reasons devices get hacked with ease.

This is a simplified explanation of the steps that pentesters usually perform. Everything seems to be reasonable and straightforward, but pentesting an IoT environment isn't as simple as it might appear [12].

7 Can Edge Analytics Protect Real-life Data

In addition to the data collection concerns, people worry more if the data collected are used to directly make money [2]. Around 93% of people considered in the survey are uncomfortable with companies selling their data. Nine out of 10 Australians are upset with companies profiting using any of their collected personal data. More importantly, the sale of biometric data, passwords, and offline conversations cause the most concern at 86% each. Hence, there is a strong

need for analytics and data processing to be performed offline on the device owned by the consumers. i.e., data should not be transmitted out of the device even for authentication or for advanced analytics purposes. In the following, we brief some of the recent works that accomplish the tasks demanded by users, offline, at the device level.

Today's DL models when efficiently deployed can convert normal IoT devices into intelligent IoT devices that can solve a wide variety of problems [20, 26]. For example, in [17], a face recognition algorithm was trained using Deep Neural Network and deployed on their modern Alexa smart speaker prototype. This model, without disturbing the smart speaker routine, can detect and identify a human face and start the Alexa voice service only when an authorized face is present in the live video frames. Similarly, in [18], a DL and Open CV based object detection model was deployed in their smart speaker. Here, whenever the user calls out the command *Alexa, ask Friday what she sees*, the smart speaker camera turns on and executes the deployed model and calls out the names of detected objects as a response to the user's command. The models optimized using methods such as Edge2train [14] and RCE-NN [15] can run on the IoT devices. Processing data at the edge level without depending on the cloud improves latency, reduces subscription & cloud storage costs, processing requirements, and bandwidth requirements. It also can address privacy and security issues by avoiding the transmission of sensitive or identifiable data over the network.

Hybrid or composite approaches involving conventional CV and DL should be used to take great advantage of the limited computing resource available at the edge [22, 24]. Such heterogeneous systems consist of a combination of multiple processors and chipsets. For example, in the Smart Hearing Aid prototype from [16], the users have integrated a DSP-based microphone array with a Linux SBC to perform edge level audio processing such as noise suppression, the direction of arrival estimation, etc., without depending on the internet. The IoT devices can be power efficient when the user assigns different workloads to the most efficient compute engine [21]. For Deep Learning (DL) use cases, the SIFT [6], SURF [1], BRIEF [5] feature descriptors need to be generally combined with traditional machine learning classification algorithms such as SVMs, SVRs, Random Forests, Decision Trees to solve the CV problems. DL is sometimes overkilling when the given problem is simple and if it can be solved by CV techniques. Algorithms like pixel counting, SIFT, simple color thresholding are not class-specific. i.e., they are very general and perform the same for any number of images.

8 The Issues with Pentesting an IoT Environment

Pentesting an IoT ecosystem presents various complicated challenges for security teams for several reasons, such as the diversity of hardware, software and protocols of the devices. Normally, pentesters perform analyses on known operating systems (such as Windows and Linux 64/x86), networking protocols (UDP, TCP, FTP,

etc.) and hardware. In the case of IoT, pentesters are obligated to have more knowledge about other architectures such as MIPS and SuperH, protocols (ZigBee, BLE, NFC), and embedded engineering. Due to the cybersecurity shortage in today's marketplace, pentesters with such capabilities are rare to be found.

It is difficult for pentesters to attack embedded devices because most of the attacks require user interaction to be completed. Due to its complexity, pentesting an IoT environment manually takes time and only produces static results (outputs including PDF reports or Excel sheets), which need to be turned into actionable insights. It will take time to resolve vulnerabilities and make business decisions.

9 Summary

In general, manual IoT pentesting takes time and demands a lot of effort from the pentester, but it puts them closer to being in the shoes of real cybercriminals. On the other hand, automated pentesting offers more efficiency and velocity. Choosing the best method to pentest an IoT ecosystem can vary from one organization to the next. Nevertheless, the overall goal is to enhance the usefulness of enterprise IoT by making it more secure.

Also we believe it is important to balance privacy and innovation. Consumers deserve to have control over their data, but at the same time, data is often essential to building new technologies and serving the common good like accurate tracking the spread of COVID-19. To strike a balance, organizations around the globe must embrace transparency and help their customers understand how their data is being used.

References

1. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: European conference on computer vision. pp. 404–417. Springer (2006)
2. Casanovas, P., Mendelson, D., Poblet, M.: A linked democracy approach for regulating public health data. Health and Technology 7(4), 519–537 (2017)
3. Distler, V., Lallemand, C., Koenig, V.: How acceptable is this? how user experience factors can broaden our understanding of the acceptance of privacy trade-offs. Computers in Human Behavior 106, 106227 (2020)
4. Jyothi, P.D., Lakshmy, K.: Vuln-check: A static analyzer framework for security parameters in web. In: Soft Computing and Signal Processing, pp. 233–247. Springer (2022)
5. Karami, E., Prasad, S., Shehata, M.: Image matching using sift, surf, brief and orb: performance comparison for distorted images. arXiv preprint arXiv:1710.02726 (2017)
6. Karami, E., Shehata, M., Smith, A.: Image identification using sift algorithm: Performance analysis against different image deformations. arXiv preprint arXiv:1710.02728 (2017)
7. Leskovec, J.: Social media analytics: tracking, modeling and predicting the flow of information through networks. In: Proceedings of the 20th international conference companion on World wide web. pp. 277–278 (2011)
8. Nakashima, R.: Ap exclusive: Google tracks your movements, like it or not (08 2018), `https://apnews.com/article/828aefab64d4411bac257a07c1af0ecb`

9. Petrenko, A.S., Petrenko, S.A., Makoveichuk, K.A., Chetyrbok, P.V.: The iiot/iot device control model based on narrow-band iot (nb-iot). In: 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). pp. 950–953. IEEE (2018)
10. Posner, E.: On cultural monopsonies and data-as-labor (01 2018), http://ericposner.com/on-cultural-monopsonies-and-data-as-labor/
11. Rollo, F., Sudharsan, B., Po, L., Breslin, J.G.: Air quality sensor network data acquisition, cleaning, visualization, and analytics: A real-world iot use case. In: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers. pp. 67–68 (2021)
12. Stanislav, M., Beardsley, T.: Hacking iot: A case study on baby monitor exposures and vulnerabilities. Rapid7 Report (2015)
13. Sudharsan, B., Breslin, J.G., Ali, M.I.: Adaptive strategy to improve the quality of communication for iot edge devices. In: 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). pp. 1–6. IEEE (2020)
14. Sudharsan, B., Breslin, J.G., Ali, M.I.: Edge2train: A framework to train machine learning models (svms) on resource-constrained iot edge devices. In: Proceedings of the 10th International Conference on the Internet of Things. pp. 1–8 (2020)
15. Sudharsan, B., Breslin, J.G., Ali, M.I.: Rce-nn: a five-stage pipeline to execute neural networks (cnns) on resource-constrained iot edge devices. In: Proceedings of the 10th International Conference on the Internet of Things. pp. 1–8 (2020)
16. Sudharsan, B., Chockalingam, M.: A microphone array and voice algorithm based smart hearing aid. CoRR abs/1908.07324 (2019), http://arxiv.org/abs/1908.07324
17. Sudharsan, B., Corcoran, P., Ali, M.I.: Smart speaker design and implementation with biometric authentication and advanced voice interaction capability. In: Proceedings for the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science (AICS). pp. 305–316 (2019)
18. Sudharsan, B., Kumar, S.P., Dhakshinamurthy, R.: Ai vision: Smart speaker design and implementation with object detection custom skill and advanced voice interaction capability. In: 2019 11th International Conference on Advanced Computing (ICoAC). pp. 97–102. IEEE (2019)
19. Sudharsan, B., Malik, S., Corcoran, P., Patel, P., Breslin, J.G., Ali, M.I.: Owsnet: Towards real-time offensive words spotting network for consumer iot devices. In: IEEE 7th World Forum on Internet of Things, 2021. p. 1 (2021)
20. Sudharsan, B., Patel, P.: Machine learning meets internet of things: From theory to practice. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (2021)
21. Sudharsan, B., Patel, P., Breslin, J., Ali, M.I., Mitra, K., Dustdar, S., Rana, O., Jayaraman, P.P., Ranjan, R.: Toward distributed, global, deep learning using iot devices. IEEE Internet Computing 25(03), 6–12 (2021)
22. Sudharsan, B., Patel, P., Breslin, J.G., Ali, M.I.: Enabling machine learning on the edge using sram conserving efficient neural networks execution approach. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (2021)
23. Sudharsan, B., Patel, P., Breslin, J.G., Ali, M.I.: Sram optimized porting and execution of machine learning classifiers on mcu-based iot devices: demo abstract. In: Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems. pp. 223–224 (2021)
24. Sudharsan, B., Patel, P., Breslin, J.G., Ali, M.I.: Ultra-fast machine learning classifier execution on iot devices without sram consumption. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). pp. 316–319. IEEE (2021)

25. Sudharsan, B., Patel, P., Wahid, A., Yahya, M., Breslin, J.G., Ali, M.I.: Porting and execution of anomalies detection models on embedded systems in iot: Demo abstract. In: Proceedings of the International Conference on Internet-of-Things Design and Implementation. pp. 265–266 (2021)
26. Sudharsan, B., Salerno, S., Nguyen, D.D., Yahya, M., Wahid, A., Yadav, P., Breslin, J.G., Ali, M.I.: Tinyml benchmark: Executing fully connected neural networks on commodity microcontrollers. In: IEEE 7th World Forum on Internet of Things (WF-IoT) (2021)
27. Sudharsan, B., Sundaram, D., Breslin, J.G., Ali, M.I.: Avoid touching your face: A hand-to-face 3d motion dataset (covid-away) and trained models for smartwatches. In: 10th International Conference on the Internet of Things Companion. pp. 1–9 (2020)
28. Sudharsan, B., Sundaram, D., Patel, P., Breslin, J.G., Ali, M.I.: Edge2guard: Botnet attacks detecting offline models for resource-constrained iot devices. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). pp. 680–685. IEEE (2021)