ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 12, Issue 3, 2024
https://zenodo.org/records/14505947

Engineering and Security Best Practices for Machine Learning-based Internet of Things Devices

Sravanth Vukanti[1], Assistant Professor[1], Department of ECE, Siddhartha Institute of Technology & Sciences, Telangana, India

Kurada Pawan Kumar Sharma[2], Assistant Professor[2], Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana, India.

ABSTRACT

Machine learning (ML) technologies on edge devices are being included into the newest generation of IoT systems. This presents significant technical difficulties in bringing ML to hardware with limited resources, as well as new hurdles in protecting user data and preserving system integrity. For machine learning enabled IoT goods, existing research recommends iterative techniques to simplify development and boost product success. These procedures are not tailored to the unique needs of machine learning or IoT devices, instead relying on the tried-and-true methods used in other, more generalised, software development domains. Through the viewpoint of the engineering lifecycle, this study aims to define engineering procedures and security practises for ML-enabled IoT devices. We conducted a survey (N=25) and interviews (N=4) with working professionals to get our data. We discovered that different businesses have different security engineering practises and procedures. When asked about the engineering cost of security research and threat modelling, respondents highlighted the trade-offs that must be made with the business's demands. If security is not mandated, engineers will spend less time and effort on it. When adopting ML for IoT devices, practitioners consistently raised concerns about IP theft and reverse engineering. Our results suggest that more effort should be put into exploring the interplay between technical constraints including cost, compliance, and security.

Keywords : Cyber Physical Systems, Embedded Systems, Software Engineering, and Internet of Things.

INTRODUCTION

IoT is a paradigm that merges the cyber and physical worlds by linking edge devices ("Things") to one another and to more powerful resources across the network ("Internet") [15]. By 2025, the number of connected devices is expected to increase by a factor of two [30, 57, 58], from the current 35 billion. Using machine learning (ML) [38, 39], IoT systems can make timely, well-considered judgement calls [8, 67]. The resultant intelligent IoT systems have the potential to revolutionise many parts of the economy [42], but there are also significant hazards involved. Engineers should use ML techniques on limited-capacity IoT devices in a safe, private manner to reduce vulnerabilities [16]. We know surprisingly little about manufacturers' engineering processes [28, 46, 53], despite the growing relevance of intelligent IoT systems to consumers, businesses, and governments.

High-profile failures such as assaults on waterworks systems resulting in contaminated water supply [55], aggressive data collecting tactics [4, 48], and vulnerabilities leading to Internet of Things botnets [1] raise concerns regarding engineering practises. Using programme analysis and failure analysis, researchers have looked at IoT software problems [46] and security issues [12, 18, 20- 23, 25, 34, 35, 47, 61] from the software's point of view. For both the creation of ML models and the creation of ML-enabled edge devices, researchers have developed general models of the secure software development life cycle (SDLC) [28, 53]. However, real-world adoption difficulties and present industrial practises remain mostly unexamined. As such, we want to research how ML is integrated into IoT devices throughout the development phase. Our overarching study topics are as follows: What procedures do companies often use when creating and overseeing IoT devices that are based on machine learning? How does security fit into product life cycle processes? We conduct a survey (N=25) and interviews (N=4) with working professionals to get to the bottom of these mysteries.

BACKGROUND

The increasing prevalence of computer systems that include network-edge intelligence has prompted this investigation due to concerns about data privacy and security. While there is no universally accepted definition of "Internet of Things" devices [56], we define them here as those that include sensors and/or actuators, are connected to a network, and have constrained memory, power, and processing capabilities. Low-budget Internet of Things systems (IoTs) integrate sensor and networking capabilities [50, 70]. Methodology of IoT engineering: Because of their scattered nature, limited resources, and combination of physical and digital elements, engineering procedures for IoT systems are notoriously difficult [68]. We based the structure of our research on the typical engineering lifecycle shown in Figure 1 for ML-based IoT devices. This lifespan integrates previous studies [2, 28, 53]. Here, the engineering of the Internet of Things is shown as a five-stage, iterative process: The product's purpose is outlined, which may place

ISSN 2321-2152
www.ijmece.com
Vol 12, Issue 3, 2024
https://zenodo.org/records/14505947

limitations on the available hardware and software options. System architecture, framework, and assessment methodology decisions are all determined during the design phase. Frameworks for development are used to put into action design choices.

Tuning hyperparameters, decreasing the model's computational complexity (such as with deep learning-based models), and adjusting network blocks are all ways to improve an ML model's performance [26, 40]. The solution is device-agnostic and aims instead at a hardware profile. The last step in the process is called "deployment," and it entails transferring the finished product to the intended computer. Pruning and other optimizations performed during deployment assist the model conform to the limitations of the IoT device [39]. The parameters of optimization algorithms are generally consistent, although they change depending on the capabilities of the target hardware [53]. The software has been installed on the hardware, and now engineers must check to see whether the requirements of the system have been satisfied. Goals for performance, fault tolerance [31, 59], and security holes may be cited as sources of concern. When designing systems, engineers think about both general threat models and ones that are unique to ML. Researchers have considered attacks such as using tampered training data [69] or deconstructing a model [49]. IoT Security: For designed systems, security is an overarching issue [51]. More and more stages of the engineering life cycle are beginning to include security [41]. Developers of IoT systems, however, often find security to be a difficult and time-consuming issue [46]. The engineering team may feel accountable for security, but they may not have a defined security procedure [9, 45, 63]. Prioritizing functionality and meeting deadlines above security is common [14, 24, 43], and implementing security on devices with limited resources reduces efficiency in those areas [11, 60]. The academic community currently lacks insight into industrial processes, but this engineering process model for ML-based IoT development is a promising start. Because of this information deficiency, we are unable to address the industry-wide issues and difficulties that must be overcome when creating and sustaining safe ecosystems. This research begins to close that void.

METHODOLOGY

Given the nature of our research concerns, we opted for an exploratory methodology [54], which combines quantitative and qualitative methods to learn more about a phenomenon and generate fresh research inquiries. We got broad information through a survey and deep understanding from individual interviews.

Survey

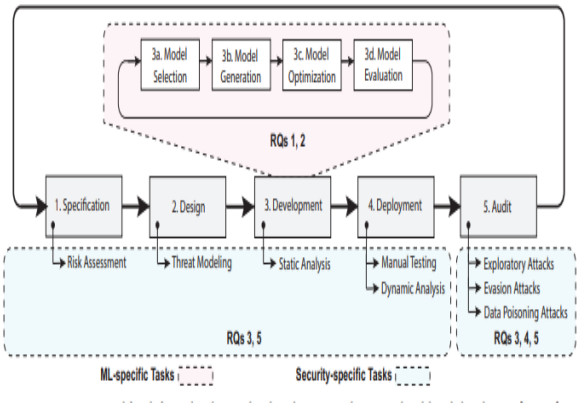
Our study objectives informed the development of a 32-question survey instrument that we administered to participants over the course of 10 minutes. Seven of the demographic questions [10, 13] were culled from the aforementioned literature, while the other questions were formulated in accordance with standard best-practices in survey design [29]. Our own professional experience implementing ML on IoT devices served as the inspiration for the first round of questions, which were later honed via dialogue with experts in the field. We sent the poll to two professionals to gauge its validity and length, and we tweaked it depending on their input.

Passing out surveys:

Due to the niche nature of the engineering security methods under investigation, we posted the poll in many places: the public forums Reddit, Hacker News, and Towards AI; our personal networks through Facebook and LinkedIn; and our department's mailing list. In addition, we requested that survey takers forward the information to their co-workers (snowball Sam pling [36]). After 5 weeks, the survey was ended after being released in the last week of March 2021. To encourage them to fill out the survey, we offered them a chance to win a $50 gift card.

Method of analysis:

With the help of Qualtrics reports, we assessed the information. Together, all of the participants' responses to each question were analysed. All survey data shown in the diagrams is expressed as a percentage of the total replies for consistency's sake.



Interviews

We developed our interview technique to be an expansion of the survey questions. We kept an eye

ISSN 2321-2152
www.ijmece.com
Vol 12, Issue 3, 2024

https://zenodo.org/records/14505947

on survey replies and crafted new inquiries in regions where respondents disagreed or offered surprise responses. Each interviewee was given 30–40 minutes to talk while answering 8 predetermined questions [27]. We conducted a trial run of the interview process with a single clinician to assess its viability and efficiency. We gathered our interviews from the group of people who filled out our survey. Participants in the survey were qualified for a more in-depth interview due to their backgrounds in machine learning and internet of things engineering. Respondents were offered a $25 gift card incentive for volunteering for a follow-up interview after completing the survey. We contacted everyone who expressed an interest and conducted interviews with those who agreed to be questioned. Confidentiality of participants: An outside firm transcribed the interview audio. Personal information was masked before analysis to protect the privacy of our participants.

Information Obtained from the Survey

Twenty-five people filled out the survey, with just fourteen filling it out in its entirety. We studied the data from partial replies in addition to the few complete responses because of the low response rate. The median responder who just partially filled out the survey did so in 42%. We conducted interviews with four professionals from various fields and levels of expertise. There was a total of 140 minutes of audio recordings made from the interviews.

FINDINGS AND DISCUSSION

In this section, we show the answers to our research questions. We combine survey and interview results for each question to streamline the presentation.

Demographics

Figure 2 shows that the majority of survey participants with bachelor's degrees in computer science, software engineering, computer engineering, or electrical engineering are employed in the consumer electronics industry (27%), the information technology and telecommunications industry (22%), the automotive industry (20%), and the healthcare and biomedical industry (15%). also gained knowledge of ML methods.

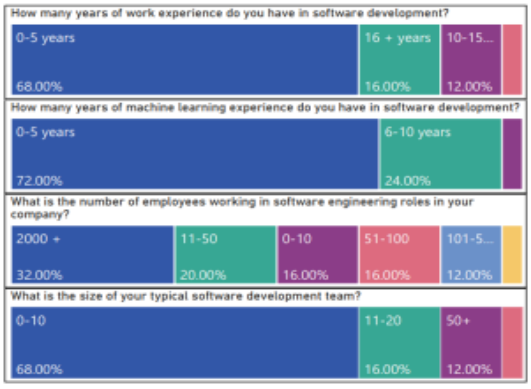


Figure 2: Demographics of survey respondents.

Table 1: Interview Subjects

Identifier	Role (Company type)	Experience
P1	Principal System Architect (HW vendor)	20 years
P2	Senior developer (HW vendor)	20 years
P3	Chief Architect (Start-up)	30 years
P4	ML Engineer (ML services)	3 years

from formal education at an academic institution (41%), self-study (37%), and on-the-job experience (20%). They are employed by businesses ranging in size from fewer than 50 people (36%) to more than 2,000 people (32%). Around 30% have more than 5 years of experience implementing ML in software engineering, whereas 70% have less than that amount of time under their belts. Nearly as many respondents said their firms have used ML in the early exploration/prototyping stage as said they had considerable expertise across many platforms (Figure 4). In Table 1, you'll see that our interviewees came from a wide variety of backgrounds and industries, including consumer electronics, military, medicine, and manufacturing.

DISCUSSION

Analogy to Previous Findings

In many ways, our results mirrored existing understanding. Participating teams used standard development tools, including ML frameworks like TensorFlow and Porch and toolchains based on the Visual Studio and Code IDEs. Each of our members engages in an ongoing process of growth. Edge-Cloud hybrids are becoming more common. It's well-known that IoT systems have difficulties on the fronts of power, memory, and computational limits. Participants in our events are well-versed on security concerns including data poisoning. Our results vary from those previously published mostly because of how we treated the topic of engineering expense. Our members, particularly those in the consumer electronics industry, compromise on safety in order to lower manufacturing costs. In a

https://zenodo.org/records/14505947

similar vein, the research literature proposes several engaging approaches to emulation, load balancing, and system validation, however the vast majority of respondents' organisations do not really use any of these strategies. Our participants, in contrast to the academics' ideals of invulnerable systems, weigh the trade-offs between the amount of security that is feasible (in terms of engineering expense) and necessary (relative to market demand). Typically, the engineering cost of suggested approaches is ignored in the research literature. Finally, there are more unreliable sources than we realised in the literature, including open-source code, academic research, and development toolchains.

Instructions for Professionals

As our research showed, there is a major chasm between academics' and professionals' views on IoT security. This has implications for training future cybersecurity professionals [7]. Secure development lifecycles are described by government recommendations (such as those from the US-NIST [5] and the EU ENISA [3]) outside of academia. For successful pre-deployment, deployment, and post-deployment phases, NIST [6] suggests doing a comprehensive analysis of the target audience, users, anticipated use cases, security concerns, and project objectives. Nobody in our sample mentioned anything like that happening. We were shocked to see that practitioners still put so much importance on code review and white-box analysis in their IoT systems, given the effectiveness of automated code analysis approaches like static analysis, black-box, and grey-box fuzzing in uncovering system vulnerabilities in IT software. We advise that these techniques be used in the field [44].

Research needs for the future

Given the difficulties reported by the practitioners we interviewed, we propose three avenues for more study. First, low-cost components and slim profit margins are typical of the Internet of Things. The technical investment required to properly secure IoT devices was a major issue for many of the people who participated in our study. Experts in the field of Internet of Things (IoT) system engineering would benefit greatly from researchers providing cost-aware engineering approaches for ensuring the security of IoT devices. Prior studies have mostly focused on balancing security with other expenses, including operating delay and energy usage [19, 65]. The results of our research highlight the significance of accounting for engineering expenses in addition to runtime effects. Our efforts also supplement the literature that aims to educate buyers about the impact security has on the price of common IoT gadgets. Second, developers and researchers alike use open-source tools and publicly available data to inform their machine learning models.

The development is sped up, but a great danger is introduced. We advise that researchers in the field of machine learning (ML) thoroughly describe their research prototypes and the constraints of their work, and that they may have a greater influence by joining community initiatives to construct exemplary ML models (e.g., Torch Vision [52] and the TensorFlow Model Garden [62]). More research is needed to figure out how to effectively duplicate and transmit ML knowledge [13, 17, 37]. Trustworthy software supply chains will increase IoT system security in general because of our members' heavy use of open-source technologies [64]. Third, one possible subject for study is the practitioners' struggles to comply with the requirements and constraints listed in Table 2. The effect of security compliance on the security outcomes of IoT applications, as well as the trade-off with engineering expense, may be investigated, for instance.

CONCLUSION

This study aimed to improve our current knowledge of machine learning and cyber security as they pertain to IoT engineering techniques. After conducting a survey and conducting interviews, we determined that the greatest difficulty engineers experience when developing an IoT device is striking a balance between engineering cost, performance, trust, and security. We discovered that businesses often rely on open-source and academic materials without verifying their legitimacy, even going so far as to include ML method prototypes developed in universities in their IoT offerings. One company even depends expressly on the open-source community to uncover vulnerabilities in their software, demonstrating the wide range of approaches to cybersecurity investment depending on available resources, engineering expense, and organisational goals. Academic studies on engineering practises and government recommendations that may help practitioners solve some of their issues have not yet been embraced. Many of the people we interviewed expressed worry about the expense of software engineering and cybersecurity projects, so we think it's important that researchers take that into account in the future.

REFERENCES

[1] 2016. Hackers Used New Weapons to Disrupt Major Websites Across U.S. https: //www.nytimes.com/2016/10/22/business/internet-problems-attack.html. Accessed June 08, 2021.

https://zenodo.org/records/14505947

[2] 2016. A Primer on Continuous Delivery. https://feeney.mba/a-primer-oncontinuous-delivery.html

[3] 2017. Baseline Security Recommendations for IoT. https://www.enisa.europa. eu/publications/baseline-security-recommendations-for-iot. Accessed June 09, 2021.

[4] 2017. Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared. https://www.nytimes.com/2017/07/25/technology/roomba-irobotdata-privacy.html. Accessed June 08, 2021.

[5] 2019. IoT Device Cybersecurity Capability Core Baseline. https://nvlpubs.nist. gov/nistpubs/ir/2020/NIST.IR.8259A.pdf. Accessed June 09, 2021.

[6] 2020. Foundational Cybersecurity Activitiesfor IoT Device Manufacturers. https: //nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8259.pdf. Accessed June 09, 2021.

[7] 2020. IoT Non-Technical Supporting Capability Core Baseline. https://nvlpubs. nist.gov/nistpubs/ir/2020/NIST.IR.8259b-draft.pdf. Accessed June 09, 2021.

[8] 2020. System brings deep learning to "internet of things" devices. https://news. mit.edu/2020/iot-deep-learning-1113.

[9] Yasemin Acar, Christian Stransky, Dominik Wermke, Charles Weir, Michelle L. Mazurek, and Sascha Fahl. 2017. Developers Need Support, Too: A Survey of Security Advice for Software Developers. Proceedings - IEEE Cybersecurity Development Conference, SecDev (2017). https://doi.org/10.1109/SecDev.2017.17

[10] Deniz Akdur, Vahid Garousi, and Onur Demirörs. 2018. A survey on modeling and model-driven engineering practices in the embedded software industry. Journal of Systems Architecture 91 (2018), 62–82. https://doi.org/10.1016/j.sysarc. 2018.09.007

[11] Sultan Alharby, Nick Harris, Alex Weddell, and Jeff Reeve. 2018. The Security Trade-offs in Resource Constrained Nodes for IoT Application. International Journal of Electrical, Electronic and Communication Sciences: 11.0. 12, 1 (2018), 56–63. https://www.researchgate.net/publication/322747058%0Ahttp:/ /www. waset.org/downloads/15/papers/18ae010177.pdf

[12] Omar Alrawi, Chaz Lever, Manos Antonakakis, and Fabian Monrose. 2019. SoK: Security Evaluation of Home-Based IoT Deployments. In 2019 IEEE Symposium on Security and Privacy, SP.

[13] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software Engineering for Machine Learning: A Case Study. In International Conference on Software Engineering: Software Engineering in Practice. https: //doi.org/10.1109/ICSE-SEIP.2019.00042

[14] Hala Assal and Sonia Chiasson. 2019. "Think secure from the beginning": A survey with software developers. Conference on Human Factors in Computing Systems - Proceedings (2019). https://doi.org/10.1145/3290605.3300519

[15] Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The Internet of Things: A survey. Computer Networks (2010). https://doi.org/10.1016/j.comnet.2010.05. 010

[16] Saurabh Bagchi, Tarek F. Abdelzaher, Ramesh Govindan, Prashant Shenoy, Akanksha Atrey, Pradipta Ghosh, and Ran Xu. 2020. New Frontiers in IoT: Networking, Systems, Reliability, and Security Challenges. IEEE Internet of Things Journal 7, 12 (2020), 11330–11346. https://doi.org/10.1109/JIOT.2020.3007690

[17] Vishnu Banna, Akhil Chinnakotla, and et al. 2021. An Experience Report on Machine Learning Reproducibility: Guidance for Practitioners and TensorFlow Model Garden Contributors. arXiv (2021).

[18] Iulia Bastys, Musard Balliu, and Andrei Sabelfeld. 2018. If This Then What?: Controlling Flows in IoT Apps. In Conference on Computer and Communications Security, CCS.

[19] Chiara Bodei, Stefano Chessa, and Letterio Galletta. 2019. Measuring Security in IoT Communications. Theoretical Computer Science 764 (April 2019), 100–124. https://doi.org/10.1016/j.tcs.2018.12.002

[20] Will Brackenbury, Abhimanyu Deora, Jillian Ritchey, Jason Vallee, Weijia He, Guan Wang, Michael L. Littman, and Blase Ur. 2019. How Users Interpret Bugs in Trigger-Action Programming. In Conference on Human Factors in Computing Systems CHI.

[21] Z. Berkay Celik, Earlence Fernandes, Eric Pauley, Gang Tan, and Patrick D. McDaniel. 2019. Program Analysis of Commodity IoT Applications for Security and Privacy: Challenges and Opportunities. ACM Comput. Surv. 52, 4 (2019).

[22] Z. Berkay Celik, Patrick D. McDaniel, Gang Tan, Leonardo Babun, and A. Selcuk Uluagac. 2019. Verifying Internet of Things Safety and Security in Physical Spaces. IEEE Secur. Priv. 17, 5 (2019).