Guidelines for Safe and Effective Engineering of Internet of Things Devices Utilizing Machine Learning

Mrs. K. RADHA[1], Mr. NALLURI SURYA NARAYANAMURTHY[2], Mr. MALE NAGESH[3]
ASSISTANT PROFESSOR[1,2,3]
DEPARTMENT OF ECE, SWARNANDHRA COLLEGE OF ENGINEERING AND TECHNOLOGY, NARASAPUR

ABSTRACT

An important part of the latest iteration of IoT systems is the use of machine learning (ML) technology on edge devices. This creates additional challenges in safeguarding user data and maintaining system integrity, in addition to substantial technical obstacles in bringing ML to hardware with low resources. To simplify development and increase product success, current research suggests iterative strategies for machine learning enabled IoT items. These processes fall back on the same old tricks employed in other, less specific, areas of software development rather than being adapted to the specific requirements of machine learning or Internet of Things devices. This research seeks to establish engineering processes and security practices for Internet of Things devices that are enabled by machine learning from the perspective of the engineering lifecycle. In order to get this information, we polled 25 working professionals and interviewed 4 more. The security engineering practises and methods used by various companies vary, as we found out. Respondents emphasized the need to balance the needs of the company with the technical expense of threat modeling and security research. Engineers won't put as much focus on security if it isn't required. Proponents of ML for IoT devices have long voiced worries about the possibility of reverse engineering and intellectual property theft. Our findings highlight the need for more investigation into the dynamics of the interaction between various technological restrictions, such as compliance, security, and cost.

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

INTRODUCTION

Linking edge devices ("Things") to one another and to more powerful resources across the network ("Internet") is the essence of the Internet of Things (IoT) concept, which unites the cyber and physical realms [15]. From its present 35 billion, the number of linked devices is projected to double by 2025, according to many sources [30, 57, 58]. With the use of machine learning (ML) [38, 39], IoT systems are able to make thoughtful decisions quickly [8, 67]. While the resulting sophisticated IoT systems may dramatically alter several economic sectors, they also pose serious risks. To lessen security risks, engineers should employ ML algorithms privately and securely on low-capacity Internet of Things devices [16]. Despite the increasing importance of intelligent IoT systems to consumers, companies, and governments, surprisingly little is known about the engineering procedures used by manufacturers [28, 46, 53].

Concerns about engineering practises have been raised by high-profile failures, such as attacks on waterworks systems that poisoned the water supply [55], aggressive data collection strategies [4, 48], and vulnerabilities that led to botnets on the Internet of Things [1]. Problems with Internet of Things (IoT) software [46] and security [12, 18, 20-23, 25, 34, 35, 47, 61] have been examined from a software perspective by researchers using failure analysis and programme analysis. Researchers have established broad models of the secure software development life cycle (SDLC) for both the building of ML models and edge devices enabled by ML [28, 53]. Nevertheless, the challenges of adoption in the real world and current industrial practices have received little attention. Therefore, we want to investigate the development phase integration of ML with IoT devices. Our main areas of research are: How do most businesses go about developing and managing machine learning-powered Internet of Things devices? What role does product life cycle security play? For this reason, we set out to solve these riddles by surveying 25 working professionals and interviewing 4 of them.

BACKGROUND

Data privacy and security concerns have motivated this examination into the increasingly common use of computer systems with network-edge intelligence. Although the term "Internet of Things" has not been agreed upon by all parties involved, we will use the following criteria: devices with limited memory, power, and computing capabilities; linked to a network; and sensors

and/or actuators. The Internet of Things (IoT) incorporates networking and sensor capabilities into low-budget devices [50, 70]. Engineering for the Internet of Things: Engineering techniques for Internet of Things (IoT) systems are infamously challenging because to their dispersed nature, finite resources, and mix of digital and physical components [68]. Our study was structured according to the normal developmental lifecycle for ML-based IoT devices, as illustrated in Figure 1. Prior research [2, 28, 53] is included into this lifetime. Here, an iterative five-stage process is shown for the engineering of the Internet of Things: There may be restrictions on the kinds of hardware and software that may be used since the product's intended use is defined. Decisions about the system's structure, architecture, and evaluation methods are made during the design process. Design decisions are put into effect via development frameworks.

Enhancing the performance of ML models may be achieved by hyperparameter tuning, reducing the computational complexity of the model (as is the case with models that rely on deep learning), and fine-tuning the network blocks [26, 40]. Instead of focusing on specific devices, the solution takes a profile-level view of hardware. Moving the final result to the target machine is what "deployment," the last stage of the process, is all about. To make the model fit the constraints of the Internet of Things device, optimizations like pruning are done during deployment [39]. Although they vary according to the capability of the target hardware, optimization method settings are often constant [53]. After engineers install software on hardware, it is their responsibility to verify that the system meets its requirements. Issues that may arise include performance goals, fault tolerance (see references 31 and 59), and potential security vulnerabilities. Engineers consider both generic threat models and ML-specific ones when they build systems. Attacks like exploiting corrupted training data [69] or breaking down a model [49] have been studied by researchers. Internet of Things Security: Ensuring the safety of developed systems is a major concern [51]. The incorporation of security measures is expanding into more and more phases of the engineering life cycle [41]. On the other hand, security is a challenging and time-consuming problem for many IoT system developers [46]. Security may not be well-defined within the technical team, despite their shared sense of responsibility for it [9, 45, 63]. Implementing security on devices with limited resources affects productivity in those areas [11, 60], and it is typical to prioritize meeting deadlines and maximizing usefulness above security [14, 24, 43]. Although there is a present knowledge gap in academia about industry processes, this engineering process model for ML-based Internet of Things development is encouraging. We can't solve the industry-wide problems and overcome the obstacles to building and maintaining safe ecosystems due to this lack of data. This study starts to fill that gap.

METHODOLOGY

Our study questions need an exploratory approach, which integrates quantitative and qualitative techniques to better understand a phenomena and to propose new lines of inquiry [54]. A survey provided us with high-level data, while in-depth interviews helped us comprehend specifics.

Survey

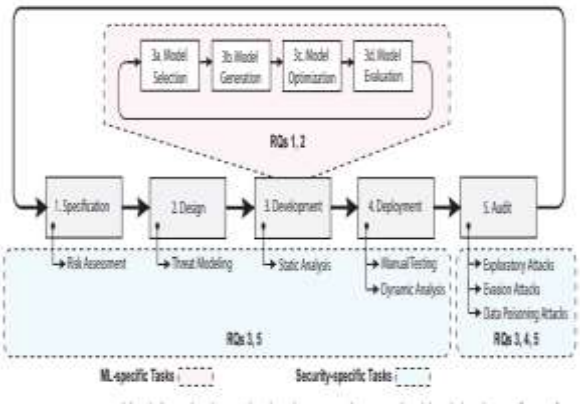
We created a 10-minute survey with 32 questions based on our research goals. Participants had 10 minutes to complete it. A total of seven demographic questions were taken directly from the aforementioned sources, while the other questions were developed using established guidelines for survey design [29]. The first set of questions was based on our professional background in applying ML to IoT devices; subsequent rounds of questioning were refined by discussions with industry specialists. We had two experts go over the survey and provide us feedback on its length and validity; we then adjusted it accordingly.

Distributing quizzes:

We announced the poll on many platforms, including public forums like Reddit, Hacker News, and Towards AI; our personal networks on LinkedIn and Facebook; and our department's mailing list, all in an effort to reach a wide audience and gauge interest in the engineering security approaches that were being studied. Snowball Sam pling [36] was another request we made to survey participants. We asked them to share the results with their colleagues. Following its publication in the last week of March 2021, the survey lasted for five weeks before being terminated. We gave them the opportunity to win a $50 gift card if they would just complete out the survey.

Analytical approach:

Specifically, we used Qualtrics reports to evaluate the data. All of the answers from each participant were taken into account while doing the analysis. For the purpose of uniformity, all survey data shown in the diagrams is given as a percentage of the total responses.



Interviews

Our method of conducting interviews evolved from an extension of the survey questions. We monitored survey responses and developed additional questions in areas where people had strong opinions or unexpected answers. Every interviewee was given 30–40 minutes to speak while responding to 8 questions that were already prepared [27]. In order to test the feasibility and efficiency of the interview process, we ran it by one clinician. Those who participated in our survey were the ones from whom we compiled our interviews. The survey takers' expertise in machine learning and IoT engineering made them eligible for a follow-up interview. In order to encourage people to volunteer for a follow-up interview after they finished the survey, we offered them a $25 gift card. We followed up with all of the interested parties and interviewed those who were willing to be interviewed. Protecting participants' privacy: The audio of the interviews was transcribed by an external company. In order to ensure the privacy of our participants, their personal information was masking before analysis.

Data Obtained through Survey

Only fourteen out of twenty-five respondents gave the survey their full attention. Because the response rate was so low, we also looked at data from incomplete replies. Half of the people who took the survey did so in a median fashion (42%). We spoke with four experts from different backgrounds and areas of expertise. The interviews were audio recorded for a duration of 140 minutes.

FINDINGS AND DISCUSSION

The section presents the results of our investigation. To make the display easier to understand, we combined survey and interview findings for each topic.

Demographics

With a bachelor's degree in computer science, software engineering, computer engineering, or electrical engineering, the majority of survey participants work in the following industries: consumer electronics (27%), information technology and telecommunications (22%), automotive (20%), healthcare and biomedical (15%), and the automotive industry, as shown in Figure 2. I learned ML techniques as well.

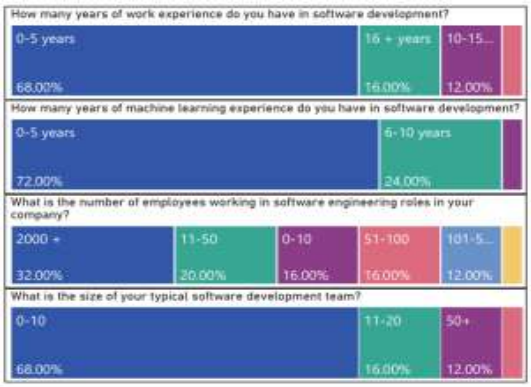


Figure 2: Demographics of survey respondents.

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

via a combination of classroom instruction (41%), independent study (37%), and work experience (20%). Companies employing them range in size from 32 (more than 2,000 employees) to 36 (less than 50 individuals). Thirty percent of software engineers have worked with ML for more than five years, whereas seventy percent have worked with it for less than five years. The number of respondents claiming to have extensive knowledge across many platforms is almost equal to the number claiming that their companies had used ML at the early exploration/prototyping stage (Figure 4). As you can see from Table 1, our respondents are from many walks of life and work in fields as diverse as consumer electronics, manufacturing, medicine, and the military.

DISCUSSION

Comparison with Prior Discoveries

Existing knowledge was reflected in our findings in several respects. Team members used industry-standard programming tools, including as ML frameworks like Porch and TensorFlow, as well as Visual Studio and Code IDE-based toolchains. Everyone here is always trying to better themselves. More and more, hybrids of the edge and cloud are cropping up. Problems with power,

memory, and computational restrictions are well-known to plague IoT devices. Data poisoning and other security issues are topics that attendees at our events are familiar with. The main reason our findings differ from other publications is the way we addressed the issue of engineering expenditure. Many of our members, especially those working in consumer electronics, cut corners on safety measures to keep production costs down. Similarly, most respondents' companies do not really implement any of the research-proposed attractive methods for emulation, load balancing, or system validation. Our participants consider both the practical (in terms of engineering cost) and essential (related to market demand) levels of security, in contrast to the academics' ideas of unbreakable systems. The research literature often disregards the engineering expense of proposed alternatives. Lastly, we were surprised to see how many untrustworthy sources, such as open-source code, academic research, and development toolchains, there are in the literature.

Guidelines for Experts

Our findings reveal a wide gap in opinion on the topic of Internet of Things security between academics and industry experts. This might affect how cybersecurity education is structured in the future [7]. Recommendations from government agencies, such as the US-NIST [5] and the EU ENISA [3]—not affiliated with academia—describe secure development lifecycles. Comprehensive examination of the target audience, users, expected use cases, security risks, and project goals is suggested by NIST [6] for effective pre-deployment, deployment, and post-deployment stages. Not a single person in our sample reported anything similar occurring. Given the efficacy of automated code analysis methodologies such as static analysis, black-box, and grey-box fuzzing in identifying system vulnerabilities in IT software, we were startled to learn that practitioners still place such a high value on code review and white-box analysis in their IoT systems. These methods should be implemented in the field, as we have advised [44].

Future research requirements

We suggest three potential research directions in light of the challenges mentioned by the practitioners we surveyed. To begin, IoT products often have cheap parts and small profit margins. Many individuals who took part in our survey expressed serious concerns about the time and effort needed to educate themselves on how to adequately protect Internet of Things (IoT) devices. Researchers offering cost-aware engineering techniques to guaranteeing the security of IoT devices would be a tremendous boon to experts in the area of Internet of Things (IoT) system engineering. Balancing security with other costs, such as operational delay and energy use, has been the primary focus of prior research [19, 65]. Our study shows that engineering expenditures, together with runtime consequences, must be considered. The literature that seeks to inform consumers about the effect of security on the cost of popular IoT devices is supplemented by our efforts as well. Second, while building machine learning models, both developers and academics rely on open-source software and freely accessible data.

While progress is accelerated, a significant threat is also presented. To make a bigger impact, ML researchers should join community efforts to build exemplary ML models (e.g., Torch Vision [52] and the TensorFlow Model Garden [62]) and provide detailed descriptions of their research prototypes and the limitations of their work. Methods for efficiently duplicating and transmitting ML knowledge need more investigation [13, 17, 37]. since a whole, the security of IoT systems will be enhanced by trustworthy software supply chains, since our members heavily use open-source technology [64]. The third point is that the difficulties practitioners have in meeting the criteria and limitations shown in Table 2 might be the topic of future research. Examples of possible areas to explore include the relationship between security compliance and the security outcomes of IoT applications, the trade-off with engineering expenditure, and so on.

CONCLUSION

Our present understanding of cyber security and machine learning as they relate to engineering methodologies for the Internet of Things (IoT) was the intended focus of this research. Finding a happy medium between engineering cost, performance, trust, and security is the biggest challenge engineers face when building an IoT device, according to our study and interviews. We found that many companies use academic and open-source resources without checking their credibility; some even include university-developed ML technique prototypes into their Internet of Things (IoT) products. Depending on available resources, engineering expenditure, and organizational objectives, there is a broad variety of ways to cybersecurity investing. One corporation even relies on the open-source community to find software vulnerabilities. Engineering practitioners still haven't fully embraced academic research on current practises or government suggestions that may assist them address their challenges. It is crucial for academics to consider the concerns raised by many interviewees about the cost of software engineering and cybersecurity initiatives moving forward.

REFERENCES

[1] 2016. Hackers Used New Weapons to Disrupt Major Websites Across U.S. https: //www.nytimes.com/2016/10/22/business/internet-assault-problems.html site. Viewed on June 8, 2021.

[2] A Comprehensive Guide to Continuous Delivery, 2016. A primer on continuous delivery can be found at https://feeney.mba.

the third 2017. Baseline Security Recommendations for IoT. "Basis-security-recommendations-for-iot" may be found at https://www.enisa.europa. eu/publications. Viewed on June 9, 2021.

[4] year 2017. While it's mapping your home, your Roomba might be gathering data that could be shared with others. Visit https://www.nytimes.com/2017/07/25/technology/roomba-irobotdata-privacy.html for this article. Viewed on June 8, 2021.

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

[6] 2020. Cybersecurity Groundwork for Internet of Things Device Producers. June 09, 2021 access to the following document: https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8259.pdf.

[7] 2020. Core Baseline for IoT Non-Technical Supporting Capabilities. You may access this document at this link: https://nvlpubs. nist.gov/nistpubs/ir/2020/NIST.IR.8259b-draft.pdf. The access date is June 09, 2021.

To "internet of things" gadgets, the system introduces deep learning in 2020. here: https://news.mit.edu/2020/iot-deep-learning-1113.

[9] Authors: Yasemin Acar, Christian Stransky, Dominik Wermke, Charles Weir, Michelle L. Mazurek, and Sascha Fahl. 2017. Additionally, Developers Require Assistance: A Review of Security Guidance for Software Engineers. Conference on Cybersecurity Development (SecDev) by the IEEE (2017). The website for this article is https://doi.org/10.1109/SecDev.2017.17.

[10] Onur Demirörs, Vahid Garousi, and Deniz Akdur (2018). An examination of the embedded software industry's model-driven engineering techniques and modeling efforts. Publication date: 2018 Jun; volume: 91; pages: 62–82. The link to the article is https://doi.org/10.1016/j.sysarc.2018.09.007.

the eleventh Citation: Alharby, Sultan, Harris, Nick, Weddell, and Reeve, Jeff. 2018. Conflicts between security and resource constraints in Internet of Things nodes. Published in the International Journal of Electrical, Electronic and Communication Sciences, volume 11.0, issue 1, pages 56–63, in 2018. This sentence cannot be paraphrased as it is a link to an online publication.URL: http://www.waset.org/downloads/15/papers/18ae010177.pdf

Alrawi, Omar, Chaz, Antonakakis, Manos, and Monrose, Fabian [12]. The year 2019. Safeguards for Internet of Things (IoT) Installations in the Home. The 2019 IEEE Symposium on Security and Privacy (SP) follows.

Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann are the authors of the cited work. The year 2019. Examining Software Engineering in the Context of Machine Learning. Session on Software Engineering in Action at the International Conference on Software Engineering. Doi: 10.109/ICSE-SEIP.2019.00042

[14] By Hala Assal and Sonia Chiasson. The year 2019. "Consider security first": a poll of programmers. Proceedings of the 2019 Conference on Human Factors in Computing Systems. This article is available online at this link: https://doi.org/10.1145/3290605.3300519.

[15] Featuring Giacomo Morabito, Luigi Atzori, and Antonio Iera. 2010. An overview of the Internet of Things. Internetworking (2010). DOI: 10.1016/j.comnet.2010.05. 010

[16] The following individuals are listed: Tarek F. Abdelzaher, Ran Xu, Akanksha Atrey, Prashant Shenoy, Ramesh Govindan, and Saurabh Bagchi. 2020. Uncharted Territory in the Internet of Things: New Difficulties in Systems, Reliability, and Security. 12(2020), 11330-11346, IEEE Internet of Things Journal, 7, 12. Please find the following link: https://doi.org/10.1109/JIOT.2020.3007690.

[17] Vishnu Banna, Akhil Chinnakotla, and others.... the year 2021. For practitioners and contributors to the TensorFlow Model Garden, this report details their experiences with machine learning reproducibility. in the year 2021.

[18] in Andrei Sabelfeld, Iulia Bastys, and Musard Balliu. 2018. This Then What?: Managing Flows in Internet of Things Applications. Presented at the CCS Conference on Computer and Communications Security.

[19] Gorgonzola, Letterio, and Bodei, Chiara. The year 2019. Measuring Security in IoT Communications. (April 2019), 100-124, Theoretical Computer Science 764. DOI: 10.1016/j.tcs.2018.12.002

In their work, [20] Will Brackenbury, Jillian Ritchey, Jason Vallee, Guan Wang, Weijia He, Michael L. Littman, and Blase Ur are cited. The year 2019. The Meaning of Trigger-Action Programming Bugs to End Users. Published at the CHI Conference on Human Factors in Computing Systems.

[21] Patrick D. McDaniel, Gang Tan, Earlence Fernandes, Earlence Z. Berkay Celik, and Eric Pauley. The year 2019. Improving the Security and Privacy of Common Internet of Things Applications: Possible Obstacles and Solutions. (2019) ACM Computing Surveys 52, 4.

[22] A. Selcuk Uluagac, Gang Tan, Z. Berkay Celik, Leonardo Babun, and Patrick D. McDaniel in collaboration. Validating the Security and Safety of Internet of Things in Real-World Environments in 2019. (2019) 17:5 in IEEE Security and Privacy.