

“We just did not have that on the embedded system”: Insights and Challenges for Securing Microcontroller Systems from the Embedded CTF Competitions

Zheyuan Ma
CactiLab, University at Buffalo
Buffalo, NY, USA

Gaoxiang Liu
CactiLab, University at Buffalo
Buffalo, NY, USA

Alex Eastman
CactiLab, University at Buffalo
Buffalo, NY, USA

Kai Kaufman
Worcester Polytechnic Institute
Worcester, MA, USA

Md Armanuzzaman
CactiLab, Northeastern University
Boston, MA, USA

Xi Tan
CactiLab, University of Colorado
Colorado Springs
Colorado Springs, CO, USA

Katherine Jesse
Worcester Polytechnic Institute
Worcester, MA, USA

Robert Walls
Worcester Polytechnic Institute
Worcester, MA, USA

Ziming Zhao
CactiLab, Northeastern University
Boston, MA, USA

Abstract

Microcontroller systems are integral to our daily lives, powering mission-critical applications such as vehicles, medical devices, and industrial control systems. Therefore, it is essential to investigate and outline the challenges encountered in developing secure microcontroller systems. While previous research has focused solely on microcontroller firmware analysis to identify and characterize vulnerabilities, our study uniquely leverages data from the 2023 and 2024 MITRE eCTF team submissions and post-competition interviews. This approach allows us to dissect the entire lifecycle of secure microcontroller system development from both technical and perceptual perspectives, providing deeper insights into how these vulnerabilities emerge in the first place.

Through the lens of eCTF, we identify fundamental conceptual and practical challenges in securing microcontroller systems. Conceptually, it is difficult to adapt from a microprocessor system to a microcontroller system and participants are not wholly aware of the unique attacks against microcontrollers. Practically, security-enhancing tools, such as the memory-safe language Rust, lack adequate support on microcontrollers. Additionally, poor-quality entropy sources weaken cryptography and secret generation. Additionally, our findings articulate specific research, developmental, and educational deficiencies, leading to targeted recommendations for researchers, developers, vendors, and educators to enhance the security of microcontroller systems.

Keywords

Embedded Systems Security, Secure Firmware Development, Capture the Flag (CTF) Competitions, Embedded Rust

1 Introduction

A microcontroller (MCU) is a compact, integrated circuit designed specifically for control tasks within embedded systems and Internet of Things (IoT) devices. Unlike general-purpose processors found in computers, a microcontroller integrates a processor core (CPU), memory (RAM and ROM), and peripheral interfaces—such

as timers, Analog-to-Digital Converters (ADC), and communication modules—onto a single chip. It is commonly used to execute firmware, which is specialized, often real-time software that manages and controls operations within embedded systems. Compared to the microprocessors used in smartphones, tablets, and desktops, microcontrollers operate at lower frequencies and have smaller memory capacities. Microcontrollers are found in a wide range of applications, including vehicles, medical devices, and industrial control systems. To highlight their prevalence, a typical mid-range automobile contains about 30 microcontrollers [25], and in 2021 alone, global shipments of microcontrollers reached approximately 31.2 billion units [3].

However, designing and implementing secure microcontroller systems is challenging. They often lack features that are standard in microprocessor architectures. For example, microcontrollers do not normally include a Memory Management Unit (MMU), which is used to implement privilege isolation, fine-grained memory access control, Address Space Layout Randomization (ASLR) [54], and many other security features. Additionally, these systems are often programmed in low-level languages like C and assembly, which lack safety features and are prone to memory corruption bugs [56–58].

Given their prevalence and use in critical applications, it is important to classify the challenges in designing and implementing secure microcontroller systems. Unfortunately, the closed-source nature of most embedded and IoT systems presents substantial barriers to comprehensive analysis. Even procuring real-world firmware samples is difficult, as highlighted in existing literature [44, 57, 66]. Additionally, we argue that understanding and addressing these challenges requires more than just *technical* analysis and solutions; it also involves grasping developer *perceptions*.

For this study, we adopted a unique approach by classifying security challenges in microcontroller systems through the lens of the MITRE Embedded Capture the Flag (eCTF) competition [15], an annual, months-long event. Each year, participants are given a theme, an insecure reference system, and specific microcontrollers to develop upon. Security is given top priority in the competition. Therefore, the security-related mistakes observed in eCTF are often

fundamental, making them highly likely to appear in real-world development, where security is frequently not as prominently prioritized [38, 62]. Indeed, many of our findings on vulnerabilities have been independently observed in research on real-world firmware analysis [44, 57].

In contrast to existing studies that typically focus on a single aspect, such as firmware analysis [44, 57, 66], and lack the capability to examine the broader perspective, our study takes a more comprehensive approach. Through the eCTF lens, we had the unique opportunity to explore the entire lifecycle of secure microcontroller system development – from design documents and source code to binary analysis and developer perceptions. As a result, we not only confirm the presence of vulnerabilities but also, for the first time, gain first-hand insights into how these vulnerabilities emerge in the first place.

Furthermore, the eCTF competition primarily attracts undergraduate students and graduate researchers at the early stages of their firmware development and security research careers. Therefore, our insights highlight key areas for enhancing security research and education for embedded developers. By addressing these gaps, we aim to equip the embedded systems development and research community with the knowledge to prevent recurring firmware security issues and guide future research directions, fostering new approaches to securing embedded systems.

Our study includes two sources of data: team submissions and post-competition interviews. To make our submission analysis broad and thorough, we attempted to identify and understand security-related mistakes and omissions made by teams. To accomplish that, we manually reviewed source code, documentation, and build tools, and we compiled every submission and examined relevant disassembly from the output. By finding omissions and mistakes, we learned about the real challenges that participants faced. We complemented the analysis with one-on-one interviews over Zoom. Whereas the source code and documentation may tell us where mistakes exist, they cannot tell us why they exist. Therefore, we used the interviews to gain a deeper understanding of participants’ security acumen and to gauge whether or not they were aware of their mistakes and omissions. This complementary approach proved to be a powerful tool for developing deep insights into which challenges are faced and why.

We break down our findings into two main categories. In the first category, we detail the *conceptual* challenges that participants faced. These challenges are the result of a lack of knowledge or a misunderstanding. In the second category, we detail the *practical* challenges, which exist even when there is abundant knowledge available. This dichotomy of results is useful because it allows the problem of securing microcontrollers to be approached from two sides. Conceptually, researchers, educators, and companies should explore better ways to bridge the knowledge gaps faced by embedded system developers. Practically, researchers and vendors should develop new methodologies and tools that not only identify and address security shortcomings but also lower the barrier for their deployment.

Our results highlight three main conceptual challenges and two main practical challenges in securing microcontroller systems. Conceptually, there is a lack of knowledge about foundational security principles, it is difficult to adapt from a microprocessor system to a

microcontroller one, and participants are not wholly aware of the unique attacks against microcontrollers or their defenses. Practically, tools that naturally enhance security, like the memory-safe language Rust, lack sufficient support on microcontrollers, and we additionally find that the lack of high-quality entropy sources leads to less secure cryptography and secret generation. The contributions of this paper are as follows:

- We present an approach to studying the challenges in securing microcontroller systems through the lens of CTF competitions, which provides an opportunity to examine the entire lifecycle of the microcontroller system development from both technical and perceptual perspectives;
- By combining in-depth technical analysis with interviews, we uncovered both experiential and systemic security challenges, revealing key conceptual and practical difficulties in securing microcontroller system development;
- We offer actionable recommendations for researchers, developers, vendors, educators, and tool maintainers to address the identified challenges, bridge existing gaps, and strengthen the security of embedded systems.

2 Background: MITRE eCTF Competition

Competition overview. The eCTF is an annual, semester-long competition organized by MITRE where teams design, build, and attack “secure” embedded software for a given microcontroller platform. Each competition has a topic, such as a firmware update system or an unmanned aerial vehicle package delivery system. The competition consists of three phases: *design/implementation*, *handoff*, and *attack* (Figure 1 in Appendix A illustrates the process). Teams have just over four months to finish them. In the design/implementation phase, teams are tasked with creating “secure” embedded software based on functional and security requirements. Teams can use the provided reference design as a starting point or create their own design from scratch. In the handoff phase, the event organizers verify that the functional requirements are met for the submitted source code. Flags are then embedded into the firmware and must be protected by the defense mechanisms employed by the team. In the attack phase, teams try to capture each other’s flags by exploiting security weaknesses.

Themes in 2023 and 2024 competitions. In the 2023 eCTF, teams were assigned the roles of car companies and were tasked with developing two sets of firmware for cars and key fobs with a remote keyless entry feature. The firmware of the car and fob runs on two development boards. The 2024 eCTF focused on an insulin pump system consisting of one controller and two components: a blood sugar monitor and a pump actuator, all operating on three boards and communicating through an I²C bus.

Reference design. In both years, the organizers provided a reference design written in C as a starting point for competitors. The reference design fulfills all the functional requirements but has no security features. For example, all communication is in plaintext, with no defenses against hardware attacks, and several exploitable buffer overflow vulnerabilities exist in the provided functions.

Microcontroller platform. The 2023 competition used a TI TM4C123GXL development board [34] equipped with two 80 MHz ARM Cortex-M4F microcontrollers. The system has 256 KB of flash

memory, 32 KB of static random access memory (SRAM), and a 2 KB EEPROM. The 2024 competition used an Analog Devices Inc. MAX78000FTHR development board [17] equipped with a 100MHz ARM Cortex-M4 microcontroller and a 60MHz RISC-V co-processor. The board has 512KB of flash, 128KB of SRAM, and no EEPROM.

Threat model. The threat model in the competition closely mirrors real-world scenarios involving embedded and IoT devices. The attacker is presumed to have physical access to the board and the communication channels between boards, enabling potential physical tampering besides software-based and network-based attacks. Additionally, the attacker has access to the firmware’s source code. It is important to note, however, that the source code does not include any secrets or flags. These elements are generated and embedded into the protected firmware by the organizer, separate from the source code made available to the participants.

3 Research Methodology

We adopted a two-pronged approach: analysis of competition submissions and interviews with participants. The analysis and interview only include teams that entered the *attack* phase, meaning they had a functionally correct submission reflecting their best effort to secure it.

3.1 Submission Analysis

We analyzed 47 unique team submissions, with 20 from the 2023 competition and 27 from 2024, referring to individual teams as T1, T2, and so on. This includes *all* teams that passed the organizers’ functionality tests and entered the attack phase. The authors of this study participated separately as teams T1 and T2, competing independently before later collaborating to summarize and develop a taxonomy of the findings. The authors’ submissions were part of the statistical analysis for completeness; however, to avoid potential bias, no examples or case studies in the paper derive from the authors’ own submissions. The submissions consisted of source code, documentation, and build instructions, and we additionally had access to teams’ posters [23] and presentations [21, 22].

The criteria for submission analysis were informed by our experience with embedded security and participation in the eCTF over several years. Through this experience, we have observed common security practices and where defenses and attacks typically occur in MCU-based systems, informing the selection of the following sections for review.

Build tools. We looked at submissions’ Makefiles [26], linker scripts, and documentation, and compiled each submission. We recorded their chosen programming language, compiler, optimization level, security-related compiler flags and linker script attributes, and any warnings issued by the compiler during compilation.

Source code and disassembly. We manually inspected teams’ source code and disassembly to learn about the defense mechanisms they used and whether their behavior matched expectations. In addition to referencing their documentation, we used `git diff` to show changes from the reference files to ensure that we inspected all the code that was changed. In addition, we examined the disassembly of the compiled firmware to identify potential compiler optimizations that could negatively impact the system’s security.

Table 1: Interview Participant Demographics.

ID	2023/ 2024*		Major [§]	Edu Level [†]	CTF Pre-Exp	Pre-Sec Courses [‡]	Embed. Pre-Exp
P1	Rust	Rust	CS	UG 2/3	Very high	Yes	None
P2	C	C	CS	UG 2/3	Very high	Yes	None
P3	Rust	-	CS,CY	Master	High	Yes	Limited
P4	C	Rust	CS	Ph.D.	High	Yes	None
P5	-	C	CS	UG 3	None	Yes	None
P6	-	C	CY	UG 3	Limited	Yes	Limited
P7	C	C	CS,MA	UG 1/2	Medium	No	None
P8	-	C	CS	Master	None	Yes	Limited
P9	C	-	CS	UG 2	Very high	No	Limited
P10	-	C	CE	UG 1	Limited	No	None
P11	-	C	CLS	UG 3/4	None	No	Limited
P12	-	C	CY	UG 4	Limited	Yes	Limited
P13	2018-2021, C		CY	MS/Ph.D.	Very high	Yes	High
P14	C	-	CY	Ph.D.	None	No	None
P15	-	C	CE,CY	Master	Limited	Yes	Limited
P16	C	-	CS	Ph.D.	None	Yes	Limited
P17	-	C	CE	Master	None	No	High
P18	Rust	Rust	CS	UG 1/2	Medium	No	Limited
P19	-	C	CY	Master	Very high	Yes	Limited
P20	-	C	CS	UG 3	None	No	None
P21	-	C	CE	Master	None	Yes	High
P22	-	C	CS	UG 4	Medium	Yes	Limited

*: Indicates the programming language used by participant’s team. “-” means the no attendance. §: CY: Cybersecurity; MA: Math; CE: Computer Engineering; CLS: Criminology, Law and Society. †: UG 2/3 means the participant was in undergraduate 2nd and 3rd year during 2023 and 2024 eCTF. ‡: Indicates whether participant had taken computer security related courses before the eCTF.

We also paid attention to specific keywords related to inline assembly, random number generation, and timing in the source code and comments, such as `asm volatile`, “random”, and “delay” while we were examining the code. We looked at random number generation because finding a reliable entropy source on a microcontroller can be challenging, and using an unreliable source can make the system less secure. We investigated timing because adding random delays on a system in which an attacker has physical access can help reduce the efficacy of side-channel attacks and glitching. Additionally, we used debugging tools to gather runtime output for any uncertainty from the static analysis.

3.2 Interviews with Participants

We conducted 22 semi-structured interviews with participants, which included a mix of undergraduate, Master’s, and Ph.D. students. Among these, 8 had never taken computer security courses, and 14 had some experience in developing embedded systems. Table 1 provides demographic information of our participants. No interviewees were from the authors’ teams.

Ethical considerations. We collaborated with the Institutional Review Boards (IRBs) at each of our institutions to ensure adherence to ethical guidelines, including informed consent, the right to withdrawal, and the anonymization of Personally Identifiable Information (PII) to protect participant privacy and confidentiality.

After a thorough review, both of our IRBs determined that an IRB exemption was appropriate for this study.

Participant recruitment. We conducted two rounds of participant recruitment to gather interviewees for our study. Both rounds focused on attack phase participants who had competed at least once in the eCTF competitions from 2021 to 2024. The first round of recruitment took place during the 2024 eCTF award ceremony [60] in April 2024, while the second round was conducted in December 2024, primarily through email outreach to expand our sample size. For all interested participants, we provided detailed study information and a consent information sheet outlining the study’s purpose, the voluntary nature of participation, and the measures taken to ensure confidentiality and data security. We recruited 10 participants in the first round and 12 in the second round, resulting in a total of 22 participants.

Interview procedure. The interviews were conducted remotely via Zoom from May to June 2024 for the first round and from January to February 2025 for the second round. They varied from 38 to 107 minutes, with an average of 74 minutes. In addition to the consent process, participants were requested to fill out an online demographic survey form before the interview to streamline the interview process, which can be found in Appendix C. This form collected essential background information, such as their academic degree, area of study, and prior experience with security courses and competitions, particularly those relevant to CTF or embedded systems. Information regarding their role within their team and contributions during the competition was also gathered.

The main body of the interview was driven by a pre-defined set of questions detailed in Appendix D, focusing on the challenges participants faced, the strategies employed to secure their systems, and their understanding of security principles and tools. The interview questions were derived from the submission analysis to examine recurring issues. Interviews were conducted in a one-to-one semi-structured way with the option to opt out of any questions. We compensate each interview participant with a \$50 gift card.

Data collection and analysis. Data collection was conducted through recorded Zoom meetings. The audio recordings were transcribed to text using Otter AI [46] services without PII. We informed all participants about the privacy practices of Otter AI and obtained their consent to use the service. To ensure the transcripts’ fidelity, one of our study members reviewed and proofread each transcript against the audio recordings, correcting any discrepancies to preserve the semantic integrity of the participant responses.

Following transcription, the data analysis process began with the development of a preliminary coding scheme. Guided by established methods [51], three of our study members initially coded a single transcript independently to foster a diverse range of codes reflecting the intricacies of the interviews. These initial codes were then discussed collectively to formulate an agreed-upon codebook.

The initial codebook included categories that encapsulated the challenges faced by participants, the strategies they employed to address them, and any unique insights they shared. This codebook guided the coding of subsequent transcripts, together with all three study members. Weekly meetings were held to discuss the coding process, resolve any conflicts, and refine the codebook. This collaborative approach negated the need for formal inter-rater reliability checks, as the codebook evolved through comprehensive consensus

among the coders [43]. The final analysis resulted in a codebook consisting of 9 themes, 44 sub-themes, and 246 codes.

3.3 Threats to Validity

In assessing the validity of our study, several limitations must be acknowledged.

First, while the eCTF competition reflects many aspects of real-world embedded system security challenges, it remains a competition, and gamification elements such as point maximization may influence participant behavior in ways that deviate from real-world scenarios. Even if our findings on vulnerabilities and mistakes align with those in real-world firmware, we do not claim that our conceptual and technical insights will fully generalize to other development environments.

Second, our focus on practical challenges in securing microcontroller systems meant that broader organizational aspects of the competition, such as team collaboration dynamics and the competition structure, were beyond the scope of our analysis.

Third, the categorization of challenges into conceptual and practical themes was intended to provide clarity and structure to the findings. However, some sub-themes may overlap or extend beyond these categories, potentially introducing nuances not fully captured in the framework. Additionally, while the study primarily centered on identified challenges, we included insights that were deemed beneficial to the community, which may have broadened the scope beyond the initial framework.

Fourth, the reliance on self-reported data collected through interviews introduces the possibility of social desirability bias. Participants may have presented themselves or their teams in a more favorable light, which could affect the accuracy of the data.

Finally, the sample size, while typical for qualitative research, was relatively small and limited to participants from the eCTF competition. This also restricts the generalizability of our findings to a broader population of microcontroller developers or other embedded systems professionals.

4 Conceptual Challenges

Conceptual challenges stem from gaps in knowledge or misunderstandings, and interviews are a highly effective tool for gaining insights into these issues. Our analysis highlighted three key conceptual challenges in security: security principles, platform adaptation, and hardware attack and defense.

4.1 Security Principles

Our study reveals two significant gaps in security principles, namely privilege separation and memory wiping, which are basic but effective mechanisms for securing a system.

4.1.1 Privilege Separation. Privilege separation is a security design principle that involves dividing a program or system into distinct components, each with different levels of privilege. Cortex-M microcontrollers support privilege separation by offering hardware features such as the Memory Protection Unit (MPU) and distinct privileged and unprivileged execution modes. The MPU can enforce read-only and non-executable memory and restrict access to configurable memory regions depending on execution mode. These

features must be enabled in code, either manually by the developer or through adequate support from the operating system.

Submission Analysis: We analyzed teams’ submissions to understand whether and how they implemented privilege separation in their designs. We found that even though privilege separation is a fundamental security concept, no teams in either year utilized it. This is particularly notable given that this concept is frequently emphasized in security courses and widely recognized in the security community. Moreover, recent studies have focused on making privilege separation easier to implement, more secure, and more efficient on microcontrollers. For example, Kage [20], a compiler and FreeRTOS-based kernel, enhances control-flow protection by isolating the kernel. Similarly, other research, such as Silhouette [68], ACES [12], and EPOXY [13], target bare-metal systems by lowering the privilege level of specific code segments.

eCTF vs. real-world firmware: Privilege separation is rarely implemented in real-world microcontroller devices, appearing in only 1.78% of firmware samples [57]. Understanding the reasons behind this is crucial.

Interviews: Through our interviews, we sought to gauge participants’ understanding of privilege separation, including whether they had any familiarity with it. When participants were aware of privilege separation, we sought to understand why they didn’t use it in their design.

Many participants showed limited awareness and understanding of privilege separation on microcontrollers, often deterred by its perceived complexity. While a few recognized the concept of least privilege, the intricacies of implementing it within the competition’s timeframe or due to unfamiliarity with the microcontroller’s low-level operations led to the neglect in their design strategies.

19 out of 22 participants were unaware of the privilege separation feature in Cortex-M devices. For instance, participant P9 understood the concept of least privilege but did not know that Cortex-M supports privilege separation. They mentioned that even if they had become aware of it, they still unsure if it was necessary for their design. Similarly, participant P13 acknowledged that privilege separation could help in certain edge cases but argued that with this added complexity in design, “*you will still potentially lose things.*”

For participants like P2, even though they understood the concept of privilege separation, the low level at which microcontrollers are programmed caused them to overlook it: “[*it*] *didn’t even come into my mind.*” They continued:

“Because I don’t know how the thing is working behind the scenes, I would just assume the level of privilege wouldn’t work in my mind.”

Participants assumed that privilege separation is only useful in an OS-based environment and questioned its relevance in a bare-metal system. P19 and P11 were both unsure about the necessity of privilege separation in a bare-metal system. P19 thought that privilege separation would be useful “*if you have different processes and threads, different levels of execution while you’re designing the system.*” Since their system followed a flat, monolithic design, they did not see the need for privilege separation:

“It wasn’t like we were trying to execute someone else’s code in an unprivileged context, and protect it from the unprivileged one. So I don’t think it was really ... useful.”

Participants believed that privilege separation would not further enhance their system’s security. Participant P12 acknowledged that using both privilege levels is generally a good security practice but did not identify any specific vulnerabilities in their design that would have been mitigated by it. P19’s team prioritized avoiding implementation bugs over security mechanisms like privilege separation, assuming that a bug-free system would mitigate security risks:

“... the philosophy was, just don’t have any bugs, and then you don’t have to have any mitigations.”

Similarly, P4 mentioned that while theoretically useful, privilege separation might not significantly enhance security in environments where “*you have physical access already as the attacker.*” They weighed the attributes of the system against the expected benefit of privilege separation, and concluded that privilege separation would not be worth the effort.

Participants who had never heard of privilege separation tended to endorse it. While privilege separation can enhance security, its effectiveness is not guaranteed and heavily depends on careful implementation and system context. Introducing it may lead to compatibility challenges and potential new vulnerabilities if not properly managed. Participant P10 said, “*it probably would be a good way of making it [the system] more secure,*” and participant P8 was also inclined to favor it:

“I didn’t know that existed. Yeah, that’s really, really cool feature. We probably would have sent the team to go hunt for that if we had known existed.”

Recommendation 1: Researchers should investigate barriers to privilege separation adoption, develop automated enforcement tools, and collaborate with educators and vendors to bridge theory and practice. **Vendors** should enhance their support by providing comprehensive documentation or demonstration projects that illustrate privilege separation on their devices. **Educators** should emphasize the importance of least privilege in system- and security-related courses, including specific strategies for implementing privilege separation in embedded systems.

4.1.2 Memory Wiping. Cryptographic secrets or sensitive data, when stored in memory, pose a security risk if exposed. Their presence in memory increases vulnerability to unauthorized access through out-of-bounds reads [1] or cold boot attacks [31]. Memory wiping is a technique used to minimize the duration that sensitive information remains in memory. However, memory wiping can fail if there is an exploitable vulnerability before the wipe, if the memory wipe is not implemented properly, or if sensitive data is duplicated in memory. Nevertheless, memory wiping is an important part of good security hygiene.

Submission Analysis: We analyzed teams’ submissions to understand whether, where, and how they used memory wiping. Similar to privilege separation, memory wiping is foundational to security and provides significant value compared to the difficulty in implementing it.

Teams attempted to use memory wiping in their designs, but their attempts were partially or fully nullified by the compiler. T12 used `memset` three times in one function to zero out local buffers (Listing 2 in Appendix B). However, the analysis of their binary revealed that the second `memset`, which zeroes out the secret AES key from the stack, is optimized away by the compiler. Indeed, T12 used `memset` 10 times in their system, and 5 of them were optimized away. Similar issues were found in T4 and T10’s submissions, in which the compiler optimized away 11 out of 30 and 2 out of 5 calls to `memset`, respectively.

Teams used library functions or custom inline functions to wipe memory, which effectively prevented the compiler from optimizing away the wipes. To effectively remove sensitive data from memory, T20 and T15 utilized the wiping function from the Monocypher library [24] (Listing 3 in Appendix B), which uses the `volatile` keyword to prevent compiler optimizations. For teams using Rust, both T14 and T2 utilized the Rust `zeroize` library to prevent optimization of buffer zeroization [19].

T18 implemented an inline function to erase sensitive data on the stack, as shown in Listing 1. After examining the resulting firmware, we observed that the compiler either unrolled the calls to this inline function for small `data_len` values or replaced them with `memset` calls for larger `data_len` values, while none of the calls were optimized away.

```
1 inline __attribute__((__always_inline__)) void
2   ↪ erase_stack_data(uint8_t *start_add, uint32_t data_len) {
3   ↪   for(uint32_t i=0; i<data_len; i++) start_add[i]=0;
4   ↪ }
```

Listing 1: The inline function to erase used data on stack implemented by T18.

Observation 1: Correctly implemented compiler optimizations cannot preserve the security-related program states that exceed the scope of semantic functionalities of language specifications [67]. This means the security-related operations need to be explicitly controlled and verified by the developer.

Interviews: While analyzing the submissions and firmware was informative, it did not tell us whether participants were aware that the compiler could alter their wiping. Therefore, our interviews sought to gauge participants’ understanding of memory wiping and their awareness of potential compiler alterations to the code.

Participants thought it was less effective to implement memory wiping because of the embedded systems’ threat model. Participant P1 explains that in their threat model, the potential of an attacker gaining arbitrary memory read capabilities would allow access to firmware directly since they are all mapped in the same address space, which will negate the benefits of wiping memory at the application level. P12 also mentioned that they designed their system under the assumption that memory could be dumped. In addition, participant P3 said, “*And if they found a way to do that [read memory content], they could probably do a lot worse than just reading intermediate memory like that.*”

Participants who used the standard library functions to erase the memory were not aware that the compiler could optimize them away. Participant P7’s and P11’s teams utilized `memset` at the

end of functions to erase leftover content. However, they were not aware that the compiler was optimizing away their `memset` calls. As participant P7 realized:

“That’s interesting. I didn’t think that it would optimize that out. Was that because of the optimization flags?”

Recommendation 2: Developers should use memory-wiping functions from trusted cryptographic libraries to ensure reliability, and should routinely verify through disassembly that these security measures have not been optimized away. **Researchers** should develop more reliable methods and tools that eliminate the need for developers to perform such manual verification. **Compiler developers** should notify users when code that has potential security implications is optimized away.

4.1.3 Stack Canary. Stack canary is a defensive mechanism designed to detect and mitigate buffer overflow attacks by inserting a known, random value—referred to as a canary—into the stack frame just before the return address. At the conclusion of a function, the integrity of this canary is verified; any modification suggests that a buffer overflow has occurred, prompting the system to take protective measures such as halting execution or invoking an exception handler. On Cortex-M microcontroller systems, the stack canary feature can be enabled by configuring compiler-level protections. Modern toolchains, such as GCC, support this mechanism through options like `-fstack-protector` or `-fstack-protector-strong`, which automatically instrument code with canary checks.

Submission Analysis: We analyzed teams’ submissions to understand whether and how they implemented stack canaries.

Among the 47 teams, only 2 (4.26%) enabled stack canary protection by activating the appropriate compiler flags. Notably, none of these teams provided additional initialization to randomize the canary value or to customize the error handler. In the absence of user-supplied initialization, the toolchain library defaults to defining the canary as a fixed constant. Consequently, if a buffer overflow corrupts the canary, the default weak-defined handler—typically designed to halt execution or trigger a system reset—will be invoked unless it is explicitly overridden by the user [4, 58]. Despite this default configuration, incorporating stack canary remains a significant step towards securing against buffer overflow attacks.

eCTF vs. real-world firmware: Stack canaries are rarely implemented in real-world microcontroller devices, with a presence rate below 0.2% in large-scale firmware samples [44, 57]. Even when implemented, Xi et al. [58] found them less effective due to the lack of canary randomization and prolonged reuse.

Interviews: In our interviews, we aimed to understand why participants did not implement stack canaries, as well as their perceptions of the potential benefits or limitations of using stack canaries in microcontroller systems.

Participants who understood but did not implement stack canaries thought they needed to be enabled by manually inserting the canary instructions. They were not aware that the stack canary feature could be enabled by configuring compiler flags. Participant P19 incorrectly believed that stack canaries are not typically available on embedded systems, and expressed interest in implementing them manually:

“... on embedded systems that this [stack canary] is not a feature that is usually present ... it would have been fun to do a little implementation of a stack canary, make like an LLVM pass that will automatically inject ... some instructions that will do that for us.”

Similarly, participant P11 was unaware that compilers can implement stack canary protections. They instead focused on securing buffers to prevent overflows in the first place via manual review.

Participants were not aware the stack canary feature could be less effective on microcontroller systems. Tan et al. [58] found that the stack canary feature is less effective on microcontroller systems due to the lack of system support for randomizing the canary value and the prolonged reuse of a single canary value. Among all participants, only P13 were aware that stack canaries on embedded systems can sometimes be static, making them vulnerable to being leaked through memory dumps or crashes. They noted that an attacker with sufficient knowledge could determine the canary value and bypass its protection:

“But realistically, if you have someone who understands how things work, I mean, you figure out what the stack canary is, and you just use that [to bypass it].”

Recommendation 3: Researchers should design new stack canary mechanisms for microcontroller systems. **Vendors** should provide options to enable stack canaries when building with their toolchains.

4.2 Platform Adaptation

System developers must be aware of the platform they’re working on, especially if they’re moving from one platform to another. For example, a developer writing secure software for a microcontroller must be aware that their hardware likely does not include an MMU.

By analyzing teams’ submission and interviewing participants, we sought to understand the conceptual challenges that developers face when moving from microprocessor systems to microcontroller systems. Our analysis revealed two key areas that are related to memory access control in which developers had challenges to adapt: non-executable stack and relocation read-only.

4.2.1 Non-executable Stack. Making the stack non-executable can effectively thwart stack-based code injection attacks [45]. In modern microprocessor systems (e.g., Cortex-A or x86/64), ensuring that the stack of applications is non-executable involves two major steps. First, during the compilation and linking stage, developers specify that they want the stack to be non-executable by using specific linking options. The latest versions of GCC and LLVM default to using a non-executable stack by setting attributes in the program header of the ELF binary [42]. Second, when the program is loaded into memory, the loader reads the ELF’s program header and asks the operating system to enforce non-executability by configuring the MMU page settings.

However, the MCU system does not load the program during runtime; instead, the raw binary, without the program header, eventually gets copied out of the ELF binary and flashed to the MCU’s storage media. Thus, the memory attributes in the ELF’s program header are already lost during this process. As a result, teams that

aimed to enforce non-executability had to devise their own techniques, utilizing the MPU to protect access to stack memory.

Submission Analysis: We sought to understand whether and how they attempted to make the stack non-executable without an operating system. We believe that this knowledge is crucial because preventing the execution of data on the stack is another hugely effective foundational security concept.

Only four submissions successfully implemented stack non-executability. Four submissions out of 47 across both years manually configured the MPU during the firmware initialization process to make the stack non-executable. This entailed enabling the MPU, configuring the memory region for the stack with the eXecute Never (XN) attribute set, and subsequently enabling the region.

A few teams tried to enable the non-executable stack but failed to do so. Two other teams attempted to make the stack non-executable, but they only completed the first step of marking the ELF header and missed the second step of honoring the request during the firmware initialization process. For instance, T18 modified the attribute of the whole SRAM to non-executable in their linker script (Listing 4 in Appendix B). T13 also explicitly specified the linker option `-z noexecstack` during the compilation of the ELF binary.

eCTF vs. real-world firmware: The real-world usage of MPU for memory protection is also minimal, with presence rates less than 2% in large-scale firmware samples [44, 57]. Understanding the reasons behind this is crucial.

Interviews: Through our interviews, we sought to understand why participants didn’t attempt to implement non-executability, and if they did, whether they knew that their implementation did not work and why it did not work.

More than half of the participants (12/22) were unaware of the advantages of a non-executable stack and how to implement it. Participant P1 was not aware of the non-executable stack feature during the competition. As their team used Rust, they were also unsure “if using the Rust compiler specifically for an embedded target will also set those memory protection flags correctly.” P2 mentioned that they were “only looking at library specific flags” when implementing the crypto, and similarly P12 admitted that “the compiler was a potential that we left on the table.”

P7 noticed some teams configured the memory pages in C code and then “setting the bits” on them after the competition, but they did not find out the specific reasons for doing so.

Most participants (21/22) thought adding the noexecstack compiler flag or modifying the attributes in the linker script would effectively make the stack non-executable on a microcontroller system. Participant P5 believed that the non-executable stack flag is a standard method to increase security against buffer overflow exploits in microcontroller environments:

“Yeah, so I think like, non executable stack is something very basic ... very much like standard and a lot more protective in terms of making it difficult to [exploit].”

Participant P4 thought it would be a good but incomplete defense when compiling embedded binary with the non-executable stack:

“So it’s like they’re definitely, obviously it’s like not a complete solution. But ... I would imagine that would still be useful as well in an embedded scenario.”

Participant P9 endorsed the idea of enabling the non-executable stack flag and thought it may work similarly to a traditional micro-processor system:

“It enables the memory protections, I think, for that region of virtual memory? As to how I work on the microcontroller, probably similar to ... a normal PC.”

Only participant P3 acknowledged that merely setting the relevant bits in the ELF files will not make the RAM region non-executable on a microcontroller system.

Recommendation 4: Researchers should explore compiler or linker extensions that automatically enforce memory attributes in microcontroller systems. Vendors might consider preserving the necessary attributes to the raw binary that is flashed to the device in their build toolchains, which enables firmware to read them during initialization to configure corresponding permissions.

4.2.2 Relocation Read-only (RELRO). In dynamically linked ELF binaries, the Global Offset Table (GOT) stores function pointers that are resolved during dynamic linking. Overwriting these function pointers has been an effective attack vector for control-flow hijacking. RELRO [55] is a binary hardening technique to mark GOT and related sections as read-only in the ELF files to mitigate control-flow hijacking. Similarly to enforcing the non-executable stack, it’s imperative for both the loader and the operating system to recognize and adhere to this marking, configuring the memory settings accordingly. However, microcontroller firmware is usually statically linked, which makes the presence of GOT and related sections uncommon, rendering RELRO ineffective for such systems.

Submission Analysis: T13 enabled the `-z relro` linker flag in their Makefile. However, since their firmware is statically linked, this flag has no effect. Unfortunately, we did not have a chance to interview them about their choice. Note that the compiler also does not give warnings when potential invalid options for a specific architecture have been enabled.

Recommendation 5: Compiler developers should warn users when they enable options that may not work on the target architecture. They should also refer users to the documentation for their architecture to ensure their security measures will behave as expected.

4.3 Hardware Attack and Defense

Compared to microprocessor systems, MCU-based embedded devices are more susceptible to hardware attacks like tamperers in physical communication channel, side-channel analysis, and fault injections due to the physical accessibility and simpler circuit design. We refer them as *embedded-prone attacks*. As a result, developers must practice good security hygiene while additionally accounting for these unique threats.

Submission Analysis: We analyzed teams’ submissions to understand whether and how defenses were implemented against embedded-prone attacks. We are careful to not include defenses against attacks that are equally practicable on the microprocessor system as they were not the focus of this study.

Across both years, only 17/47 teams (36.17%) implemented any defense against embedded-prone attacks. Rather, the majority of competitors focused on securing their communication and defending against common software vulnerabilities. For example, most competitors were aware of buffer overflow and brute-force vulnerabilities, and chose to avoid using unsafe functions like `gets` as a result.

Of the teams that implemented any defenses, asynchronous physical communication channel tampering was the least frequently defended against, at 7/17 (41.18%). Asynchronous means the attacker can intercept the communication for offline tampering for as long as they want. For example, T15 designed its communication protocol to require timely response from the other side. They implemented a timeout mechanism such that messages must have been received within a certain time frame to be accepted as valid. If attackers want to manipulate the message in the physical channel, they need to modify the intercepted messages in a tiny time window, which might not be feasible.

Of the teams that implemented any defenses, side-channel analysis was the most frequently defended against, at 13/17 (76.47%). These defenses effectively prevent timing side-channel analysis on password comparison by completing the process in constant time regardless of the input. For example, T7 used the `timing_safe_strcmp` function from the `bCrypt` [28] library, while teams that did not implement a defense used the `strcmp()` or `memcmp()` library functions, which do not operate in constant-time. Other teams, such as T18, implemented their own constant-time comparison functions (Listing 5 in Appendix B).

Of the teams that implemented any defenses, fault injection attacks were defended against by 9/17 teams (52.94%). Mitigation inserts random delays to make it harder for an attacker to inject a fault at the correct time. T15 implemented the random delay macro for delaying 1 to 255 CPU cycles and check for possible glitches at the end (Listing 6 in Appendix B). Using the macro can also avoid potential unwanted compiler optimizations [52].

Interviews: We sought to gauge participants’ knowledge of the unique attack surface in the microcontroller environment and understand why they either did not implement defenses against embedded-prone attacks or did not implement all possible defenses. **Participants acknowledged the larger attack surface of embedded systems due to physical accessibility.** As mentioned by participant P4, *“we had to assume that attackers had physical access with the board,”* which influenced their defensive strategies. They need to additionally consider physical attacks, such as side-channel, and assume the attacker can do *“a lot of things that cryptography alone isn’t really equipped to handle.”*

P11 believed that the embedded systems are simpler but easier to attack because *“it does not have an OS, and everything is pretty much streamlined, so you do have a lot of good regularities in that systems.”* In addition, they are more susceptible to physical attacks as *“you are exposing to the hardware immediately.”*

Participants gave less priority to defending against hardware attacks than software vulnerabilities. Participant P8 initially aimed to secure the design against side-channel analysis. However, due to their knowledge gap in the side-channel analysis, and *“we realized that the number of people who actually know how to run these*

attacks is actually quite small,” they decided to focus on preventing buffer overflows instead.

Participant P9 admitted that the fault injection attacks were “more difficult compared to some of the low hanging fruit,” and they lacked awareness of straightforward protective strategies. As a result, they instead focused on defending attacks that they thought “would be easier to execute.”

P13 admitted that it is “technically easier” to conduct hardware attacks on embedded devices, and attributed the lower priority of hardware defenses to the knowledge gap among developers:

“It’s harder to do hardware stuff because you’re missing the knowledge. It’s easier to do software because more people ... understand how software attacks work.”

Participant P19 felt that software security is generally easier to implement and provides higher defensive value compared to hardware-based protections:

“The potential upsides of defending from a software perspective are both higher and cheaper for ... developer or manufacturer than on the hardware side.”

Observation 2: The knowledge gap in hardware attacks was the dominant factor that prevents participants from implementing effective defenses.

Even if participants were aware of embedded-prone hardware attacks, they had challenges implementing defenses.

Participant P2 discussed the challenges of designing persistent memory on microcontrollers due to the paged flash access:

“We had to find pages in memory, that we’re not going to overwrite anything else that was useful, which usually in general-purpose, that is done for you ... It’s already there in the logistics of the system.”

Participant P15 admitted that they “don’t know if there is a defend mechanism against glitching” and thus skipped the defense. Similarly, P8 acknowledged the theoretical possibility of fault injection attacks but lacked the resources and tools to effectively explore and mitigate such threats:

“For the CS guys, we had no clue this [fault injection] was a thing. And even for the embedded systems [CE] guys, they were like, we know this theoretically could be done, but we don’t quite know how to do it. And we don’t know how to prevent against it.”

Additionally, besides being harder to execute, P18 believed that the hardware attacks are “more difficult to mitigate” than software attacks. They also noted that compilers provide little to no protection against hardware attacks and may even “work against you” by introducing unintended security weaknesses.

Recommendation 6: Researchers should develop frameworks and tools that enable systematic testing of hardware attack mitigations on embedded systems (e.g., reference implementations for side-channel and glitching defenses). They should also investigate compiler-level or build-system features to integrate hardware-focused protections more seamlessly. **Vendors** can improve hardware or SDK support for logging anomalies and detecting tampering events at runtime. **Educators** should incorporate hardware attack scenarios into security curricula. Hands-on lab exercises

and simplified testing tools would help students gain the skills needed to implement or evaluate hardware-level defenses.

5 Practical Challenges

Practical challenges are those faced even when there is an abundance of information. Based on our analysis, two practical challenges emerged as particularly significant: embedded Rust adaption and sources of entropy.

5.1 Memory Safe Language

Microcontroller systems require permissive access to memory to properly and efficiently interface with hardware and peripherals. Accordingly, they are usually programmed in C, but the difficulty of safely and correctly accessing memory often leads to vulnerabilities.

Rust, a memory-safe language, aims to eliminate or reduce memory corruption vulnerabilities with compile-time safety checks. Because Rust is checked for issues at compile time, it maintains a speed similar to that of C. Unfortunately, the unique requirements of microcontroller systems present practical challenges to adopting Rust, which is shown in our interviews and submission analysis.

Submission Analysis: In 2023, 5 out of 20 teams made into the attack phase used Rust as their primary programming language, whereas in 2024, only 2 out of 27 teams adopted Rust in their design. This decrease is attributed to the fact that in 2024, the vendor-provided C Software Development Kit (SDK) was considerably larger in size compared to 2023. Since most teams adopting Rust still needed to compile the vendor’s C SDK with their Rust code, this resulted in an excessive binary size that could not fit into the flash memory, leading to fewer teams adopting Rust that year.

Interviews: We prepared interview questions (Appendix D.3) to understand participants’ views on using a memory-safe language within the microcontroller environment and to learn about their attitudes and practices regarding code safety.

5.1.1 Perspectives and Hesitations About Employing Rust. While most participants did not choose Rust accounting for the overall team familiarities, several participants described practical factors that limited or dissuaded them from incorporating Rust.

Observation 3: Besides familiarity, participants decided against Rust for reasons spanning the lack of direct vendor support, insufficient library support for microcontroller systems, and the steep learning curve. Participants who did not use Rust also expressed misconceptions about its memory safety features.

Participants cited the lack of direct vendor support and insufficient library support for microcontroller systems as major obstacles to switching to Rust. For both years, the SDK for interfacing with the hardware and peripherals was only provided in C. To use Rust, teams had to write their own implementations of SDK functions or use Rust-to-C bindings to interact with hardware.

P5, for instance, realized that the “[vendor] library support is not present in Rust,” which made hardware interactions, such as I2C or flash memory, exceedingly difficult without rewriting entire driver layers from scratch.

Participant P13’s team also focused on language compatibility with the target architecture, explaining that having a functional compiler and toolchain was a deciding factor for language choice:

“Okay, is there a compiler for that particular architecture which allows us to compile binary? And the other thing was, like, you don’t need only a compiler, but you need also, like the whole toolchain.”

They believed Rust was not widely adopted in commercial embedded development due to the extra effort required in the toolchain:

“I have something like 800 [market] embedded IoT devices ... No one in their right mind would use Rust.”

The steep learning curve was also a major hurdle for those who hesitated to adopt Rust. P4 recognized Rust’s learning curve as a significant time investment, especially nuances like handling unsafe code which were unfamiliar to those experienced in C/C++:

“... especially working with unsafe code and stuff, there’s a lot of ... weird idiosyncrasies with Rust that I think what it slowed us down in the development.”

P18 found Rust more complex than C due to features like the ownership model, which “people have to understand” to take its full advantage. Participant P9 acknowledged the advantages of Rust in terms of security and robustness but expressed concerns as it “would be more difficult to develop rapidly.”

Participants who had not actively used Rust sometimes saw it as completely eliminating memory vulnerabilities. Five participants who did not use Rust expressed misconceptions about its memory safety features on embedded systems. Participant P10 said that in their understanding, Rust takes care of memory safety; it is “already memory safe” compared to other languages. P12 also thought that Rust by nature “doesn’t let you write the function if it’s not memory safe, or it kind of optimizes it out if it’s not memory safe.” While it is generally true that Rust enforces strong safety guarantees, these participants did not realize the challenges posed by unsafe code blocks or low-level memory interactions, which are discussed in the next section.

Recommendation 7: Educators should emphasize that, while Rust offers memory safety, its application in microcontroller systems may still pose challenges, particularly when interfacing with existing C libraries and requiring the use of unsafe blocks.

5.1.2 Adoption Experiences and Challenges Among Rust Users. In contrast, some participants did integrate Rust into their designs and encountered a different set of hurdles.

Observation 4: Participants used Rust mainly faced challenges in compiling Rust without the standard library, efficiently implementing the hardware abstraction layer, and managing unsafe operations properly.

Participants encountered challenges when compiling Rust without the standard library on the embedded device. The Rust standard library provides abstractions, types, operations, and other quality-of-life utilities for developing in Rust [50]. However, it is often necessary to use Rust without the standard environment in situations where an OS is not available.

P3 told us that transitioning to using Rust in a `no_std` [7] environment posed a unique set of challenges distinct from general-purpose programming on the microprocessor systems. They had to ensure that any dependencies they wanted to use were compatible with a `no_std` environment, which narrowed the range of usable

libraries. Besides compatibility, participant P4 also noted that any inclusion of the Rust standard library risked bloating the binary.

Participants highlighted the challenges in implementing a secure Hardware Abstraction Layer (HAL). A common practice in embedded development is to perform hardware interactions through an intermediate HAL, which hides the low-level implementation details from the rest of the system. The HAL used for the reference system was provided as an SDK that offered C interfaces to various components of the microcontroller used in the competition. However, a similar HAL was not available for Rust.

Participant P1 told us that in 2023, they created many Rust bindings to the original C SDK, which was compiled with their Rust code that only implements the high-level protocol:

“... if you look at our 2023 codebase, it’s mostly ... written in C because we just collect the entire C library for the Tiva [C SDK]. That’s instruments driver, and then we compile that in with our like, small Rust code.”

In the 2024 event, P1’s team avoided compiling the vendor’s large C SDK alongside their Rust code (which inflated firmware size), and instead wrote a minimal HAL from scratch:

“... it lets us have a more holistic view of the device ... It lets us get more Rust experience and make sure that we’re also avoiding any bugs that could potentially occur in a C library.”

In order to develop the HAL that directly interacts with peripherals, P1’s team utilized the `svd2rust` [49] tool to automatically generate Rust structures from CMSIS-SVD [5] files, which describes the memory-mapped registers of peripherals and is available in the vendor-provided SDKs. This automation reduced the need for manual volatile reads/writes by providing pre-generated bindings for hardware interactions. After this, with the Rust peripheral access crate [8], they were able to abstract the direct register manipulations into more manageable function calls in HAL.

Observation 5: Automated conversion of hardware descriptor files into Rust structures can accelerate embedded HAL development, but its fidelity and security require further research.

P1 also mentioned that they encountered inconsistencies between the vendor-provided CMSIS-SVD files and the device user manual. They believe “it’s probably because they [the vendor] took some previous SVD file copied over and change.”

Teams also struggled with effectively compartmentalizing unsafe operations. It is possible to call C functions from Rust code, but only within an unsafe block. Unsafe blocks in Rust allow for certain operations that the compiler cannot guarantee to be safe, such as dereferencing raw pointers or calling external C code [9]. This bypasses the rigorous safety checks normally enforced by Rust, potentially leading to security vulnerabilities such as buffer overflows or access violations if not carefully managed.

Some teams, such as T11 led by P3, used unsafe blocks extensively for calling C SDK functions, primarily due to their incremental development model, which allowed for parts of the system to be gradually ported to Rust while continuously testing and validating functionality. They explained:

“... the teams that chose to rewrite everything in scratch from Rust would have had an intermediate state, your

part of the code was ported over to Rust, part of it was in C, but they were separate. And there was no way to know if it worked until the Rust rewrite was done.”

In contrast, P18’s team aimed to confine unsafe code to low-level crates that interact with hardware, while they tried not to use unsafe code “especially like anywhere in our logic.”

Recommendation 8: Researchers might further explore ways to automate or reduce unsafe code by refining Rust-based abstractions, improving static analysis, and clarifying best practices for HAL construction. Vendors should consider enhancing their support for Rust in embedded systems by providing full Rust SDKs or Rust-to-C bindings for their existing C libraries. Additionally, the hardware descriptor files should be made accurately to assist developers in implementing Rust-based HAL effectively.

5.2 Entropy Sources

Like microprocessor systems, microcontrollers use Pseudo-Random Number Generators (PRNGs) to generate cryptographic secrets. PRNGs take a seed as an input to deterministically generate an output. Since the output is deterministic, the input must be comprised of harvested randomness. However, there are fewer high-quality sources of entropy on a microcontroller system than a microprocessor system, complicating the generation of pseudo-random numbers and impacting the robustness of cryptographic operations [33].

Submission Analysis: We analyzed teams’ submissions to understand how they used randomness in their designs, their source(s) of entropy, and shortcomings that might have arisen from their combination of the two.

Teams were not able to maintain unpredictability in their use of randomness. In 2023, 1 out of 20 teams neglected the inclusion of entropy in their cryptographic design, while 5 teams relied on hard-coded seeds or entropy generated at build time. Although some teams updated the seed with each use of the PRNG, the ability of attackers to reset the seed by re-flashing the firmware allowed them to predict the PRNG’s output. In 2024, 18 out of 27 teams utilized the vendor-provided True Random Number Generator (TRNG), while 5 teams did not include randomness in their design.

In addition to hard-coded seeds, some teams opted for entropy sources such as the SysTick counter, built-in timers, or CPU cycle counters. While these sources represent an improvement over static seeds, using them without other sources still introduces vulnerabilities. Attackers could execute the firmware and repeatedly perform specific operations, cataloging PRNG outputs to construct a comprehensive database. A sufficiently extensive database increases the likelihood that the output of a future operation, if conducted at a precise time, could coincide with a database entry, thereby facilitating replay attacks.

Teams utilized different approaches to overcome the challenge and obtain sufficient entropy. Addressing the challenge of limited randomness and entropy in microcontroller systems, a viable strategy involves aggregating multiple samples and, when feasible, incorporating various sources of entropy [29]. This approach mixes random bits across the collected data, thereby maximizing the entropy of randomness.

Several teams adopted this technique by consolidating entropy samples into a pool and subsequently employing a hashing algorithm to derive their seed, mirroring the methodology employed by the Linux PRNG algorithm [30]. Teams such as T14 and T4 exemplified this practice, while T15 adopted an alternative approach, hashing values sampled from an internal temperature sensor and combining them through XOR operations. Both strategies align with NIST recommendations for preserving inputted randomness and mitigating risks associated with insufficient randomness [35].

Observation 6: Although microcontroller systems usually have limited entropy sources, combining multiple available sources can effectively improve the robustness of the generated randomness.

Interviews: We sought to understand the challenges that participants faced with regard to finding and using a reliable source of entropy. We asked participants about why they chose their entropy source and how they tested its effectiveness.

Participants faced challenges in implementing RNG on the microcontroller system due to limited entropy sources. Participant P4 admitted to having very low confidence in their RNG implementation, primarily due to the lack of appropriate entropy sources and knowledge gaps in implementing RNG on a microcontroller system:

“... most implementations of RNG that we could find relied on OS level calls to like urandom ... and we just did not have that on the embedded system.”

They struggled with identifying reliable entropy sources, such as SRAM at boot time or temperature sensors, and had difficulties in ensuring adequate entropy was sampled for true randomness:

“We didn’t really know how to ... get solid entropy sources to seed the random number generator to make sure that it’s random.”

Participants questioned the reliability of the vendor-provided TRNG but did not conduct rigorous tests. In 2024, the development board was equipped with a vendor-provided TRNG implementation, which could be accessed through the SDK function calls. However, the board’s documentation only indicated that the TRNG gathered randomness from various sources on the board, without providing any details about the reliability and quality of the generated randomness [18].

P1 chose a PRNG seeded by both TRNG and the CPU clock because “we were not totally sure how much we could trust the TRNG.” Eventually, due to the TRNG’s slow performance and the potential for attacks against hardware TRNGs, they decided to mix multiple entropy sources to seed a PRNG.

Participant P2 also didn’t trust the TRNG initially and tried to confirm it with the vendor to understand how it was implemented. However, the vendor “didn’t tell us everything,” and they had to rely on the vendor’s assurances.

Most participants did not rigorously test the vendor-provided TRNG; instead, they relied on the vendor’s assurances. According to P2, since the vendor did not disclose sufficient information and because of the competition schedule, they chose to trust the vendor’s certainty and did not perform a robustness test. Similarly, 18 out of 22 participants admitted to performing only rudimentary

tests, such as checking the first few outputs to ensure they appeared random, as mentioned by P7:

“We did not test. We just ... printed it ten times and until they it’s different each time. So we have it good enough.”

To further assess the RNG’s robustness, several teams conducted ad-hoc tests. However, they struggled to interpret the results, as many tests yielded inconclusive outcomes. Participant P11 and P19’s teams experimented the board with different environmental conditions, such as by freezing the board, *“to see if the TRNG would break or produce predictable values,”* as P19 explained. They also mentioned a basic but effective test they conducted to ensure the TRNG’s reliability:

“If you print out an image of the [output] of the TRNG, it becomes quite obvious ... sometimes you can see boxes or marks where there is some repeatable pattern.”

Others, like P11, applied *“cryptographic distinguishers to see if it’s random or not”* and observed no obvious patterns. P5’s team *“generated around 2 million random numbers from the TRNG”* for preliminary statistical checks. They all mentioned that their statistical approaches were taken from existing cryptographic research.

P4 stated that they utilized NIST’s suit of statistical randomness tests [48] to evaluate the RNG in both years, gathering around a million samples for testing. They encountered difficulties in interpreting the test results, with many tests returning inconclusive outcomes that did not definitively indicate the RNG’s reliability. In 2024, they also employed the dieharder test suite [10], which is more comprehensive but also more demanding in terms of the sample size required for conclusive results. P4 admitted:

“We were more so trusting that the proprietary non-disclosure TRNG was cryptographically secure ... the test was more so just a very quick sanity check to make sure it’s not just egregiously bad.”

Recommendation 9: Developers should diversify their entropy sources and rigorously test RNG implementations. Vendors should clearly delineate and recommend available entropy sources for specific boards, especially when a TRNG is absent, and provide detailed documentation on the implementation and performance of provided TRNGs.

6 Summary and Future Work

Summary of our findings. From our observations, several significant trends related to embedded development and the effectiveness of our security curriculum have emerged.

First, the responsibility for writing secure code largely falls on developers. Although modern languages and recent research offer promising advancements, implementing these technologies poses both conceptual and practical challenges. The lack of vendor and library support further hinders the adoption of new technologies, often leaving embedded developers to either create their own solutions or abandon the problem entirely.

Second, MCU developers face a unique set of demands compared to their counterparts working on microprocessor systems. They need an in-depth understanding of their specific platform, including how their code interacts with the compiler and underlying hardware. This requires not only proficiency in application code

but also a thorough knowledge of hardware features and how to use them to enforce security boundaries. The often opaque and counter-intuitive nature of compiler operations adds to the complexity, placing a substantial burden on developers as they navigate multiple layers of the software/hardware stack.

Lastly, current curricula, particularly those for CS majors, fall short in preparing students for the specific challenges of embedded development. Our study participants were notably puzzled by how to defend against physical threats, highlighting a significant gap in their education regarding practical security measures for embedded systems.

Future work. Overall, the demands placed on MCU developers are substantial, requiring a unique combination of skills that span software development, computer engineering, and security best practices. As the ubiquity, importance, and connectivity of microcontroller systems continue to grow, there is an increasing need for education, tools, and frameworks that can assist developers in navigating these challenges and reducing the likelihood of security vulnerabilities in embedded code.

7 Related Work

CTF research and user study. Previous papers on CTF experiences [32, 36, 61, 63, 64] have a primary focus on educational purposes. Vigna et al. [63] introduced a framework built on a decade’s worth of experience in organizing the international Capture the Flag (iCTF) [53], which was further developed to offer a CTF-as-a-Service solution [61]. Similarly, Vykopal et al. [64] emphasized the advantages of using CTF challenges as hands-on assignments to enhance students’ skills. In addition to the educational benefits, researchers have also examined the challenges and obstacles associated with the CTF model itself. Such research sheds light on strategies for addressing various challenges and ensuring a successful CTF experience for participants. Crispin et al. [16] described their experience in the Defcon CTF. Chung et al. [11] discussed methods to overcome the pitfalls and hurdles commonly encountered in organizing CTFs. Fulton et al. [27] conducted user studies with senior software developers who have worked with Rust to understand the benefits and challenges of using Rust in their projects. They identified drawbacks including the steep learning curve, limited library support, and concerns about the ability to hire additional Rust developers in the future.

Unlike the existing literature, our study uniquely combines submission analysis with participant interviews to provide a dual-perspective understanding of both the technical and human factors influencing security practices.

Securing microcontroller systems. Many technical solutions have been proposed to protect microcontroller systems against attacks. These include privilege separation and compartmentalization [6, 12, 14, 37], control-flow integrity (CFI) techniques [2, 20, 41, 58, 59, 65, 68], randomization methods [40, 54], Return-Oriented Programming (ROP) gadget removal techniques [39], etc. For a comprehensive review of the research on defensive approaches, please refer to Tan et al. [57]. For real-world firmware analysis, FirmXRay [66], Nino et al. [44], and Tan et al. [57] all presented datasets of microcontroller firmware for IoT devices. These studies conducted static analyses to assess the security properties of

these datasets. Even though the results of these works reveal many concerning issues in real-world firmware, none of them studies the perceptual challenges associated with adopting security mitigation for microcontroller systems.

Unlike previous purely technical studies, we examine the primary challenges associated with designing and implementing security mitigation for microcontroller systems from both technical and perceptual perspectives.

8 Conclusion

In this study, we investigated the security practices and challenges faced by participants in the 2023 and 2024 MITRE eCTF competitions. Through a detailed analysis of competition submissions and interviews with participants, we uncovered both conceptual and practical security gaps in the development of embedded systems. Our findings indicate that despite the participants’ familiarity with basic security concepts, there is a significant disconnect when applying these principles to embedded systems, compounded by a lack of adequate support for robust, embedded-specific security practices. We hope this paper spurs further discussion and improvement within the educational frameworks and industrial practices surrounding embedded systems security.

Acknowledgment

The authors extend their gratitude to Ya-Hui Chang and Dan Walters for their valuable discussions and suggestions. This material is based upon work supported in part by National Science Foundation (NSF) grants (2237238, 2329704, 2512972, 2508320, 2422242, and 2523436). Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of United States Government or any agency thereof.

References

- [1] The Heartbleed Bug. <https://heartbleed.com/>.
- [2] Naif Saleh Almkhhdhub, Abraham A Clements, Saurabh Bagchi, and Mathias Payer. μ RAI: Securing embedded systems with return address integrity. In *Network and Distributed System Security Symposium (NDSS)*, 2020.
- [3] Thomas Alsop. Microcontroller unit (MCU) shipments worldwide from 2015 to 2021. <https://www.statista.com/statistics/935382/worldwide-microcontroller-unit-shipments/>, 2023.
- [4] GitHub anza xyz/newlib. `newlib/newlib/libc/ssp/stack_protector.c` at `bpf-port`. https://github.com/anza-xyz/newlib/blob/bpf-port/newlib/libc/ssp/stack_protector.c.
- [5] ARM. CMSIS - Common Microcontroller Software Interface Standard. https://arm-software.github.io/CMSIS_6/latest/General/index.html.
- [6] Zelalem Birhanu Aweke and Todd Austin. uSFI: Ultra-lightweight software fault isolation for IoT-class devices. In *IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2018.
- [7] The Embedded Rust Book. A no_std Rust Environment. <https://docs.rust-embedded.org/book/intro/no-std.html>.
- [8] The Embedded Rust Book. Memory Mapped Registers. <https://docs.rust-embedded.org/book/start/registers.html>.
- [9] The Embedded Rust Book. Unsafe Rust. <https://doc.rust-lang.org/book/ch19-01-unsafe-rust.html>.
- [10] Robert G Brown, Dirk Eddelbuettel, and David Bauer. Dieharder. *Duke University Physics Department Durham, NC*, pages 27708–0305, 2018.
- [11] Kevin Chung and Julian Cohen. Learning obstacles in the capture the flag model. In *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE)*, 2014.
- [12] Abraham A Clements, Naif Saleh Almkhhdhub, Saurabh Bagchi, and Mathias Payer. {ACES}: Automatic compartments for embedded systems. In *27th USENIX Security Symposium (USENIX Security 18)*, 2018.
- [13] Abraham A Clements, Naif Saleh Almkhhdhub, Khaled S Saab, Prashast Srivastava, Jinkyu Koo, Saurabh Bagchi, and Mathias Payer. Protecting bare-metal embedded systems with privilege overlays. In *2017 IEEE Symposium on Security and Privacy (SP)*, 2017.
- [14] Abraham A Clements, Naif Saleh Almkhhdhub, Khaled S Saab, Prashast Srivastava, Jinkyu Koo, Saurabh Bagchi, and Mathias Payer. Protecting bare-metal embedded systems with privilege overlays. In *IEEE Symposium on Security and Privacy (S&P)*, 2017.
- [15] MITRE Corporation. Embedded Capture the Flag (eCTF). <https://ectf.mitre.org/>.
- [16] Crispin Cowan, Seth Arnold, Steve Beattie, Chris Wright, and John Viega. Defcon capture the flag: Defending vulnerable code from intense attack. In *Proceedings DARPA Information Survivability Conference and Exposition*, volume 1, pages 120–129. IEEE, 2003.
- [17] Analog Devices. MAX78000FTHR. <https://www.analog.com/en/resources/evaluation-hardware-and-software/evaluation-boards-kits/max78000fthr.html>.
- [18] Analog Devices. UG-7456: MAX78000 User Guide (Rev.1). <https://www.analog.com/media/en/technical-documentation/user-guides/max78000-user-guide.pdf>.
- [19] Docs.rs. zeroize - Rust. <https://docs.rs/zeroize/latest/zeroize/>, 2024.
- [20] Yufei Du, Zhuojia Shen, Komail Dharsee, Jie Zhou, Robert J Walls, and John Criswell. Holistic Control-Flow Protection on Real-Time Embedded Systems with Kage. In *USENIX Security Symposium*, 2022.
- [21] MITRE Engenuity. eCTF: 2023 Awards Ceremony. <https://www.youtube.com/watch?v=QDfEKUtrf8o>.
- [22] MITRE Engenuity. eCTF: 2024 Awards Ceremony. <https://www.youtube.com/watch?v=an5utl--nto>.
- [23] MITRE Engenuity. Past Competitions | Embedded Capture the Flag (eCTF). <https://ectf.mitre.org/past-competitions/>.
- [24] Daniel J. Bernstein et al. Monocypher. <https://monocypher.org/>, 2024.
- [25] Igor Fedorov, Ryan P Adams, Matthew Mattina, and Paul Whatmough. Sparse: Sparse architecture search for cnns on resource-constrained microcontrollers. *Advances in Neural Information Processing Systems*, 32, 2019.
- [26] Stuart I. Feldman. Make—a program for maintaining computer programs. In *Software: Practice and experience* 9.4, 1979.
- [27] Kelsey R Fulton, Anna Chan, Daniel Votipka, Michael Hicks, and Michelle L Mazurek. Benefits and drawbacks of adopting a secure programming language: Rust as a case study. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021.
- [28] Ricardo Garcia and Alex Young. bCrypt. <https://github.com/rg3/libbcbcrypt/tree/master>, 2011. Last access date: 2024.
- [29] Jacob Grycel and Robert J Walls. Erhard-rng: A random number generator built from repurposed hardware in embedded systems. *arXiv preprint arXiv:1903.09365*, 2019.
- [30] Zvi Gutterman, Benny Pinkas, and Tzachy Reinman. Analysis of the linux random number generator. In *2006 IEEE Symposium on Security and Privacy (S&P'06)*, 2006.
- [31] J Alex Halderman, Seth D Schoen, Nadia Heninger, William Clarkson, William Paul, Joseph A Calandrino, Ariel J Feldman, Jacob Appelbaum, and Edward W Felten. Let's we remember: cold-boot attacks on encryption keys. *Communications of the ACM*, 52(5):91–98, 2009.
- [32] Hugrun Hannesdottir. *Teaching IoT Security using CTF Problems*. PhD thesis, Carnegie Mellon University, 2021.
- [33] Nadia Heninger, Zakir Durumeric, Eric Wustrow, and J Alex Halderman. Mining your ps and qs: Detection of widespread weak keys in network devices. In *21st USENIX Security Symposium (USENIX Security 12)*, 2012.
- [34] Texas Instruments. EK-TM4C123GXL. <https://www.ti.com/tool/EK-TM4C123GXL>.
- [35] Meltem Sonmez Turan John M. Kelsey, Kerry McKay. Recommendation for the entropy sources used for random bit generation. <https://doi.org/10.6028/NIST.SP.800-90B>, 2018.
- [36] Alexandre Oliveira Junior, Gustavo Funchal, Jonas Queiroz, Jorge Loureiro, Tiago Pedrosa, Javier Parra, and Paulo Leitao. Learning Cybersecurity in IoT-based Applications through a Capture the Flag Competition. In *International Conference on Industrial Informatics (INDIN)*. IEEE, 2022.
- [37] Chung Hwan Kim, Taegyu Kim, Hongjun Choi, Zhongshu Gu, Byoungyoung Lee, Xiangyu Zhang, and Dongyan Xu. Securing Real-Time Microcontroller Systems through Customized Memory View Switching. In *Network and Distributed System Security Symposium (NDSS)*, 2018.
- [38] Mark Kreitz. Security by design in software engineering. *ACM SIGSOFT Software Engineering Notes*, 44(3):23–23, 2019.
- [39] Donghyun Kwon, Jangseop Shin, Giyeol Kim, Byoungyoung Lee, Yeongpil Cho, and Yunheung Paek. uXOM: Efficient eXecute-Only Memory on Arm Cortex-M. In *USENIX Security Symposium*, 2019.
- [40] Lan Luo, Xinhui Shao, Zhen Ling, Huaayu Yan, Yumeng Wei, and Xinwen Fu. FASLR: Function-Based ASLR via TrustZone-M and MPU for Resource-Constrained IoT Systems. *IEEE Internet of Things Journal*, 2022.
- [41] Zheyuan Ma, Xi Tan, Lukasz Ziarek, Ning Zhang, Hongxin Hu, and Ziming Zhao. Return-to-non-secure vulnerabilities on arm cortex-m trustzone: Attack and defense. In *ACM/IEEE Design Automation Conference*, 2023.

- [42] Linux manual page. `execstack(8)`. <https://man7.org/linux/man-pages/man8/execstack.8.html>.
- [43] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for csw and hci practice. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–23, 2019.
- [44] Nicolas Nino, Ruibo Lu, Wei Zhou, Kyu Hyung Lee, Ziming Zhao, and Le Guan. Unveiling {IoT} security in reality: A {Firmware-Centric} journey. In *33rd USENIX Security Symposium (USENIX Security 24)*, 2024.
- [45] Aleph One. Smashing the stack for fun and profit. *Phrack magazine*, 7(49):14–16, 1996.
- [46] Otter.ai. Otter.ai - AI Meeting Note Taker & Real-time AI Transcription. <https://otter.ai/>.
- [47] Otter.ai. Privacy & Security. <https://otter.ai/privacy-security>.
- [48] Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Alan Heckert, et al. *A statistical test suite for random and pseudorandom number generators for cryptographic applications*, volume 22. US Department of Commerce, Technology Administration, National Institute of . . . , 2001.
- [49] Rust. Peripheral API generator from CMSIS-SVD files. <https://docs.rs/svd2rust/1atest/svd2rust/>.
- [50] Rust. `std` - The Rust Standard Library. <https://doc.rust-lang.org/std/>.
- [51] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. SAGE publications Ltd, 2021.
- [52] Michael Savage. C++ tricks: macro to disable optimisations. <https://mikesavage.co.uk/cpp-tricks-disable-optimisations-macro/>.
- [53] Shellphish. iCTF: the International Capture The Flag Competition. <https://shellphish.net/ictf/>.
- [54] Jiameng Shi, Le Guan, Wenqiang Li, Dayou Zhang, Ping Chen, and Ning Zhang. HARM: Hardware-Assisted Continuous Re-randomization for Microcontrollers. In *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2022.
- [55] Huzaiifa Sidhpurwala. Hardening ELF binaries using Relocation Read-Only (RELRO). <https://www.redhat.com/en/blog/hardening-elf-binaries-using-relocation-read-only-relro>, 2019.
- [56] Laszlo Szekeres, Mathias Payer, Tao Wei, and Dawn Song. Sok: Eternal war in memory. In *IEEE Symposium on Security and Privacy*, pages 48–62. IEEE, 2013.
- [57] Xi Tan, Zheyuan Ma, Sandro Pinto, Le Guan, Ning Zhang, Jun Xu, Zhiqiang Lin, Hongxin Hu, and Ziming Zhao. SoK: Where’s the “up”?! A Comprehensive (bottom-up) Study on the Security of Arm Cortex-M Systems. In *USENIX WOOT Conference on Offensive Technologies*, 2024.
- [58] Xi Tan, Sagar Mohan, Md Armanuzzaman, Zheyuan Ma, Gaoxiang Liu, Alex Eastman, Hongxin Hu, and Ziming Zhao. Is the Canary Dead? On the Effectiveness of Stack Canaries on Microcontroller Systems. In *ACM/SIGAPP Symposium On Applied Computing*, 2024.
- [59] Xi Tan and Ziming Zhao. Sherlock: Secure and holistic control-flow violation detection on embedded systems. In *ACM Conference on Computer and Communications Security*, 2023.
- [60] The MITRE Embedded Capture the Flag Competition. 2024 Award Ceremony. https://ectfmitre.gitlab.io/ectf-website/2024/events/award_ceremony.html.
- [61] Erik Tricket, Francesco Disperati, Eric Gustafson, Faezeh Kalantari, Mike Mabey, Naveen Tiwari, Yeganeh Safaei, Adam Doupe, and Giovanni Vigna. Shell We Play A Game? CTF-as-a-service for Security Education. In *ASE@USENIX Security Symposium*, 2017.
- [62] Elaine Venson, Reem Alfayez, Marília MF Gomes, Rejane MC Figueiredo, and Barry Boehm. The impact of software security practices on development effort: An initial survey. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 1–12. IEEE, 2019.
- [63] Giovanni Vigna, Kevin Borgolte, Jacopo Corbetta, Adam Doupe, Yanick Fratantonio, Luca Invernizzi, Dhilung Kirat, and Yan Shoshitaishvili. Ten years of ictf: The good, the bad, and the ugly. In *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE)*, 2014.
- [64] Jan Vykopal, Valdemar Švábenský, and Ee-Chien Chang. Benefits and pitfalls of using capture the flag games in university courses. In *ACM Technical Symposium on Computer Science Education*, pages 752–758, 2020.
- [65] Yujie Wang, Cailani Lemieux Mack, Xi Tan, Ning Zhang, Ziming Zhao, Sanjoy Baruah, and Bryan C. Ward. InsectACIDE: Debugger-Based Holistic Asynchronous CFI for Embedded System. In *IEEE Real-Time and Embedded Technology and Applications Symposium*, 2024.
- [66] Haohuang Wen, Zhiqiang Lin, and Yinqian Zhang. Firmxray: Detecting bluetooth link layer vulnerabilities from bare-metal firmware. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2020.
- [67] Jianhao Xu, Kangjie Lu, Zhengjie Du, Zhu Ding, Linke Li, Qiushi Wu, Mathias Payer, and Bing Mao. Silent Bugs Matter: A Study of Compiler-Introduced Security Bugs. In *USENIX Security Symposium*, 2023.
- [68] Jie Zhou, Yufei Du, Zhuojia Shen, Lele Ma, John Criswell, and Robert J Walls. Silhouette: Efficient protected shadow stacks for embedded systems. In *USENIX Security Symposium*, 2020.

Ethics and Open Science Compliance

Adhering to ethical standards and despite receiving an exemption from our IRBs, our study strictly followed the designated policies and procedures applicable to human research as specified by our institutions and study protocol. Prior to conducting the interviews, we ensured all participants fully understood their involvement through a clear consent process and upheld their right to withdraw from the study at any point without repercussion. To safeguard participant privacy and confidentiality, all PII was anonymized during the transcription of interviews. Additionally, participants were afforded the freedom to withhold any information they preferred not to disclose during all stages of data collection.

For transcription of interview recordings, we used Otter AI [46], an online service that converts audio to text. At the time of conducting interviews, the transcription tool had not been finalized, and participants were not explicitly informed about the use of Otter AI. Following peer review feedback, we promptly contacted all participants to disclose this information and clarified that Otter AI’s privacy policy [47] complies with our IRB protocol. No manual review of transcripts or audio was conducted by Otter AI staff, and no PII was shared with the service. Additionally, according to Otter AI’s privacy policy, data used for model training is further de-identified to ensure individual users cannot be identified.

In alignment with the open science policy, our study is committed to enhancing the transparency of our research findings. We make available all research artifacts that are permissible within the constraints of our IRB study protocol, including the demographic survey, interview questionnaire, and the codebook used for the thematic analysis. Detailed interview transcripts, however, cannot be shared publicly due to privacy and confidentiality constraints specified by our IRB study protocol.

A The General Process of eCTF

Figure 1 illustrates the general process of eCTF. The *handoff* phase typically begins one and a half months after the *design* phase starts. After successfully completing function tests in the *handoff* phase, a team enters the *attack* phase immediately. The competition ends approximately one and a half months after the *handoff* phase begins.

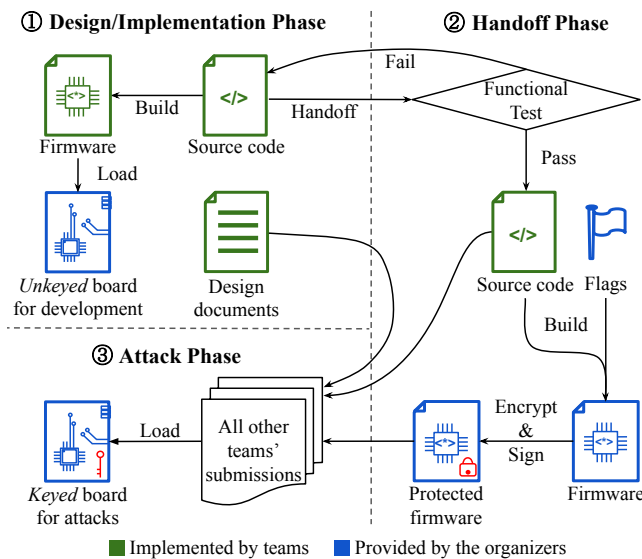


Figure 1: The general process of eCTF.

B Code Examples

Listing 2 shows the optimized away `memset` in a C function by T12. Listing 3 shows the `crypto_wipe` function defined in the Mono-cypher library. These code samples are referenced in Section 4.1.2.

Listing 4 shows the modified linker script by T18, as detailed in Section 4.2.1.

Listing 5 and 6 show the constant-time password comparison function from T18 and the random delay macro from T15, which are referenced by Section 4.3.

```

1 void unlockCar(FLASH_DATA *fob_state_ram) {
2     if (fob_state_ram->paired == FLASH_PAIRED) {
3         ...
4         MESSAGE_PACKET message;
5         char buffer[64];
6         memset(buffer, 0, 64);
7         ...
8         message.buffer = buffer + 1;
9         struct tc_aes_key_sched_struct s;
10        ...
11        tc_aes_encrypt((message.buffer)-1, (message.buffer)-1, &s);
12        ...
13        memset(&s, 0, sizeof(struct tc_aes_key_sched_struct));
14        memset(message.buffer, 0, 64);
15    }
16 }

```

Listing 2: The C code of a function that uses `memset` to zeroize local buffers. In the compiled firmware, the second `memset` was optimized away.

```

1 #define ZERO(buf, size) FOR(_i_, 0, size) (buf)[_i_] = 0
2
3 void crypto_wipe(void *secret, size_t size) {
4     volatile u8 *v_secret = (u8*)secret;
5     ZERO(v_secret, size);
6 }

```

Listing 3: The `crypto_wipe` function defined in the Mono-cypher library. The `volatile` keyword prevents potential compiler optimizations.

```

1 MEMORY
2 {
3     FLASH (rx) : ORIGIN = 0x00008000, LENGTH = 0x00038000
4     SRAM (rw) : ORIGIN = 0x20000000, LENGTH = 0x00008000
5 }

```

Listing 4: The linker script snippet showing that the SRAM region is set to read and write only by T18.

```

1 int ConstantCompare(const uint8_t* a, const uint8_t* b, int length)
2 ↪ {
3     int i;
4     int compareSum = 0;
5     for (i = 0; i < length; i++) {
6         compareSum |= a[i] ^ b[i];
7     }
8     return compareSum;
9 }

```

Listing 5: The constant-time password comparison function from T18. Given a fixed length argument, the execution time of this function is constant, and is independent of the similarity of the two string inputs.

```

1 #define RAND_STALL() \
2     rand_ret = -1; \
3     rand_ret = fillEntropyBuf(rand_rbt, 2); \
4     if (rand_ret == -1) halt_and_catch_fire(); \
5     rand_i = 0; \
6     rand_y = 0; \
7     for (rand_i = 0; rand_i < rand_rbt[0]; rand_i++) \
8         rand_y += 1; \
9     rand_ret = ((rand_i == rand_y) && (rand_rbt[0] == rand_y))

```

Listing 6: The random delay macro from T15.

C Background Survey

C.1 Personal background questions

- (1) Which year of the MITRE eCTF did you participate?
- (2) What was your degree type during the eCTF? For example, high school, undergraduate, Master's, Ph.D., or other.
- (3) What was your program of study (major)?
- (4) Does your institution offer embedded- or IoT-specific security courses? If so, did you take those courses before the competition? Have you taken them since the competition?
- (5) Does your institution offer other security courses? If so, did you take those courses before the competition? Have you taken them since the competition?
- (6) Have you previously participated in the eCTF competition? If so, when?
- (7) What is your experience with other security competitions?
- (8) What was your design or implementation experience of the embedded systems before the eCTF?

C.2 Team-related questions

- (1) How many active members were on your team? Please only include members who regularly contributed during the competition.
- (2) How many students expressed interest in competing but did not remain active until the end of the competition? If so, what are some of the reasons that they became less involved?
- (3) How much was your teams' faculty advisor involved in the eCTF? (give some hints)
- (4) Did your institution offer class credit for participation in this competition? If yes, how many credits?
- (5) What was your role and contribution in the team?
- (6) What were the degree-level and program of study for the other participants on the team?
- (7) What was the general level of experience among your team members?

D Interview Questions

Prior to each interview, we ensured that interviewees read and understood the consent information sheet and were fully informed about the study's scope, their rights as interviewees, and our confidentiality measures. At the beginning of each session, verbal consent was recorded to confirm their willingness to participate.

D.1 Software-related questions

- (1) *The principle of least privilege is a computer security practice that gives tasks least access rights based on their job. The Microcontroller Unit (MCU) used in the competition supports two privilege levels. However, our analysis found that none of the competing teams utilized this privilege separation feature.*
 - (a) Did your team know that this MCU offers two privilege levels?
 - (b) Why didn't your team use both privilege levels of the MCU?
 - (c) Do you think using both privilege levels would have improved the security of your design?

- (2) *The compiler and linker provide multiple security enforcement options that can be enabled through editing the Makefile or linker script.*
 - (a) Do you know the benefits of making the stack non-executable?
 - (b) Do you know there are compiler flags to make the stack non-executable, and you can also mark the SRAM region as non-executable in the linker script?
 - (c) Do you know how these compiler flags or linker script attributes work under the hood?
 - (d) Did you know the memory protection unit (MPU)? How did you know it?
- (3) *Stack canaries are a security feature that detects stack buffer overflows by placing a random value before the return address on the stack. When the function returns, the canary value is checked to see if it has been modified.*
 - (a) Did you know the stack canary feature before the competition?
 - (b) Why didn't your team use the stack canary feature?
- (4) *Using the C standard library on embedded systems can be tricky due to limited resources and specific requirements.*
 - (a) Before the competition, did you aware that some of the C standard library functions perform differently on the embedded system compared to a general purpose system?
- (5) (Some teams) *We found in your team's design, the non-executable stack and **relro** compiler flags were added in the Makefile.*
 - (a) What was the intention in using the flags?
 - (b) Were these flags effective at meeting the intentions?
- (6) (Some teams) *We found in your team's design, the SRAM attribute in the linker script was changed to read/write only.*
 - (a) What was the intention in changing this attribute?
 - (b) Were these changes effective at meeting the intentions?

D.2 Questions related to memory wiping

- (1) *Sensitive data, such as encryption keys, can be stored in some secure storage at rest, and then be loaded to the main memory for computation. To minimize the time window for sensitive data residing in the main memory, one can clear out the data in the main memory after use. This protection is called memory wiping.*
 - (a) Did your team use or attempt to use the memory wiping as a defense technique?
 - (i) If so,
 - how did your team implement it?
 - what was your intention in using it?
 - how did you confirm that your use met the intentions?
 - (some teams) we found in your team's design, some of the memset function calls were optimized out by the compiler in the resulting assembly. How do you think of this?
 - (ii) If not, were you aware that these techniques existed before the competition?
 - (iii) Did you know that the compiler may optimize away some code, which in the compiler's view, has no effect on the later code?

- (iv) Did you know that the compiler may treat some of the security-related code as unnecessary and optimize it away?
- (2) (Some teams) *We found in your team’s design, you implemented your own inline wiping function instead of using those provided by the library.*
 - (a) Why did your team choose to do this?
 - (b) Do you know your inline function to erase the memory is translated into memset in the resulting assembly?
 - (c) Did you intend to do this, or you didn’t notice what happened during the compilation?
- (3) (Some teams) *We found in your team’s design, you used the memory wiping functions provided by the crypto libraries.*
 - (a) Why did your team choose to use the wiping functions provided by the crypto libraries, instead of using the libc standard functions such as memset?
 - (b) Did you know that libc’s wiping functions could be optimized?

D.3 Questions related to memory-safe programming

- (1) *Embedded systems are usually developed in memory unsafe low-level languages such as C, this is also the language that the reference design used. A memory-safe language, such as Rust, aims to significantly reduce memory corruption vulnerabilities in a program while keeping the performance high if used in a proper way.*
 - (a) What were the factors you consider when you choose which language to use?
 - (b) (Teams used C/C++) *We found your team did not choose to use a memory-safe language, such as Rust.*
 - (i) What stopped you from using a memory-safe language?
 - (ii) Would you consider using it in the future? Why or why not?
 - (iii) Do you believe that utilizing a memory-safe language like Rust can eliminate memory vulnerabilities in embedded systems?
 - (skip the rest of questions in this section)
 - (c) (Teams used Rust) *We found your team used the Rust programming language in the design.*
 - (i) What was your level of experience in using the Rust programming language?
 - (ii) What were the challenges when using Rust on this microcontroller-based system?
 - (iii) Would you use it again if you still participate in the future? Why and why not?
- (2) (Some teams) *We found in your team’s design, you mixed C and Rust for the implementation.*
 - (a) What motivated your Rust/C hybrid design?
- (3) *The Rust programming language also provides the unsafe code block which allows developers to violate the memory safety rules. Abusing the unsafe code block could void the memory safety feature of Rust. But sometimes it is necessary to use it for embedded system development.*
 - (a) When or where did you feel you have to use unsafe blocks?

- (b) How did you approach the use of unsafe blocks in your Rust code? Did you have any security concerns when using them, and did you try to minimize their usage?
- (c) What approaches did you use to minimize the number of lines of code in unsafe blocks?
- (4) Did you or your team run into any significant roadblocks in using Rust, such as excessively large binaries, that you had to work around? Did you have to make any compromises in the process?

D.4 Questions related to RNG

- (1) *In cryptography, a Random Number Generator (RNG) is essential for creating secure cryptographic keys, generating nonces, and ensuring randomness in various security protocols.*
 - (a) Did your team use the random number generator in the design?
 - (i) If so,
 - did your team implemented your own RNG, or just used one that provided by the library/SDK? Why?
 - how confident were you in your team’s RNG implementation?
 - did you have any concerns about the robustness of the vendor-provided RNG?
 - what entropy sources did your team choose to seed the RNG? Why?
 - how confident were you in your entropy design?
 - how did your team test the effectiveness of the RNG design?
 - (ii) If not, why didn’t your team consider including randomness in the design?
 - (b) (2023) Did you know the board had an internal temperature sensor that could serve as an entropy source? How did you know it?
 - (c) (2024) Did you know the board had a True Random Number Generator (TRNG)? How did you know it?

D.5 Questions related to threat model

- (1) What is the threat model of your team’s design?
- (2) What are the particular types of attack that you considered to address in your design?
- (3) For each of the attacks that was not mentioned in the previous answer, e.g., brute-force, MITM, cold boot attack (to dump the memory), side-channel and fault injection, ask them the following questions:
 - (a) What is your understanding of this attack?
 - (b) Did you consider this attack in your threat model? Why not?
 - (c) How did your design mitigate this attack? How effective was that defense?
 - (d) How did the embedded nature of the system influence your approach?
- (4) In terms of attack surface, do you think the embedded system has a larger or smaller attack surface compared to a desktop system? Why?

- (5) Regarding the software vulnerabilities and potential hardware attacks, which one would you give more priority to defend against? Why?
- (6) *Questions related to comparing the user-controlled input with secret values.*
 - (a) When checking PIN value, which method or function did your team use to compare two values?
 - (b) Why did you choose this method/function?
 - (c) Did you know how the strcmp/memcmp function works? If so, how? (stop at the first difference)
 - (d) What issues do you think exist in these functions?
- (7) (Some teams) *We found in your team's design, random delays were added between some operations.*
 - (a) What was the purpose of adding these random delays?
 - (b) Where do you think are the best locations of adding these delays?

D.6 Questions related to vulnerability discovery

- (1) *There are many ways to find vulnerabilities in a system, such as checking the source code, analyzing the disassembly code of the compiled firmware, dynamically debugging the programs during the runtime, fuzzing, and so forth.*
 - (a) How did your team evaluate the correctness and security of your own design?
 - (b) How did your team do debugging when implementing your design on the board?
 - (c) How did your team find vulnerabilities in other teams' design during attack phase?
 - (d) Did your team use any static or dynamic techniques to detect errors?

D.7 Ending questions

- (1) Are there any other notable practical challenges that you encountered during the competition?
- (2) Is there anything else you would like to share with us?

E Codebook

The codebook contains theme names, theme descriptions, and associated sub-themes, as detailed in Table 2.

Table 2: Codebook.

Theme Name	Theme Description	Sub-theme
Knowledge of security principles	Statements that describe the awareness and understanding of general security knowledge.	<ul style="list-style-type: none"> * Participants were not aware of privilege separation. * Those that knew about privilege separation were unaware that it is offered by Cortex-M. * Other participants were not confident that privilege separation would improve the security of the system. * They believed that their logic lacked flaws, so privilege separation was not needed. * Participants were unaware of memory wiping as a defense technique. * Participants didn't believe that memory wiping would improve the security of their system. * Participants were not aware that the compiler optimized their memory wiping out. * Some were aware that the compiler optimizes out unnecessary code, but none were aware that memory setting would be optimized out.
Knowledge of tools/features	Statements that describe the understanding embedded system development tools/features.	<ul style="list-style-type: none"> * Participants were not aware that the NX compiler flag did not make memory non-executable on the microcontroller. * Participants were unaware that they could use the MPU to set regions of memory as non-executable. * Participants were not aware that some C standard library functions performed differently on the embedded system compared to a general-purpose system (e.g., time(), printf()).
Knowledge of cryptography	Statements that describe participants' knowledge of using cryptography on an embedded system.	<ul style="list-style-type: none"> * Participants were not aware of the different modes offered by AES, and therefore did not choose the best mode for their system. * Participants were not aware of the best cryptographic algorithms for their embedded system. * Participants might not know that standard cryptographic operations on a general-purpose system might be infeasible on a microcontroller due to processor capability (e.g., clock speed).
Attack and defense	Statements that describe the awareness of embedded-prone attacks and how participants defend them.	<ul style="list-style-type: none"> * Participants were not always aware of embedded-prone attacks that they must implement defenses for. * Even if participants were aware of embedded-prone attacks, they are not always aware how to defend against them. * Even if participants are aware of embedded-prone attacks and they try to implement a defense, their defense is conceptually flawed.
Embedded Memory	Statements that describe challenges of using embedded system memory.	<ul style="list-style-type: none"> * Participants mentioned that memory bugs are more severe because everything (Flash, SRAM, peripherals) is accessible in a single memory space. * Participants mentioned that storing sensitive data is more difficult on microcontroller systems. * Participants mentioned that the lack of dynamic memory management and OS on the microcontroller systems poses security challenges and may downgrade the security design. * Participants mentioned that the limited flash and SRAM space in statically linked microcontroller systems makes it harder to use third-party libraries. * Participants mentioned that testing and debugging is harder on a microcontroller compare to the microprocessor systems.
Embedded Rust support	Statements that describe participants' challenges of using Rust for embedded system development.	<ul style="list-style-type: none"> * Participants thought Rust is useful because it eliminates virtually all memory safety bugs. * Participants mentioned having to remove the Rust standard library because there is no operating system support on the embedded system (e.g., memory allocation). * Participants mentioned some libraries not running because the Rust standard library wouldn't work on the embedded system. * Participants mentioned having no direct vendor support for getting Rust to compile to the board. * Participants mentioned encountering excessive stack memory consumption because of how the Rust compiler constructs objects. Because embedded systems are resource constrained, using excessive stack memory will eventually cause issues. * Participants mentioned having to write a HAL, perhaps using unsafe blocks, in order to let Rust access board hardware.
Experience with Rust	Statements that describe participants' understanding of the Rust language.	<ul style="list-style-type: none"> * Participants thought Rust is useful because it eliminates virtually all memory safety bugs. * Participants that didn't use Rust cited lack of experience / programming knowledge as a reason. * Despite their lack of knowledge about the language, participants believed that Rust eliminates memory bugs.
Security-related compiler flags	Statements that describe the knowledge of security-related compiler flags.	<ul style="list-style-type: none"> * Participants did not notice any warnings from the compiler when security-related flags have no effect. * Participants tried to use NX, or tried to modify the linker script, but none of them knew that these were not effective. * Some participants were aware of security concepts like NX, but were not sure how to apply it to their design. * Participants did not know that because the ELF header is removed when the firmware is flashed to the board, any bits that are set in the header are always discarded.
Lack of entropy source	Statements describe the awareness and knowledge of entropy source of embedded systems and how participants get random numbers.	<ul style="list-style-type: none"> * Participants aware that an entropy source is needed to seed the random number generator for cryptographic operations. * Participants may unaware that general-purpose systems get their entropy from system time, cursor location, sources of noise, etc. * Participants were not aware the existence of the CPU temperature sensor which was a good source of entropy. * Participants who aware the CPU temperature sensor all used it as the entropy source. * Participants mentioned the use of mixed entropy sources. Build time entropy supplied by host, run time entropy from systick timer, cycle count, uninitialized SRAM (didn't work). * Participants were aware the board vendor provided an API for true random number generator. * Participants mentioned only used one secure of entropy, such as build-time entropy, or no entropy. Some teams used rand() and time(), which doesn't work on microcontroller. * Participants had challenges in testing their RNG design on the microcontroller. * Participants were unaware that the same library might work different between a microprocessor system and a microcontroller, e.g., the libc standard library time() function, and no warning is given to the user.