Abstract— Recent technological advancements have proliferated the use of small embedded devices for collecting, processing, and transferring the security-critical information. The Internet of Things (IoT) has enabled remote access and control of these network-connected devices. Consequently, an attacker can exploit security vulnerabilities and compromise these devices. In this context, the secure boot becomes a useful security mechanism to verify the integrity and authenticity of the software state of the devices. However, the current secure boot schemes focus on detecting the presence of potential malware on the device but not on disinfecting and restoring the software to a benign state. This manuscript presents CARE - the first secure boot framework that provides detection, resilience, and onboard recovery mechanism for the compromised devices. The framework uses a prototype hybrid CARE: Code Authentication and Resilience Engine to verify the software state and restore it to a benign state. It uses Physical Memory Protection (PMP) and other security enforcing techniques of RISC-V processor to provide resilience from modern attacks. The state-of-the-art comparison and performance analysis results indicate that the proposed secure boot framework provides promising resilience and recovery mechanism with very little (8%) performance and resource overhead.

Keywords— hardware and system security, HW/SW co-design, SoC, secure boot, attack resilient, smart recovery, small embedded and IoT devices security system, RISC-V

I. Introduction

The recent technological advancement has catastrophically increased the utilization of small embedded and IoT devices in applications ranging from industrial control systems, distributed sensing and actuation, vehicular and home automation systems. This increased utilization and inter-connectivity for collecting, transferring, and processing the security-critical information has made the small embedded and IoT devices, attractive targets for attacks. Prominent examples are the Rootkit [1], the bios and secure boot attacks [2], the Stuxnet [3], and the Jeep hack [4]. By-enlarge such attacks modify the targeted device’s software state to leak, steal, tamper, or misuse the security-critical information for malicious activities. Such attacks can render the device into an unusable state. These types of attacks are commonly referred to as malicious code modification attacks or malware infestation.

Secure boot process verifies the integrity and authenticity of devices’ software state during boot time and ensures that the device boots-up with a known good code. Several secure boot methods have been proposed based on hardware [5,6], software [7,8], and hardware/software co-design [9,12]. While prior secure boot techniques focus on the detection of malicious code modification attacks, the problem of disinfecting the affected devices has been totally overlooked. The conventional device needs over-the-air or manual code re-flash to restore its normal operational state. A smart attacker can fail over-the-air code re-flash by corrupting the networking stack. This necessitates manual intervention. Sometimes the manual re-flash becomes relatively difficult due to the placement (in home security sensors and cameras, industrial and automotive control systems, ships) of the devices.

To bridge this gap, the proposed work presents CARE - first lightweight secure boot framework that provides detection, resilience, and onboard recovery mechanism for small embedded and IoT devices.

![Fig. 1](image.png)

**Fig. 1**

**DEPICTS THE PLACEMENT OF THE CUSTOM CARE MODULE (PLACED IN BETWEEN THE FIRST STAGE BOOT CODE (ROM) AND SECOND STAGE BOOT FLASH) THAT MONITORS THE CODE INTEGRITY AND AUTHENTICITY DURING SECURE BOOT, AND RESTORES THE CORRUPTED FLASH MEMORY REGION WITH THE GOLDEN RECOVERY DATA FROM SECURE ROM.**
again and continues the secure boot process. This method ensures that irrespective of any malicious code modification attacks, the device will self recover and always boot up with a known good code.

The design, implementation, and evaluation of the proposed framework provides the following research contributions:

- **Code Integrity and Authenticity Measurement (CA) Tool:** It demonstrates the lightweight implementation of integrity and authenticity measurement tools by reusing the same underlying hardware cryptographic core.
- **Resilience Engine (RE):** It demonstrates the first implementation of onboard resilience and recovery engine for small embedded and IoT devices.
- **Lightweight Secure Boot Architecture:** It provides FPGA prototype implementation of a lightweight, secure boot framework CARE for small embedded and IoT devices. It enhances the attack resilience and security of the system by leveraging Side-Channel Analysis (SCA) and fault injection attack protection features of the RISC-V processor.

II. Background and Related Work

Arbaugh et al. has proposed the first secure boot mechanism [13], which measures the integrity of the system by verifying the integrity of the boot software code(stages). It performs a measured boot in which every stage verifies the subsequent stage’s integrity before it gets executed. Authenticated boot verifies that software running on the system is coming from an authorized vendor. The Unified Extensible Firmware Interface (UEFI) specification since version 2.2 [6] defines secure boot as a process to verify the integrity of each stage of the boot process by digest computation and comparing the result with a cryptographic signature. It requires access to a trustworthy public key database to verify the signature. The majority of the previous implementation of the secure boot systems performs either measured or authenticated boot, and very few perform both.

One of the popular methods for the secure boot is to use a discrete co-processor called the Trusted Platform Module (TPM) [9]. TPM has a special purpose registers called Platform Configuration Registers (PCRs), which cannot be overwritten. PCR’s can only be extended by hashing the software measurements together with the previous values of PCR. TPM can sign the PCRs with a private attestation key to generate a piece of attestation evidence. However, TPM is not suitable for small embedded or IoT devices due to space, size, and cost constraints. Intel’s processor supports two modes of the secure boot - measured and verified and uses microcode as root-of-trust [14]. For measured boot, it uses TPM, and for verified boot, each component is signed by the manufacturer’s key, and signatures are verified before loading the component. Microsoft’s fTPM [10] provides a use-case of Arm TrustZone based secure boot and attestation. RISC-V based Sanctum [7] uses software-based secure boot and remote attestation. SMART [8] provides dynamic root-of-trust architecture for low-end devices. Keystone [12] showcases a use case of Trusted Execution Environment (TEE) with enclaves. Haj et al. [5] presents lightweight hardware-based secure boot architecture for RISC-V based SoC. Google’s recent open-source root-of-trust project Opentitan [11], provides a sample implementation of secure root-of-trust.

However, none of the available solutions have a secure recovery mechanism. Recently implementation Healed [15] and [16] demonstrates recovery mechanisms. however, they both lack in proper secure boot implementation. To the best of our knowledge, the proposed work is the first implementation of a lightweight secure boot architecture with onboard resilience and recovery engine for small embedded and IoT devices.

III. Adversarial Model & Protection Axioms

The proposed system assumes that the adversary can control the entire software code and data. The adversary can modify any writable memory and read memory region that is not protected by access control policies (using PMP) and secureIbex hardware features. The adversary can re-locate malware from one memory segment to another for hiding it from being detected. It also has full control over all Direct Memory Access (DMA) to access the main memory directly (e.g., RAM, flash or ROM) without going through the processor core.

The proposed solution blocks un-authorized read, write, and code execution triggered from non-secure flash memory (to ROM) by applying PMP access control rules. It also leverages the special security feature (secureIbex) of the Ibex processor to protect the device from data independent timing - side-channel attack and fault injection attacks.

IV. SYSTEM OVERVIEW

A. Architecture and Design Choices

The proposed secure boot system is built upon the Ibex RISC-V processor. The system is equipped with hardware-accelerated code integrity and authentication measurement (CA) unit, recovery engine (RE), secure boot, secure memory (ROM), and dedicated SPI bus as shown in Fig 2. Notice that secure storage ROM has numbers 1 and 2, because it is used to store secure signing key, device information, and the recovery data. The flash controller module is used to translate read, erase, and program requests to low-level protocol signaling and timing. The proposed framework has incorporated the following security enhancing design features: 1. **Secure Storage ROM** is used for storing the device information such as vendor ID, Unique Device Identification (UUID), firmware revision, symmetric cryptographic shared key (K), and trusted recovery image. 2. **Secure SPI bus** is used for communication between the ROM, flash, and CARE module to protect the device from attacks launched using a shared internal bus [18]. However, if the SPI tool’s hardware or software gets corrupted, it can render and transfer incorrect or corrupted flash data. Therefore, the proposed design divides the flash image into 1 KB frame/chunks.
covers the details of the frame data structure. Note that, this design choice is used for proof of concept implementation only and user can parameterize it to optimize the system performance. **3 Code Integrity & Authentication (CA) Unit** is implemented by reusing the same underlying hardware lightweight cryptographic-core (HMAC-SHA256), which performs both integrity and authenticity checks. This hardware reuse makes the proposed framework lightweight and resource-efficient, suitable for small embedded and IoT devices. **4 Resilience Engine (RE)** is implemented in software and it re-flashes the affected flash memory region during the secure boot. It applies access control policies to protect the device from future attacks. **Ibex Core** The Ibex core provides memory protection and access control using PMP. It also provides resilience from fault injection, data independent timing attacks by leveraging secureIbex feature. It inserts dummy instructions (such as NOP) at random interval to protect the system from side channel attacks. It performs ECC checking of flash blocks to protect the system from fault-injection attacks.

### B. System Operation

The architecture design of the proposed framework is shown in Fig. 3 and the system operation is divided into three main steps: (1) System Initialization; (2) Code Integrity and Authenticity Check (Bootstrapping); and (3) Resilience Engine (RE).

1) **System Initialization:** Upon power-on, the system locates and executes the First Stage Boot Loader (FSBL) code from secure ROM to initialize the SPI and flash controllers. It then applies memory protection (using PMP) rules, and blocks unauthorized code read, write, and execution initiated from unprotected flash memory. It reads the chip information such as - device UUID, board version, and symmetric share key, generates derived keys, and hands off the control to the second stage boot code called the bootstrap.

2) **Bootstrapping System:** The secure boot starts with the bootstrapping process. It can be triggered by the hardware reset, power-on, or triggering from the external host via the General Purpose Input Output (GPIO) pin seven in the proposed framework. When the bootstrap is activated, the executable flash image is broken down into 1 KB frame chunks and sent sequentially to the host over the SPI bus. Each frame consists of a header and associated payload, as illustrated in Fig. 4. The header contains the signed digest of the frame data. The offset location field indicates the flash memory location, which will be used for code re-flash. The payload contains 968 bytes of the data for each frame.

The proposed framework has leveraged Hashed based Message Authentication Code (HMAC)'s feature HMAC-SHA256 for signing the data and perform signature verification and SHA256 for digest computation. Few APIs were developed to reuse the same underlying cryptographic hardware HMAC-SHA256. The HW-SHA256 module computes the digest of each 1KB frame and compares it with the pre-computed hash for an integrity check. The HMAC-SHA256 uses a derived key to sign the computed digest.
matches it with the "Hash" field in the frame header for authenticity check. The framework follows the same bootstrapping process for each subsequent frame. If everything passes, the device boots up with verified code, else it triggers the resilience engine.

3) Resilience Engine (RE): The resilience engine acts as follows: (1) It identifies the frame number and offset location of the corrupted frame, and locates the corresponding golden frame data from the secure EEPROM. (2) It exclusively re-flashes the corrupted flash memory region with the known good code. (3) It locks unauthorized read-write access to the memory using the PMP mechanism. These steps ensure that the proposed device will always boot up with a known good code irrespective of any memory modification attacks.

V. Evaluation

This section describes the chain-of-trust theory, resource utilization, and performance analysis for each sub-module and overall framework in the proposed framework, and presents state-of-the-art comparison results.

A. Chain-of-Trust

The proposed framework breaks down the entire flash image into 1 KB chunks/frames and measures the code's integrity and authenticity at the frame level. The following equation represents the chain-of-trust:

\[ V_0 = True \]

\[ V_{i+1} = V_i \land S_R(f_{i+1}) \land I_R(f_{i+1}) \]  

(1)

\[ V_i \] is the boolean value representing the software state verification (both integrity and authenticity) of the ith frame, and \( \land \) is the boolean AND operation. \( I_R \) and \( S_R \) represents the integrity and signature (authenticity) verification functions respectively. \( I_R \) takes frame data \( f_{i+1} \) as input argument, performs a cryptographic hash, compares the result with golden digest value, and returns a boolean result. The \( S_R \) signs the digest, matches with frame header value, and gives a boolean result. The following equation\ref{eqn} calculates the estimated increase in boot time \( (T_{\Delta}) \) for proposed secure boot with CARE.

\[ T_{\Delta} = t_{im,0}(I_0(f_1) + S_0(f_1)) + t_{im} \sum_{i=1}^{n} [I_i(f_{i+1}) + S_i(f_{i+1})] \]  

(2)

Where \( t_{im,0} \) and \( t_{im} \) are the execution time for the first and all other frames, respectively. By design, the framework, first matches the frame number of the received frame and clears the flash region to re-flash it with a trusted code. Therefore, the first frame processing requires more time than the remaining frames. In case of verification failure of any frame, the framework triggers RE sub-module, which intern re-flashes the corrupted frame with know good recovery data. This process increases the boot time by a fraction, as discussed in subsection \ref{sec:resilience_engine}. The system analysis is carried out by evaluating performance, hardware-software resource utilization, and energy consumption of the CA unit, RE, and overall system.

B. Code Integrity and Authentication (CA) Unit

The cryptographic-core (HMAC-SHA256) is the key component of the Code Integrity and Authentication (CA) unit. The test setup first uses both hardware and software implementation of cryptographic-core running on FPGA for performance evaluation, as shown in Table 1. The system computes the digest of 256 Bytes of data for performance and energy efficiency evaluation. Table 1 shows the performance increase of 16x with 92% less power utilization using hardware-based cryptographic-core. Furthermore, the proposed cryptographic-core is lightweight and consumes less energy than the recent state-of-the-art HMAC-SHA256 implementations, as depicted in Table 2. As seen from Table 2, \ref{table:cross_comparison} requires a relatively low area but consumes high energy. \ref{table:cross_comparison} presents base and DPA resilient cryptographic-core but uses more area and power. The work presented in \ref{table:cross_comparison} optimizes the core for high throughput while compromising area and energy consumption, which makes both of them unsuitable for small embedded and IoT devices. The cryptographic-core used in the proposed work is an area and energy-optimized version of opentitan \ref{table:cross_comparison}.

C. Resilience Engine (RE)

The Resilience Engine (RE) sub-module is implemented in software for the Proof-Of-Concept (POC) work. The test application of 5.6 KB is used for POC validation. RE requires \ref{table:resilience_engine} additional lines of code (C language) for secure boot, and increases the secure ROM by 5 KB to store recovery data. The Resilience Engine (RE) requires 968 bytes of recovery data for every 1 KB of the flash image. To limit the size of the recovery data storage on the ROM, the system developer can select the necessary code modules for

\begin{table}  
\centering  
\caption{Cross Comparison of Crypto-core on FPGA.}  
\begin{tabular}{|c|c|c|c|}  
\hline  
Work & Area & Freq. & Energy Con. \  
\hline  
This work & 2591 & 100 & .012 \  
Opentitan \cite{21} & 2693 & 100 & .022 \  
He et al. \cite{20} & 7219 & 116.24 & 1.20 \  
He et al. \cite{20} & 10918 & 87 & 1.80 \  
Juliato et al. \cite{22} & 2347 & 138.10 & .48 \  
Juliato et al. \cite{22} & 4281 & 67 & .285 \  
Juliato et al. \cite{22} & 6874 & 41.25 & .431 \  
\hline  
\end{tabular}  
\end{table}
recovery to bring the system to a minimum working state. Although this feature is not implemented in the presented work due to small test applications.

D. System Performance

The system divides a test application of 5.6 KB into six 1 KB frames and performs integrity and authenticity checks for the performance evaluation. The total boot-time and energy consumption with and without CARE based secure boot is calculated for the test application running on FPGA. The timing analysis details are depicted in Table III. The framework uses equation 2 to calculate the total execution time \( T \). As explained earlier, the first frame requires more cycles and time. The rest of the frames consume an equal number of cycles. The secure boot with CARE consumes 8% more energy and requires additional \( D_\Delta = 529\mu\text{sec} \) boot-time. The proposed RE sub-module requires an additional 334.475 \( \mu\text{sec} \) to re-flash 968 bytes of data for each affected frame. This performance overhead (only 8% for the test application) is insignificant compared to the security and resilience, it provides.

E. Comparison with the state-of-the-art solutions

The majority of the available secure boot implementations focus on detecting and preventing malicious code modification attacks. They generally stop the code execution or resets the system to protect it from attacks. These systems largely lack in providing recovery mechanism. Furthermore, our architecture was design using open-sourced RISC-V ISA, which is relatively new, and we did not found any secure boot implementation on RISC-V that provides the recovery mechanism. Therefore, we chose to compare the (quantitative) hardware footprint requirements of the recently proposed RISC-V based secure boot architectures with this work. The qualitative and quantitative comparison of the proposed secure boot framework with state-of-the-art solutions are presented in Table IV and Table V.

Qualitative Comparison: Table IV shows that CARE, [11] and [23] are hybrid secure boot systems. All three uses cryptographic-core SHA256 for integrity checking, CARE and [11] uses HMAC-SHA256, and [23] uses AES for authentication check. [23] uses a discrete TPM module connected to FPGA for secure boot.

Haj et al.[5] is a pure hardware-based secure boot with TEE and resource-heavy cryptographic-cores (ECDSA Table V and sha3). Another implementation sanctum [7] provides a software-based secure boot by using secure enclaves. All three of them ([23], [5], [7]) are resource heavy and not suitable for our targeted small embedded and IoT devices. Only CARE and [11] opentitan are lightweight solution. However, the baseline opentitan does not have support for cryptocore (HMAC-SHA256) for authenticity check, PMP, secureIbex register, and onboard recovery engine such as CARE. In addition, reuses the same hardware HMAC-SHA256 reuse for both integrity and authenticity check makes CARE lightweight and suitable for our targeted devices.

Quantitative Comparison: Table V enumerates the quantitative comparison of CARE, [11] opentitan and Haj et al.[5] systems. Since [23] uses discrete TPM module attached to FPGA for secure boot, the architecture design becomes different and heavy. Therefore, it is not suitable for quantitative analysis. Note that Haj et al.[5] does not provide complete secure boot SoC hardware footprint. Therefore, that row has N/A - Not Available for LUTs, Regs, and Cells fields.

Table V shows percentage hardware overhead of the cryptographic-core for all three solutions. (Note: "Optitani" name is used for Opentitan representation only). The ECDSA core from Haj et al. requires 90% more hardware resources than HMAC from CARE. In-fact, the area required by Haj et al. based cryptographic-core is 14x larger than that of CARE’s cryptographic-core. The comparison of asymmetric and symmetric cryptographic hardware requirements provide an initial estimation of overall hardware overhead requirements. Additionally, Haj et al.[5] requires two 64 bit RISC-V cores for Trusted Execution Environment (TEE) implementation, hardware SHA3 for hashing, and configurable LFSR-based Physical Unclonable Function (CoLPUF) for key generation, boot sequencer, and key
management unit. These makes it a resource-heavy solution and unsuitable for small embedded and IoT devices. The percentage (%) overhead of complete SoC is calculated between CARE and opentitan [11] only, as Haj et al. [5] does not provide details of SoC hardware footprint. Table 6 depicts that the complete SoC with CARE utilizes 29.66% less area. Table 7 shows extended ROM region by 5 KB, and Table 8 depicts performance and energy consumption overhead of 8% each compared with opentitan secure boot SoC.

VI. DISCUSSION

A possible alternate of this work is a method that allows the device to boot from an SD card or a USB flash drive. However, providing physical security to an external card is extremely difficult. It can result in malicious actors replacing any hardware or software module in the system without security controls if the external card is lost or stolen. Second, the practicality of storing the golden image on EEPROM and not on the flash, as ROM’s cost decreases (~ .50 cents for 32 KB, when bought in bulk), making it an affordable alternative with more security. The CARE based SoC design is limited to support for small embedded and IoT devices, which does not require frequent application code update. However, the design can use EEPROM to update the recovery image if the system needs it.

VII. CONCLUSION

This paper has presented a lightweight, secure boot framework with an onboard recovery and protection mechanism for small embedded and IoT devices, to protect it from malicious code modification attacks. It provides code modification attack detection, recovery, and prevention tools that assure the user that the device will always boot with a known good code. The framework achieves these by using a prototype CARE module. It reuses the same cryptographic-core for authenticity and integrity check. The comparison of the proposed solution with the state-of-the-art secure boot implementations demonstrate that the proposed framework shows promising resilience and recovery methods with only 8% performance and energy consumption overhead and a minimal increase in hardware-resource utilization.

REFERENCES

[1] Comodo, “Comodo creating Trust online,” https://enterprise.comodo.com/rootkit-definition/, 2019.
[2] A. Furtak, Y. Bulygin, O. Bazhanik, J. Loucaides, A. Matrosov, and M. Gorobet, “Bios and secure boot attacks uncovered,” 2014.
[3] J. Vijayan, “Stuxnet renews power grid security concerns,” https://www.computerworld.com/article/2519574/stuxnet-renews-power-grid-security-concerns.html, June 2010.
[4] D. Schneider, “Jeep hacking 101,” http://spectrum.ieee.org/cars-that-think/transportation/systems/jeep-hacking-101.html, 2015.
[5] J. Haj-Yahya, M. M. Wong, V. Pudi, S. Bhaisin, and A. Chattopadhyay, “Lightweight secure-boot architecture for risc-v system-on-chip,” in 20th International Symposium on Quality Electronic Design (ISQED), March 2019, pp. 216–223.
[6] NSA Cyber Report, “UEFI DEFENSIVE PRACTICES GUIDANCE,” https://www.nsa.gov/Portals/70/documents/what-we-do/cybersecurity/professional-resources/ctr-uefi-defensive-practices-guidance.pdf, 2017.
[7] I. Lebedev, K. Hogan, and S. Devadas, “Invited paper: Secure boot and remote attestation in the sanctum processor,” in 2018 IEEE 31st Computer Security Foundations Symposium (CSF), July 2018, pp. 46–60.
[8] M. M. Wong, J. Haj-Yahya, and A. Chattopadhyay, “Smarts: secure memory assurance of risc-v trusted soc,” 06 2018, pp. 1–8.
[9] “Trusted Platform Module by Trusted Computing Group,” https://trustedcomputinggroup.org/work-groups/trusted-platform-module/, 2010.
[10] H. Raj, S. Saroiu, A. Wolman, R. Aigner, J. Cox, and P. England, “Ftpm: A software-only implementation of a TPM chip,” in 25th USENIX Security Symposium (USENIX Security 16). Austin, TX: USENIX Association, Aug. 2016, pp. 841–856. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/ra.
[11] “Opentitan: Secure boot SoC design,” https://opentitan.org/, 2019.
[12] D. Lee, D. Kohlbrenner, S. Shinde, D. X. Song, and K. Asanovic, “Keystone: A framework for architecting tees,” ArXiv, vol. abs/1907.10119, 2019.
[13] W. A. Arbaugh, D. J. Farber, and J. M. Smith, “A secure and reliable bootstrap architecture,” in Proceedings. 1997 IEEE Symposium on Security and Privacy (Cat. No.97CB36097), May 1997, pp. 65–71.
[14] X. Ruan, “Boot with integrity, or dont boot,” in Platform Embedded Security Technology Revealed, Springer, 2014.
[15] A. Ibrahim, A.-R. Sadeghi, and G. Tsudik, “Healed: Healing and attestation for low-end embedded devices,” in Financial Cryptography, 2019.
[16] D. Perito and G. Tsudik, “Secure code update for embedded devices via proofs of secure erasure,” in ESORICS, 2010.
[17] ETH zurich, “Ibex : lowrisc 32-bit processor core,” https://ibex-core.readthedocs.io/en/latest/, 2018.

[18] N. Jacob, J. Heyszl, A. Zankl, C. Rolles, and G. Sigl, “How to break secure boot on fpga socs through malicious hardware,” IACR Cryptology ePrint Archive, vol. 2017, p. 625, 2017.

[19] G. Inc, “Google chromiumos third party cryptoc factory grub 8652 b - sha256,” https://chromium.googlesource.com/chromiumos/third-party/cryptoc/-/factory-grub-8652-B/sha256.c, 2016.

[20] Z. He, L. Wu, and X. Zhang, “High-speed pipeline design for hmac of sha-256 with masking scheme,” in 2018 12th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID), 2018, pp. 174–178.

[21] M. Juliato and C. H. Gebotys, “Fpga implementation of an hmac processor based on the sha-2 family of hash functions,” 2011.

[22] M. Juliato and C. Gebotys, “An efficient fault-tolerance technique for the keyed-hash message authentication code,” in 2010 IEEE Aerospace Conference, 2010, pp. 1–17.

[23] A. S. Siddiqui, G. Shirley, S. Bendre, G. Bhagwat, J. Plusquellic, and F. Saqib, “Secure design flow of fpga based risc-v implementation,” in 2019 IEEE 4th International Verification and Security Workshop (IVSW), 2019, pp. 37–42.