

Article

Secure Boot for Reconfigurable Architectures

Ali Shuja Siddiqui ¹, Yutian Gui ¹ and Fareena Saqib ^{1,*}

¹ Dept. of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC 27599, USA; asiddiq6@uncc.edu; ygui@uncc.edu

* Correspondence: fsaqib@uncc.edu

Abstract: Reconfigurable computing is becoming ubiquitous in the form of consumer-based Internet of Things (IoT) devices. Reconfigurable computing architectures have found their place in safety-critical infrastructures such as the automotive industry. As the target architecture evolves, it also needs to be updated remotely on the target platform. This process is susceptible to remote hijacking where the attacker can maliciously update the reconfigurable hardware target with tainted hardware configuration. This paper proposes an architecture of establishing Root of Trust at the hardware level using cryptographic co-processors and Trusted Platform Modules (TPMs) and enable over the air updates. The proposed framework implements secure boot protocol on Xilinx based FPGAs. The project demonstrates the configuration of the bitstream, boot process integration with TPM and secure over-the-air updates for the hardware reconfiguration.

Keywords: Secure boot; Cyber-physical system security; Embedded systems; FPGA; Hardware primitives; IoT security

1. Introduction

Internet of Things (IoT) are ubiquitous devices with limited functionality and computational resources, enabled with networking features and connectivity to the internet. These devices have long life cycle, where the updates/changes are deployed through software or firmware updates. Software updates can be provided either manually or physically by using cables and programmer interfaces. Physical access to a device for firmware upgrade is always not possible and requires the configuration of the devices in the field through secure channels. The connected nature of IoTs makes them accessible remotely, and firmware updates for devices can be provided using Over the Air (OTA) firmware updates [1][2].

Software firmware updates change the functionality of the software; however, the scope is limited by the hardware capabilities and its architecture. Depending on the application and the expected life of a product, reconfigurable hardware architectures can improve the device updates and secure boot processes. Reconfigurability allows the manufacturer to update the hardware design while the machine is in the field. Using OTA, hardware updates allow for updating the device hardware configuration without the need for physically replacing it.

Over the air (OTA) updates are critical in the embedded system consumer domain such as the cellular phone and the automotive industry. The requirements for availability and quality of service are high in safety-critical applications. For the vehicular domain, the requirement for availability is directly tied to safety. Furthermore, it is not feasible for a car owner to take the car to a service station, whenever there is a software update available. Instead, firmware updates can be provided remotely with OTA updates that can be transferred via the cellular network directly by the manufacturer [3].

Root of Trust is an anchor for implementing trust in a device [4]. Maintaining Root of Trust (RoT) is crucial once a device has been deployed in an untrusted field. Tainted firmware updates can break the trust. The software domain employs various techniques for maintaining Root of Trust. Some popular examples include Universal Extensible Firmware Interface (UEFI) secure boot extensions [5] and Microsoft Windows' Secure Boot [6]. The former provides maintaining Root of Trust at the boot

level and the latter extends it to the operating system (OS) level. We discuss novel scheme of root of trust at using cryptographic processors, such as Trust Platform Modules (TPM).

Reconfigurable hardware has become pervasive in the Internet of Things domain, this is a requirement for extending the root of trust to the hardware. Current commercial FPGA vendors provide limited security to the programmable logic fabric and those security mechanisms are limited in application, as is shown in section II. Additionally, the provided methods have closed access which the users can use in their systems but cannot inspect themselves. There is a need for an open and reliable security structure for the programmable fabric that can be integrated into the device's Root of Trust.

In this work, we present a framework to establish the Root of Trust for secure boot and OTA updates for reconfigurable hardware. Mutual trust is established between an FPGA device and the server or the content provider. A symmetric encryption key update mechanism is proposed to enable key provisioning for updates. The key update mechanism is tied to the integrity of the bitstream. If there is an unintended change in the integrity of the bitstream, the device will no longer be able to successfully decrypt continuing updates from the server. The proposed scheme extends the integration of TPM with the FPGA boot process to assist in secure boot, key provisioning, and secure communication. Additionally, we present schemes for runtime mitigation of malicious logic insertion. To the best of the authors' knowledge, this is the first work that practically provides the integration between TPMs and FPGAs at the First Stage Boot Loader stage. The organization of the paper is as follows. Section II discusses the relevant background studies. Section III covers the proposed architecture and analysis of the methodologies. Section IV describes the implementation details and security analysis is discussed in Section V.

2. Background Studies

Secure Boot in FPGAs

The security features in FPGAs target various aspects of security, such as secure boot, encryption, data integrity and secure communication. Commercial FPGA vendors have increasingly incorporated new security features over the years, but they do not follow any standard and are not consistent across the devices. In Zynq 7000 FPGAs, the FPGA SoC provides an AES 256 based encryption engine. Additionally, RSA asymmetric authentication is used to ensure an authenticated source. These functions are implemented as a hardware-based function on the FPGA fabric and their configuration is integrated in the EDA tools. The implementation of this core is not open to the end-user [7]. Both the encryption and authentication processes use keys which are defined before deployment. In the Zynq 7000 architecture, there are two ways of key storage, namely Battery-Backed RAM (BBRAM) and one-time programmable fuses. BBRAM is an on-chip volatile memory region, that has to be battery powered. One-time programmable efuses are configurable fuses that are burned into the fabric once they are programmed. There are caveats of using these technologies for holding keys. The BBRAM is dependent on the presence of a dependable power source. In the case there is any fluctuation in the power supply, the keys will be lost. On the other hand, the efuses are once programmable. In the case the programmed keys leak, there is no way to refresh them. Another serious flaw is that these keys are stored unencrypted in the memory. Access can be gained to them through physical means [8] and through side channel attacks [9].

These cryptographic processes are invoked as part of the secure boot process implemented in the Xilinx provided zeroth stage boot loader called BootROM [10]. BootROM code exists in non-volatile memory on the FPGA fabric. It is executed by the Processing System on FPGA at the beginning of the boot cycle. Once the execution is completed by the BootROM code, control is passed to user-defined code, which can be an operating system or a user level application. The vendor provides no access to the code implementation of BootROM and neither is there any read or write access to the memory holding the BootROM code.

Much recently, in the newer Zynq UltraScale architecture, Xilinx has built upon the schemes in the Zynq 7000 architecture to add more security solutions [11][12]. The major revision is in the

addition of Physical Unclonable Function (PUF) based key support. Whenever keys are generated for encrypting bitstreams, they are given as input to the on-board PUF to generate an encrypted key. This key can be stored onto the efuses, the BBRAM or externally in any unencrypted space. Since an attacker does not have access to the PUF implementation, they will not be able to decrypt the encryption key. Once the host system sends an encrypted bitstream to the FPGA board, the encrypted key is decrypted at runtime to reveal the bitstream decryption key. This key is used by the encryption subsystem to decrypt the bitstream.

Attacks on vendor provided security, such as Xilinx's Secure Boot have also been documented. Outsourced trusted third-party IPs, based on their application also have access to the main memory of the system. Since the main memory is also used by secure boot to store the unencrypted Second Stage Boot Loader (e.g. U-boot) used by the system, a hardware trojan can target the main memory to manipulate the Second Stage Boot Loader to execute a malicious executable. [13] suggests wrapper logic to be added to IPs before the IP is connected to the system bus. Hardcoded configuration of the wrapper defines the scope of memory access by the connecting IP.

Xilinx FPGA bitstreams consist of sequence of commands that are sent to the FPGA's Internal Configuration Access Port (ICAP) interface [14]. The ICAP interface is used to interact with the Programmable Logic fabric. There are several registers that an ICAP register provides. The bitstream header instructs if the bitstream is encrypted or not. This sets the FPGA to decrypt the bitstream data using AES-CBC mode. In [15], authors exploit CBC malleability to alter the bitstream. The malicious modification routes the decrypted bitstream configuration to be written to the fabric, to an alternative ICAP register. The leaked configuration is then collected to retrieve the entire confidential bitstream configuration. Recent work on the secure bitstream configuration at the boot level proposes the use of PUF technology and on-board peripherals for FPGA bitstream secure boot [16]. The ICAP is used to retrieve the configuration of the current programmable logic. A Hardware-Embedded Delay Physical Unclonable Function (HELPUF [17]) uses SHA3 digest as the challenge input. The generated output is used as a key for decrypting the image of the operating system and the application software to realize self-authentication. This enables the second stage boot loader to program the programmable logic PL and processing system PS. Since, vendor-provided secure boot is not used with this system, ICAP also allows readback of on-chip memory elements such as the block RAM and registers. If an adversary can capture the readback process, they will also be able to read the PUF output as well and therefore will have access to decryption keys. Another scheme implements self-authentication of the logic fabric design and extends the protection to include processor design [18]. As opposed to [16], this work has employed the use of Elliptic Curve Cryptography based asymmetric keys with Diffie Helman key exchange for the key generation for encryption. A fuzzy key extractor is used to use extract hardware-based variations and to generate a key. This extracted key is used in the design by various cryptographic functions such as remote attestation and encryption key generation.

Dynamic Partial Reconfiguration (DPR) allows reconfiguration of pre-defined sections of the FPGA fabric during runtime. Static design reconfiguration requires the FPGA to be shut down before it can be reprogrammed. Following current security standards followed by FPGA manufacturers [19][7], the static encryption key must also be shared with the third party. Thus, adding more actors who would have access to the key. In [20], authors have suggested improving on the situation, with the involvement of the hardware vendor in the distribution process. Each end device has a Unique ID. This ID is exchanged with the IP provider during an IP purchase transaction. The IP provider passes the unencrypted IP bitstream and the obtained ID of the end device to the vendor. The vendor has a database of IDs of end devices and their corresponding encryption keys. The vendor encrypts the bitstream using the stored an end device's stored encryption key before it is transmitted to the device. The solution does limit the issue of distribution of the encryption key between parties, however, promotes the use of static symmetric encryption keys.

DPR based approaches, such as [21] and [22], suggest dividing the Programmable Logic (PL) fabric into two zones, the static / security partition and the application partition. The security partition implements functions required for authentication of the device, e.g. a PUF. It also supports

cryptographic functions such as SHA or AES. Whereas, the application partition houses the application bitstream. The shortcoming is the area overhead for application logic, which adds a constraint over the application logic design. Additionally, since the security function lies with the application partition in the same fabric at the same time, thus it creates a possibility for secret leakage between the two partitions.

Trusted Platform Modules

Trusted Platform Modules (TPMs) are hardware-based cryptographic processors. These processors provide hardware assisted capabilities of secure encryption, signing and digital certificates processes and its storage on tamper-resistant Non-Volatile Memory (NVM). The TPM follows Trusted Computing Group's (TCG) TPM Profile Specification [23]. Trusted Platform Module provides constructs for key generation, True random number generators, Hash, and a set of registers known as Platform Configuration Registers (PCR). In TPM 2.0 there are two sets of PCRs, SHA 1 and SHA256. PCRs are special registers used to hold hash values for a block of input data and are commonly used in Secure Boot applications. The hash is computed using the TPM's TPM2_PCR_EXTEND function, for a block size of 256 bits. Each time the hash is computed, it is added to the previous value of a PCR value register.

$$PCR_n = PCR_n \parallel SHA256(x) \quad (1)$$

In equation (1), n is the PCR number, and x is the 256 bits input value. In the secure boot process, computed PCR values are compared with reference values at different times points for verification. Hash functions are one-way functions that are used for the integrity checking [24]. PCR registers commute the cumulative hash, with any bit flip in the configuration will result in an incorrect response. Other capabilities of TPM include support of symmetric and asymmetric cryptographic cores to enable confidentiality and availability.

There are two specification versions, namely 1.2 and 2.0. TPM 2.0 supports several cryptographic cores that are not available in the predecessor version 1.2. A complete list of supported cryptographic engines and the comparison between the two is given in Table 1.

Table 1. Comparison between TPM 1.2 and TPM 2.0.

Algorithm	RSA 1024	RSA 2048	ECC NISTP256	ECC BNP256	AES 128	AES 256	SHA-1	SHA- 256
TPM 1.2	Yes	Yes	No	No	Optional	Optional	Yes	No
TPM 2.0	Optional	Yes	Yes	Yes	Yes	Optional	Yes	Yes

3. Threat Model for Secure Boot of FPGA Bitstreams

The FPGA market is dominated by proprietary tools, Intellectual Properties (IPs), and closed hardware implementation. There are a handful of vendors that provide varying architectures and interfaces to those architectures. The reconfigurable fabric on FPGAs is programmed using bitstreams. The bitstream configures the Look Up Tables (LUTs) in the logic fabric. These LUTs act as combinatorial logic and sequential data paths for the hardware design. Bitstreams also configure other fabric elements, e.g. on-chip memory, Digital System Processing (DSP), clocking blocks and wire connections. An attack on the bitstream can affect the entire system operation of a device on the field. This work focusses its efforts on providing bitstream security on the device and on providing security between a content provider and a device.

Bitstream Spoofing

Bitstream spoofing updates the victim device with an update that seem to come from an authorized source. Bitstream spoofing may compromise the security of the victim device using relay and replay attacks [25][26]. An adversary acts as a man in the middle between a bitstream content provider and a device. Once an authenticated session is set up between the two nodes, the adversary replaces the original bitstream with a malicious one. In the case where the victim device is using one single key for bitstream encryption, replay attacks can be used by an adversary. An attacker can forward an older copy of the bitstream which has limited functionality compared to the current version.

Runtime Malicious Modification

Once a bitstream has been programmed onto an FPGA programmable logic fabric, an FPGA device may provide interfaces to the outside world for readback and modification of the running bitstream [27]. Using these interfaces, faults or trojans can be introduced in the design [28]. Additionally, the same interfaces can be used to make unauthorized modifications to the original design. Our work focusses on mitigating malicious logic insertion in the bitstream during runtime.

Nonsecure Communication with Content Provider

For an FPGA device placed in the field, bitstream updates can be provided manually physically by an engineer, through a physical update mechanism or using remote updates over a network. If a content provider over a network is not secure, an adversary may spoof its identity to become a content provider. Therefore, an adversary may be able to push malicious updates to the client. On the other hand, an adversary can also impersonate a device on the field to download bitstream updates from a content provider not meant for it.

4. The Root of Trust Architecture

In the proposed scheme there are two actors, the content provider or the server and client nodes. The content provider is a server that exists over a network for all client nodes. It serves the latest bitstream to all clients. A client node is any authorized FPGA device that is connected to the network of nodes. The authentication and authorization privileges for a client to join the network is open to the implementation and is left at the discretion of the designer. There are two components in this framework, namely the proposed hardware and the Secure First Stage Boot Loader (SFSBL).

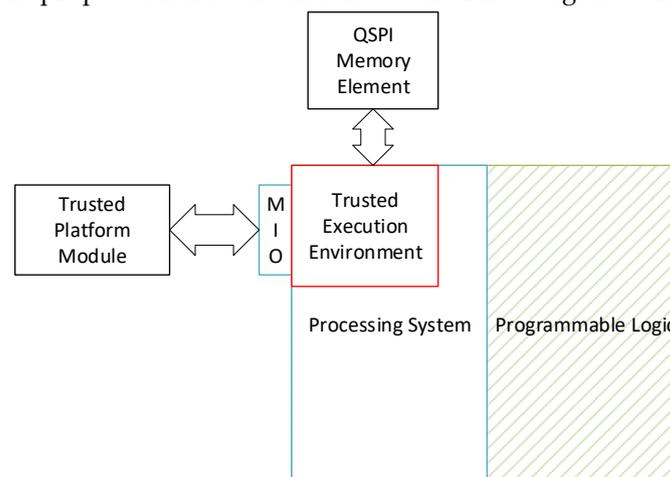


Figure 1. Proposed System Architecture

4.1. Hardware Overview

The reference reconfigurable platform is an SRAM based FPGA SoC that has a hard-core Processing System (PS) and a Programmable Logic (PL) fabric. The PS is a trusted verifier and the PL design is the prover. The verifier is a trusted source, whereas the prover needs to prove its legitimacy

to the verifier. We investigate the secure co-processor integration with the SoC and export the security functions and key provisioning onto an external cryptographic processor, such as the Trusted Platform Module (TPM). To provide isolation for a higher level of security functions that is for Secure OTA update functions, the framework uses a Trusted Execution Environment (TEE), such as ARM's TrustZone. TPM is interfaced using the Multiplexed Input /Output (MIO) port of the FPGA via serial SPI interface on Xilinx Zynq 7000 series FPGA. An MIO port is directly connected to the hard core of the PS and does not require any additional routing from the PL. The MIO based connection to the TPM is essential since in the case PL based configuration is used, the PL will be needed to be programmed first, thus introducing a possible attack vector for the target platform. To mitigate any damage caused due to privilege escalation attacks, PL programming port and the TPM are configured to be accessible only through TrustZone's secure world. The prototype is tested on the Xilinx platform that incorporates external Quad Serial Peripheral Interface (QSPI) based flash memory to hold the SFSBL and the bitstream. The write access to the flash memory is limited to the flash memory is also limited to TrustZone configured secure world. The hardware is described in Figure 1.

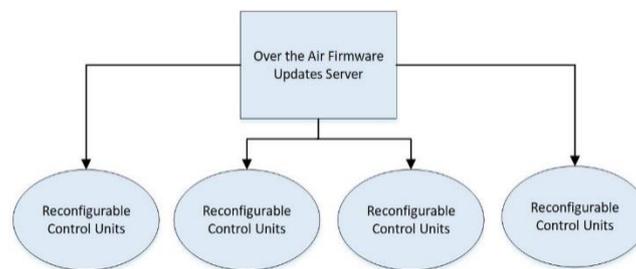


Figure 2. Content Provider connected with clients in an untrusted field.

4.2. Establishing Source of Trust

A content provider/server may serve multiple client nodes, as illustrated in Figure 2. The server is assumed to be secure. To establish the identity of the content provider and a client node, asymmetric digital keys are used. The proposed framework uses Elliptic Curve Digital Signature Algorithm based NISTP256 curve keys [29] for data signing. A pair of asymmetric signing keys are generated on the server as well as each client node. The client nodes use the equipped TPM to generate the key pairs. Additionally, a unique shared symmetric key is generated for each client. This key acts as a base encryption key (K_b) and is updated during the bitstream update process. The update process of the symmetric encryption key is discussed in detail in the following subsections. A client's own private key (K_{cpr}), public key (K_{cpb}), the server's public key (K_{ps}) and the base encryption key are stored on a client's TPM, in its NVM. The exchange of keys occurs in a trusted environment, where the server's private key (K_{rs}) does not leave the system. The server maintains a database of all its client's public keys and the base encryption keys. This system supports key renewal; however, the lifetime of the keys is left at the discretion of the system designer. To mitigate chances for key theft, it is assumed that the renewal process occurs in a trusted environment. Once deployed, the public key of the server and private key of the client cannot be retrieved from the TPM, however, they can be addressed for use within the TPM. The key exchange process is illustrated in Figure 3.

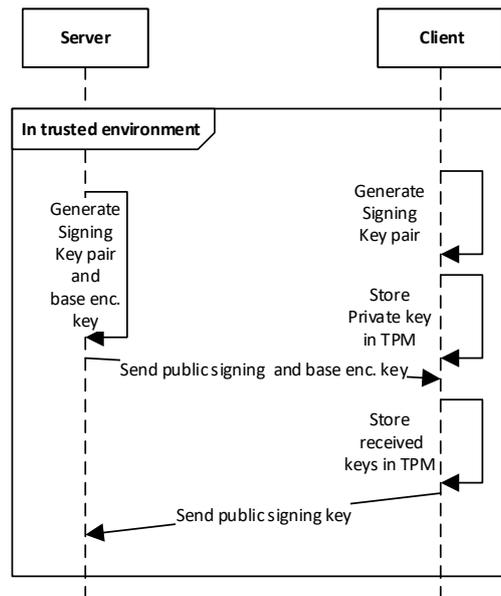


Figure 3. Key exchange in a trusted environment.

Bitstreams are stored in the QSPI Flash on the hardware. The initial bitstream is stored by the vendor in the trusted field. The proposed framework uses Platform Configuration Registers (PCR) to measure the boot process and to maintain prolonged integrity of the bitstream. PCR register, PCR_t is used to store the cumulative SHA256 of all the bitstreams loaded onto the client. Initially, the value of PCR_t is set to the SHA256 of the initial bitstream. The server also calculates the same cumulative hash locally (SHA_t). For use during secure boot at the client, the SHA256 of the bitstream (SHA_c) is stored on the NVM in the TPM. Figure 4 shows the keys on a client node and shared between a server.

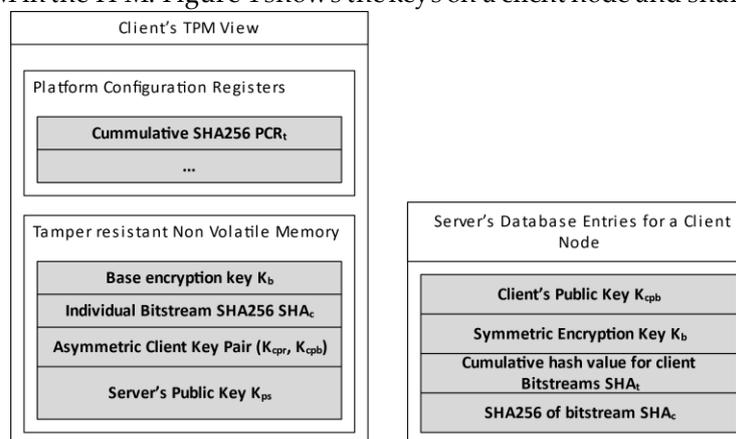


Figure 4. Keys shared between a client and server.

4.3. Secure Over the Air (OTA) Update Mechanism

When the updated bitstream available on the server, the connected and authorized client nodes receive a notification from the server. This triggers a client to initiate secure communication with the server. At a client, the mode of operation switches from normal mode to the update mode. The update mode is handled by the trusted execution environment at the client. In the case of TrustZone, the processor mode switches from the non-secure world to the secure world. The update process is assigned its own memory area, isolated from the non-secure world. The update procedure consists of the following processes:

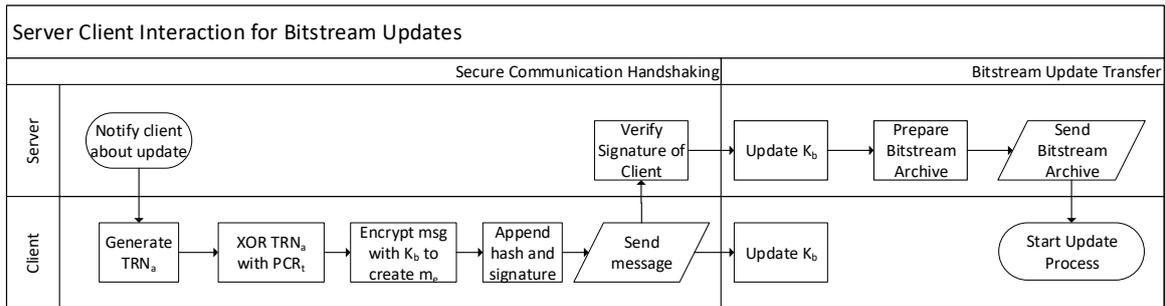


Figure 5. Server Client Interaction for Bitstream Updates.

4.3.1. Secure Communication Handshaking

To enable secure communication between the server and the client, and integrity of the bitstreams is maintained between two subsequent updates, we propose a handshaking scheme. The scheme is summarized in Figure 5. A client generates a True Random Number (TRN_a) using the True Random Number Generator equipped on the TPM. The PCR register value PCR_t is XORed with TRN_a and is then symmetrically encrypted using the base encryption key K_b .

$$m_e = AES128((PCR_t \wedge TRN_a), K_b) \quad (2)$$

The construction of the encrypted message (m_e) is shown in Equation 2. Hash of m_e is asymmetrically signed using K_{cp} . Message m_e and its hash is sent to the server. After transmission, the K_b is updated client side by taking a XOR with TRN_a . K_b on the TPM is also replaced with the newly computed value. Equation 3 illustrates the update process.

$$K_b = K_b \wedge TRN_a \quad (3)$$

The server using its copy of K_{cpb} recalculates the hash of the received message m_e . Once the client has been authenticated, the server decrypts m_e using the stored key K_b . To retrieve TRN_a , the server computes the XOR of SHA_t with the decrypted message. Like the client node, K_b is updated at the server.

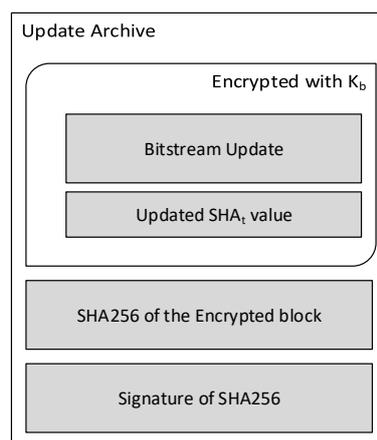


Figure 6. Bitstream Update Archive.

4.3.2. Secure Update Packaging and Transfer

The bitstream update is compiled into an archive package before it is sent to a client. Using the bitstream update and SHA_t , the server calculates the new cumulative hash and updates SHA_t . The updated bitstream and SHA_t are combined and encrypted using the updated key K_b . This encrypted block is copied to the update archive package along with its SHA256 value. The SHA256 value is signed using its own private key K_{rs} . The signature is appended to the package. The update archive

is sent to the client. The client stores the package in its NVM and triggers the update process. Figure 6 shows the archive package.

4.3.4. Applying the Update

The process of applying the update on the target platform is handled by the secure update process running on the client FPGA. Once the bitstream package has been downloaded, SHA 256 of the package is recalculated client side and is compared with that in the download package. The signature of the computed hash is recomputed using the key K_{ps} on the TPM. This signature is compared against the signature packaged in the update archive. Once the identity of the server has been verified, the encrypted archive is decrypted using K_b . Using the TPM PCR functions, PCR_t is updated and verified against the archived SHA_t value. If the updated PCR_t and SHA_t values are equal, the initial bitstream on the NVM is replaced by the update, otherwise if in case any check fails during the update process, the node is assumed to be compromised and the server is notified. On completion of the update process, the individual SHA256 of the bitstream calculated and SHA_c is updated. The process is described in the flow chart in Figure 7.

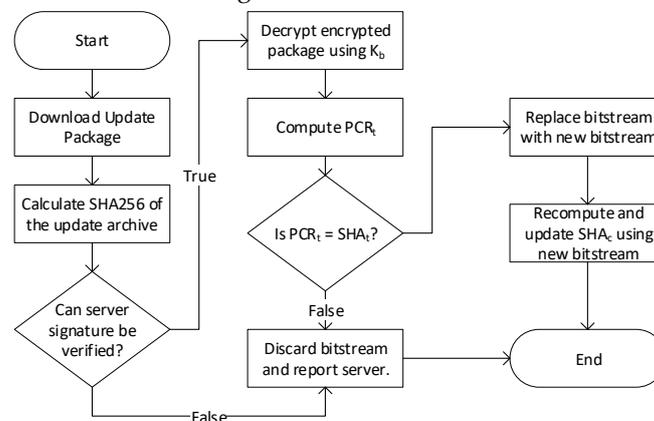


Figure 7. Secure Bitstream Update Process.

4.3.5. Secure FSBL Boot Process

We have designed a Secure First Stage Boot Loader (SFSBL) that is integrated with the Trusted Platform Module with custom device drivers that enable the device to communicate with the TPM before the device has booted. In an FPGA, the zeroth stage boot loader referred to as BootROM is executed by the Processing System on FPGA at the beginning of the boot cycle. BootROM code exists in a non-volatile memory of the FPGA fabric and is assumed to be secure. Once the BootROM has completed its execution, the control is passed to the proposed SFSBL that exists in the on-board QSPI memory. The SFSBL code extends and integrates the security functions to the FSBL. FSBL configures the device with the hardware bitstream and configures the processing system with a bare-metal application or loads the second stage boot loader to the RAM. This, in turn, loads the operating system. The vendor FSBL design does not verify the integrity of the bitstream or what it is programming the processing system. This lack of security at the boot process can be circumvented with the proposed secure FSBL that integrates key management and security features to validate the configuration files before programming the device.

The proposed SFSBL begins execution with initializing on-board components and memory elements. Once the basic devices have been initialized, the TrustZone is set up. The access to the MIO port connected to the TPM as well as the QSPI ports is both configured to be accessible only by TrustZone's secure world. The secure world code is copied onto the RAM. This segment of memory is only addressable through secure applications. Once the setup is complete, the control is switched to the secure world. The secure world initiates the process of secure FPGA boot. To mitigate exploitation of Time of Check Time of Use (ToCtoU)[30] by an attacker, the processing system recomputes the cumulative SHA256 of the bitstream on the TPM. This value is compared with SHA_c .

stored on the TPM during the bitstream update process. If the values are different, the boot process stalls and the violation is recorded for later reference. Otherwise, the bitstream is configured onto the FPGA programmable logic fabric and the boot process continues.

The secure world sets up service hooks for the non-secure world. These hooks can be used to provide security functions for applications. One such service is the bitstream update service, in which the control is switched to the secure world. The SFSBL process continues with loading the non-secure world application, which can be either a bare-metal application or a second stage boot loader, e.g. U-Boot[31]. The SFSBL process finishes by switching back to the non-secure world and passing the control of execution to the non-secure world.

5. Results and Analysis

The proposed framework has been implemented on a Xilinx Zedboard FPGA board equipped with a Zynq-7000 XC7Z020-CLG484 [32]. The FPGA has an embedded ARM Cortex A9 hard processor. The FPGA is integrated with the Infineon TPM SLB9670, a secure coprocessor for the key management and secure boot processes [33]. The processor is equipped with ARM's TrustZone for Trusted Execution Environment (TEE). Additionally, the on-board QSPI memory is used for holding the SFSBL implementation and the bitstream package extracted from the bitstream update process. Experimental setup of the proposed framework is shown in Figure 8.

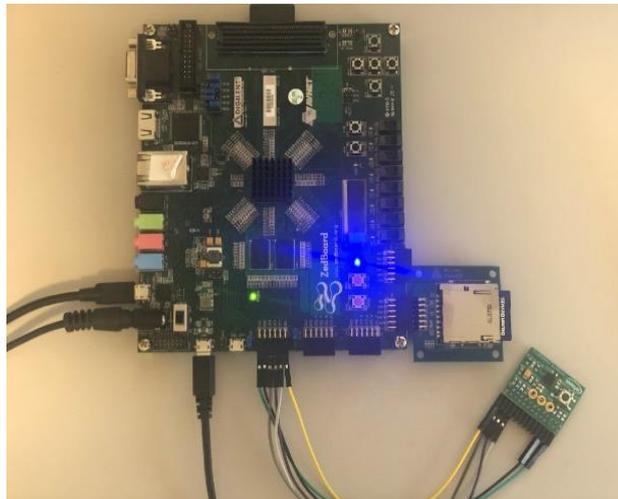


Figure 8. Experimental setup featuring TPM connected with an FPGA.

The FPGA board communicates with the TPM via the Serial Peripheral Interface (SPI) over the dedicated MIO port. The proposed SFSBL implementation is the extension of the Xilinx provided First Stage Boot Loader (FSBL). The secure extensions are added to the existing FSBL that uses the custom device driver library written as part of this research.

The MIO ports are accessible through the secure world configured using ARM processor TrustZone. [34]. The total RAM on the Xilinx Zedboard is 512 MB. We have configured upper 64MB of the RAM space to be used by the secure world. This setting is controlled using the TZ_DDR_RAM register. Additionally, to limit access to the QSPI memory, the QSPI_S_APB bit in the SECURITY6_APB_SLAVES register is set to 0. Typically, for the configuration of peripherals in TrustZone, the value '0' signifies that a peripheral is set to be secure or only accessible from the secure world.

To incorporate the TPM with the FPGA board at the FSBL level, we have implemented a device driver library. This library provides all necessary functions to set up the TPM and implements security functions on the TPM. To the best knowledge of the authors, this is the first library of its kind. The SPI interface accessible through the secure world implements the TPM transfer function driver to communicate with the TPM device. This communication is required at the FSBL level to perform secure boot, which is not supported until the second stage boot loader in the traditional

design flow. Figure 9 shows the transfer function that is part of the TPM driver library to establish the SPI interface.

```
int tpm_xfer(uint32_t addr, const uint8_t * out, uint8_t * in, uint16_t len){
    u8 tx_buf[MAX_BUFFER_SIZE];
    u8 rx_buf[MAX_BUFFER_SIZE];
    int transfer_len, ret;

    if (in && out){
        fsbl_printf(DEBUG_GENERAL, "tpm_xfer: in and out cannot both be not NULL.\r\n");
        return XST_FAILURE;
    }

    while(len){
        transfer_len = (len<MAX_BUFFER_SIZE) ? len: MAX_BUFFER_SIZE;
        tx_buf[0] = (in ? 0x80 : 0) | (transfer_len -1);
        tx_buf[1] = 0xD4;
        tx_buf[2] = (addr) >>8;
        tx_buf[3] = addr;

        if (out){
            memcpy(&tx_buf[4], out, transfer_len);
            out+=transfer_len;
        }

        ret = XspiPs_PolledTransfer(&SpiInstance, tx_buf, rx_buf,4 + transfer_len );
    }
}
```

Figure 9. Screenshot the tpm_xfer function.

The device driver library is open source and is made available online for public use[35]. The services provided by the library can be divided into two categories, device power-up services, and cryptographic functions. TPM 2.0 architecture has five layers. These layers signify the boot stage for a target platform and are termed as localities. Each locality offers specific functionalities and implements privileges with restrictions of allowed functions, for example, the PCR registers are resources limited to specific localities. Our library implementation provides access to different secure boot specific functions at all localities. In the proposed framework implementation, the TPM is only accessible from the trusted secure world, the library exists in the scope of the secure world and is not accessible from the non-secure world.

Timing overhead for the proposed solution is dependent on the data rate of the SPI interface and the wait time for each operation. The data rate of the SPI interface is dependent on the host and the TPM device. The TPM2 specifications only specify a maximum timeout for a message transfer and timeout for primitive operations such as requesting a locality and checking the ownership of a locality, etc.

```
int tpm_pcr_extend(int loc, u32 index, const uint8_t *digest){
    u8 * pcr_cmd_buf;
    unsigned int offset = 33;
    int ret=0;
    u8 response[500];

    pcr_cmd_buf = malloc(sizeof(tpm2_pcr_extend) + TPM2_DIGEST_LEN);
    memcpy(pcr_cmd_buf, tpm2_pcr_extend, sizeof(tpm2_pcr_extend));

    put_unaligned_be32(33+TPM2_DIGEST_LEN, pcr_cmd_buf +2); //copy digest length
    put_unaligned_be32(index, pcr_cmd_buf +10); //copy index
    memcpy(pcr_cmd_buf + 31, sha256_alg, sizeof(sha256_alg)); //copy sha type
    //copy digest
    if (digest ==NULL)
        return XST_FAILURE;

    memcpy(pcr_cmd_buf + sizeof(tpm2_pcr_extend), digest, 32);
    fsbl_printf(DEBUG_GENERAL, "tpm_pcr_extend: INFO Command Bytes: ");
    for(int i=0; i<sizeof(tpm2_pcr_extend) + TPM2_DIGEST_LEN; i++){
        fsbl_printf(DEBUG_GENERAL, "%02x.", pcr_cmd_buf[i]);
    }
    fsbl_printf(DEBUG_GENERAL, "\r\n");

    ret = tpm_send(loc, pcr_cmd_buf, sizeof(tpm2_pcr_extend) + TPM2_DIGEST_LEN);
    fsbl_printf(DEBUG_GENERAL, "tpm_pcr_extend(): INFO tpm_send() return: %d\r\n", ret);

    memset(response, 0xFF, sizeof(response));
    ret = tpm_recv(loc, response, 500);
    fsbl_printf(DEBUG_GENERAL, "tpm_pcr_extend(): INFO return size: %d\r\n", ret);
    if (ret>0){
        fsbl_printf(DEBUG_GENERAL, "tpm_pcr_extend(): INFO PCR_READ(REG:%d): ", index);
        for (int i=0; i<ret; i++){
            fsbl_printf(DEBUG_GENERAL, "%02x.", response[i]);
        }
        fsbl_printf(DEBUG_GENERAL, "\r\n");
    }
    free(pcr_cmd_buf);
}
```

Figure 10. Snapshot of TPM's Driver Extend Function.

The secure boot process, loads the bitstream, as shown in the Figure 12. The bitstream is sent to the TPM for the integrity checking, where the cumulative hash is computed using the PCR registers with the scheme discussed in the section 3. The TPM driver loads the PCR register with the cumulative hash for the 3.85 MB bitstream with the 256 bit segments along with the feedback from the PCR cumulative hash. The novel bare-metal TPM driver library uses locality 4 interface to access the PCR registers and implements TPM_PCR_READ structure with the function `tpm_pcr_read`. The hash value computed by the `tpm_pcr_read` function call is compared with the golden reference value stored on the tamper resistant storage. In the following example, the reference hash is not equal to the computed hash, therefore the boot process halts, and results into diagnostics using the fallback process. In the fallback process the device is only capable of bare metal functionality with limited networking capabilities to mitigate the impact of the compromised device over the network.

6. Security Analysis

Reconfigurable computing is becoming ubiquitous combined in the IoT devices that embedded devices that are embedded devices with long life cycles. The primary goal of SFSBL framework is to enable security features to establish secure boot and TPM based over the air secure hardware configuration updates on the field to enable these devices to evolve and update with the change in device requirements, such as security patches. In this section, we discuss the security properties of the proposed framework.

- Digital certificates are used for the identification of the server and participating nodes. The certificates are shared before deployment by the trusted authority and can only be modified by the trusted facility. Thus, impersonation and data spoofing are not possible in the proposed scheme.
- The encrypted configuration bitstream is packaged with the SHA256 of the encrypted bitstream, along with the bitstream package signature signed off with the private key of the server. Any changes in the bitstream will fail the hash comparison and device will discard the updates. This prohibits the man in the middle, and data spoofing attacks and provides an authentication mechanism for securely transfer the configuration files.
- The keys are stored on the tamper-resistant storage inside TPM on server and device nodes, eliminating invasive and probe attacks to break the key. Furthermore, the private keys never leave the premises thus mitigates the man in the middle attack.
- The symmetric encryption key is updated on every bitstream. It ensures freshness of encryption keys as well as mitigates replay attacks. Once the key is used it is never used again and hence any repeated keys can be used to identify attacks.
- During the execution, the device cannot be altered, and the keys can be accessed with the hardware isolation protection of the SPI interface. The TPM interface is only accessed in the secure mode.

7. Conclusion

In this paper the security extensions of reconfigurable logic based embedded device are proposed to enable secure boot processes and the firmware updates to reconfigure the hardware and software to run on the device in an untrusted field. We have integrated the TPM capabilities at the boot level with the custom drivers that are integrated at the first stage boot loader to verify the bitstream before the programmable logic is configured or the software is loaded to identify any malicious modifications in the device configuration files. The framework integrates the existing security features such as the trusted execution environment and enables the trusted platform module capabilities with the custom drivers that interface with the FSBL to implement secure boot process. We demonstrate the secure boot and authentication processes for the over the air updates to the firmware in the field.

Funding: This research was funded by NSF, grant number 1814420, 1819694 and 1819687.

References

- [1] S. Schmidt, M. Tausig, M. Hudler, and G. Simhandl, "Secure Firmware Update Over the Air in the Internet of Things Focusing on Flexibility and Feasibility," in *Internet of Things Software Update Workshop (IoTSU), At Dublin*, 2016, no. June.
- [2] H. Chandra, E. Anggadajaja, P. S. Wijaya, and E. Gunawan, "Internet of Things: Over-the-Air (OTA) firmware update in Lightweight mesh network protocol for smart urban development," in *Proceedings - Asia-Pacific Conference on Communications, APCC 2016*, 2016, pp. 115–118.
- [3] C. E. Andrade *et al.*, "Managing massive firmware-over-the-air updates for connected cars in cellular networks," in *CarSys 2017 - Proceedings of the 2nd ACM International Workshop on Smart, Autonomous, and Connected Vehicular Systems and Services, co-located with MobiCom 2017*, 2017, pp. 65–72.
- [4] V. Zimmer and M. Krau, "Establishing the root of trust," 2016.
- [5] R. Wilkins and B. Richardson, "Uefi Secure Boot in Modern Computer Security Solutions," *UEFI Forum*, 2013.
- [6] "Secure the Windows 10 boot process | Microsoft Docs." [Online]. Available: <https://docs.microsoft.com/en-us/windows/security/information-protection/secure-the-windows-10-boot-process>. [Accessed: 04-Oct-2018].
- [7] Xilinx Inc., "Zynq-7000 All Programmable SoC Secure Boot," 2014. [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug1025-zynq-secure-boot-gsg.pdf. [Accessed: 11-Apr-2018].
- [8] S. M. Trimberger and J. J. Moore, "FPGA security: Motivations, features, and applications," *Proc. IEEE*, vol. 102, no. 8, pp. 1248–1265, Aug. 2014.
- [9] S. Tajik, H. Lohrke, J. P. Seifert, and C. Boit, "On the power of optical contactless Probing: Attacking bitstream encryption of FPGAs," in *Proceedings of the ACM Conference on Computer and Communications Security*, 2017, pp. 1661–1674.
- [10] T. R. Manual, "Zynq-7000 All Programmable SoC," *Xilinx*, vol. 585, pp. 1–1863, 2015.
- [11] Xilinx Inc., "Zynq UltraScale+ Device Technical Reference Manual UG1085," 2018. [Online]. Available: https://www.xilinx.com/content/dam/xilinx/support/documentation/user_guides/ug1085-zynq-ultrascale-trm.pdf.
- [12] Xilinx Inc., "Developing Tamper-Resistant Designs with Zynq UltraScale+ Devices," 2018. [Online]. Available: https://www.xilinx.com/support/documentation/application_notes/xapp1323-zynq-usp-tamper-resistant-designs.pdf. [Accessed: 11-Apr-2018].
- [13] N. Jacob, J. Heyszl, A. Zankl, C. Rolfes, and G. Sigl, "How to break secure boot on FPGA SoCs through malicious hardware," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10529 LNCS, pp. 425–442.
- [14] Xilinx Inc., "UG 470 - 7 Series FPGAs Configuration." [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug470_7Series_Config.pdf.
- [15] M. Ender, A. Moradi, and C. Paar, "The Unpatchable Silicon: A Full Break of the Bitstream Encryption of Xilinx 7-Series FPGAs," 2020.
- [16] D. Owen Jr. *et al.*, "An Autonomous, Self-Authenticating, and Self-Contained Secure Boot Process for Field-Programmable Gate Arrays," *Cryptography*, vol. 2, no. 3, p. 15, Jul. 2018.
- [17] W. Che, F. Saqib, and J. Plusquellic, "PUF-based authentication," in *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2015, pp. 337–344.

- [18] I. Lebedev, K. Hogan, and S. Devadas, "Invited paper: Secure boot and remote attestation in the sanctum processor," in *Proceedings - IEEE Computer Security Foundations Symposium*, 2018, vol. 2018-July, pp. 46–60.
- [19] Xilinx and Inc, "Using Encryption and Authentication to Secure an UltraScale/UltraScale+ FPGA Bitstream Application Note (XAPP1267)," *XAPP1267*, 2017. [Online]. Available: https://www.xilinx.com/support/documentation/application_notes/xapp1267-encryp-efuse-program.pdf. [Accessed: 01-Feb-2018].
- [20] A. Carelli, C. A. Cristofanini, A. Vallerio, C. Basile, P. Prinetto, and S. Di Carlo, "Securing bitstream integrity, confidentiality and authenticity in reconfigurable mobile heterogeneous systems," in *2018 IEEE International Conference on Automation, Quality and Testing, Robotics, AQTR 2018 - THETA 21st Edition, Proceedings*, 2018, pp. 1–6.
- [21] J. Vliegen, M. M. Rabhani, M. Conti, and N. Mentens, "SACHa: Self-Attestation of Configurable Hardware," in *Proceedings of the 2019 Design, Automation and Test in Europe Conference and Exhibition, DATE 2019*, 2019, pp. 746–751.
- [22] G. Pocklassery, W. Che, F. Saqib, M. Areno, and J. Plusquellic, "Self-authenticating secure boot for FPGAs," in *2018 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2018, pp. 221–226.
- [23] Trusted Computing Group, "TCG PC Client Platform TPM Profile (PTP) Specification Family '2.0' TCG Public Review," 2017.
- [24] W. Arthur, D. Challener, and K. Goldman, *A Practical Guide to TPM 2.0*. Apress, 2015.
- [25] S. Drimer, "Volatile FPGA design security – a survey," *Univ. Cambridge*, pp. 1–51, 2008.
- [26] J. Zhang, Y. Lin, and G. Qu, "Reconfigurable Binding against FPGA Replay Attacks," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 20, no. 2, pp. 1–20, Mar. 2015.
- [27] Xilinx, "AXI Hardware ICAP." [Online]. Available: https://www.xilinx.com/products/intellectual-property/axi_hwicap.html. [Accessed: 11-May-2019].
- [28] R. S. Chakraborty, I. Saha, A. Palchoudhuri, and G. K. Naik, "Hardware trojan insertion by direct modification of FPGA configuration bitstream," *IEEE Des. Test*, vol. 30, no. 2, pp. 45–54, Apr. 2013.
- [29] NIST, "RECOMMENDED ELLIPTIC CURVES FOR FEDERAL GOVERNMENT USE," 1999.
- [30] A. Francillon, Q. Nguyen, K. B. Rasmussen, and G. Tsudik, "A minimalist approach to Remote Attestation," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2015.
- [31] "WebHome - U-Boot - DENX." [Online]. Available: <https://www.denx.de/wiki/U-Boot>. [Accessed: 29-Oct-2018].
- [32] AVNET, "ZedBoard | Zedboard." [Online]. Available: <http://zedboard.org/product/zedboard>. [Accessed: 19-Jan-2018].
- [33] "SLB 9670VQ2.0 - Infineon Technologies." [Online]. Available: <https://www.infineon.com/cms/en/product/security-smart-card-solutions/optiga-embedded-security-solutions/optiga-tpm/slb-9670vq2.0/>. [Accessed: 03-Nov-2018].
- [34] Xilinx and Inc, "Programming ARM TrustZone Architecture on the Xilinx Zynq7000 All Programmable SoC User Guide (UG1019)," 2014.
- [35] A. S. Siddiqui and F. Saqib, "HEADS-UNCC/TPM_baremetal_drivers: TPM Baremetal for FPGAs and other embedded systems." [Online]. Available: https://github.com/HEADS-UNCC/TPM_baremetal_drivers. [Accessed: 18-Jun-2020].