

TALUS: Reinforcing TEE Confidentiality with Cryptographic Coprocessors

Dhiman Chakraborty, Michael Schwarz, and Sven Bugiel

CISPA Helmholtz Center for Information Security, Saarbruecken, Germany
{dhiman.chakraborty,michael.schwarz,bugiel}@cispa.de

Abstract. Platforms are nowadays typically equipped with trusted execution environments (TEEs), such as Intel SGX or ARM TrustZone. However, recent microarchitectural attacks on TEEs repeatedly broke their confidentiality guarantees, including the leakage of long-term cryptographic secrets. These systems are typically also equipped with a cryptographic coprocessor, such as a TPM or Google Titan. These coprocessors offer a unique set of security features focused on safeguarding cryptographic secrets. Still, despite their simultaneous availability, the integration between these technologies is practically nonexistent, which prevents them from benefitting from each other’s strengths.

In this paper, we propose *TALUS*, a general design and a set of three main requirements for a secure symbiosis between TEEs and cryptographic coprocessors. We implement a proof-of-concept of *TALUS* based on Intel SGX and a hardware TPM. We show that with *TALUS*, the long-term secrets used in the SGX life cycle can be moved to the TPM. We demonstrate that our design is robust even in the presence of transient execution attacks, preventing an entire class of attacks due to the reduced attack surface on the shared hardware.

1 Introduction

The need for stronger protection of data and computations has led to the advent of *secure enclaves*, CPU-provided isolated *Trusted Execution Environments* (TEE) that secure general-purpose computations. Prevalent technologies are Intel SGX [27, 20, 16], ARM TrustZone [1], or Keystone [42] and MI6 [5] for RISC-V.

The security offered by these *secure enclaves* for code and data isolation depends on several high value cryptographic credentials (e.g., Launch and Provisioning Key for Intel SGX, AMD PSP infrastructure key for AMD SEV, manufacturer root keys for ARM TrustZone). Enclave programs, in turn, depend on credentials derived from those long-term secrets, e.g., for secure storage of enclave data. Unfortunately, enclave technology shares hardware, e.g., CPU cores, between trusted and untrusted code, opening an attack surface. Especially for Intel SGX, this attack surface has been exploited in microarchitectural attacks [49], some of which leak confidential data from CPU buffers [7, 63, 64, 72, 4].

Our key observation is that virtually all platforms today are additionally equipped with specialized cryptographic or security-oriented coprocessors that

protect cryptographic credentials, access control secure storage, or monotonically count. For instance, Trusted Platform Modules (TPM) [68] are available on effectively all desktop and server machines, and more solutions become available, such as Google’s Titan, Microsoft’s Cerberus, or AMD’s PSP [76]. In contrast to general purpose application processors with security extensions for TEEs, those coprocessors have been designed for the primary goal to safeguard cryptographic credentials and secret data. Integration between secure enclaves and cryptographic coprocessors creates a stronger security solution in which enclaves can use the complementary coprocessor features. Concrete use-cases would benefit from this integration, e.g., impeding microarchitectural attacks against enclaves based on TPM features. Unfortunately, such an integration is currently, if it exists, very limited. We ask the following fundamental questions: *Which security guarantees does the combination of CPU-provided TEEs with secure coprocessors provide that each of the technologies cannot provide on their own? What are the requirements to combine the advantages of both technologies without introducing new security problems or large performance overheads?*

To answer these questions, we introduce a hardware/software co-design, *TALUS*, to combine CPU-provided TEEs with cryptographic coprocessors. Enclave code can directly invoke the coprocessor *only* via the CPU firmware and bus connections to make use of the coprocessor’s facilities, such as counters or key management. We identify three core requirements to realize our idea: a *secure communication channel* between processors and coprocessors, *vertical access control* to distinguish between enclave and non-enclave code, and *horizontal access control* to distinguish between different enclaves. To understand how SGX can be integrated with an on-board hardware TPM, we built a proof-of-concept integration between Intel SGX and a hardware TPM on commodity hardware. We show that a combination of Intel SGX (emulated through KVM-SGX [34] and QEMU-SGX [35]) with hardware TPM is feasible with firmware changes and demonstrate through different use cases the security benefits of this symbiosis.

We show that TPM fills a gap in the trusted-computing features of SGX that is due to a lack of replay-protected secure non-volatile memory. Several previously published defenses for attacks against SGX provide their full strength only if such building blocks are available [65, 57, 13]. Furthermore, preventing recent microarchitectural attacks against TEEs [75, 7, 63, 64], including under-volting [61, 56, 38] is only effective if an enclave can store a persistent state to limit the number of attack attempts. In addition to the possibility of preventing attacks against enclaves, we demonstrate that all high-value secrets used during the lifetime of enclaves can be safely stored in the TPM without ever reaching a shared hardware element. We can actively mitigate existing attacks and harden an enclave against potential future attacks by reducing the amount of high-value secrets stored in the enclave. Our proof-of-concept implementation shows that the expected overhead of an average 21.6% is amortized in typical use cases, as only rarely used operations suffer from a slowdown of several milliseconds.

In summary, we make the following contributions:

1. We introduce *TALUS*, a hardware/software co-design to combine CPU-provided TEEs with cryptographic coprocessors.
2. We show that *TALUS* provides extended features, like rollback protected TPM NV-storage for persistent counters to limit execution control attacks against enclaves.
3. We demonstrate that *TALUS* significantly reduces the attack surface for microarchitectural attacks.
4. We analyze *TALUS* for real use cases, showing that its performance overhead is amortized in many use cases while providing strong security guarantees.

2 Background

2.1 Intel Software Guard eXtension (SGX)

SGX is an extension to the x86 instruction set that allows a user-space process to create and manage a protected isolated memory region called an *enclave* within its own address space, even protected from OS and hypervisor access [48, 26]. SGX assumes that the CPU, including its microcode, is the only trusted element in the system. Enclave data are stored encrypted in DRAM and unencrypted in the CPU caches and registers. An external party can verify an enclave by (remote) attestation of the enclave code and meta-data [66, 6].

Intel supplies two infrastructure enclaves, the *launch enclave* (LE) and *quoting enclave* (QE), on which SGX is heavily dependent. The LE is responsible for handling and launching user-space enclaves with a token called `EINITTOKEN` that is generated using i) the measurement of the static content of the enclave (`MRENCLAVE`) and ii) the enclave-author validation (`MRSIGNER`). The LE requires a 128-bit *Launch Key* (LK) to derive the `EINITTOKEN`. The QE is designed to validate local attestation reports by enclaves generated with an asymmetric private key that a remote verifier can verify. Both the LE and the QE are entrusted with long-term high-value cryptographic credentials.

2.2 Trusted Platform Module (TPM)

TPM by the Trusted Computing Group is the most widely deployed trusted computing technology on commodity platforms used by, e.g., Microsoft Windows management instrumentation, Intel Trusted eXecution Technology (TXT) [31], Microsoft Bitlocker [51], or Google Chrome [19]. A TPM contains a small non-volatile memory block, a set of platform configuration registers (PCR), an on-board processor to execute TPM code in isolation from the other hardware, co-processing for standard cryptographic algorithms, a secure clock, and a random number generator. TPMs can reliably report internal data to a third-party verifier, i.e., remote attestation based on a pre-installed endorsement key. Typically, a TPM is available as a hardware chip soldered to the mainboard, traditionally connected via the Low Pin Count (LPC) bus or on newer platforms via the SPI bus, making it only available through memory-mapped I/O (MMIO) registers protected by the chipset. Intel also implements a firmware TPM called Platform Trust Technology (Intel-PTT) [25] housed inside Intel CSME [28].

3 Requirements Analysis

In this section, we define three fundamental requirements for a secure integration of CPU-provided Trusted Execution Environments with onboard secure coprocessors: *secure communication channel*, *horizontal access control*, and *vertical access control*. We systematically compare how SGX and TPM meet those requirements and how well these two technologies can be integrated, as demonstrated later by our proof-of-concept implementation. In the full technical report [11], we have extended this comparison further to other secure coprocessors and TEEs.

Communication Channel (CC). For a secure integration between the security coprocessor and the application processor (AP), the communication channel between them must be secured from eavesdropping even in case of physical attacks, e.g., bus sniffing (**CC1**), and there should not be any dependencies on buffers vulnerable to microarchitectural attacks that can leak sensitive data transferred via the channel (**CC2**). TPM and SGX fulfill **CC1** since TPM and Intel CPUs support end-to-end encryption of the communication between them [68, 17, 31]. However, this channel does not avoid insecure buffers, and decrypted data on the CPU side might still pass through such buffers. As demonstrated by recent attacks, none of the TEEs, including SGX, is free of insecure buffers. Therefore, SGX inherently fulfills **not CC2**, and we show in Section 5 how we overcome this limitation in combination with TPM.

Horizontal Access Control (HC). TEEs can host multiple tenants. For example, SGX supports multiple (parallel) enclaves. Horizontal access control ensures that the AP and the coprocessor can distinguish between requests from mutually untrusted tenants inside a TEE. For instance, one enclave should not be able to access another enclave’s data within the coprocessor. A trusted entity, such as an AP, must create access or identity tokens that can identify TEE tenants. The tokens must be securely communicated to the coprocessor. The coprocessor must also understand those tokens to control access to managed data and secrets. Hardware TPM and firmware TPM employ extended authorization policies (EAP) that can use these access tokens for access control to TPM-managed objects, like TPM-internal storage and keys. All AP-based TEEs can fulfill this requirement because they can uniquely identify the different enclave codes they host. They can provide this information on calls to the coprocessor. For example, in SGX, this would be the code measurement of the enclave by the CPU.

Vertical Access Control (VC). AP-based TEE technologies and the coprocessor should support access control based on different security levels (e.g., application, OS, or hardware) to prevent non-enclave code from accessing enclave-owned entities in the coprocessor. The access token to distinguish between different security levels needs to be generated and handled by a secure piece of code and be securely communicated to the coprocessor. Hardware and firmware TPM offer *Locality* to distinguish between TPM commands originating from different

security levels. Still, the locality of a command must be communicated to the TPM by the CPU or firmware. Furthermore, SGX registers when it executes in enclave mode, but this security level is only used CPU-internally and not for Locality.

TALUS: Integrating Intel SGX and Hardware TPM. The main issue of vanilla SGX is the lack of confidentiality- and integrity-protected tamper-resistant storage. As we are unaware of any non-volatile memory inside a CPU, we do not see how SGX can be improved by only updating the firmware and without adding new components (like a TPM) to the TCB. Vanilla SGX can use PTT for certain trusted computing use cases. However, PTT is housed inside the CSME [28] and connected through the DMI interface without any security around the communication channel. Moreover, although CSME employs its own OS with its own security ring, completely segregated from the platform security, the command buffer for PTT is configured by untrusted software, such as the OS, and PTT recently suffered from access control errors [29, 30] that completely undermine its security and are currently unfixable in production devices. Additionally, secrets typically flow through the memory hierarchy on the CPU where untrusted code can run in parallel, observing side effects of the secret processing, e.g., when unsealing data from disk. Furthermore, in SGX, support for counters depends on the Platform Service Enclave and Intel ME, which are often not available in SGX production deployments and have already been deprecated [32]. Moreover, these counters can be reset by reinstalling the SGX platform software [46]. As SGX stores counters inside the BIOS flash storage, they do not persist across system resets [46]. The unavailability of integrity-protected, tamper-resistant storage does not allow SGX to store a secure counter, which limits the possibility of enclaves to enforce a number of enclave executions, as exploited in interrupt-based attacks [7].

Based on our requirements analysis, we found that the combination of SGX with hardware TPM is highly amenable for integration and allows to fill those gaps in SGX with TPM functionality. Due to the historical relationships between Intel CPUs and TPM, they can create an encrypted channel between them. Additionally, SGX can identify (i.e., measure) enclave code while TPM can use this identity in its access control policies. Therefore, our proof-of-concept implementation for our TALUS design is based on SGX and a **hardware** TPM.

4 High-level TALUS Overview

Our systematization (Section 3) underlines the intuition that the TPM, when integrated as a coprocessor with SGX, can provide desirable features to secure enclaves, such as physically isolated processing of cryptographic secrets, a secure clock, or persistent counters. The basic idea is to retrofit SGX with a *direct* communication channel to the TPM chip *without* going through the host OS. With such a communication channel, enclaves can leverage the TPM facilities as building blocks, e.g., to implement secure monotonic counters (cf. Section 5).

This section provides more details on the security benefits, requirements, and challenges of integrating SGX enclaves with a TPM. The high-level overview of *TALUS* is available in the extended version of the paper [11].

4.1 Threat Model

The threat model for *TALUS* is the union of the coprocessor and enclave threat model. Only the coprocessor (including firmware) and the processor (including microcode) are trusted. We assume that the coprocessor does not suffer from implementation [54] or platform integration flaws [22]. Similarly, we assume that the enclaves are not malicious [50] and are free of classical software vulnerabilities [74, 43, 71, 14]. Microarchitectural attacks [69], such as classical side channels and transient execution attacks, are in scope. We allow physical attacks in line with the TPM and SGX specifications, e.g., bus tapping, bus sniffing, or similar physical layer attacks [3, 37, 40]. We exclude physical attacks outside of a reasonable attacker model for SGX and consumer-grade hardware, such as bus snooping on high-speed or address buses [41], against which SGX also fails to defend.

4.2 Design of TALUS

Integrating a coprocessor (e.g., TPM) with a secure enclave technology, such as SGX, poses both security (SC) and functional (FC) challenges. In this section, we detail the challenges and how we design *TALUS* to solve these challenges.

SC1. Secure Communication Channel. CPU and coprocessor must exchange data securely. Ideally, the coprocessor is physically integrated with the CPU package (e.g., similar to AMD PSP), and the communication channel is physically secured against eavesdropping. If the coprocessor is an additional hardware element, a secure connection via the usually insecure bus is required. For TPM and SGX, TPM is connected to the CPU via the unprotected LPC or SPI bus. Thus, *TALUS* relies on symmetric authenticated cryptography to establish a secure channel between the coprocessor and the CPU while ensuring confidentiality and integrity despite an untrusted OS and a physical attacker.

SC2. Authorization of Commands. A coprocessor, such as TPM, is often shared between various entities on the system, such as firmware, OS, and user-space applications. Further, the enclave technology might support multiple mutually-untrusted tenants. Thus, the coprocessor has to manage the credentials for different enclaves (differentiated using, e.g., *MRSIGNER*, *MRENCLAVE*, *PROVID* and *SVN*). Moreover, the coprocessor is also used by non-enclave code, e.g., the OS, firmware, or user-space application. Consequently, it is crucial to have authorization of coprocessor commands to control access to coprocessor entities (like keys or NVM) to ensure that every enclave and non-enclave code only ever has access to its own coprocessor entities. *TALUS* with SGX and TPM ensures authorization using locality and EAP. Authorization to TPM entities between different actors in the system, e.g., OS, third-party software, or hardware, is based on the

TPM locality. Different enclaves running on the same system authorize via their identities through TPM EAP [11].

SC3. Avoiding Shared Hardware. It is often necessary to securely (SC1) send secret data, e.g., session keys, to the CPU while reducing the amount of shared hardware involved in the communication. Recent transient-execution attacks showed that a software-only attacker can read stale entries in various internal CPU buffers [7, 64, 63, 9, 62]. Thus, *TALUS* provides strict isolation of coprocessor-released data, ensuring that data does not pass (in plaintext) through shared hardware elements with (known) vulnerabilities. *TALUS* implements the entire communication using only CPU registers as storage.

Besides those security challenges, we identify the following functional challenges (FC) that influence *TALUS*.

FC1. Functionality Mapping. Enclave functionalities require a corresponding faithful command mapping offered by the coprocessor, e.g., to generate and use keys with the same authorization policies. The coprocessor driver logic for these commands can be implemented in CPU microcode [33] without requiring hardware changes. The microcode changes have to support only minimal amounts of ephemeral storage for policy sessions and session handles, both of which can be stored in the insecure BIOS flash.

FC2. Attestation. Enclaves depend on attestation to convince (remote) parties that they are communicating with the intended enclave. If the coprocessor supports attestation and management of attestation secrets, the attestation can be outsourced to the coprocessor. Thus, attestation secrets are never stored in shared hardware. A TPM supports remote attestation of TPM internal data. However, this poses the challenge of faithfully integrating the TPM attestation protocols with SGX. *TALUS* achieves this by extending TPM PCR21 with a measurement of an SGX secret (e.g., measurement of the QE). PCR21 is protected using EAP to ensure that only the microcode can access it, and the PCR21 measurement is attested through TPM-based attestation to a remote verifier.

FC3. Asynchronous Execution. When outsourcing cryptographic commands to a potentially much slower coprocessor, we face the problem that the coprocessor execution is asynchronous to the enclave execution. For example, the enclave might be interrupted before the coprocessor finishes executing an issued command by the enclave. Thus, *TALUS* ensures proper scheduling between enclave execution and coprocessor execution to handle asynchronous execution by storing secrets in the special-purpose registers and encrypting them during interrupts, preventing the register content from leaking through unprotected buffers. Interrupts already require a significant amount of microcode execution in the CPU, e.g., SGX stores registers in the SSA and resets the register values to non-secret values. Hence, adding encryption is feasible in microcode.

5 TALUS Implementation

This section briefly introduces the implementation details of a proof-of-concept of *TALUS* based on SGX and a hardware TPM. An in-depth discussion is avail-

able in the extended version [11]. We show the functionality and all the security guarantees using the Intel SGX emulator [45] and a hardware TPM, allowing us to implement the entire design of *TALUS*. For the performance evaluation, we instead use a hardware SGX enclave in combination with the same hardware TPM, with the limitation that the communication channel is not protected against a malicious OS. All evaluations are performed on an Intel i7-7820X running Ubuntu 16.04.04 with kernel 5.0.0. As the TPM, we use an Infineon SLB 9670 that supports TPM 2.0 (HTPM). The size of the enclave used for performance evaluation is 52 kB.

5.1 Connecting SGX and TPM

Channel between SGX and TPM (SC1). Typically, the OS provides the TPM as an MMIO device to the system and user-space software. However, *TALUS* cannot rely on the untrusted OS for communication. For our proof-of-concept implementation, we rely on the end-to-end encrypted programmed I/O channel between the CPU and the hardware TPM. To prevent untrusted system software from interfering with the channel, we distinguish between MMIO and DMA requests. The channel is controlled by Intel TXT using an access control mechanism called *Locality* offered by the TPM through *TPM Locality Address Mapping* [31]. TPM localities indicate the source of the command within the platform. Locality 0 is full public access, locality 1 is the OS, and higher localities (up to locality 4) correspond to the highest privilege levels, *i.e.*, hardware and microcode, including SGX. In *TALUS*, localities ensure the vertical access control to the TPM (e.g., software, OS), while command authorization (cf. Section 5.2) ensures the horizontal access control (*i.e.*, different enclaves).

The channel directly stores data in the CPU registers. Cole and Prakash [15] showed that, in addition to general-purpose registers, sensitive data can also be stored in the Intel MPX `bnd` registers. As Linux or GCC no longer supports Intel MPX [60], these registers can be used by an enclave without conflicting with any other existing software.

Interrupt Handling (FC3). On an interrupt, SGX performs an *Asynchronous Enclave Exit (AEX)* to save the enclave execution state in the *State Save Area (SSA)* before invoking the OS exception handling. Although architecturally secure, RIDL [63], ZombieLoad [64], and $\text{\AE}PIC$ [4] showed that storing registers in the SSA leaves copies of the values in internal CPU buffers from where they can be leaked. Forcing SGX to dump registers to the SSA is always possible, as an attacker can inject interrupts at any time during enclave execution [70].

TALUS does not allow the registers (`BND0-BND3`) holding potentially secret data to be saved directly to the SSA. In our proof-of-concept implementation, we encrypt the registers on `EEXIT`, `EREMOVE`, or `AEX` before storing them. We use AES in counter mode, with the SGX sealing key as the encryption key and the number of asynchronous exits as the counter. Using the number of asynchronous exits as a counter has the advantage that an attacker has only one shot at

both CPU and TPM, SGX sends a secret derived from `SEAL_FUSES` to the TPM as input to the TPM key generation. Other enclave-related information are available in the `SECS` created by SGX for every enclave. More details on the command mapping between SGX and TPM are available in the extended version [11].

Enclave Authorization (SC2). *TALUS* uses TPM’s extended authorization policies (EAP) to ensure that one enclave cannot have unauthorized access to another enclave’s TPM entities. EAP policies are set during the creation of a TPM entity, such as a key. The CPU in *TALUS* dictates the EAP of newly created TPM entities. It handles the policy sessions with the TPM, supplying the necessary information for authorization from the key-derivation material. With EAP, we can represent the same policies reflected in the key-derivation material selection in default SGX. For example, if a key is created with `MRSIGNER` selected but not `MRENCLAVE`, *i.e.*, it can be derived by all enclaves of the same developer, we represent this in an EAP that requires the enclave’s `MRSIGNER` value. When using the key, the CPU supplies the current enclave’s `MRSIGNER` value to the TPM policy session. Only if it matches the value set in the EAP at key creation time can that enclave use the key.

5.3 Limitations of the *TALUS* Implementation

Our proof-of-concept implementation demonstrates that TPM and SGX are very amenable for integration, leading to improved enclave security (cf. Section 6). Our security discussion motivates further research into more secure integration of coprocessors with CPUs. In our proof-of-concept implementation, the CPU uses an end-to-end encrypted channel with pre-shared keys to the TPM (`TPM_TakeOwnership`). Hence, we rely on a non-compromised chipset to, e.g., prevent cuckoo attacks [58]. A coprocessor physically integrated into the CPU, such as Microsoft Pluton [52], can remove the dependency on the chipset for a secure, authenticated connection. While we did not attempt such a tighter integration for the proof-of-concept in this paper, we provide functional objectives and requirements for a secure integration between a coprocessor and an enclave. More details are available in the extended version [11].

6 Case Studies

In this section, we present two case studies using *TALUS*. We demonstrate how *TALUS* protects the enclave life cycle by storing all long-term secrets in the TPM. We also show how to strengthen mitigations against microarchitectural attacks by reducing the amount of data to protect and limiting enclave restarts.

6.1 *TALUS*-backed Enclave Management

Enclave Creation. Figure 2.a shows the two-step process of TPM-backed enclave creation: (i) allocating enclave pages in EPC and addition of code and

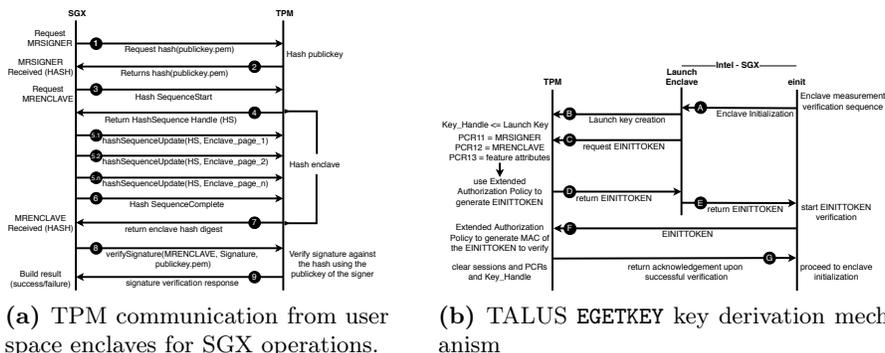


Fig. 2: Enclave-related use-cases for TALUS

data to those pages, and (ii) measuring page contents (MRENCLAVE) and verification of the measurement against a signed reference value. With *TALUS*, the TPM creates and verifies MRSIGNER and MRENCLAVE. These operations require hashing of MRSIGNER using TPM commands like `TPM2_HashSequenceStart`, `TPM2_HashSequenceUpdate` and lastly `TPM2_HashSequenceComplete`. The TPM returns the hash of the measured enclave pages, *i.e.*, MRENCLAVE. SGX verifies the measurement of the enclave code (using the command `TPM2_VerifySignature`) with the reference value signed by the creator of the enclave using the creator’s public key. If the values are the same, the enclave creation is successful.

Enclave Launch. A successfully created enclave is launched using the `EINIT` command. Vanilla SGX employs a complex launch-control mechanism involving the LE, which requires a launch key (LK) [16]. By default, the LK is derived using the same key derivation used for sealing keys, and transferred between the trusted runtime and LE via microarchitectural buffers. Transient-execution attacks [7, 64] attacked these buffers to extract the launch key. *TALUS* replaces this unprotected buffer transfer by encapsulating the key inside the TPM and releasing it upon successful authorization. We implement the launch control using TPM (cf. Figure 2.b). The launch process starts when `EINIT` requests an enclave initialization (cf. Figure 2.b) from the LE. The LE issues an LK request to the TPM with the `TPM2_CreatePrimary` command. Note that this process can also be ported to Intel DCAP.

The related enclave information from Enclave SECS is passed to the TPM. The TPM creates a key using the `EINITTOKEN` KDM as supplied by the CPU. SGX also resets TPM PCRs and extends the enclave information into those PCRs (e.g., PCR 11-13). The PCR extension is a well-known procedure used in, e.g., Flicker [47], other solutions for proof-of-execution [59], and measured boot mechanisms [31]. After the TPM returns a key handle, an `EINITTOKEN` generation request is issued, wrapped in an EAP session using the enclave identity information as policy. Therefore, the authorization succeeds only if the correct enclave information was extended into the PCRs. The TPM creates the `EINITTOKEN`, an

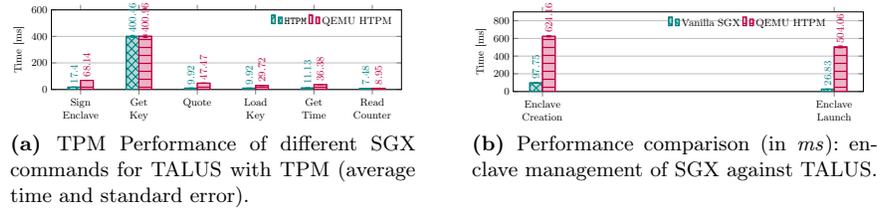


Fig. 3: 2 TALUS performance evaluation

HMAC of the enclave identity information, using the launch key loaded into the TPM. The `EINITTOKEN` is returned to `EINIT` (D) from the LE. `EINIT` receives the `EINITTOKEN` and sends it to the TPM for verification (F). After verification, the TPM returns an acknowledgment of success to `EINIT` (G) to proceed, setting the enclave’s `INIT` attribute to true. This enables a ring 3 application to execute the enclave’s code using SGX instructions. The used PCRs are reset to their predefined values, which is possible because the code runs at locality 4.

Performance of Enclave Management using *TALUS*. Figure 3.b shows the performance of the TPM-backed functions. Enclave creation, which includes allocating enclave pages, measuring page contents, and verifying the measurements, takes on average 624.16 ms with *TALUS* and a hardware TPM (QEMU-HTPM). Compared to vanilla SGX, which also takes 97.75 ms, this is only an overhead of 526.41 ms. Given that the creation of an enclave is a one-time event in the life cycle of an enclave and does not affect any operation at runtime, this overhead is likely amortized over the runtime of the enclave.

SGX Attestation (FC2). For SGX, attestation is implemented in the QE. SGX employs local attestation to prove an enclave’s identity to the QE. The QE uses the attestation keys provisioned to the platform to attest the platform information and the attested enclave’s `MRENCLAVE`. A TPM naturally supports attestation using attestation keys, however, only of TPM-internal data (e.g., PCR values or TPM entities). With *TALUS*, we adapt the mechanism implemented by Intel and AMD for DRTM/Late-launch, where the platform attests with the TPM a small piece of code measured by the CPU. DRTM uses PCR17 of the TPM for measurement attestation. The CPU can only reset PCR17 at locality 4. Hence, a verifier is assured that the attested measurement in PCR17 can only come from the CPU during DRTM. In *TALUS*, we designate PCR21 for SGX attestation and set an EAP on this PCR that allows only locality 4 to read, extend, and reset this PCR. The TPM can attest this policy to a remote verifier to ensure them about this policy. During SGX attestation, the microcode resets PCR21 and extends it with the measurement of the QE (i.e., `MRENCLAVE` of the QE) and the report generated by the QE. A remote verifier can use the attested PCR21 value to check for a trusted QE and the proper report, i.e., `MRENCLAVE` and optionally supplied data to the report. Note that the EPID attestation used by SGX [66] is an extended version of TPM’s DAA and can be modeled entirely using DAA [6]. Simply extending the enclave `MRENCLAVE` into a PCR and attest-

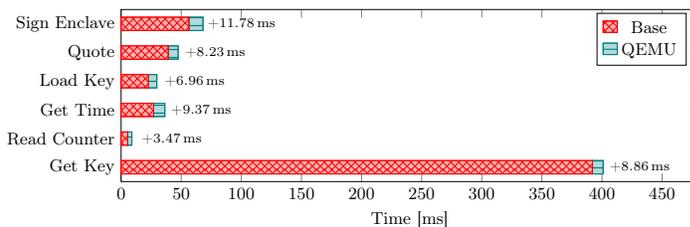


Fig. 4: The total runtime of the commands split into base execution time and the overhead added by QEMU.

ing this PCR is insufficient without ensuring that the `MRENCLAVE` is correct and reported by a trusted entity.

Performance of Other Co-processor Functions We evaluate the runtime of `Sign Enclave`, `Get Key`, `Quote`, `Load key`, `Get Time` and `Read Counter` provided by *TALUS*. As a baseline, we measure the time it takes the hardware TPM (HTPM) to execute these primitives. Figure 3.a shows the average execution time over 1000 measurements and a 95 % confidence interval. Communication between the TPM and SGX adds a small average overhead between 0.49 ms (generating a 2048-bit RSA key) and 50.77 ms (enclave signing).

TALUS running with a hardware TPM adds an average overhead of $98.61 \text{ ms} \pm 1.95 \text{ ms}$. Note that the overall runtime overhead of an enclave depends on its workload, *i.e.*, how often these commands are executed.

Data Encryption using *TALUS*. We evaluate a real-world use case that encrypts data using AES without leaking the key, even in the presence of transient execution attacks (cf. Section 6.2). Our application uses a 128-bit AES key securely stored in the TPM, only fetched when encrypting user-provided data. To ensure no leakage of round keys via the SSA [64], we execute the round-key derivation and encryption within a hardware transaction [21]. The total runtime of encrypting 4 kB of data and cleaning up any secret state is $1.66 \mu\text{s} \pm 0.001 \mu\text{s}$, excluding fetching the key from the TPM. The overhead from *TALUS*, *i.e.*, securely getting the key, is $58.43 \text{ ms} \pm 1.45 \text{ ms}$. As a baseline, we compare the runtime to a variant where the key is not fetched from the TPM but unsealed from the disk. This (insecure) variant has an average runtime of $199.21 \mu\text{s} \pm 0.45 \mu\text{s}$. Note that the one-time overhead is amortized if the enclave runtime increases, *e.g.*, if larger amounts of data are encrypted.

Since only Intel can implement a native version of *TALUS*, and there is no cycle-accurate emulator that supports SGX, we can only provide an estimate for such a version. Figure 4 shows the overhead added by QEMU for the *TALUS* commands, adding an overhead between 5 ms to 10 ms (avg. of 6.82 ms). This overhead constitutes between 2.21 % to 38.77 % (avg. of 21.60 %). We assume that commands in a native *TALUS* implementation are around 20 % faster.

6.2 Impeding Microarchitectural Attacks

SGX enclaves are a constant target of microarchitectural attacks [69, 49]. The property that enclaves can be started arbitrarily often makes it challenging to write side-channel-resilient code [49]. Furthermore, with transient-execution attacks such as Foreshadow [7], Spectre [39], RIDL [63], ZombieLoad [64], and architectural vulnerabilities such as $\text{\AE}PIC$ [4], attackers can leak sensitive data from internal CPU buffers despite side-channel-resilient code.

Preventing Transient-Execution Attacks. TPMs are assumed to be resilient against other forms of microarchitectural attacks since no untrusted code can access the hardware of a TPM. Further, by design, TPM does not release any secret keys managed by TPM to the outside, but only *key handles*. However, sometimes the TPM needs to release secret data to the enclave (e.g., a decrypted symmetric key). With *TALUS*, data is loaded directly into CPU registers. No transient-execution attack against CPU general-purpose registers has been demonstrated [8]. Note that Meltdown attacks were only shown against system registers [23, 8] and floating-point and the upper half of SIMD registers in specific scenarios [67, 53, 24]. Hence, as long as a secret is only stored in, e.g., an MPX register (`BND0-BND3`), it cannot be leaked using a transient-execution attack. Otherwise, Meltdown mitigations, such as KPTI, would also be ineffective.

Proof-of-concept Evaluation. As a proof of concept, we reproduce the AES-NI encryption from ZombieLoad [64]. With *TALUS*, we can load the AES key from the TPM directly into the CPU registers without requiring a memory load. Hence, the attack vector used by Schwarz et al. [64] is mitigated. To mitigate the remaining attack vector, the storing and loading of the `XMM` registers in the SSA, we rely on Cloak [21] to not leak any intermediate results from the registers to memory. We verify that the plain AES key is never stored in memory by inspecting the memory. Further, we are certain that the key is not stored in any vulnerable microarchitectural element used for interacting with the memory, such as the store buffer or line-fill buffer, preventing leakage via transient-execution attacks. However, we cannot exclude the existence of unknown buffers that are on the data path in Cloak [21] and that might become vulnerable in the future.

Limiting Precise Execution Control & Strengthening Countermeasures. Due to the strong attacker model, SGX enclaves can be interrupted at an arbitrary point, allowing precise execution control [49]. With SGX-step [70], enclaves can be interrupted after every instruction, allowing to amplify side-channel leakage. Constant interruptions result in constantly storing and loading of the enclave state, resulting in more reliable transient-execution attacks [7, 64]. By design, *TALUS* does not store secrets stored in the MPX registers in plain memory, preventing leakage of these values (cf. Section 5.1). While *TALUS* cannot directly prevent precise execution control, its persistent storage can track how often an enclave was interrupted. Although enclaves can detect interrupts via overwritten values in the SSA [57, 12], they cannot store this information across enclave restarts. With *TALUS*, an enclave can track the number of in-

interrupts across enclave restarts. Due to this persistent storage, an enclave can refuse to start if it suffers from an excessive number of interrupts.

Generally, *TALUS* allows enclaves to keep information across restarts, strengthening state-of-the-art countermeasures against microarchitectural attacks. T-SGX [65], Varys [57], or Déjà Vu [13] drastically reduce the observable leakage during one enclave run. However, since they cannot prevent arbitrary enclave restarts, leakage is still possible [36]. Using secure counters of *TALUS* strengthens such countermeasures to prevent an enclave from starting if too many abnormal events have been observed during execution.

Proof-of-concept Evaluation. We implement the restart limitation in the sample enclave of T-SGX [65]. The enclave first increments a counter stored in the TPM and retrieves the current value. This value is the number of times the enclave has been started. Only if the current counter value is below an enclave-defined threshold the enclave continues to provision the secrets. The limit can be obtained from a remote server to increase the number of allowed executions over time gradually. Contrary to the number of enclave executions, storing this threshold in a sealed data blob is possible. A rollback attack would only decrease the number of remaining enclave executions, providing no advantages to an attacker. As the check only happens once at enclave startup, this is a one-time overhead. With T-SGX, the time it takes to create and launch the enclave is $19.66\text{ ms} \pm 0.016\text{ ms}$ ($n = 1000$). Increasing, reading, and comparing the timer with *TALUS* takes on average $17.45\text{ ms} \pm 0.23\text{ ms}$.

7 Other Platforms

TALUS shows how a co-processor can be integrated with a TEE on x86. Other platforms, such as ARM and RISC-V, can also benefit from our requirement analysis. For example, ARM TrustZone supports co-processors such as Google Titan or Apple T2 but with limited use cases such as disk encryption, key generation or encryption. On RISC-V, Keystone Enclaves and RoCC (Rocket chip coprocessor) are available on the Boom core [10] and Rocket core [2]. Hence, also on RISC-V, integrating the co-processor with enclaves can provide better security guarantees. A detailed discussion on how other platform can benefit from a *TALUS* implementation is available in the extended version [11].

8 Conclusion

We showed that secure enclaves, such as SGX, can benefit from secure coprocessors, such as a TPM, if they are securely integrated. With *TALUS*, we presented a design that supports secure side-channel-resilient communication between TEEs and cryptographic coprocessors. We presented a proof-of-concept implementation based on a hardware TPM and SGX, demonstrating how a TPM can protect the SGX infrastructure credentials during enclave building and launching, and how such a design impedes microarchitectural attacks on SGX. From our prototype, we derive crucial requirements for secure integration

between TEEs and coprocessors. We believe that the identified and solved challenges leading to our design of *TALUS* are valuable for future systems, such as integrating Microsoft’s Pluton with enclaves, and can be transferred to other combinations of enclave technology and coprocessors, such as AMD PSP or ARM TrustZone.

References

1. Arm Limited: The trustzone hardware architecture. <https://developer.arm.com/documentation/100935/0100/The-TrustZone-hardware-architecture-?lang=en> (2021), accessed: 01.05.2021
2. Asanović, K., et al.: The rocket chip generator. Tech. rep., EECS Department, University of California, Berkeley (Apr 2016), <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-17.html>
3. Boone, J.: Tpm genie: Attacking the hardware root of trust for less than \$50 (2018), accessed: 02/13/2019
4. Borrello, P., et al.: \mathbb{A} EPIC Leak: Architecturally leaking uninitialized data from the microarchitecture. In: USENIX Security 22 (2022)
5. Bourgeat, T., et al.: Mi6: Secure enclaves in a speculative out-of-order processor. In: MICRO ’52 (2019)
6. Brickell, E., Li, J.: Enhanced privacy id: A direct anonymous attestation scheme with enhanced revocation capabilities. IEEE Transactions on Dependable and Secure Computing (2012)
7. Bulck, J.V., et al.: Foreshadow: Extracting the keys to the intel SGX kingdom with transient out-of-order execution. In: 27th USENIX Security Symposium (USENIX Security 18) (2018)
8. Canella, C., et al.: A Systematic Evaluation of Transient Execution Attacks and Defenses. In: USENIX Security Symposium (2019)
9. Canella, C., et al.: Fallout: Leaking data on meltdown-resistant cpus. In: CCS (2019)
10. Celio, C., et al.: Boom v2: an open-source out-of-order risc-v core. Tech. rep., EECS Department, University of California, Berkeley (Sep 2017), <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-157.html>
11. Chakraborty, D., Schwarz, M., Bugiel, S.: Talus: Reinforcing tee confidentiality with cryptographic coprocessors (technical report) (2023)
12. Chen, G., et al.: Defeating Speculative-Execution Attacks on SGX with HyperRace. In: Dependable and Secure Computing (DSC) (2019)
13. Chen, S., Zhang, X., Reiter, M.K., Zhang, Y.: Detecting privileged side-channel attacks in shielded execution with déjà vu. In: AsiaCCS (2017)
14. Cloosters, T., Rodler, M., Davi, L.: Teerex: Discovery and exploitation of memory corruption vulnerabilities in {SGX} enclaves. In: USENIX Security Symposium (2020)
15. Cole, M., Prakash, A.: Simplex: Repurposing intel memory protection extensions for information hiding. arXiv:2009.06490 (2020)
16. Costan, V., Devadas, S.: Intel sgx explained. IACR Cryptology ePrint Archive (2016)
17. Futral, W., Greene, J.: Intel Trusted Execution Technology for Server Platforms: A Guide to More Secure Datacenters. Apress, Berkely, CA, USA, 1st edn. (2013)

18. Garmany, B., Müller, T.: PRIME: Private RSA Infrastructure for Memory-less Encryption. In: ACSAC (2013)
19. Google: Tpm usage - the chromium project. <https://www.chromium.org/developers/design-documents/tpm-usage> (2019)
20. Greene, James: Intel® smi transfer monitor (stm) user guide. <https://firmware.intel.com/content/smi-transfer-monitor-stm> (2016)
21. Gruss, D., et al.: Strong and Efficient Cache Side-Channel Protection using Hardware Transactional Memory. In: USENIX Security Symposium (2017)
22. Han, S., Shin, W., Park, J.H., Kim, H.: A bad dream: Subverting trusted platform module while you are sleeping. In: USENIX Security 18 (2018)
23. Intel: Rogue system register read (2018), <https://software.intel.com/content/www/us/en/develop/articles/software-security-guidance/advisory-guidance/rogue-system-register-read.html>
24. Intel: Vector Register Sampling / CVE-2020-0548 / INTEL-SA-00329 (2020), <https://software.intel.com/content/www/us/en/develop/articles/software-security-guidance/advisory-guidance/vector-register-sampling.html>
25. Intel Corporation: Strengthening security with intel®platform trust technology. <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/enterprise-security-platform-trust-technology-white-paper.pdf> (2014)
26. Intel Corporation: Intel® 64 and ia-32 architectures software developer’s manual. <https://www.intel.com/content/dam/www/public/us/en/documents/manuals/64-ia-32-architectures-software-developer-instruction-set-reference-manual-325383.pdf> (2016)
27. Intel Corporation: Intel® software guard extensions (intel sgx). <https://software.intel.com/en-us/sgx> (2016), accessed: 15.07.2019
28. Intel Corporation: Intel converged security and management engine (intel csme). <https://www.intel.com/content/dam/www/public/us/en/security-advisory/documents/intel-csme-security-white-paper.pdf> (2020)
29. Intel Corporation: Intel-sa-00219 sgx sw developer guidance. <https://www.intel.com/content/dam/www/public/us/en/security-advisory/documents/the-intel-csme-dam-vulnerability-cve-2018-3659-and-cve-2018-3643-whitepaper.pdf> (2020), accessed: 11/12/2021
30. Intel Corporation: The intel® converged security and management engine iommu hardware issue – cve-2019-0090 and cve-2020-0566. <https://www.intel.com/content/dam/www/public/us/en/security-advisory/documents/cve-2019-0090-whitepaper.pdf> (2020), accessed: 11/12/2021
31. Intel Corporation: Intel® trusted execution technology (intel® txt) software development guide measured launch environment developer’s guide. <http://www.intel.com/content/www/us/en/software-developers/intel-txt-software-development-guide.html> (2021)
32. Intel Corporation: Unable to find alternatives to monotonic counter application programming interfaces (apis) in intel® software guard extensions (intel® sgx) for linux* to prevent sealing rollback attacks. <https://www.intel.com/content/www/us/en/support/articles/000057968/software/intel-security-products.html> (2021), accessed: 10.01.2022
33. Intel Corporation: Xucode: An innovative technology for implementing complex instruction flows. <https://software.intel.com/content/www/us/en/develop/articles/software-security-guidance/secure-coding/xucode-implementing-complex-instruction-flows.html> (2021)

34. Intel Corporation: Kvm sgx. <https://github.com/intel/kvm-sgx> (2022), accessed: 01/06/2020
35. Intel Corporation: Qemu sgx. <https://github.com/intel/qemu-sgx> (2022), accessed: 01/06/2020
36. Jiang, J., Soriente, C., Karame, G.: Monitoring performance metrics is not enough to detect side-channel attacks on intel sgx. arXiv:2011.14599 (2020)
37. Kauer, B.: Oslo: Improving the security of trusted computing. In: Proc. 16th USENIX Security Symposium (SEC '07). USENIX Association (2007)
38. Kenjar, Z., et al.: VOLTpwn: Attacking x86 Processor Integrity from Software. In: USENIX Security Symposium (2020)
39. Kocher, P., et al.: Spectre attacks: Exploiting speculative execution. In: 2019 IEEE Symposium on Security and Privacy (SP) (2019)
40. Lawson, N.: Tpm hardware attacks. <https://rdist.root.org/2007/07/16/tpm-hardware-attacks/> (Jul 2007), accessed: 06.08.2018
41. Lee, D., Jung, D., Fang, I.T., Tsai, C.C., Popa, R.A.: An off-chip attack on hardware enclaves via the memory bus. In: USENIX Security 20 (2020)
42. Lee, D., Kohlbrenner, D., Shinde, S., Asanović, K., Song, D.: Keystone: An open framework for architecting trusted execution environments. In: Proceedings of the Fifteenth European Conference on Computer Systems (2020), <https://dl.acm.org/doi/abs/10.1145/3342195.3387532>
43. Lee, J., Jang, J., Jang, Y., Kwak, N., Choi, Y., Choi, C., Kim, T., Peinado, M., Kang, B.B.: Hacking in Darkness: Return-oriented Programming against Secure Enclaves. In: USENIX Security Symposium (2017)
44. Li, M., Zhang, Y., Wang, H., Li, K., Cheng, Y.: CIPHERLEAKS: Breaking Constant-time Cryptography on AMD SEV via the Ciphertext Side Channel. In: USENIX Security Symposium (2021)
45. M., J.: Virtualizing intel® software guard extensions with kvm and qemu. <https://software.intel.com/en-us/articles/virtualizing-intel-software-guard-extensions-with-kvm-and-qemu> (2019), accessed: 01/06/2020
46. Matetic, S., et al.: ROTE: Rollback protection for trusted execution. In: Proc. 26th USENIX Security Symposium (SEC' 17). usenix (2017)
47. McCune, J.M., Parno, B.J., Perrig, A., Reiter, M.K., Isozaki, H.: Flicker: An execution infrastructure for tcb minimization. SIGOPS Oper. Syst. Rev. (2008)
48. McKeen, F., Alexandrovich, I., Berenzon, A., Rozas, C.V., Shafi, H., Shanbhogue, V., Savagaonkar, U.R.: Innovative instructions and software model for isolated execution. Hasp@ isca **10**(1) (2013)
49. Michael Schwarz, D.G.: How Trusted Execution Environments Fuel Research on Microarchitectural Attacks. In: Security & Privacy (2020)
50. Michael Schwarz, Samuel Weiser, D.G.: Practical Enclave Malware with Intel SGX. In: DIMVA (2019)
51. Microsoft Corporation: Bitlocker. <https://docs.microsoft.com/en-us/windows/security/information-protection/bitlocker/bitlocker-overview> (2018)
52. Microsoft Corporation: Meet the microsoft pluton processor – the security chip designed for the future of windows pcs. <https://www.microsoft.com/security/blog/2020/11/17/meet-the-microsoft-pluton-processor-the-security-chip-designed-for-the-future-of-windows-pcs/> (2021)
53. Moghimi, D., et al.: Medusa: Microarchitectural Data Leakage via Automated Attack Synthesis. In: USENIX Security Symposium (2020)
54. Moghimi, D., et al.: TPM-FAIL: TPM meets Timing and Lattice Attacks. In: 29th USENIX Security Symposium (USENIX Security 20) (2020)

55. Müller, T., Freiling, F.C., Dewald, A.: Tresor runs encryption securely outside ram. In: USENIX Security Symposium (2011)
56. Murdock, K., et al.: Plundervolt: Software-based Fault Injection Attacks against Intel SGX. In: S&P (2020)
57. Oleksenko, O., Trach, B., Krahn, R., Silberstein, M., Fetzer, C.: Varys: Protecting SGX Enclaves from Practical Side-channel Attacks. In: USENIX ATC (2018)
58. Parno, B.: Bootstrapping trust in a "trusted" platform. In: Proceedings of the 3rd Conference on Hot Topics in Security. HOTSEC'08 (2008)
59. Perez, R., Sailer, R., van Doorn, L., et al.: vtpm: virtualizing the trusted platform module. In: USENIX Security 06. pp. 305–320 (2006)
60. Phoronix: Intel MPX Support Is Dead With Linux 5.6 (2020), https://www.phoronix.com/scan.php?page=news_item&px=Intel-MPX-Is-Dead
61. Qiu, P., Wang, D., Lyu, Y., Qu, G.: Voltjockey: Breaching trustzone by software-controlled voltage manipulation over multi-core frequencies. In: Proceedings of the 2019 ACM CCS 2019 (2019)
62. Ragab, H., et al.: CrossTalk: Speculative Data Leaks Across Cores Are Real. In: S&P (2021)
63. van Schaik, S., et al.: RIDL: Rogue in-flight data load. In: S&P (May 2019)
64. Schwarz, M., et al.: ZombieLoad: Cross-privilege-boundary data sampling. In: CCS (2019)
65. Shih, M.W., Lee, S., Kim, T., Peinado, M.: T-SGX: Eradicating controlled-channel attacks against enclave programs. In: NDSS (2017)
66. Simon Johnson and Vinnie Scarlata and Carlos Rozas and Ernie Brickell and Frank Mckeen: Intel software guard extensions: Epid provisioning and attestation services. <https://software.intel.com/sites/default/files/managed/57/0e/ww10-2016-sgx-provisioning-and-attestation-final.pdf> (2016)
67. Stecklina, J., Prescher, T.: LazyFP: Leaking FPU Register State using Microarchitectural Side-Channels. arXiv:1806.07480 (2018)
68. Trusted Computing Group: Tpm 2.0 library specification. <https://trustedcomputinggroup.org/resource/tpm-library-specification/> (2016)
69. Van Bulck, J.: Microarchitectural Side-Channel Attacks for Privileged Software Adversaries. Ph.D. thesis, KU Leuven (2020)
70. Van Bulck, J., Piessens, F., Strackx, R.: Sgx-step: A practical attack framework for precise enclave execution control. SysTEX'17 (2017)
71. Van Bulck, J., et al.: A tale of two worlds: Assessing the vulnerability of enclave shielding runtimes. In: CCS (2019)
72. Van Bulck, J., et al.: LVI: Hijacking Transient Execution through Microarchitectural Load Value Injection. In: S&P (2020)
73. Wampler, J., Martiny, I., Wustrow, E.: Exspectre: Hiding malware in speculative execution. In: NDSS (2019)
74. Weichbrodt, N., Kurmus, A., Pietzuch, P., Kapitza, R.: AsyncShock: Exploiting Synchronisation Bugs in Intel SGX Enclaves. In: ESORICS (2016)
75. Xu, Y., Cui, W., Peinado, M.: Controlled-channel attacks: Deterministic side channels for untrusted operating systems. In: ssp15. IEEE Computer Society (2015)
76. Zhang, F., Zhang, H.: Sok: A study of using hardware-assisted isolated execution environments for security. In: Proceedings of the Hardware and Architectural Support for Security and Privacy 2016. HASP 2016 (2016)
77. Zhang, T., Koltermann, K., Evtyushkin, D.: Exploring branch predictors for constructing transient execution trojans. In: ASPLOS (2020)