Formal Security Analysis of the AMD SEV-SNP Software Interface

Petar Paradžik, University of Zagreb Faculty of Electrical Engineering and Computing, petar.paradzik@fer.hr
Ante Derek, University of Zagreb Faculty of Electrical Engineering and Computing, ante.derek@fer.hr
Marko Horvat, University of Zagreb Faculty of Science, marko.horvat@math.hr

Abstract—AMD Secure Encrypted Virtualization technologies enable confidential computing by protecting virtual machines from highly privileged software such as hypervisors. In this work, we develop the first, comprehensive symbolic model of the software interface of the latest SEV iteration called SEV Secure Nested Paging (SEV-SNP). Our model covers remote attestation, key derivation, page swap and live migration. We analyze the security of the software interface of SEV-SNP by verifying critical secrecy, authentication, attestation and freshness properties, and find that the platform-agnostic nature of messages exchanged between SNP guests and the AMD Secure Processor firmware presents a weakness of the design. We show multiple ways of exploiting this weakness, including the compromise of attestation report integrity, and suggest slight modifications to the design which let third parties detect guest migrations to vulnerable platforms.

Index Terms—trusted execution environments, formal security analysis, SEV-SNP, system verification

I. INTRODUCTION

An increasing number of cloud providers are beginning to rely on trusted execution environments (TEEs) to ensure the safety of user data while it is being processed by applications and services on foreign platforms. TEEs are tamper-resistant environments intended to provide the confidentiality and integrity of user code and data in use, and isolate them from untrusted, yet highly privileged software such as operating system kernels and hypervisors. This makes TEEs suitable for processing secrets on remote devices and platforms such as public clouds. Modern TEEs also provide useful features such as remote attestation, a mechanism by which a remote party can obtain evidence that a particular process or virtual machine (VM) is running correctly with the expected configuration.

Hardware-based TEE solutions typically involve a security kernel executed on a main processor with security extensions, or on a coprocessor. The security kernel is a piece of software hardwired on a chip (ARM TrustZone [1]), or it comes in the form of microcode (Intel SGX [2]) or firmware (AMD SEV [3]) that resides in an isolated area within the processor, which is not directly accessible or modifiable by external code. The software implements an interface as an instruction set that can be used to establish a secure environment.

Historically, the first of the hardware-based TEE solutions from AMD was named Secure Virtualization (SEV) [3]. Developed in 2016, SEV was made available for the first generation of the AMD EPYC brand of x86-64 microprocessors, based on the Zen 1 microarchitecture codenamed Naples. SEV enhanced VM security through VM memory encryption and isolation.

It also offered live VM migration and launch attestation features. The latter represented a rather limited form of remote attestation, allowing only the guest owner to attest the integrity of their guest VMs during the guest launch procedure.

A year later, AMD introduced SEV Encrypted State (SEV-ES) [4], making it available for the second generation of AMD EPYC microprocessors based on the Zen 2 microarchitecture codenamed Rome. SEV-ES encrypts and protects the integrity of CPU registers that store information about the VM runtime before handing over control to a hypervisor, thereby preventing the hypervisor from reading sensitive information of the guest as well as tampering with the control flow of the guest.

In this paper, we focus on AMD SEV Secure Nested Paging (SEV-SNP), which came out in 2020 and represents the latest iteration of the AMD SEV technologies. It is currently available for the third and fourth generation of the AMD EPYC microprocessors, based on Zen 3 and Zen 4 microarchitectures codenamed Milan and Genoa, respectively. With the added precaution of treating the hypervisor as fully untrusted, AMD SEV-SNP introduces memory integrity protection as well as other security and usability enhancements to the existing SEV and SEV-ES functionalities, including more flexibility of essential features such as remote attestation and live migration. AMD SEV-SNP is used by popular cloud providers such as Microsoft Azure [5], AWS [6], and Google Cloud [7] to help provide Infrastructure as a Service.

There has been a substantial amount of analysis of the SEV technology from both academia [8, 9, 10, 11, 12, 13] and industry [14]. However, to the best of our knowledge, there is no work that formally analyzes the software interface of any AMD SEV technology, including SEV-SNP.

AMD SEV-SNP is a complex system which supports a large number of guest policies and features, with most having multiple variants, and it uses an intricate key schedule. This makes it very difficult to ascertain whether an adversary could launch an interaction attack by using the interface in an unexpected way. The main research question that we would like to answer in this work is the following:

Can an adversary use the AMD SEV-SNP software interface to force the system into an undesirable state?

In this work, we employ the TAMARIN PROVER (TAMARIN) protocol verification tool to meticulously model and analyze the software interface of AMD SEV-SNP. The model encom-

passes all key features and captures many subtle behaviors of SEV-SNP. We have specified and analyzed nearly a hundred properties and find several attacks. Specifically, our contributions are as follows:

- We develop a formal model of the AMD SEV-SNP software interface. The model is close to being fully comprehensive; it covers protected guest launch, remote attestation, key derivation, page swap and live migration. To the best of our knowledge, there exists no prior formal model of the SEV-SNP software interface.
- We give automated formal proofs for critical secrecy, authentication, attestation, and freshness properties, including the proof of correct stream cipher usage—no key can ever be reused with the same nonce.
- We exploit the platform-agnostic nature of SNP-protected guest messages and find attacks on several authentication and attestation properties, including the compromise of attestation report integrity. We also discover a vulnerability that allows a cloud provider to trick a third party into incorrectly believing that an SNP-protected guest is running on a platform with secure, up-to-date firmware.
- We suggest slight modifications to the design which let third parties detect guest migrations to vulnerable platforms.

Paper Outline. We first give the necessary background related to AMD SEV-SNP and the TAMARIN tool in Section II. Next, we explain the minutiae of our formal model of AMD SEV-SNP in Section III. We provide the details of our analysis in Section IV and thoroughly discuss both the positive and negative results in Section V; we also suggest possible countermeasures. Finally, we give an overview of the related work in Section VI and conclude in Section VII.

II. BACKGROUND

The core ideas behind SEV-SNP were outlined in a white paper published in January 2020 [15]. Subsequently, the technology was officially specified; the current revision is 1.55 and was released in September 2023 [16]. In June 2022, Linux 5.19-rc1 was released with SEV-SNP support [17], and in August 2023, AMD made the SEV-SNP Genoa firmware source code publicly available on GitHub [18].

SEV-SNP provides a novel application binary interface (ABI) intended for a hypervisor to bootstrap and manage virtual machines within a secure environment. It extends SEV technologies by enhancing memory integrity protection, attestation, and virtualization capabilities. In particular, it

- enables virtual machines to run in an isolated environment where their memory contents is not only encrypted, but also authenticated and protected from unauthorized accesses by hypervisors;
- utilizes Trusted Computing Base (TCB) versioning to guarantee that guests run under up-to-date firmware;
- adds a versatile remote attestation mechanism where a third party may establish trust in a guest during runtime of the guest;
- supports generating guest key material from different sources;

- enhances the flexibility of live virtual machine migration by introducing an entity called a *Migration Agent*;
- provides several additional features such as secure nested virtualization.

SEV-SNP prevents a hypervisor from compromising the memory integrity of an SNP-protected guest. The potential attacks include replacing VM memory with an old copy (*replay attack*) or mapping two different guest physical addresses to a single system physical address or DRAM page (*memory aliasing attack*). It does so by utilizing a data structure called Reverse Map Table (RMP) and a page validation mechanism to track page ownership and enforce proper page access control.

In previous SEV instances, a hypervisor was assumed not to be malicious, but potentially buggy. Like before, from the perspective of a single SNP-protected guest, other VMs—whether non-SNP (legacy) or SNP-protected—are also regarded as untrusted entities. However, unlike before, SEV-SNP treats the hypervisor as fully untrusted, capable of tampering with page tables, injecting arbitrary events, and providing false system information.

The TCB of SEV-SNP comprises the AMD System on Chip, which includes the AMD (Platform) Security Processor (AMD-[P]SP), and the software running on top of it, i.e. the microcode, bootloader, operating system, and the SNP firmware which implements the SNP ABI. Each TCB software component can be upgraded and its security version numbers are included in the TCB_VERSION structure that is associated with each SNP VM image.

SEV-SNP introduces an interface that the SNP-protected guest may utilize to request services from the firmware during runtime. This enables the guest to obtain—via so called *guest messages*—attestation reports and derived keys, for instance. These messages are encrypted and integrity protected by the Virtual Machine Platform Communication Key (VMPCK). The firmware generates and installs this key into both the guest context that it maintains and the guest memory pages during the guest launch.

In SEV and SEV-ES, the attestation procedure was confined to the guest launch, limiting its use to a singular entity—the guest owner. SEV-SNP introduces a more versatile approach by replacing launch attestation with remote attestation, enabling any third party to acquire an attestation report and establish trust in a guest at any given moment.

The attestation report comprises various information about the guest, such as its migration policy and image digest (measurement), and well as the platform it operates on, such as the chip identifier and TCB version. It is signed using a private ECDSA P-384 key that is either a Versioned Chip Endorsement Key (VCEK)—a machine-specific key derived from chip-unique secrets and a digest of TCB_VERSION—or a Versioned Loaded Endorsement Key (VLEK), which is a cloud provider-specific key derived from a seed maintained by the AMD Key Derivation Service (KDS). The inclusion of the TCB version allows a third-party to reject the signature if it originated from an unpatched AMD-SP.

In order to further demonstrate the authenticity of the attestation report signature, the VCEK/VLEK is verified against the AMD Signing Key (ASK), which is further verified against

the AMD Root Key (ARK), both of which are 4096-bit RSA keys.

SEV-SNP introduces a robust key derivation mechanism. It enables an SNP-protected guest to instruct the firmware to derive a key rooted in either VCEK, VLEK, or a Virtual Machine Root Key (VMRK). Moreover, the guest has the option to request adding further data, such as its launch digest and TCB version, into the key derivation process. Such keys may be used for data sealing and other purposes.

Similar to previous SEV technologies, SEV-SNP offers an interface for secure swapping wherein the guest may be saved on a disk and later resumed. The firmware ensures confidentiality and authenticity of the guest memory pages by utilizing an Offline Encryption Key (OEK). The swapping plays a crucial role in live migration.

Live migration is a mechanism by which guests may be seamlessly and securely transferred to another physical system, without having to shut themselves down first. Whereas in SEV and SEV-ES the firmware on the source machine was responsible for authenticating the firmware on the destination machine prior to guest context transfer, in SEV-SNP this task is facilitated by an entity called a Migration Agent.

A Migration Agent is an SNP-protected guest that is responsible for enforcing guest migration policies and providing the firmware with guest-unique secrets (VMRK). Whereas a single guest may be associated with at most one migration agent, a single migration agent may be associated with and manage multiple guests concurrently. This association is indicated in each attestation report. SEV-SNP does not specify the behaviour of migration agents nor the manner by which the guest context is securely transferred over the network.

A. Tamarin Prover

TAMARIN PROVER (TAMARIN) [19] is an automated symbolic verification tool for security protocols. It is based on multiset-rewriting; more precisely, its semantics comprises a labeled transition system whose states are multisets of *facts*, and the transitions between them are specified by *rules* which prescribe the behavior of protocol participants and the adversary. Each rule has a *left-hand side* (facts that must be available in the current global state for the rule to execute), *actions* (labels which are logged in the trace and used to express the desired security properties), and a *right-hand side* (facts that will be added to the state).

The left-hand and right-hand side contain multisets of facts, each fact being either *linear* or *persistent* (the latter are prefixed with an exclamation point). Facts can be *produced* (when executing a rule with the fact on the right-hand side) and *consumed* (if it is both linear and on the left-hand side). While there is no bound on the number of times a fact can be produced, linear facts model limited resources that can only be consumed as many times as they are produced, and persistent facts model unlimited resources, which (once produced) can be consumed any number of times.

Fresh variables, denoted by the prefix "~" (or suffix ":fresh"), indicate freshly generated names. They are suitable for modelling randomly generated values such as keys and

thread identifiers. The built-in fact $Fr(\cdot)$ can be utilized to generate such names. *Public* variables, identified by the prefix "\$" (or suffix ":pub"), are used to represent publicly known names such as agent identities and group generators. Additionally, *temporal* variables, prefixed with "#", signify timepoints.

The TAMARIN *builtins* include equational theories for Diffie-Hellman operations, (a)symmetric encryption, digital signatures, and hashing. The theory *multiset* can be used to model monotonic counters. Additionally, TAMARIN supports user-defined function symbols and equational theories. For example, in our model, two ternary function symbols, snenc and sndec are defined, along with the equation

```
sndec(snenc(msg, nonce, key), nonce, key) = msg
```

which models a symmetric cipher with a nonce as an initialization vector; incrementing the nonce is modelled by using the built-in *multiset* theory.

Consider the following rule (taken from our SEV-SNP model and simplified) that enables an SNP-protected guest to request an attestation report.

```
[ StateGuest('RUNNING', $image, ~key, nonce, ~ptr)
2, In(rd), Fr(~newPtr) ]
3 —[ ReportRequest($image, rd) ]
4 [ Out(snenc(\( 'MSG_REPORT_REQ', rd \), key, nonce))
5, StateGuest('WAIT', $image, ~key, nonce, ~newPtr) ]
```

The rule requires that a guest is in the RUNNING state before it can be executed: it consumes the linear fact StateGuest('RUNNING',\$image, \sim key, nonce, \sim ptr) from the global state, receives an arbitrary message rd (which will be included in the attestation report) from the network using the In(·) fact, and generates a new pointer newPtr to its updated state using the Fr(·) fact.

Subsequently, the guest constructs a plaintext, which is an ordered pair ('MSG_REPORT_REQ', rd) comprising the message tag and received data, encrypts it with a symmetric key using a nonce and obtains the ciphertext snenc(('MSG_REPORT_REQ', rd), key, nonce). sends the ciphertext to the network using the Out(·) fact, and a new linear fact StateGuest('WAIT',\$image, ~key, nonce) is produced. This fact can now be consumed by a rule wherein the guest receives the attestation report.

TAMARIN assumes a Dolev-Yao adversary who carries all messages between protocol participants. More precisely, the KU and KD facts represent the adversary knowledge and its ability to receive messages from the network and send messages to the network, respectively, by using $Out(x) - [\rightarrow KD(x)]$ and $KU(x) - K(x) \rightarrow In(x)$ communication rules (as with all actions, the K action is part of the TAMARIN property specification language and can be used to express that the adversary necessarily knows the term it sends to the network). The adversary can try to deconstruct messages it received (constructing any keys needed) in order to gain knowledge of additional terms (e.g. decrypt a ciphertext with the help of the rule KD(snenc(x,y)), $KU(y) - \mapsto KD(x)$ and then switch via $KD(x) - \bigcup KU(x)$ to constructing new messages by applying cryptographic operations on known messages (e.g. applying a hash function can be done with the rule $KU(x) - [\rightarrow KU(h(x))]$.

The trace (security) properties of a protocol encoded as a set of multiset-rewriting rules are specified as temporal first-order formulas (*lemmas*) over the rule labels. For example, the following secrecy property states that, for all SNP firmware threads (chipTID), guest VM contexts (gctx), VM Platform Communication Keys (vmpck), and timepoints #i and #j, if a chipTID generates vmpck and installs it in gctx at some timepoint #i. and if the adversary knows vmpck at #j, then it necessarily holds that there exists some timepoint #k such that the adversary corrupted vmpck at #k and #k precedes #j:

TAMARIN proves trace properties by *falsification*. In order to prove that a property is true, TAMARIN tries to find a counterexample execution—an alternating sequence of states and transitions that satisfies the negation of the property in question. If TAMARIN halts the analysis and succeeds, then the resulting execution represents an attack; if TAMARIN halts the analysis and fails, then it provides a proof of the property in the form of a tree. Lastly, TAMARIN may not terminate as verifying security properties is in general undecidable. There are several ways to avoid non-termination and to improve verification time. For example, we can customize the ranking of proof goals by writing scripts called (*proof*) oracles and use supporting lemmas to prove other lemmas.

Finally, *restrictions* can be used to filter out traces that need not be considered during the security analysis or to enforce certain behaviour such as verification of signatures or branching. One such restriction in our model states that a memory pointer that is read and released at any point in time must be read before it is released.

```
1 \forall ptr #i #j. Read(ptr)@i \land Free(ptr)@j \Rightarrow #i < #j
```

III. FORMAL MODEL OF AMD SEV-SNP

Our goal is to model and analyze the software interface of AMD SEV-SNP. We aim to capture all of its principal features, including remote attestation, key derivation, page swapping and live migration, while ignoring the low-level details such as the memory encryption and RMP structure. We consider a powerful Dolev-Yao adversary that has full control over the communication network in an idealized cryptography setting. Before we delve into the intricacies of our model, we briefly explain the methodology we use.

We selected TAMARIN PROVER as a modeling language over existing symbolic verification tools, such as PROVERIF, for the following reasons. First, TAMARIN does not use abstractions to overestimate the capabilities of the adversary. This is sometimes desirable, as it may lead to fully automated analysis at the cost of incompleteness (e.g., false attacks may be found). Instead, we take advantage of fine-grained control over the execution model and rely heavily on the interactive analysis, oracle scripts and supporting lemmas to provide fully automated proofs. Second, TAMARIN is able to handle protocols with mutable global state. This feature is essential to

properly model the behavior of SNP-protected guests, as their execution may temporarily be suspended due to swapping.

To facilitate development, we use the m4 general-purpose macro processor. This allows for efficient prototyping (e.g., by disabling certain precomputations) and provides the flexibility to extract multiple variants of the model. Some variants, such as the one obtained with the flag —DIGNORE_ROOT_MD_ENTRY, are intentionally not aligned with the specification; we use them for *sanity checks*, i.e., to confirm the existence of supposed attack traces.

Our model is based on the specification SEV Secure Nested Paging Firmware ABI Specification, revision 1.55, published in September 2023 [16]. AMD recently published the source code of the SEV-SNP Genoa firmware on GitHub [18]. In scenarios where the specification was not clear enough, we consulted with the implementation.

Although AMD has published in the SEV-SNP white paper [15] a set of security threats (properties) that it addresses, we do not consider them in our analysis. This is because they account for a low-level behavior, such as memory aliasing, rather then the specific manner in which the interface is utilized. Since no security properties are defined in the specification either, we have ended up with our own set of properties.

While we believe that the formal attacks we uncovered are consistent with both the specification and the implementation, we want to point out that we have not attempted to perform them on actual hardware. Our model, complete with proofs and documentation is available on Gitlab [20].

A. Entities

The model can be described in terms of five interacting state machines. In particular, we have:

- a state machine that describes the behavior of the SNPprotected Guest (GVM) as depicted in Figure 2a;
- a state machine that describes the behavior of the Guest Owner (GO) as depicted in Figure 2b;
- a state machine that describes the behavior of the Migration Agent (MA) as depicted in Figure 4;
- state machines that describe the behavior of the SNP Firmware (FW): one that responds to GVM requests, and one that launches GVMs as depicted in Figure 3 and Figure 1, respectively.

The state machines offer a somewhat simplified representation of the model wherein FW maintains not one, but two separate states to manage two distinct guest contexts for instance. Nevertheless, we find them useful in demonstrating what can be done, rather then how it is done. The latter will be explained in the subsequent sections.

Our model of the SEV-SNP software interface can be interacted with indefinitely by several entities making calls in an arbitrary order. To support such interaction, we allow an unbounded execution of each *thread*, regardless of the program it executes (i.e., the entity it belongs to). For example, a single MA thread can manage, i.e. be associated with, an unbounded

number of GVMs. If a thread is capable of executing multiple tasks, it can execute them in any order. For example, a GVM thread can execute any sequence of attestation report and key derivation requests. In some cases, the order in which calls are made may result in substantially different traces as they may update a shared state. For example, if the GVM gets swapped out immediately after it receives a derived key, its encrypted page (state) will contain the key.

B. Key Distribution Service

The rules KDSCreateARK and KDSCreateASK model a part of the AMD Key Distribution Service (KDS). This includes the generation of the AMD Root Key, denoted as privARK, and the AMD Signing Key, denoted as privASK. Each long-term key (including privVCEK which is discussed next) is available via the fact !LTK and certified by the next one in the hierarchy, except for privARK, which represents the root of trust and is self-signed. For the purpose of producing and verifying digital signatures, we employ the built-in message theory signing, which exports function symbols sign, verify, pk, true; they are related by the equation verify(sign(m,sk),m,pk(sk)) = true.

We publish the certificate of each long-term key to the network using Out facts, and make the root certificate available to a guest owner (GO) via !Cert(...) fact. Furthermore, we allow the adversary to compromise the KDS and extract the keys using the RevealARK and RevealASK rules.

C. Platform

We model the creation, initialization (SNP_INIT) and configuration (SNP_CONFIG) of the platform by using a single rule called PLCreate. This rule binds a firmware to a specific chip, which is denoted by chipID.

```
rule PLCreate:
let
privVCEK = KDF('VCEK', CEK, $version)
pubVCEK = pk(privVCEK)
data = ('VCEK', askID, $chipID, pubVCEK)
cert = ( data, sign(data, privASK) )
in
[!LTK('ASK', askID, privASK), Fr(CEK)]
-[Uniq(('VCEK', $chipID)),...]
[!LTK('VCEK', $chipID, privVCEK)]
!Cert('VCEK', $chipID, cert), Out(cert),...]
```

The rule uses a freshly generated, chip-unique, long-term secret cek, the value of a public variable \$version (the TCB version of the platform) and a key derivation function KDF to derive an attestation signing key, privVCEK. We model side-channel attacks on its confidentiality with the help of the ExtractCEK rule.

Note that firmware updates (DOWNLOAD_FIRMWARE) are not supported; we represent all of CurrentTcb, ReportedTcb and LaunchTcb by version, whose value (once initialized) persists throughout the lifetime of a chip, and consequently the lifetime of privVCEK.

The per-chip uniqueness of privVCEK is enforced by applying the restriction Unique.

```
restriction Unique:
2 \forall x #i #j. Uniq(x)@i \land Uniq(x)@j \Rightarrow #i = #j
```

It requires that the Uniq action shown above be injective, so the restriction ensures that every instance of the PLCreate rule yields a distinct chipID.

The fact !LTK('VCEK', ...) may be used to start any number of FW threads, all running on the same chip; the relevant FWLaunchesMA rule is outlined in the Initialization subsection.

D. Measurement

A guest image digest, also called a *measurement*, is computed by the guest owner prior to image deployment and afterwards by the firmware during guest launch. In the latter case, the firmware constructs the measurement by initializing a load digest via an SNP_LAUNCH_START call and subsequently updating it with SNP_LAUNCH_UPDATE calls. Each update saves a hash of the PAGE_INFO structure, which transitively binds the contents of an individual page to the digest, and thereby the contents, of all previous pages.

When SNP_LAUNCH_FINISH is called, assuming that ID_BLOCK_EN is set, the firmware checks whether the computed measurement matches the one specified by the guest owner; if it does not, the firmware refuses to launch a guest and returns BAD_MEASUREMENT.

Note the dual nature of a guest image as both a program, which is intended to be executed on a virtual machine, and data that can be measured. In order for us to directly model both the dynamic and static nature of the image, the modelling language would need to support metaprogramming features that would allow us to treat programs as data. To the best of our knowledge, none of the current security protocol verification tools offer such support.

We employ the following approach: to model an image as data, we use a public variable \$image, and public constants '5XPYKIAXFS060' and '3A9B8C7D1E2F' to represent specific images assigned to a guest (GVM) and a migration agent (MA), respectively. To model an image as a program, we use a set of multiset rewriting rules and prefix each rule belonging to the respective images with GVM and MA. By doing so we assign a predefined load digest to each image we consider (e.g., h(('VM_IMAGE', '5XPYKIAXFS060'))).

Whenever a FW launches a GVM or MA, the behavior of the launched virtual machine is governed by a set of multiset rewriting rules which is fixed in advance. Therefore, tampering with the images or launching any other image that would behave differently is not supported. Instead, the images have the ability to publish secrets and the adversary may use this to its advantage.

E. Initialization

Upon startup and before initiating any primary guest launch, FW spawns a Migration Agent (MA) background thread capable of managing guests associated with it. The MA thread should be able to migrate only those guests associated with it during launch, where the association with a particular guest depends on the migration policy of that guest. However, because it is constantly running in the background, it may attempt to initiate migration of any guest at any time and potentially violate attestation report integrity if the guest migration policy

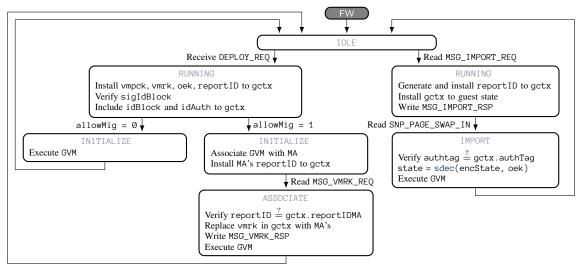


Fig. 1: AMD Security Processor Firmware State Machine (guest launch)

is not properly enforced. We prove that such violations are not possible in our model (AttestationReportIntegrityNoAssocMA).

By allowing a single MA thread to concurrently manage multiple guests, we address scenarios where the adversary may replay the MA thread messages to multiple guests associated with the thread. For instance, the adversary might be able to reuse the MSG_VMRK_REQ message and install the same Virtual Machine Root Key (vmrk) in more than one guest context, thus making each of the guests later derive the same key rooted in vmrk; protecting against such behavior is important because the derived key may be used e.g. for key sealing. We prove that the described behavior is not possible in our model (FreshKeyDerivedFromFWVMRKIsGuestUnique).

The FWLaunchesMA rule is an abstraction of all the SNP_LAUNCH commands. It models the launch of an MA thread by FW. In order to execute the rule, a platform must be created and configured, and there must be a pair of migration agents imageTID and assocImageTID capable of securely migrating guests across platforms.

The MA thread context mctx is initialized with several fresh values: a firmware thread identifier chipTID, which uniquely identifies a FW thread; an attestation report identifier reportID, which binds an attestation report to a specific guest instance; a secret Virtual Machine Platform Communication Key vmpck and the corresponding message counter nonce, which are used to establish a secure communication channel between GVM and FW; and a pointer ptrMA to mctx which is freshly generated upon each mctx update.

```
1 rule FWLaunchesMA:
2 let nonce = N('0')
3     image = '3A9B8C7D1E2F'
4     mctx = ( ~vmpck, ~reportID, nonce, ~ptrMA,... )...
5 in
6 [!LTK('VCEK', chipID, privVCEK)
7, MA(~imageTID, ~assocImageTID,...)
8, Fr(~vmpck), Fr(~reportID), Fr(~chipTID), Fr(~ptrMA) ]
9 -[...] →
10 [!StateFWMA('RUNNING', ~ptrMA, mctx,...)
11, StateMA(~chipTID, 'IDLE', ~vmpck, nonce, ~imageTID,...)
12, VMPCK(~chipTID, ~imageTID, ~vmpck),
13, Out(( ~reportID, ~chipTID )) ]
```

The rule produces the state facts that establish a secure communication channel between FW and MA. More precisely, the persistent fact !StateFWMA(...) represents the part of the FW thread state relevant to launching guests and enforcing migration policies. It contains the MA thread context mctx, which is read and updated during execution (e.g., to increment the message counter). Both that fact and the linear fact StateMA(...) include a state name (e.g., 'IDLE') as a parameter that is updated as the state machine is executed (cf. Figure 4). The rule also outputs a VMPCK(...) fact, which enables the adversary to corrupt and extract the communication key, and publishes all fresh values not meant to be secret to the network.

F. Guest Launch

A simplified guest launch procedure is depicted in Figure 1. An SNP-protected guest (GVM) is launched using the sequence of rules FWInitializesGVM, FWAssociatesGVMWithMA and FWInstallsMASVMRK, or the sequence of rules FWInitializesGVM and FWLaunchGVMNoAssocMA, depending on whether migration is allowed or not (allowMig). The launch procedure has similarities to that of MA; here we only emphasize the differences.

FW reads an image (\$image), an ID Block (idBlock), and an ID Authentication Information Structure (idAuth) from the guest owner (GO) using the FWInitializesGVM rule. An idBlock includes the migration policy (allowMig) for the guest. An idAuth pair consists of an ID block signature (idBlockSig) and the public part of a GO-provided identity key (pubIDK), which

is used to produce the signature. The FW validates the launch digest (as described in Section III-D) and enforces sigunature verification to be successful using the Eq(...) action fact and Equality restriction.

```
restriction Equality:

\forall x y \#i . Eq(x, y)@i \Rightarrow x = y
```

The FW installs in a guest context gctx several (fresh) values, including: an Offline Encryption Key oek, which is used to encrypt contents of guest pages that have been swapped-out; a Virtual Machine Root Key vmrk, which the derived keys may be rooted in; a pointer ptrGVM to a guest state which is freshly generated for each guest state machine transition; a platform identifier pid which persists throughout the lifetime of a guest on a particular platform; and a session identifier sid which is freshly generated upon each guest swap. Additionally, we use the variable isMig as a flag to indicate whether the guest has been migrated or not, and the variable authTag to store the authentication tag of the swapped-out guest state.

The specification mandates that the firmware rejects the pages donated by a hypervisor via an SNP_GCTX_CREATE call if they are not in the Firmware state. Given that a single MA thread may manage multiple GVMs concurrently, each MA message includes a guest context address to distinguish individual guests, as can be seen in the Figure 4. We link a guest context with an address \$gctxAddr and ensure that is not already assigned on the platform (chipID) by applying the Unique restriction; otherwise, the adversary can replay the MSG_VMRK_REQ guest message, prompting FW to install the same VMRK in multiple contexts.

The produced state facts StateGVM(...) and StateFWGVM(...) allow GVM to request remote attestation and key derivation services via a secure communication channel, and FW to provide these services (note that FW uses two separate state facts for execution: StateFWGVM(...) and !StateFWMA(...)). The two states are bound by FW, who simply installs the reportID of MA into the guest context as reportIDMA. Whenever FW receives a request from MA to manage a guest, FW checks whether the reportID in the MA context matches the reportIDMA in the GVM context and refuses to service the request otherwise. The check is performed by the rule FWInstallsMASVMRK, wherein FW receives the vmrk request from the MA and finalizes the guest launch procedure.

G. Guest messages

SNP-protected guests can utilize the <code>SNP_GUEST_REQUEST</code> command to securely communicate with firmware. All exchanged messages are tagged, encrypted and integrity-protected using AES-256-GCM with the Virtual Machine Platform Communication Key (VMPCK) that is injected into guest pages through the <code>SNP_LAUNCH_UPDATE</code> call. Guests, whether primary or migration agents, employ messages for tasks like obtaining attestation reports, deriving keys, managing migration, and other purposes.

We leverage the public constant '0' and built-in multiset theory in TAMARIN to represent nonces. The theory features an associative-commutative union operator +, which suffices to model nonces via monotonically increasing counters. In addition, we employ a unary function symbol N to enforce a nonce type in order to prevent partial deconstructions, i.e. help TAMARIN resolve fact sources.

GVM and MA each maintain their own message counter msgSeqNo, while FW maintains two separate message counters denoted as msgCount — one for each of them. They are initialized to zero (N('0')) and incremented upon each successful message receipt (as in, e.g., FWInstallsMAsVMRK).

H. Guest Context

FW threads use !StateFWMA(...) state facts to store contexts of MA threads for the purpose of guest management. In addition, each FW thread may spawn any number of StateFWGVM(...) state facts used to store the contexts of GVM threads for the purpose of providing services to guests.

Note that we represent each guest management state by a persistent fact; otherwise, if we use a linear fact, then in a rule which involves reading the state, such as FWInitializesGVM, the same fact must appear on both sides of the rule (if the rule needed to update the state instead, the fact on the right-hand side would contain the updated values). This results in non-termination when inductively proving some true statements, such as SupNoKeyNonceReuseMA, where TAMARIN perpetually loops over the rule FWInitializesGVM in an unsuccessful effort to apply the induction hypothesis.

However, using a persistent fact makes it harder to update guest management states (although a new version of the fact can be added to the global state, the old version can never be removed). Updates are required, for example, to increment a message counter when FW services the MSG_VMRK_REQ request via the rule FWInstallsMASVMRK. We can fortunately employ a trick used for this purpose by Cremers et al. [21]: we generate a fresh value that serves as a pointer to the persistent fact, i.e. the memory represented by the fact, which can then be read via Read(...) actions and released via Free(...) actions. We can then simply allocate new memory by generating a new persistent fact and a fresh pointer to it.

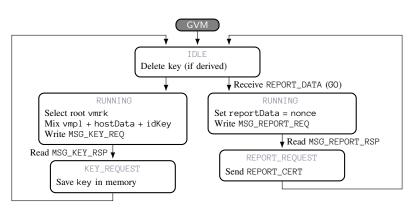
To ensure that memory is always read before it is released and that it cannot be released more than once, we use the respective two restrictions.

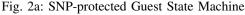
```
restriction FreedMemoryCannotBeRead:
v ptr #i #j. Read(ptr)@i ∧ Free(ptr)@j ⇒ #i < #j
restriction MemoryCanBeFreedOnlyOnce:
v ptr #i #j. Free(ptr)@i ∧ Free(ptr)@j ⇒ #i = #j</pre>
```

I. Remote Attestation

GVM may request, via the rule GVMRequestsReport, any number of attestation reports from FW through the MSG_REPORT_REQ guest message, regardless of whether the guest is migrated or not. Each request includes the reportData field — an arbitrary value provided by the GVM and uninterpreted by the FW. The GO utilizes it to ensure attestation report freshness. This is shown in Figure 2a and Figure 2b.

While the specification accommodates guests with an option to select either VCEK or VLEK as a report signing key, we only ever use VCEK for that purpose (i.e., privVCEK in the model).





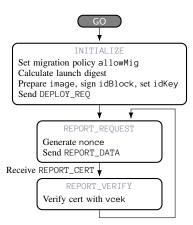


Fig. 2b: Guest Owner State Machine

FW assembles an attestation report, as illustrated in Figure 3, by utilizing the FWGeneratesReport rule. The report comprises various information from the guest context including a chip identifier (chipID), a migration policy (allowMig), and a report identifier (reportID). An attestation report is always signed with privVCEK and forwarded by GVM to the public network where it is available for inspection by GO.

J. Key Derivation

GVM may also ask FW to derive and provide a key by sending the MSG_KEY_REQ guest message, as depicted again in Figure 2a, using the GVMRequestsKey rule. The keys are derived by applying the ternary function symbol KDF to a root key and additional data. This is also shown in Figure 3.

The root key is the VMRK of MA or FW, depending on whether GVM is associated with MA or not.

The acquired key is deleted by GVM before the subsequent request. However, the adversary can gain knowledge of the OEK-encrypted key if it swaps out GVM while it is still in the IDLE state. In fact, the confidentiality of a derived key depends on the confidentiality of five different keys as we will see later.

K. Page Swap

The model incorporates a swapping mechanism where an adversary may use SNP_PAGE_SWAP_OUT and SNP_PAGE_SWAP_IN commands to swap out and swap in GVM any number of times. The corresponding state machine is visualized in Figure 3. We do not explicitly model memory pages of guests or VM Encryption Keys (VEK) used to encrypt them. Instead, we represent GVM memory through state facts StateGVM(...), and we model decryption and encryption with VEK through consumption and production of the StateGVM(...) state fact. FW utilizes four rules FWSwapsOutGVM(BM) and FWSwapsInGVM(BM) to swap out and swap in GVM; employing two almost identical pair of rules for each purpose (BM stands for Before Migration) facilitates our modelling of the migration procedure.

We abstract away from the details of stream cipher encryption with the <code>OEK</code>. Instead, we use the built-in theory <code>symmetric-encryption</code> which defines two binary function symbols <code>senc</code> and <code>sdec</code> related by the equation <code>sdec(senc(m,k),k)=m</code>. Moreover, we introduce an irreducible

binary function symbol mac for the purpose of producing authentication tags.

During the swap-out process, FW encrypts the contents of the GVM state with OEK to obtain the payload ciphertext, which it subsequently uses to calculate the authentication tag authTag. While both payload and authTag are then transmitted over the network, authTag is saved in the GVM context as rootMDEntry. Additionally, StateGVM(...) and StateFWGVM(...) state facts are parameterized with a fresh session identifier ~sid to facilitate backward search. Swapping out GVM is not allowed once it is migrated, as indicated by the isMig flag.

Conversely, during the swap-in process, FW first receives payload and authTag from the network. It then decrypts the GVM state contents with OEK and verifies authTag by comparing it against rootMDEntry using the Equality restriction. If the check succeeds, the GVM thread may continue its execution afterwards. We can demonstrate (as described in Section IV-A) that ignoring this check leads to a rollback attack where the same key and nonce gets reused.

For termination reasons, we prohibit the FW from swapping out a GVM state more then once. Each of the two rules FWSwapsOutGVM(BM) generates a fresh pointer, denoted as ptrGVM, and produces a StateGVM(..., ptrGVM, ...) fact which uniquely identifies a particular GVM state; additionally, the Swap(ptrGVM) action is logged in the trace. To ensure that the state is swapped out at most once, we enforce the following restriction.

```
1 restriction MemoryCanBeSwappedOnlyOnce:
2 ♥ ptr #i #j. Swap(ptr)@i ∧ Swap(ptr)@j ⇒ #i = #j
```

We do not consider this to be a limitation since, in our model, the adversary gains no advantage by obtaining the same ciphertext multiple times.

For performance reasons, we also prohibit swapping out GVM while it is in the RUNNING state by employing the following restriction.

```
restriction GuestStateDuringASwap:

∀ state #i. SwapState(state)@i ⇒ state = 'IDLE' ∨

state = 'KEY_REQ' ∨ state = 'REPORT_REQ'
```

We also do not regard this as a limitation since a GVM IDLE state, which may be swapped out, differs from a RUNNING state

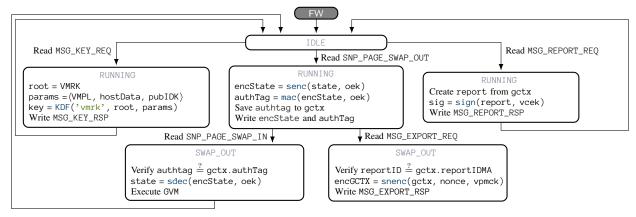


Fig. 3: AMD Security Processor Firmware State Machine (guest management)

only in the deletion of a derived key (modelled by replacing the key with the KDF('NULL', 'NULL', 'NULL') fact).

L. Live Migration

SEV-SNP offers several methods for migrating SNP-protected guests, depending on whether an assistance of a migration agent or a guest is utilized. In our model, we do not consider guest-assisted migration.

The MA state machine is depicted in Figure 4. The MA thread is tasked with providing a VMRK during the launch and ensuring compliance with the guest migration policy. Each migratable GVM, as indicated by the allowMig flag, is assigned to a single MA on a particular platform. Conversely, a single MA thread can manage an arbitrary number of primary guests. MA itself is not migratable.

As mandated by the specification, GVM is swapped out prior to migration using the FWSwapsOutGVMBM rule. Migration is then initiated by MA, which sends the MSG_EXPORT_REQ guest message to FW. Upon receipt of the message via the FWExportsGVM rule, the FW verifies whether the MA thread is assigned to the specific guest by comparing reportID in the context of the MA thread with reportIDMA in the context of GVM. Assuming they are equal, FW responds by transmitting the guest context via MSG_EXPORT_RSP guest message. The exported context encompasses all GVM data except for the reportIDMA, which will be replaced with the reportID of MA on the destination machine.

The specification indicates that the context is sent to a migration agent on the destination machine through a secure channel. However, unlike in previous SEV instances, the specific mechanism by which this transmission is achieved is outside the scope of the specification.

In our model, we assume that each MA thread on the source machine (source MA), denoted by <code>imageTIDMA</code>, is capable of establishing a secure communication channel with an MA thread on the destination machine (target MA), denoted by <code>assocImageTIDMA</code>. This communication channel is modeled via two rules, <code>ComChannelOut</code> and <code>ComChannelIn</code>.

Upon receiving the GVM context from the source MA, the target MA initiates the import procedure by sending the MSG_IMPORT_REQ guest message to the target FW. Subsequently,

the FW utilizes the FWImportsGVM rule to add the guest context to the StateFWGVM(...) state fact and update it with a freshly generated reportID. Additionally, the FW establishes the association between MA and GVM by incorporating the reportID value of the MA into the guest context as reportIDMA. Following this, GVM is swapped in using the FWSwapsInGVMAM rule and launched afterward.

While the GVM policy migrates with its context, the model permits for only one migration. Despite being a formal limitation, it suffices to showcase a certain kind of vulnerability, as we will see later. Moreover, we also prohibit swapping out GVM after it is migrated to keep the verification time manageable.

M. Secure Channel

Migration agents employ ComChannelOut and ComChannelIn rules to enable the secure transfer of guest contexts between platforms. This process involves using ComChan__(...) state facts to link OutChan(...) and InChan(...) state facts, for the purpose of sending and receiving a guest context. The use of state facts ensures that the adversary can neither modify nor learn messages that are sent over the channel. Furthermore, the use of linear facts ensures that the messages sent cannot be replayed at a later point in time.

```
rule ComChannelOut:
2 [ OutChan(~A, ~B, msg) ]
3 —[ ComChanOut(~A, ~B, msg) ]→
4 [ ComChan__(~A, ~B, msg) ]
5
6 rule ComChannelIn:
7 [ ComChan__(~A, ~B, msg) ]
9 —[ ComChanIn(~A, ~B, msg) ]→
9 [ InChan(~A, ~B, msg) ]
```

The <code>-DENABLE_REPLAY_OVER_COMM</code> flag allows extraction of a model variant that permits message replay over the channel. In this model, a persistent fact <code>!ComChan__(...)</code> is used instead of a linear fact <code>ComChan__(...)</code>. We demonstrate how the adversary may exploit this to its advantage and compromise <code>vmpck</code> by utilizing the <code>ExeAttackReplayOverCommChan</code> executable lemma.

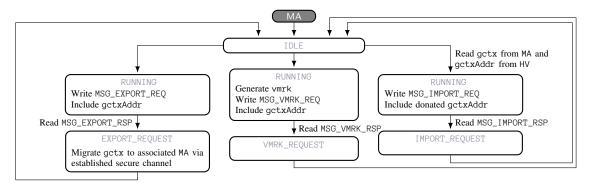


Fig. 4: Migration Agent State Machine

N. Adversarial Model

In accordance with the AMD SEV-SNP threat model, we assume that the adversary has control over a hypervisor and a cloud provider and is able to launch an arbitrary number of SNP-protected VMs, execute the ABI commands in any order, spy on the communication outputs, and tamper with the communication inputs. Moreover, the adversary may corrupt both the SNP firmware and guests, and extract keys from repeated keystreams. The Dolev-Yao adversary of TAMARIN has most of these functionalities built in by default; we only had to manually enable the last two behaviors.

We introduce various Extract and Reveal rules that disclose, respectively, secrets generated during launch and long-term keys to the adversary. For instance, the adversary may use the following ExtractVMPCK rule to corrupt either GVM or MA and extract vmpck.

```
rule ExtractVMPCK:

[ VMPCK(chipTID, imageTID, key) ]

3 —[ CorruptVMPCK(key),...]→

4 [ Out(key) ]
```

We allow the extraction of any key generated during launch, excluding a specific finite set of keys bound to a particular thread identifier (e.g., imageTID). Due to the swapping and migration mechanisms, the confidentiality of a given key may rely on that of other keys, thereby introducing additional potential attack vectors.

Similarly, the adversary may employ the ExtractCEK rule to extract cek from FW, which it might be able to do in practice by using side-channel attacks. It can also reveal the AMD Root Key privARK and the AMD Signing Key privASK.

The Galois/Counter mode stream cipher is extensively utilized within AMD SEV-SNP. However, considering the fact that the guest state may be swapped and the guest context migrated, it is not clear whether this cipher is always correctly employed. We overapproximate a worst-case scenario (in which the adversary might be able to recover part of the plaintext) by letting the adversary recover vmpck from two distinct guest messages, both encrypted with that same key using the same nonce:

Here, we use the Inequality restriction to ensure that messages m1 and m2 are distinct.

```
1 restriction Inequality:
2 \forall x #i. Neq(x,x)@i \Rightarrow \bot
```

Note that we exclusively model the stream cipher for guest messages, even though it is also employed for guest pages.

O. Summary

Our complete TAMARIN model includes 43 rewrite rules and 94 lemmas in total. The model encompasses approximately 5,800 lines of code without comments, with 1,100 lines dedicated solely to the rules. Our model has several constraints:

- (1) only one migration per thread is permitted;
- (2) swapping is not possible after migration;
- (3) the stream cipher is only employed for guest messages;
- (4) firmware updates are not supported;
- (5) key derivation always utilizes VM Root Key;
- (6) we do not model the Versioned Loaded Endorsement Key.

IV. ANALYSIS

We classify the specified properties, also referred to as lemmas, into seven distinct groups: *source*, *executability*, *supporting*, *secrecy*, *authentication*, *attestation*, and *freshness* lemmas. The corresponding counts of lemmas and the analysis time for each group are outlined in Table I.

A. Executability Lemmas

Executability lemmas claim the existence of particularly formed traces and are identified with the "exists—trace" keyword. The use of this keyword instructs TAMARIN to mark the lemma as true if there exists at least one execution which satisfies the underlying formula. This is in contrast with the security lemmas, which must hold for all possible executions.

PROPERTIES	NUMBER	\sim TIME (MIN)	AUTO?
source	1	39	
executability	12	122	
supporting	48	119	
secrecy	13	133	✓
authentication	14	18	
attestation	5	384	
freshness	4	36	
Σ	96	480	

TABLE I: Summary of analyzed properties

We use executability lemmas in two ways. First, we establish the functional correctness of our model by verifying them. Namely, these lemmas encompass a range of potential behaviors within the model. For instance, the ExeMAManagesTwoGVMs lemma specifies that a single MA thread can be associated with two GVMs. In each such lemma we limit the number of rule instances and prohibit any key compromise to reduce verification time.

Second, we prove the existence of attacks in seemingly vulnerable model variants. The rationale behind this is to verify whether certain checks and assumptions are indeed necessary for security properties to hold. Specifically, we employ the m4 flags —DIGNORE_ROOT_MD_ENTRY and —DENABLE_REPLAY_OVER_COMM, respectively, to disable authTag verification during swap in and to enable replaying messages via CommChan rules. Both lead to attacks where the same key and nonce gets reused. We leverage the ExeAttack lemmas, such as ExeAttackSwapInRollbackBM, to demonstrate this.

B. Supporting Lemmas

Due to the complexity of our model, including loops which tend to make unbounded verification challenging, proving most of the security properties directly is not feasible. Therefore, we employ supporting lemmas to

- provide entry points for loops
- enforce termination
- improve verification time

We allow loops to execute an unbounded number of times; this models for instance the perpetual servicing of GVM requests by FW and may lead to non-termination of backward search. We can remedy this by proving that each loop has an entry point, and we do so through inductive reasoning [22]. See for instance the supporting lemma SupFWInitializesGCTX.

We also leverage supporting lemmas to prove other supporting and secrecy lemmas. For instance, we may prove certain invariants that hold for every loop iteration. An example of such a property is the lemma SupGVMMessageCounterMustBeEven, which states that the GVM message counter always has an even value. We utilize proof oracles to guide the TAMARIN proof procedure in a direction where such lemmas can be applied. Note that certain lemmas, such as SupFWGVMResponseIsUnique, are split into several variants to facilitate prototyping.

Most of the supporting lemmas are utilized to prove the following lemma which states that honest agents will never reuse a nonce with the same encryption key (i.e., vmpck).

```
1 rule SupNoKeyRevealFromNonceReuse:
2 ∀ nonce key #i. ReuseNonceKey(nonce, key)@i
3 ⇒ ∃ #j. KU(key)@j ∧ #j < #i</pre>
```

The difficulty arises as TAMARIN considers all possible traces wherein two distinct messages are encrypted with the same vmpck using the same nonce. Moreover, the ability to swap a guest any number of times further complicates matters. The SupNoKeyRevealFromNonceReuse lemma is necessary to prove most of the secrecy lemmas.

The fresh variable pid is used to enforce that StateFWGVM and StateGVM fact symbols have *injective instances* [23]. Our analysis relies on the capability of TAMARIN to under-approximate a set of such symbols. For instance, the following lemma can be proven only if we assume that StateGVM has injective instances.

This lemma states that no guest request handling is possible after the guest has been swapped out, prior to export. Most of the subsequent supporting lemmas, including the previously mentioned SupNoKeyRevealFromNonceReuse, depend on this property.

C. Secrecy Lemmas

We verify the *perfect forward secrecy* of each key that we use within the model, including long-term, generated, and derived keys. In all security properties we permit the adversary to corrupt dishonest agents, and exploit key and nonce reuse of any agent.

We specify security properties so that each accommodates a single guest migration policy, rather than all of them at once. This approach enables us to verify certain properties in the presence of a more powerful adversary. For instance, when analyzing the secrecy of oek, we consider two lemmas: SecGVMOEKIsSecret and SecGVMOEKIsSecretNoAssocMA. These lemmas correspond to scenarios where association with MA is allowed or disallowed, respectively. The lemmas differ in that the latter allows the adversary to corrupt even the MA thread that is running in the background on the same platform.

Key secrecy is usually contingent on several other keys. Take, for instance, Lemma 1a, which specifies the secrecy of the key derived from the vmrk of MA. With this property, we consider the scenario wherein the vmrk is generated by the MA thread imageTIDMA and installed by the FW thread chipTID within the context of the GVM thread imageTID. If we assume that the adversary knows the key derived from vmrk, then it must necessarily be the case that either vmrk, vmpck or oek of imageTID is corrupted, or one of the associated MAS, imageTIDMA or assocImageTIDMA, is corrupted.

Here, the secrecy of the key critically depends on the secrecy of five other keys. To illustrate this, consider the various way in which key could potentially be compromised.

```
1 lemma SecKeyDerivedFromMAVMRKIsSecret:
                                                               1 lemma AuthFWGVMMsgAgreeForAttestAssocMA:
2 ∀ chipTID imageTIDMA assImageTIDMA gctxAddr imageTID
                                                               2 ∀ isMig imageTID chipTID vmpck response imageTIDMA
     vmpck vmrk key info params #i #j #k #l.
                                                                    assImageTIDMA #i #j #k.
     EstablishSecureChan(imageTIDMA + assImageTIDMA)@i
                                                                    EstablishSecureChanMA(imageTIDMA + assImageTIDMA)@i
5 ∧ GenerateVMRK(imageTIDMA, chipTID, gctxAddr, vmrk)@j
                                                               5 ∧ AssociateMAGVM(imageTID, imageTIDMA)@j
6 ∧ InstallVMRK(chipTID, vmpck, imageTIDMA, imageTID,
                                                               6 ∧ FWReceiveGVMRequest('FW_GENERATE_REPORT', isMig,
                 gctxAddr, vmrk)@k
                                                                                     chipTID, imageTID, vmpck, request)@k
8 \( \text{key} = KDF(info, vmrk, params)
                                                               8 \Rightarrow (\exists \#k. (\#k < \#j))
                                                                       ∧ GVMIssueRequest('GVM_REPORT_REQUEST', isMig,
9 ∧ KU(key)@1
                                                               a
10 \Rightarrow (\exists \#m. (m < 1) \land CorruptVMRK(vmrk)@m)
                                                                                           imageTID, chipTID, request)@k )
                                                              11 ∨ (∃ #k. (k < j) ∧ CorruptVMPCK(vmpck)@k)
11 ∨ (∃ #m. (m < 1) ∧ CorruptVMPCK(vmpck)@m)
12 ∨ (∃ #m. (m < 1) ∧ CorruptImageOEK(imageTID)@m)
                                                                 ∨ (∃ #k. (k < j) ∧ CorruptImageOEK(imageTID)@k)</pre>
                                                              13 ∨ (∃ #k. (k < j) ∧ CorruptImageVMPCK(imageTIDMA)@k)
13 ∨ (∃ #m. (m < 1) ∧ CorruptImageVMPCK(imageTIDMA)@m)
14 ∨ (∃ #m. (m < 1) ∧ CorruptImageVMPCK(assImageTIDMA)@m)
                                                             14 ∨ (∃ #k. (k < j) ∧ CorruptImageVMPCK(assImageTIDMA)@k)
```

Lemma 1a: Secrecy of keys derived from VMRK of FW

Lemma 1b: Agreement on the MSG_REPORT_REQ guest message

First, if the adversary corrupts the vmrk, it can clearly construct the key independently. Second, the adversary can obtain the key by compromising the vmpck of the GVM thread and decrypting the guest message MSG_KEY_RSP. Third, the adversary can swap out the state of the GVM thread and acquire either the vmpck or key directly (assuming it has not 8 \(\lambda\) digestIDK = h(\(\lambda\)' ID_KEY', pubIDK \(\rangla\) been deleted and the adversary possesses the corresponding oek). Finally, corrupting the MA thread on either the source $10^{-10} \stackrel{10}{\rightarrow} 3$ chipTID isMig vmpck imageTID msgSeqNo #m. or destination platform lets the adversary obtain vmrk from a 12 MSG_EXPORT_REQ or MSG_IMPORT_REQ guest message, respectively. 13

Note that most of the specified secrecy properties can 14 VIIIPCK, IN(IIISGSGGIRG), 150505 SECRETARY (IIISGSGGIRG), 150505 SEC additionally be viewed as supporting lemmas, because they 16 ∨ (∃ #k. (k < j) ∧ CorruptImageOEK(imageTID)@k) facilitate the verification of authentication, attestation, fresh- 17 \vee ($\frac{3}{2}$ #k. (k < j) \wedge CorruptImageVMPCK(imageTIDMA)@k) ness, and other secrecy lemmas.

D. Authentication Lemmas

We utilize authentication lemmas to verify whether the firmware and guest, either GVM or MA, agree on the messages exchanged. Specifically, we are interested in whether the messages received by the guest, running on a particular platform, indeed originate from the firmware on that same platform, and vice-versa.

Take for instance Lemma 1b, which specifies agreement on the guest message MSG_REPORT_REQ. It considers a GVM thread with the thread identifier imageTID, launched by a FW thread with the thread identifier chipTID, under a policy that permits migration. Here, we want to verify whether the attestation report request the FW obtains was indeed issued by the GVM after launch, assuming both honest agents are uncompromised.

E. Attestation Lemmas

The attestation lemmas specify authenticity and integrity properties of attestation reports. Both kinds of property consider the scenario where GO verifies an attestation report using the signing key that apparently belongs to a particular chip. The authenticity properties then affirm the existence of a FW thread, operating on that chip, which previously generated the same report. In contrast, the integrity properties are concerned with the state of the guest bound to that report, asserting that the guest is indeed running on the designated platform, with the correct policy and configuration. Lemma 2 is an example of such a property.

```
1 lemma AttestationReportStrongIntegrityAssocMA:
 2 ∀ ownerID ownerTID privIDK report imageID allowMig
     reportData ld digestIDK pubIDK reportID reportIDMA
     chipID hostData #1.
     GOVerifiesReport(ownerID, ownerTID, privIDK, report)@l
 \wedge report = \langle imageID, allowMig, reportData, ld, digestIDK,
                reportID, reportIDMA, chipID, hostData>
9 ∧ pubIDK
                = pk(privIDK)
                = h(\langle 'VM_IMAGE', image \rangle)
        (#m < #1)
     ∧ GVMReportReq(chipID, chipTID, imageID, imageTID,
         vmpck, N(msgSeqNo), reportData, isMig)@m
18 \vee (\exists #k. (k \prec j) \wedge CorruptImageVMPCK(assocImageTIDMA)@k)
```

Lemma 2: Strong Integrity of Attestation Report

F. Freshness Lemmas

The freshness lemmas comprise several *uniqueness* lemmas. In particular, these lemmas allow us to determine whether two separate GVM threads can acquire the same derived key, whether GO can verify outdated certificates, or if the launch procedure can be manipulated to enforce FW to install the same vmrk into multiple guests; the latter is shown as Lemma 3.

As we mentioned previously, guests have the option, via MSG_KEY_REQ, to choose the root key-VCEK, VLEK, or VMRK—and provide additional data for key derivation. Opting for VMRK as the root key should ensure that each guest instance will obtain a different key; see lemma FreshKeyDerivedFromFWVMRKIsGuestUnique. We note here, however, that it is not possible for a guest to provide a random sequence of bytes to be included in SEV-SNP key derivation (so unlike in Intel SGX, key wear-out protection is not supported).

V. RESULTS AND DISCUSSION

The analysis of the model was conducted using Debian 11, running on an AMD EPYC 7713 processor, equipped with 16 cores and 32 threads, with a 2.0GHz base clock. The system is complemented by 256GB of memory. We employed TAMARIN version 1.9.0 for the analysis.

Lemma 3: Freshness of Attestation Report

All of the specified properties were automatically analyzed in about 8 hours, as detailed in Table I. Throughout the analysis, TAMARIN utilized no more than 32GB of memory.

We analyzed a total of 96 properties, including 36 security properties. The summary of the results is presented in Table II. All but five security properties were successfully verified.

The analysis of the model was conducted using Debian 11, running on an AMD EPYC 7713 processor, equipped with 16 cores and 32 threads, with a 2.0GHz base clock. The system is complemented by 256GB of memory. We employed TAMARIN version 1.9.0 for the analysis.

All of the specified properties were automatically analyzed in about 8 hours, as detailed in Table I. Throughout the analysis, TAMARIN utilized no more than 32GB of memory.

In this section, we discuss the results in detail and suggest a mitigation for one of the discovered weaknesses.

A. Positive Results

On a positive note, all secrecy and freshness properties, and the majority of authentication and attestation properties, have been successfully proven. The analysis was done considering an unbounded number of sessions and an adversary capable of corrupting the keys of each dishonest agent. The verification was heavily facilitated through supporting lemmas and proof strategies based on the oracle rankings. Both of them are utilized to either enforce termination or enhance overall verification time. Furthermore, many security properties were specified in a sufficiently granular manner to be very useful in proving other security properties.

A substantial amount of work went into proving the lemma FreshNoKeyNonceReuse (also called SupNoKeyNonceReuse); the majority of the supporting lemmas were specifically leveraged for this purpose. The main reason for this is because we permit for the GVM state to be swapped out and swapped in unbounded number of times. Consequently, we had to prove that the FW and GVM message counters are synchronised with each swap operation, i.e. they are either equal or the FW counter exceeds the GVM counter by two. The corresponding lemmas are prefixed with SupFWAndGVMMsgCountersAreInSync.

As previously mentioned, we opted to overapproximate the adversary capabilities by assuming a remarkably weak stream cipher — one where the reuse of key and nonce compromises the former. If, by any chance, the lemma FreshNoKeyNonceReuse did not hold, we would consider a more realistic scenario, such as the one where the described reuse allows the adversary to recover the message plaintexts rather than the key.

PROPERTIES	LEMMA	MODEL
	SecMAVMPCKIsSecret	/
	SecGVMOEKIsSecretNoAssocMA	/
	SecGVMOEKIsSecret	/
	SecGVMVMPCKIsSecretNoAssocMA	/
	SecGVMVMPCKIsSecret	/
0	SecMAVMRKIsSecret	/
Secrecy	SecFWVMRKIsSecret	✓
	SecKeyDerivedFromMAVMRKIsSecret	/
	SecKeyDerivedFromFWVMRKIsSecret	✓
	SecARKIsSecret	✓
	SecASKIsSecret	✓
	SecCEKIsSecret	✓
	SecVCEKIsSecret	✓
	AuthFWGVMMsgAgreeForAttestNoAssocMA	✓
	AuthFWGVMMsgAgreeForAttestAssocMA	Х
	AuthFWGVMMsgAgreeForKeyDerNoAssocMA	✓
	AuthFWGVMMsgAgreeForKeyDerAssocMA	Х
	AuthGVMFWMsgAgreeForAttestNoAssocMA	✓
	AuthGVMFWMsgAgreeForAttestAssocMA	X
Authentication	AuthGVMFWMsgAgreeForKeyDerNoAssocMA	✓
	AuthGVMFWMsgAgreeForKeyDerAssocMA	X
	AuthFWMAMsgAgreeForVMRK	✓
	AuthFWMAMsgAgreeForExport	✓
	AuthFWMAMsgAgreeForImport	✓
	AuthMAFWMsgAgreeForVMRK	✓
	AuthMAFWMsgAgreeForExport	✓
	AuthMAFWMsgAgreeForImport	✓
Attestation	AttestationReportAuthenticityNoAssocMA	✓
	AttestationReportAuthenticityAssocMA	✓
	AttestationReportIntegrityNoAssocMA	✓
	AttestationReportWeakIntegrityAssocMA	✓
	AttestationReportStrongIntegrityAssocMA	X
	E IN K N D	,
	FreshNoKeyNonceReuse	√
Freshness	FreshAttestationReportFreshness	/
	FreshMAVMRKInstallationIsUnique	✓
	FreshKeyDerivedFromFWVMRKIsGuestUnique	✓

TABLE II: A summary of the analyzed security properties.

B. Platform Confusion Attacks

We identified five *formal attacks* in our model; four on authentication properties and one on an attestation property. We would like to emphasize that we have not attempted to execute them in practice.

The authentication attacks stem from exploiting the platform-agnostic nature of guest messages. These messages lack any inherent binding to a specific platform with even vmpck as it gets reused across platforms. In the process of migration, the guest context is reinstalled, and guest pages are seamlessly swapped in, allowing the guest to resume operations as usual. Therefore, there is a possibility that guest messages, processed by both the guest and firmware on the current platform, may actually originate from the previous platform.

Consider the GVM that requests a service via either MSG_REPORT_REQ or MSG_KEY_REQ, and waits for a response on the source platform. Two distinct authentication attacks per type of request may arise, depending on whether the FW receives the request and replies with a response, or it does not receive the request at all.

In the former scenario, the adversary migrates the GVM to a destination platform without first forwarding the response to it. Consequently, the GVM, upon resuming its execution, remains in a state where it waits for a response. It is reasonable to expect that guests will typically only accept messages originating from the firmware operating on the platform where they are currently running. However, if the adversary now forwards the response message from the FW on the source platform to the GVM, it will accept it.

The steps to reproduce the attacks in the latter scenario are similar and can be seen in Figure 5; the adversary forwards the initial GVM request from the source platform to the FW on the destination platform. Consequently, the FW incorrectly believes that the GVM sent the message while it was executing on the (current from the perspective of the FW) destination platform, whereas in reality the GVM was executing on the source platform at the time. This makes Lemma 1b false as there is no agreement on the value of chipTID in general; to find an attack trace faster, we add several preconditions to the lemma that coarsely restrict the set of considered traces.

The adversary can leverage the previous attack to compromise the integrity of attestation reports. In particular, when GO reads the report to be verified, it may falsely believe that, at the time the report was requested, the GVM was executing on a destination platform, when, in reality, the GVM was executing on the source platform. As a consequence, Lemma 2 is false because chipID can differ from the anticipated value.

C. Practical Implications and Countermeasures

Ensuring that the guest consistently operates on secure platforms is a critical concern, at least for the guest owner. Nevertheless, issues arise as platforms may become insecure over the time; there may be instances where the current platform or the one eligible for migration hosts vulnerable firmware that has not been security-patched. In other words, a malicious cloud provider may execute the attacks described in the previous subsection to successfully deceive a guest owner or a third party into believing that a guest is executing in a secure environment, when, in reality, it is not.

While considering changes to the model that would prevent the described platform-confusion attacks, we decided against changes that appear to translate poorly to practice. For example, making guest messages platform-specific by employing a new vmpck would require modifications to guest pages upon each guest import. However, it is unclear to us how to do this while preserving seamless migration, without having to shut down the guest.

Note that the described issue persists regardless of Lemma 2: a malicious cloud provider could migrate the guest back and forth between platforms every time an attestation report is requested by a third party, such as the guest owner. Specifically, the cloud provider could run the guest on a vulnerable platform, but obtain attestation reports by briefly migrating the guest to the secure platform, thus creating an illusion that the guest is consistently managed by secure, upto-date firmware.

To fix this problem of back-and-forth migration, we recommend adding a migration-enabled field to both the guest context and the ATTESTATION_REPORT structure. It is a one-bit field that is set to 0 at guest launch, and updated to 1

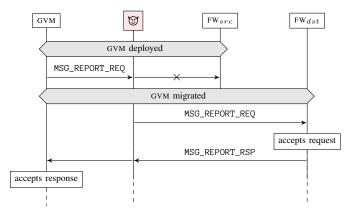


Fig. 5: Platform Confusion Attack: FW_{dst} erroneously believes that the message MSG_REPORT_REQ originated from the GVM while it was executing on the same (destination) platform. In reality, the GVM was executing on the previous (source) platform at that time.

during import if CurrentTcb is strictly less than the migrated LaunchTcb. This allows a third party to find out if, at some point in time after it had verified the last attestation report, a guest was run on a machine whose firmware is not up to date.

VI. RELATED WORK

There is a vast body of literature devoted to TEEs in general and AMD SEV technologies in particular, including more than a hundred works that mention SEV-SNP last year alone (mostly comprising overviews, side-channel and fault injection attacks, technologies building upon/using/replacing SEV-SNP, etc.).

However, to the best of our knowledge, there has been no attempt to build a comprehensive symbolic model and analyze the security of SEV-SNP, including its firmware ABI. Some previous works in formal verification do cover pre-SNP solutions by AMD, such as Antonino et al. [24], or systems that were built around TEEs, such as one by Cremers et al. [25] whose principal focus is to ensure both post-compromise security and the unlinkability of user secrets.

Some oversights underlying SEV and SEV-ES, which are the initial 2016 version and the subsequent 2017 version of AMD SEV technologies, have been identified in a smaller number of works, including [8, 9, 10]. While it seems that AMD has improved their previous designs and implementations with SEV-SNP, some issues still remain [11], and new ones have been identified since the inception of SEV-SNP in 2020. For instance, in a CCS paper [12] from 2021, Buhren et al. show how an insider attacker can derive valid endorsement keys for arbitrary firmware versions. In an S&P paper [13] from 2022, Li et al. describe how in SEV-SNP leaking register values of a VM breaks constant-time cryptographic implementations.

VII. CONCLUSION

AMD SEV-SNP is the latest in a line of AMD hardwarebased TEE technologies, which is intended to improve on the previous SEV and SEV-ES designs by providing stronger protection of virtual machines from malicious hypervisors.

In this paper, we developed the first, comprehensive formal model of the AMD SEV-SNP software interface. The model covers the guest launch, remote attestation, key derivation, page swap and live migration features.

With the help of TAMARIN, we produced automated formal proofs for the most important secrecy, authenticity, attestation, and freshness properties, including the proof of correct stream cipher usage.

Additionally, we identified weaknesses in the design that enable platform confusion attacks. We have shown that some of these attacks are a direct consequence of the platformagnostic nature of guest messages, and that they violate desirable authentication and attestation properties, including attestation report integrity.

Moreover, we discovered that platform confusion attacks are possible even if guest messages can not be received on a different platform; a guest might be run on a machine with unpatched firmware, and occasionally be migrated to a secure platform just to request and send a misleading attestation report.

We discussed the practical security implications if such back-and-forth guest migrations go undetected; a malicious cloud provider could trick a third party into mistakenly believing that a guest is executing in a secure environment. We suggested a way to enable third parties to monitor such unwanted behavior through slightly augmented attestation reports.

ACKNOWLEDGMENTS

This work has been supported by the European Union through the European Regional Development Fund, under the grant KK.01.1.1.01.0009 (DATACROSS), and the project AutoDataLog, a cooperation between the Faculty of Science and AVL-AST d.o.o. Croatia.

REFERENCES

- "Building a Secure System using TrustZone Technology," white paper, https://documentation-service.arm.com/static/5f212796500e883a b8e74531, ARM Limited, 2005.
- [2] "Overview on Signing and Whitelisting for Intel® Software Guard Extension (Intel® SGX) Enclaves," white paper, https://www.intel.com/content/dam/develop/external/us/en/documents/ overview-signing-whitelisting-intel-sgx-enclaves.pdf, Intel Corporation, 2018
- [3] "Secure Encrypted Virtualization API Version 0.24 (Revision 3.24)," technical preview, https://www.amd.com/content/dam/amd/en/docume nts/epyc-technical-docs/programmer-references/55766_SEV-KM_API _Specification.pdf, Advanced Micro Devices, Inc., 2020.
- [4] D. Kaplan, "Protecting VM Register State With SEV-ES," white paper, https://www.amd.com/content/dam/amd/en/documents/epyc-businessdocs/white-papers/Protecting-VM-Register-State-with-SEV-ES.pdf, 2017.
- [5] "Azure Confidential VM options," https://learn.microsoft.com/en-us/azu re/confidential-computing/virtual-machine-solutions, Microsoft Corporation.
- [6] "User guide for Linux Instances: AMD SEV-SNP," https://docs.aws.ama zon.com/AWSEC2/latest/UserGuide/sev-snp.html, Amazon.com, Inc.
- [7] "Oh SNP! VMs get even more confidential," https://cloud.google.com/blog/products/identity-security/rsa-snp-vm-more-confidential, Google LLC.
- [8] M. Li, Y. Zhang, Z. Lin, and Y. Solihin, "Exploiting unprotected I/O operations in AMD's secure encrypted virtualization," in 28th USENIX Security Symposium (USENIX Security 19). Santa Clara, CA: USENIX

- Association, Aug. 2019, pp. 1257–1272. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/li-mengyuan
- [9] M. Li, Y. Zhang, and Z. Lin, "CrossLine: Breaking "Security-by-Crash" Based Memory Isolation in AMD SEV," in *Proceedings of the* 2021 ACM SIGSAC Conference on Computer and Communications Security, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, Nov. 2021, p. 2937–2950. [Online]. Available: https://doi.org/10.1145/3460120.3485253
- [10] M. Li, Y. Zhang, H. Wang, K. Li, and Y. Cheng, "TLB Poisoning Attacks on AMD Secure Encrypted Virtualization," in *Annual Computer Security Applications Conference*, ser. ACSAC '21. New York, NY, USA: Association for Computing Machinery, Dec. 2021, p. 609–619. [Online]. Available: https://doi.org/10.1145/3485832.3485876
- [11] ——, "CIPHERLEAKS: Breaking Constant-time Cryptography on AMD SEV via the Ciphertext Side Channel," in 30th USENIX Security Symposium (USENIX Security 21). USENIX Association, Aug. 2021, pp. 717–732. [Online]. Available: https://www.usenix.org/conference/ usenixsecurity21/presentation/li-mengyuan
- [12] R. Buhren, H.-N. Jacob, T. Krachenfels, and J.-P. Seifert, "One glitch to rule them all: Fault injection attacks against amd's secure encrypted virtualization," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, Nov. 2021, p. 2875–2889. [Online]. Available: https://doi.org/10.1145/3460120.3484779
- [13] M. Li, L. Wilke, J. Wichelmann, T. Eisenbarth, R. Teodorescu, and Y. Zhang, "A Systematic Look at Ciphertext Side Channels on AMD SEV-SNP," in 2022 IEEE Symposium on Security and Privacy (SP), Jul. 2022, pp. 337–351.
- [14] "AMD Secure Processor for Confidential Computing: Security Review," technical Report, https://googleprojectzero.blogspot.com/2022/05/releas e-of-technical-report-into-amd.html, Google Project Zero and Google Cloud Security, 2022.
- [15] "Strengthening VM isolation with integrity protection and more," white paper, https://www.amd.com/system/files/TechDocs/56860.pdf, Advanced Micro Devices, Inc., 2020.
- [16] "SEV Secure Nested Paging Firmware ABI Specification (Revision 1.55)," https://www.amd.com/en/support/tech-docs/sev-secure-nested-paging-firmware-abi-specification, Advanced Micro Devices, Inc., Sep. 2023.
- [17] Linus Torvalds, "Linux 5.19-rc1," https://lore.kernel.org/lkml/CAHk-= wgZt-YDSKfdyES2p6A_KJoG8DwQ0mb9CeS8jZYp+0Y2Rw@mail.g mail.com/T/#u, 2022.
- [18] "AMD-ASPFW," https://github.com/amd/AMD-ASPFW, 2023.
- [19] S. Meier, B. Schmidt, C. Cremers, and D. Basin, "The TAMARIN prover for the symbolic analysis of security protocols," in *Computer Aided* Verification: 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings 25. Springer, 2013, pp. 696–701.
- [20] P. Paradzik, A. Derek, and M. Horvat, "Formal Security Analysis of the AMD SEV-SNP Software Interface," 2023. [Online]. Available: https://gitlab.com/sev-snp-abi-security-analysis/sev-snp-abi-security-analysis
- [21] C. Cremers, B. Kiesl, and N. Medinger, "A formal analysis of IEEE 802.11's WPA2: Countering the kracks caused by cracking the counters," in 29th USENIX Security Symposium (USENIX Security 20). USENIX Association, Aug. 2020, pp. 1–17. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/cremers
- [22] S. Meier, "Advancing automated security protocol verification," 2013, doctoral thesis, ETH Zurich, Switzerland.
- [23] Tamarin-Prover Manual, The Tamarin Team. [Online]. Available: https://tamarin-prover.com/manual/develop/tex/tamarin-manual.pdf
- [24] P. Antonino, A. Derek, and W. A. Woloszyn, "Flexible Remote Attestation of Pre-SNP SEV VMs Using SGX Enclaves," *IEEE Access*, vol. 11, pp. 90839–90856, Aug. 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3308850
- [25] C. Cremers, C. Jacomme, and E. Ronen, "Tokenweaver: Privacy preserving and post-compromise secure attestation," Cryptology ePrint Archive, Paper 2022/1691, Jun. 2022, https://eprint.iacr.org/2022/1691. [Online]. Available: https://eprint.iacr.org/2022/1691
- [26] "Tamarin prover GitHub repository," https://github.com/tamarin-prover/ tamarin-prover, 2023.
- [27] B. Schmidt, "Formal analysis of key exchange protocols and physical protocols," 2012, doctoral thesis, ETH Zurich, Switzerland.
- [28] P. Maene, J. Götzfried, R. de Clercq, T. Müller, F. Freiling, and I. Verbauwhede, "Hardware-based trusted computing architectures for isolation and attestation," *IEEE Transactions on Computers*, vol. 67, no. 3, pp. 361–374, Jan. 2018.
- [29] C. Jacomme, S. Kremer, and G. Scerri, "Symbolic models for isolated execution environments," in 2017 IEEE European Symposium on Secu-



Petar Paradžik received an MSc in Computer Science and Mathematics from the Department of Mathematics, Faculty of Science, University of Zagreb. He is currently a PhD student and teaching assistant at the Faculty of Electrical Engineering and Computing in Zagreb. His research interests include formal methods, automated verification, and applied cryptography.



Ante Derek (Member, IEEE) is currently an Assistant Professor with the Faculty of Electrical Engineering and Computing, University of Zagreb. He participates in a number of national and EUfunded projects in the area of computer security. His research interests include the area of applying formal methods to problems in computer security, privacy, and cryptography.



Marko Horvat received a DPhil in Computer Science from the University of Oxford, UK. He is currently working as Assistant Professor at the Department of Mathematics, Faculty of Science, University of Zagreb, Croatia. His research interests range from formal verification of security protocols to computable analysis and topology.