AMD SEV Overview

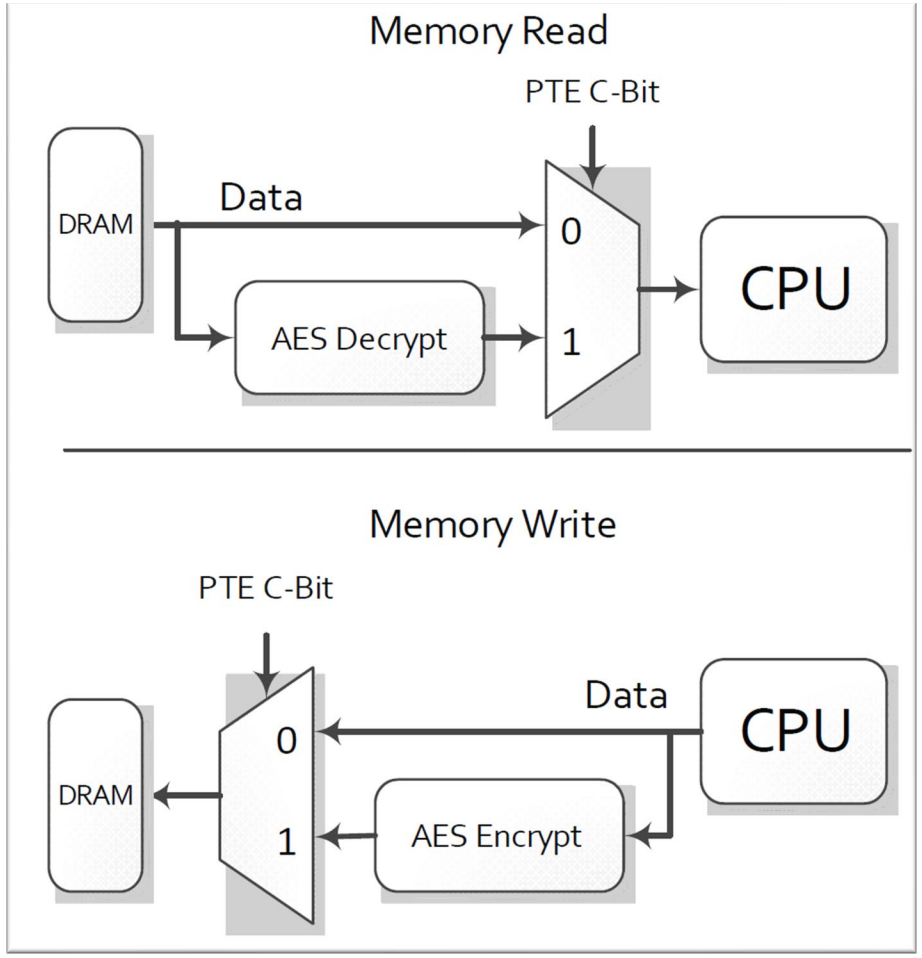
秦浩翔

2021.10.29

Outlines

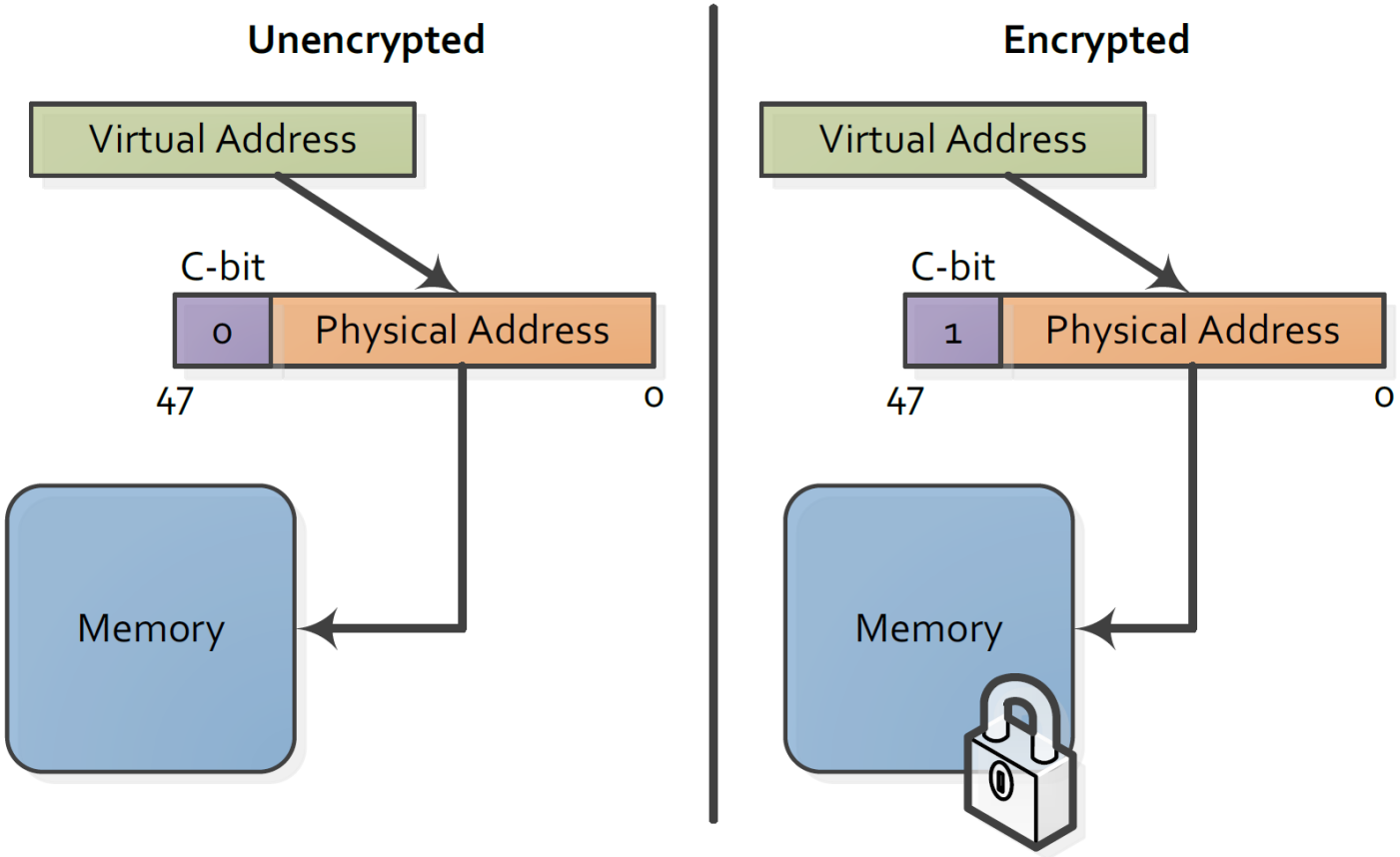
- AMD SEV
  - SME & SEV
  - SEV-ES
  - SEV-SNP
- Vulnerabilities
  - Unencrypted VMCB
  - SEV I/O Operation
  - Lack of Memory Integrity Protection
  - CrossLine: ASID Abuse
  - CipherLeak: Ciphertext Side Channel
- Discussion

Secure Memory Encryption, SME



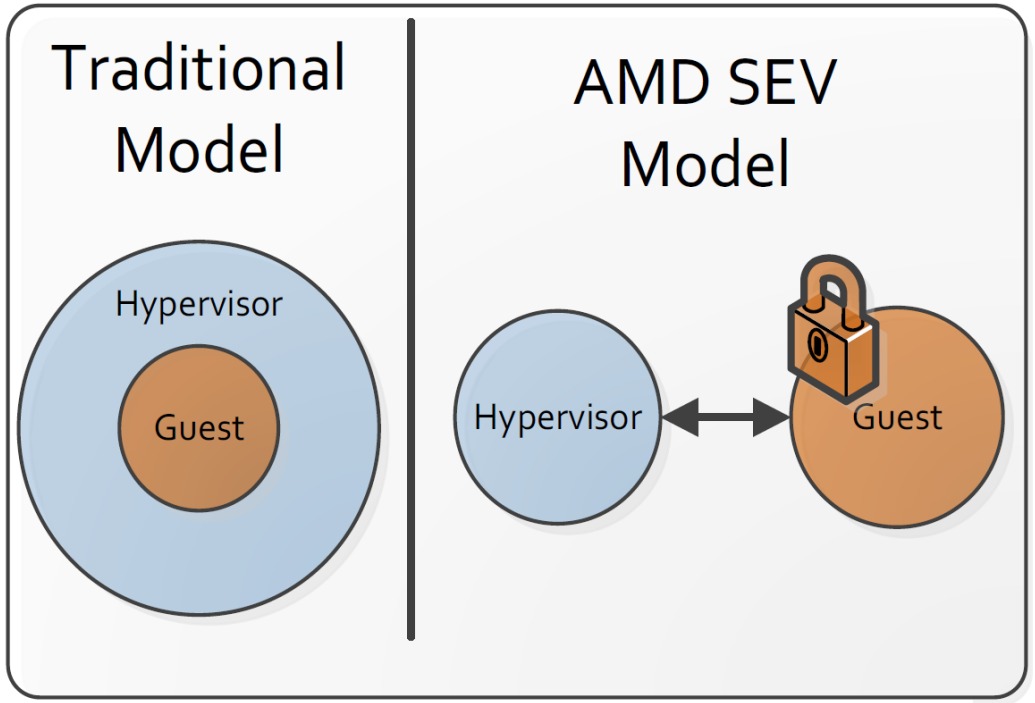
- 内存控制器中包含 AES 引擎
- 写入内存时加密，读内存时解密
- 密钥由 AMD-SP 安全处理器管理
  - 32-bit ARM Cortex A5
- 通过页表控制哪些内存页加密 (bit 47)

Secure Memory Encryption, SME



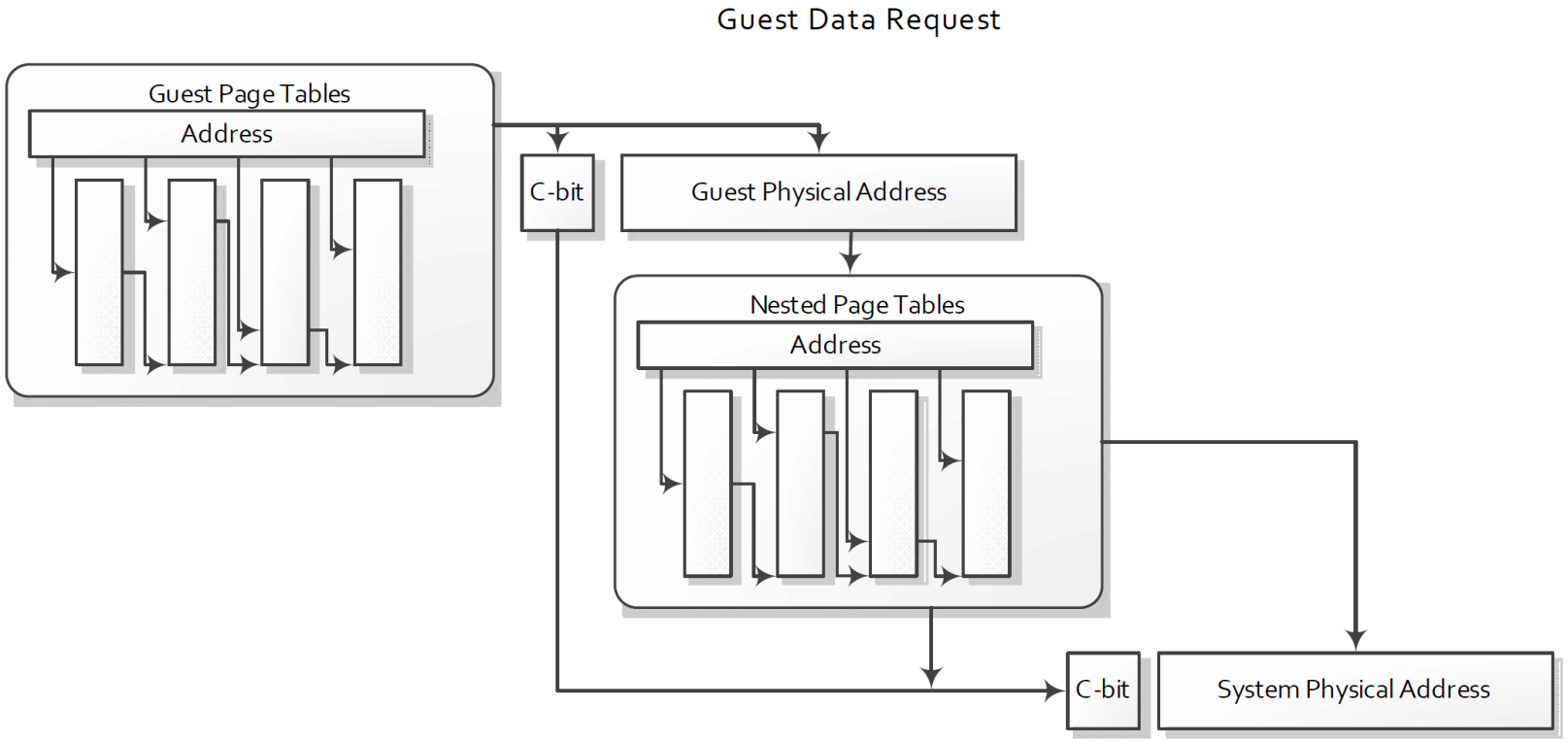
- 使用模式
  - 全加密、部分加密、透明加密
- 支持 DMA
  - 设备向加密内存发送 DMA 请求也要将 C-bit 置位
- 抵御 cold boot 等物理访问攻击

Secure Encrypted Virtualization, SEV



- AMD-V 与 SME 结合
- 使用 ASID 标记代码和数据属于哪个 VM
- 使用 ASID 索引加密密钥
- 同一 SEV VM 的所有 VCPU 分配相同的 ASID，每个 SEV VM 对应一个加密密钥
- 密钥由 AMD-SP 管理

Secure Encrypted Virtualization, SEV



- Guest 通过设置 gPT 中的 C-bit 控制内存页的加密状态
- 代码和页表所在页强制加密
- gPA 中的 C-bit 被保存并用于嵌套地址转换后的 sPA
- Guest 设置为共享的内存使用 hypervisor 的密钥加密

Secure Encrypted Virtualization, SEV

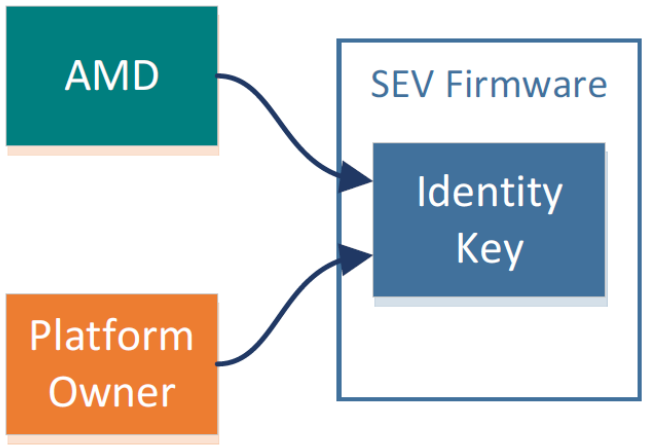


Figure 11: Authenticating the Firmware

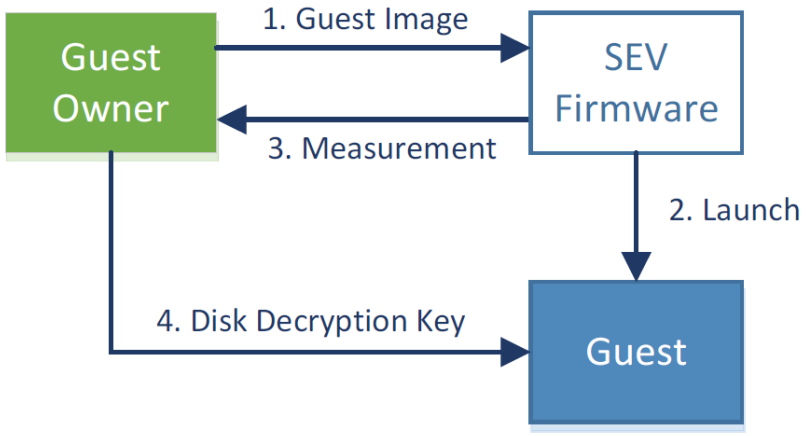


Figure 12: Guest Attestation Example

- 运行在 AMD-SP 中的 SEV 固件向 HV (hypervisor) 提供密钥管理接口，HV 调用接口实现 guest 启动、运行、快照和迁移等
- SEV 提供的安全属性
  - 平台身份验证
  - 已启动 guest 的认证
  - Guest 数据机密性

Uencrypted VMCB

- SEV 对内存加密而未加密 VMCB（虚拟机控制块）
- VM Exit 时，HV (hypervisor) 可以读写 VMCB，提取或控制 VM 状态
  - 任意读写内存

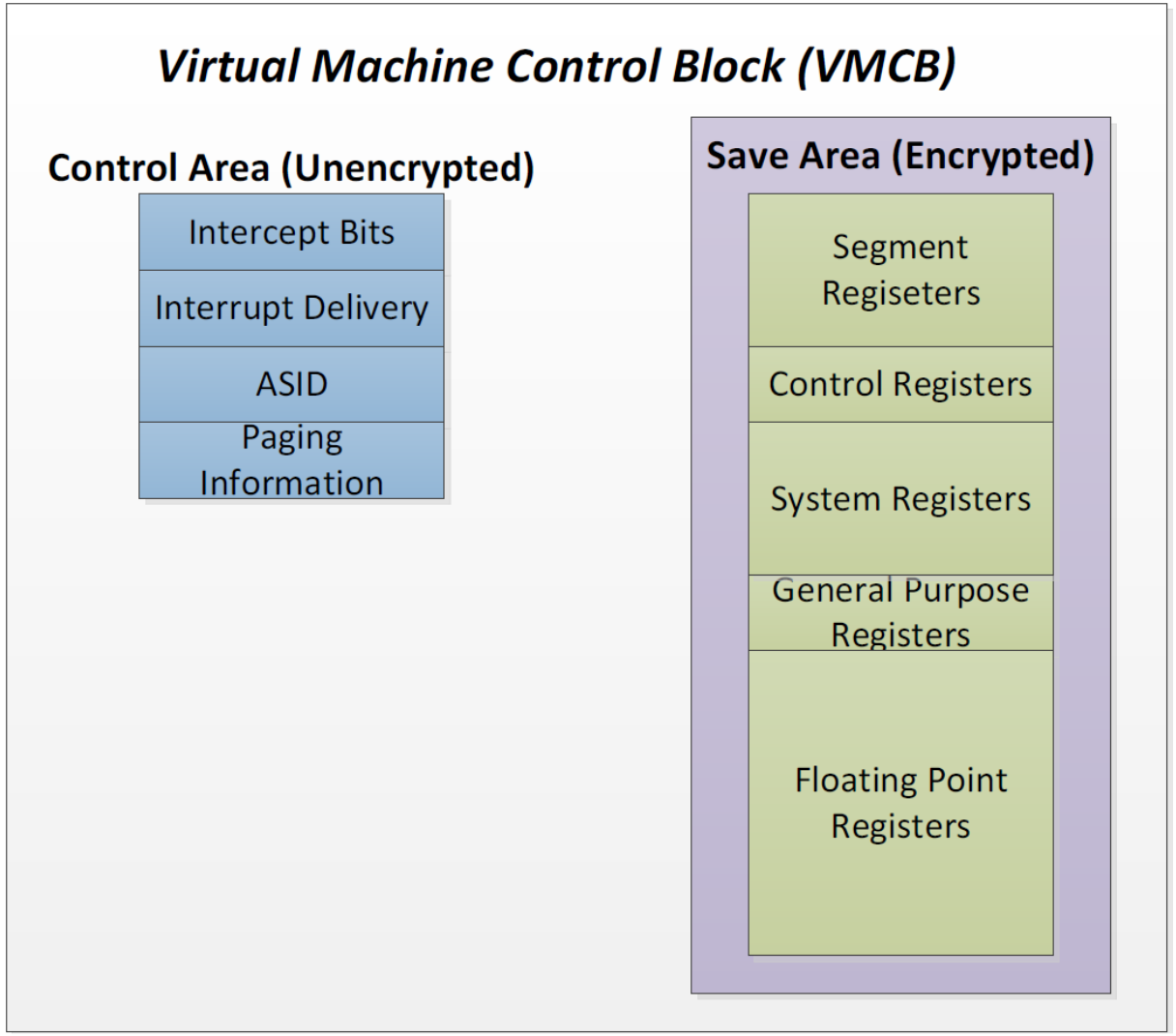
```
mov edi, dword ptr [rbx]
hlt
```

  - 修改 guest 页表，关闭内存加密保护

Felicitas Hetzelt, Robert Buhren, Security Analysis of Encrypted Virtual Machines, VEE 2017.

SEV with Encrypted State, SEV-ES

- 将 VMCB 分为两部分，用于保存寄存器状态的 VMSA 在 VM Exit 时会加密保存并提供完整性保护
- VMSA 指针存在 VMCB 中，物理地址
- AMD-V 中，保存和恢复 VM 寄存器状态分几步完成
  - VMRUN 将控制权转移给 HV，只加载部分状态
  - VMLOAD 恢复额外的寄存器
  - 可能还需要使用 XRSTOR 恢复浮点寄存器状态
- SEV-ES 将以上操作整合到 VMRUN 中，还会校验 VMSA 完整性



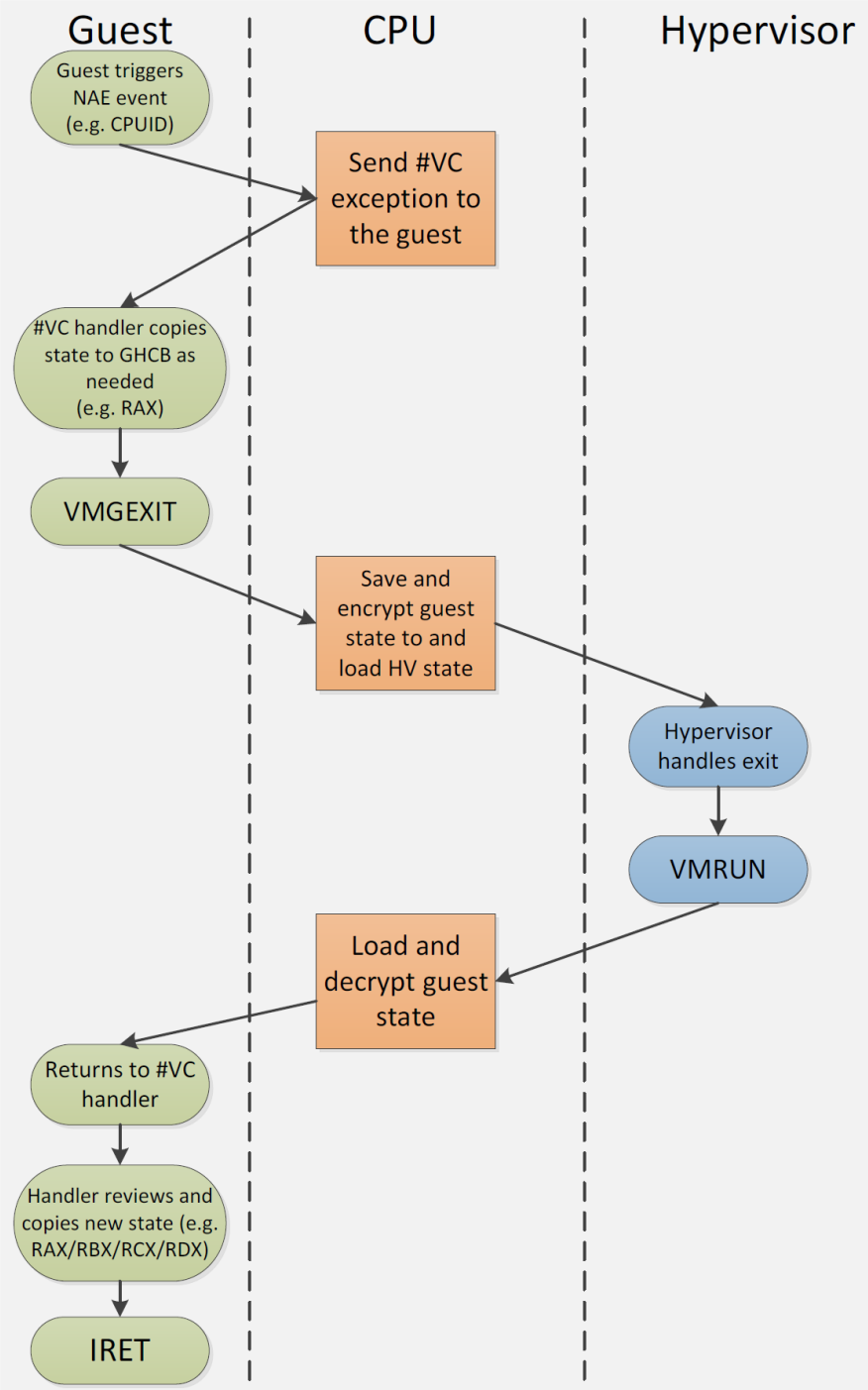
SEV with Encrypted State, SEV-ES

- 将 VM Exit 分为两类
  - Automatic Exits, AE
  - Non-Automatic Exits, NAE
- AE 事件触发后，控制权交给 HV，硬件加密保存 VMSA，HV 处理后执行 VMRUN 恢复 guest
- NAE 事件需要 HV 提供模拟支持

Code	Name	Notes
52h	VMEXIT_MC	Machine check exception
60h	VMEXIT_INTR	Physical INTR
61h	VMEXIT_NMI	Physical NMI
62h	VMEXIT_SMI	Physical SMI
63h	VMEXIT_INIT	Physical INIT
64h	VMEXIT_VINTR	Virtual INTR
77h	VMEXIT_PAUSE	PAUSE instruction
78h	VMEXIT_HLT	HLT instruction
7Fh	VMEXIT_SHUTDOWN	Shutdown
8Fh	VMEXIT_EFER_WRITE_TRAP	See section 15.35.10
90h -9Fh	VMEXIT_CR[0-15]_WRITE_TRAP	See section 15.35.10
400h	VMEXIT_NPF	Only if PFCODE[3]=0 (no reserved bit error)
403h	VMEXIT_VMGEXIT	VMGEXIT instruction
–1	VMEXIT_INVALID	Invalid guest state
–2	VMEXIT_BUSY	Busy bit was set in guest state (see Section 15.36.16)

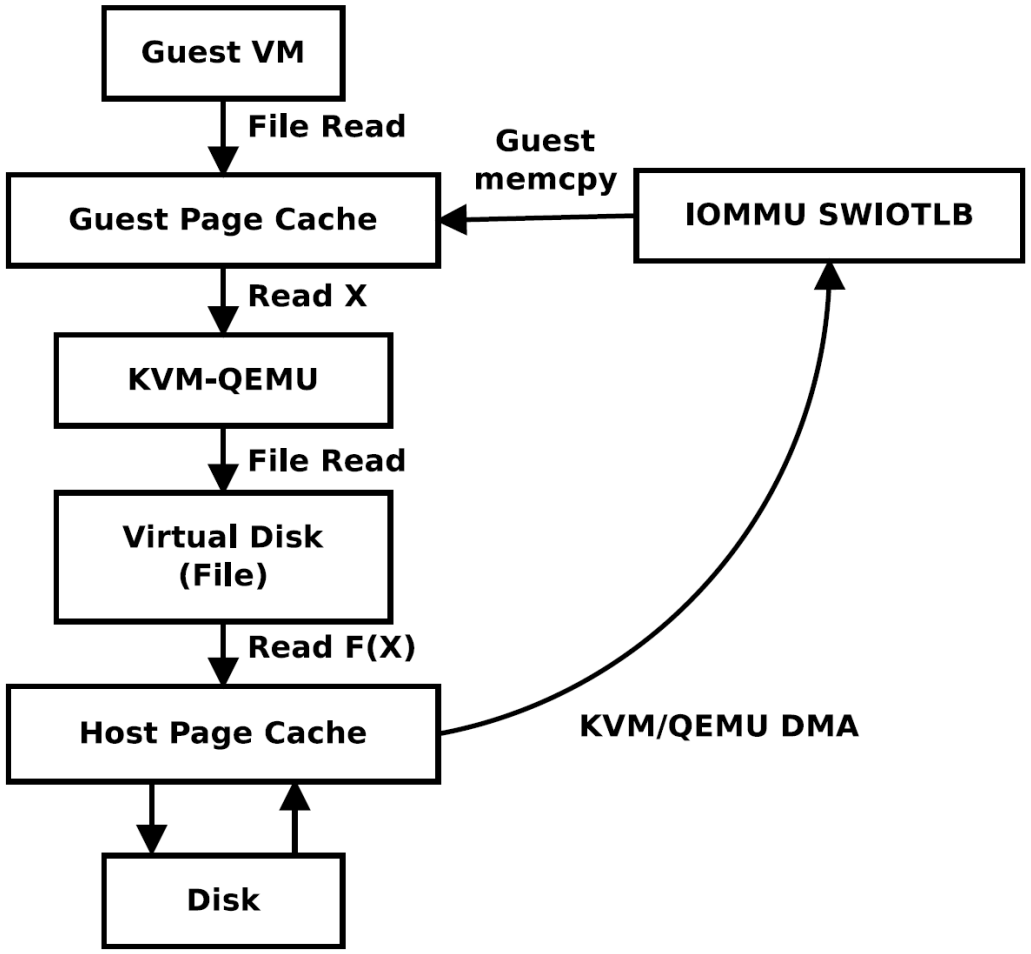
SEV with Encrypted State, SEV-ES

- VM 和 HV 使用 Guest Hypervisor Communication Block, GHCB 通信，设置为共享内存
- NAE 事件触发会产生异常 #VC，由 guest 中的 VC handler 处理，负责响应和请求 HV 的服务
- VC handler 将请求和相关状态信息写入到 GHCB，使用 VMGEXIT 将控制权转移到 HV，HV 读取 GHCB，执行模拟，将返回值写回 GHCB，执行 VMRUN 恢复到 VC handler 执行
- VC handler 检查 GHCB 中的返回值，更新 guest 状态



SEV I/O Operation

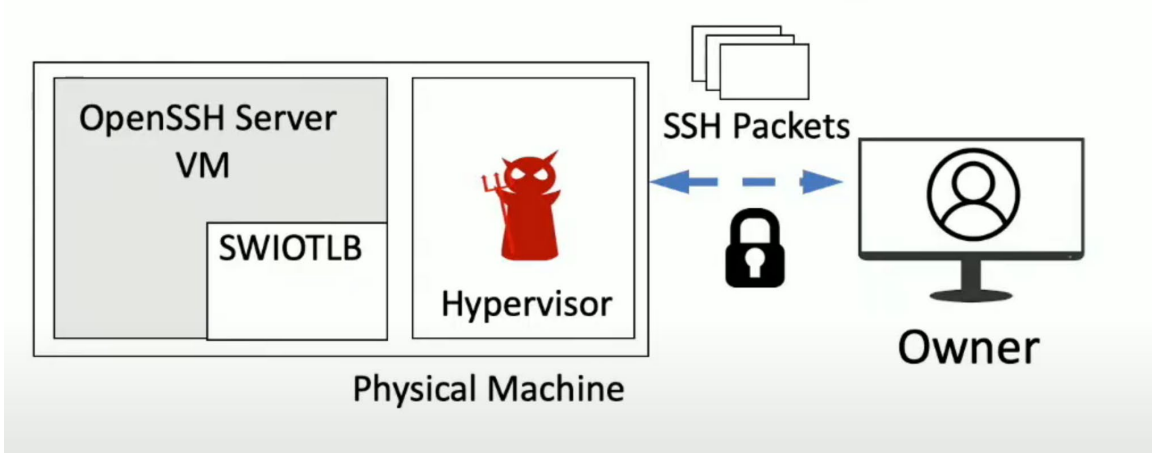
- SEV 的 IOMMU 仅支持带 HV 加密密钥的 DMA，SEV 的 I/O 操作需要 SEV VM 中的共享内存区域 Software I/O TLB, SWIOTLB
- 数据通过 QEMU-KVM 的 DMA 从磁盘读到 SWIOTLB 中，然后由 guest 复制到磁盘驱动的 I/O 缓冲区
- I/O 数据对 HV 可见，HV 可以监控或修改 I/O，即使有磁盘加密或 TLS



SEV I/O 支持：virtio、passthrough、sr-iov

Unprotected I/O

- 共享内存对 HV 可见，且 HV 可以修改加密内存
- HV 将要解密的内存放到要 guest 复制的地址 Bp
- Guest 将数据复制到 SWIOTLB 后，内容被解密



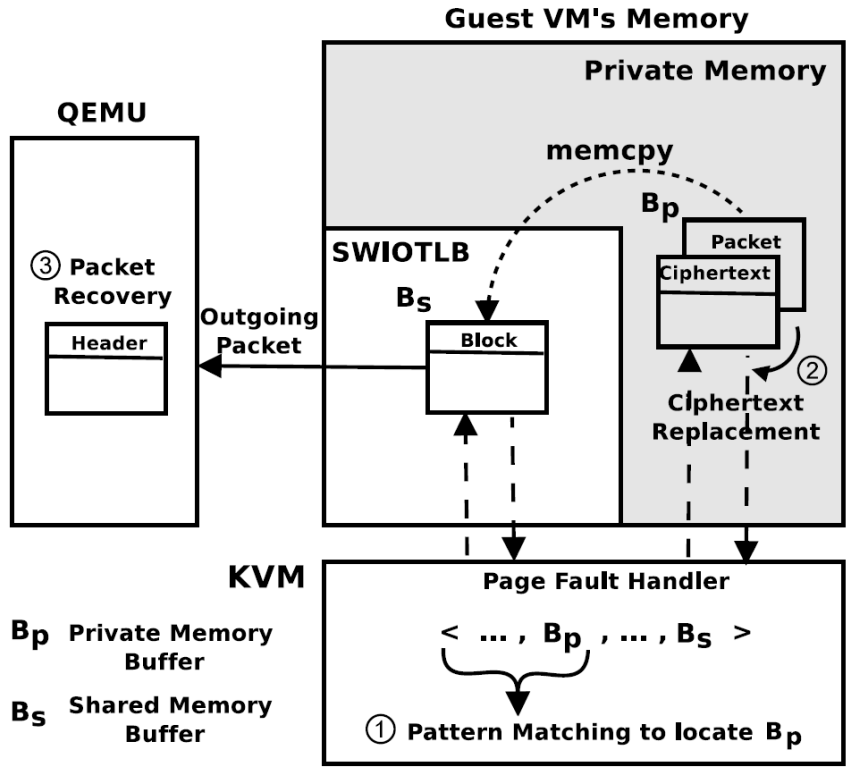


Figure 3: A decryption oracle. Step ①, the hypervisor conducts pattern matching using page-fault side channels to determine the address of B_p. Step ②, the hypervisor replaces a ciphertext block in B_p with the target memory block, which will be decrypted when copied to B_s. Step ③, QEMU recovers the network packet headers.

Mengyuan Li, Yinqian Zhang, and Zhiqiang Lin, Exploiting Unprotected I/O Operations in AMD's Secure Encrypted Virtualization, USENIX 2019.

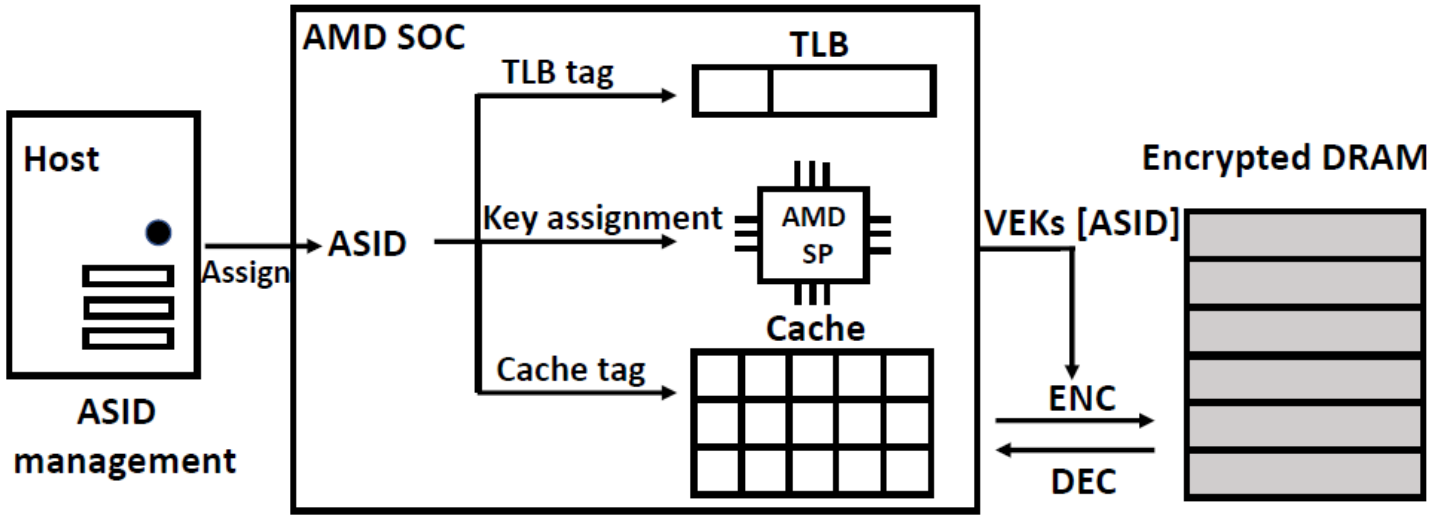
Lack of Memory Integrity Protection

- SEV 缺乏内存完整性保护
- HV 可以访问/修改加密后的内存
  - 重放攻击
  - 恶意篡改
- 修改 nPT，破坏 gPA 到 sPA 的单射
  - 内存别名，多个 gPA 对应一个 sPA
  - 内存重映射，将 gPA 映射到多个 sPA

Mathias Morbitzer, Manuel Huber, Julian Horsch, and Sascha Wessel. SEVered: Subverting AMD's virtual machine encryption. EuroSec@EuroSys 2018.
Mathias Morbitzer, Manuel Huber, and Julian Horsch. Extracting secrets from encrypted virtual machines. CODASPY 2019.

ASID Abuse: ASID-based Isolation in SEV

- ASID 作为内存、缓存和 TLB 访问控制的标识符。
  - 索引内存加密密钥
  - Non-SEV VM 中，每个 VCPU 可以有不同的 ASID，HV 动态分配
  - SEV VM 所有 VCPU 共用一个 ASID，理论上在VM 整个生命周期保持不变
  - ASID 和 C-bit 都包含在 Cache tag 中
- ASID 由 HV 管理



ASID Abuse: ASID Management

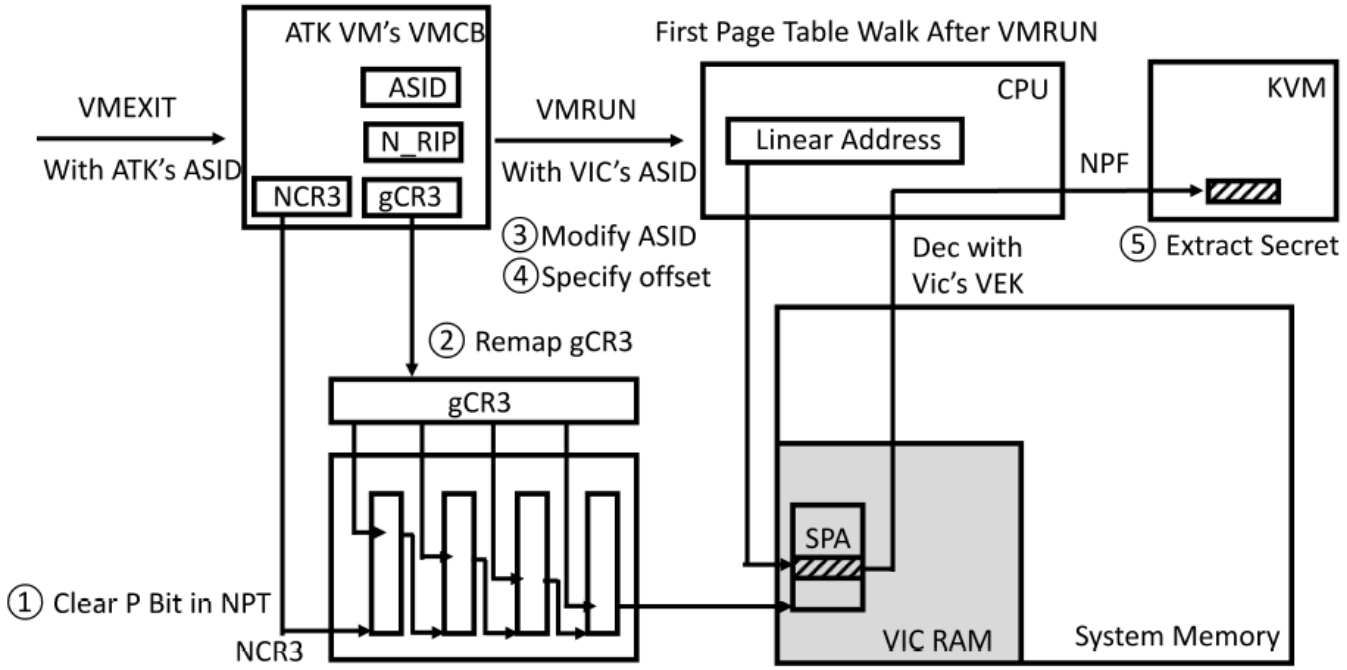
- HV 维护 ASID 池，SEV-ASID 池的范围由 CPUID 8000_001F[ECX] 指定。HV 使用 ACTIVATE 通知 AMD-SP 为给定的 guest 绑定一个 ASID。
- ASID 在 VMCB 的未加密区域，在 VMEXIT 期间，HV 可以修改 ASID 字段。
- “Security-by-crash”：AMD 认为，当 HV 修改 SEV VM 的 ASID 字段后并恢复执行后，guest 将使用不正确的密钥解密内存，导致 VM 崩溃。
- 虽然 VM 会崩溃，但从恢复执行到崩溃的这段足够攻击者 VM 冒充受害者 VM 从而破坏机密性和完整性——CrossLine 攻击，提取由受害者密钥加密的内存内容。
- 两个变体，V1 可以解密受害者 gPT 或符合 PTE 格式的内存块（SEV & SEV-ES），V2 可以执行受害者 VM 的指令构造加密或解密的 oracle（SEV only）。

Mengyuan Li, Yinqian Zhang, and Zhiqiang Lin. CROSSLINE: Breaking “Security-by-Crash” based Memory Isolation in AMD SEV. CCS 2021.

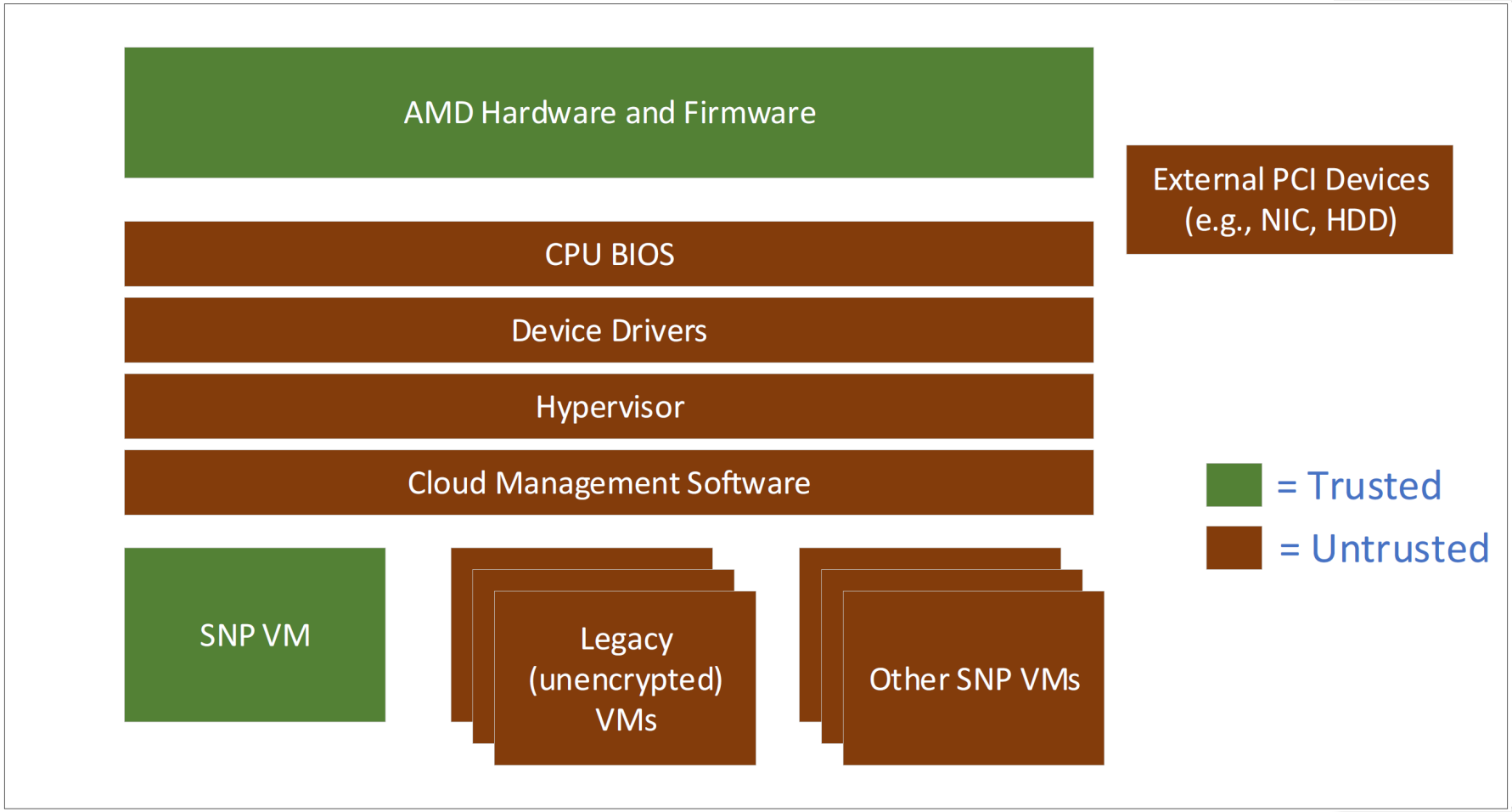
ASID Abuse: CrossLine V1

两个 SEV VM，攻击者和受害者，目标内存物理页帧为 sPFN0，目标页上的一个 8-byte 对齐地址 sPA0

1. 修改 nPT，清零攻击者页的 P-bit，访存会触发 Nested Page Fault, NPF
2. 修改 nPT，将攻击者 VM 当前进程的 gCR3 重映射到 sPFN0，并将 P-bit 置位
3. 修改攻击者 ASID 为受害者 ASID
4. 修改 NRIP 指定要解密的页偏移，利用页目录表每次解密 8-byte
5. VMRUN 恢复执行，取指，从 gCR3 开始地址转换，使用受害者 ASID 索引的密钥解密 sPFN0。根据指令虚拟地址的 47-39 bit 索引 8-byte 内存块作为页表项加载，触发 NPF 的 gPA 会保存在 VMCB 的 EXITINFO2 字段。

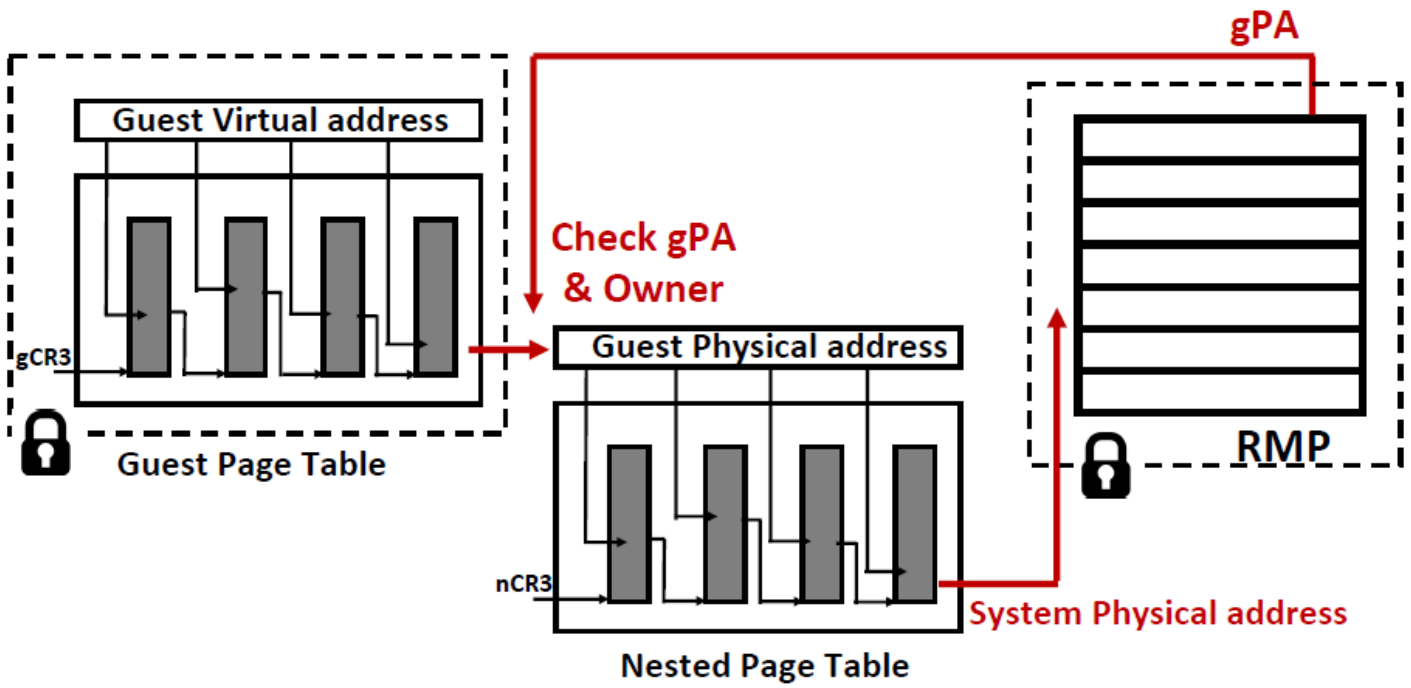


SEV with Secure Nested Paging, SEV-SNP



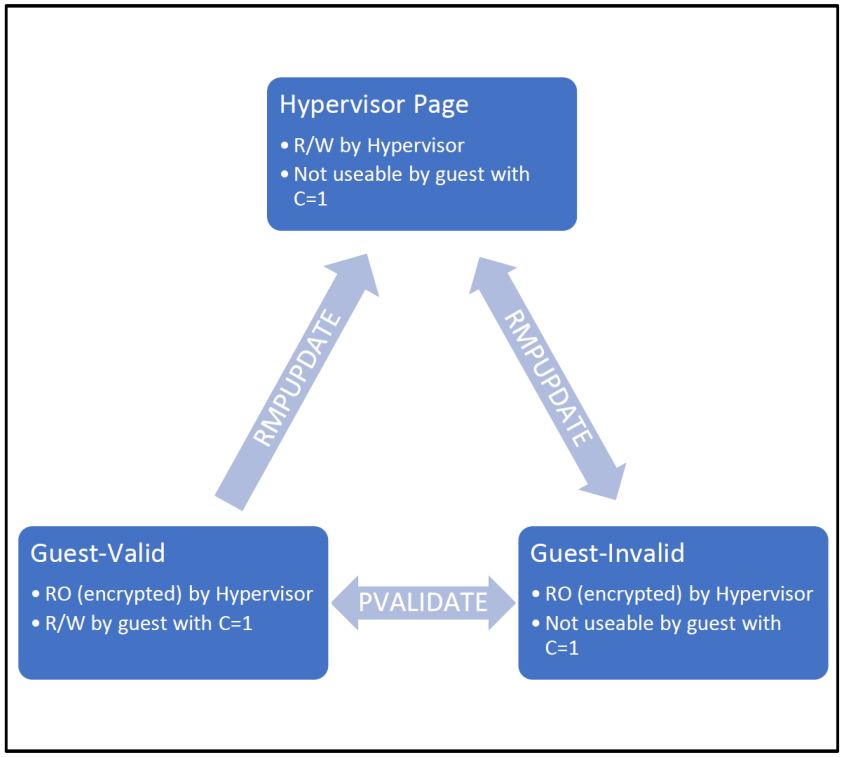
- SEV-SNP 威胁模型
- 提供完整性保护和额外的可选安全增强功能

SEV-SNP



- RMP 是整个系统共享的单一数据结构，由 sPA 索引，每个 4K 页对应一个条目。
- RMP 条目包含物理页状态 (HV、guest-invalid、guest-vali)、所有权 (VM ASID、gPA)。
- 完成地址转换后检查 RMP
- 不是所有的内存访问都要检查
  - SEV-SNP VM私有内存读写
  - 任何模式下的写访问
- RMP 不能被软件直接写入，

SEV-SNP



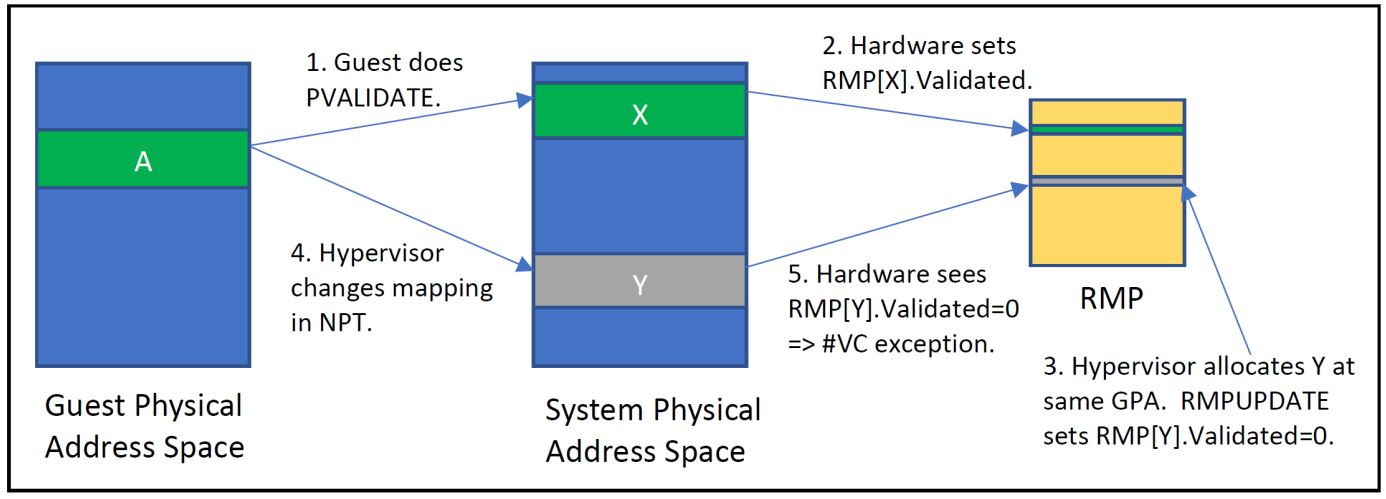
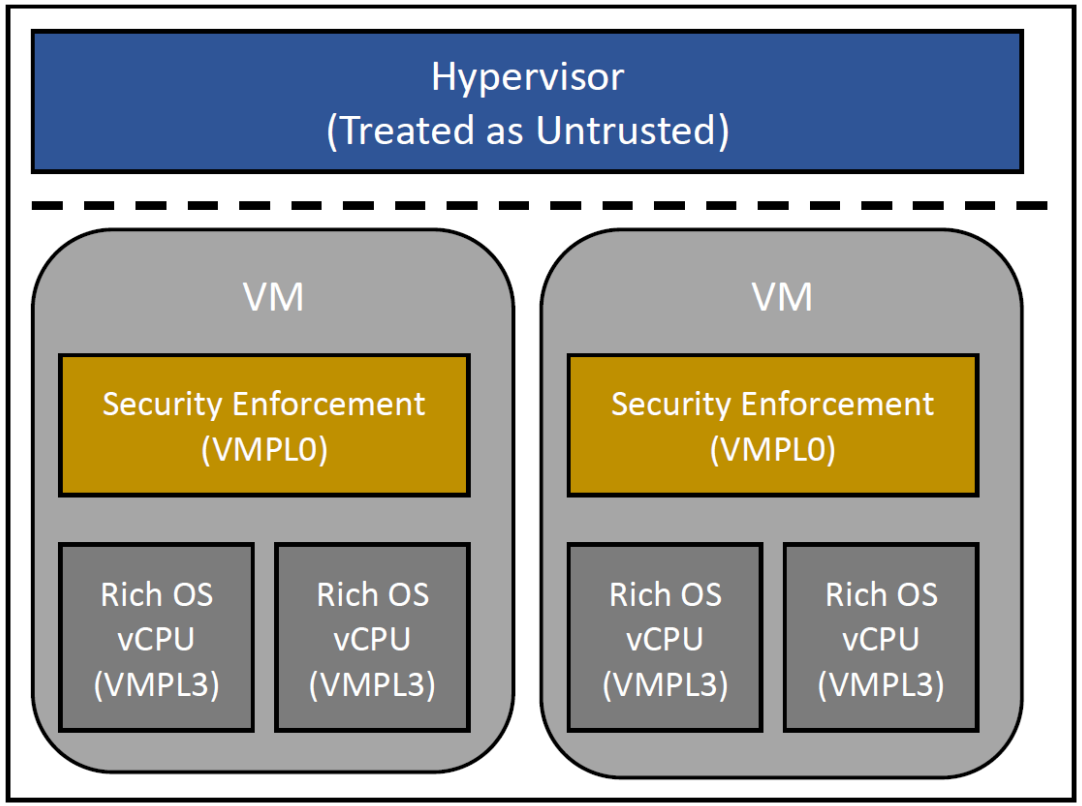


FIGURE 5: PAGE RE-MAPPING ATTACK

- 页验证操作
- RMP 条目验证标志位
- HV 为 guest 分配新内存页：
  - HV 使用 RMPUPDATE 将页分配给 guest，RMP 验证位置零，页状态为 guest-invalid
  - Guest 中使用 PVALIDATE 验证页，将标志位置位，guest-valid
- 一经验证，HV 对页的任何取消分配、重新分配或重映射都将使页面转换为 guest-invalid 状态
- Guest 对所有未验证页的访问会触发 #VC 异常

SEV-SNP



可选安全增强功能

- 虚拟机特权级 VMPL
  - 为 VCPU 分配 VMPL
  - 设置 VMSA 的 VMPL 字段
  - Guest 使用 RMPADJUST 修改 VMPL
- 中断/异常保护、可信平台信息等
- BTB 保护

Table 15-38. VMPL Permission Mask Definition

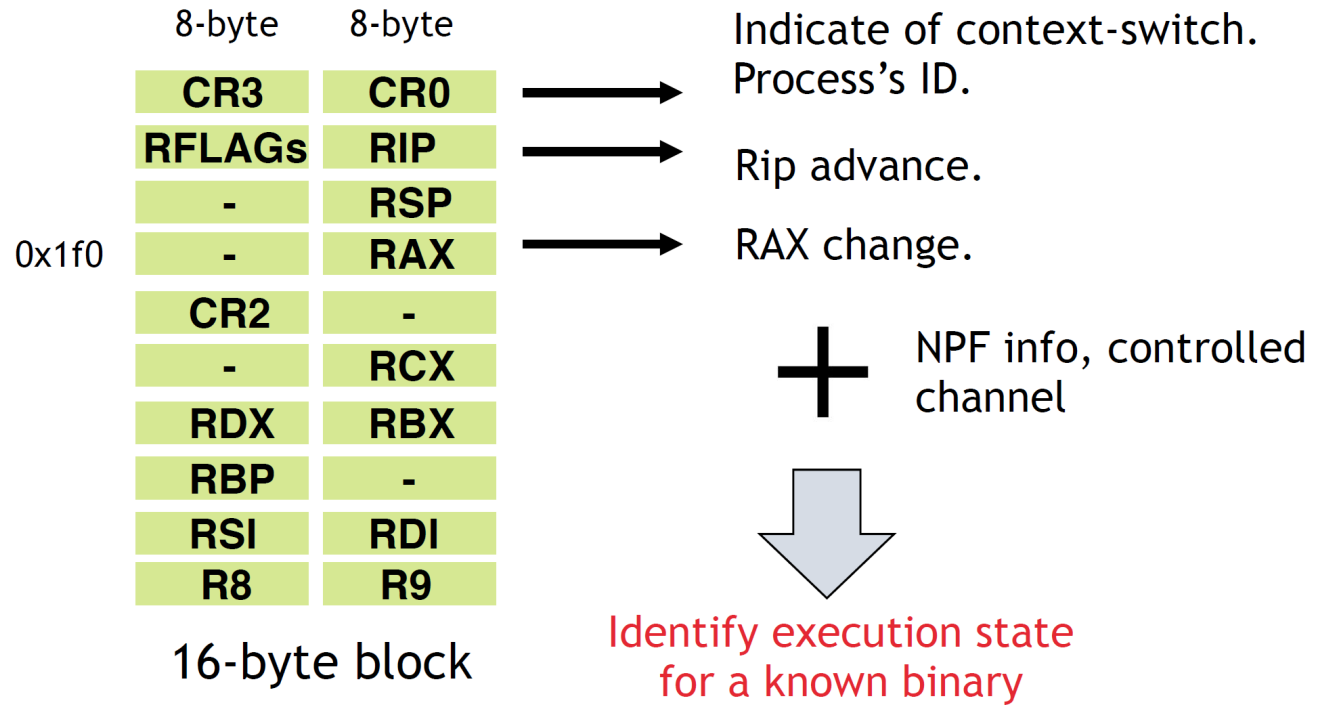
Bit	Name	Settings
0	Read	0: Reads cause #VMEXIT(NPF) 1: Reads are allowed
1	Write	0: Writes cause #VMEXIT(NPF) 1: Writes are allowed
2	Execute-User	0: Execution at CPL 3 causes #VMEXIT(NPF) 1: Execution at CPL 3 is allowed
3	Execute-Supervisor	0: Execution at CPL < 3 causes #VMEXIT(NPF) 1: Execution at CPL < 3 is allowed
4-7	Reserved	SBZ

Ciphertext Side Channel

- 128-bit AES XEX

$$C = \mathrm{T}(\mathrm{sPA_m}) \oplus Enc(\mathrm{m} \oplus \mathrm{T}(\mathrm{sPA_m}))$$

- 在 VM 生命周期中，同一物理地址的相同明文对应的密文是不变的
- HV 可以读密文
- SEV-ES 中，VMCB 中有指向 VMSA（4K页）的指针（物理地址），HV 监控 VMSA 相应字段密文的变化推断明文的修改，进而推断 SEV VM 执行状态
- HV 直接通过物理地址访问，SEV-SNP 中 RMP 的访问控制不会限制读 VMSA



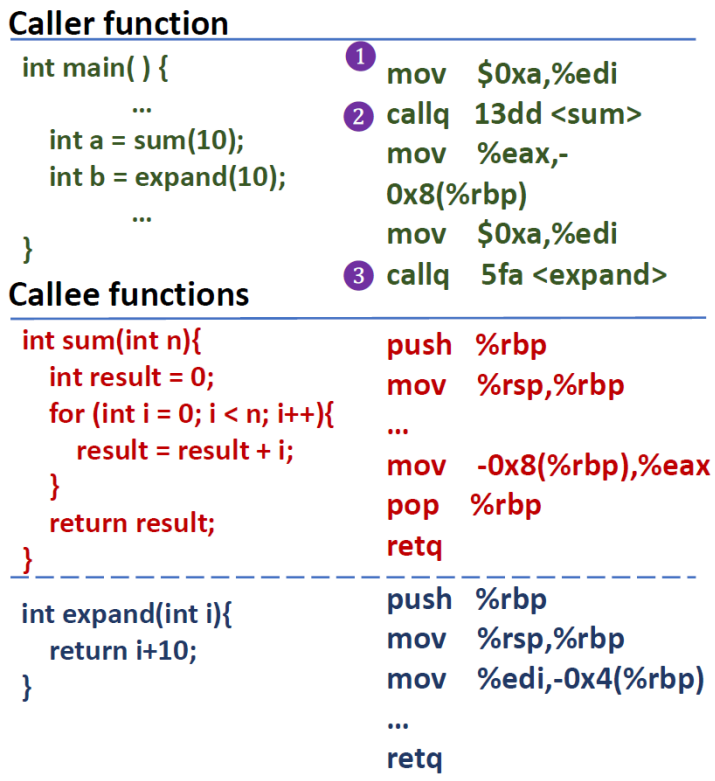
Mengyuan Li, Yinqian Zhang, and Huibo Wang. CIPHERLEASK: Breaking Constant-time Cryptography on AMD SEV via the Ciphertext Side Channel. USENIX 2021.

Ciphertext Side Channel

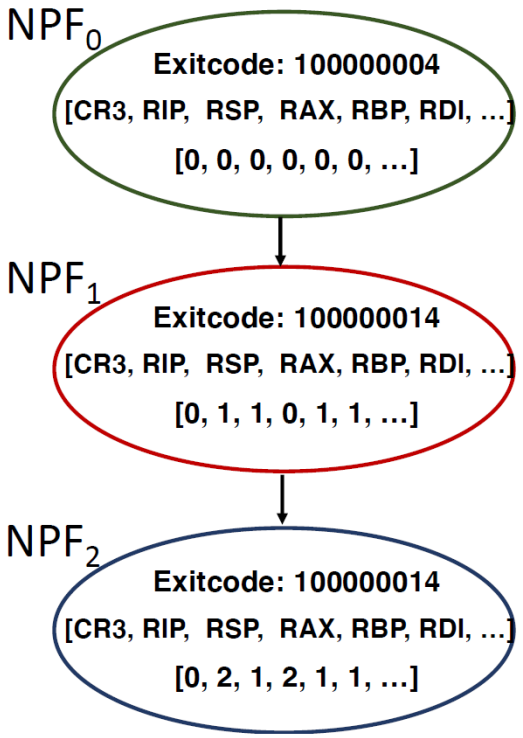
推断 SEV VM 执行状态

1. 清空 nPT 的 P-bit
2. 访存（取指）触发 NPF，HV 记录 VMSA、时间戳和 EXITCODE，将 P-bit 置位，恢复执行
3. 收集一系列的密文块和时间戳，比较 CR3&CR0 字段，将每次 NPF 与进程关联

NPF 中，VMCB 的 EXITINFO2 字段泄露函数的物理地址



(a) C source code with assembly code.



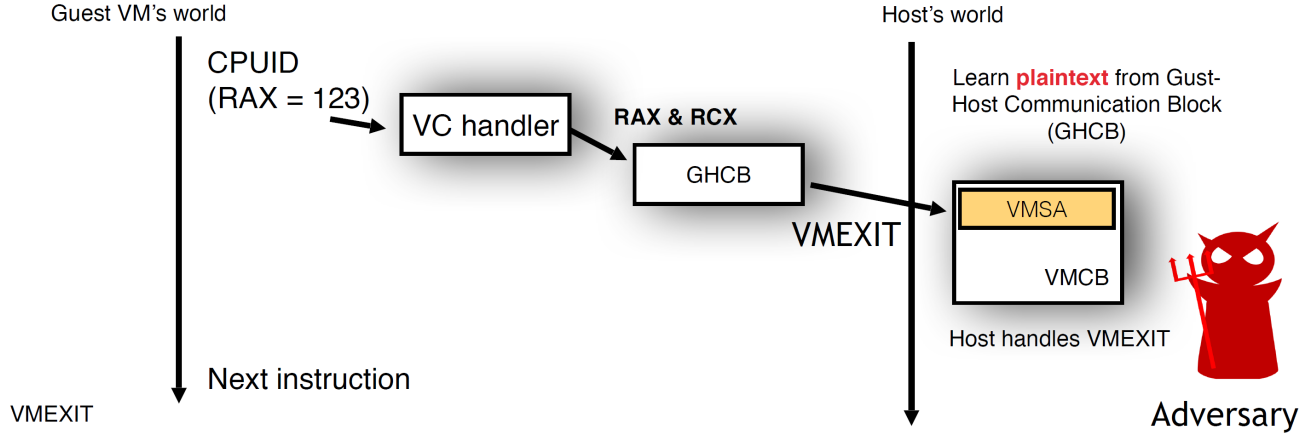
(b) Ciphertext blocks.

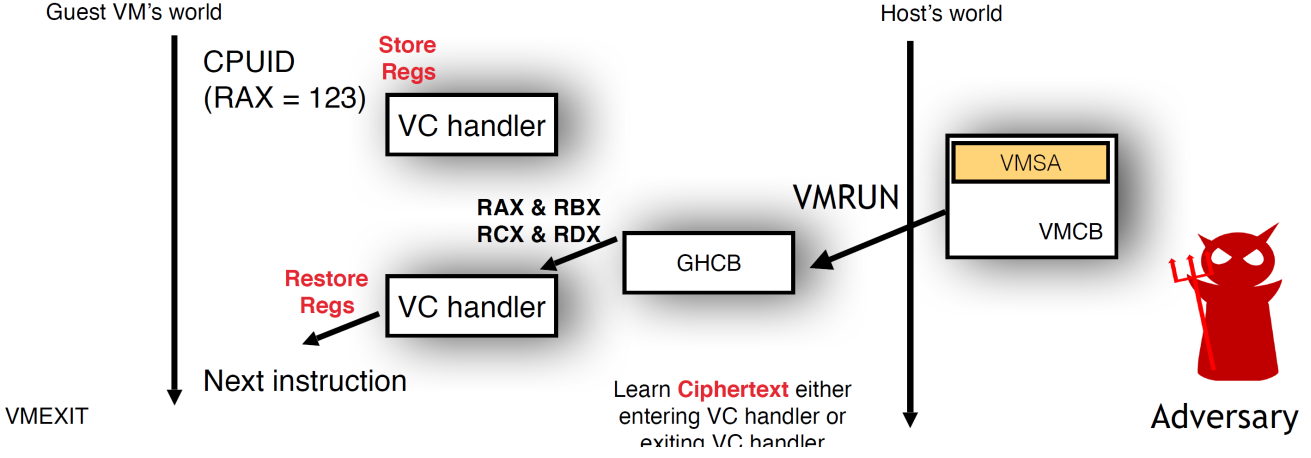
Mengyuan Li, Yinqian Zhang, and Huibo Wang. CIPHERLEASK: Breaking Constant-time Cryptography on AMD SEV via the Ciphertext Side Channel. USENIX 2021.

Ciphertext Side Channel

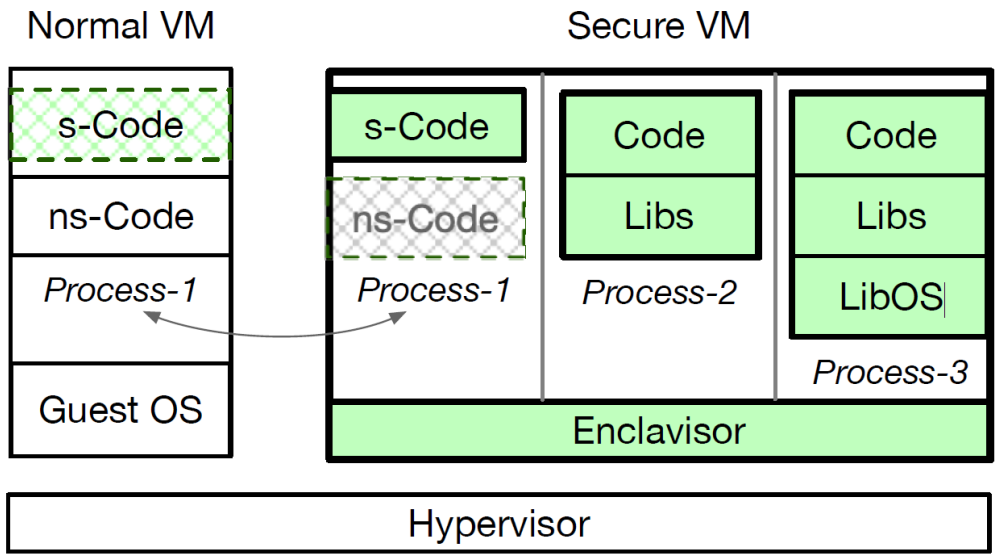
恢复明文

- NAE 事件中，根据 GHCB 和 VMSA 的对应值构造指定寄存器的明密文对字典
- 通过 Boot 阶段的 NAE (CPUID, IOIO_Port) 可以收集到 RAX [0-127] 对应的密文
- 攻击案例
  - 根据已知 RAX [0-31] 对应的密文可以破解 OpenSSL RSA 的常数时间实现
  - 根据 RAX [0-1] 对应的密文可以破解 ECDSA 签名的常数时间实现





Enclavisor



- SGX on SEV
- 系统调用重定向到 guest OS
- 限制 HV 对安全内存页的访问
  - 在 HV 通特权级部署可信组件，HV 需要调用可信组件修改

A comparison on Intel SGX, AMD SEV and Enclavisor from different dimensions.

	Memory encryption	Memory integrity	Memory size limitation	Enclave number	Enclave granularity	Fast boot	Efficient interaction	Remote attestation
SGX	Yes	Yes	All 256MB EPC	Unlimited	Fine-grained	No	No	Pre/post-boot
SEV	Yes	No	All phy-mem	Limited (15)	Coarse-grained	No	No	Pre-boot only
Enclavisor	Yes	Partial [*]	All phy-mem	Unlimited	Multiple	Yes	Yes	Pre/post-boot

Jinyu Gu, Xinyue Wu, Bojun Zhu, Yubin Xia, Binyu Zang, Haibing Guan, and Haibo Chen. Enclavisor: A Hardware-software Co-design for Enclaves on Untrusted Cloud. TC 2020.
Yuming Wu, Yutao Liu, Ruifeng Liu, Haibo Chen, Binyu Zang, Haibing Guan. Comprehensive VM Protection against Untrusted Hypervisor through Retrofitted AMD Memory Encryption. HPCA 2018.

Discussion

- Application
- Intel SGX
  - Features Intel SGX, the most deployed trusted execution environment, with up to 512GB enclave per processor.
- Intel TDX
  - SEAM mode, Intel TDX module
- Flexible and Scalable TEE

References

- AMD Memory Encryption. White paper. 2016.
- Protecting VM Register State with SEV-ES. White paper. 2017.
- AMD SEV-SNP: Strengthening VM Isolation with Integrity Protection and More. White paper. 2020.
- AMD Programmer's Manual. Volume 2. Section 15.34-15.36.
- Felicitas Hetzelt, Robert Buhren, Security Analysis of Encrypted Virtual Machines, VEE 2017.
- Mengyuan Li, Yinqian Zhang, and Zhiqiang Lin, Exploiting Unprotected I/O Operations in AMD's Secure Encrypted Virtualization, USENIX 2019.
- Mathias Morbitzer, Manuel Huber, Julian Horsch, and Sascha Wessel. SEVered: Subverting AMD's virtual machine encryption. EuroSec@EuroSys 2018.
- Mathias Morbitzer, Manuel Huber, and Julian Horsch. Extracting secrets from encrypted virtual machines. CODASPY 2019.
- Mengyuan Li, Yinqian Zhang, and Zhiqiang Lin. CROSSLINE: Breaking "Security-by-Crash" based Memory Isolation in AMD SEV. CCS 2021.
- Mengyuan Li, Yinqian Zhang, and Huibo Wang. CIPHERLEASK: Breaking Constant-time Cryptography on AMD SEV via the Ciphertext Side Channel. USENIX 2021.
- Jinyu Gu, Xinyue Wu, Bojun Zhu, Yubin Xia, Binyu Zang, Haibing Guan, and Haibo Chen. Enclavisor: A Hardware-software Co-design for Enclaves on Untrusted Cloud. TC 2020.
- Yuming Wu, Yutao Liu, Ruifeng Liu, Haibo Chen, Binyu Zang, Haibing Guan. Comprehensive VM Protection against Untrusted Hypervisor through Retrofitted AMD Memory Encryption. HPCA 2018.

Q&A