Microsoft

FOSDEM

Confidential Virtual Machines Demystified: A Technical Deep Dive into Linux Guest OS Enlightenment

Ankita Pareek

Software Engineer, Azure Linux, Microsoft

Archana

Software Engineer, Azure Linux, Microsoft

Microsoft

What to expect?

- Understanding Confidential VMs
- Public cloud provider trends
- Concerns wrt Cloud CVMs & expectations from system stack (Guest OS)
- Concepts: UKI, Secure Boot, Measured Boot, Full disk encryption & Attestation
- Configurations for Guest OS enlightenment (TDX, SEV-SNP)

Microsoft

Confidential Virtual Machines (CVMs)

- A use case of confidential computing
- Running workload and its data in use is protected from higher privilege layers of the software and hardware stack
- Most flexible approach, easy lift and shift
- Provide remote attestation facilities
- Backed by CPU vendors: **AMD SEV-SNP**, **Intel TDX**, IBM Z SE, OpenPower PEF, ARM CCA
- Cloud provider adoptions
  - Google — N2D (AMD SEV-SNP), C3D/C2D/N2D (AMD SEV), C3 (Intel TDX) Machine Series
  - Microsoft — DCasv5/DCadsv5/ECasv5/ECadsv5 (AMD SEV-SNP), DCesv5/DCedsv5/ECesv5/ECedsv5 (Intel TDX), NCCadsH100v5 (AMD SEV-SNP and NVIDIA H100 Tensor Core GPUs CVM with Confidential GPU)
  - AWS — M6a, C6a, and R6a instance types
  - Alibaba
  - IBM
  - Oracle
- Allow Cloud Providers to Guarantee: No access to customer data - even from privileged layers of the virtualization stack

Microsoft

What does memory encryption in CVM entail?

	Normal VM	Confidential VM
Management	Full VMM control	Restricted VMM access, secure processor involvement
VM Initialization & Boot	Standard boot process	Secure measurement of components, cryptographic attestation
VM VMM communication	VMEXIT events	VMGXIT, TDCALL, specialized communication protocol
I/O and MMIO	Direct access	Explicit VMM calls required, exceptions for normal I/O #VC, #VE
Direct Memory Access (DMA)	Open device access	External devices cannot access guest memory, dedicated bounce buffer

Microsoft


I promise confidentiality for the guest OS.
It's truly private, right!
It's truly private, right?

Microsoft

CVMs and Guest OS, Where is the problem?

- Confidentiality guarantee does not apply for guest OS.
- Protecting guest OS is important. Why?
- Need for Full Disk Encryption
  - **without relying on host**
  - supporting **unattended boot**
- Using vTPM
  - Attestation
  - Need to trust cloud providers
- Remote attestation

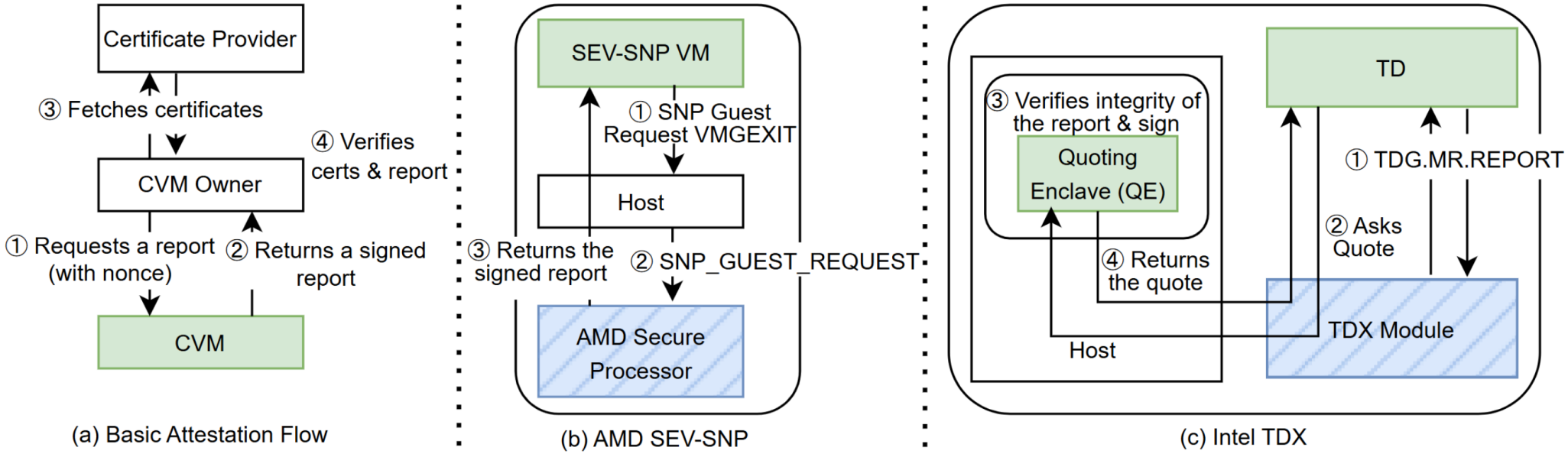
Microsoft

I don’t trust the CLOUD

- Attest
- How would a consumer know **it is** what **it is**?
  - How is evidence generated?
    - PSP, QE – Hardware based security module
    - vTPM – Software based security module
  - What does the evidence contain?
  - Authenticity of the attestation report (verification) – Local & Remote
- I don’t trust the verifier (MAA, Google attestation agent, ITA, etc. etc.) – You can build your own!

Microsoft

Remote attestation



Microsoft

What does the Guest OS need to take care of?

- Guest Kernel (with patches for AMD SEV-SNP or Intel TDX)
- Kernel Configurations
- Full disk encryption
- Secure Boot
- Measured Boot

NOTE: We will mainly be focusing on these concepts from the point of view of rpm-based distros

Microsoft

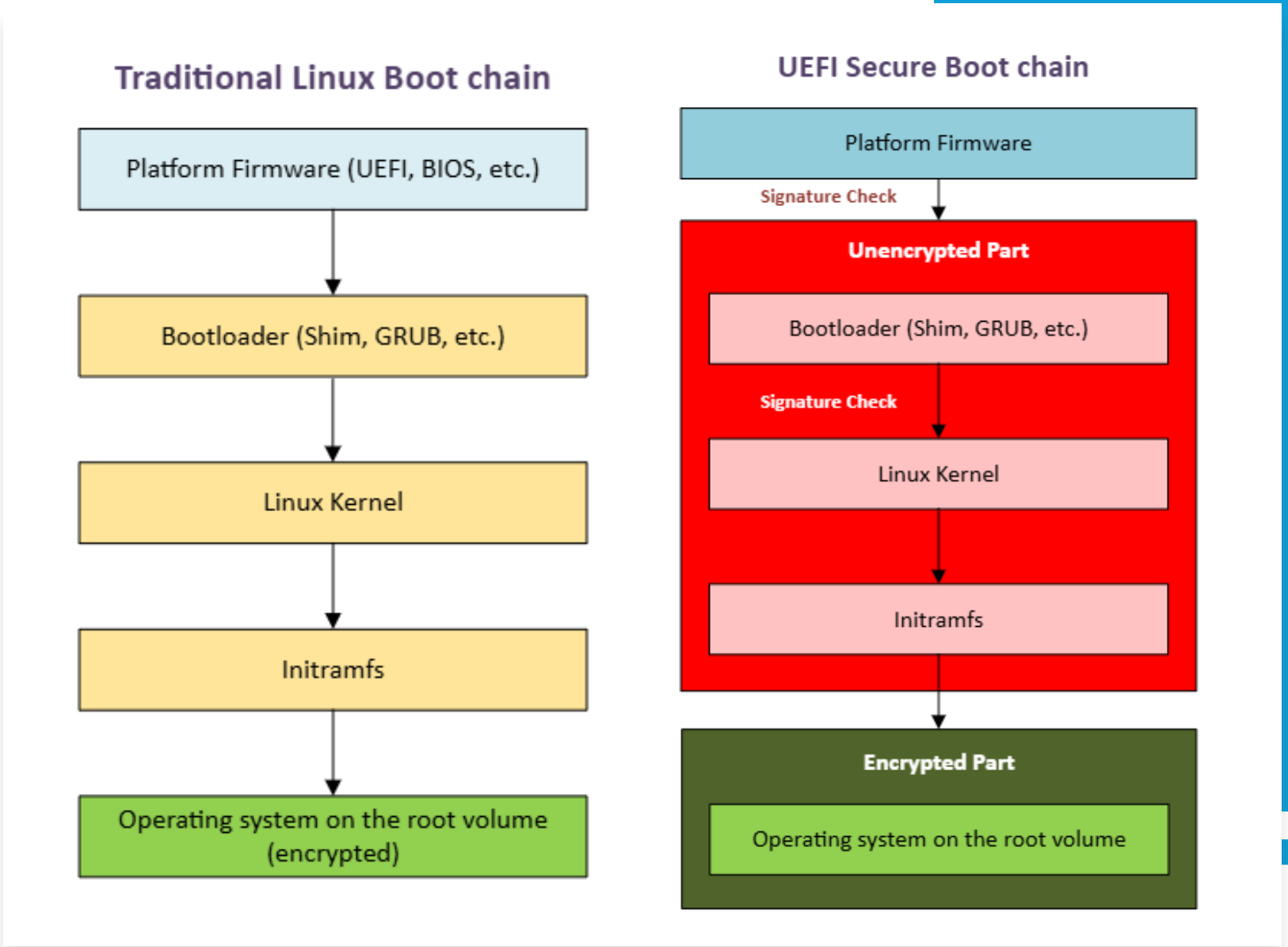
Kernel Configurations

AMD SEV-SNP	Intel TDX
CONFIG_AMD_MEM_ENCRYPT Kconfig for enabling support for memory encryption on AMD processors	**CONFIG_INTEL_TDX_GUEST** Kconfig for supporting running as a guest under Intel TDX processor
CONFIG_SEV_GUEST KConfig for supporting the driver which provides userspace interface to communicate securely with the PSP for requesting attestation report and more.	**CONFIG_TDX_GUEST_DRIVER** Kconfig for enabling the driver which provides userspace interface to communicate with the TDX module to request the TDX guest details like attestation report.

NOTE: We assume that the guest kernel has all the required patches for supporting a confidential VM on the respective processors

Microsoft

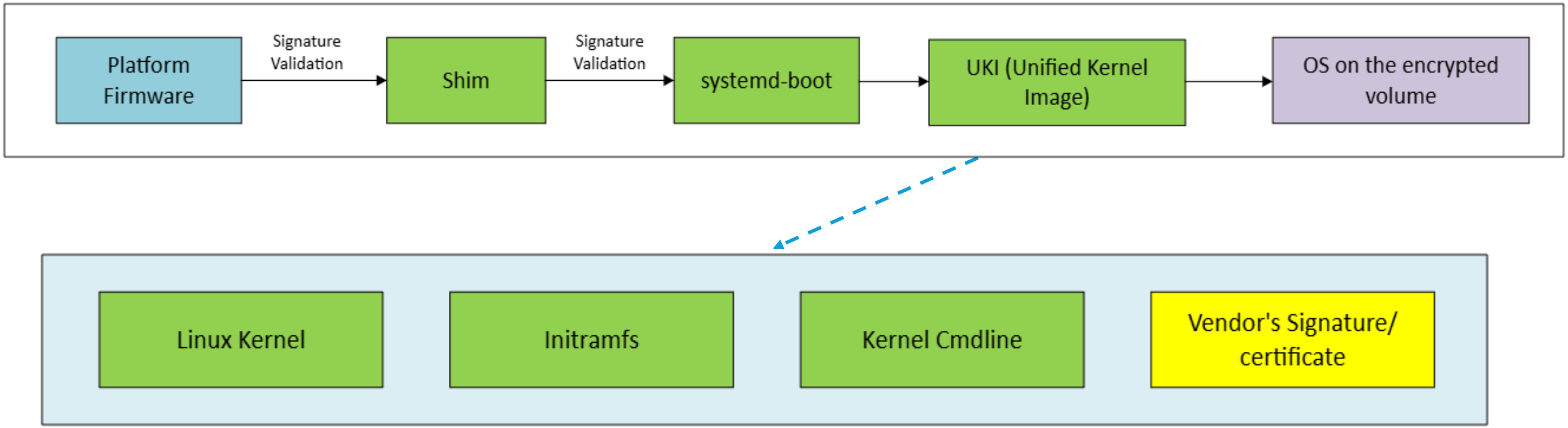
Booting Linux in a Secure Fashion



Microsoft

Boot simplified with UKI

CVM Secure Boot Flow



[Unified Kernel Image (UKI)]

Microsoft

Why UKI?

✓ Immutable Bundle

✓Single Signature Coverage

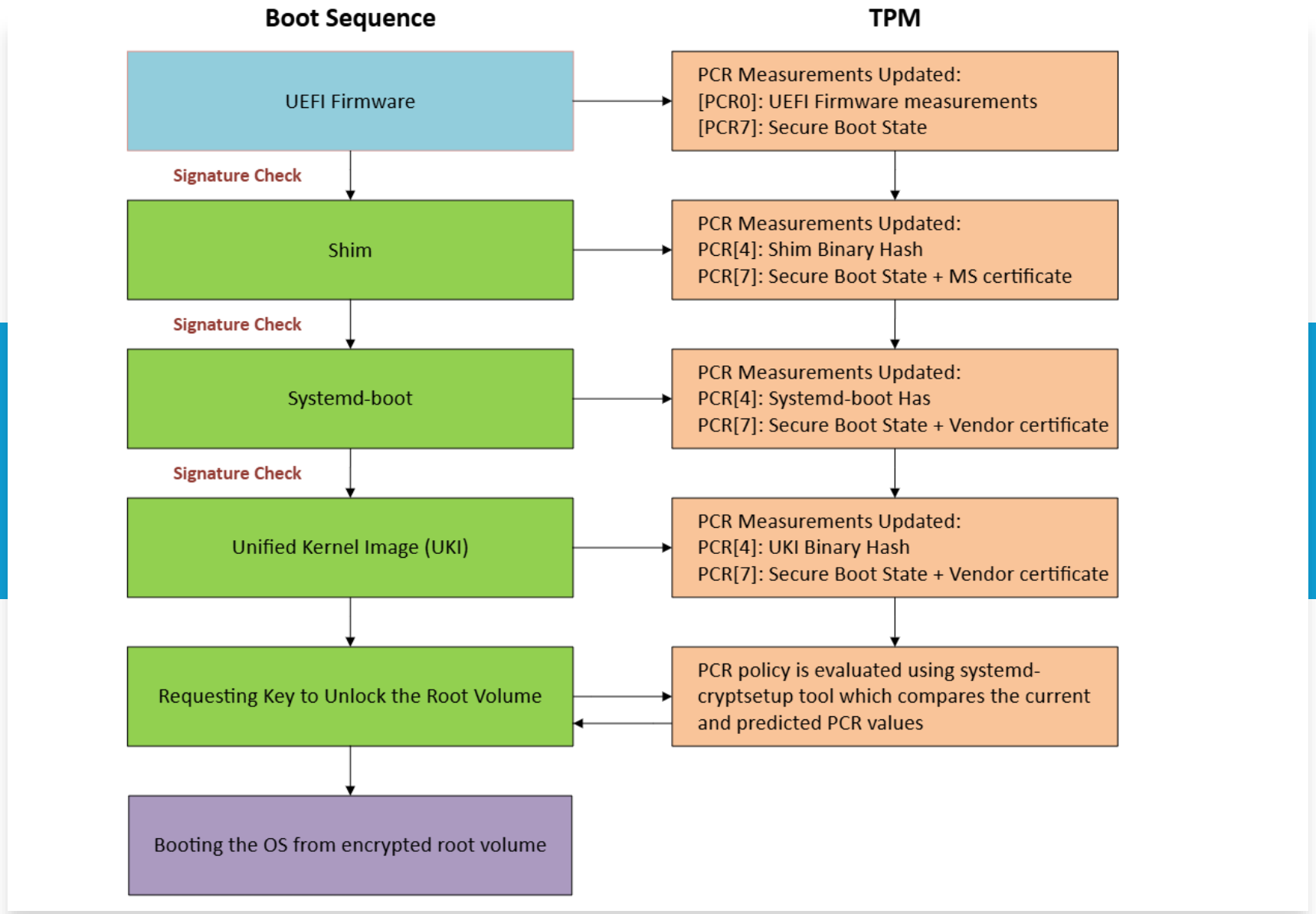
✓ Secure Boot Integration

Microsoft

UKI: The Security v/s Flexibility Trade-off

- **Static Initramfs** - UKI requires vendor-built static initramfs, preventing runtime module additions or customization
- **Immutable Kernel Command Line** - Kernel command line becomes fixed in UKI, requiring alternative methods for root volume discovery
- **Bootloader Interface Changes -** UKI binary loads directly from firmware/Shim, removing traditional bootloader UI and user interaction

Microsoft

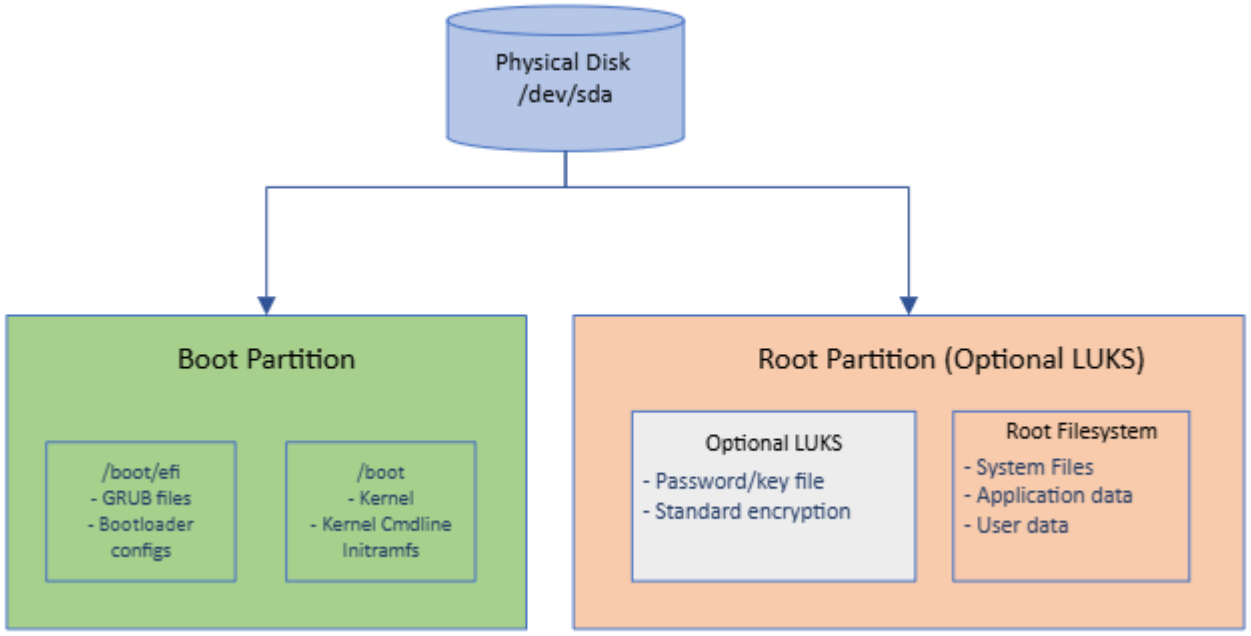


Measured Boot

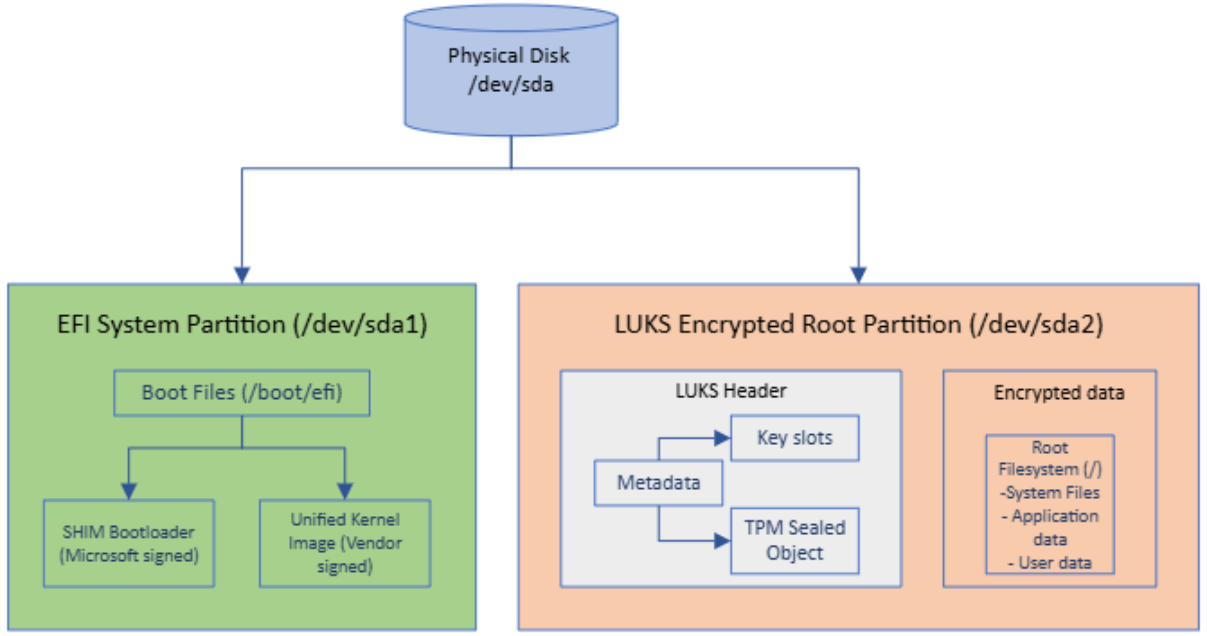
The standard way for checking the authenticity of the boot chain is called **Measured Boot** :a distinct hash value from each component in the chain is recorded in TPM's Platform Configuration Registers (PCRs) and can be examined later.

Microsoft

Disk Layout (Standard v/s Encrypted)



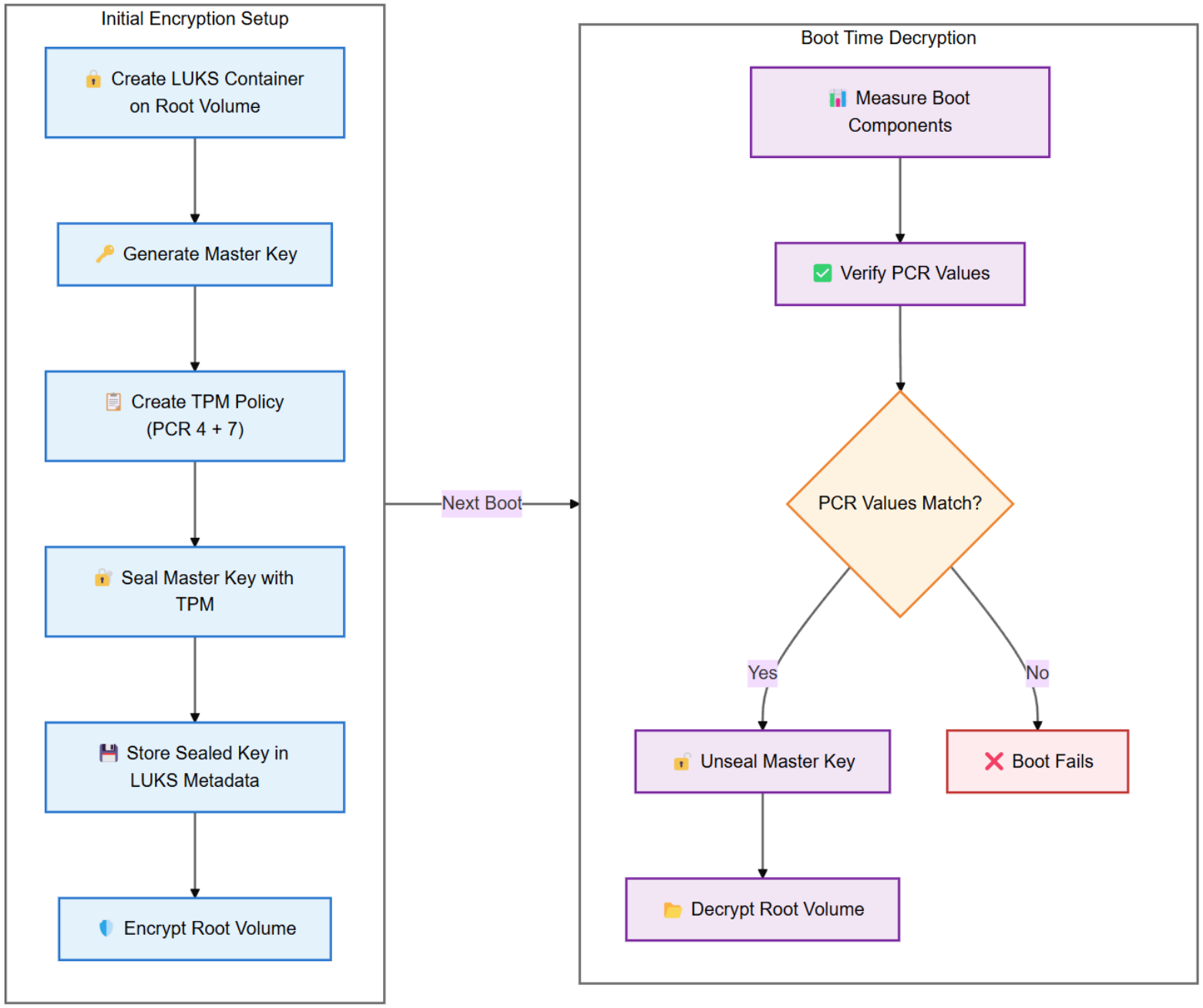
Standard Linux VM Partition Layout in Guest OS



Confidential Linux VM Partition Layout in Guest OS

Microsoft

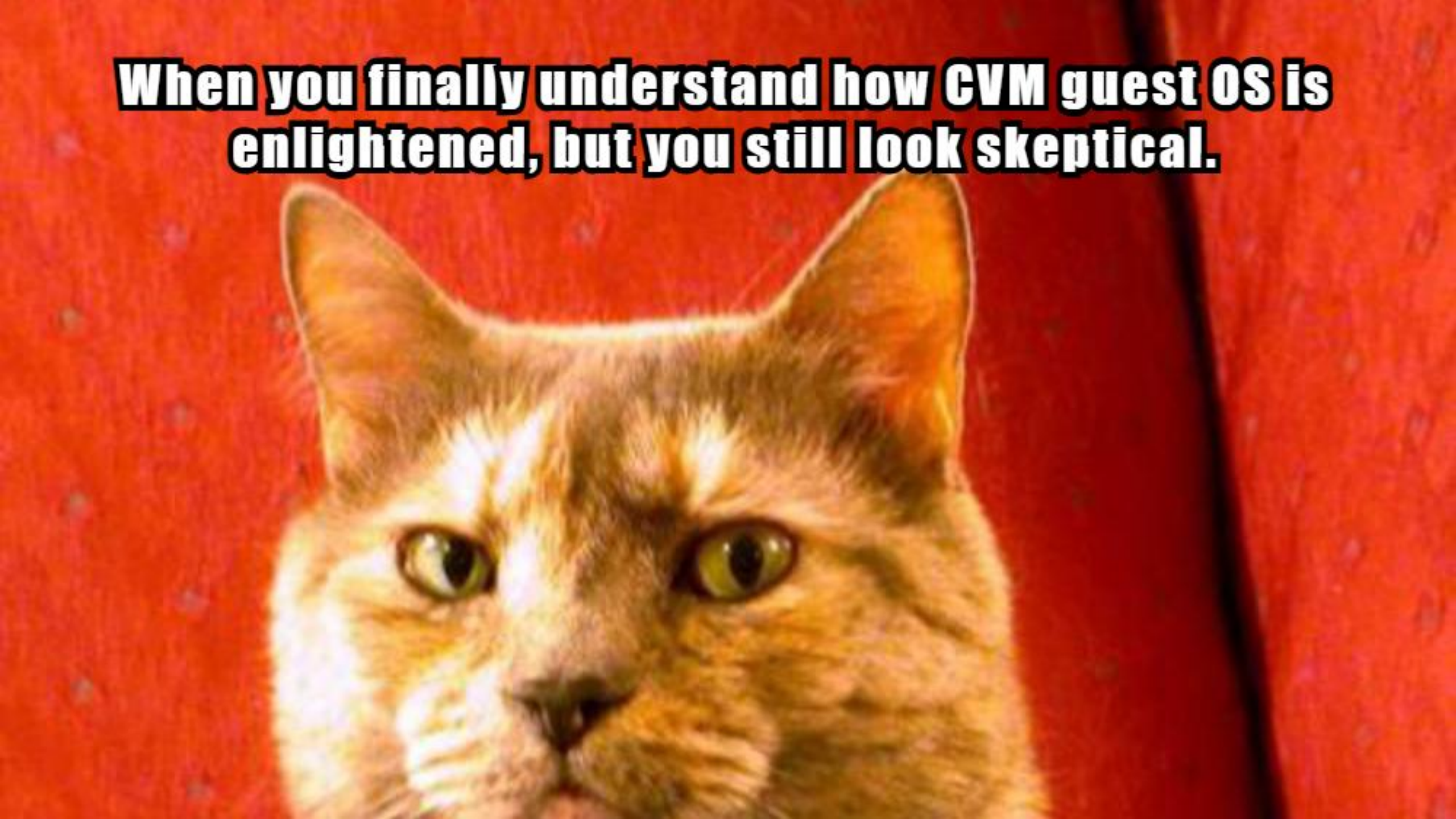
Full Disk Encryption and Decryption Flow



Microsoft

Key Takeaways

- ✓ CVMs – Protect the VM from Host and Hypervisor
- ✓ Don't need to fully trust the cloud provider
- ✓ **Regulatory Compliance** - Ensures the confidentiality and integrity of the system
- ✓ **Hardware agnostic** : UKI and other concepts can be applied to craft and secure any Linux Guest OS image


When you finally understand how CVM guest OS is enlightened, but you still look skeptical.

Microsoft

References

- Confidential VMs Explained: An Empirical Analysis of AMD SEV-SNP and Intel TDX
- Encrypted Virtual Machine Images for Confidential Computing - James Bottomley, IBM & Brijesh Singh – YouTube
- Securing Linux VM boot with AMD SEV measurement
- Introduction to confidential virtual machines
- RHEL confidential virtual machines on Azure: A technical deep dive
- FOSDEM 2024 - Linux on a Confidential VM in a cloud: where's the challenge?
- FOSDEM 2025 - Confidential VMs on public clouds and on-premise: a long way towards zero trust – 9:30 AM, SUNDAY

Microsoft

THANK YOU!