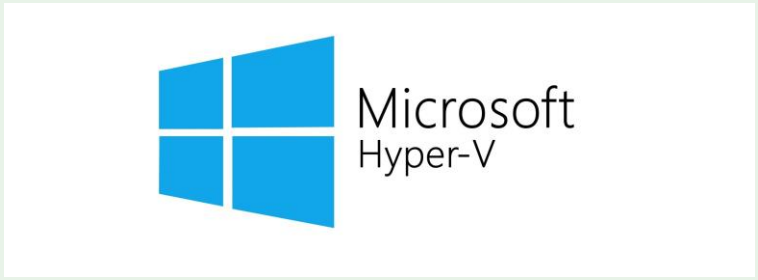
Linux Plumbers Conference 2024

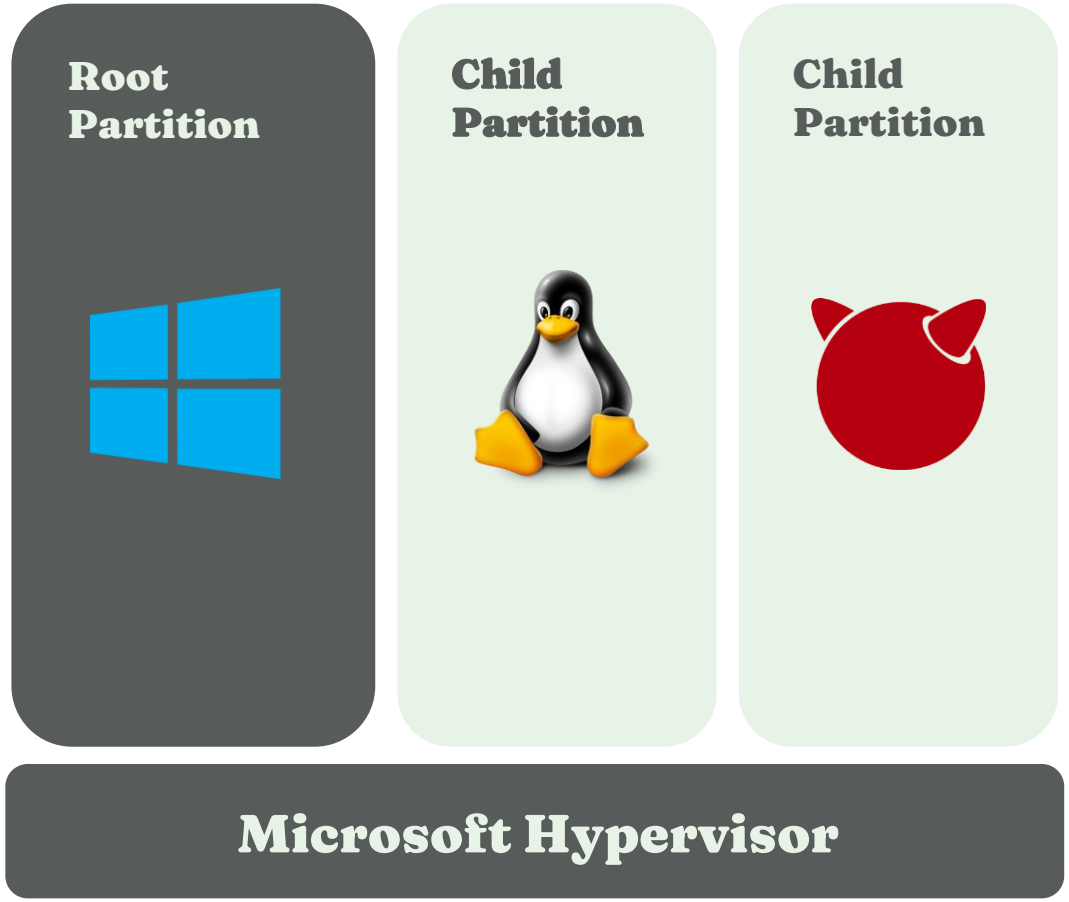
Beneath the Surface: Analyzing Nested CVM Performance on KVM/QEMU and Linux Root Partition for Microsoft Hyper-V/Cloud-Hypervisor

Jinank Jain
Muminul Islam
Linux Systems Group - Microsoft

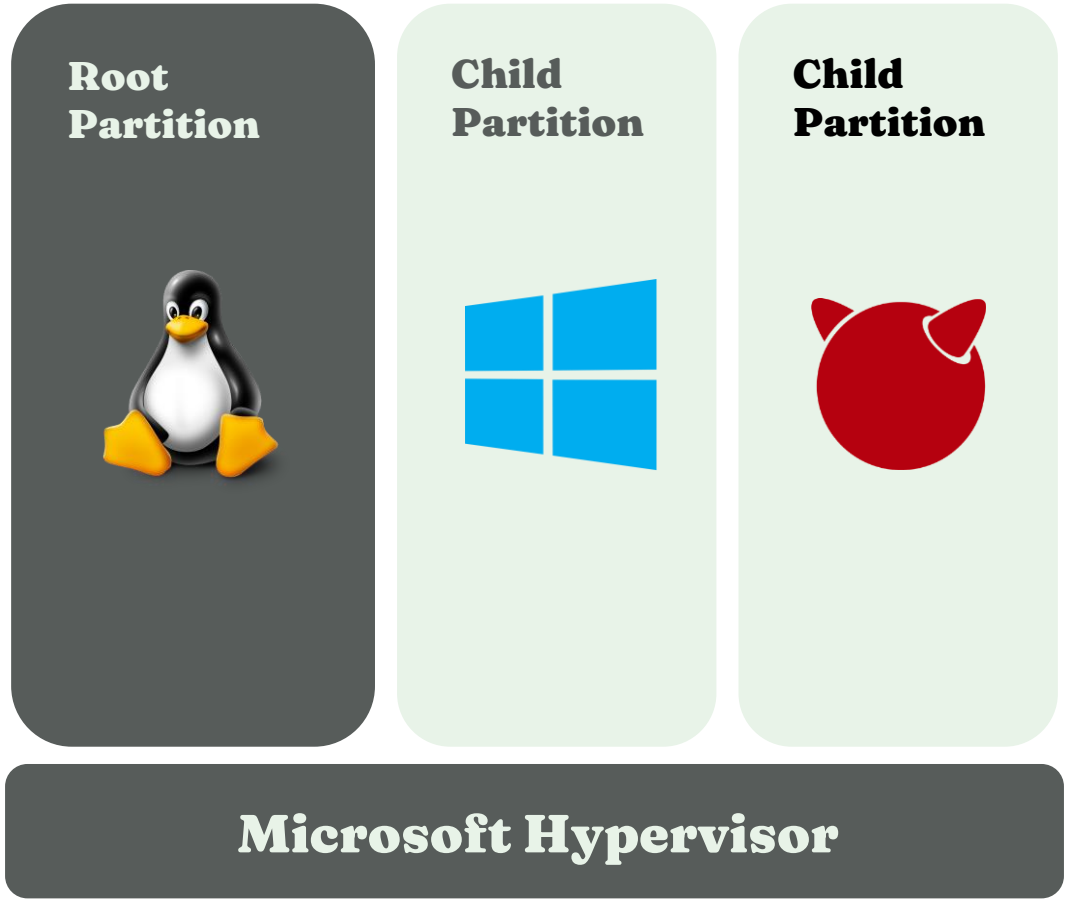


Microsoft Hypervisor Virtualization Stack

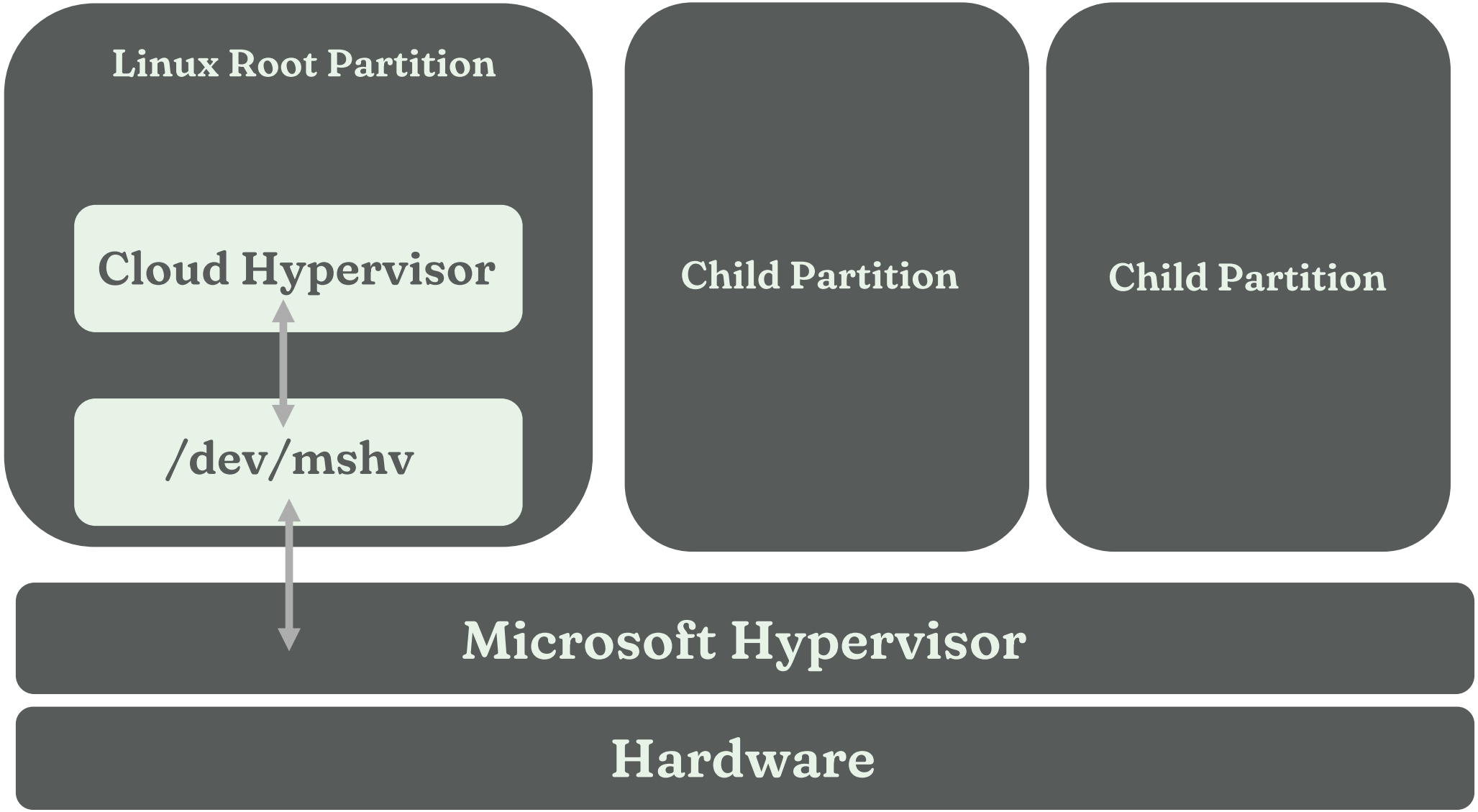
Windows Root Partition



Linux Root Partition

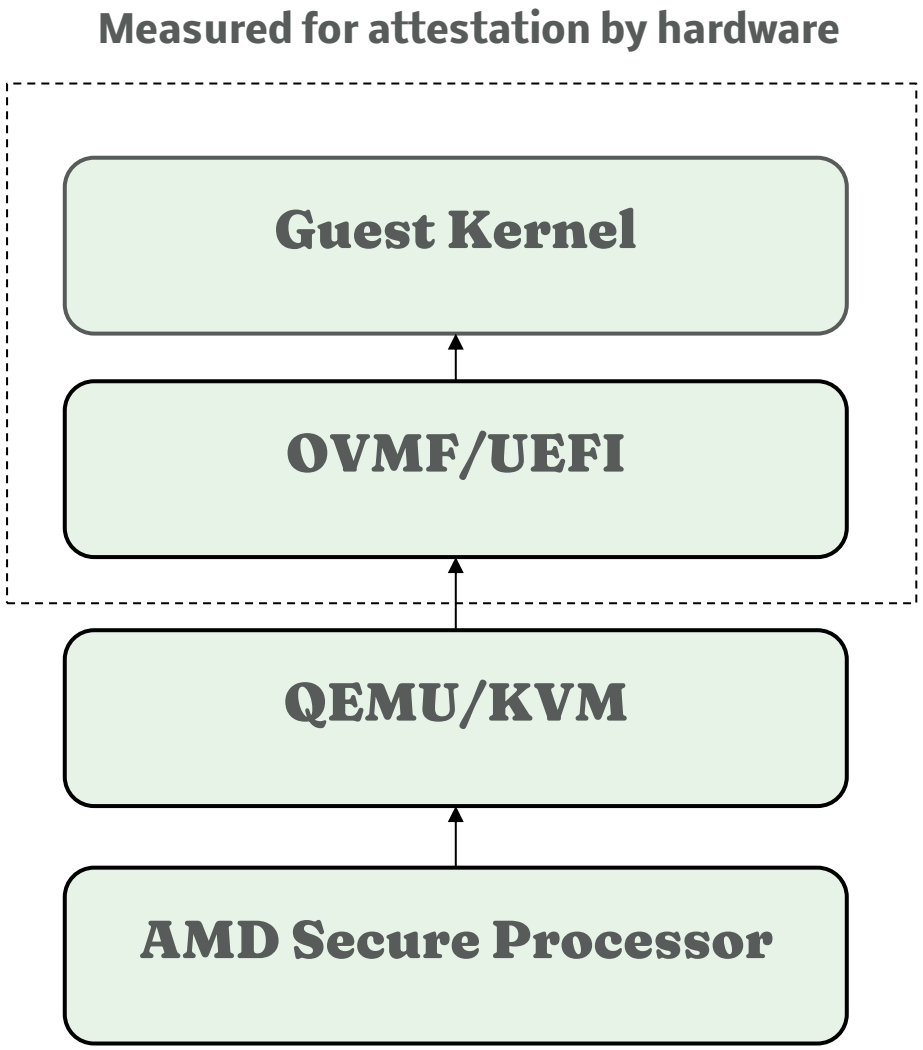


Linux MSHV Virtualization Stack



LKML: https://lore.kernel.org/lkml/1692309711-5573-1-git-send-email-nunodasneves@linux.microsoft.com/

Confidential VMs on KVM/QEMU



Attestation Report Generation

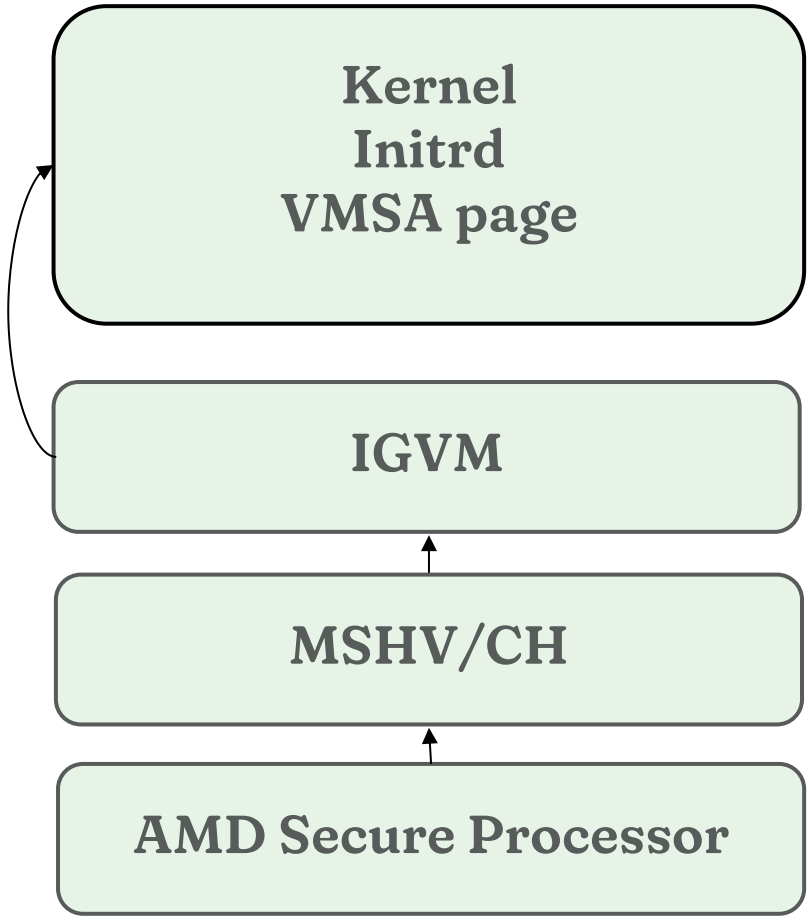
- Open source tools like **virtee/sev-snp-measure**

  $ sev-snp-measure --mode snp --vcpus=1 --vcpu-type=EPYC-v4 --ovmf=OVMF.fd --kernel=vmlinuz --initrd=initrd.img --append="console=ttySo loglevel=7“

  $1c8bf2f320add50cb22ca824c17f3fa51a7a4296a4a3113698c2e31b50c2dcfa7e36dea3ebc3a9411061c30acffc6d5a

- Need **access to OVMF source** or **binary blob** to generate the attestation report

Confidential VMs on MSHV/CloudHypervisor



Attestation Report Generation

- Open-source tools to generate IGVM file **microsoft/igvm-tooling**

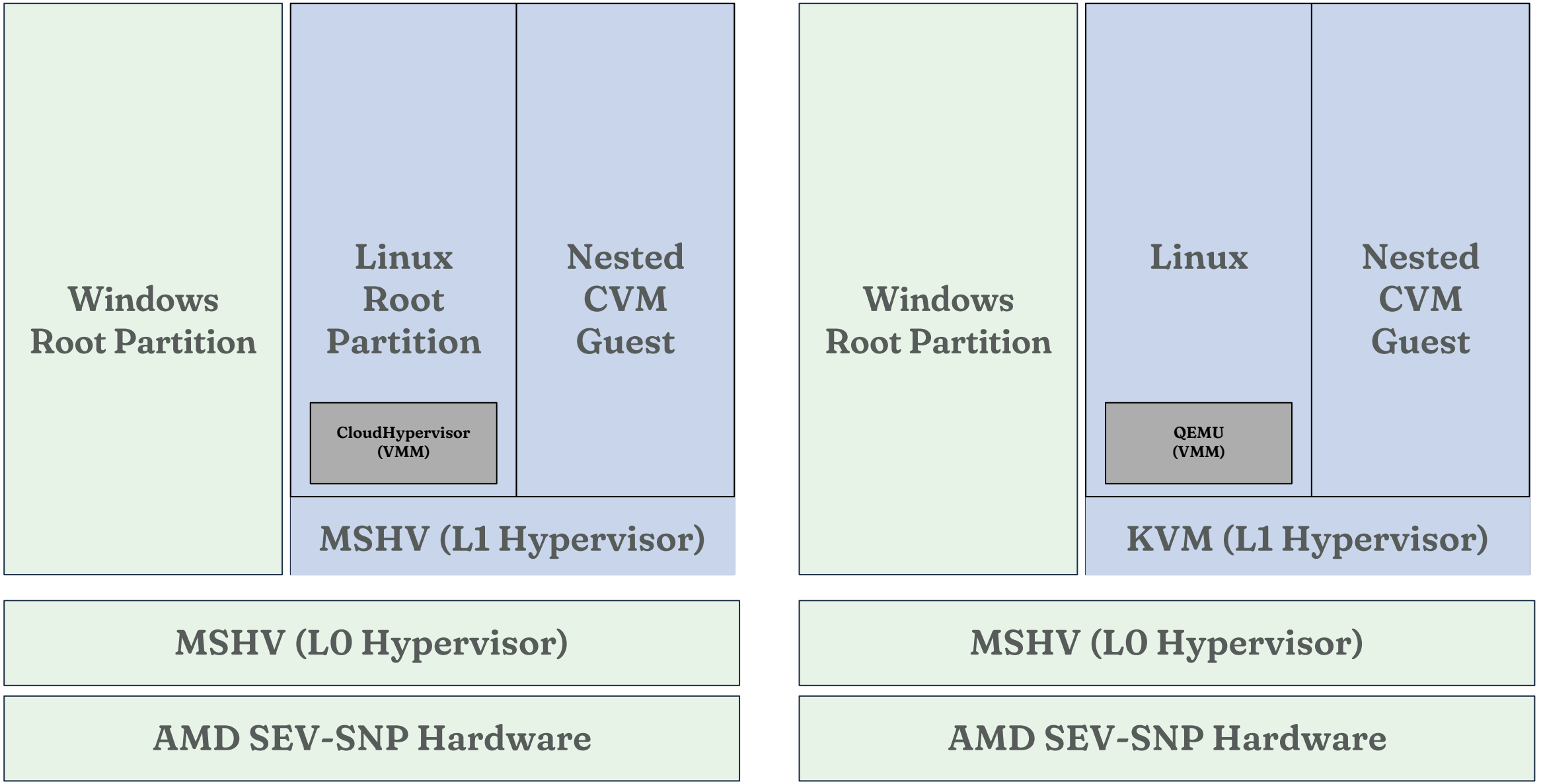
  $ python3 igvm/igvmgen.py -kernel $KERNEL -append $CMDLINE -boot_mode x64 -vtl 0 -o linux.bin -svme 1 -encrypted_page 1 -pvalidate_opt 1 -acpi $ACPI_DUMP

  ${'x-ms-sevsnpvm-guestsvn': '1', 'x-ms-sevsnpvm-launchmeasurement': '3d27b68ec2d5bdb630cb05edocbdad115c30caefd96ac2ae8753106813101f4948778661a91b09ce1f8dab865b8a79ce'}

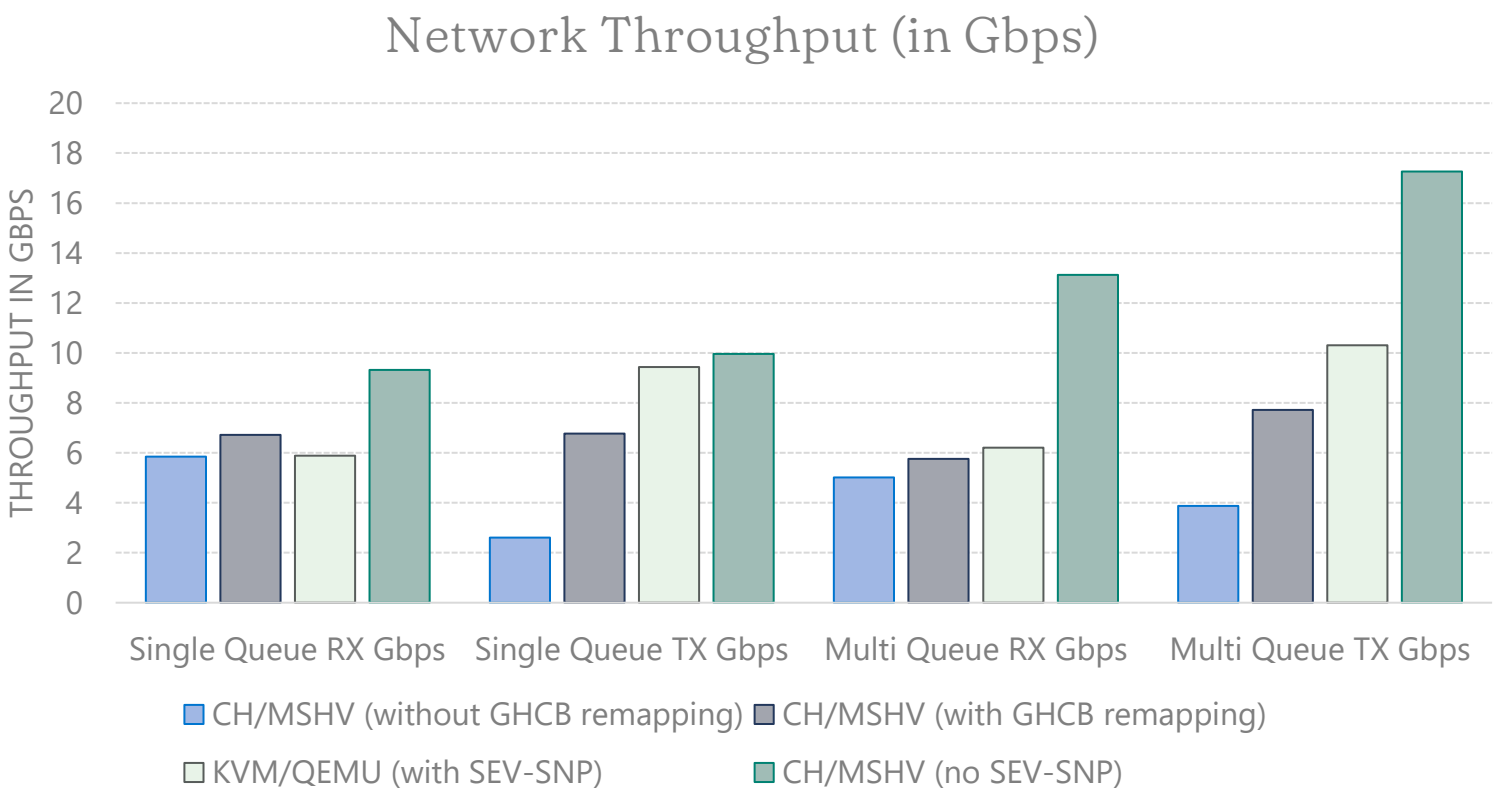
- Supports **direct kernel boot** and measurement file generated by the tool along with IGVM file creation
- Need **another tool generate ACPI table dump** which would be also required in this process.

Performance comparison between Linux/KVM/QEMU vs Linux/MSHV/CloudHypervisor

Performance test setup

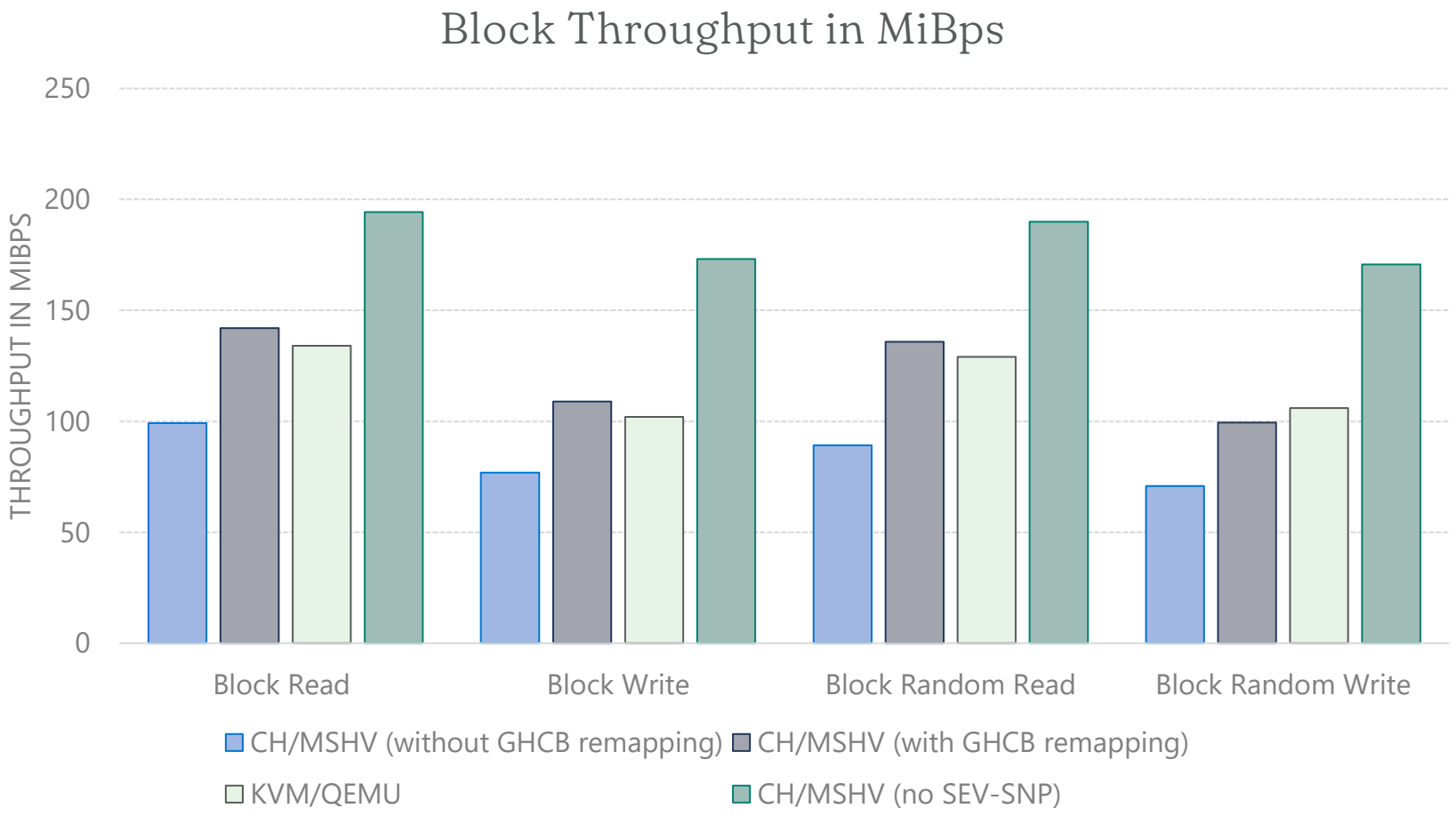


Network Throughput



- ## Hardware Setup
  - Processor: AMD EPYC 7763 64-Core Processor 2.45 GHz
  - Installed RAM: 1.00 TB
  - L1 VM: 16 CPU and 32 GiB RAM
- ## Single Queue Setup
  - L2 VM: 1 CPU and 4 GiB
- ## Multi Queue Setup
  - L2 VM: 8 CPU and 8 GiB

Block I/O Throughput



Hardware Setup

Processor: AMD EPYC 7763 64-Core

Processor 2.45 GHz

Installed RAM: 1.00 TB

L1 VM: 16 CPU and 32 GiB RAM

L2 VM: 8 CPU and 8 GiB

Disk Size: 4GB

CPU Stress Test

$ sysbench --test=cpu --cpu-max-prime=20000 run

KVM/QEMU Guest:

Number of threads: 1

Prime numbers limit: 20000

CPU speed:

events per second: 621.41

General statistics:

total time: 10.0021s

total number of events: 6216

Latency (ms):

min: 0.92

avg: 1.61

max: 4.13

95th percentile: 1.70

sum: 9995.15

Threads fairness:

events (avg/stddev): 6216.0000/0.00

execution time (avg/stddev): 9.9951/0.00

MSHV/CH Guest:

Number of threads: 1

Prime numbers limit: 20000

CPU speed:

events per second: 699.22

General statistics:

total time: 10.0013s

total number of events: 6994

Latency (ms):

min: 0.92

avg: 1.43

max: 6.03

95th percentile: 1.64

sum: 9994.40

Threads fairness:

events (avg/stddev): 6994.0000/0.00

execution time (avg/stddev): 9.9944/0.00

Discussion Points/Future Work

- Measurement of ACPI tables in guest attestation report
- Alternative CVM-native firmware like td-shim
- Boot time optimizations - hashing rate is 2ms per page, large guest take a lot of time to boot
- Live migration support for AMD SEV-SNP CVM on CloudHypervisor

Q & A?