Multi-Hypervisor Nested Virtual Machines
Yaohui Hu, Siddhesh Phadke, Kartik Gopalan, Michael R. Hines*
Computer Science, Binghamton University
*IBM Research Labs, Yorktown Heights
Contacts: yhu15@binghamton.edu, kartik@binghamton.edu

Traditional Nested Virtualization

Objective
- Develop systems support to run unmodified Nested VMs simultaneously on multiple hypervisors.

Level-2
- V1, V2, V3, and V4 are Level-2 VMs.
- V2 runs on H1 and H2.
- V3 runs on H2, H3, and H4.
- V4 runs on H4, and L0.

Level-1
- H1, H2, H3, and H4 are Level-1 hypervisors.
- V1 runs only on H1 as a traditional nested VM.

Level-0
- H0 Hypervisor

Benefits
- Users can run their own hypervisors in IaaS clouds.
- Live migration of hypervisors + guest VMs as a single entity.
- Hypervisor-level intrusion detection/prevention.

Limitation
- Presently, an L2 VM can run on only one L1 hypervisor at a time.
- L2 VM cannot run on multiple co-located L1 hypervisors.
- E.g. An L2 VM cannot simultaneously run on a commodity L1 hypervisor and another L1 hypervisor providing intrusion-detection.

Prototype and Performance

Current Status
- Developed prototype to run L2 VM on two L1 hypervisors running KVM/QEMU (as in V2)
- Memory mapping uses shadow-on-EPT mode.
- Ongoing work on Nested EPT mode
- Virtio-based I/O distribution.
- Ongoing work on direct assignment.

Resource Distribution

Multi-Hypervisor Nested VM (at L2)
- Process 1
- Process 2
- Process N
- VCPU 0, VCPU 1, vDevice 1
- Hypervisor 1 (at L1)
- Hypervisor 2 (at L1)
- L0 HPA

Memory Mapping

Shadow-on-EPT Mode
- L2 VA
- L2 GPA
- L1a GPA
- L1b GPA
- EPTa
- EPTb

Nested EPT Mode
- L2 VA
- L2 GPA
- L1a GPA
- L1b GPA
- Shadow EPTa
- Shadow EPTb

Challenges
1. Sharing the L2 memory across multiple L1s
   - Two modes: Shadow-on-EPT and Nested EPT.
2. Distributing L2 vCPUs across multiple L1s
3. Forwarding inter-processor interrupts (IPIs) across L1s
4. Distributing I/O using virtio or direct device assignment.

Prototype and Performance

Kernbench
- Run time (sec)
  - Std dev.
  - 136.15 8.09
  - 146.31 1.13
  - 634.70 8.79
  - 674.79 9.68
- % overhead vs. host
  - 7.5
  - 366.2
  - 395.6
- % overhead vs. guest
  - 333.8
  - 361.2
- % overhead vs. nested
  - 63
- %CPU
  - 97
  - 90
  - 100
  - 100

netperf
- Throughput (Mbps)
  - 940.5
  - 0.38
  - 343.92
  - 311.36
- % degradation vs. host
  - 1.1
  - 63.4
  - 66.9
- % degradation vs. guest
  - -
  - 63.3
  - 66.5
- % degradation vs. nested
  - -
  - -
  - 9.5

Funded by NSF