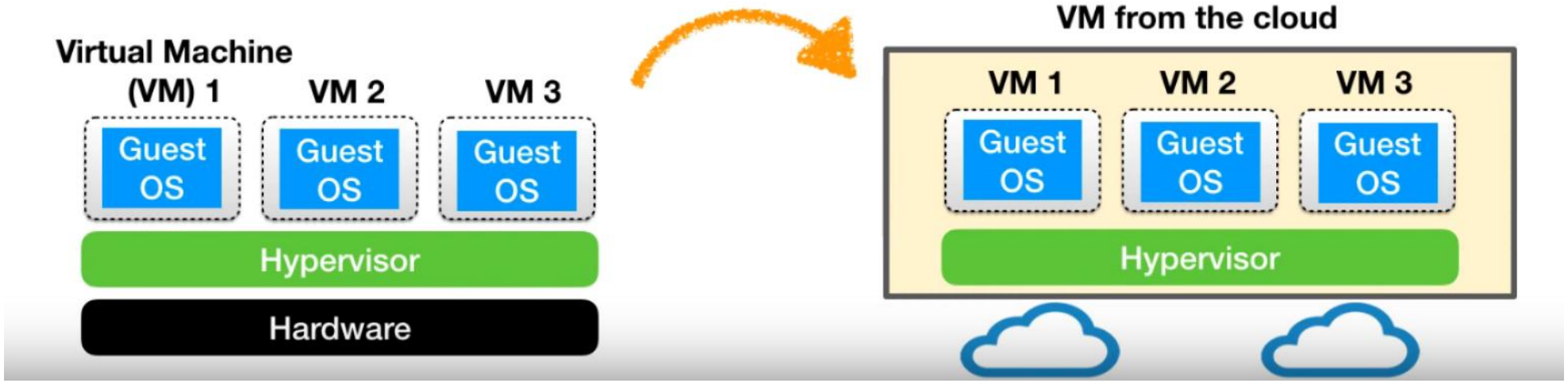
Optimizing Nested Virtualization Performance Using Direct Virtual Hardware

SOSP'20

Jin Tack Lim & Jason Nieh

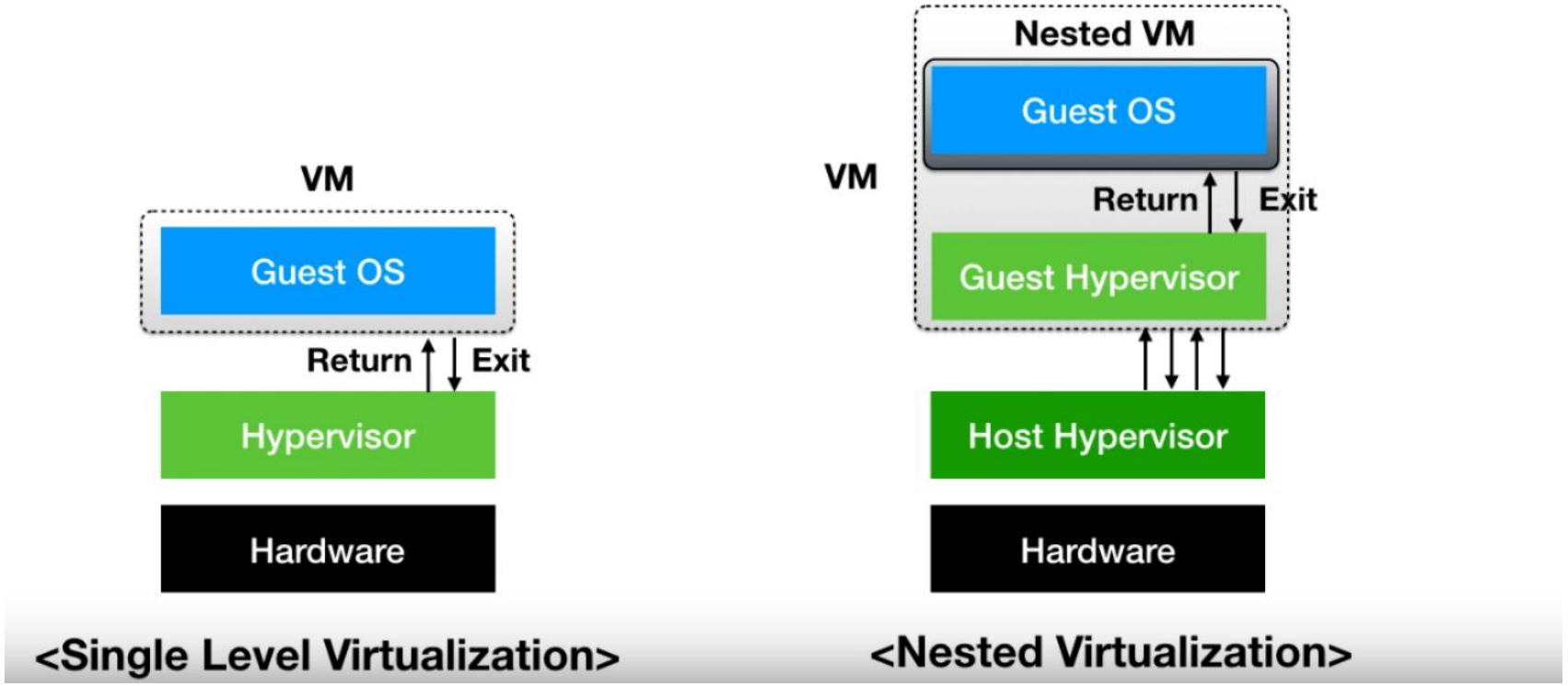
Nested Virtualization

- A technique to run a virtual machine (VM) inside a VM
- A key technology for cloud computing
  - Migrate workloads already having VMs to the cloud
  - Run OSes already leveraging virtualization in the cloud



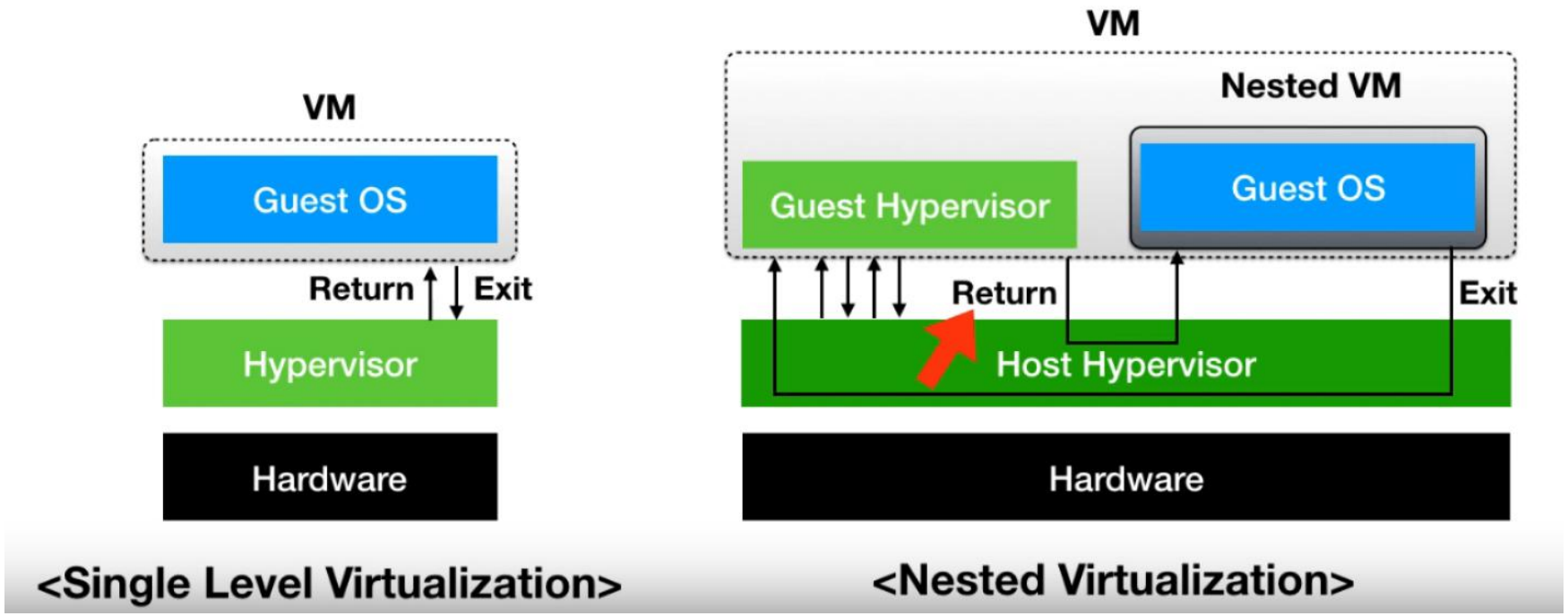
Nested Virtualization Performance

- Many times slower compared to native execution
- Single-level virtualization performs close to native execution



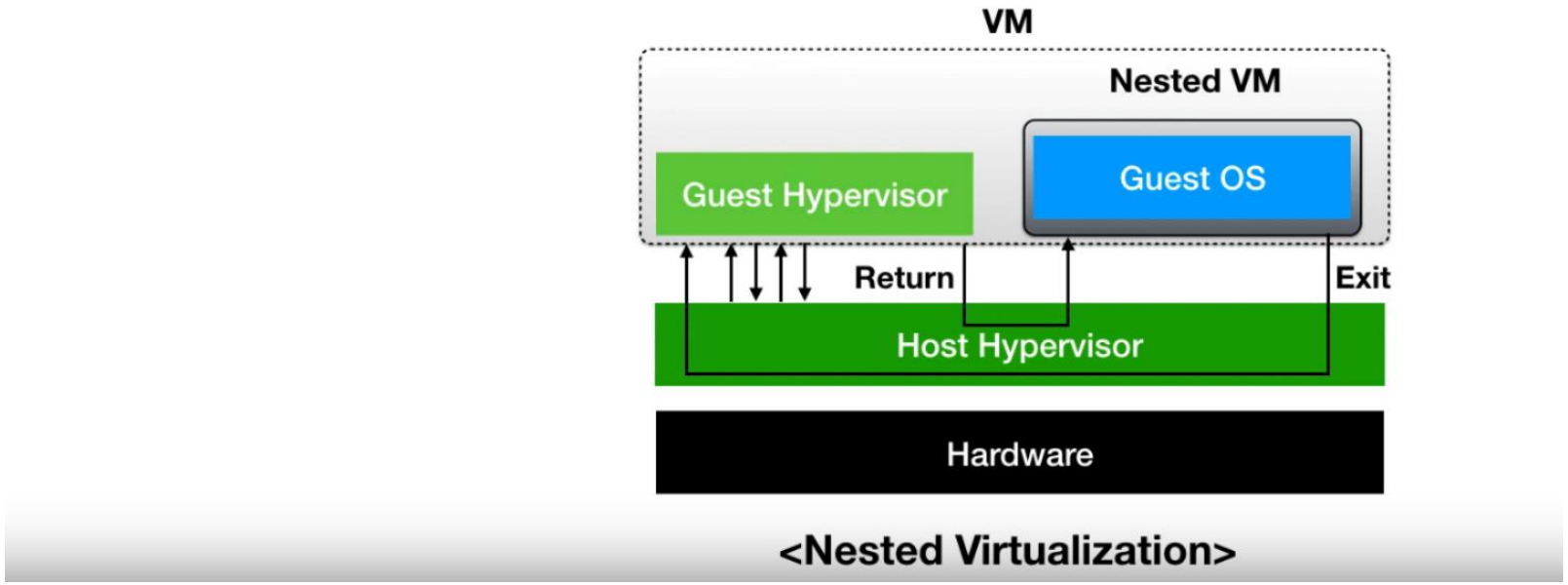
Nested Virtualization Performance

- Many times slower compared to native execution
- Single-level virtualization performs close to native execution



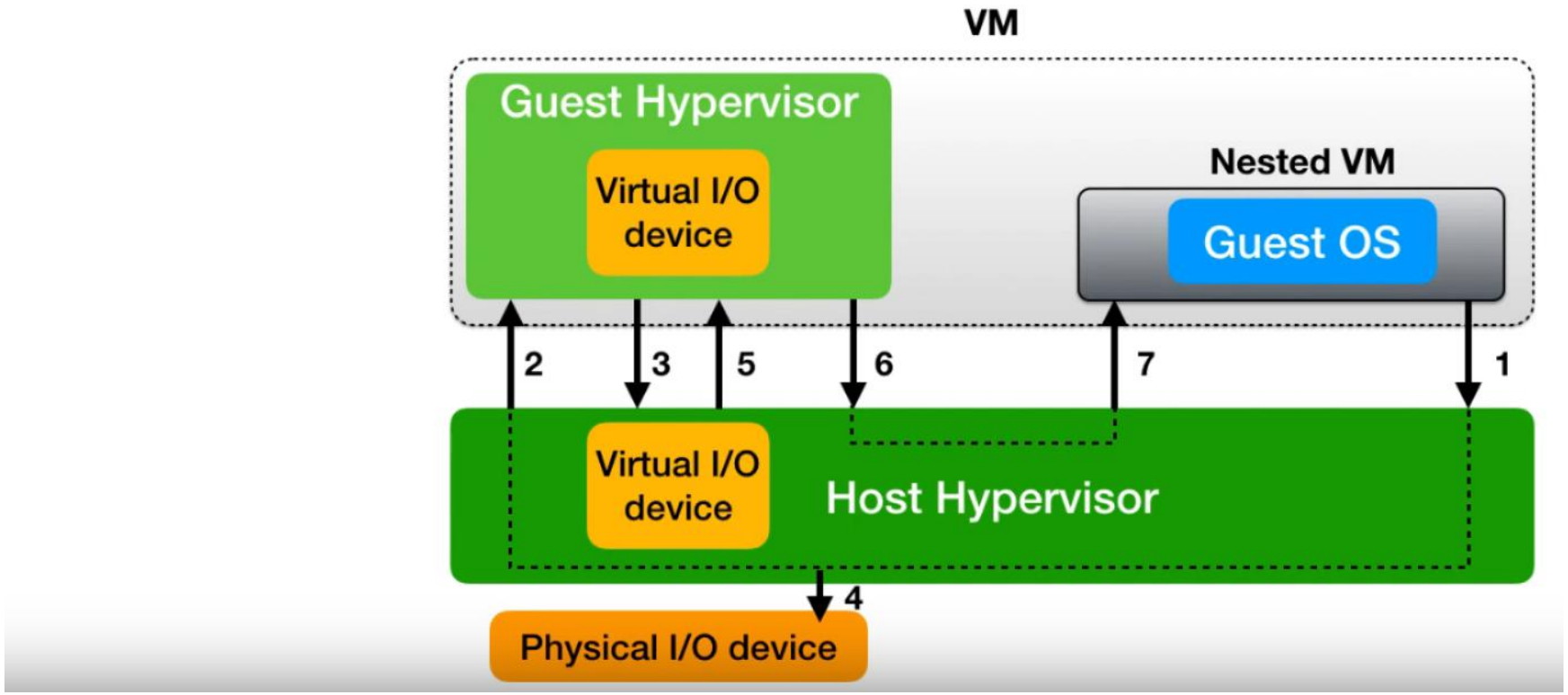
Exit Multiplication

- A single exit from a nested VM results in multiple exits to the host hypervisor



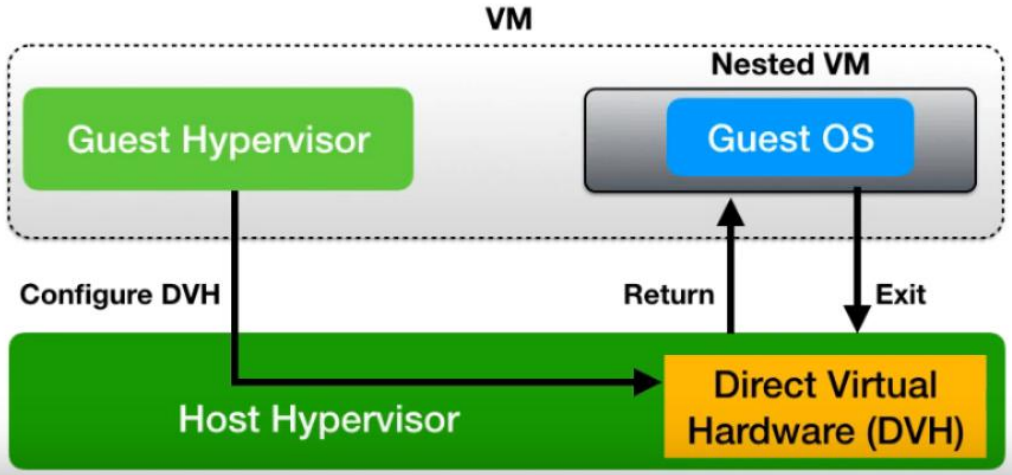
Virtual I/O Device for Nested Virtualization

- Sending data over network is expensive



Direct Virtual Hardware

- The host hypervisor directly provides **virtual** hardware to a nested VM
  - Only a single exit required
- The guest hypervisor configures the additional virtual hardware
- Transparent to a nested VM



Direct Virtual Hardware Benefits

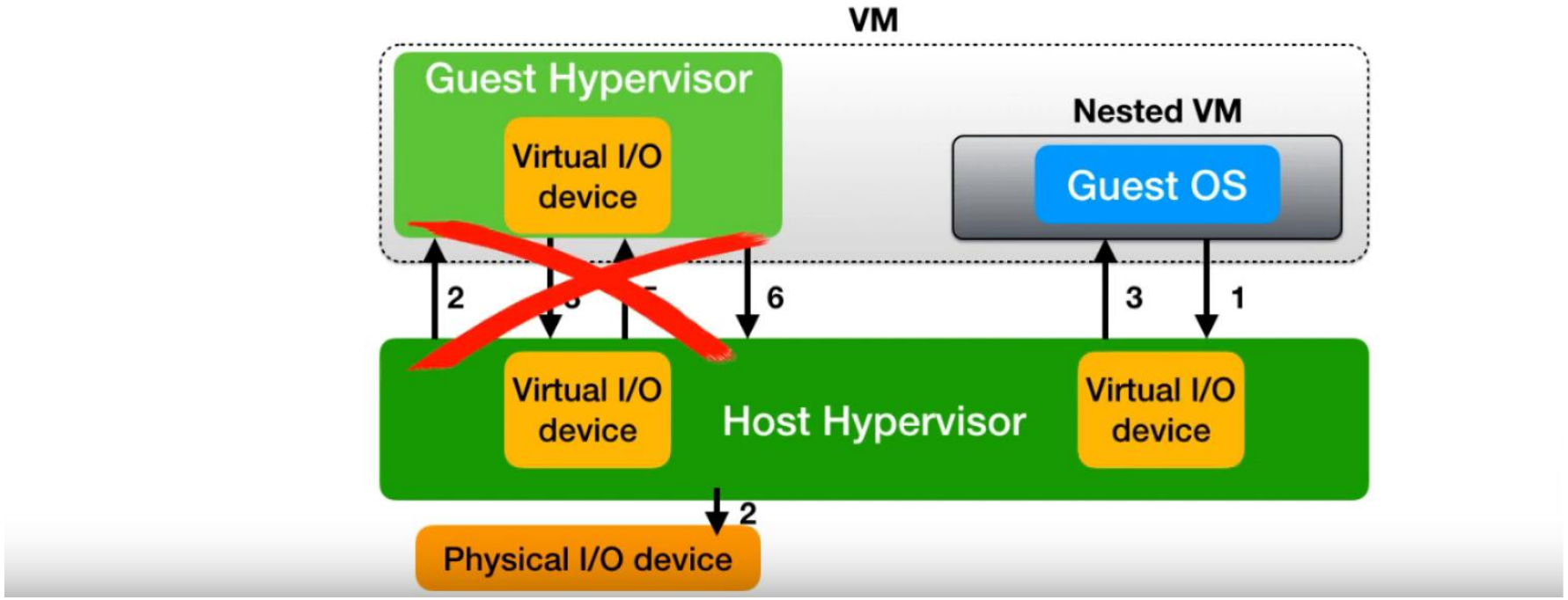
- Performance improvement with no exit multiplication
- Interposition in the host hypervisor
- Software-only - easy to deploy and scale

Direct Virtual Hardware Mechanisms on Intel x86

- **Virtual-passthrough**
- **Virtual timer**
- Virtual inter-processor interrupts (IPIs)
- Virtual idle

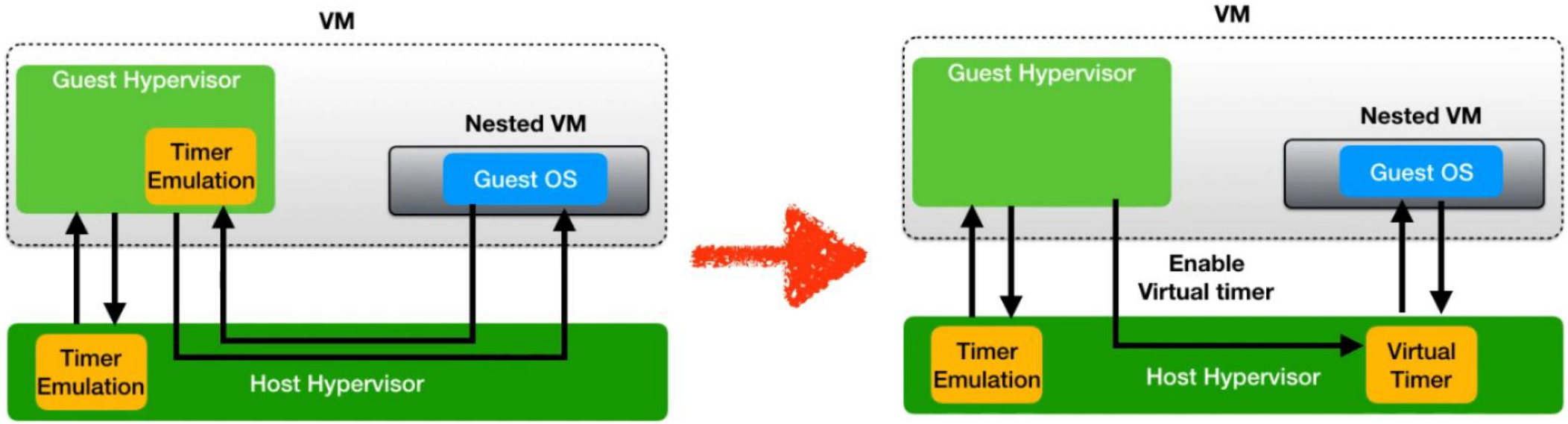
Virtual Passthrough

- Allow a nested VM to interact with the virtual I/O device provided by the host hypervisor
- Similar to passthrough, but with virtual I/O device instead of physical one



Virtual Timers

- Trapping instruction: programming timer (LAPIC timer)
- DVH solution: the host hypervisor provides Virtual LAPIC timer



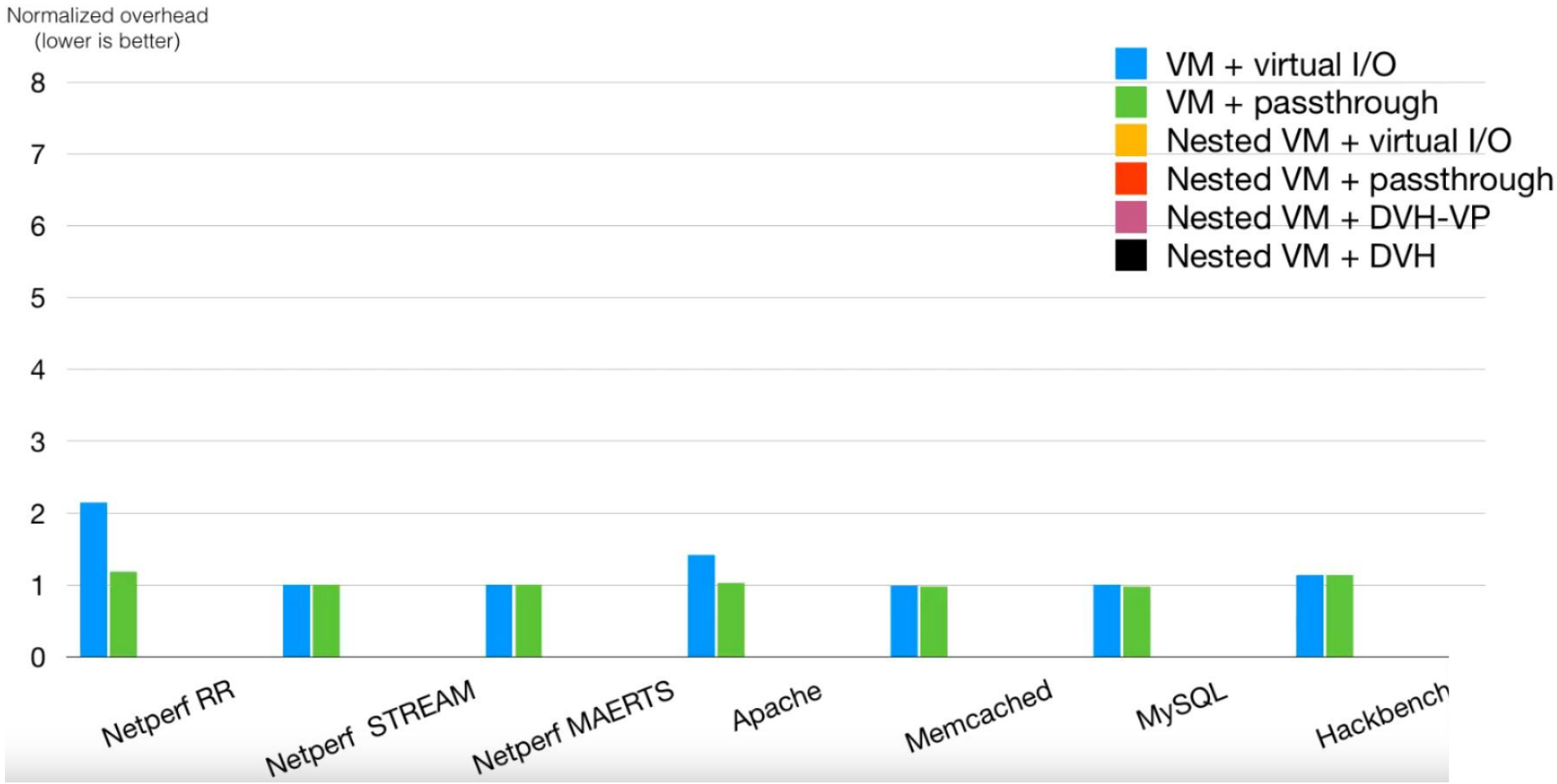
Application Benchmarks

Application	Description
Netperf TCP_RR	Network latency
Netperf TCP STREAM	Network bandwidth
Netperf TCP MAERTS	Network bandwidth
Apache	Web server
Memcached	Key-Value store
MySQL	Database management
Hackbench	Scheduler stress

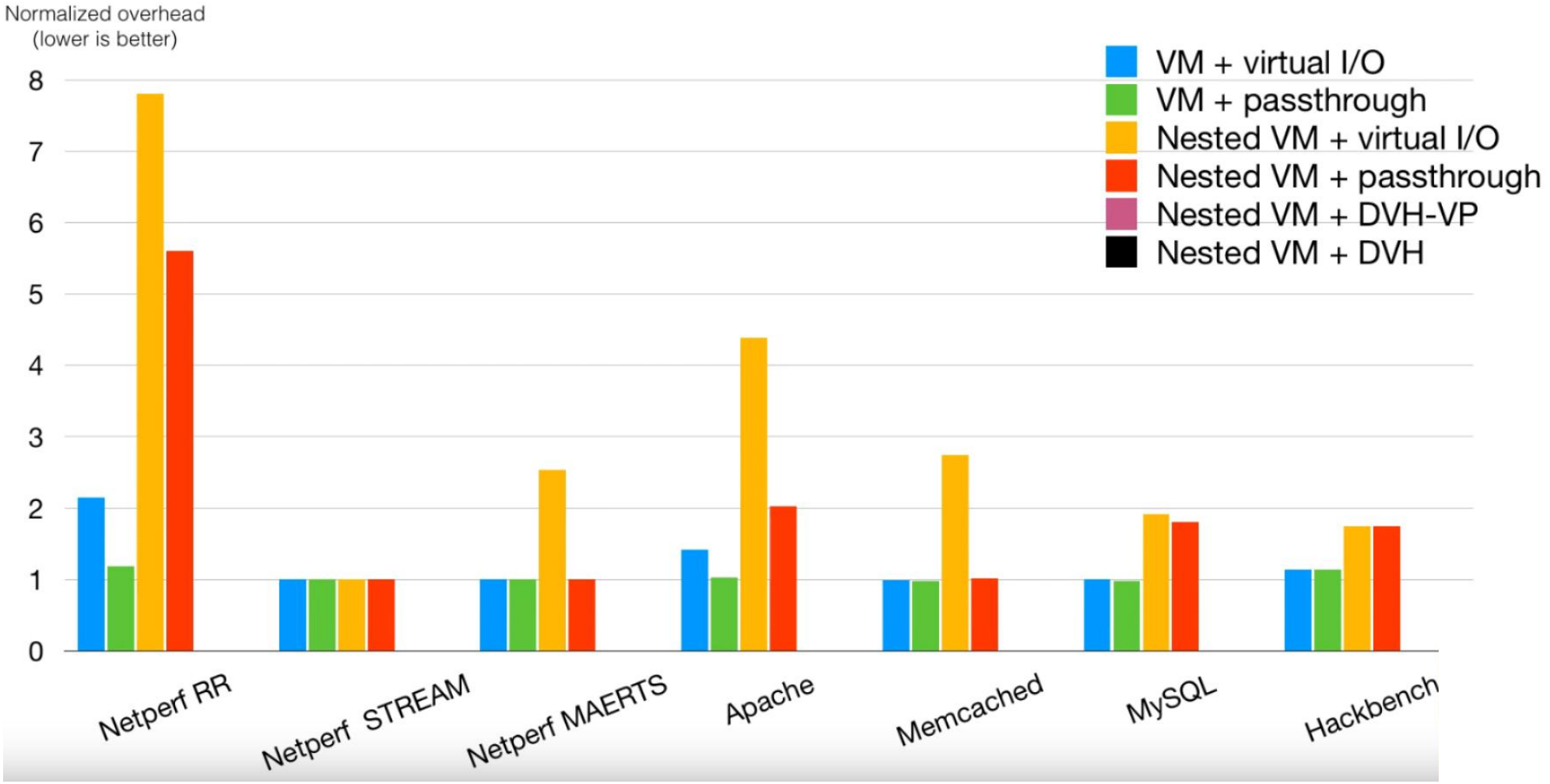
Experimental Setup

- Hardware
  - Intel Xeon Silver 4114, having VMCS shadowing
  - Intel X520-DA2 10Gb NIC
- Experiment configurations
  - 4-way SMP
  - KVM/QEMU, Linux
  - Virtio for virtual I/O devices

Application Performance



Application Performance



Application Performance