Accelerating Virtual Machine Storage I/O for Multicore Systems

Dongwoo Lee, Junghan Kim, Junghoon Kim, Changwoo Min, Young Ik Eom

Motivation and Goal

I/O workloads in VM suffer performance degradation due to virtualization overhead.

	I/O Request Cost	Device Emulation Cost	Completion Notification Cost
Direct Cost	VM-to-VMM world switching (Issue I/O requests to VMM)	Domain or user-kernel mode switching (Emulate requests in a separated domain or process)	VM-to-VMM world switching (Inject interrupts to guest and signal the completion of interrupt)
Indirect Cost	Cache pollution TLB flush	Cache pollution TLB flush	Cache pollution TLB flush
Synchronous Cost	ALL processes on a VM are stopped.		A process on a CPU is stopped.

Goal : Accelerating I/O performance in VM by reducing the cost induced by exits

Key Techniques

Reducing the I/O virtualization overhead by exploiting multicore architecture

I/O Request

Exitless I/O Request

- Using para-virtualized device driver that communicates with VMM through shared request queue.
- Issuing an I/O request is just enqueing a request in the shared queue.
- IOCore thread checks if there are new requests and dequeues them for further processing.

Device Emulation

In-VMM Device Emulation

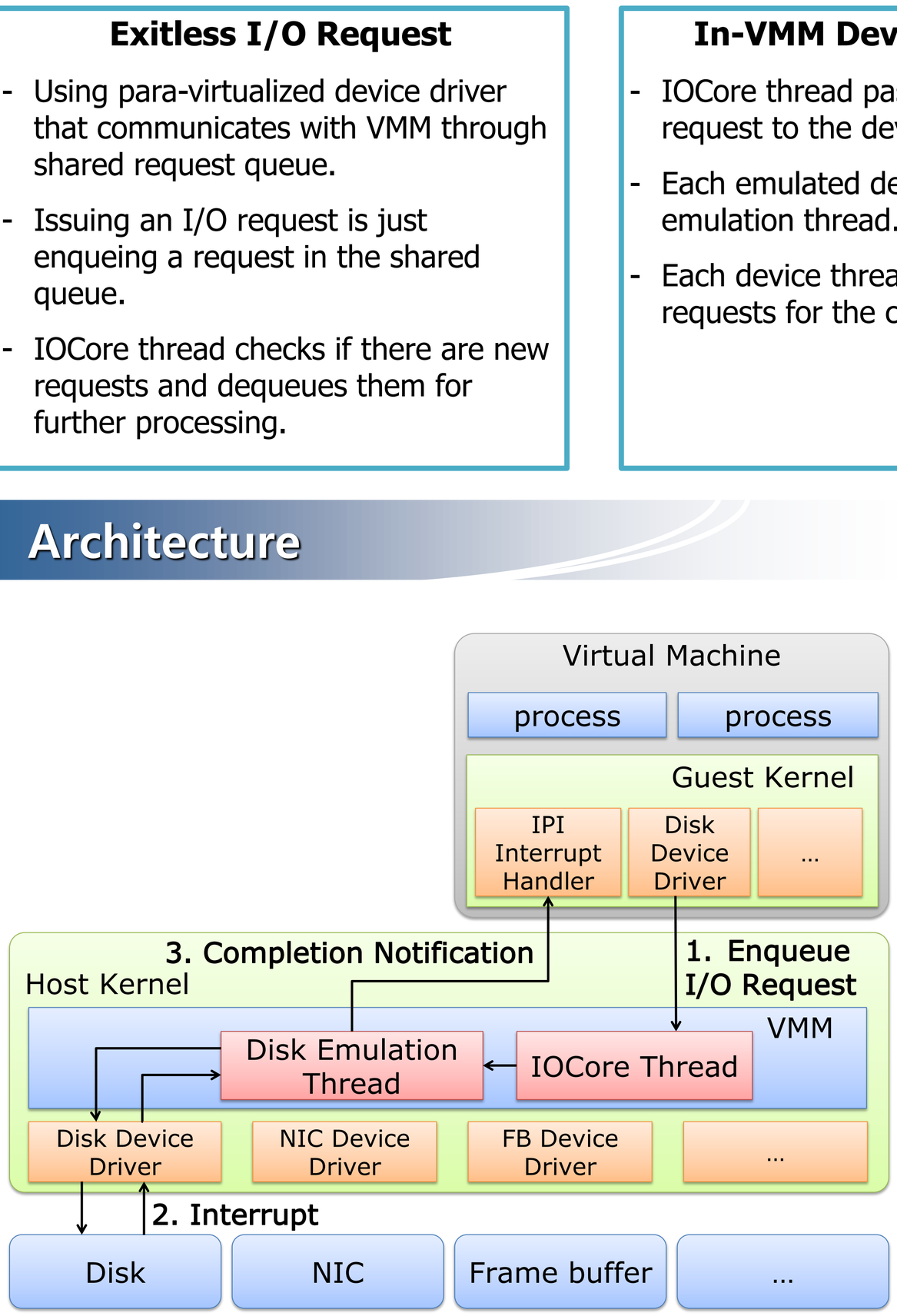
- IOCore thread passes the dequeued request to the device emulation thread.
- Each emulated device has its own emulation thread.
- Each device thread handles the I/O requests for the corresponding device.

Completion Notification

Exitless Completion Notification

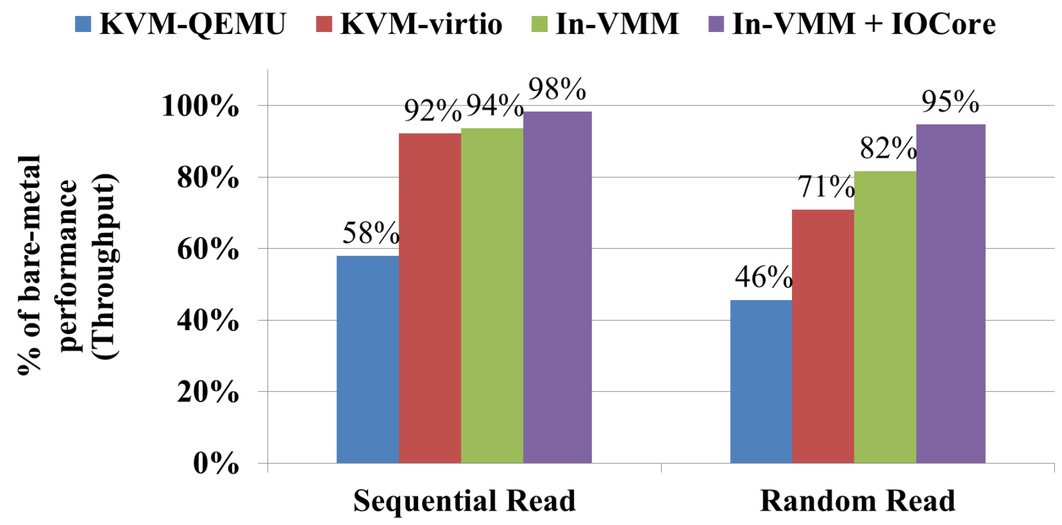
- IDT in guest is shadowed and configured to directly deliver IPI to the guest OS.
- Configuring x2APIC to directly expose EOI register.
- Guest OS writes the EOI registers without exits when guest completes interrupt handling.

Architecture



Experimental Results

Sequential reads and random reads on high-end SSD



Our approach highly close to the bare-metal performance
Seq. Reads by 98% and Rnd. Reads by 95%

In-VMM device emulation benefit: 2~11%

Exitless I/O Request benefit: 4~13%

Considering our prototype is **not optimized** and **exitless completion notification** is not implemented yet, the performance result is quite encouraging

For Further Information, please contact Dongwoo Lee (dwlee@ece.skku.ac.kr)