



Xen VT status and TODO lists for Xen-summit

**Arun Sharma, Asit Mallick,
Jun Nakajima, Sunil Saxena**

Outline

- **VMX Guests Status Summary**
- **Status**
 - Domain0 restructuring – PCI/IOAPIC
 - X86-64
 - VMX guests enhancements
 - QEMU Device Models
 - Paravirtualized drivers
 - Guest FW
- **TODO list**

VMX Guests Status Summary

- **VMX:**
 - Full support of 32-bit unmodified Linux guests without FW on IA-32 platforms
 - Includes support of standard PC platform devices
 - 32-bit unmodified Linux guest support without FW added to EM64T
- **Making Good progress on (Q2)**
 - Domain0 restructuring – PCI/IOAPIC
 - X86-64 support (64-bit unmodified guests)
 - VMX guests enhancements
 - Device models, paravirtualized drivers
 - Guest FW
- **In plan (Q3)**
 - SMP support for VMX guests, IPF VT guests, performance, installation

Domain0 restructuring – PCI/IOAPIC

- **Xen IO architecture today**
 - Xen enumerates PCI devices
 - Xen partitions the devices among domains
- **Adds complexity to Xen**
 - Complex PCI configurations, need for ACPI
 - Many workarounds for PCI hardware bugs needed
- **Proposed Solution**
 - Move all PCI device enumeration code to Domain0
 - Including full support of ACPI
 - Also requires enabling PCI, Local and I/O APIC

Domain0 restructuring – PCI/IOAPIC (cont.)

- **Splitting of Resources and Device Enumeration**

- Xen
 - Owns Local APIC, IOAPIC including assignment of interrupt vectors
 - ACPI tables parsing only (no AML interpreter)
- Domain0
 - Full ACPI support
 - Need to map ACPI tables into domain0
 - Xen passes ACPI table pointer in `shared_info_t`
 - Local APIC
 - Stubs to satisfy MP parsing
 - IO APIC
 - Accessed by Domain0 using hypercalls
 - Primarily to set interrupt redirection entry
 - PCI Device Enumeration

Domain0 restructuring – PCI/IOAPIC (Status)

- **A patch against xeno-unstable.bk exists and boots on some machines**
- **Need to validate on a wider variety of machines**
 - Especially ones with complex PCI configurations
- **Other outstanding works**
 - MSI support
 - Need to covert Domain0 IRQ to vector (hypercall)
 - Partitioning of devices between domains
 - PCI Hotplug support
 - Hotplug of I/O APICs?

X86-64 Status

- **It's in unstable BK tree**
- **Stable and usable**
 - Multi-user mode, e.g. login from remote and build the kernel
- **What's left**
 - DomU support. Tested to build 32-bit VMX domain
 - SMP support, 32-bit binary support, IOMMU, writable page tables, NUMA support?
 - Performance tuning
 - Running on big machines (# of CPUs, memory, I/O)
- **Need help from community**
 - Actively build, test, use.
 - Finish what's left

X86-64 – Restructuring

- **Need to have common chipset/platform code among {x86-32, x86-64} x {Linux, Xenolinux} for better code maintenance**
 - No compelling reasons for them to be slightly different: time.c, smp.c, smpboot.c, several headers
 - Shared files with x86 as much as possible, but not enough
 - Timer, PCI, microcode, etc.
 - We can use the same code for x86-32 and x86-64 more for xenolinux
 - Initialization
 - Chipset/platform support – PCI, APIC, ACPI, interrupt management, most of trap handling, timer management
 - Processor specific part – CPU detection, most of context switching
 - SMP support

X86-64 Status – VMX guests

- **32-bit VMX guests are available on both 32-bit and 64-bit VT-x Xen**
- **Working on:**
 - 64-bit guests support
 - 4-level/3-level shadow mode support
 - Enhanced 32-bit support
 - PAE, eXecute Disable (XD) support on 64-bit VT-x Xen
 - Extend shadow code to support 4-level cleanly to support both types of guests
 - SMP guest support
 - Local and I/O APIC support
 - Test more on MP systems

VMX Guests Enhancements: Device Models

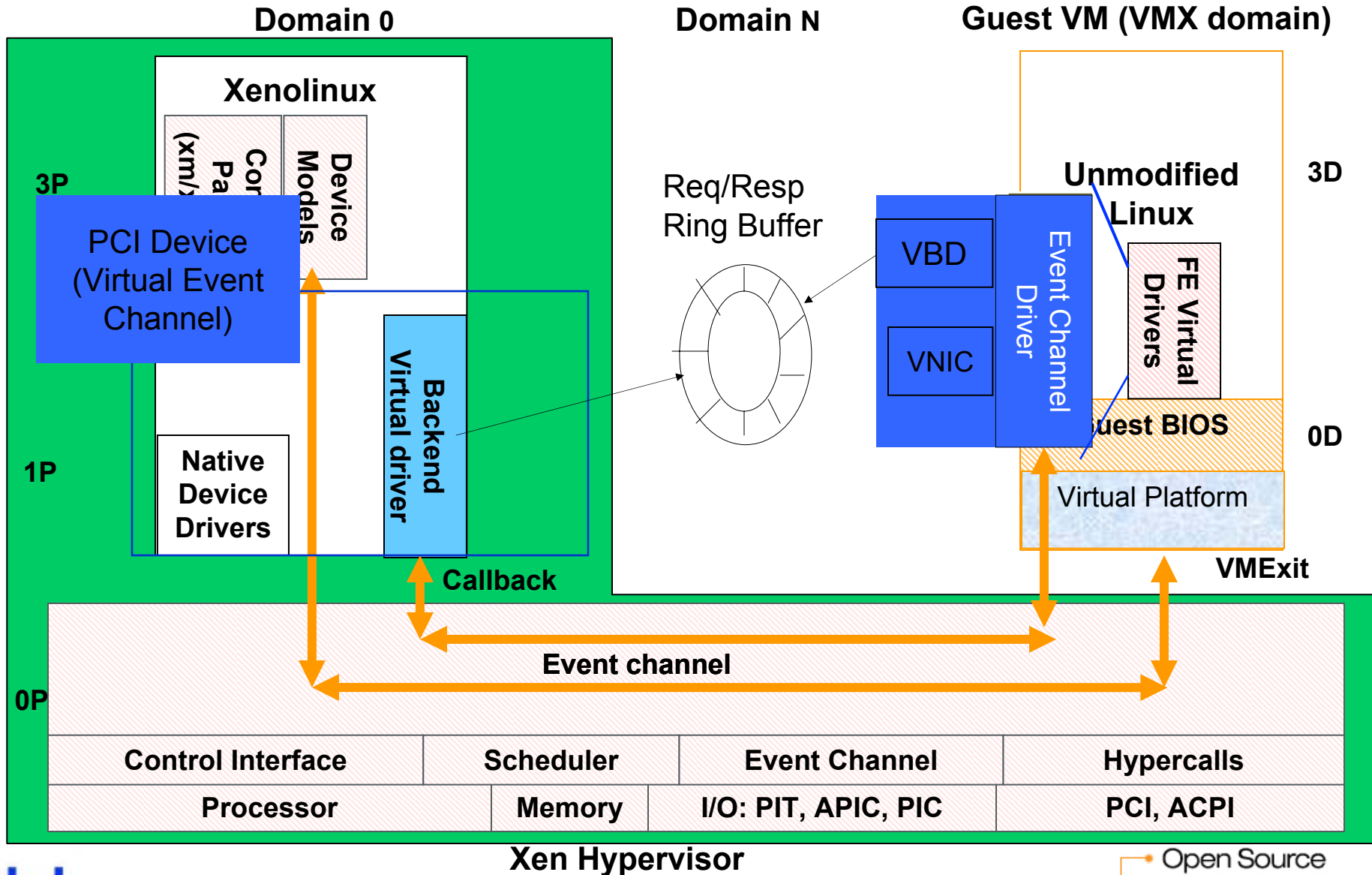
- **What we have today**

- Bochs device models providing
 - PS/2 kbd/mouse, CMOS/RTC, Timer (8254), Generic PCI bridge, EIDE, VGA, Serial ports, NMI Ports, Floppy, NIC NE2K (ISA), Local APIC, IOAPIC, CDROM, USB
 - Gaps: **SCSI, PCI NIC, IDE-DMA**

- **What we are planning to do: switch to QEMU**

- Get PCI NIC and IDE-DMA for free
- Code written in C
- Need to implement some functionality (for eg: APIC) either from – unstable or Bochs
- Status: Implementation complete, under regression/performance testing.

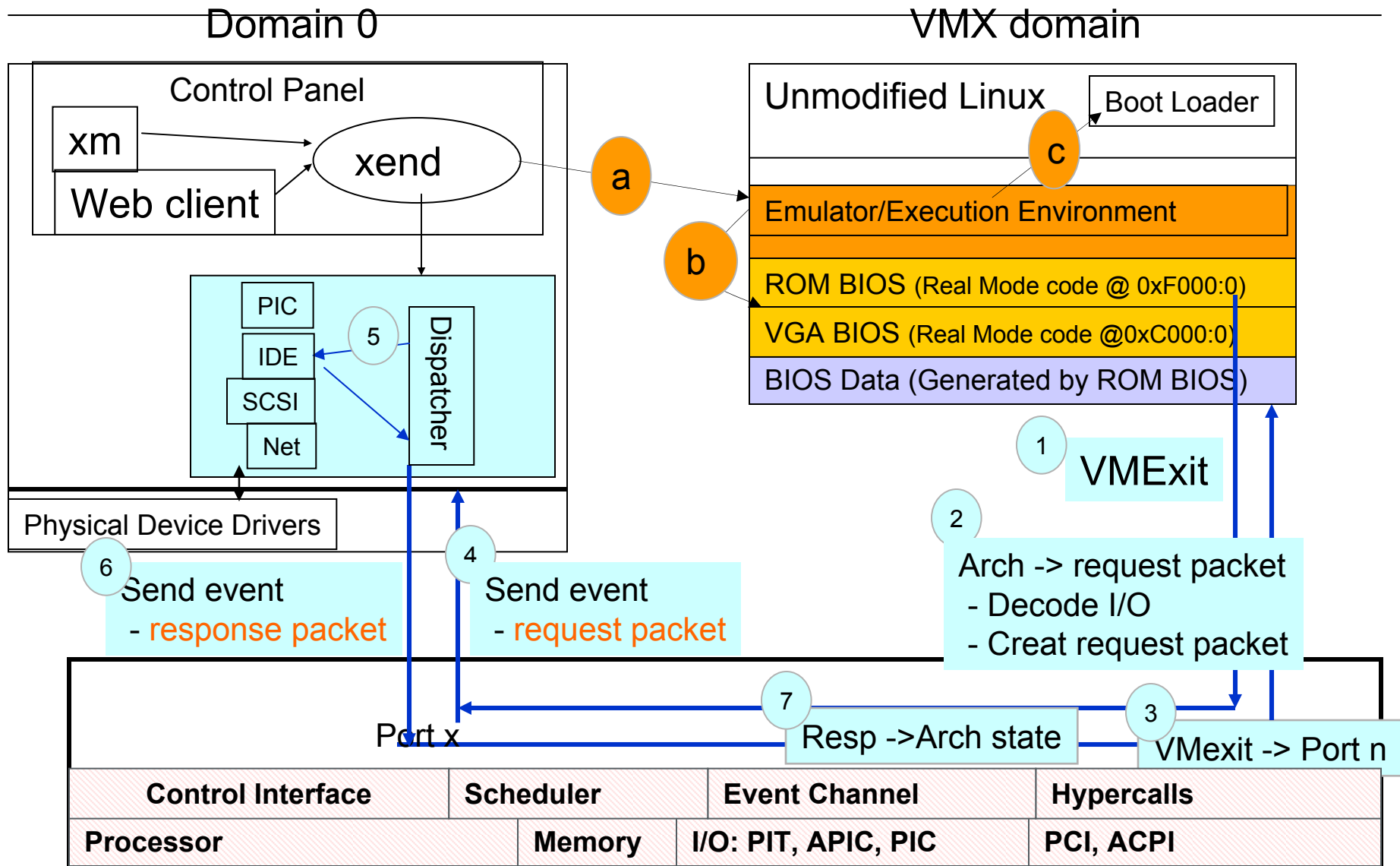
VMX Guests Enhancements: VBD/VNIC



VMX Guests Enhancements: VBD/VNIC Status

- **Basic implementation complete**
 - Not integrated with PCI virtual event channel device model
 - Testing/performance analysis in progress
- **Work remaining**
 - Use of grant tables instead of mpa

Guest FW Architecture



Hypervisor

Guest FW Status

- **Real Mode Emulator**

- Implementation complete and integrated into Guest FW execution environment

- **ROM BIOS**

- Using BOCHS/Plex86 ROM BIOS
- Extended to E820 memory map to be dynamic
- Adding ACPI support

- **VGA BIOS**

- Using BOCHS/Plex86 VGA BIOS

- **Control Panel Enhancements**

- Enhanced to load Execution environment, ROM BIOS and VGA BIOS in a VMX domain

TODO

- **> 4 Gig memory support for 32-bit Xen and 32-bit domains**
 - PAE support
- **Discontiguous memory support in Xen**
- **32-bit/64-bit MP Guests**
- **Debugger / Crash dump**
- **Performance monitoring tools**
- **Scheduler enhancements for load balancing**
- **Power-management**