Xen, KVM Virtualization

Κώστας Δρόγγος
costas.drogos@gmail.com

Hellug Event, Πανεπιστήμιο Πειραιά

26 Μαρτίου 2010

Γιατί Virtualization;

- Οικονομία.
- Ευελιξία.
- Fun :)

Ποιο απ'όλα όμως;

- Hardware-Assisted Virtualization - KVM
- Paravirtualization - Xen
- Shared-Kernel Virtualization - openVZ
- Full Virtualization - Qemu, Virtualbox

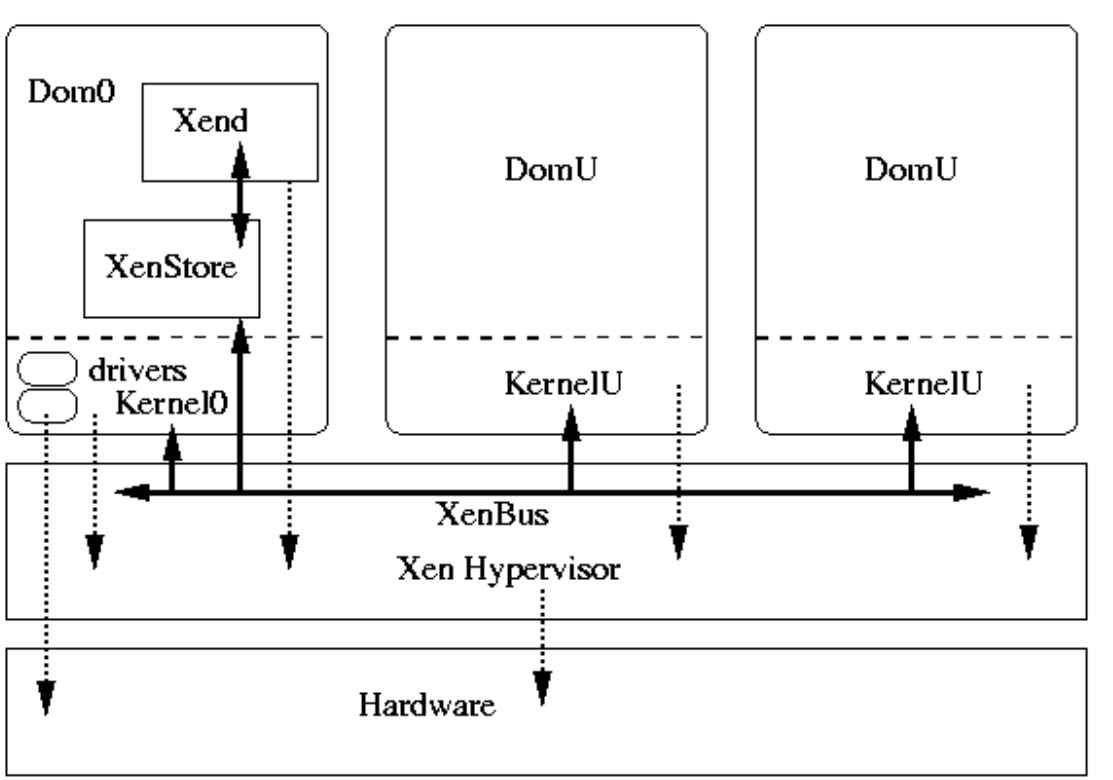
Paravirtualization

“Paravirtualization is a virtualization technique that presents a software interface to virtual machines that is similar but not identical to that of the underlying hardware.”

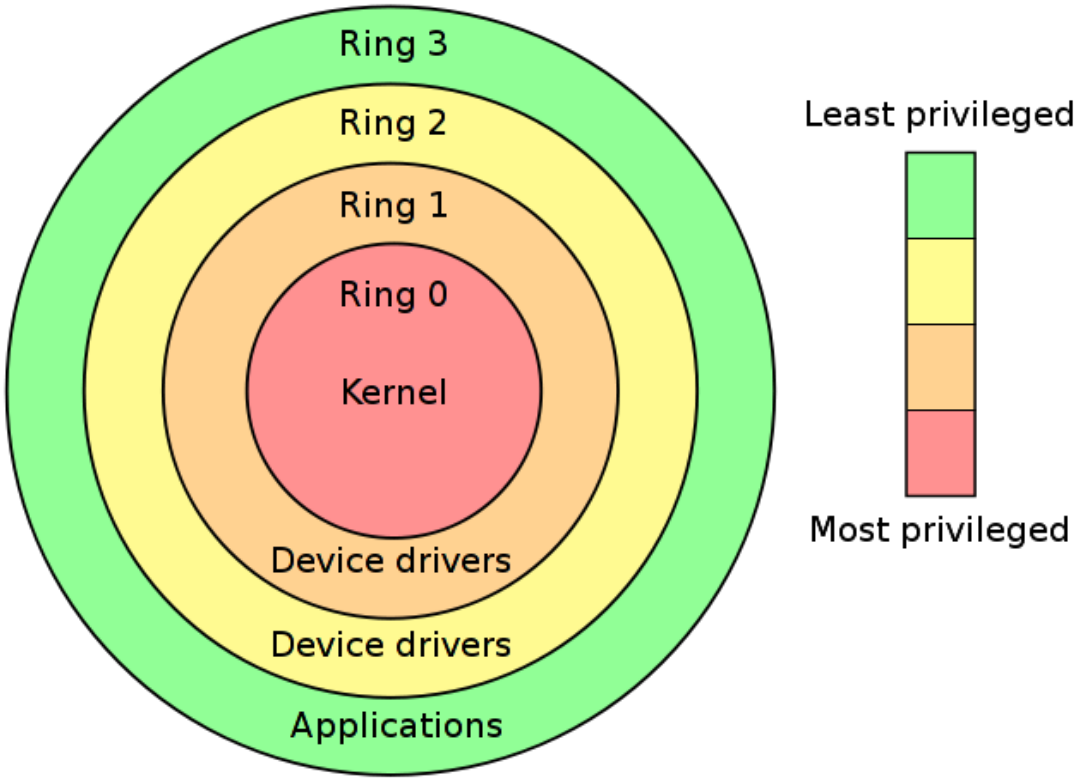
Δομή

- Hypervisor
- Host(dom0) OS
- Virtual hosts (domUs)

Δομή


Dom0
Xend
XenStore
drivers
Kernel0
DomU
KernelU
DomU
KernelU
XenBus
Xen Hypervisor
Hardware

Execution Rings x86



Hypervisor

- Ένα μικρό λειτουργικό
- Υπεύθυνο για scheduling, memory management
- Αλλιώς γνωστό και ως VMM (Virtual Machine Monitor)
- Τρέχει στο ring 0

dom0

Ένα privilleged VM.

- Το λειτουργικό boot-άρει σε protected mode.
- Τρέχει στο ring 1 (x86)
- Device Drivers
- Userspace Interface to Hypervisor
- Xen Drivers Backend

pv domUs

- Τρέχουν στο ring 3
- Δεν έχουν κανονικές συσκευές
- Network & Block Drivers Frontend

HVM domUs

- Hardware assisted machines running inside modified-qemu
- Optional paravirtualized drivers
- Guest OS can be a non-xenified one (e.g. Windows)

Hypervisor - Domains Communication

Το αντίστοιχο των syscalls για την επικοινωνία hypervisor και domains.

- Στο kernel userspace, τα syscalls εκτελούνται με το soft interrupt 0x80
- Οι xen kernels, έχουν άλλο ένα soft interrupt, το 0x82 που οδηγεί σε εκτέλεση hypercall (deprecated)
- Πλέον, χρησιμοποιείται ένα page (hypercall_page), σε κάθε guest
- 38 Hypercalls και 8 ελεύθερα για μελοντική χρήση.

Inter-Domain Communication

- XenBus
- xenstore
- Shared pages (Grant Tables)
- Asynchronous Event Channel

Inter-Domain Communication

- XenBus
- xenstore
- Shared pages (Grant Tables)
- Asynchronous Event Channel

Inter-Domain Communication

- XenBus
- xenstore
- Shared pages (Grant Tables)
- Asynchronous Event Channel

Inter-Domain Communication

- XenBus
- xenstore
- Shared pages (Grant Tables)
- Asynchronous Event Channel

Two-parts pv Drivers

Network and Block device

- Backend on dom0 (usually)
- Frontend on domU (usually)

VT extensions

Intel

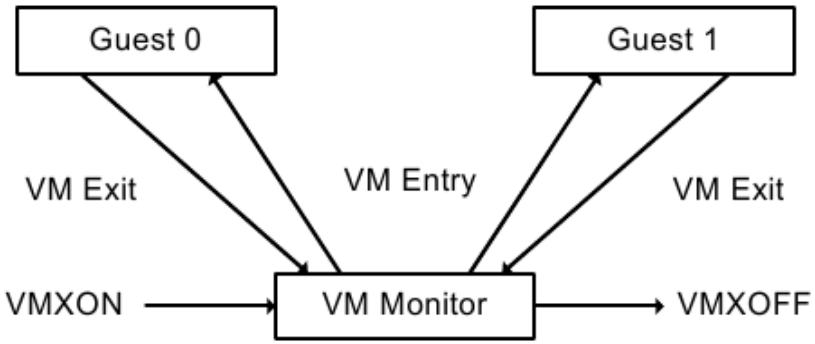
- grep vmx /proc/cpuinfo
- VT-x
- VT-d
- Extended Page Tables

AMD

- grep svm /proc/cpuinfo
- AMD-V (Pacifica)
- AMD-Vi
- Rapid Virt. Indexing

Λειτουργία - Intel

- VMX Root Operations
- VMX Non-Root Operations
- VM exits, VM entries
- VMCS Regions



One day in VMM paradise

1. VMXON: Switch into VMX mode: To VMM(#modprobe kvm)
2. VMLAUNCH VM1: Start executing VM1 in VMX non-root operation (start VM1)
3. VM1 Exits: Go back to VMM
4. VMLAUNCH VM2: Start executing VM2 (start VM2)
5. VM2 Exits: Go back to VMM
6. VMRESUME VM2: Switch to VM2 again (VM2 needs attention)
7. VM2 Exits: Go back to VMM (VM2 doesn't need attention)
8. VMRESUME VM2: Switch to VM2 (VM2 needs attention again)
9. VMRESUME VM1: VM2 exits, VM1 switched in (VM1 needs attention)
10. VM1 exits: Go back to VMM (VM1 finished)
11. VMXOFF: Get back to Regular mode (#rmmod kvm)

Δομή

- Linux Kernel as a Hypervisor
- VMM in the hardware (η CPU δημιουργεί ring -1)
- Guests

virtIO - pv drivers

- Paravirtualized Drivers
- Hypervisor agnostic
- Included in vanilla kernel

Linux components

- kvm.ko & kvm-intel.ko or kvm-amd.ko -> /dev/kvm
- modified qemu (named kvm)
- libvirt tools

KVM Sharedpage Merging - KSM

- Βρίσκει κοινές σελίδες σε processes, επομένως δουλεύει και εκτός KVM
- Αντικαθιστά τα κοινά pages των processes με ένα ro page προσβάσιμο από περισσότερα VMs
- Τρέχει σε 1 kernel thread

Ερωτήσεις & Things to read

Xen

- Xen: http://xen.org και ειδικά, http://wiki.xen.org
- Book: ”The definitive guide to Xen Hypervisor”

KVM

- http://linux-kvm.org
- http://wiki.libvirt.org/page/Virtio
- Book: ”Intel Software Developers Manual Volume 3B, System Programming Guide, Part 2”

Ευχαριστώ! :)