Lecture #3   CS695.   13.1.2023

(*) Popek & Goldberg 1974

Req. for VMM design
- efficiency
- control
- equivalence

Appl'ns

OS

VMM

} 3 contendors for CPU

Designs of CPU virtualization with VMs by VMMs.

(1) Trap & Emulate/Virtualize

Native

Ring 3 (user space) Appl'ns

Ring 0 (Kernel mode) OS

VMM

Appl'ns	3
OS	1
VMM	0

In the x86 ISA

S: set of sensitive instructions
- they influence "system" behaviour
- they need certain special privileges

⇒ CPU always operates with diff. PL configs.

CPL — current execution PL

⇒ CPL = 0 or CPL < req. PL

if not enough privilege generate a trap.

OS handles ~~[illegible]~~ trap.

(*) Guest OS issues a write to CR_3
- trap!
- VMM can handle the trap.
- x : ~~[illegible]~~ guestos

X+200: VMM

Key to virtualizing CR3
- error checks
- bound checks

Issues. (1) VMM needs to know semantics of OS state! e.g.: interrupt handlers.

② sensistive instructions do not generate a trap!

all

C: critical instructions : S which donot generate a trap on not enough privileges.

~ popf ~ which pops top of stack & update the EFLAGS registers

bit #9 enable/ disable interrupts.

w/o privileges

- no trap
- no update.

⇒ breaks equivalence & missed virtualization opportunity.

③ - mov %cs, mem locⁿ.
%ss

} examine the 2 LSB bits to find the CPL.

- sidtr ~ stores locⁿ. of IDT to a memory addr.

④ info crosses layers with no/invalid semantics.

---

efficiency ~ ?

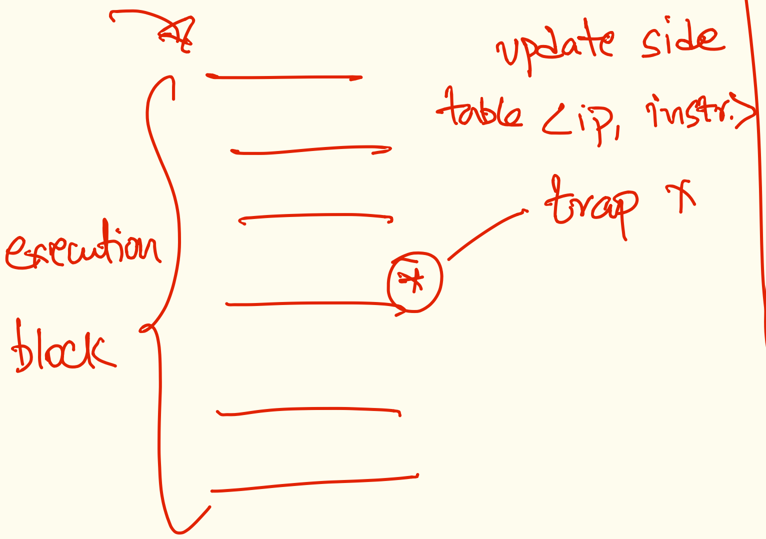
control — × / ?

equivalence — [illegible] ×

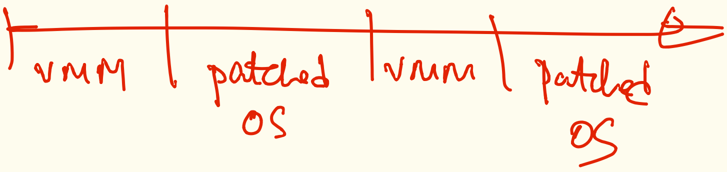
② scan-and-patch

binary translation.

= vmware ~~ex~~ esx



cpu



equivalence ✓
control ✓
efficiency. ?

③ para-virtualization

- xen

~do not issue critical instructions.

⇒ request hypervisor for all sensitive tasks.

⇒ (Guest) OS knows that the h/w ~~it~~ is being virtualized.

hypercall ~ (ABI)

(mmu-update (CR3 value, ...)

equivalence ✗
control ✓
efficiency. ✓ ✓

④ Hardware-assisted, cpu virtualization.

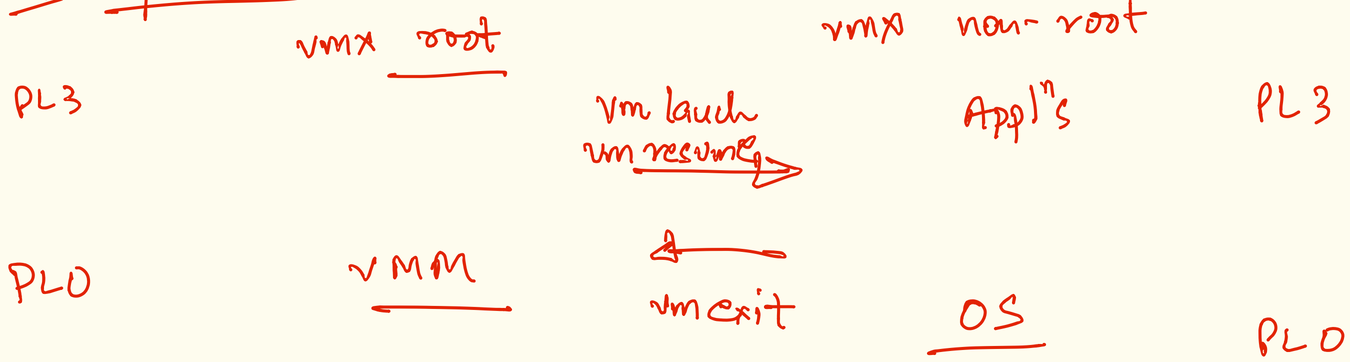
Intel VT-x }
AMD -V } ~

(i) vmx modes
- root: PL3, PL0
- non-root: PL3, PL0

(ii) vmx instructions

(iii) vmx state

(i) operations



(i) non-root is configured by the VMM.
(execution context)

(ii) program/configure the vmexit conditions.