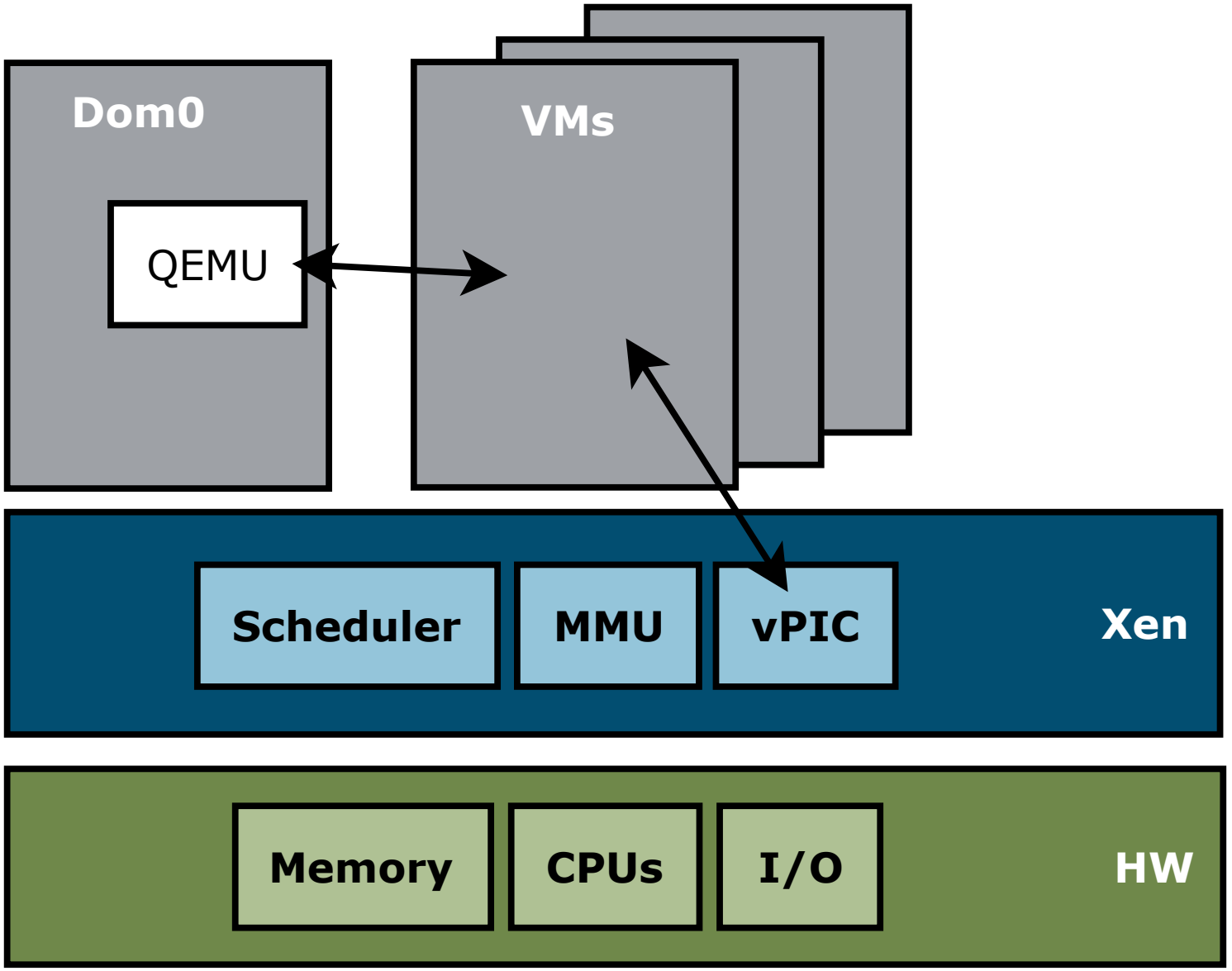
Enhance protection from security bugs in the Xen hypervisor

Anthony PERARD

Xen architecture



Emulation in hypervisor

- For performance
- Examples:
  - Interrupt controller
  - Real mode emulation
  - Timers

Emulation in hypervisor

- For performance
- Examples:
  - Interrupt controller
  - Real mode emulation
  - Timers
- Have same privilege as the hypervisor

Reduce severity of bugs

- Deprivilege emulator execution
- Different memory space
- User mode

Deprivileged mode

- Prepare page tables for user access

Deprivileged mode

- Prepare page tables for user access
- Emulator code into different section
  - .hvm_deprivileged_enhancement.text

Deprivileged mode

- Prepare page tables for user access
- Emulator code into different section
- Have context switch:
  - vmx_ctxt_switch_from()
  - Save EFER, then allow sysret/syscall
  - Save registers
  - Setup new stack for depriv mode
  - Sysret
  - Now in user mode, call the function

Deprivileged mode

- Prepare page tables for user access
- Emulator code into different section
- Have context switch
- Jump table for switch statement issue
  - .rodata

Bad behavior

- What if there is a bug in the emulator?
  - Access other memory?
  - Infinite loop?
  - Other exception?

Bad behavior

- Trap handlers for exception:
  - Page fault
  - General exception
  - ...

Bad behavior

- Trap handlers for exception:
  - Page fault
  - General exception
  - ...

    → crash domain!

Bad behavior

- Trap handlers for exception:
  - Page fault
  - General exception
  - ...

    → crash domain!
- Infinit loop?
  - Watchdog

Bad behavior

- Trap handlers for exception:
  - Page fault
  - General exception
  - ...

    → crash domain!
- Infinit loop?
  - Watchdog → crash domain

Syscall from depriv mode

- Do privileged command while in depriv mode

Syscall from depriv mode

- Do privileged command while in depriv mode
  - Set a number in a register, then syscall

Syscall from depriv mode

- Do privileged command while in depriv mode
  - Set a number in a register, then syscall
- Problem, syscall use same return path
  - Have a syscall number for actual return

Conclusion

- Optimisation
- Benchmark
- Do not trust depriv mode
- Work in progress

Conclusion

- Optimisation
- Benchmark
- Do not trust depriv mode
- Work in progress
- Proof-of-concept by Ben Catterall
- Look for “deprivileged mode” in xen-devel archive
  - http://lists.xen.org/archives/html/xen-devel/

Question?

- Look for “deprivileged mode” in xen-devel archive
  - http://lists.xen.org/archives/html/xen-devel/