



# Saving and Restoring HVM Domain

Jun Nakajima

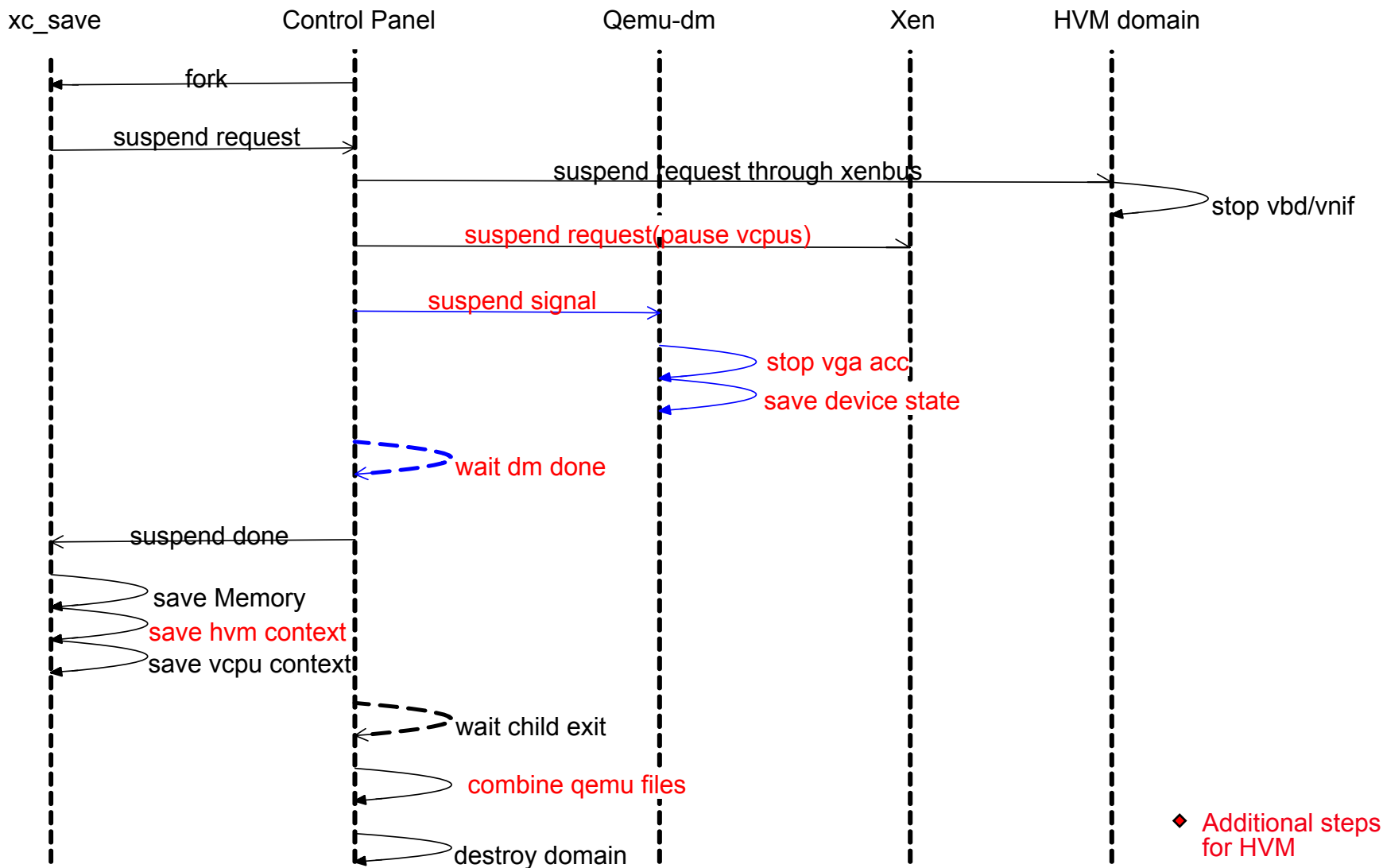
Edwin Zhai



# Basic Strategy

- Extend the existing infrastructure for PV guests
- Use the existing save/restore feature of original QEMU as much as possible
- Establish the basis for live migration

# Process For Saving HVM Domain



# Device Models State

- Reuse existing qemu's save/restore frame work
  - Call each device's save/restore handler to transfer device state to/from a file
- State of device models in Xen (PIC, PIT, Local APIC, I/O APIC)
  - Save/restore these states by new hypercalls
- VGA
  - Save VGA memory in the device model
  - Acceleration support: stop VGA acceleration when saving DM (decrease related memory) and restart it when restore
- NIC
  - MAC address(es) is also saved/restored

# HVM CPU State

- Save VMCS/VMCB fields vCPU has more state in HVM domain (e.g. GDT, IDT, TSS, etc.) than in PV domain.
  - No need for host state
- Save natural fields as 64-bit even for 32-bit guest to achieve a uniform image format

# SMP Support

- For domU we unplug/plug all vCPUs except vCPU0 when saving/restoring
  - Save vcpu0's context only
- To make save/restore transparent to HVM guest, all vCPU context need to be saved/restored
- HVM hardware context (e.g. VMCS) of all CPUs needs to saved/restored
- Bring up all vCPUs one by one at restore time

# Issues

- Time adjustment
  - Can't inject all the lost PIT interrupts to guest after restore
  - Need guest's cooperation e.g. update RTC or NTP
- Restoring on a different machine – generic issues with save, restore, and live migration
  - CPU speed difference
  - CPU feature difference

# VMX Image (Tentative) Format

xen signature
xen config len
xen config
num of pfns
all pfns
.....
hvm ctxt len
hvm ctxt
num of vcpu
vcpu ctxt len
vcpu ctxt
.....
qemu signature
qemu dm state

record len
hvm signature
hvm version
xen changeset
cpuid
device states
....

ldstr_len
idstr
Instance_id
Version_id
record size
record data

fpus
user_regs
ctrl_regs
.....
vmcs_ctxt

guest area vmcs
valid flag

msr_items
cpu_state

eip
esp
eflags
cr0
cr3
cr4
idtr
gdr
cs
ds
es
ss
fs
gs
tr
ldtr

sysenter cs/eip/esp  
Long mode state

- ◆ VMX extra fields
- ◆ domU fields



# Current Status and Next Steps

## Current Status:

- All combinations works well for UP
  - {32-bit, PAE, 64-bit guests} on {32-bit, PAE, and 64-bit host}
- Works for Linux SMP Guest as well

## Next Steps:

- Sort out hypercalls and image file format
- Get the code into the base
- More guests
  - X64 Windows
  - SMP Windows
- Port to non x86 architectures
- Live migration