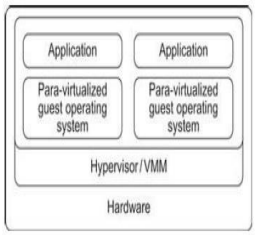
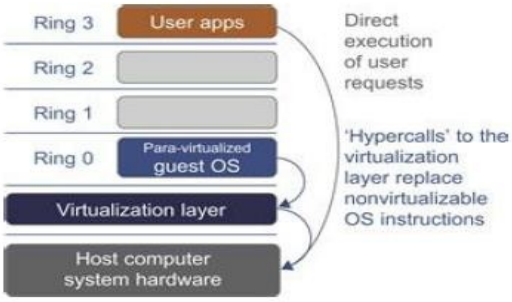


FIGURE 3.7 Para-virtualized VM architecture, which involves modifying the guest OS kernel to replace nonvirtualizable instructions with hypercalls for the hypervisor or the VMM to carry out the virtualization process (See Figure 3.8 for more details.)



3.7 VIRTUALIZATION OF CPU, MEMORY, AND I/O DEVICES

Part –B	1. Discuss how virtualization is implemented in CPU, Memory and I/O Devices.

- To support virtualization, processors such as the x86 employ a special running mode and instructions, known as hardware-assisted virtualization.
- In this way, the VMM and guest OS run in different modes and all sensitive instructions of the guest OS and its applications are trapped in the VMM.
- To save processor states, mode switching is completed by hardware. For the x86 architecture, Intel and AMD have proprietary technologies for hardware-assisted virtualization.

HARDWARE SUPPORT FOR VIRTUALIZATION

- Modern operating systems and processors permit multiple processes to run simultaneously.
- Therefore, all processors have at least two modes, user mode and supervisor mode, to ensure controlled access of critical hardware.
- Instructions running in supervisor mode are called privileged instructions.
- Other instructions are unprivileged instructions.
- In a virtualized environment, it is more difficult to make OS and applications run correctly because there are more layers in the machine stack.

SOFTWARE:

- The VMware Workstation is a VM suite for x86 and x86-64 computers.
- This software suite allows users to set up multiple x86 and x86-64 virtual computers and to use one or more of these VMs simultaneously with the host operating system.
- The VMware Workstation assumes the host-based virtualization.
- Xen is a hypervisor for use in IA-32,x86-64, Itanium, and PowerPC 970 hosts.

- Actually, Xen modifies Linux as the lowest and most privileged layer, or a hypervisor.
- One or more guest OS can run on top of the hypervisor.
- KVM (Kernel-based Virtual Machine) is a Linux kernel virtualization infrastructure. KVM can support hardware assisted virtualization and para virtualization by using the Intel VT-x or AMD-v and Virtual IO framework, respectively.
- The Virtual IO framework includes a para virtual Ethernet card, a disk I/O controller, a balloon device for adjusting guest memory usage, and a VGA graphics interface using VMware drivers.

3.7.1. CPU VIRTUALIZATION

- A VM is a duplicate of an existing computer system in which a majority of the VM instructions are executed on the host processor in native mode.
- Thus, unprivileged instructions of VMs run directly on the host machine for higher efficiency.
- The critical instructions are divided into three categories:
    - Privileged Instructions,
    - Control-Sensitive instructions, And
    - Behavior-Sensitive Instructions.
- Privileged instructions execute in a privileged mode and will be trapped if executed outside this mode.
- Control-sensitive instructions attempt to change the configuration of resources used. Behavior-sensitive instructions have different behaviors depending on the configuration of resources, including the load and store operations over the virtual memory.
- When the privileged instructions including control- and behavior-sensitive instructions of a VM are executed, they are trapped in the VMM.
- In this case, the VM Macts as a unified mediator for hardware access from different VMs to guarantee the correctness and stability of the whole system.
- However, not all CPU architectures are virtualizable.
- RISC CPU architectures can be naturally virtualized because all control and behavior-sensitive instructions are privileged instructions. On the contrary, x86 CPU architectures are not primarily designed to support virtualization.
- This is because about sensitive instructions, such as SGDT and SMSW, are not privileged instructions. When these instructions execute in virtualization, they cannot be trapped in the VMM.
- On a native UNIX-like system, a system call triggers the 80h interrupt and passes control to the OS kernel.
- The interrupt handler in the kernel is then invoked to process the system call.

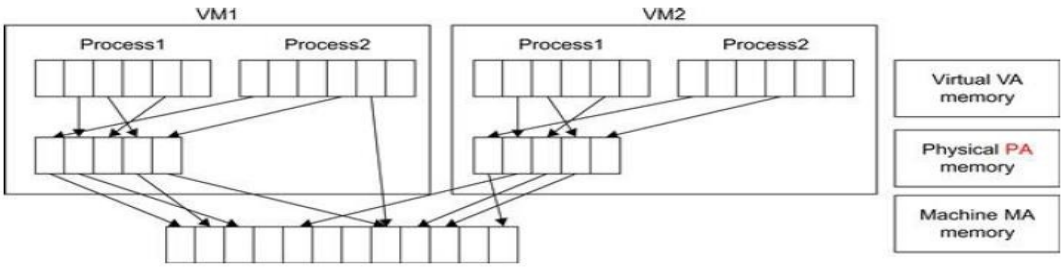
HARDWARE-ASSISTED CPU VIRTUALIZATION

- This technique attempts to simplify virtualization because full or para virtualization is complicated.
- Intel and AMD add an additional mode called privilege mode level to x86 processors.
- Therefore, operating systems can still run at Ring0 and the hypervisor can run at Ring -1. All the privileged and sensitive instructions are trapped in the hypervisor automatically. This technique removes the difficulty of implementing binary translation of full virtualization. It also lets the operating system run in VMs without modification.

3.7.2 MEMORY VIRTUALIZATION

- Virtual memory virtualization is similar to the virtual memory support provided by modern operating systems.

- In a traditional execution environment, the operating system maintains mappings of virtual memory to machine memory using page tables, which is a one-stage mapping from virtual memory to machine memory.
- All modern x86 CPUs include a memory management unit (MMU) and a translation look aside buffer (TLB) to optimize virtual memory performance.
- However, in a virtual execution environment, virtual memory virtualization involves sharing the physical system memory in RAM and dynamically allocating it to the physical memory of the VMs.
- That means a two-stage mapping process should be maintained by the guest OS and the VMM, respectively: virtual memory to physical memory and physical memory to machine memory.
- MMU virtualization should be supported, which is transparent to the guest OS. The guest OS continues to control the mapping of virtual addresses to the physical memory addresses of VMs.
- But the guest OS cannot directly access the actual machine memory. The VMM is responsible for mapping the guest physical memory to the actual machine memory.



- Since each page table of the guest OSes has a separate page table in the VMM corresponding to it, the VMM page table is called the shadow page table.
- Nested page tables add another layer of indirection to virtual memory. The MMU already handles virtual-to-physical translations as defined by the OS.
- Then the physical memory addresses are translated to machine addresses using another set of page tables defined by the hypervisor.
- VMware uses shadow page tables to perform virtual-memory-to-machine-memory address translation. Processors use TLB hardware to map the virtual memory directly to the machine memory to avoid the two levels of translation on every access.
- When the guest OS changes the virtual memory to a physical memory mapping, the VMM updates the shadow page tables to enable a direct lookup.
- The AMD Barcelona processor has featured hardware-assisted memory virtualization since 2007.
- It provides hardware assistance to the two-stage address translation in a virtual execution environment by using a technology called nested paging.

3.7.3 I/O VIRTUALIZATION

- I/O virtualization involves managing the routing of I/O requests between virtual devices and the shared physical hardware. These are three ways to implement I/O virtualization: full device emulation, para-virtualization, and direct I/O.
- Full device emulation is the first approach for I/O virtualization. Generally, this approach emulates well-known, real-world devices.
- All the functions of a device or bus infrastructure, such as device enumeration, identification, interrupts, and DMA, are replicated in software.
- This software is located in the VMM and acts as a virtual device.
- The I/O access requests of the guest OS are trapped in the VMM which interacts with theI/O devices.