Virtualization for Data Center Automation

Y. V. Lokeswari AP/ CSE

SSN College of Engineering

Reference: Distributed and Cloud Computing

K. Hwang, G. Fox and J. Dongarra

Virtualization for Datacenter Automation
to serve millions of clients, simultaneously

- Server Consolidation in Virtualized Datacenter
- Virtual Storage Provisioning and De-provisioning
- Cloud Operating Systems for Virtual Datacenters
- Trust Management in virtualized Datacenters

Data Center Virtualization

- **Data-center automation** means that **huge volumes** of **hardware**, **software**, and **database resources** in these data centers can be **allocated dynamically** to millions of Internet users simultaneously, with guaranteed **QoS** and **cost-effectiveness**

- Latest virtualization development highlights **high availability (HA), backup services, workload balancing, and further increases in client bases.**

Virtualization for Datacenter Automation
to serve millions of clients, simultaneously

- **Server Consolidation in Virtualized Datacenter**
- Virtual Storage Provisioning and De-provisioning
- Cloud Operating Systems for Virtual Datacenters
- Trust Management in virtualized Datacenters

Server Consolidation in Data Centers

- In data centers, a large number of **heterogeneous workloads** can **run** on **servers** at various times.

- **Chatty workloads** may **burst** at some **point** and return to a **silent** state at some **other point**.
- A **web video** service is an example of this, whereby a lot of people use it at night and few people use it during the day.

- **Non-Interactive** workloads do **not** require **people's efforts** to make progress after they are submitted.
- **High-performance** computing is a typical example of this.

Server Consolidation in Data Centers

- The **workload** is statically allocated **enough resources** so that peak demand is satisfied.
- Therefore, it is common that **most servers** in data centers are **underutilized**
- Among several **server consolidation** techniques such as **centralized** and **physical consolidation**, **virtualization-based server consolidation** is the most powerful.

Server Consolidation in Data Centers

- Data centers need to **optimize** their **resource management.**
- Use of VMs **increases resource management** complexity. Server Consolidation is a solution.

Properties of Sever Consolidation

- Consolidation enhances **hardware utilization**. Consolidation also facilitates **backup services** and **disaster recovery**.

- This approach enables more agile **provisioning** and **deployment** of resources. In a virtual environment, the **images** of the **guest OSes** and their applications are readily **cloned** and **reused**.

Server Consolidation in Data Centers

Properties of Sever Consolidation

- The total **cost** of **ownership** is **reduced**
- This approach **improves availability** and **business continuity**. The **crash** of a **guest OS** has **no effect** on the **host OS** or any other guest OS.
- It becomes easier to **transfer** a **VM** from **one server** to **another**, because **virtual servers** are **unaware** of the underlying **hardware**.

Server Consolidation in Data Centers

- In virtualized data centers, an **efficient**, **on-demand**, **fine-grained scheduler** is one of the key factors to improve **resource utilization**

- Two methods **Autonomic Resource Allocation**.

1. **Automatically adjusting** resource overhead based on **varying workloads** in hosted services.
2. **Two-level resource management** system to handle the complexity involved.
   - A **local controller** at the **VM level** and a **global controller** at the **server level** are designed.
   - They implement autonomic resource allocation via the **interaction** of the **local** and **global** controllers

Virtualization for Datacenter Automation
to serve millions of clients, simultaneously

- Server Consolidation in Virtualized Datacenter
- **Virtual Storage Provisioning and De-provisioning**
- Cloud Operating Systems for Virtual Datacenters
- Trust Management in virtualized Datacenters

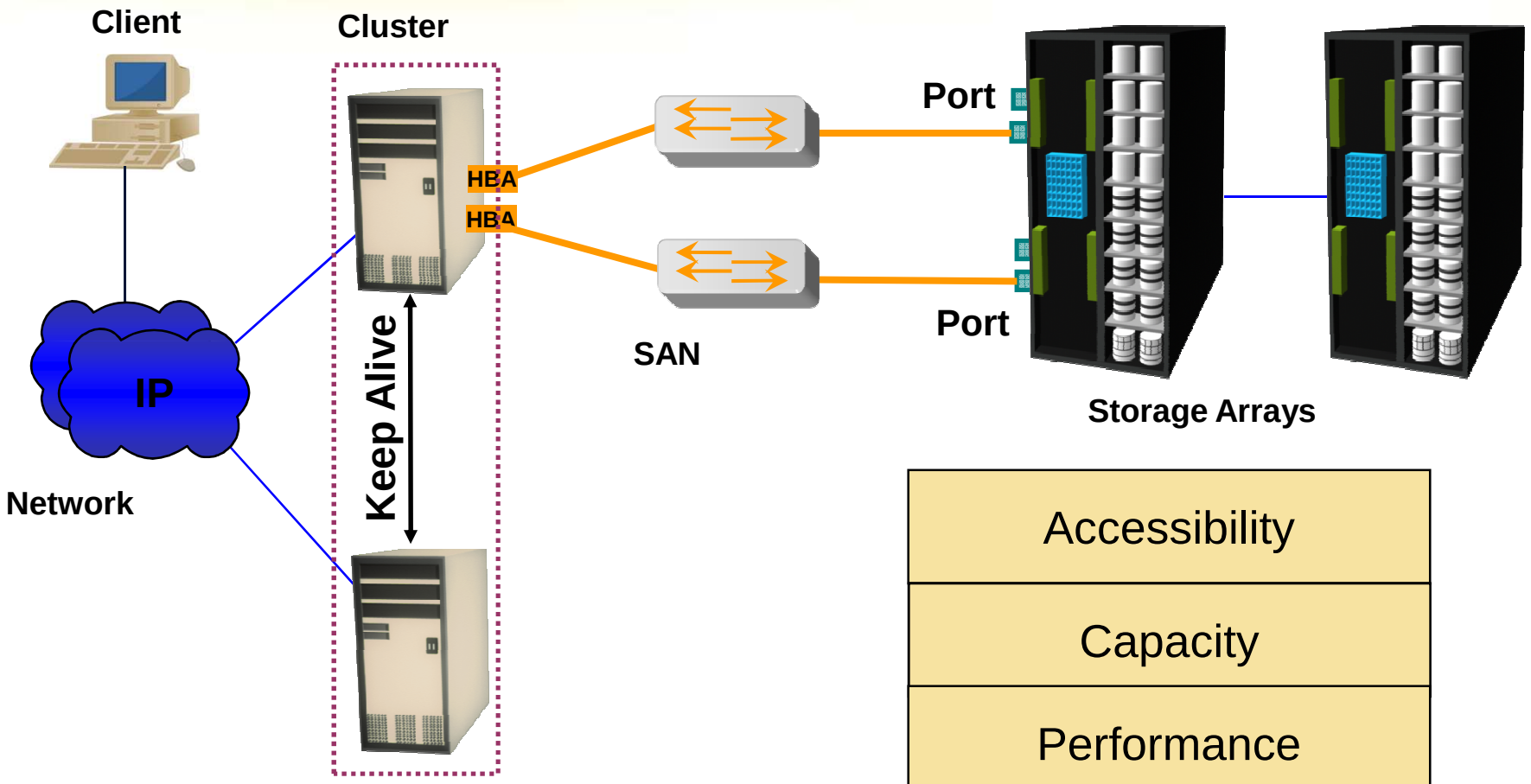
Virtual Storage Management

- **Storage virtualization** was largely used to describe the **aggregation** and **repartitioning** of **disks** for use by physical machines.
- Generally, the **data stored** in this environment can be classified into two categories: **VM images** and **application data**.
- The most important aspects of system virtualization are **encapsulation** and **isolation**.
- Traditional **operating systems** and **applications** running on them can be **encapsulated** in **VMs**.
- Only **one operating system** runs in a virtualization while **many applications** run in the operating system.
- System virtualization allows **multiple VMs** to run on a **physical machine** and the **VMs** are completely **isolated**.

Virtual Storage Management

- **Virtualization** procedure **complicates storage** operations.
- Operations such as **remapping volumes** across hosts and **check-pointing disks** are frequently clumsy
- In data centers, there are often **thousands** of **VMs**, which cause the **VM images** to become flooded
- Lot **research** is going on in **Storage Virtualization**, to make **management easy** while **enhancing performance** and **reducing** the **amount** of **storage** occupied by the **VM images**.
- **Parallax** is a **distributed storage system** customized for virtualization environments.
- **Content Addressable Storage (CAS)** is a solution to reduce the **total size** of **VM images**.

Monitoring Storage Infrastructure



- Accessibility
- Capacity
- Performance
- Security

Parallax Providing Virtual Disks to Client VMs from a Large Common Shared Physical Disk

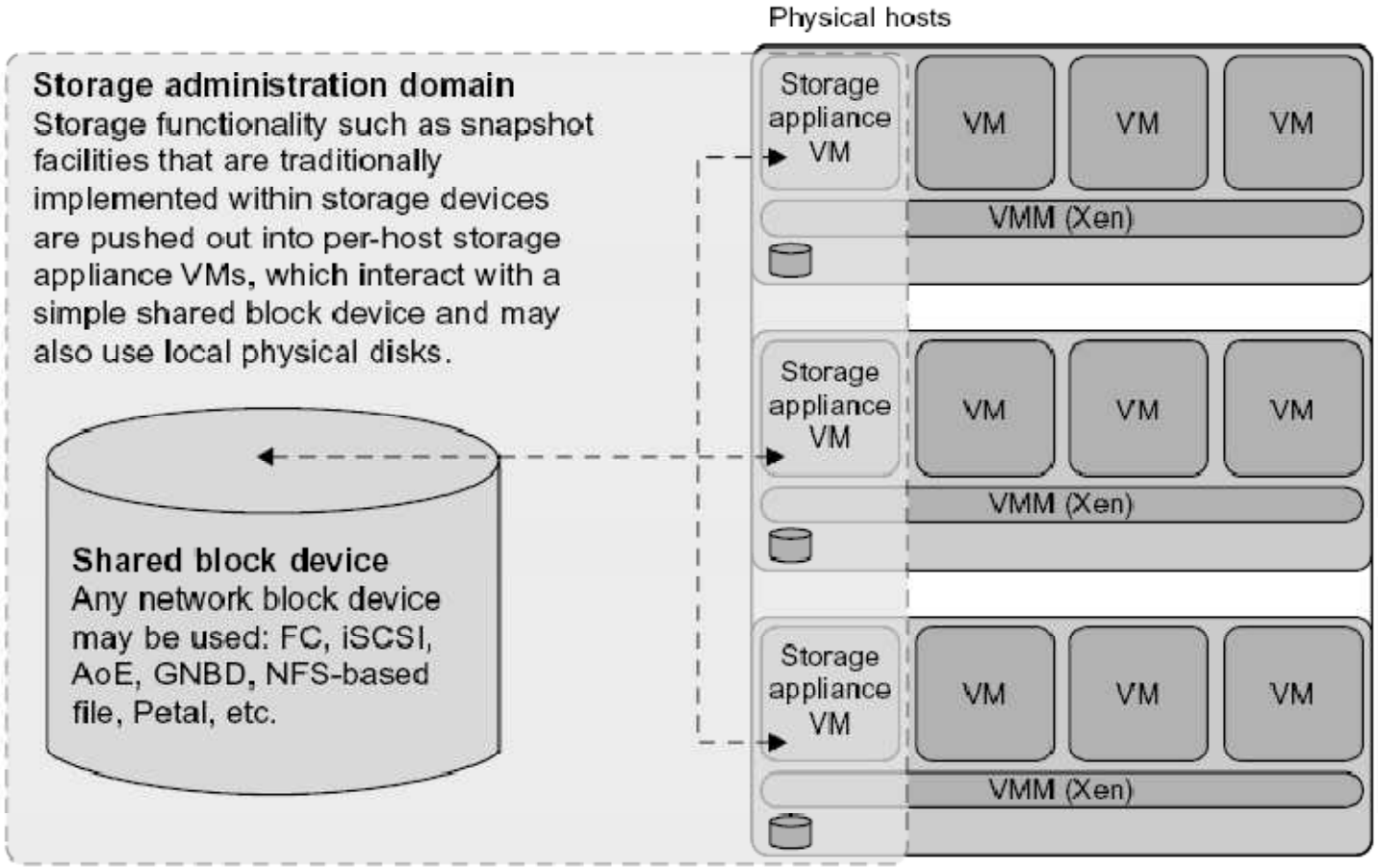


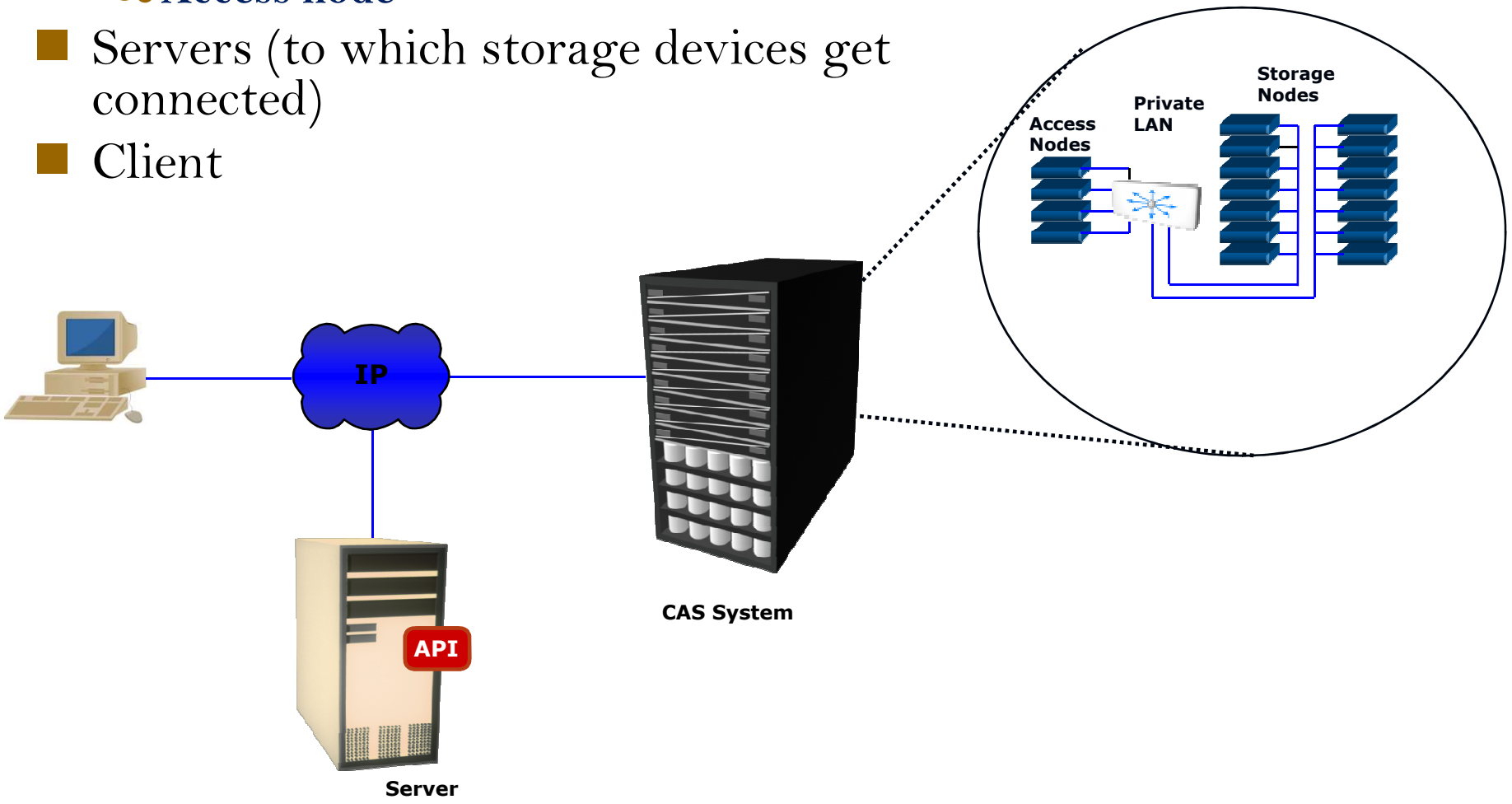
FIGURE 3.26

Parallax is a set of per-host storage appliances that share access to a common block device and presents virtual disks to client VMs.

(*Courtesy of D. Meyer, et al. [43]*)

CAS – Content Addressable Storage

- Storage devices (CAS Based)
  - **Storage node**
  - **Access node**
- Servers (to which storage devices get connected)
- Client



Physical Tape Library

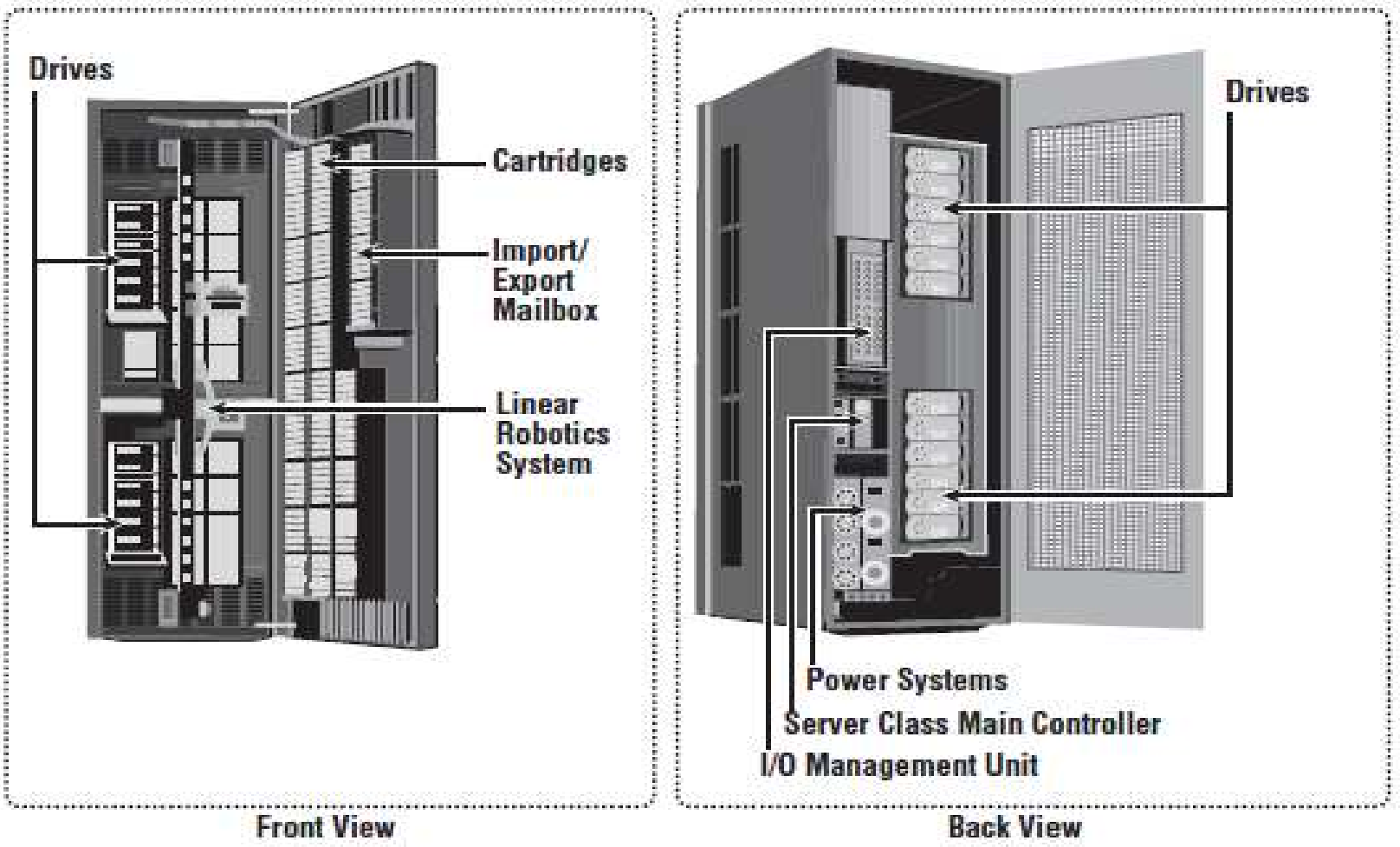


Figure 12-15: Physical tape library

Virtualization for Datacenter Automation
to serve millions of clients, simultaneously

- Server Consolidation in Virtualized Datacenter
- Virtual Storage Provisioning and De-provisioning
- **Cloud Operating Systems for Virtual Datacenters**
- Trust Management in virtualized Datacenters

Cloud OS for Building Private Clouds

Table 3.6 VI Managers and Operating Systems for Virtualizing Data Centers [9]

Manager/ OS, Platforms, License	Resources Being Virtualized, Web Link	Client API, Language	Hypervisors Used	Public Cloud Interface	Special Features
Nimbus Linux, Apache v2	VM creation, virtual cluster, www .nimbusproject.org/	EC2 WS, WSRF, CLI	Xen, KVM	EC2	Virtual networks
Eucalyptus Linux, BSD	Virtual networking (Example 3.12 and [41]), www .eucalyptus.com/	EC2 WS, CLI	Xen, KVM	EC2	Virtual networks
OpenNebula Linux, Apache v2	Management of VM, host, virtual network, and scheduling tools, www.opennebula.org/	XML-RPC, CLI, Java	Xen, KVM	EC2, Elastic Host	Virtual networks, dynamic provisioning
vSphere 4 Linux, Windows, proprietary	Virtualizing OS for data centers (Example 3.13), www .vmware.com/ products/vsphere/ [66]	CLI, GUI, Portal, WS	VMware ESX, ESXi	VMware vCloud partners	Data protection, vStorage, VMFS, DRM, HA

Eucalyptus : An Open-Source OS for Setting Up and Managing Private Clouds

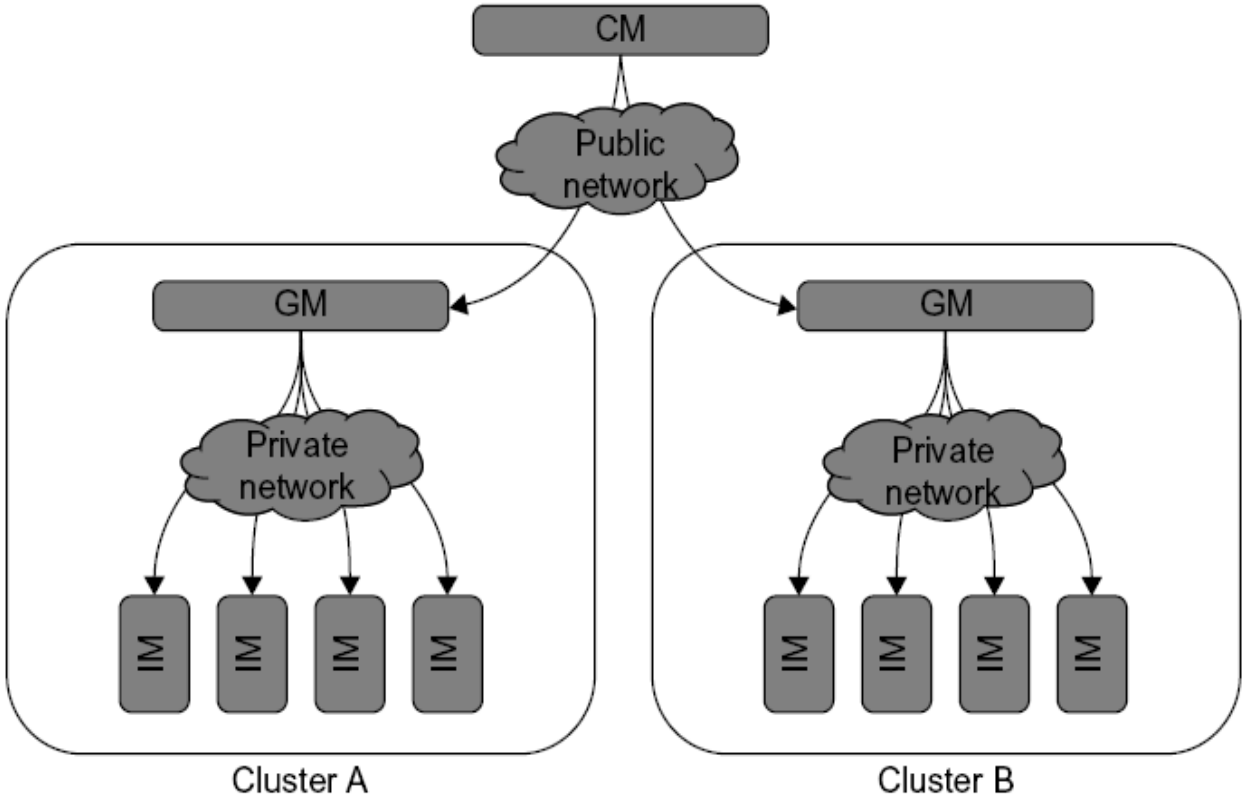


FIGURE 3.27

Eucalyptus for building private clouds by establishing virtual networks over the VMs linking through Ethernet and the Internet.

(*Courtesy of D. Nurmi, et al. [45]*)

VMware vSphere 4 as a Commercial Cloud OS

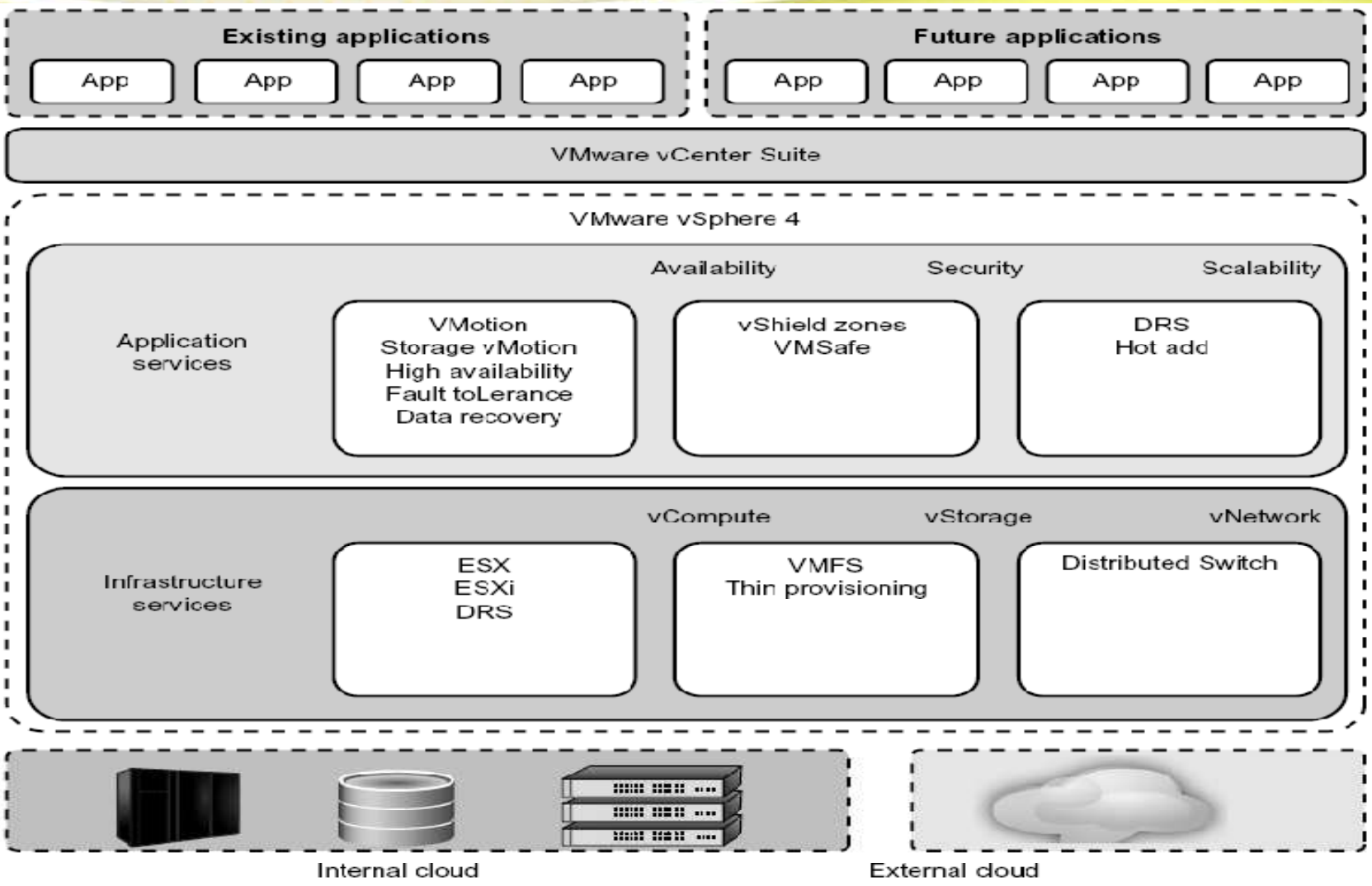


FIGURE 3.28

vSphere/4, a cloud operating system that manages compute, storage, and network resources over virtualized data centers.

(Courtesy of VMware, April 2010 [72])

Virtualization for Datacenter Automation
to serve millions of clients, simultaneously

- Server Consolidation in Virtualized Datacenter
- Virtual Storage Provisioning and De-provisioning
- Cloud Operating Systems for Virtual Datacenters
- **Trust Management in virtualized Datacenters**

Trust Management in Virtualized Data Center

- **VMM** can provide **secure isolation** and a VM accesses hardware resources through the control of the VMM, so the **VMM** is the **base** of the **security** of a virtual system.
- Normally, **one VM** is taken as a **management VM** to have some **privileges** such as creating, suspending, resuming, or deleting a VM.
- **If a hacker successfully enters the VMM or management VM, the whole system is in danger.**
- A **subtler problem** arises in protocols that rely on the "**freshness**" of their **random number** source for generating **session keys**
- For example, the **reuse** of **TCP** initial **sequence numbers** can raise **TCP hijacking** attacks.

VM-Based Intrusion Detection

- **Intrusions** are **unauthorized access** to a certain **computer** from local or network users
- **intrusion detection** is used to **recognize** the **unauthorized access**.
- An **intrusion detection system (IDS)** is built on operating systems, and is **based on** the **characteristics** of **intrusion actions**.
- IDS can be classified as a **host-based IDS (HIDS) or a network-based IDS (NIDS),** depending on the data source.
- A **HIDS** can be implemented on the **monitored system**.
- A **NIDS** is based on the **flow** of **network traffic** which can't detect fake actions..

VM-Based Intrusion Detection

- **Implementing a VM-based IDS**
  - **IDS** is an **independent** process in each VM or a high-privileged VM on the VMM;
  - **IDS** is **integrated** into the **VMM** and has the same privilege to access the hardware as well as the VMM
- An **analysis** of the **intrusion action** is extremely important **after** an **intrusion** occurs.
- Traditionally, most computer systems use **logs** to **analyze** attack actions, but it is **hard** to ensure the **credibility** and **integrity** of a **log**.
- Besides IDS, **honeypots** and **honeynets** are also prevalent in **intrusion detection**
- They attract and **provide** a **fake system view** to **attackers** in order to **protect** the **real system**. In addition, the **attack action** can be **analyzed**, and a **secure IDS** can be **built**.

VM-Based Intrusion detection

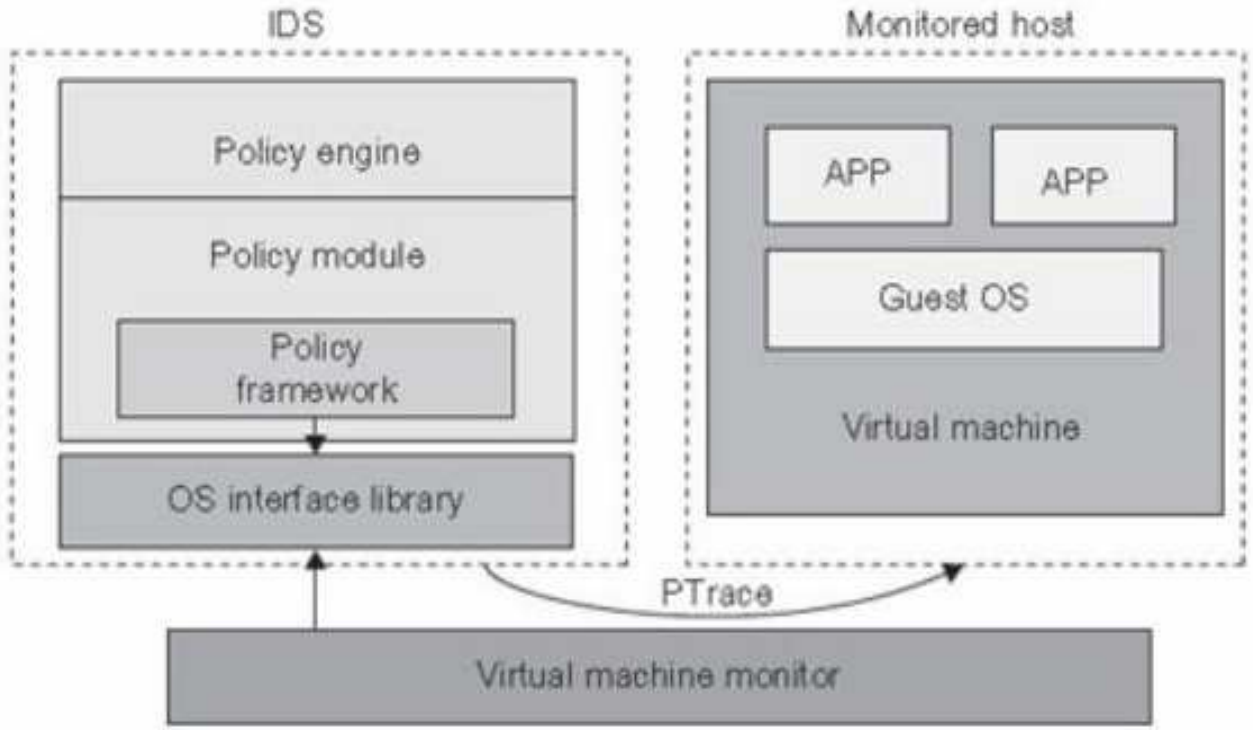


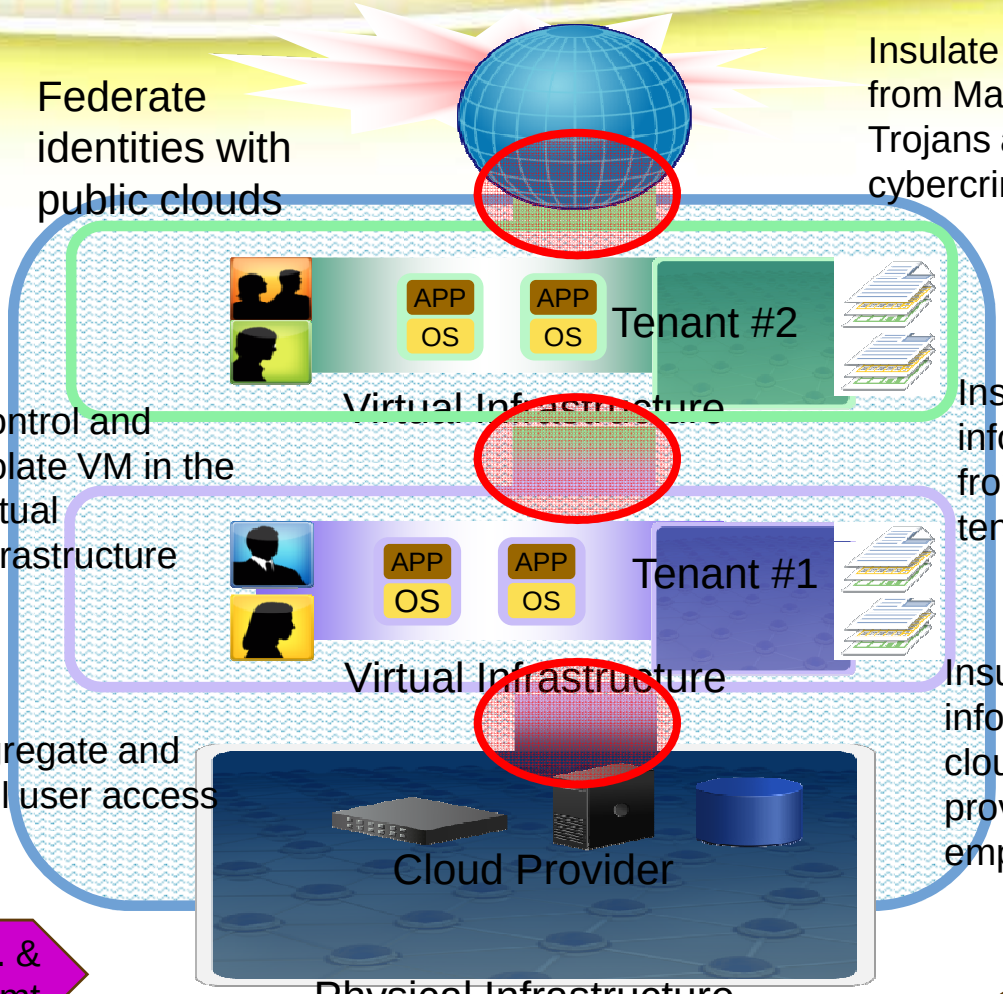
FIGURE 3.29

The architecture of livewire for intrusion detection using a dedicated VM.

(Courtesy of Garfinkel and Rosenblum, 2002 [17])

Garfinkel and Rosenblum have proposed an IDS to run on a VMM as a high-privileged VM.

Trusted Zones for VM Insulation



(Courtesy of L. Nick, EMC 2008)

Summary

- Server Consolidation in Virtualized Datacenter
- Virtual Storage Provisioning and De-provisioning
- Cloud Operating Systems for Virtual Datacenters
- Trust Management in virtualized Datacenters

Thank You