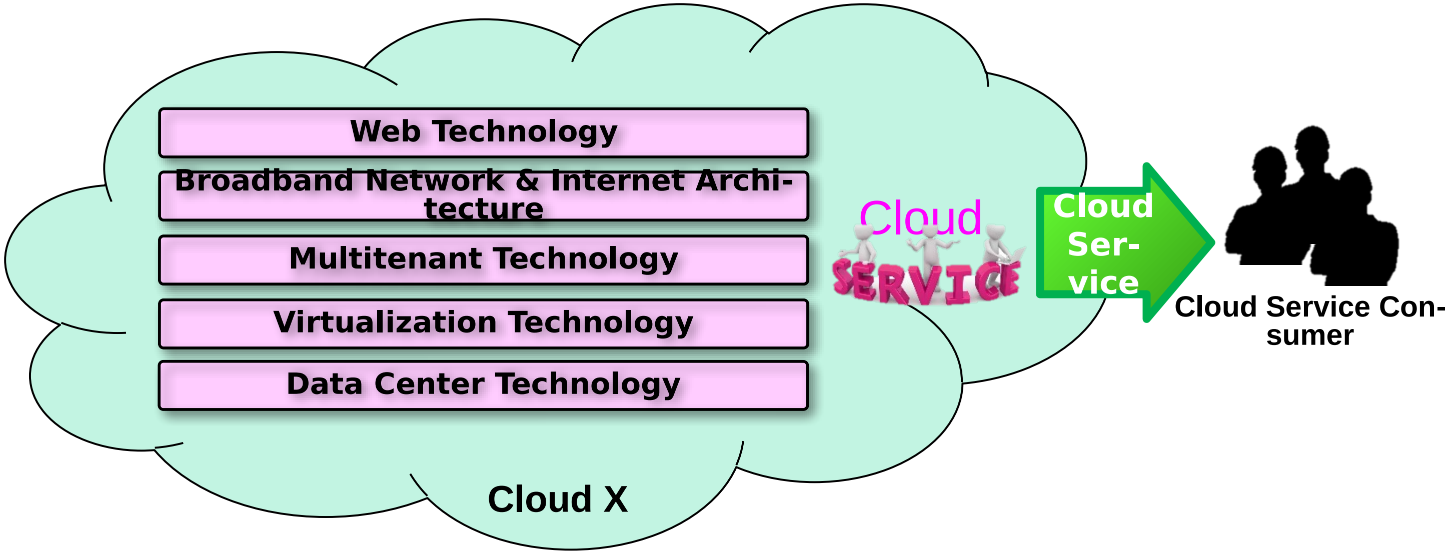
Cloud Computing

Cloud Enabling Technology

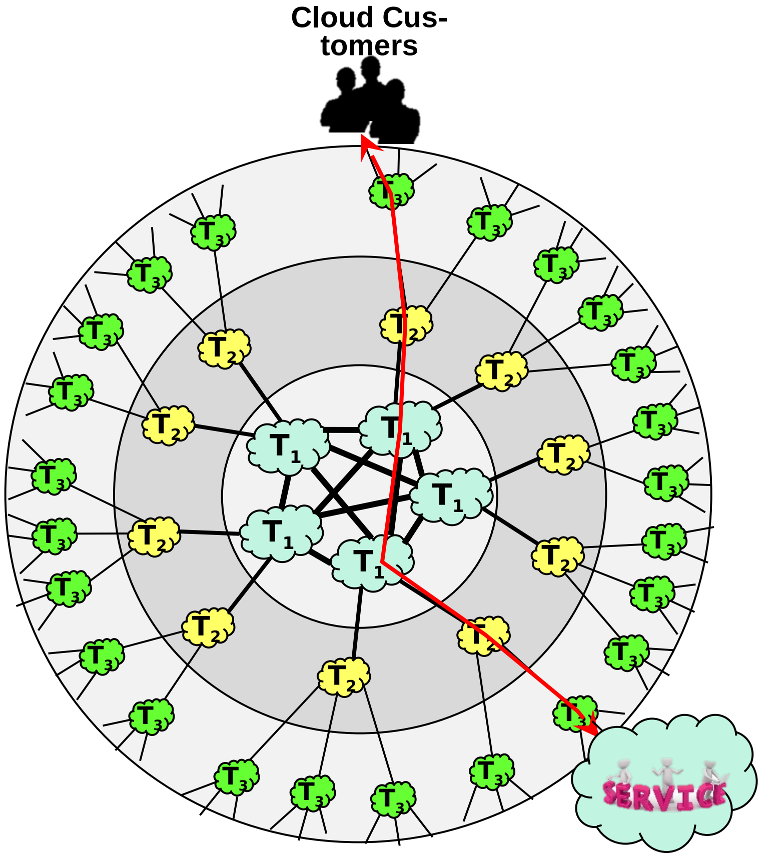
Cloud Enabling Technology

- ❑ **Integrated technology**
  - Not something entirely new – combined of & integrated from a number of existing technologies
  - Integrating a number of existing core technologies into a single service – already matured and some of them more evolved on the way
- ❑ **Existing technologies enabled cloud computing include:**
  - Broadband networks & internet architecture
  - Data center technology
  - Virtualization technology
  - Web technology
  - Multitenant technology
  - Service technology



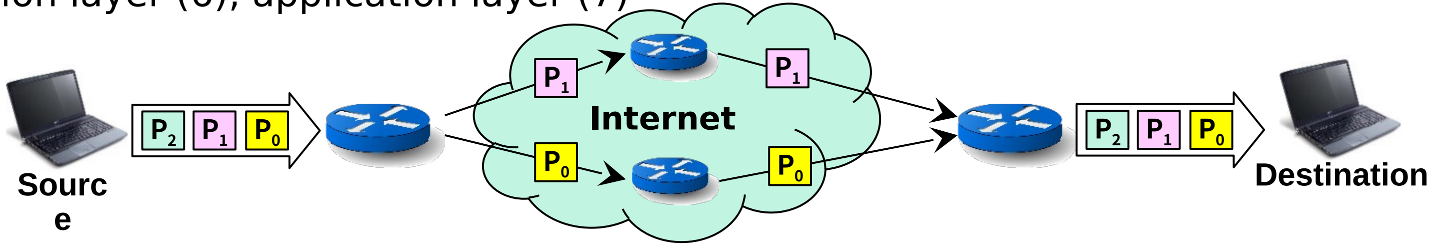
Broadband Networks & Internet Architecture – 1/11

- **Cloud service**
  - Requires remotely accessible service by definition – network connections are inevitable
  - Implies inherent dependency on internet technology
  - Enables remote provisioning of IT resources via ubiquitous network access (VPN or public network)
  - Advances in accordance with the advancements of internet technology and QoS
- **Internet Service Providers (ISPs)**
  - An organization providing national-wide or world-wide internet access service
  - Governed by Internet Corporations for Assigned Names and Numbers (ICANN)
  - No comprehensive governing by ICANN – ISPs freely deploys, operates and manages their own networks based on basically decentralized provisioning and management models
  - Fundamental governmental and regulatory laws applied within national borders
  - Internet topology – a dynamic and complex aggregate of ISPs highly interconnected via its core protocols
  - Worldwide connectivity via a hierarchical topology composed of Tier 1(large-scale international ISPs), Tier 2 (large regional ISPs) and Tier 3 (local ISPs)
  - Two fundamental components of internetworking architecture: connectionless packet switching vs. router-based interconnectivity



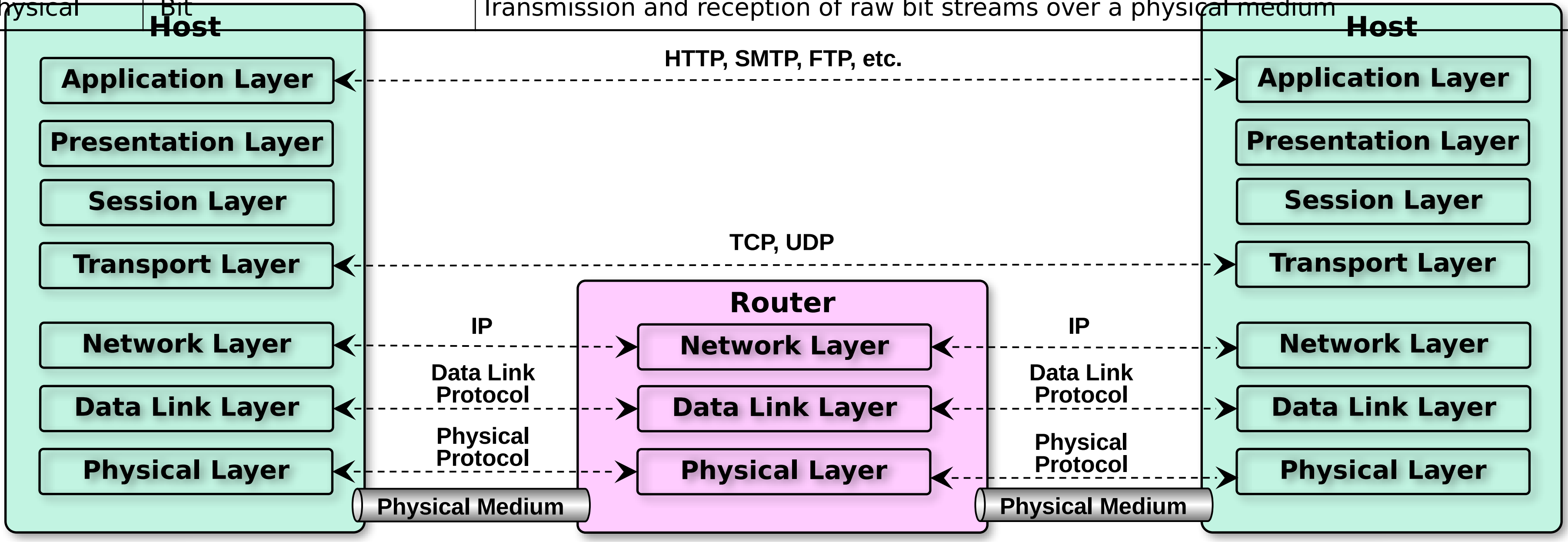
Broadband Networks & Internet Architecture – 2/11

- **Connectionless packet switching (datagram network)**
  - End-to-end (sender-receiver pair) data message divided into packets of limited size
  - Each packet processed through network switches and routers, queued and forwarded from one intermediary node to the next
  - Necessary transfer information carried by each packet in accordance with corresponding protocols such as Internet Protocol (**IP**) address or Media Access Control (**MAC**) address
- **Router-based interconnectivity**
  - A router – a device connected to multiple networks through which it forwards packets
  - Each packet transferred (stored & forwarded at each router) to destination individually via possibly different routes from each other □ routing information (IP addresses of the source & the destination, sequential number, etc.) included in each packet
  - Packets reassembled into a message on the destination node (at the network layer)
  - Each router responsible for finding the most efficient hop for packet delivery at runtime
  - Possibly multiple ISP networks between a cloud customer and its cloud provider
  - 7 abstraction layer model defined in **OSI** (Open Systems Interconnection) project by **ISO/IEC 7498-1**
    - ➢ physical layer (1), data link layer (2), network layer (3), transport layer (4), session layer (5), presentation layer (6), application layer (7)



Broadband Networks & Internet Architecture – 3/11

Layer	Protocol Data Unit (PDU)	Function
7. Application	Data	High-level APIs, including resource sharing, remote file access – HTTP, FTP etc.
6. Presentation		Translation of data between a networking service and an application; including character encoding, data compression and encryption/decryption
5. Session		Managing communication sessions, i.e. continuous exchange of information in the form of multiple back-and-forth transmissions between two nodes
4. Transport	Segment (TPC) / Datagram (UDP)	Reliable transmission of data segments between points on a network, including segmentation, acknowledgement and multiplexing
3. Network	Packet	Structuring and managing a multi-node network, including addressing, routing and traffic control
2. Data Link	Frame	Reliable transmission of data frames between two nodes connected by a physical layer
1. Physical	Bit	Transmission and reception of raw bit streams over a physical medium



Broadband Networks & Internet Architecture – 4/11

- **OSI 7 layer model**
  - Physical layer (Layer 1)
    - Defines for the electrical and physical specifications of the data connection
    - Defines the relationship between a device and a physical transmission medium (e.g., a copper or fiber optical cable, radio frequency) including the layout of pins, voltages, line impedance, cable specifications, signal timing and similar characteristics for connected devices and frequency (5 GHz or 2.4 GHz etc.) for wireless devices
    - Responsible for transmission and reception of unstructured raw data in a physical medium
    - Defines the network topology as bus, mesh, or ring being some of the most common
    - Includes **Parallel SCSI**, **Ethernet** & other local-area networks such as **token ring**, **FDDI**, **ITU-T G.hn**, and **IEEE 802.11** (Wi-Fi)
    - Defines personal area networks such as **Bluetooth** and **IEEE 802.15.4** as well
    - Defines low-level networking equipment, such as network adapters, repeaters, network hubs, modems, and fiber media converters
    - Protocol independent - never concerned with protocols or other such higher-layer items
  - Data link layer (layer 2)
    - Provides node-to-node data transfer – a link between two directly connected nodes
    - Detects and possibly corrects errors that may occur in the physical layer
    - Defines the protocol to establish and terminate a connection between two physically connected devices as well as the protocol for flow control between them
    - High-speed local area networking over existing wires (power lines, phone lines and coaxial cables) defined by The **ITU-T G.hn** standard in this data link layer, providing both error correction and flow control by means of a selective-repeat sliding-window protocol

Broadband Networks & Internet Architecture – 5/11

  - Divided into two sublayers by IEEE 802:
    - **Media Access Control** (**MAC**) layer - responsible for controlling how devices in a network gain access to medium and permission to transmit it
    - **Logical Link Control** (**LLC**) layer - responsible for identifying Network layer protocols and then encapsulating them and controls error checking and frame synchronization
  - Includes the MAC and LLC layers of IEEE 802 networks such as **802.3 Ethernet**, **802.11 Wi-Fi**, and **802.15.4 ZigBee**
  - Defines the Point-to-Point Protocol (**PPP**) that can operate over several different physical layers, such as synchronous and asynchronous serial lines

- Network layer (Layer 3)
  - Provides the functional and procedural means of transferring variable length data sequences (called **datagrams**) from one node to another connected to the same "network"
  - A network – a communication medium to which many nodes with **addresses** (e.g., IP) can be connected, allowing each member node to transfer a message to any other member nodes via **address resolution** or **routing** through intermediate nodes
  - Large messages divided into several fragments before sending and reassembled again upon receiving at the network layer
  - May report delivery errors – message delivery at the network layer is not necessarily guaranteed to be reliable; a network layer protocol may provide reliable message delivery, but it need not do so.
  - Defines a number of layer-management protocols (a function defined in the *management annex*, ISO 7498/4) including routing protocols, multicast group management, network-layer information and error, and network-layer address assignment – determined by the payload that makes these belong to the network layer, not the protocol that carries them

Broadband Networks & Internet Architecture – 6/11

- Transport layer (Layer 4)
  - Provides the functional and procedural means of transferring variable-length data sequences from a source to a destination host via one or more networks, while maintaining the quality of service functions - Transmission Control Protocol (TCP) usually built on top of the Internet Protocol (IP) is an example of a transport-layer protocol in the standard Internet stack
  - Controls the reliability of a given link through flow control, segmentation/desegmentation, and error control
  - Some protocols are state- and connection-oriented implying that the transport layer can keep track of the segments and re-transmit those that fail.
  - Also provides the acknowledgement of the successful data transmission and sends the next data if no errors occurred
  - Creates packets out of the message received from the application layer. Packetizing is a process of dividing the long message into smaller messages.
  - Five classes of connection-mode transport protocols defined by OSI, ranging from class 0 (which is also known as TP0 and provides the fewest features) to class 4 (TP4, designed for less reliable networks, similar to the Internet)
    - Class 0: contains no error recovery and designed for use on network layers that provide error-free connections
    - Class 4: closest to TCP, although TCP contains functions, such as the graceful close, which OSI assigns to the session layer
    - All OSI TP connection-mode protocol classes provide expedited data and preservation of record boundaries.
  - Similar to a post office which deals with the dispatch and classification of mail and parcels sent

Broadband Networks & Internet Architecture – 7/11

- Packets are then encapsulated into higher level protocols, such as cryptographic presentation services that can be read by the addressee only.
- Non-IP tunneling protocols operating at the transport layer: IBM's **SNA**, Novell's **IPX** over an IP network, or end-to-end encryption with **IPsec**
- While Generic Routing Encapsulation (GRE) might seem to be a network-layer protocol, if the encapsulation of the payload takes place only at endpoint, GRE becomes closer to a transport protocol that uses IP headers but contains complete frames or packets to deliver to an endpoint.
- L2TP carries PPP frames inside transport packet.
- Although not developed under the OSI Reference Model and not strictly conforming to the OSI definition of the transport layer, the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) of the Internet Protocol Suite are commonly categorized as layer-4 protocols within OSI.

Feature Name	TP0	TP1	TP2	TP3	TP4
Connection-oriented Network	Yes	Yes	Yes	Yes	Yes
Connectionless Network	No	No	No	No	Yes
Concatenation and Separation	No	Yes	Yes	Yes	Yes
Segmentation and Reassembly	Yes	Yes	Yes	Yes	Yes
Error Recovery	No	Yes	Yes	Yes	Yes
Reinitiate Connection*	No	Yes	No	Yes	No
Multiplexing / Demultiplexing over Single Virtual Circuit	No	No	Yes	Yes	Yes
Explicit Flow Control	No	No	Yes	Yes	Yes
Retransmission on Timeout	No	No	No	No	Yes
Reliable Transport Service	No	Yes	No	Yes	Yes
* If an excessive number of PDUs are unacknowledged					

Broadband Networks & Internet Architecture – 8/11

- Session layer (Layer 5)
  - Controls the dialogues (connections) between computers - establishing, managing and terminating the connections between the local and remote application
  - Provides for full-duplex, half-duplex, or simplex operation, and establishes checkpointing, adjournment, termination, and restart procedures
  - Also provides graceful close of sessions which is a property of the Transmission Control Protocol and session checkpointing and recovery which is not usually used in the Internet Protocol Suite
  - Commonly implemented explicitly in application environments that use remote procedure calls
- Presentation layer (Layer 6)
  - Establishes context between application-layer entities, in which the application-layer entities may use different syntax and semantics if the presentation service provides a mapping between them
  - Encapsulates presentation service data units into session protocol data units that are then passed down the protocol stack If a mapping is available
  - Provides independence from data representation (e.g., encryption) by translating between application and network formats
  - Transforms data into the form that the application accepts - formatting and encrypting data to be sent across a network (sometimes called the syntax layer)
  - The original presentation structure used the Basic Encoding Rules of Abstract Syntax Notation One (ASN.1) with capabilities such as converting an EBCDIC-coded text file to an ASCII-coded file, or serialization of objects and other data structures from and to XML.

Broadband Networks & Internet Architecture – 9/11

- Application layer (Layer 7)
  - The OSI layer closest to the end user which means both the OSI application layer and the user interact directly with the software application
  - Interacts with software applications (outside the scope of the OSI model) that implement a communicating component
  - Includes functions such as identifying communication partners, determining resource availability, and synchronizing communication
  - Determines the identity and availability of communication partners for an application with data to transmit when identifying communication partners
  - Must decide whether sufficient network resources for the requested communication are available when determining resource availability

Broadband Networks & Internet Architecture – 10/11

- **Technical and business considerations**
  - Connectivity issues
    - Traditional deployment model
      - Via the corporate network (VPN) which provide uninterrupted Internet connectivity
      - Completely controlled by the organizations with their own safeguard based on firewalls and various monitoring tools
      - Each organization responsible for deploying, operating and managing their IT resources and Internet connectivity
    - Cloud deployment model
      - Continuous access to centralized servers and applications granted to end-user devices as long as they are connected to the network through the Internet in the cloud
      - Centralized IT resources accessible using the same network protocols regardless of whether users reside inside or outside of a corporate network
      - Cloud IT resources configured by cloud providers to be accessible for both external and internal users through an Internet connection and for cloud consumers to provide Internet-based services to external users

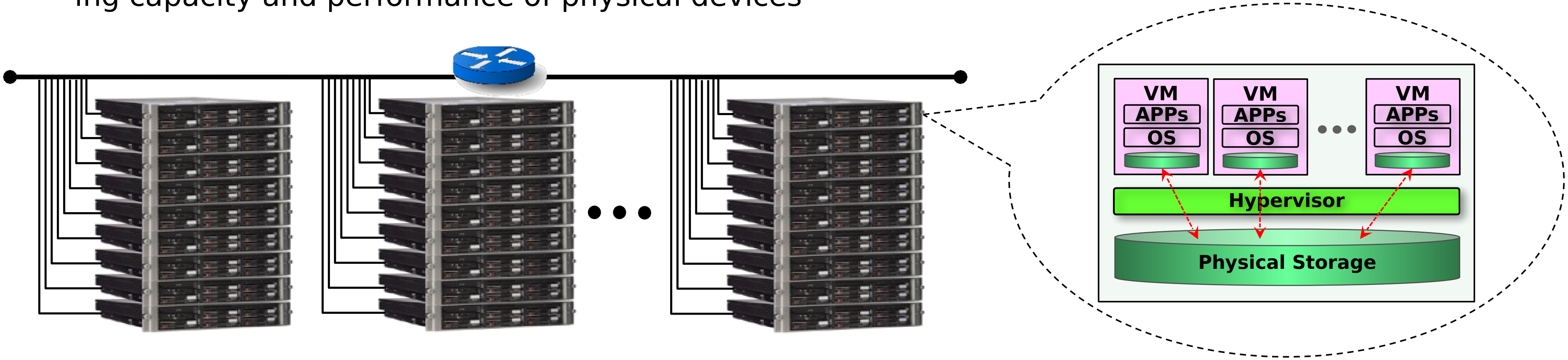
On-premise IT Resources	Cloud-based IT Resources
Internal end-user devices access corporate IT services through the corporate network.	**Internal end-user devices access corporate IT services through an Internet connection.**
Internal users access corporate IT services through the corporate Internet connection while roaming in external networks.	**Internal users access corporate IT services while roaming in external networks through the cloud provider's Internet connection.**
External users access corporate IT services through the corporate Internet connection.	**External users access corporate IT services through the cloud provider's Internet connection.**

Broadband Networks & Internet Architecture – 11/11

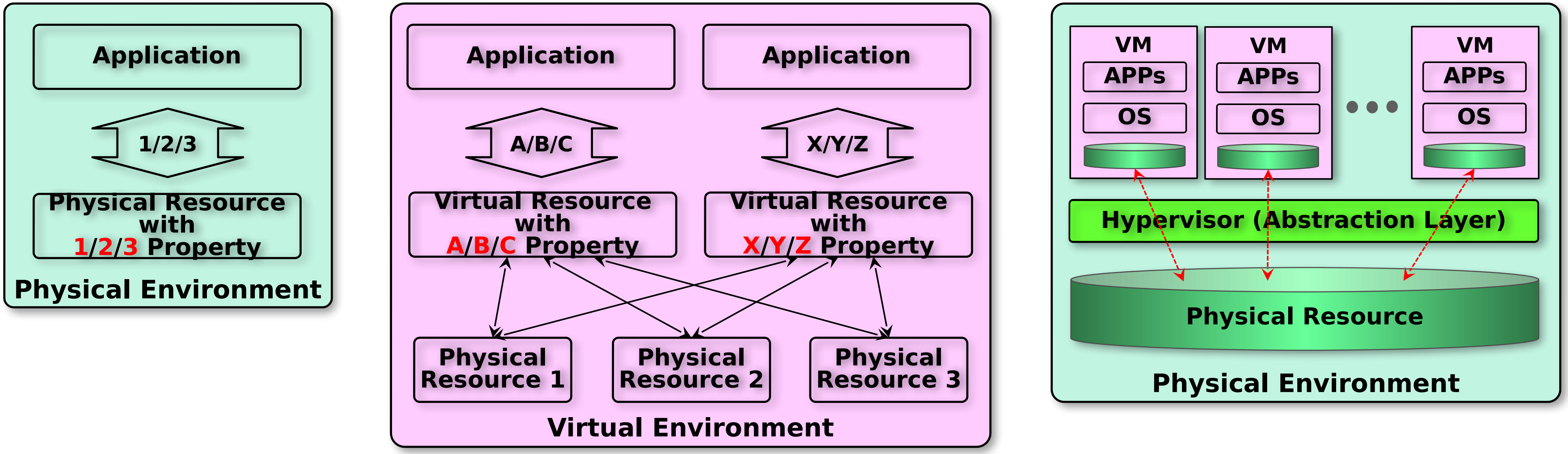
- Network bandwidth and latency issues
  - Network QoS: bandwidth, latency, jitter
  - **Bandwidth** – how much data can be transferred within a unit time
    - End-to-end bandwidth determined by the transmission capacity of the shared data links that connect intermediary nodes
    - Attempt to improve end-to-end bandwidth by ISPs with technologies such as broadband network technology & web acceleration technologies – dynamic caching, compression, pre-fetching, etc.
    - Critical for applications requiring substantial amount of data transfer
  - **Latency** – how fast a request can be satisfied (time for a packet to travel from one node to another)
    - The longer a packet travels the larger the latency is – web caching technology can apply
    - The more network traffic is the larger the latency is – more queuing delay at each hop
    - Critical for applications with a business requirement of fast response time
  - **Jitter** – how consistent the given latency is
    - A gap between the smallest latency and the largest latency
    - Response time less than a millisecond in general, but frequently more than several seconds
    - Internet-wide QoS control required to guarantee small jitter
  - QoS of the underlying network inherited to QoS of the given cloud service
- Cloud carrier and cloud provider selection
  - Involves multiple cloud carriers to achieve the necessary level of connectivity and reliability for the given cloud applications resulting in additional costs
  - QoS determined by multiple ISPs involved & required collaboration of the cloud carriers
  - Wise to adopt more relaxed latency and bandwidth requirements

Data Center Technology – 1/6

- **Data center**
  - Grouping IT resources in close proximity with one another (rather than having them geographically dispersed) for power sharing, higher efficiency in shared IT resource usage and improved accessibility for IT personnel – reason for popularizing data center concept
  - Characterized for centralized IT resources such as servers, storages, databases, networking & telecommunication devices and software solutions via applying a number of technologies
- **Standardization and modularity**
  - Built upon standardized commodity hardware and designed with modular architecture
  - Aggregating multiple identical building blocks of facility infrastructure and equipment to support scalability, growth and speedy hardware replacements
  - Reduces investment and operational costs as they enable economies of scale for the procurement, acquisition, deployment, operation and maintenance processes
  - IT resource consolidation favored by common virtualization strategies and the constantly improving capacity and performance of physical devices



Data Center Technology – 2/6



❑ **Virtualization**

- Virtualization: an abstraction layer with mapping or redirection capability
- Physical IT resources: the facility infrastructure that houses computing/networking systems and equipment, together with hardware systems and their operating systems
- Virtual IT resources: comprised of operational and management tools that are often based on virtualization platforms that abstract the physical computing and networking IT resources as virtualized components that are easier to allocate, operate, release, monitor and control (more details later)

❑ **Automation**

- Reduces operational costs and the error rate in data center via automated management without human supervision – provisioning, configuration, patching and monitoring
- Enables self-configuration and self-recovery – basis of automatic computing technology

Data Center Technology – 3/6

- **Remote operation and management**
  - Most operational and administrative tasks of IT resources in data center can be commanded through the network's remote (within data center boundary in general) consoles and management systems.
  - Most operational and administrative tasks carried out from the control room in data center except for those requiring physical operations such as hardware jobs or cabling
  - Remote operation from outside of data center boundary strictly prohibited in general.
- **High availability**
  - All resources in data center are subject to fail anytime based on current hardware and software technologies.
  - Most resource failures affect service continuity and underlying business as well.
  - In general, data center provides fail-safe technologies mainly based on redundancy in every possible layer – fault-tolerant or fault-resilient technologies on top of redundant resources: power supply, cabling, networking, servers, storages and software licenses.
  - Fault-avoidance technologies: load balancing, scaling-up/down, etc.
- **Security-aware design, operation and management**
  - The level of security determines the credibility of the given data center.
  - Security issue is the main concern that prohibits many organizations from migrating their IT resources from on-premise to cloud-based.
  - Security threats that make organizations hesitate to outsource IT environment are two-fold: possible malicious attack from outside and anxiety about keeping data outside of organization's physical boundary (not only business-wise but also legality-wise).
  - Various levels of protection and security mechanisms: network isolation, firewalls and monitoring tools – big data analysis recently

Data Center Technology – 4/6

- **Facilities**
  - Typically custom-designed computing resources, storages and network equipment for the given purpose
  - Several functional layout areas based on power supplies, cabling, environmental control stations that regulate heating, ventilation, air conditioning, fire protection, (physical) security & access control system, monitoring system, etc.

- **Computing hardware**
  - Mainly composed of standardized commodity servers with a number of computing hardware technologies such as:
    - Rack technology – standardized rack with interconnects for power, network, and internal cooling
    - CPU architecture – support for various CPU types: x86-32bits, x86-64bits, RISC, CISC, etc.
    - Multi-core CPU architecture – hundreds of physical & logical processing core in single unit of standardized racks
    - Redundancy & hot-swap technology – hard disks, power supplies, network interfaces, storage controller cards, etc.
  - Blade server technologies with rack-embedded physical interconnections (blade enclosures), fabrics (switches), power supply units, cooling fans, etc.
  - Maximizes and enhances inter-component networking and management while optimizing physical space & power via individual server hot-swapping, scaling, replacement and maintenance
  - Benefits the deployment of fault-resilient (tolerant) systems based on cluster technology

Data Center Technology – 5/6

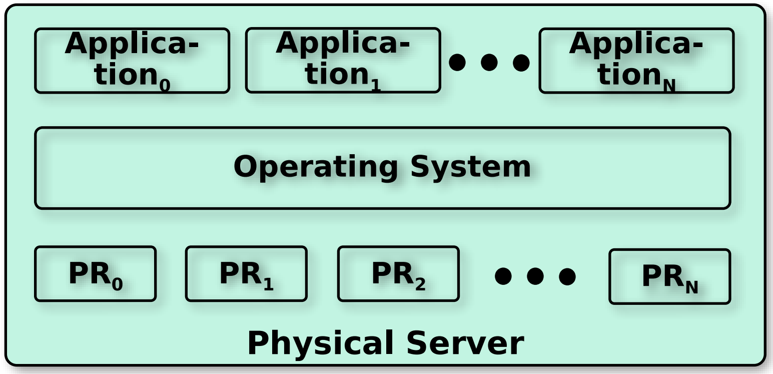
  - Several industry-standard and proprietary operational and management software tools that configure, monitor, and control hardware IT resources from remote & centralized consoles – self-provisioning
  - Hundreds or even thousands of physical or virtual servers (IT resources) operated by a single operator
- **Storage hardware**
  - Needs to deal with tons of data created every day – easily reaching PBs of total scale in general
  - One of the most difficult task to deal with in data center and many different levels of technologies for fast access, data availability, massive data accommodation, etc.:
    - RAID (Redundant Array of Independent/Inexpensive Disks) – integrating hundreds of individual HDD to provide fast, reliable, massive storage space
    - IO caching – at different layers: storage controllers, each physical/virtual servers, separate caching servers
    - Hot-swapping – replacing faulty HDD without requiring prior power down (a part of RAID technology)
    - Storage virtualization – abstracted storage layer creating virtual storage device free from the physical property of member storage devices
    - Data replication – memory snapshot, volume cloning, mirroring, DR, CDP, etc.
    - Distributed storage – file, block, object-level distributed storage: HDFS, Ceph, etc.
  - Storage Topology
    - DAS (Direct Attached Storage): storages directly attached to a host system via block-level channel protocol such as SCSI/FC
    - NAS (Network Attached Storage): storages attached to a number of host systems via file-level network protocols such as NFS/CIFS/SMB – while providing file-level data sharing among multiple hosts
    - SAN (Storage Area Network): storages attached to multiple hosts via block-level network protocols such as Fibre Channel, Infiniband, iSCSI, etc.

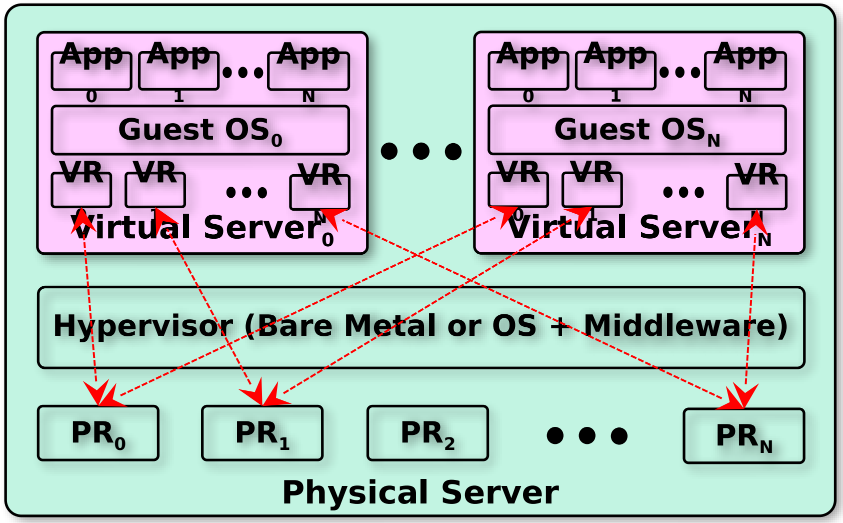
Data Center Technology – 6/6

- **Network hardware**
  - One of the most important IT capabilities for data center to support remote IT access – broken down into five network subsystems in general
    - Carrier & external network interconnection internetworking infrastructure comprised o backbone routers that provide routing between external WAN connections and LANs in the given data center including firewalls and VPN gateways
    - Web-tier load balancing and acceleration for even distribution of web traffics and acceleration of web protocols comprised of XML pre-processors, encryption/decryption appliances (web acceleration), layer 7 switching devices (content-aware load balancing), etc.
    - LAN fabric intranetworking infrastructure comprised of multiple layer 4 or lower switching devices up to ~10G bandwidth providing several virtualization functions such as LAN segregation into VLANs, link aggregation, control routing between networks, load balancing, failover (redundant connectivity), etc.
    - SAN fabric data networking infrastructure composed of multiple SAN switching devices based on data networking protocols such as Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), Infiniband (IB), Internet Small Computer Systems Interface (iSCSI)
    - NAS gateway shared file-transfer networking infrastructure composed of a number of NAS-based storage devices based on file-transfer protocols such as NFS and SMB/CIFS (Samba)
  - Basically redundant and/or fault-tolerant networking configurations for scalability and high availability
  - DWDM (Dense Wavelength Driven Multiplexing) devices for ultra high-speed networking and improved resiliency in general for the purpose of high-speed real-time data replication between data centers
- **Other consideration**
  - Technological obsolescence, heterogeneity, security, vast quantities of data and their backup, etc.

Virtualization Technology – 1/5

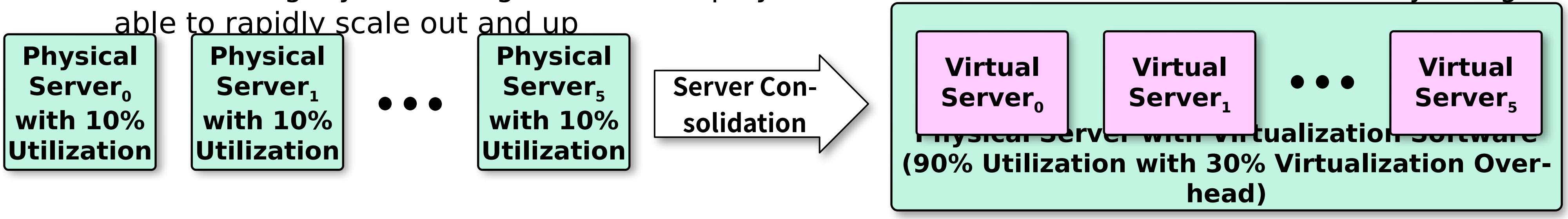
- **Basic concepts**
  - A process to convert a physical IT resource into a virtual IT resource
  - Typical IT resources that can be virtualized: memory, CPUs, network, storage, power, **server**
  - Server virtualization software: middleware vs. bare metal
    - Create a set of virtual IT resources on top of physical IT resources and maintain mapping between a set of virtual resources and physical resources
    - Create guest (user) operating systems
    - Guest OSs and application programs running on both physical or virtual environments **without any modification a vital characteristic of virtualization**
    - Virtual machine manager, virtual machine monitor or **hypervisor**





Virtualization Technology – 2/5

- **Hardware independence**
  - Creating standardized soft (virtual) copies of physical IT resources eliminating hardware dependency
  - Easy automated VM migration or failover between different physical servers
- **Server consolidation**
  - Creating different multiple virtual servers on a single physical server
  - Mainly for Increasing server utilization, load balancing or optimizing IT resource utilization
  - Support for common cloud features such as on-demand usage, resource pooling, elasticity, scalability and resiliency
- **Resource replication**
  - Virtual servers are implemented as virtual disk images (configuration, memory state, etc.) containing binary file copies of hard disk content and being accessible via simple file operations such as copy and move host OS.
  - Easy to be replicated, migrated, backed up and manipulated enabling:
    - Easy creation of standardized VM images with guest OS and pre-packaged application software in virtual disk images for instantaneous deployment
    - Increased agility in the migration and deployment of a virtual machine's new instance by being able to rapidly scale out and up

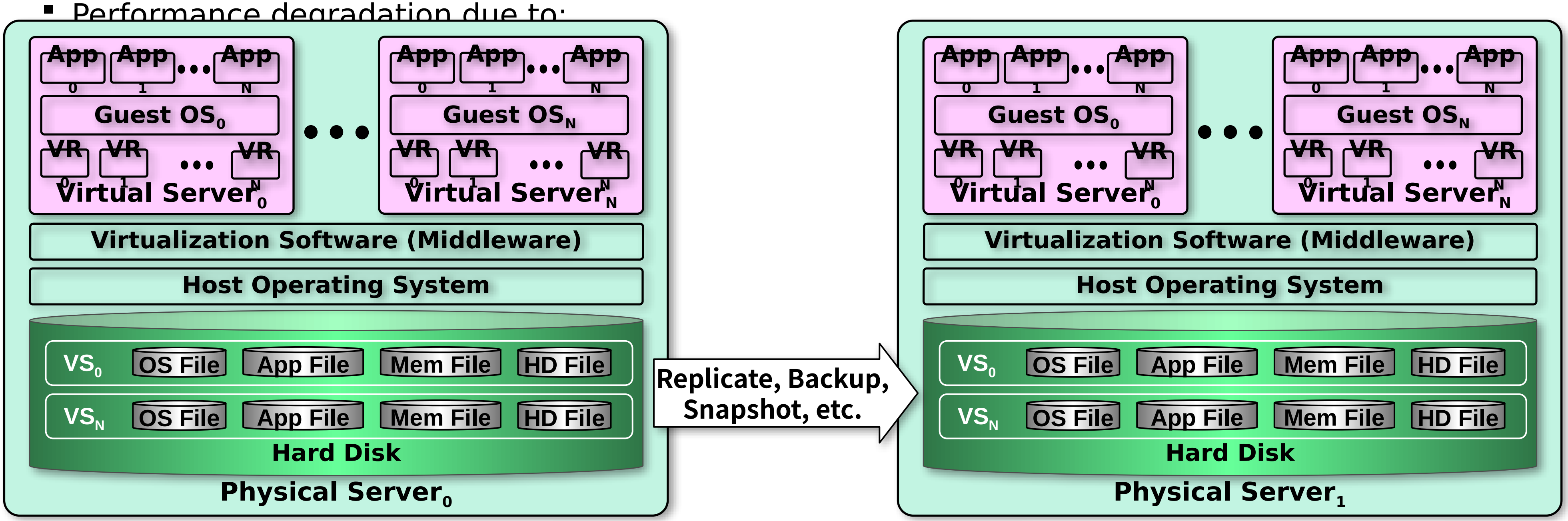

Physical Server$_0$ with 10% Utilization
Physical Server$_1$ with 10% Utilization
• • •
Physical Server$_5$ with 10% Utilization
Server Consolidation
Virtual Server$_0$
Virtual Server$_1$
• • •
Virtual Server$_5$
Physical Server with Virtualization Software
(90% Utilization with 30% Virtualization Overhead)

Virtualization Technology – 3/5

- ➢ Ability to roll back for instantaneous creation of VM snapshot by saving the state of the virtual server’s memory and hard disk image to a host-based file
- ➢ Easy implementation of business continuity with efficient backup and restoration

❑ **Operating system-based virtualization**

- ▪ Install virtualization software in a pre-existing operating system (host vs. guest)
- ▪ Act as an application or more precisely as a middleware
- ▪ Easy to deal with hardware compatibility issues even with absence of a specific hardware driver
- ▪ Host OS services to be utilized: backup/recovery, integration to directory service, security management
- ▪ Performance degradation due to:



App$_0$ App$_1$ … App$_N$
Guest OS$_0$
VR$_0$ VR$_1$ … VR$_N$
Virtual Server$_0$
…
App$_0$ App$_1$ … App$_N$
Guest OS$_N$
VR$_0$ VR$_1$ … VR$_N$
Virtual Server$_N$
Virtualization Software (Middleware)
Host Operating System
VS$_0$ OS File App File Mem File HD File
VS$_N$ OS File App File Mem File HD File
Hard Disk
Physical Server$_0$
Replicate, Backup, Snapshot, etc.
App$_0$ App$_1$ … App$_N$
Guest OS$_0$
VR$_0$ VR$_1$ … VR$_N$
Virtual Server$_0$
…
App$_0$ App$_1$ … App$_N$
Guest OS$_N$
VR$_0$ VR$_1$ … VR$_N$
Virtual Server$_N$
Virtualization Software (Middleware)
Host Operating System
VS$_0$ OS File App File Mem File HD File
VS$_N$ OS File App File Mem File HD File
Hard Disk
Physical Server$_1$

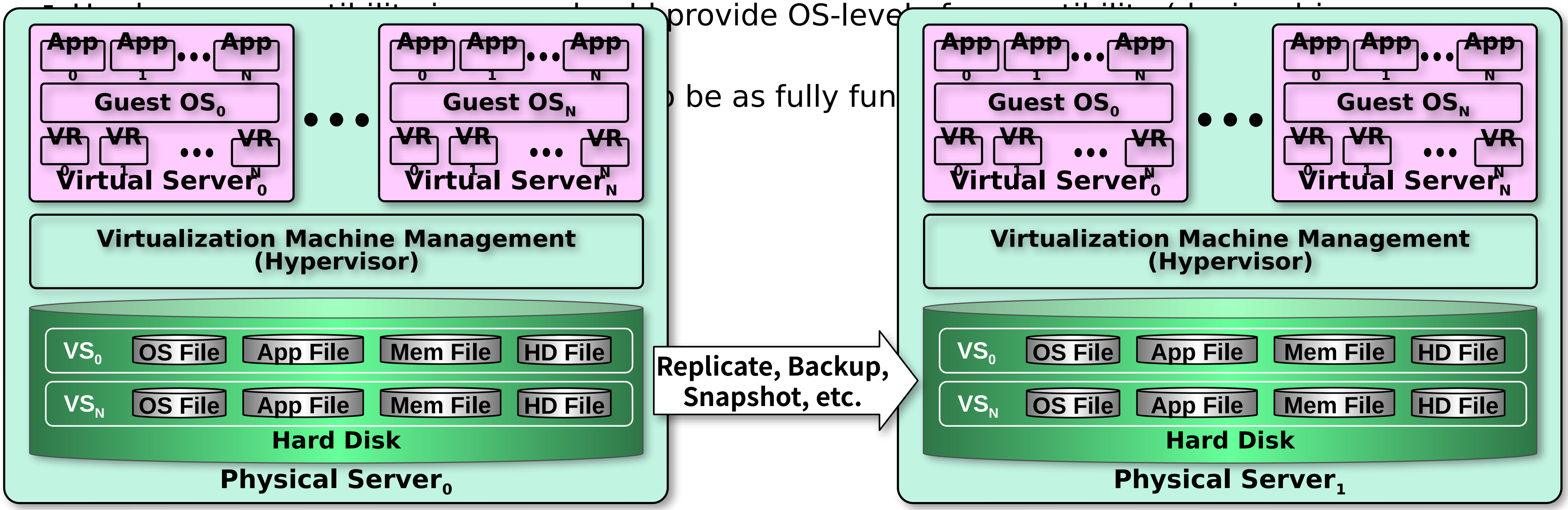
Virtualization Technology – 4/5

    - IT resource (CPU, Memory, etc.) sharing with host and guest OSs
    - Several additional traverse for each system call
  - Additional license cost for host OS (Windows license or Linux subscription)

- **Hardware-based virtualization (Bare Metal)**
  - Install virtualization software (**Hypervisor**) on physical (bare metal) host hardware directly (no host OS)
  - Eliminating one layer between host hardware and virtual servers reducing performance overhead
  - Optimized and minimized thin software layer that handles hardware management functions to provide virtualization management - a sort of designated system, not a general purpose operating system

provide OS-leve

be as fully fun



Virtualization Technology – 5/5

- **Virtualization management**
  - Easier to administrate virtual servers than physical servers
  - Many administration tasks automated by virtualization software
  - VIM (Virtualization Infrastructure Management) tools – collectively manage virtual IT resources from a centralized & dedicated management computer (controller)
- **Other consideration**
  - Performance overhead
    - Not ideal for complex systems with heavy workload
    - Excessive or unnecessary performance overhead with poorly formulated virtualization plan
    - Para-virtualization APIs – modified to reduce the guest OS’s processing overhead need to customize guest OSs to adapt them at the cost of sacrificing portability
  - Special hardware compatibility
    - There are many vendors supplying specialized hardware devices and not all of them are compatible with the given virtualization software.
    - Old and existing software may not support those hardware recently released.
    - Solution: standardization, commoditization and frequent virtualization software update/upgrade
  - Portability
    - Poor portability due to automated & programmatic VS management interfaces for their own
    - Demand for international standard such as OVF (Open Virtualization Format) for standardization of virtual disk formats in order to insure a wide range of VS management portability

Cloud Computing

End of Lecture Note