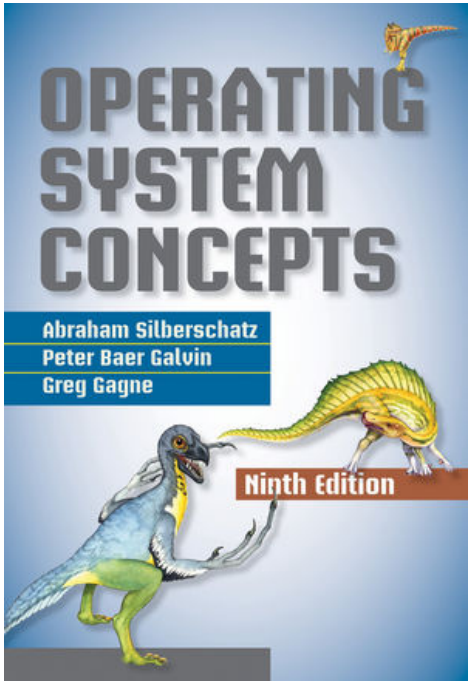
10 – Networks and Distributed Systems



EECE 315 (101)
ECE – UBC
2013 W1

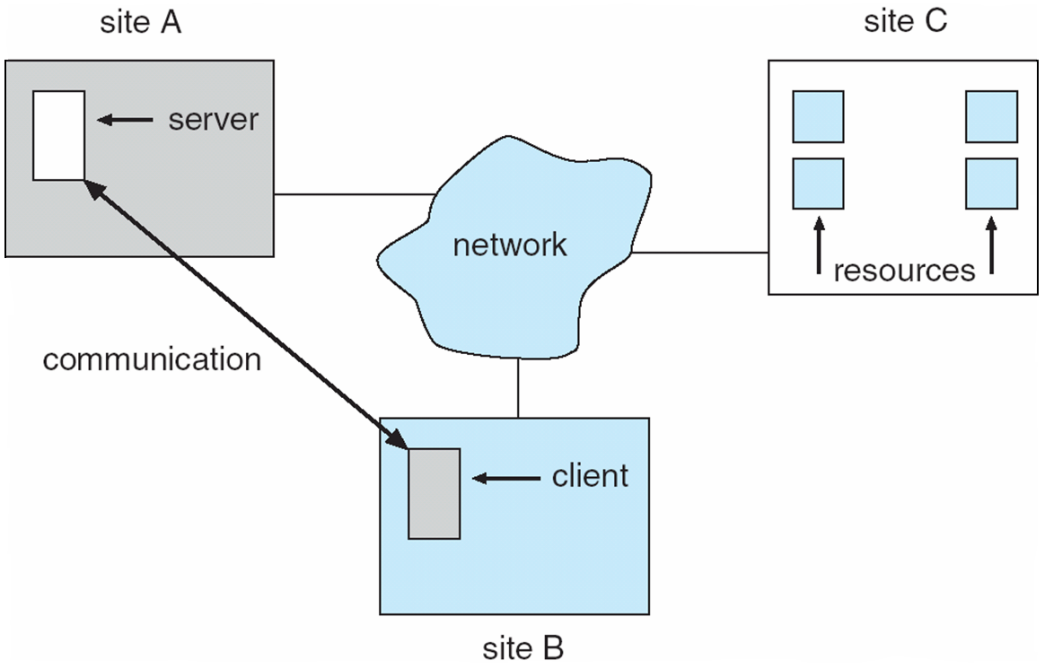
Acknowledgement: This set of slides is partly based on the PPTs provided by the Wiley's companion website (including textbook images, when not explicitly mentioned/referenced).

Chapter 17: Distributed Systems

- **Motivation**
- Types of Network-Based Operating Systems
- Network Structure
  - Communication Structure
  - Communication Protocols
- TCP/IP
- Distributed File System

Motivation

- A **Distributed system** is the collection of loosely coupled nodes interconnected by a communications network
- The nodes in a distributed system may vary in size and function. They are called *nodes, computers, machines, hosts*
  - *Site* is the location of a machine, while *node* refer to a specific system at a site.



Motivation: A Distributed System

- There are four reasons for building distributed systems:
  - **Resource sharing**
    - sharing and printing files at remote sites
    - processing information in a distributed database
    - using remote specialized hardware devices
  - **Computation speedup** – load sharing
  - **Reliability** – detect and recover from site failure, function transfer, reintegrate failed site
  - **Communication** – message passing

- Types of Distributed Operating Systems
  - Network Operating Systems
  - Distributed Operating Systems

Chapter 17: Distributed Systems

- Motivation
- **Types of Network-Based Operating Systems**
- Network Structure
  - Communication Structure
  - Communication Protocols
- TCP/IP
- Distributed File System

Types of Distributed Operating Systems

- Network Operating Systems (e.g. general purpose OS …)
  - Users are aware of multiplicity of machines.
  - Access to resources of various machines is done explicitly by:
    - Remote logging into the appropriate remote machine (telnet, ssh)
    - Remote Desktop (Microsoft Windows)
    - Transferring data from remote machines to local machines, via some file transfer protocol mechanism
- Distributed Operating Systems (e.g. NFS network file system)
  - Users not aware of multiplicity of machines
    - Access to remote resources similar to access to local resources
  - Data Migration (transfer data): by transferring entire file, or transferring only those portions of the file necessary for the immediate task
  - Computation Migration: transfer the computation, rather than the data, across the system

Distributed-Operating Systems (Cont.)

- Process Migration: execute an entire process, or parts of it, at different sites
  - **Load balancing** – distribute processes across network to even the workload
  - **Computation speedup** – subprocesses can run concurrently on different sites
  - **Hardware preference** – process execution may require specialized processor
  - **Software preference** – required software may be available at only a particular site
  - **Data access** – run process remotely, rather than transfer all data locally

Chapter 17: Distributed Systems

- Motivation
- Types of Network-Based Operating Systems
- **Network Structure**
  - Communication Structure
  - Communication Protocols
- TCP/IP
- Distributed File System

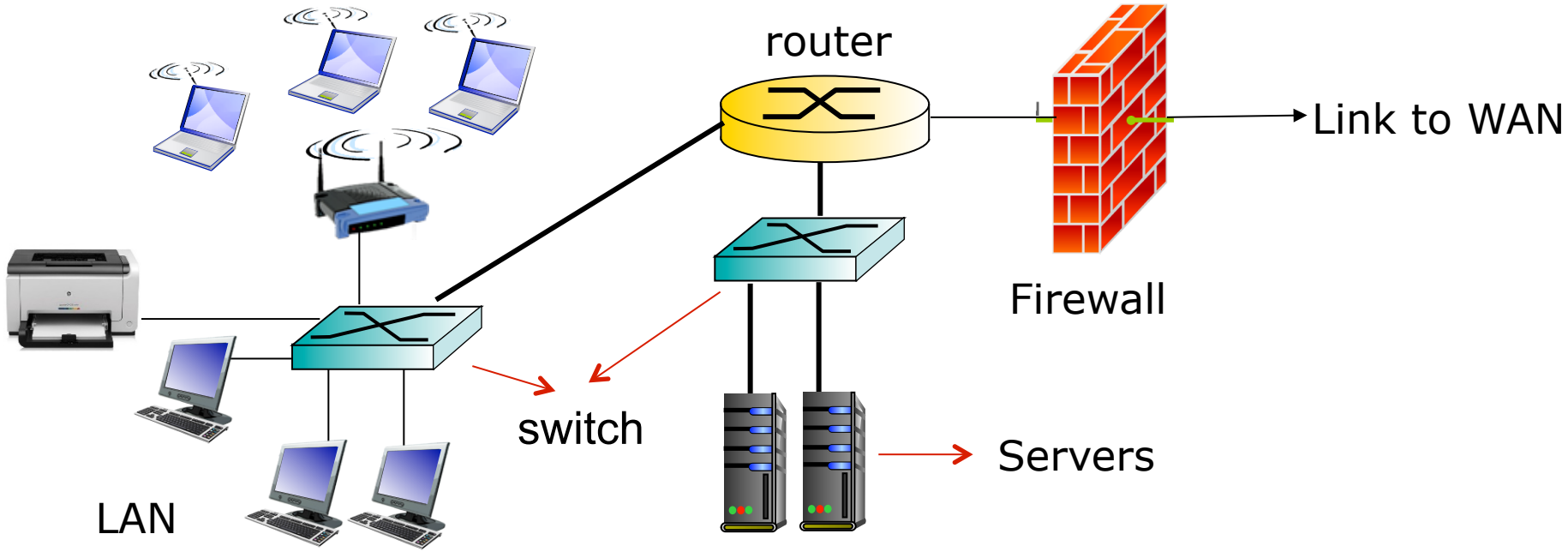
Network Structure

- There are two basic types of networks based on how they are geographically distributed:
  - Local-Area Network (LAN)
  - Wide-Area Network (WAN)

- **Local-Area Network** (**LAN**) – designed to cover small geographical area.
  - Multi-access bus, ring, or star network
  - Speed: ~ 10 megabits/sec to a few gigabits/sec
  - Broadcast is fast and cheap
  - Nodes:
    - usually workstations and/or personal computers
    - a few (usually one or two) larger computers/servers

An Example of a Typical LAN

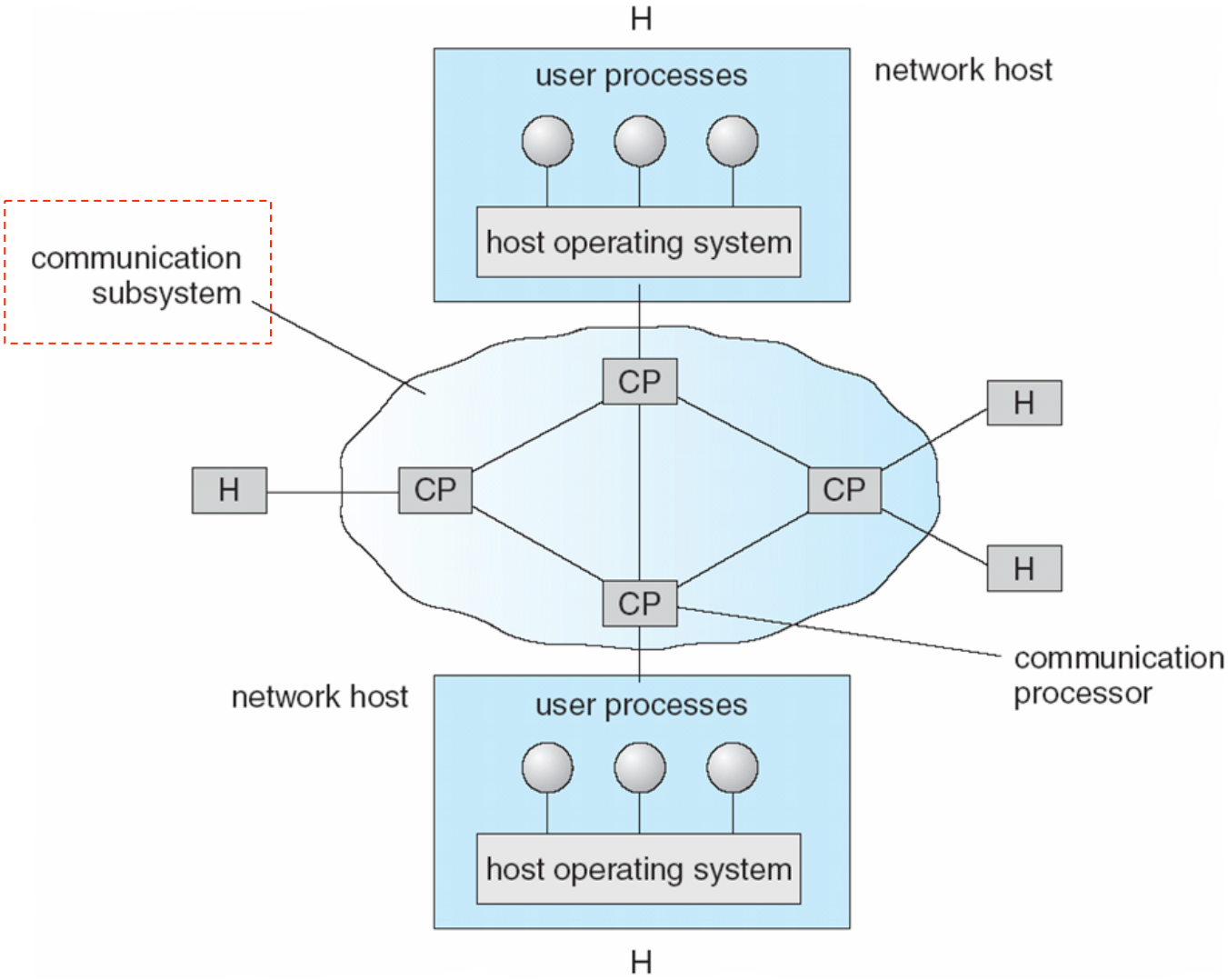
- A typical LAN may consist of a number of different
  - computers or such
  - various shared peripherals
  - switches, routers and links (wired or WiFi))
  - firewall, …



Network Types (Cont.)

- **Wide-Area Network** (**WAN**) – links geographically separated sites
  - Point-to-point connections over long-haul lines (often leased from a phone or cable company)
  - Speed varies depending on the link: DSL, cable, optical, …
  - Broadcast usually requires multiple messages
  - Nodes:
    - usually include a high percentage of large computing devices

Routers in a Wide-Area Network



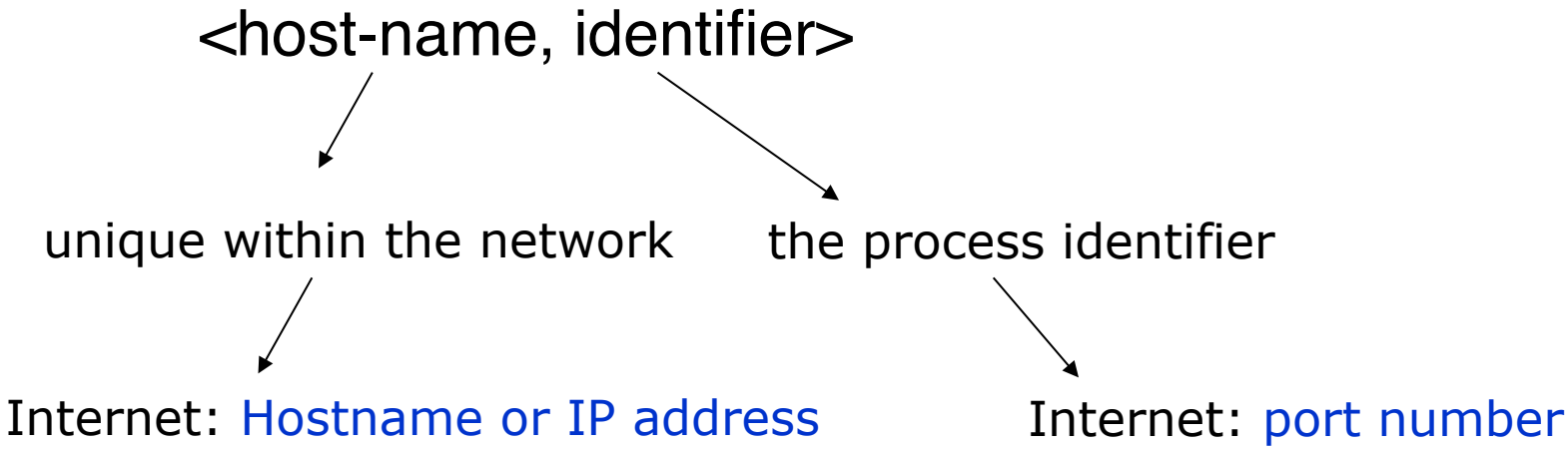
Communication Structure

The design of a *communication* network must address four basic issues:

- **Naming and name resolution** - How do two processes locate each other to communicate?
- **Routing strategies** - How are messages sent through the network?
- **Connection strategies** - How do two processes send a sequence of messages?
- **Contention -** The network is a shared resource, so how do we resolve conflicting demands for its use?
- We will focus on the Internet to discuss these issues.

Naming and Name Resolution

- For a process at node A to exchange information with another process at node B, each must be able to specify the other.
- Identify processes on remote systems by the pair:



- **Domain name service** (**DNS**) – specifies the naming structure of the hosts, as well as name to address resolution (Internet)

Routing Strategies

- We can choose between fixed routing or dynamic routing.
- **Fixed routing** - A path from *A* to *B* is specified in advance.
- **Dynamic routing** - The path used to send a message form site *A* to site *B* is chosen only when a message is sent
  - Usually a site sends a message to another site on the link least used at that particular time or the shortest path
  - Adapts to load changes by avoiding routing messages on heavily used path
  - Messages may arrive out of order
    - This problem can be remedied by appending a sequence number to each message

Connection Strategies

- A method called packet switching is used in the Internet for tis connection strategy.
- **Packet switching** - Messages of variable length are divided into fixed-length packets which are sent to the destination
  - Each packet may take a different path through the network
  - The packets must be reassembled into messages as they arrive
- An alternative is circuit switching which is similar to the normal telephony model.

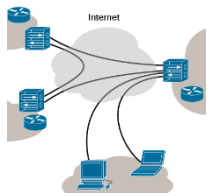
Contention

Several nodes may want to transmit information over a link simultaneously. A widely used technique to avoid repeated collisions is CSMA/CD:

- **CSMA/CD** - Carrier sense with multiple access (CSMA); collision detection (CD)
  - **Human analogy**: the polite conversationalist (listen before speaking + stop speaking if the other party starts to speak too)
  - A site determines whether another message is currently being transmitted over that link. If two or more sites begin transmitting at exactly the same time, then they will register a CD and will stop transmitting
  - When the system is very busy, many collisions may occur, and thus performance may be degraded
- CSMA/CD is used successfully in the Ethernet system, the most common network system

Communication Networks: Internet

- A computer network (such as the Internet) is usually based on a layered architecture.
  - Each layer is a well-defined, specific part of a large and complex system.
  - Each layer is associated with a network protocol.
  - The layered architecture would allow dividing the functionality into layers that are easier to design, implement and upgrade.
    - Simplification, abstraction and modularity
- Wireless Internet



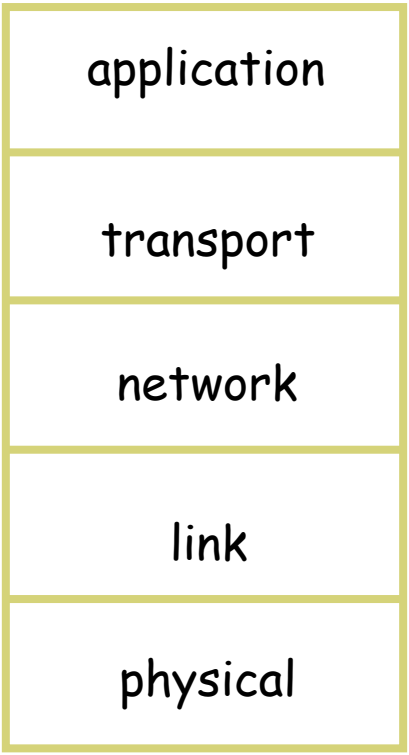
Internet

Email
Web
Skype
Facebook
Twitter
...

+ Wireless Technology = Wireless Internet

Internet Protocol Stack

- A protocol defines the format and order of messages exchanged. The protocols of various layers of the Internet is called the Internet protocol stack (shown in the figure).
  - application: supporting network applications
    - FTP, SMTP, HTTP
  - transport: process-process data transfer
    - TCP, UDP
  - network: routing of datagrams from source to destination
    - IP, routing protocols
  - link: data transfer between neighboring network elements
    - Ethernet
  - physical: bits “on the wire or air”



Chapter 17: Distributed Systems

- Motivation
- Types of Network-Based Operating Systems

- Network Structure
  - Communication Structure
  - Communication Protocols

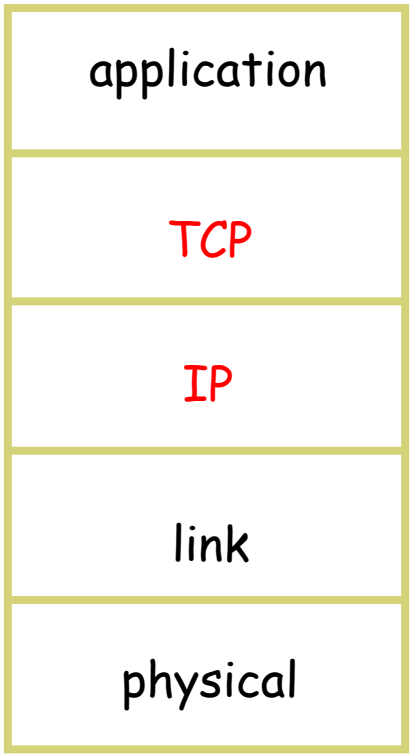
- **TCP/IP**

- Distributed File System

The TCP/IP Protocol Layers

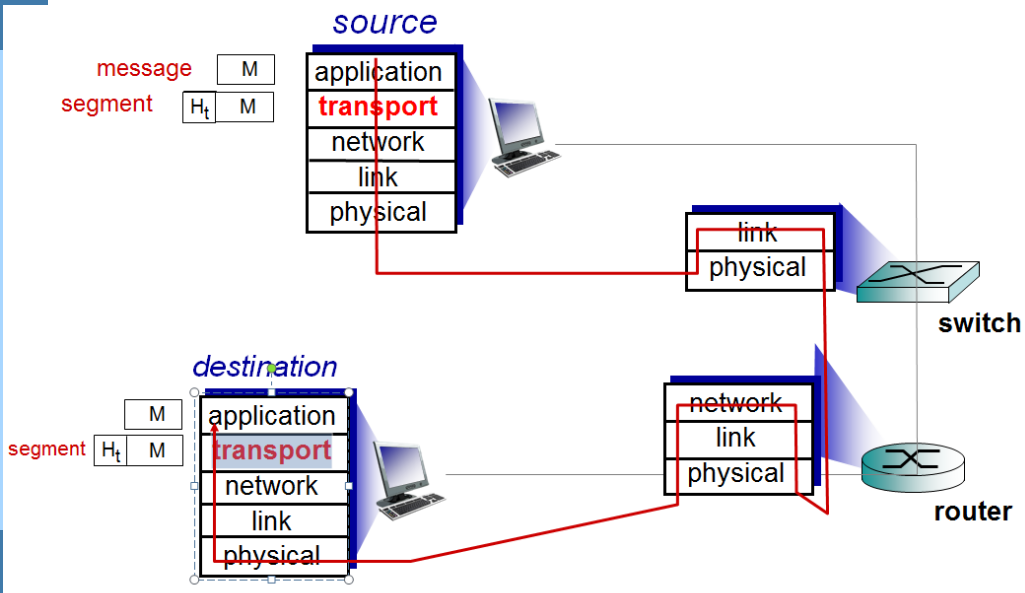
- TCP/IP (transmission control protocol/ Internet protocol) is the heart of the protocol hierarchy of the Internet.

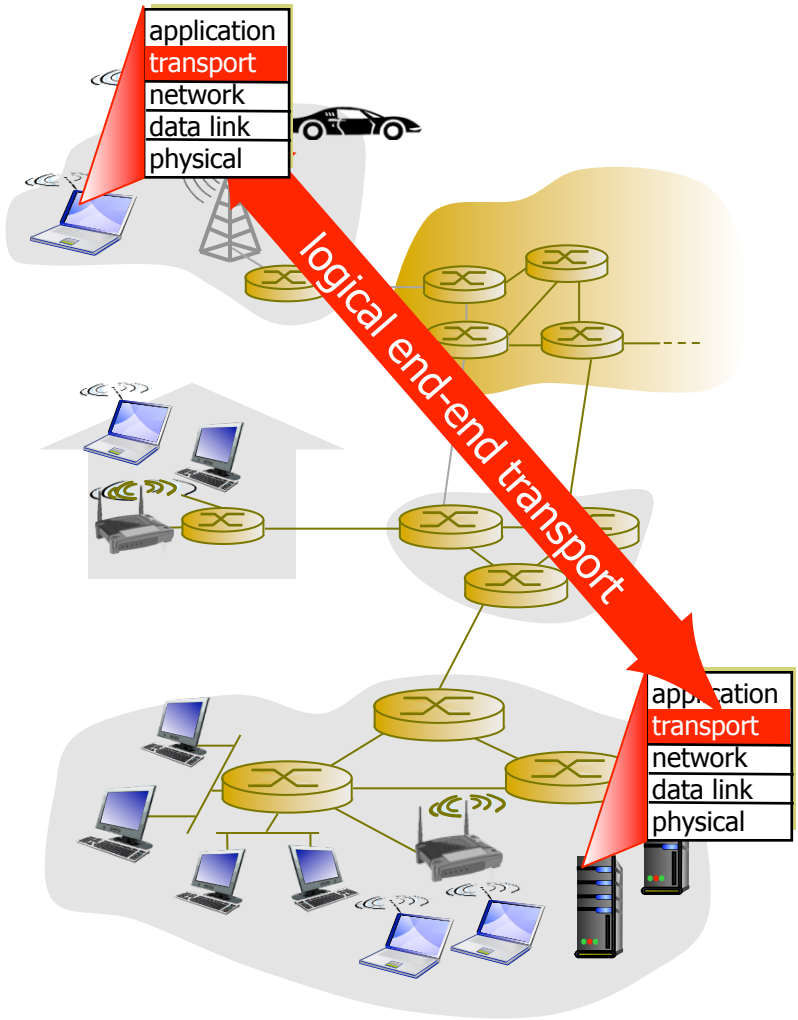
- TCP (at the transport layer) provides data transport from a process on a source machine to a process on a destination machine with a desired level of reliability.
  - It provides the abstractions that applications need to use the network.
  - It also provide flow-control and congestion control services and reliable transmission.



TCP (Transmission Control Protocol)

- Logical end-end transport
  - Process-to-process





By logical communication, we mean that from an application's perspective, it is as if the hosts running the processes were directly connected; in reality, the hosts may be on opposite sides of the planet, connected via numerous routers and a wide range of link types.

Network Layer: IP

- The network layer provides the host-to-host communication service.
  - unlike the transport and application layers, there is a piece of the network layer in each and every host and router in the network.

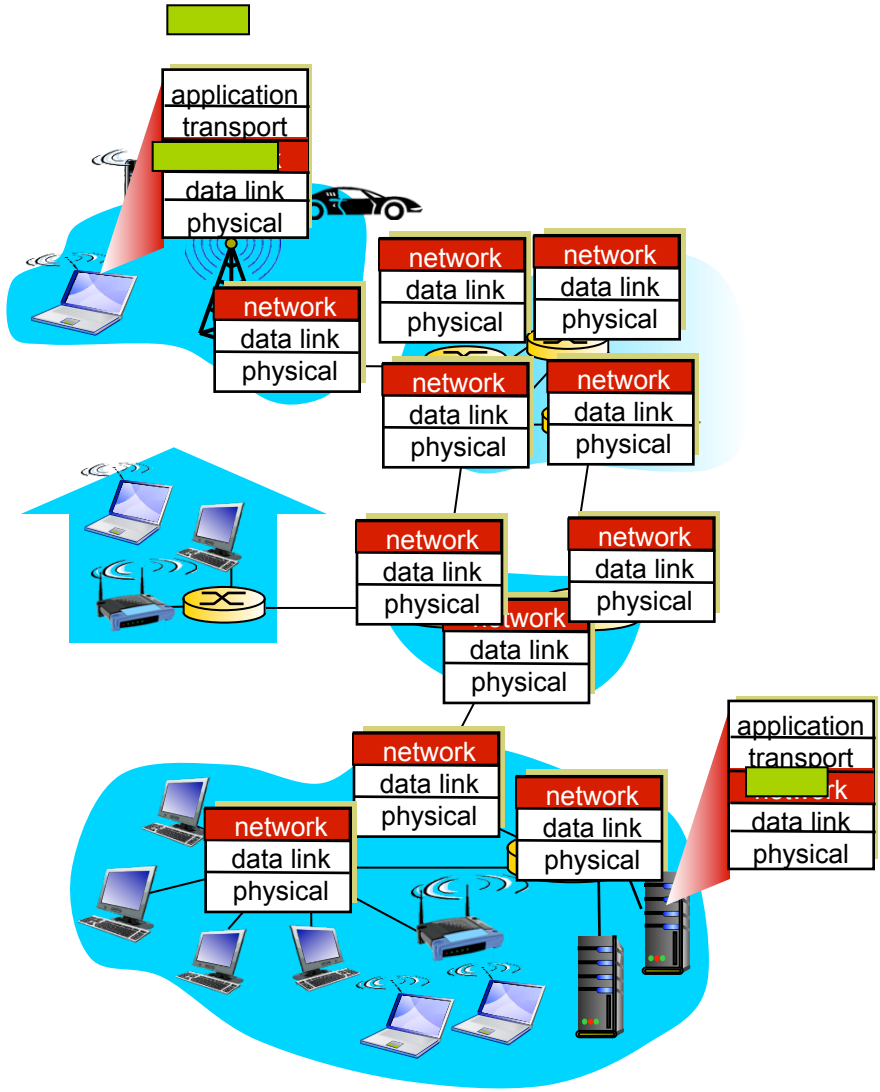
- IP (Internet Protocol) is the Internet's famous network layer protocol:
  - the IP service model is a best-effort delivery service.
  - this means that IP makes its "best effort" to deliver segments between communicating hosts, but it makes no guarantees.

IP (Internet Protocol)

- so IP does not guarantee segment delivery, it does not guarantee orderly delivery of segments, and it does not guarantee the integrity of the data in the segments.
  - For these reasons, IP is said to be an *unreliable service*.
- every host has at least one network-layer address, a so-called IP address.
- the network layer is one of the most complex layers in the protocol stack.

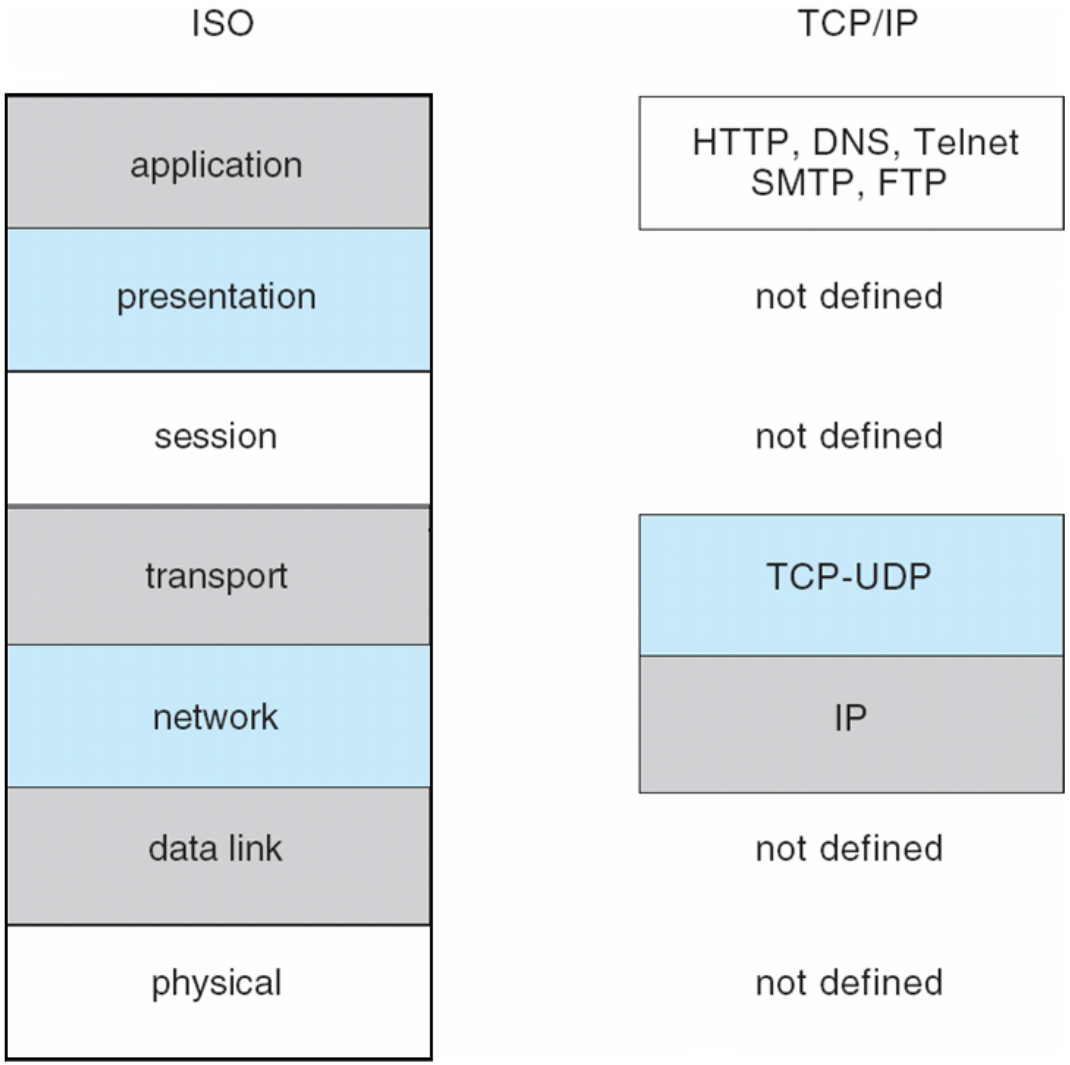
Network layer

- The network layer protocols reside in *every* host and router
- It transports the segments from sending to receiving host
- router examines the header fields in all IP datagrams that are passing through it



Internet: The TCP/IP Protocol Layers

OSI and TCP/IP protocol stack comparison:



OSI Protocol Layers

The OSI model (by international standard organization) defines 7 layers for a communication network:

- **Physical layer** – handles the mechanical and electrical details of the physical transmission of a bit stream

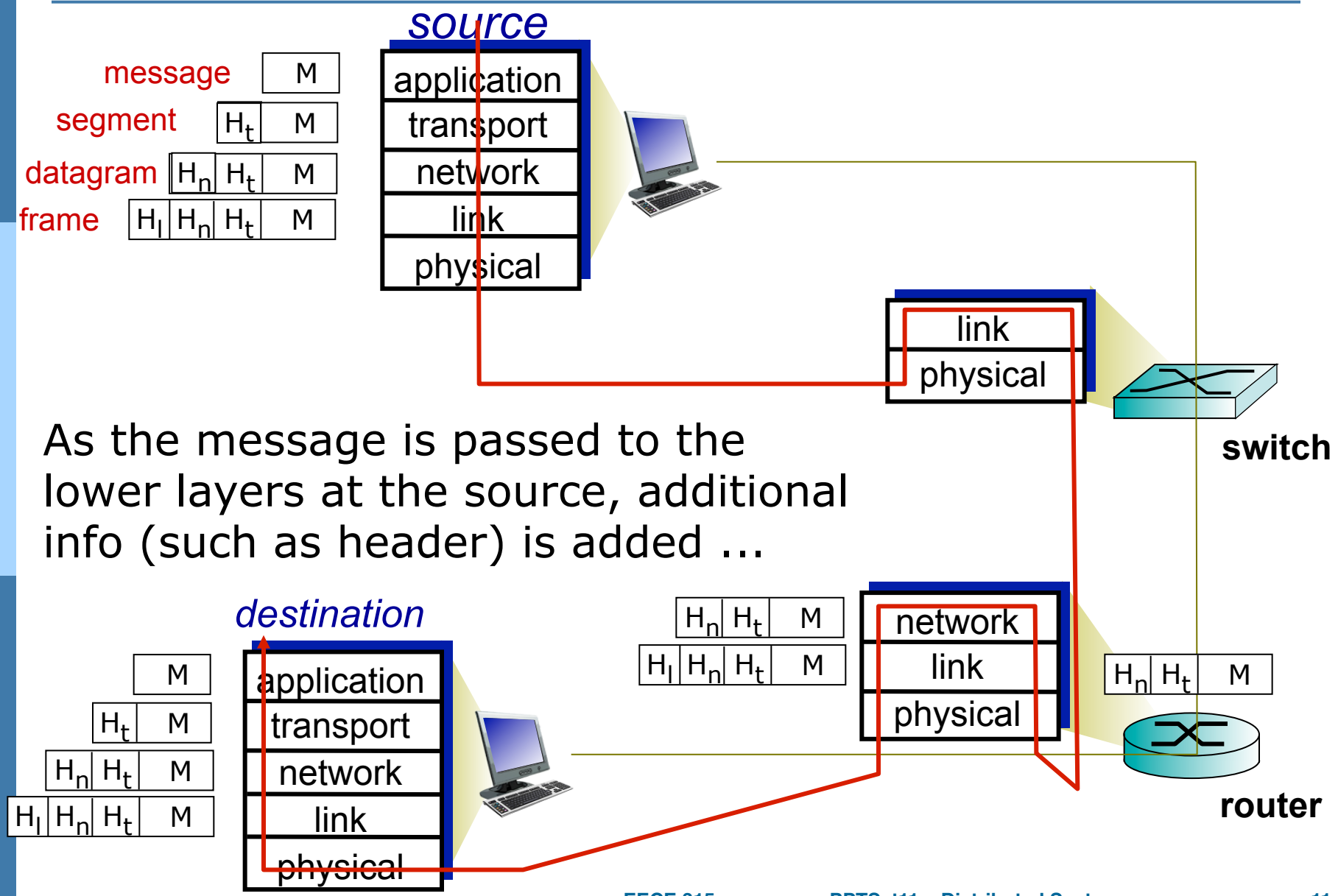
- **Data-link layer** – handles the *frames*, or fixed-length parts of packets, including any error detection and recovery that occurred in the physical layer

- **Network layer** – provides connections and routes packets in the communication network, including handling the address of outgoing packets, decoding the address of incoming packets, and maintaining routing information for proper response to changing load levels

Communication Protocol (Cont.)

- **Transport layer** – responsible for low-level network access and for message transfer between clients, including partitioning messages into packets, maintaining packet order, controlling flow, and generating physical addresses
- **Session layer** – implements sessions, or process-to-process communications protocols
- **Presentation layer** – resolves the differences in formats among the various sites in the network, including character conversions, and half duplex/full duplex (echoing)
- **Application layer** – interacts directly with the users' deals with file transfer, remote-login protocols and electronic mail, as well as schemas for distributed databases

Packets in the Internet



Example: Networking

- The transmission of a network packet between hosts on an Ethernet network
- Every host has a unique IP address and a corresponding Ethernet (MAC) address
- Communication requires both addresses
- Domain Name Service (DNS) can be used to acquire IP addresses
- Address Resolution Protocol (ARP) is used to map MAC addresses to IP addresses
- If the hosts are on the same network, ARP can be used
  - If the hosts are on different networks, the sending host will send the packet to a *router* which routes the packet to the destination network

Chapter 17: Distributed Systems

- Motivation
- Types of Network-Based Operating Systems

- Network Structure
  - Communication Structure
  - Communication Protocols

- TCP/IP

- **Distributed File System**

Distributed File system

- Naming and Transparency
- Remote File Access

Background

- **Distributed file system** (**DFS**) is a distributed implementation of the classical time-sharing model of a file system, where multiple users share files and storage resources
- A DFS manages a set of dispersed storage devices
- Overall storage space that is managed by a DFS is composed of different, remotely located, smaller storage spaces
- There is usually a correspondence between constituent storage spaces and sets of files

DFS Structure

- **Service** – software entity running on one or more machines and providing a particular type of function to a priori unknown clients
- **Server** – service software running on a single machine
- **Client** – process that can invoke a service using a set of operations that forms its client interface
- A client interface for a file service is formed by a set of primitive file operations (create, delete, read, write)
- Client interface of a DFS should be transparent, i.e., not distinguish between local and remote files

Naming and Transparency

- **Naming** – mapping between logical and physical objects
- **Multilevel mapping** – abstraction of a file that hides the details of how and where on the disk the file is actually stored
- A **transparent** DFS hides the location where in the network the file is stored
- For a file being replicated in several sites, the mapping returns a set of the locations of this file's replicas; both the existence of multiple copies and their location are hidden

Naming Structures

- **Location transparency** – file name does not reveal the file's physical storage location

- **Location independence** – file name does not need to be changed when the file's physical storage location changes

Naming Schemes — Three Main Approaches

1. Files named by combination of their host name and local name; guarantees a unique system-wide name

2. Attach remote directories to local directories, giving the appearance of a coherent directory tree; only previously mounted remote directories can be accessed transparently

3. Total integration of the component file systems
   - A single global name structure spans all the files in the system
   - If a server is unavailable, some arbitrary set of directories on different machines also becomes unavailable

Remote File Access

- **Remote-service mechanism** is one transfer approach

- Reduce network traffic by retaining recently accessed disk blocks in a cache, so that repeated accesses to the same information can be handled locally
  - If needed data not already cached, a copy of data is brought from the server to the user
  - Accesses are performed on the cached copy
  - Files identified with one master copy residing at the server machine, but copies of (parts of) the file are scattered in different caches
  - **Cache-consistency problem** – keeping the cached copies consistent with the master file
    - Could be called **network virtual memory**

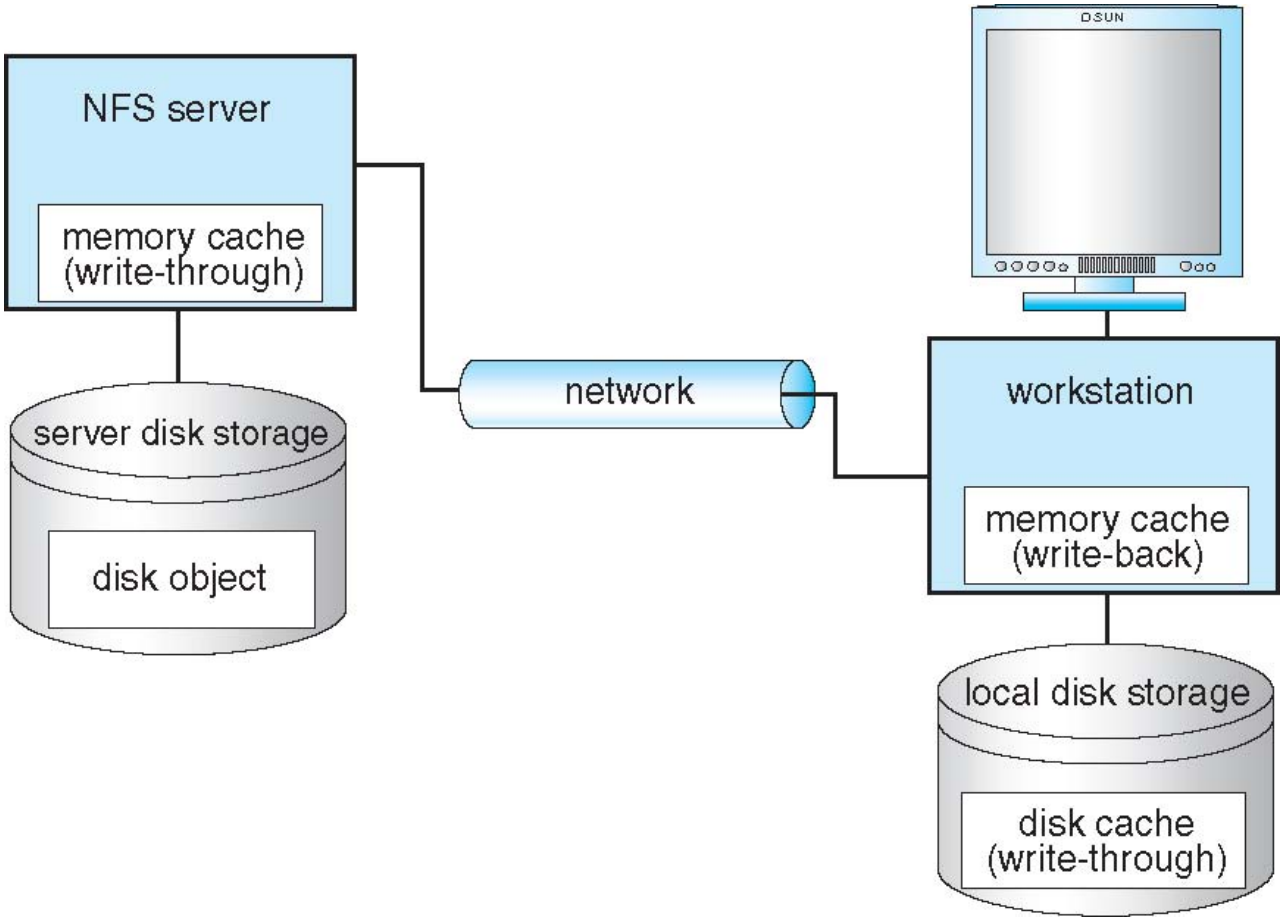
Cache Location – Disk vs. Main Memory

- Advantages of disk caches
  - More reliable
  - Cached data kept on disk are still there during recovery and don't need to be fetched again
- Advantages of main-memory caches:
  - Permit workstations to be diskless
  - Data can be accessed more quickly
  - Performance speedup in bigger memories
  - Server caches (used to speed up disk I/O) are in main memory regardless of where user caches are located; using main-memory caches on the user machine permits a single caching mechanism for servers and users

Cache Update Policy

- **Write-through** – write data through to disk as soon as they are placed on any cache
  - Reliable, but poor performance
- **Delayed-write** – modifications written to the cache and then written through to the server later
  - Write accesses complete quickly; some data may be overwritten before they are written back, and so need never be written at all
  - Poor reliability; unwritten data will be lost whenever a user machine crashes
  - Variation – scan cache at regular intervals and flush blocks that have been modified since the last scan
  - Variation – **write-on-close**, writes data back to the server when the file is closed
    - Best for files that are open for long periods and frequently modified

CacheFS and its Use of Caching



Consistency

- Is locally cached copy of the data consistent with the master copy?

- **Client-initiated approach**
  - Client initiates a validity check
  - Server checks whether the local data are consistent with the master copy

- **Server-initiated approach**
  - Server records, for each client, the (parts of) files it caches
  - When server detects a potential inconsistency, it must react

Comparing Caching and Remote Service

- In caching, many remote accesses handled efficiently by the local cache; most remote accesses will be served as fast as local ones
- Servers are contracted only occasionally in caching (rather than for each access)
  - Reduces server load and network traffic
  - Enhances potential for scalability
- Remote server method handles every remote access across the network; penalty in network traffic, server load, and performance
- Total network overhead in transmitting big chunks of data (caching) is lower than a series of responses to specific requests (remote-service)

Caching and Remote Service (Cont.)

- Caching is superior in access patterns with infrequent writes
  - With frequent writes, substantial overhead incurred to overcome cache-consistency problem
- Benefit from caching when execution carried out on machines with either local disks or large main memories
- Remote access on diskless, small-memory-capacity machines should be done through remote-service method
- In caching, the lower intermachine interface is different form the upper user interface
- In remote-service, the intermachine interface mirrors the local user-file-system interface

Chapter 17: Distributed Systems

- Motivation
- Types of Network-Based Operating Systems

- Network Structure
  - Communication Structure
  - Communication Protocols

- TCP/IP

- Distributed File System