Distributed Systems – CS249

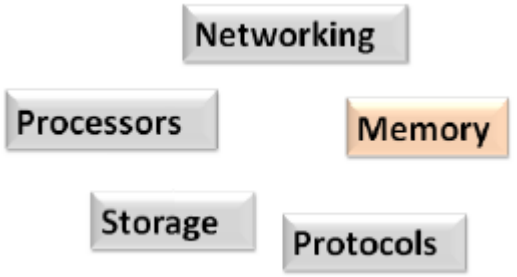
Introduction to Distributed Systems

Agenda

- Course Overview
- Characterization and Motivation of Distributed Systems
- Architecture Models
- Networking and Inetrnetworking

Course Overview

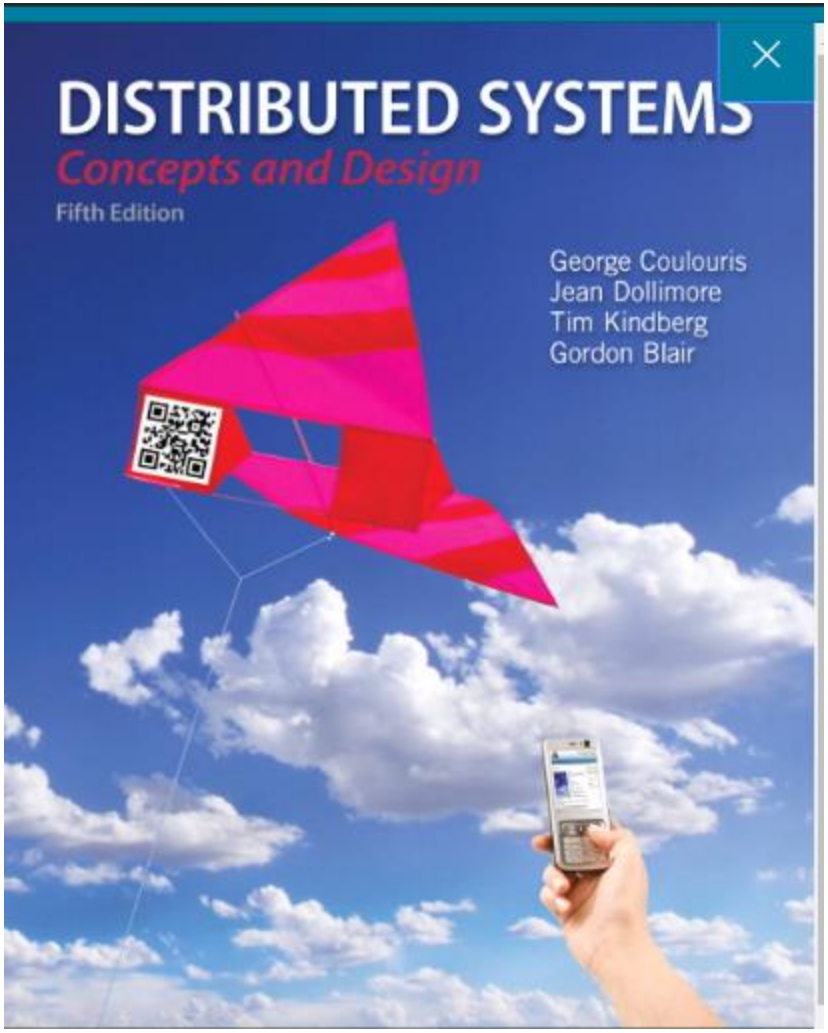
Course Overview



- Introduction to Distributed Systems
- Inter-process Communication
- Distributed Objects and Components
- Security
- Distributed Shared Memory
- Distributed File System
- Name Services
- Time and Global States
- Coordination and Agreement
- Transactions and Concurrency Control
- Replication
- **Case Study:** Google, Paxos, MapReduce, Zookeeper.

Textbook

- **Distributed Systems: Concepts and Design,** *by* George Coulouris, Jean Dollimore, Tim Kindberg and Gordon Blair, *5th Edition, Addison Wesley/Pearson (2012)*
  - ***ISBN-13:*** *978-0-13-214301-1 or*
    ISBN-10: *0-13-214301-1.*


DISTRIBUTED SYSTEMS
Concepts and Design
Fifth Edition
George Coulouris
Jean Dollimore
Tim Kindberg
Gordon Blair

Grading

Total	100%	Comments
Homework (5)	20%	Individual scores
Project (1)	10%	Group Project
Research Paper (1)	15%	Group Paper
Random Quizzes	10%	Individual scores
Midterm	20%	Individual scores
Final (Project Demo)	25%	Individual (10) + Group (15)

- Final Letter grade is based on relative grading within the class.
- Participation in the class is expected and is important.
- Assignments late submission will result in losing 10% of the assignment grade per day late and assignments are not accepted after 5 days late.
- YOU ARE RESPONSIBLE TO REMEMBER DUE DATES (SYLLABUS)

Can Not Attend?

- **If you**
  - cannot finish an assignment, or
  - cannot write any quiz or exam,

- you must send me an email **before** the event
  - explaining the reason and be accepted by myself and I need to approve it.
  - Late explanation will not be accepted

Emailing

- **Subject** line:
  - must start with **SJSU – CS249 Sec 1, Spring18**,
  - followed by your student ID#, then the actual subject line
- **Body**: always include your full name
- Remind me if you do not receive any reply within 24 hrs, and you think I might have missed it.

Sample Email

FROM: Joe Mohammed

TO: Ahmed.Ezzat@sjsu.edu

Subject: SJSU CS249, Section 1, Spring18 - About assignment #1

Body:

Dear Professor Ezzat,

I am Joe Mohammed (student ID # A0001) in class CS249, Spring 2018. I have a question about …

Course Home Page

https://sjsu.instructure.com/courses/1257602

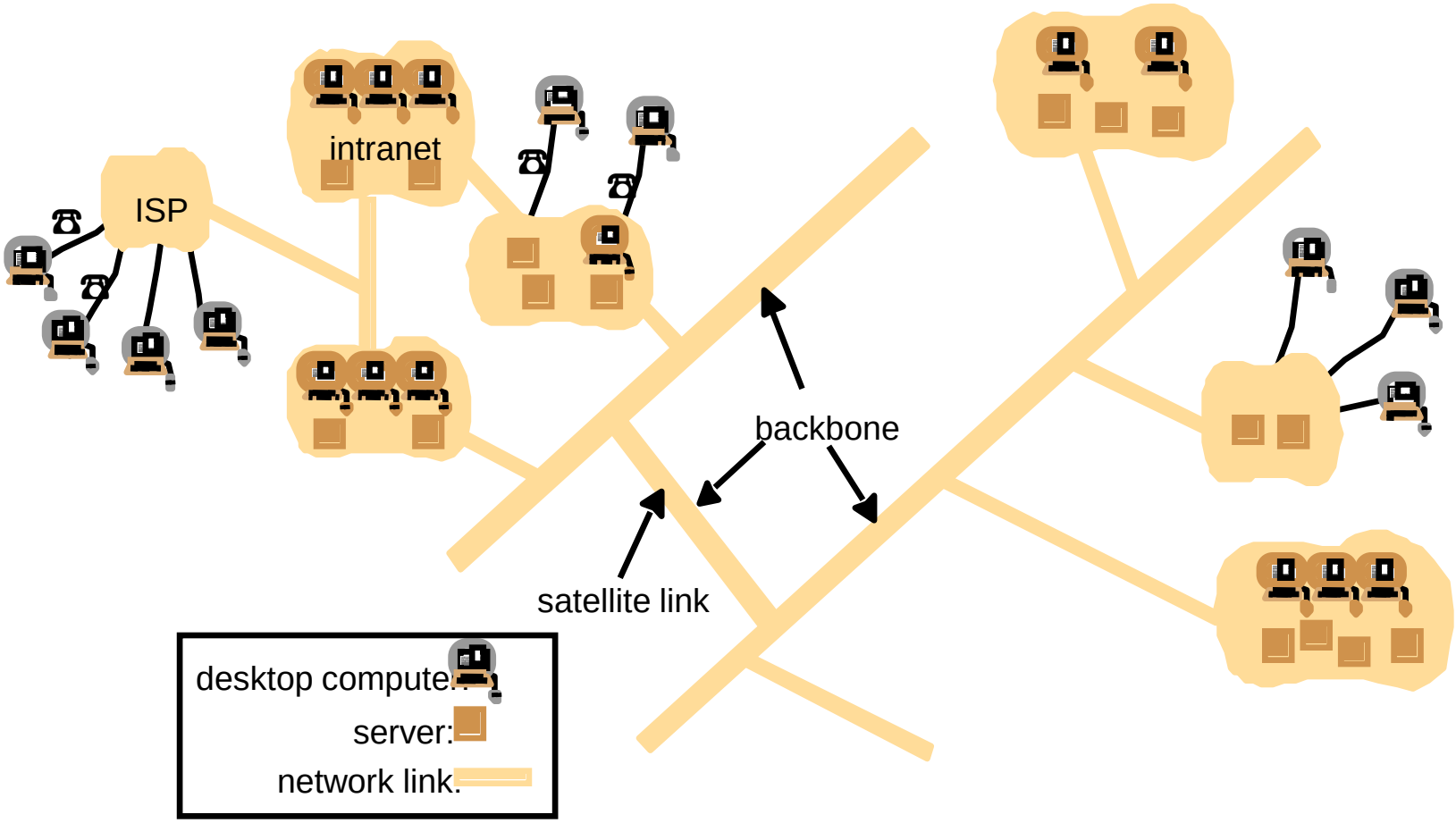
All class material are posted including Homework Assignments, Some class lecture notes (PDF) presentations, Syllabus and the Green sheet.

Characterization and Motivation of Distributed Systems

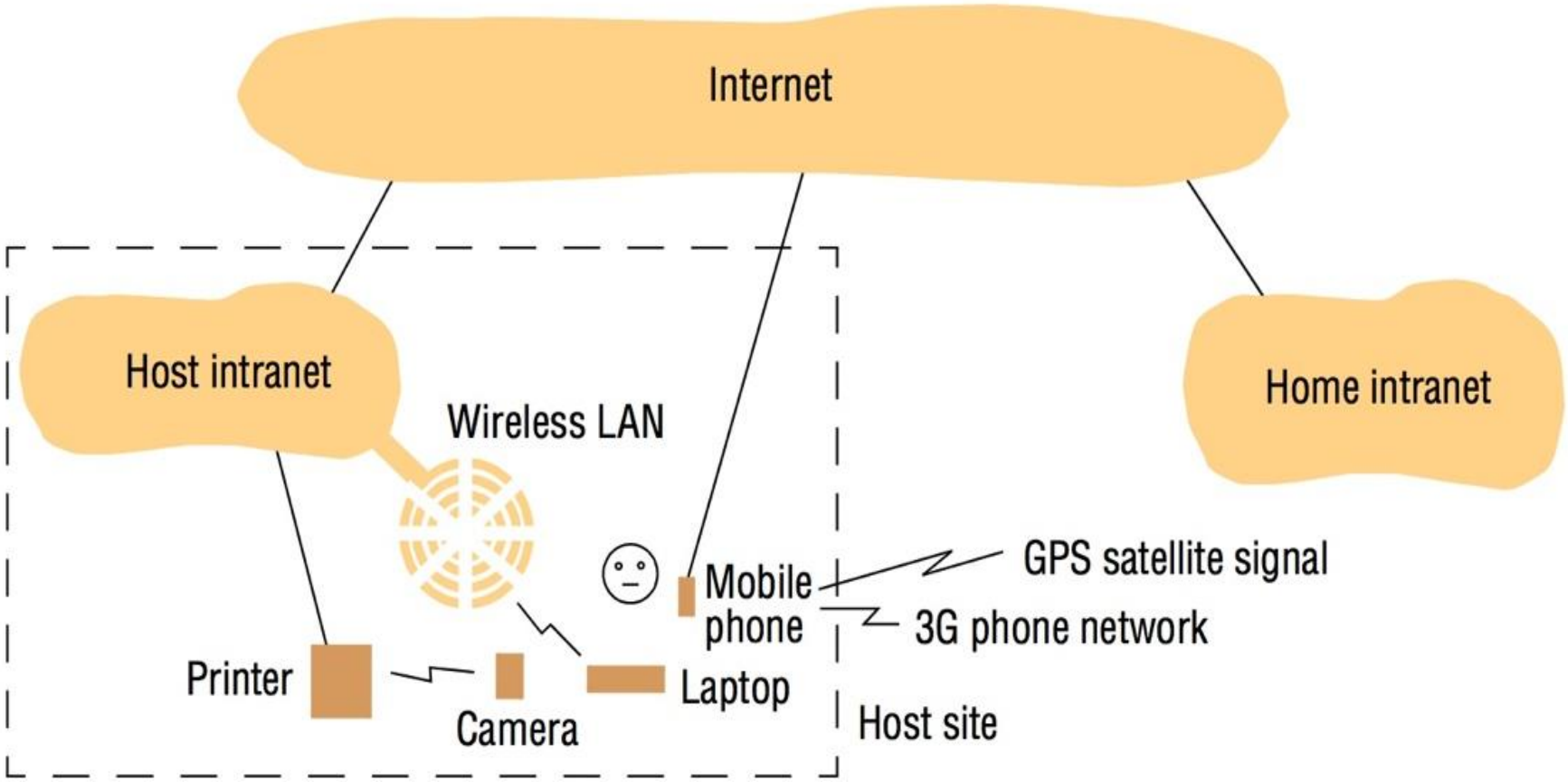
Selected Application Domains and Associated Networked Applications

Finance and commerce	eCommerce e.g. Amazon and eBay, PayPal, online banking and trading
The information society	Web information and search engines, ebooks, Wikipedia; social networking: Facebook and MySpace.
Creative industries and entertainment	online gaming, music and film in the home, user-generated content, e.g. YouTube, Flickr
Healthcare	health informatics, on online patient records, monitoring patients
Education	e-learning, virtual learning environments; distance learning
Transport and logistics	GPS in route finding systems, map services: Google Maps, Google Earth
Science	The Grid as an enabling technology for collaboration between scientists
Environmental management	sensor technology to monitor earthquakes, floods or tsunamis

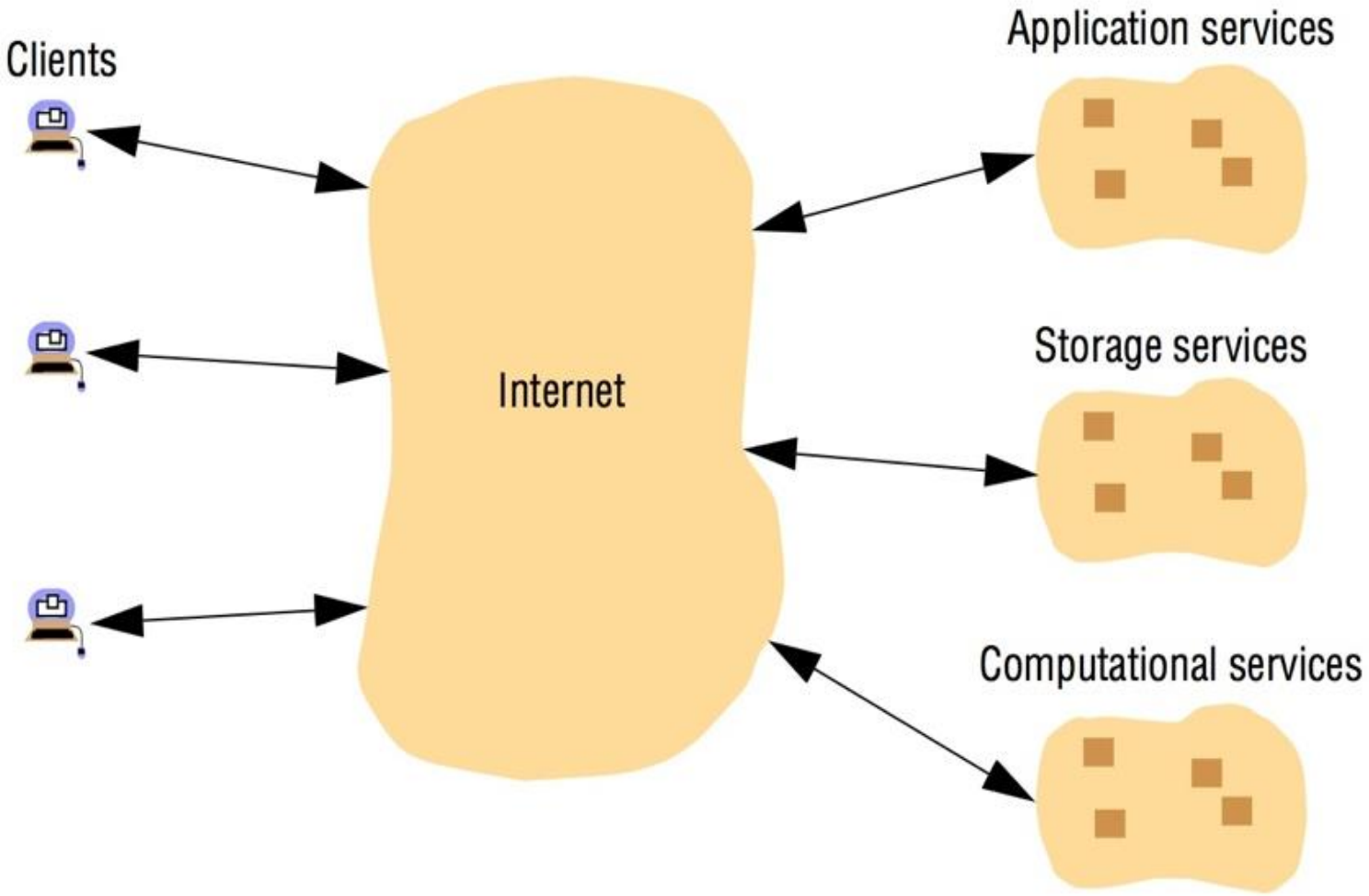
A Typical Portion of the Internet



Portable and Handheld Devices in Distributed System



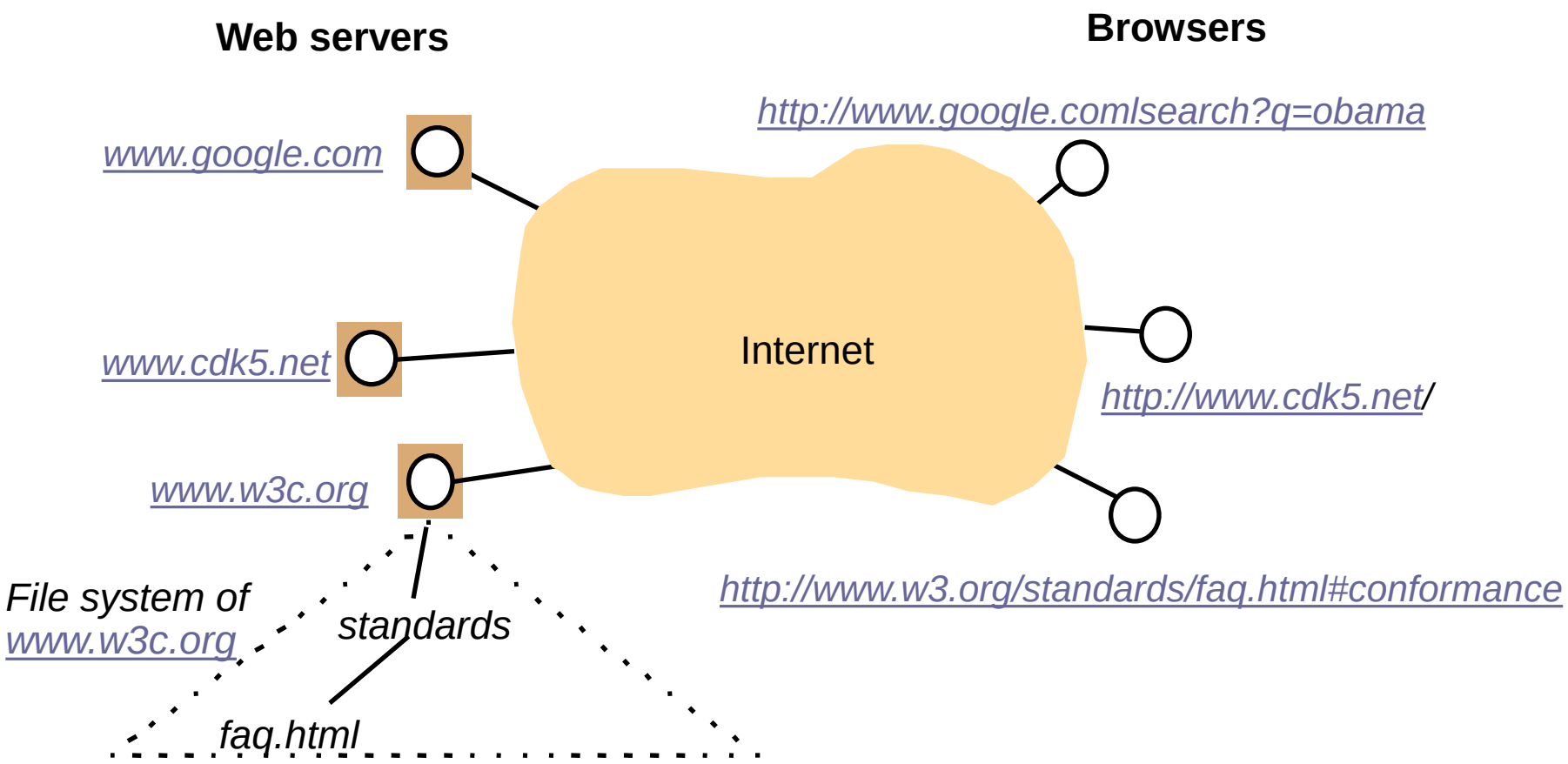
Cloud Computing



Growth of the Internet (Computers and Web Servers)

Date	*Computers*	*Web servers*	*Percentage Web servers / Computers*
1993, July	1,776,000	130	0.008
1995, July	6,642,000	23,500	0.4
1997, July	19,540,000	1,203,096	6
1999, July	56,218,000	6,598,697	12
2001, July	125,888,197	31,299,592	25
2003, July	~200,000,000	42,298,371	21
2005, July	353,284,187	67,571,581	19

Web Servers and Web Browsers



Distributed Systems Definition

- *"You know you have a distributed system when the crash of a computer you've never heard of stops you from getting any work done."*
  (**Leslie Lamport**, Distribution email, May 28, 1987, http://research.microsoft.com/users/lamport/pubs/distributed_systems.txt).
- "*A collection of computers that do not share common memory or a common physical clock, that communicate by a messages passing over a communication network, and where each computer has its own memory and runs its own operating system. Typically the computers are semi-autonomous and are loosely coupled while they cooperate to address a problem collectively*"

Distributed Systems Definition (Contd.)

(**M. Singhaland N. Shivaratri**, Advanced Concepts in Operating Systems, New York, McGraw Hill, 1994)

- "*A collection of independent computers that appears to the users of the system as a single coherent computer*." (**A. Tanenbaum** and **M. Van Steen**, Distributed Systems: Principles and Paradigms, Upper Saddle River, NJ, Prentice-Hall, 2003)

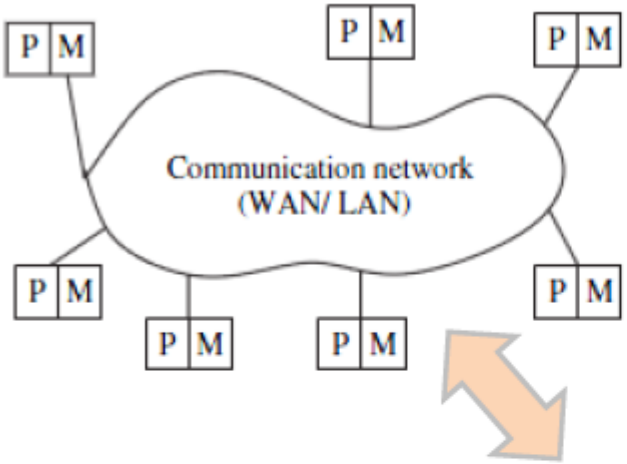
Distributed Systems Characteristics

- **A distributed system** - a collection of autonomous processors which communicate through a network and has the following characteristics:
  - **There is no common physical clock** ➔ asynchronies between processors
  - **There is no shared memory** ➔ m*essage-passing* mechanism for communication Obs. Distributed systems can supply an abstraction regarding a common address space via *distributed shared memory*
  - **Geographical separation:**
    - Is not required for processors to be in the same WAN
    - NOW/COW (Network/Cluster of Workstations) from a LAN are popular because of low costs and higher transfer speed

Distributed Systems Characteristics

- **Autonomy and heterogeneousness:**
  - The processors are *loosely coupled :*
    - They have different speeds and they may use different operating systems
    - They do not belong to a dedicated system but they cooperate in order to solve a problem

Distributed Systems Characteristics



P processor(s)
M memory bank(s)

Figure. A distributed system that links nodes through a communication network

[A.Kshemkalyani , M. Singhal , Distributed Computing]

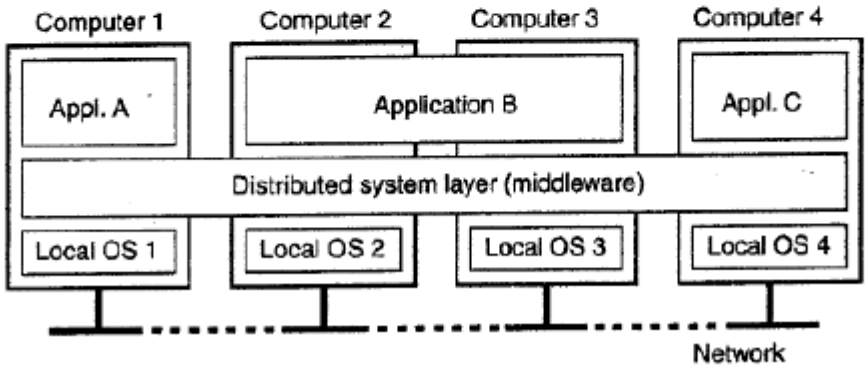


Figure. Distributed system - a generic architecture

Distributed Systems Characteristics

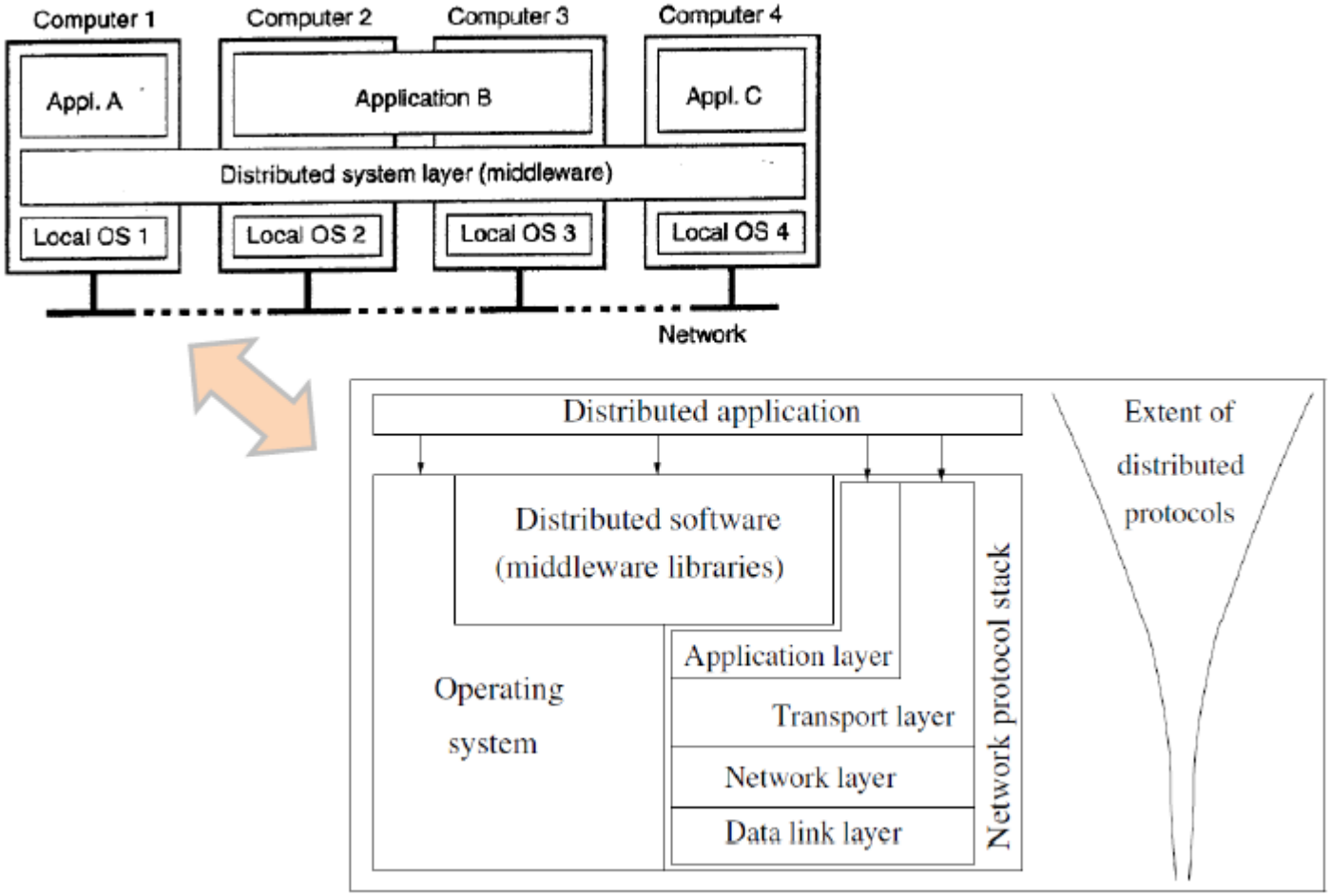


Figure. The interaction between software components

Distributed Systems Characteristics

- **Factors in the Design of a Distributed Systems:**
  - Resources Accessibility
  - Transparency: Access, location, concurrency, replication, failure, migration, performance, and scalability
  - Openness: interoperability, portability
  - Extensibility
  - CAP Theorem (Brewer): Consistency, Availability, Partition Tolerance). Only two out of the three can be supported.
  - False Assumptions: Network is reliable, network is secure, network is homogeneous, topology does not change, latency is zero, bandwidth is infinite, transport cost is zero!

Transparencies

- **Access transparency:** enables local and remote resources to be accessed using identical operations.
- **Location transparency:** enables resources to be accessed without knowledge of their physical or network location (for example, which building or IP address).
- **Concurrency transparency:** enables several processes to operate concurrently using shared resources without interference between them.

Transparencies

- **Replication transparency:** enables multiple instances of resources to be used to increase reliability and performance without knowledge of the replicas by users or application programmers.
- **Failure transparency:** enables the concealment of faults, allowing users and application programs to complete their tasks despite the failure of hardware or software components.
- **Mobility transparency:** allows the movement of resources and clients within a system without affecting the operation of users or programs.

Transparencies

- **Performance transparency:** allows the system to be reconfigured to improve performance as loads vary (elastic computing).
- **Scaling transparency:** allows the system and applications to expand in scale without change to the system structure or the application algorithms.

Distributed Systems Motivation

- Access to geographically remote data and resources
- Enhanced reliability
  - Availability
  - Integrity
  - Fault-tolerance
- Increased Performance/cost
- Scalability
- Modularity and incremental expandability

Architecture Models

Generations of Distributed Systems

Distributed systems:	*Early*	*Internet-scale*	*Contemporary*
Scale	Small	Large	Ultra-large
Heterogeneity	Limited (typically relatively homogenous configurations)	Significant in terms of platforms, languages and middleware	Added dimensions introduced including radically different styles of architecture
Openness	Not a priority	Significant priority with range of standards introduced	Major research challenge with existing standards not yet able to embrace complex systems
Quality of service	In its infancy	Significant priority with range of services introduced	Major research challenge with existing services not yet able to embrace complex systems

Communication Entities and Communication Paradigms

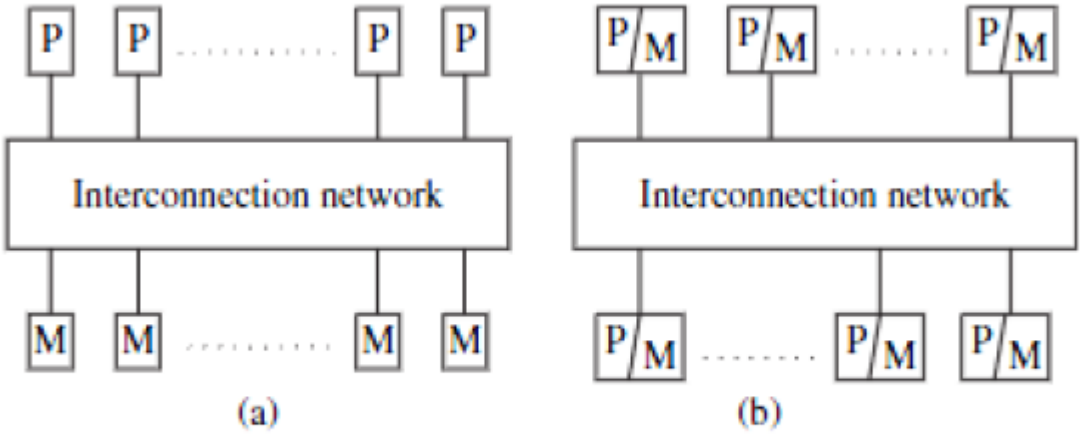
Communicating entities (what is communicating)		*Communication paradigms (how they communicate)*		
System-oriented entities	*Problem-oriented entities*	*Interprocess communication*	*Remote invocation*	*Indirect communication*
Nodes	Objects	Message passing	Request-reply	Group communication
Processes	Components	Sockets	RPC	Publish-subscribe
	Web services	Multicast	RMI	Message queues
				Tuple spaces
				DSM

Concepts and Architecture Aspects

- **Message Passing vs. Shared Memory Systems:**
  - **Shared memory systems:**
    - Shared address space
    - Communication between process is through the shared memory
    - We can have distributed shared memory (NUMA) – access remote memory is expensive
  - **Message-passing (MP) mechanism**
    - More scalable solution

Concepts and Architecture Aspects

- **Shared memory systems:** *Parallel multiprocessor vs. Multicomputer Systems*

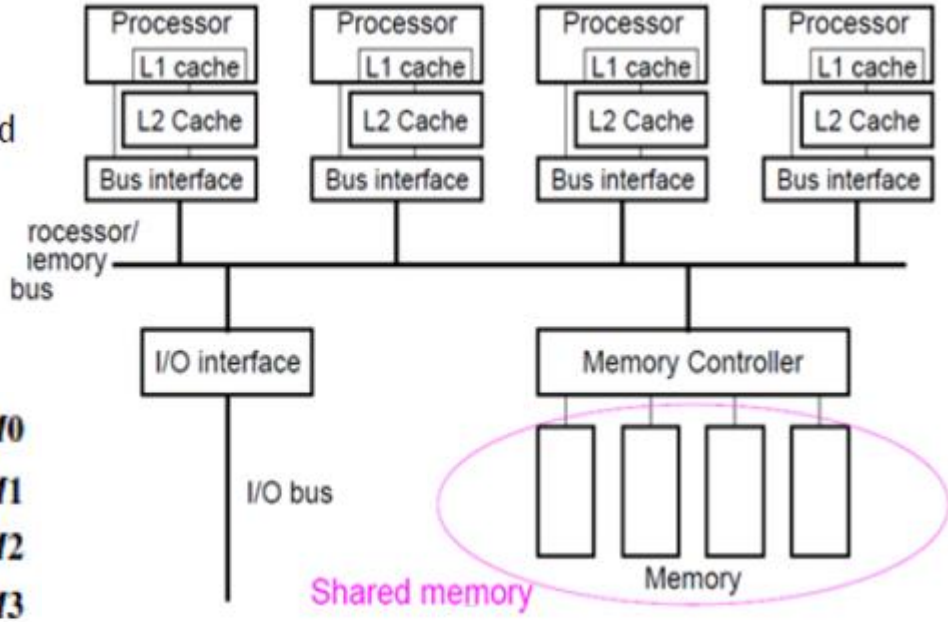


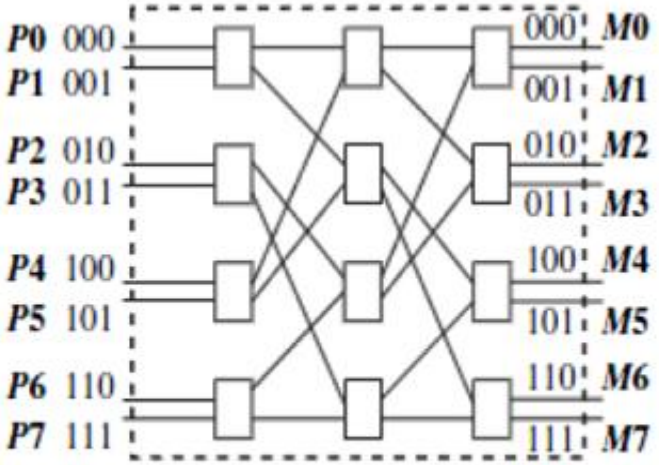
(a) Uniform memory access(UMA)

(b) Non-Uniform memory access (NUMA)

Concepts and Architecture Aspects:
Shared Memory Systems

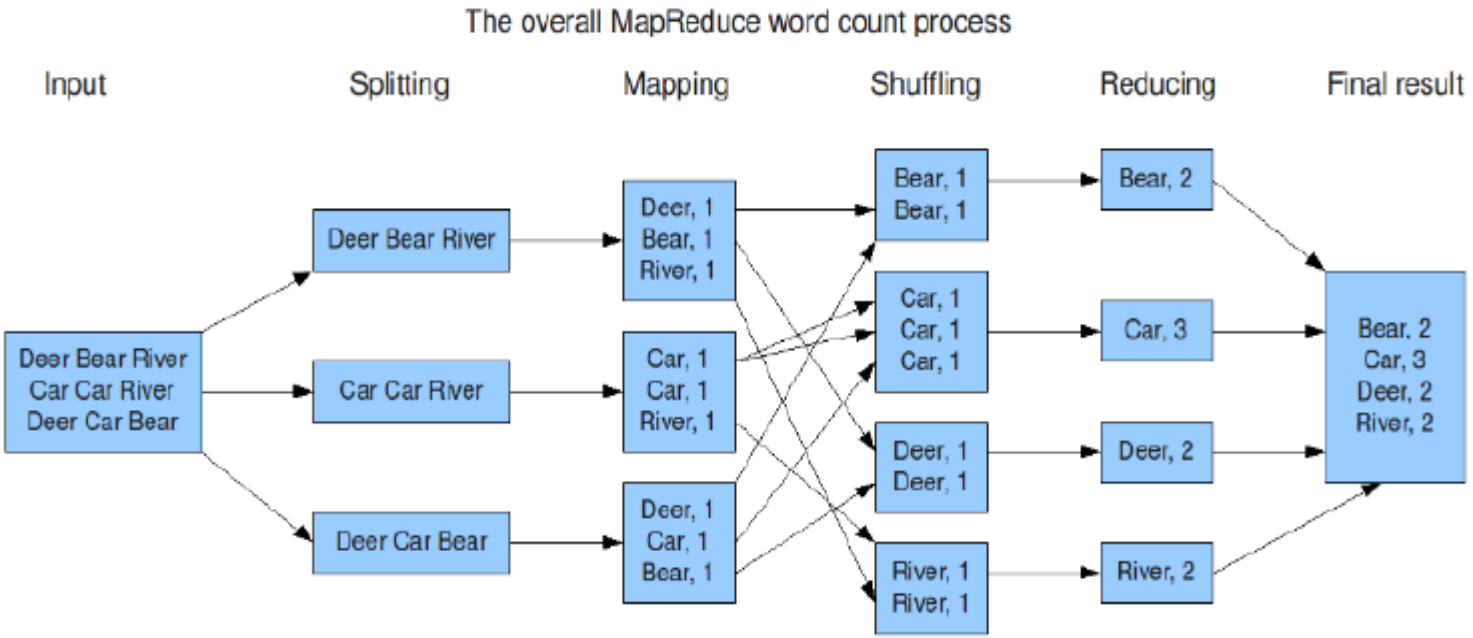
Example:
Multiprocessor Quad Pentium Shared Memory





Uniform memory access (UMA) – Omega Network

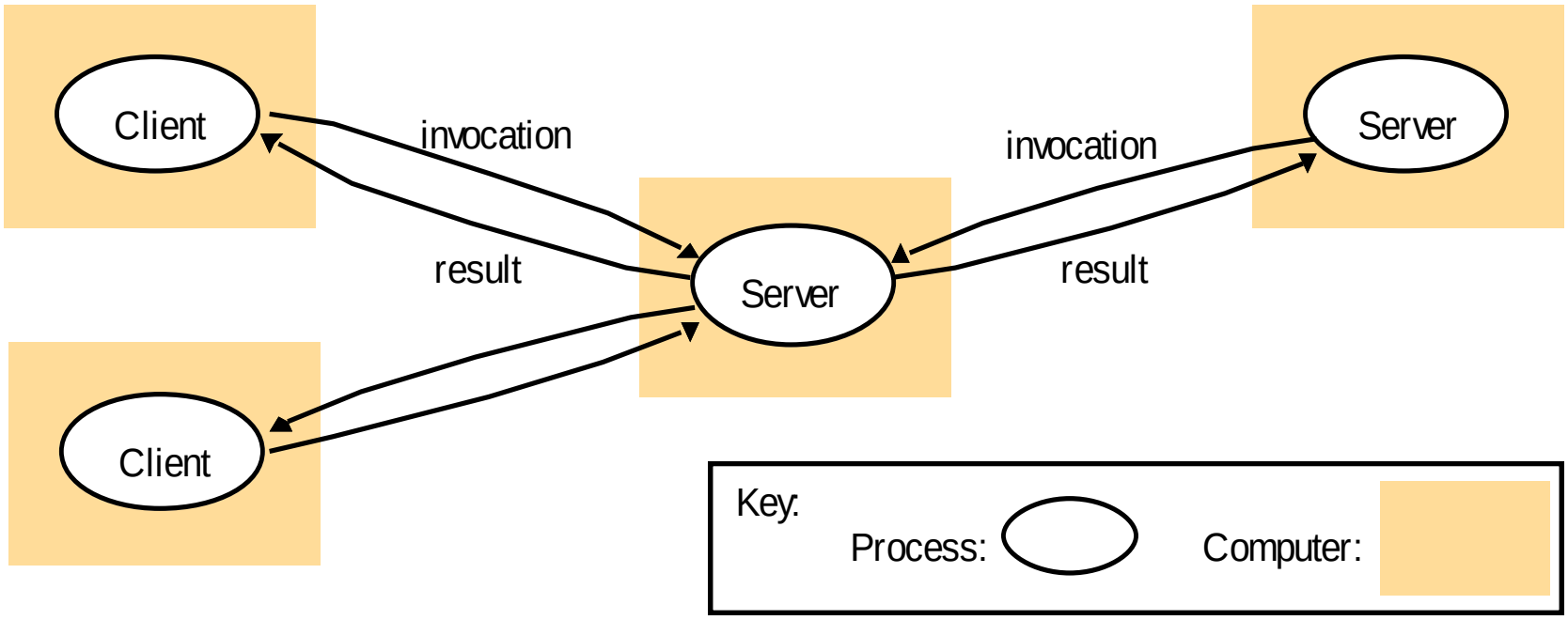
Concepts and Architecture Aspects: Distributed Systems



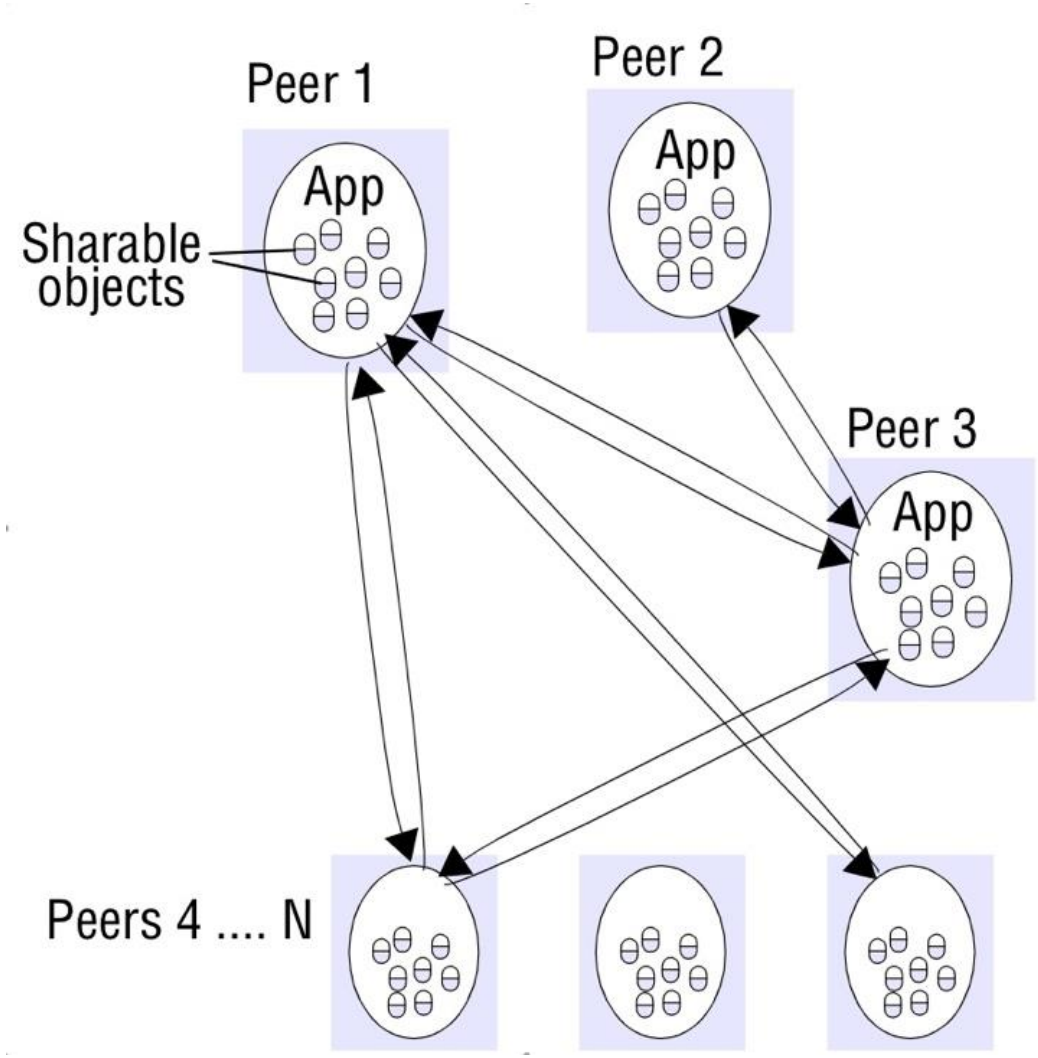
Concepts and Architecture Aspects: Processing Modes

- **Single instruction stream, single data stream (SISD):** traditional uniprocessor
- **Single instruction stream, multiple data stream (SIMD):** Vector processing – Supercomputers
- **Multiple instruction stream, single data stream (MISD):** execute different operations in parallel on the same data – not common
- **Multiple instruction stream, Multiple data stream (MIMD):** processors execute different instructions on different data – many distributed systems fall in this category

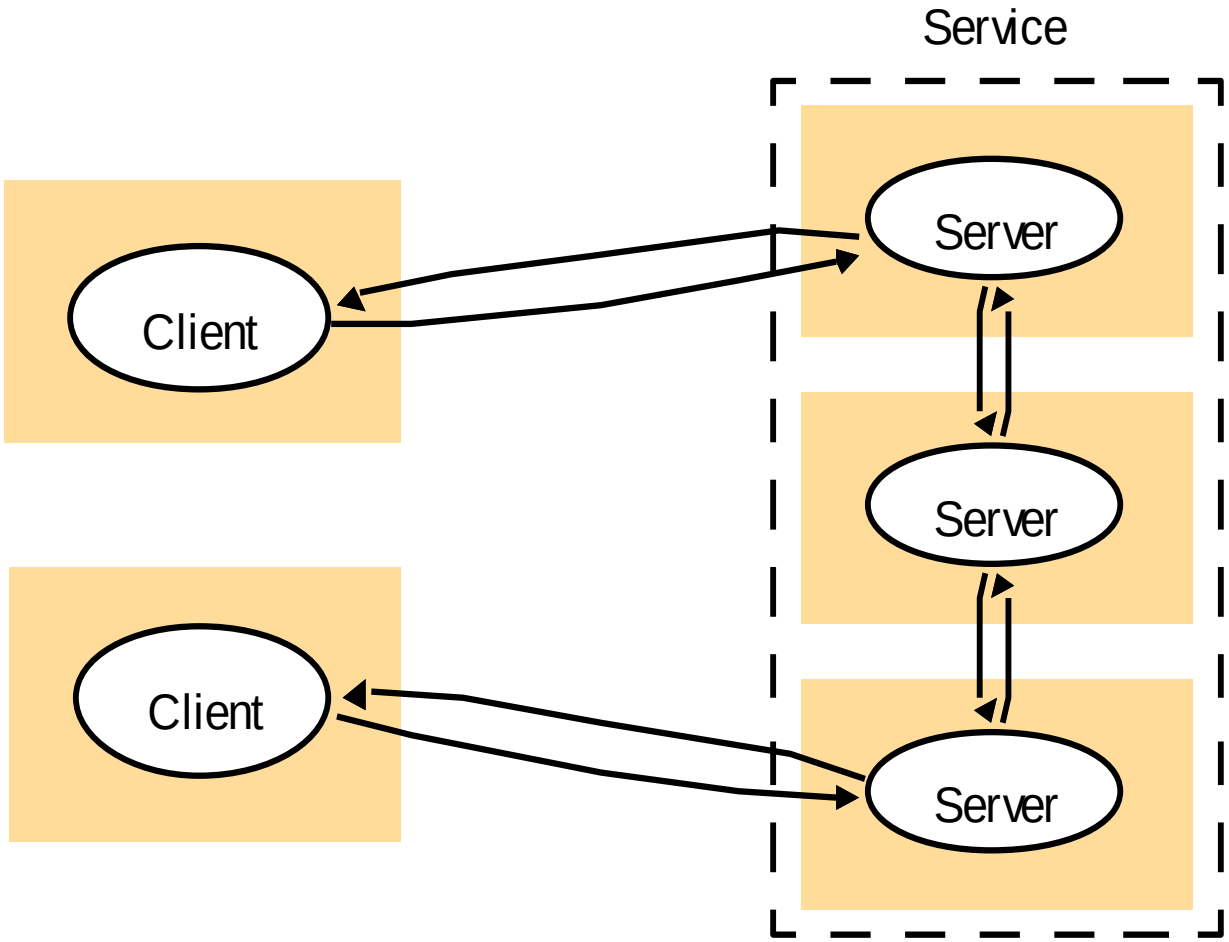
Clients Invoke Individual Servers



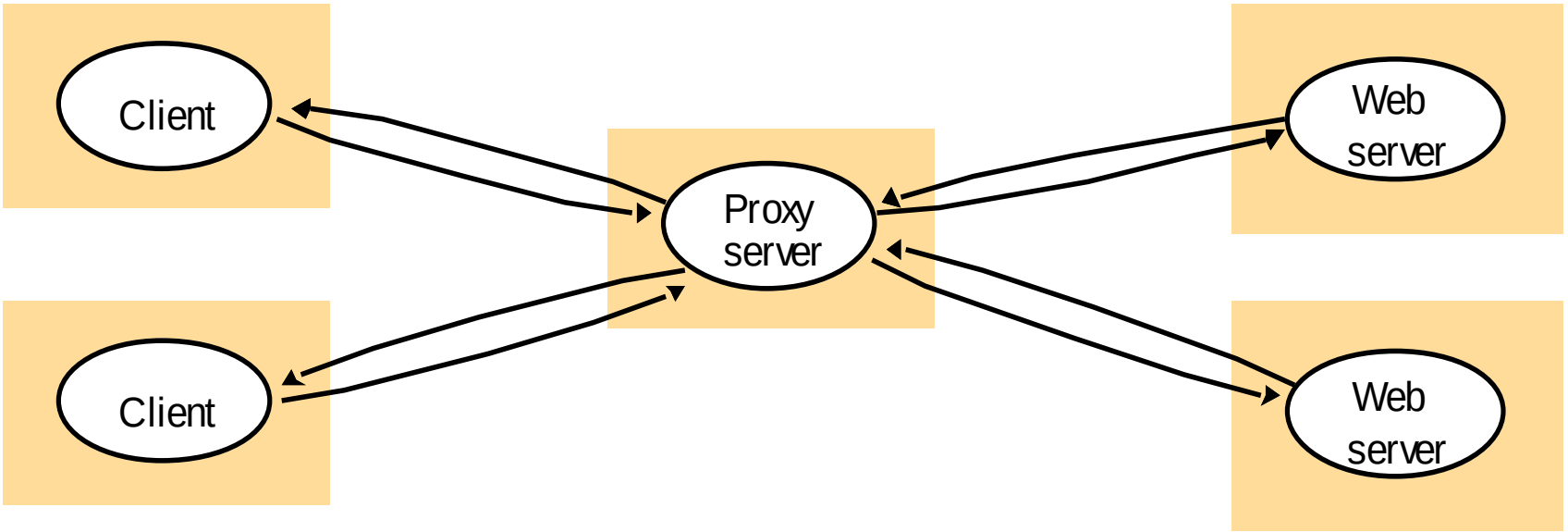
Peer-to-Peer Architecture



A Service Provided by Multiple Servers

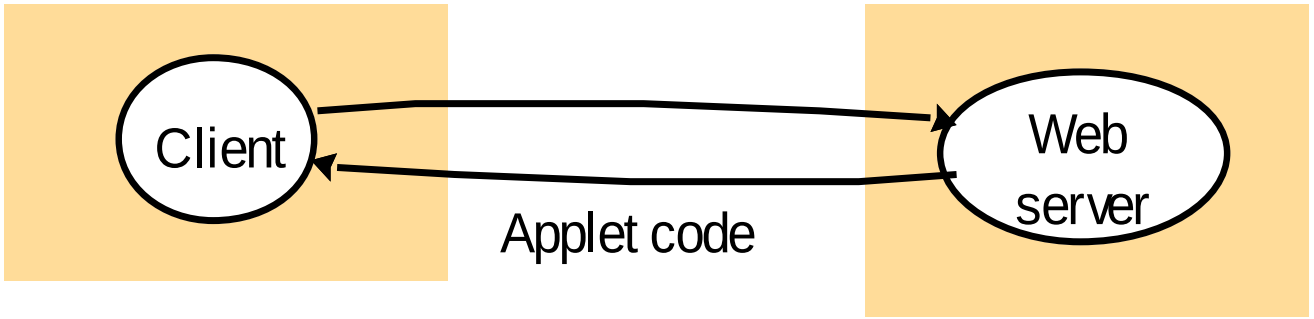


Web Proxy Server



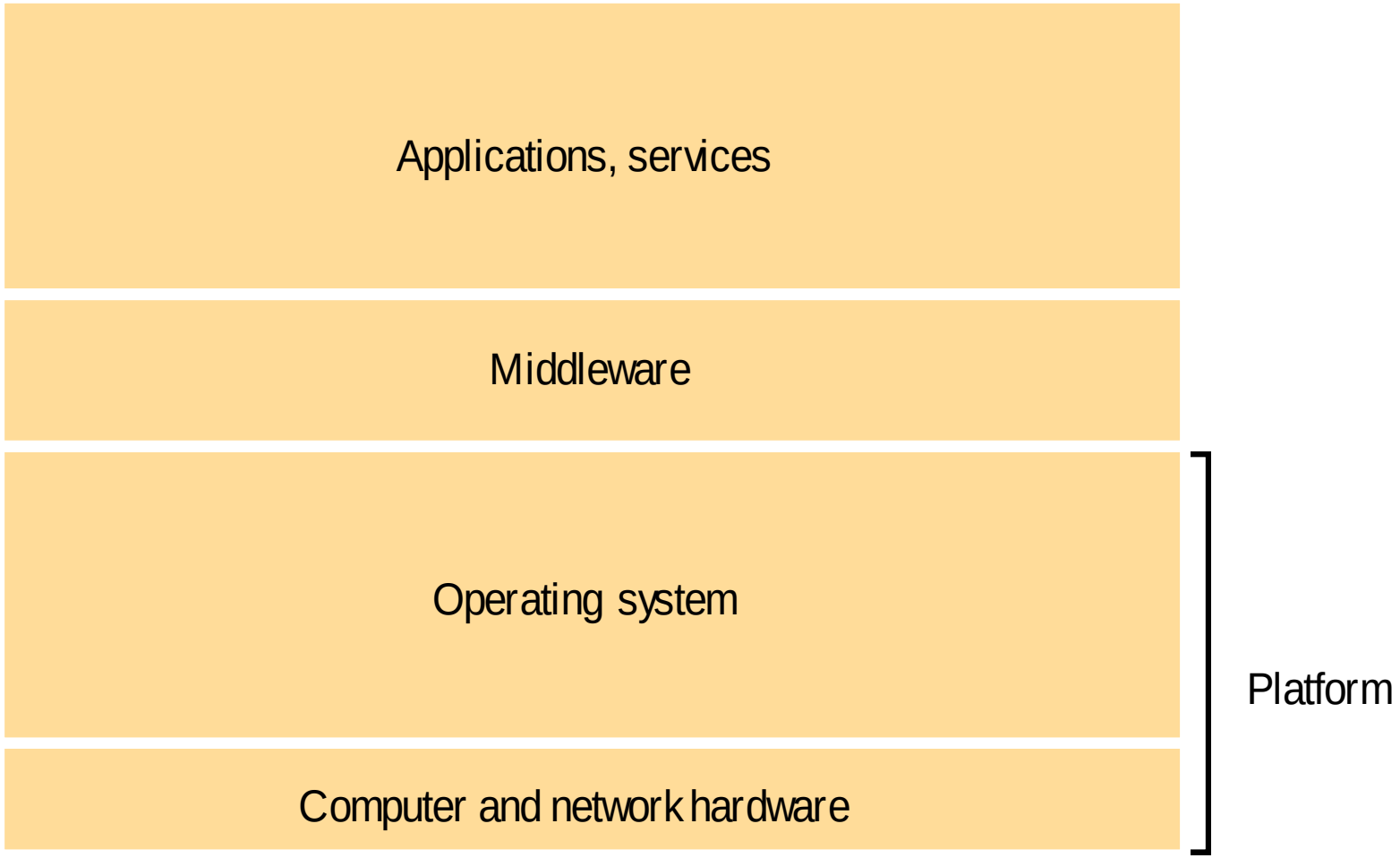
Web Applets

a) client request results in the downloading of applet code

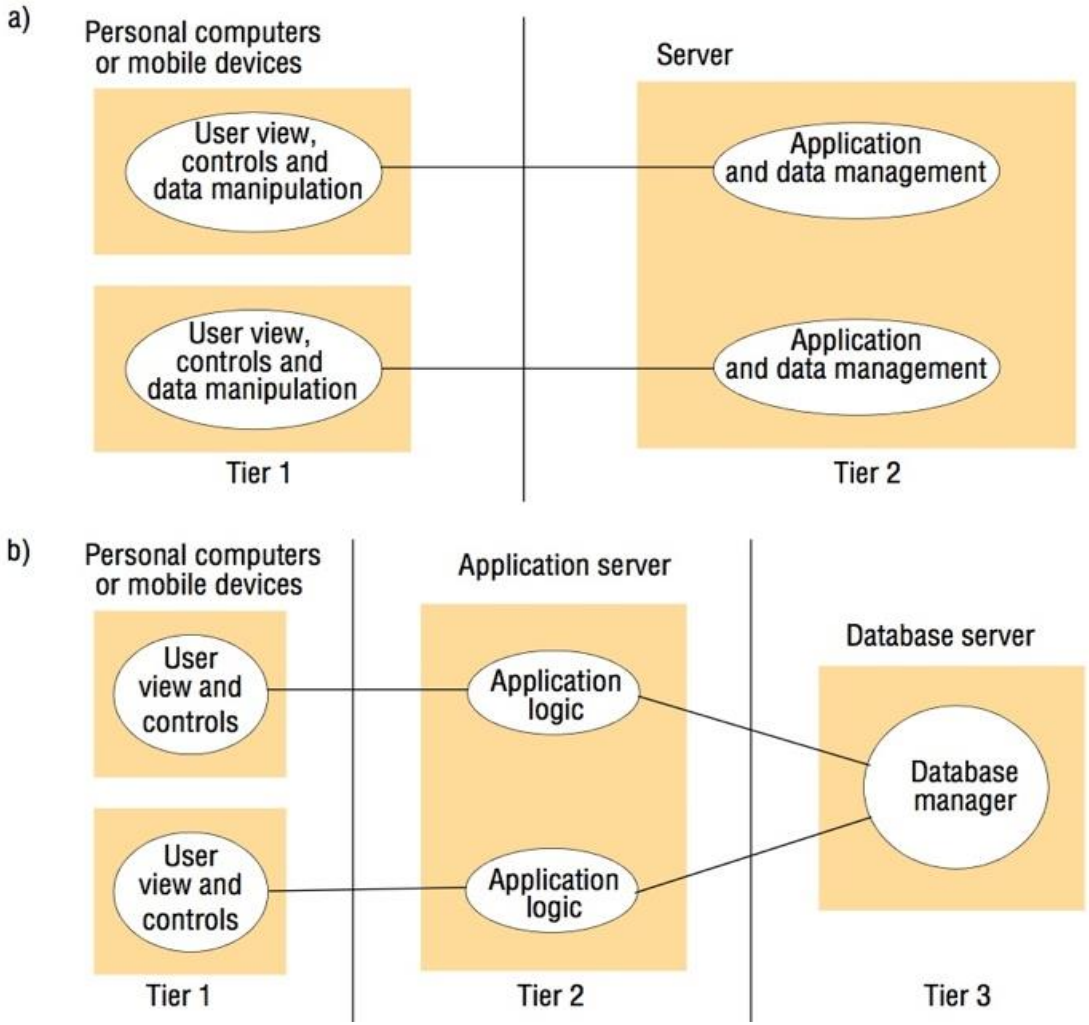


b) client interacts with the applet

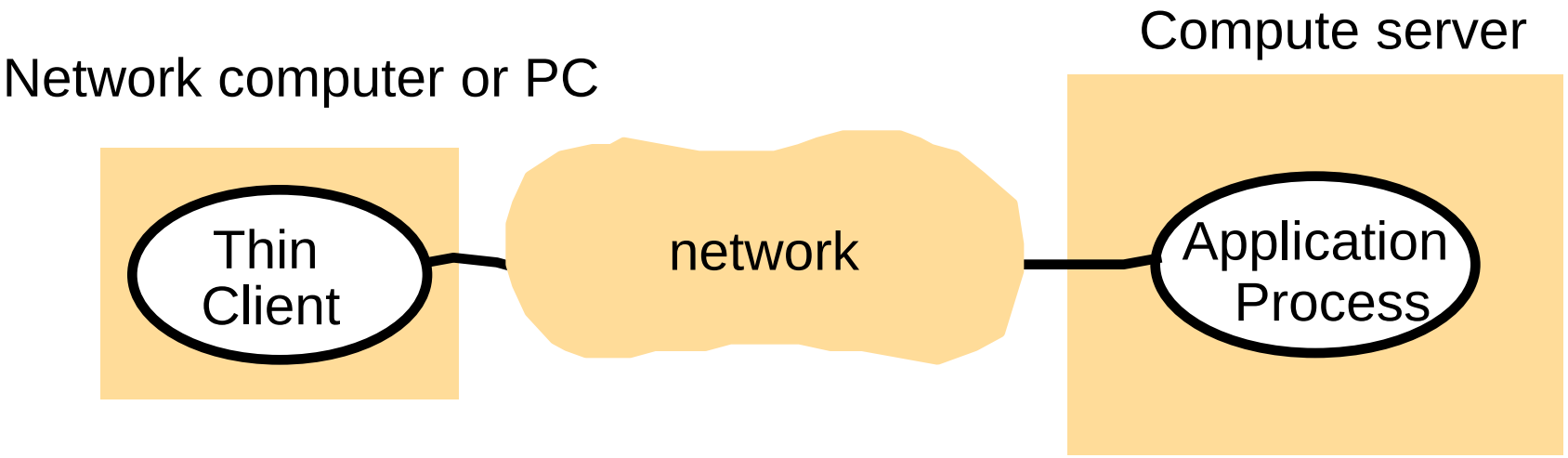
Software and Hardware Service Layers in Distributed Systems



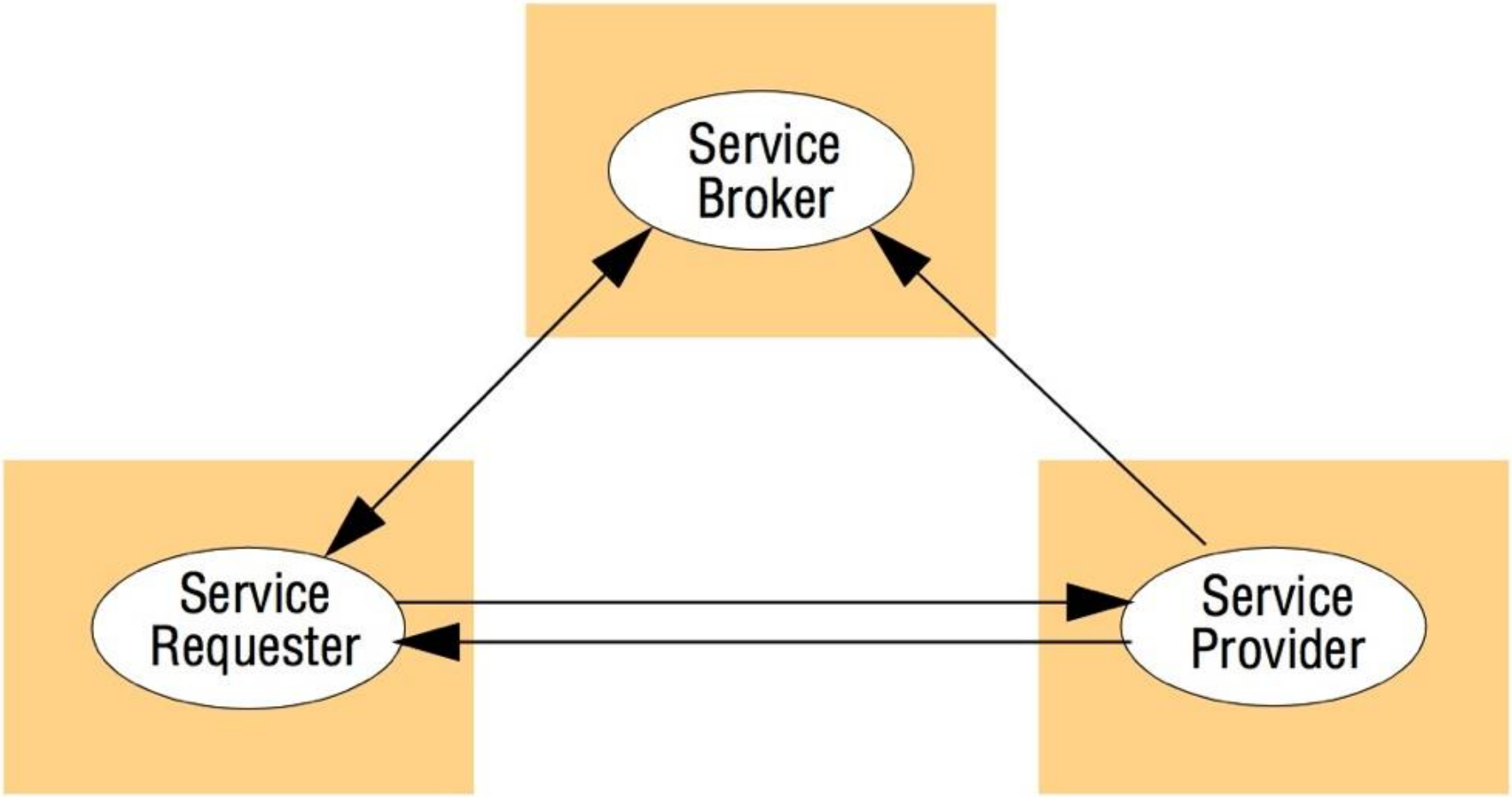
Two-tier and Three-tier Architecture



Thin Clients and Compute Servers



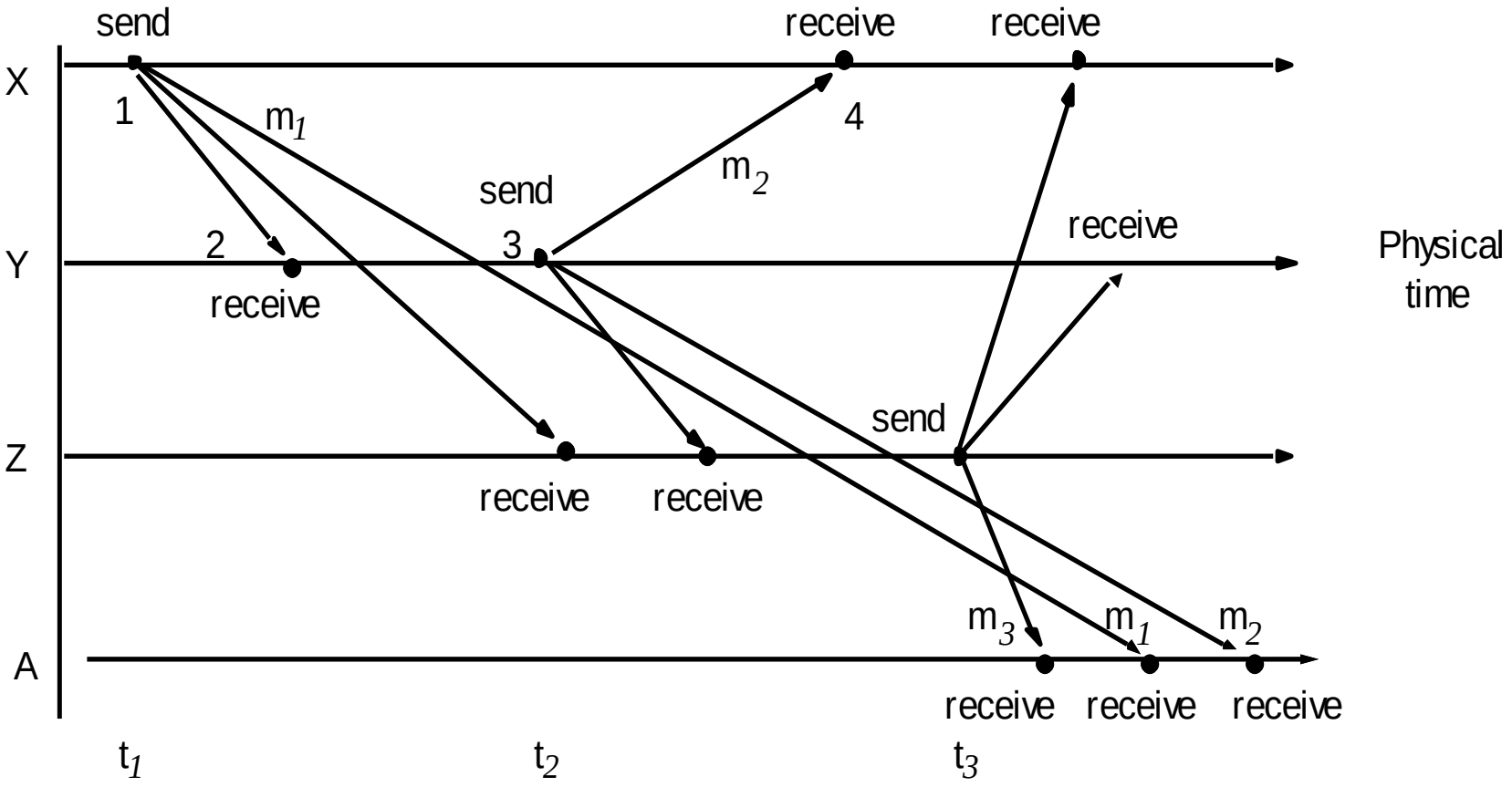
The Web Service Architectural Pattern



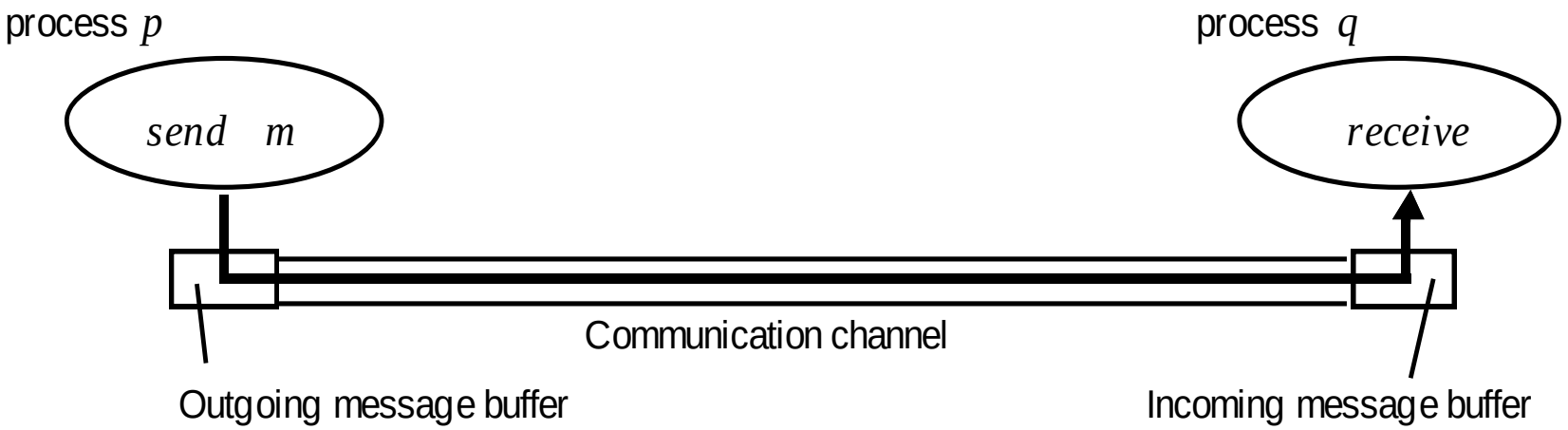
Categories of Middleware

Major categories:	*Subcategory*	*Example systems*
Distributed objects (Chapters 5, 8)	Standard	RM-ODP
	Platform	CORBA
	Platform	Java RMI
Distributed components (Chapter 8)	Lightweight components	Fractal
	Lightweight components	OpenCOM
	Application servers	SUN EJB
	Application servers	CORBA Component Model
	Application servers	JBoss
Publish-subscribe systems (Chapter 6)	-	CORBA Event Service
	-	Scribe
	-	JMS
Message queues (Chapter 6)	-	Websphere MQ
	-	JMS
Web services (Chapter 9)	Web services	Apache Axis
	Grid services	The Globus Toolkit
Peer-to-peer (Chapter 10)	Routing overlays	Pastry
	Routing overlays	Tapestry
	Application-specific	Squirrel
	Application-specific	OceanStore
	Application-specific	Ivy
	Application-specific	Gnutella

Real-time Ordering of Events



Processes and Channels



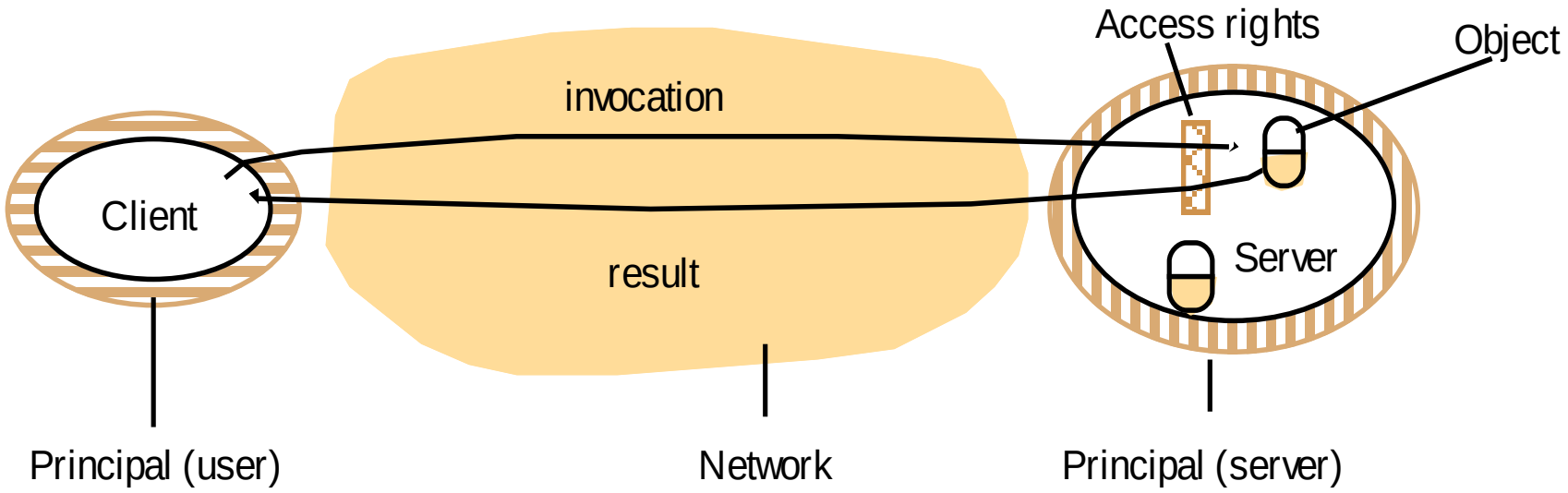
Omission and Arbitrary Failure

Class of failure	*Affects*	*Description*
Fail-stop	Process	Process halts and remains halted. Other processes may detect this state.
Crash	Process	Process halts and remains halted. Other processes may not be able to detect this state.
Omission	Channel	A message inserted in an outgoing message buffer never arrives at the other end's incoming message buffer.
Send-omission	Process	A process completes a *send,* but the message is not put in its outgoing message buffer.
Receive-omission	Process	A message is put in a process's incoming message buffer, but that process does not receive it.
Arbitrary (Byzantine)	Process or channel	Process/channel exhibits arbitrary behavior: it may send/transmit arbitrary messages at arbitrary times, commit omissions; a process may stop or take an incorrect step.

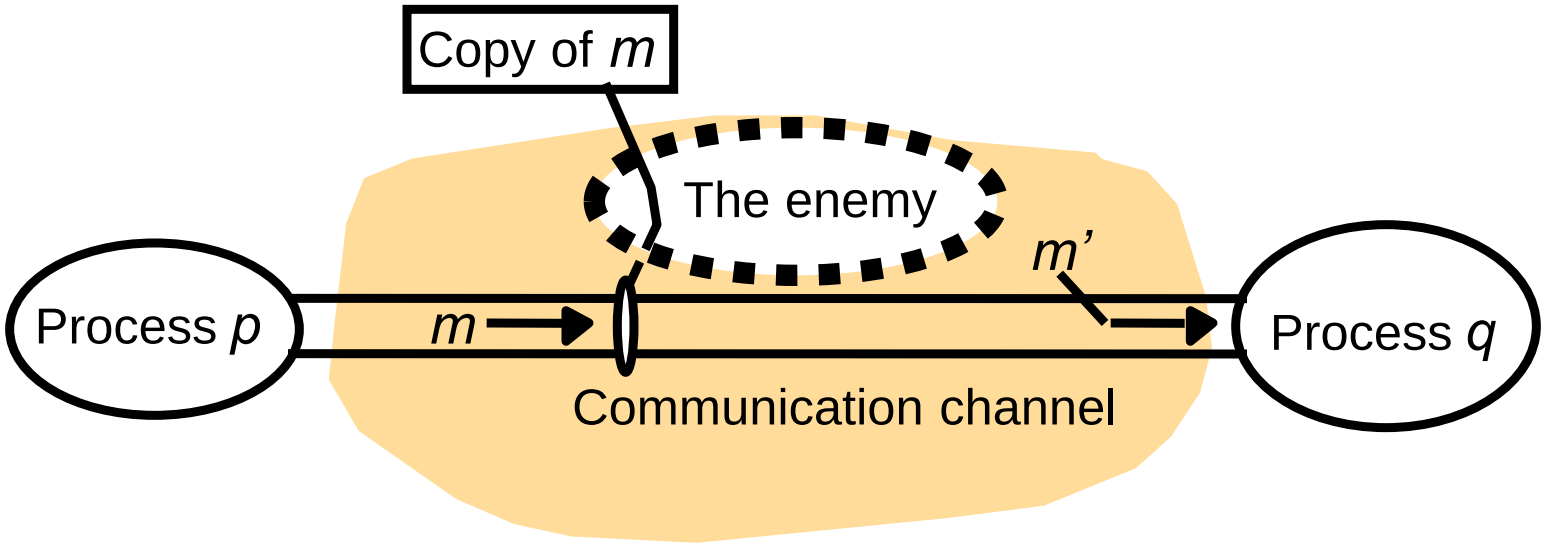
Timing Failures

Class of Failure	*Affects*	*Description*
Clock	Process	Process's local clock exceeds the bounds on its rate of drift from real time.
Performance	Process	Process exceeds the bounds on the interval between two steps.
Performance	Channel	A message's transmission takes longer than the stated bound.

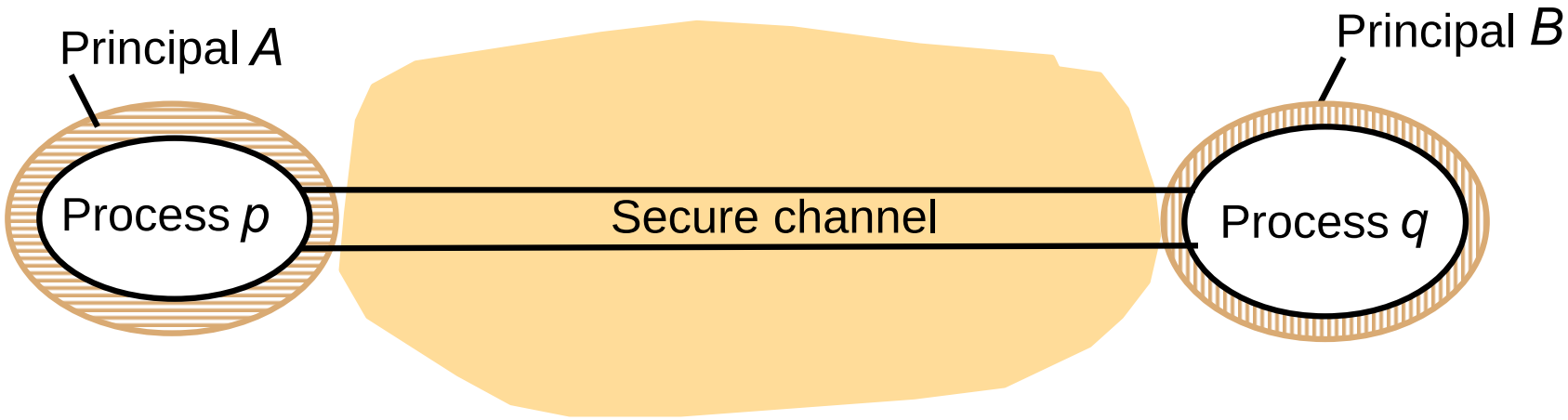
Objects and Principals



The Enemy


Copy of m
The enemy
m'
Process p
m
Process q
Communication channel

Secure Channels



Networking and Inter-networking

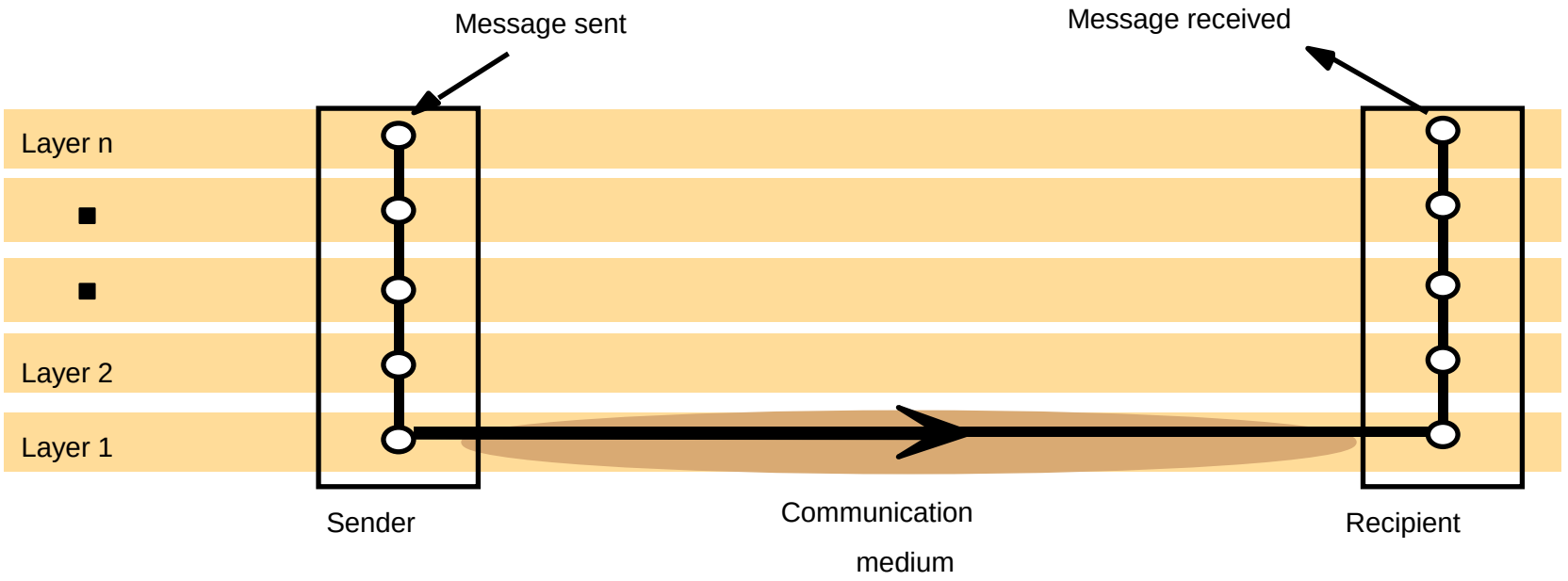
Network Performance

Network types include: personal area network, LAN, WAN, Metropolitan area network, and wireless. Internetwork such as the Internet is constructed from networks of all these types.

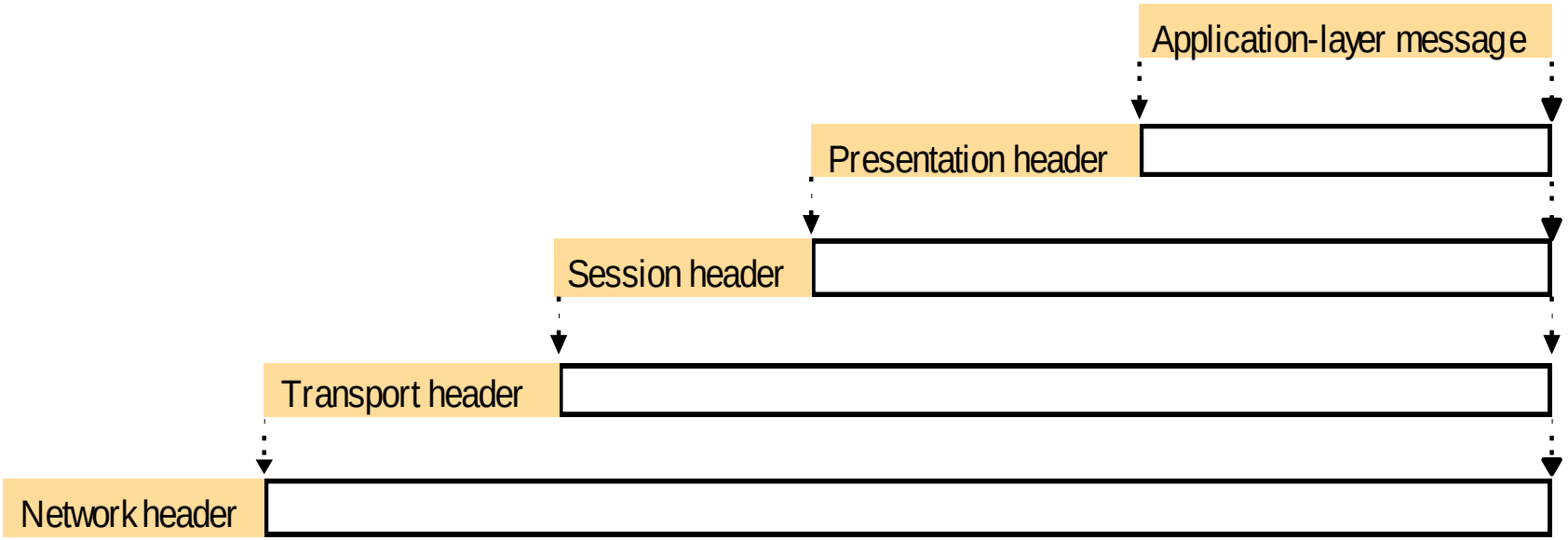
	Example	*Range*	*Bandwidth (Mbps)*	*Latency (ms)*
Wired:				
LAN	Ethernet	1–2 kms	10–10,000	1–10
WAN	IP routing	worldwide	0.010–600	100–500
MAN	ATM	2–50 kms	1–600	10
Internetwork	Internet	worldwide	0.5–600	100–500
Wireless:				
WPAN	Bluetooth (IEEE 802.15.1)	10–30m	0.5–2	5–20
WLAN	WiFi (IEEE 802.11)	0.15–1.5 km	11–108	5–20
WMAN	WiMAX (IEEE 802.16)	5–50 km	1.5–20	5–20
WWAN	3G phone	cell: 1–-5	348–14.4	100–500

km

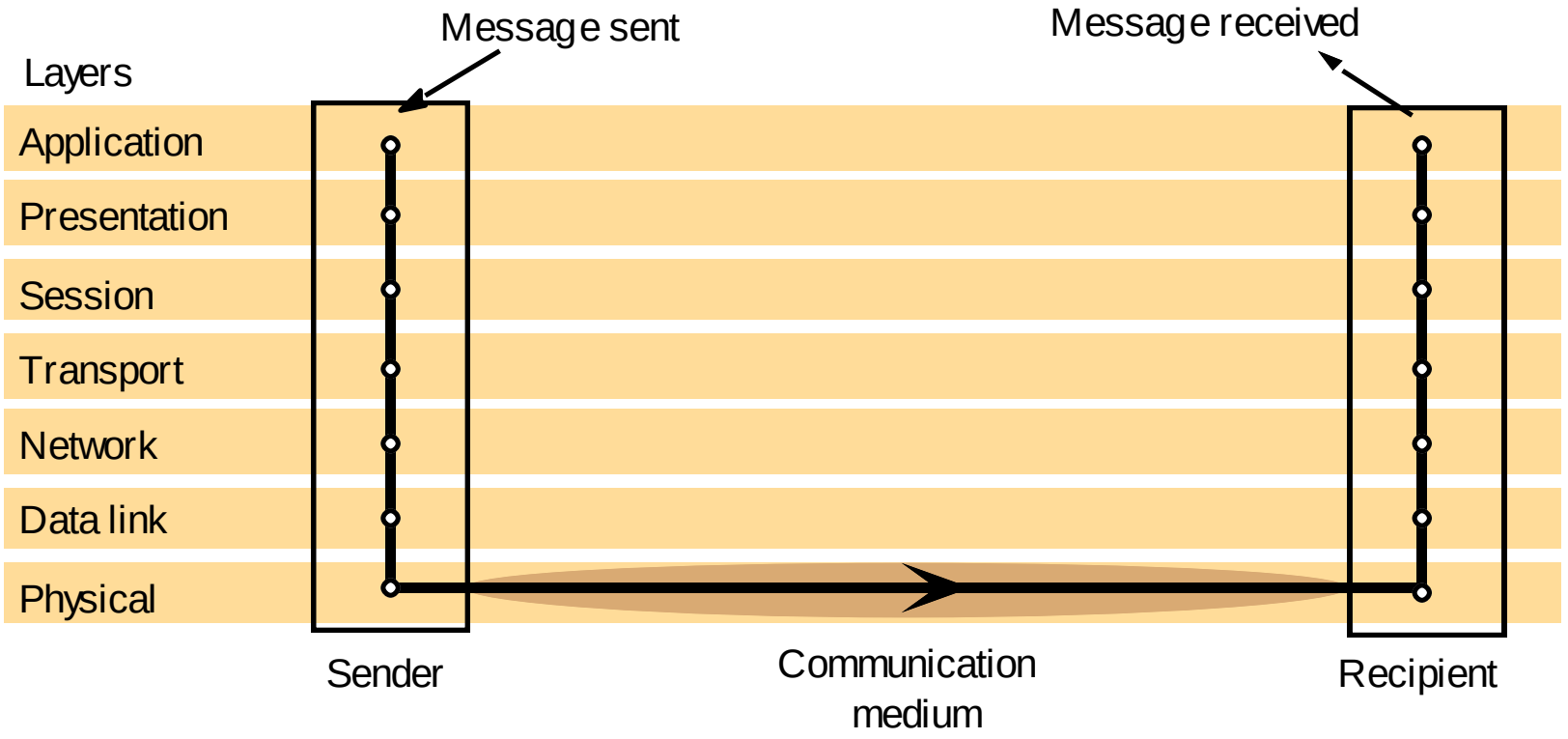
Conceptual Layering of Protocol Software



Encapsulation as it is Applied in Layered Protocols



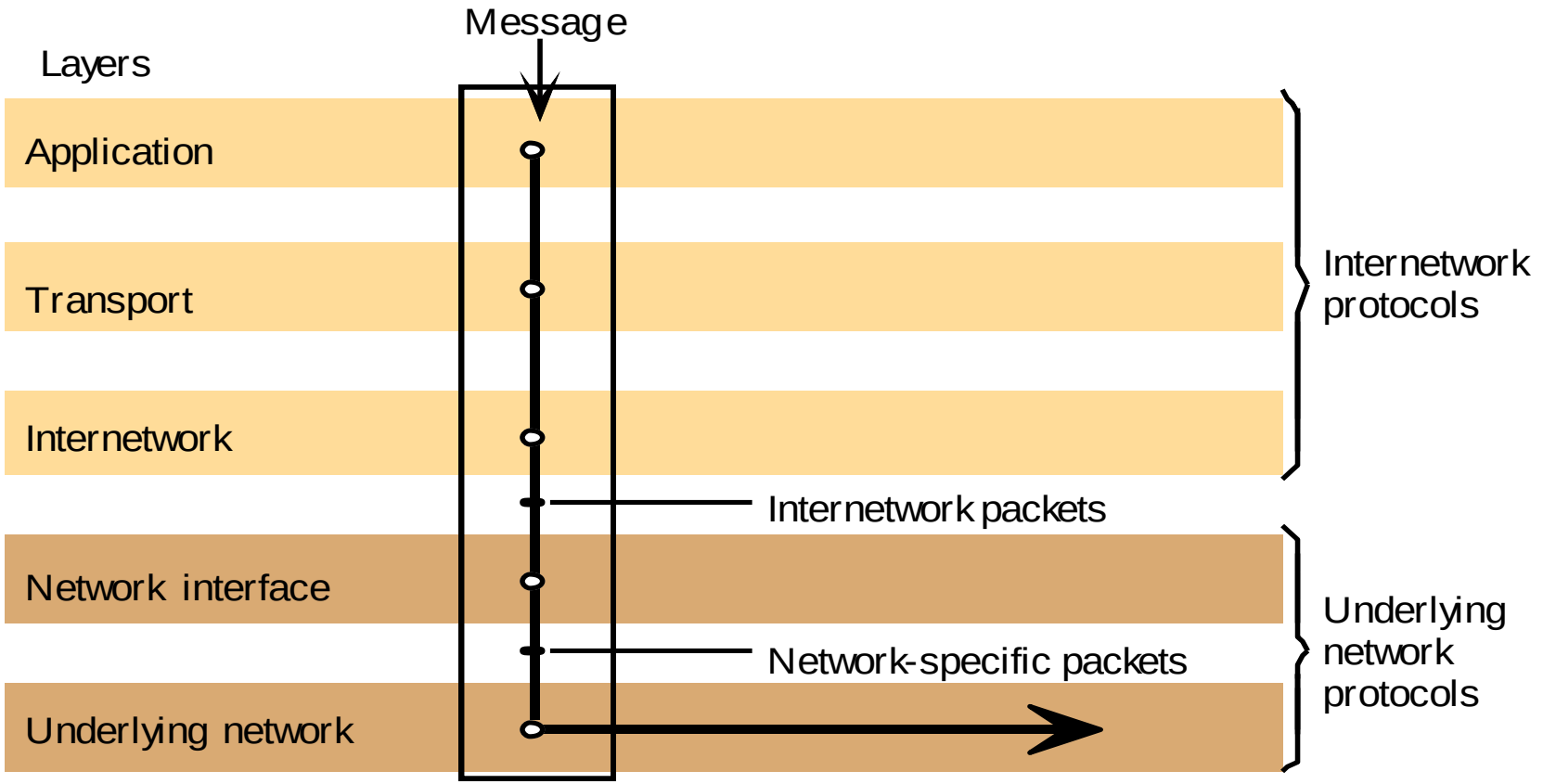
Protocol Layers in the ISO Open Systems Interconnection (OSI) Model



OSI Protocol Summary

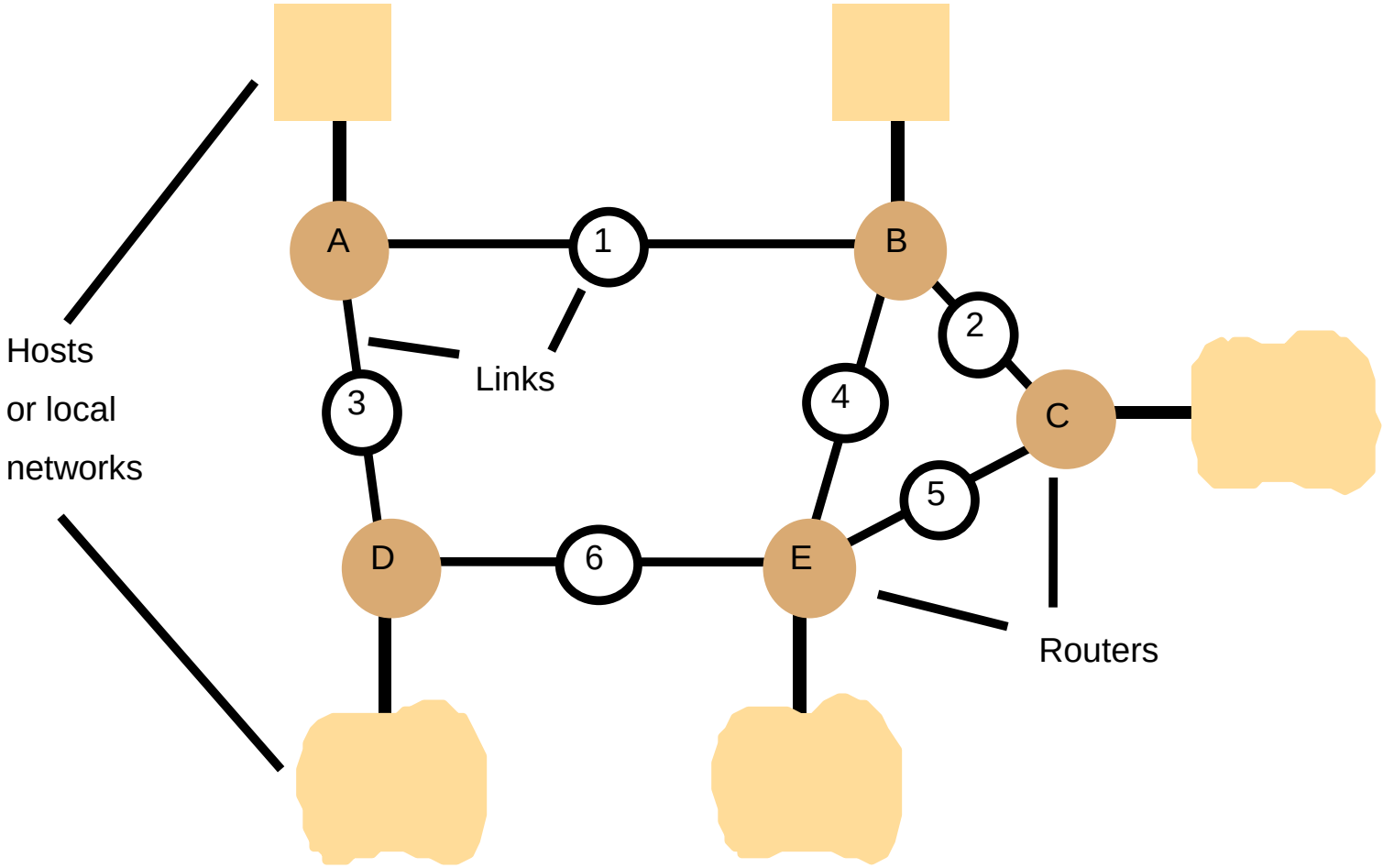
Layer	*Description*	*Examples*
Application	Protocols that are designed to meet the communication requirements of specific applications, often defining the interface to a service.	HTTP, FTP SMTP, CORBA IIOP
Presentation	Protocols at this level transmit data in a network representation that is independent of the representations used in individual computers, which may differ. Encryption is also performed in this layer, if required.	Secure Sockets (SSL),CORBA Data Rep.
Session	At this level reliability and adaptation are performed, such as detection of failures and automatic recovery.	
Transport	This is the lowest level at which messages (rather than packets) are handled. Messages are addressed to communication ports attached to processes, Protocols in this layer may be connection-oriented or connectionless.	TCP, UDP
Network	Transfers data packets between computers in a specific network. In a WAN or an internetwork this involves the generation of a route passing through routers. In a single LAN no routing is required.	IP,ATM virtual circuits
Data link	Responsible for transmission of packets between nodes that are directly connected by a physical link. In a WAN transmission is between pairs of routers or between routers and hosts. In a LAN it is between any pair of hosts.	Ethernet MAC, ATM cell transfer, PPP
Physical	The circuits and hardware that drive the network. It transmits sequences of binary data by analogue signalling, using amplitude or frequency modulation of electrical signals (on cable circuits), light signals (on fibre optic circuits) or other electromagnetic signals (on radio and microwave circuits).	Ethernet base- band signalling, ISDN

Inter-network Layers



Transport is the lowest level at which messages (rather than packets) are handled. Network interface layer accepts internetwork packets and convert them into suitable packets to the specific underlying network.

Routing in a Wide-Area Network



Routing Tables for the Last Network

Routings from A

To	*Link*	*Cost*
A	local	0
B	1	1
C	1	2
D	3	1
E	1	2

Routings from B

To	*Link*	*Cost*
A	1	1
B	local	0
C	2	1
D	1	2
E	4	1

Routings from C

To	*Link*	*Cost*
A	2	2
B	2	1
C	local	0
D	5	2
E	5	1

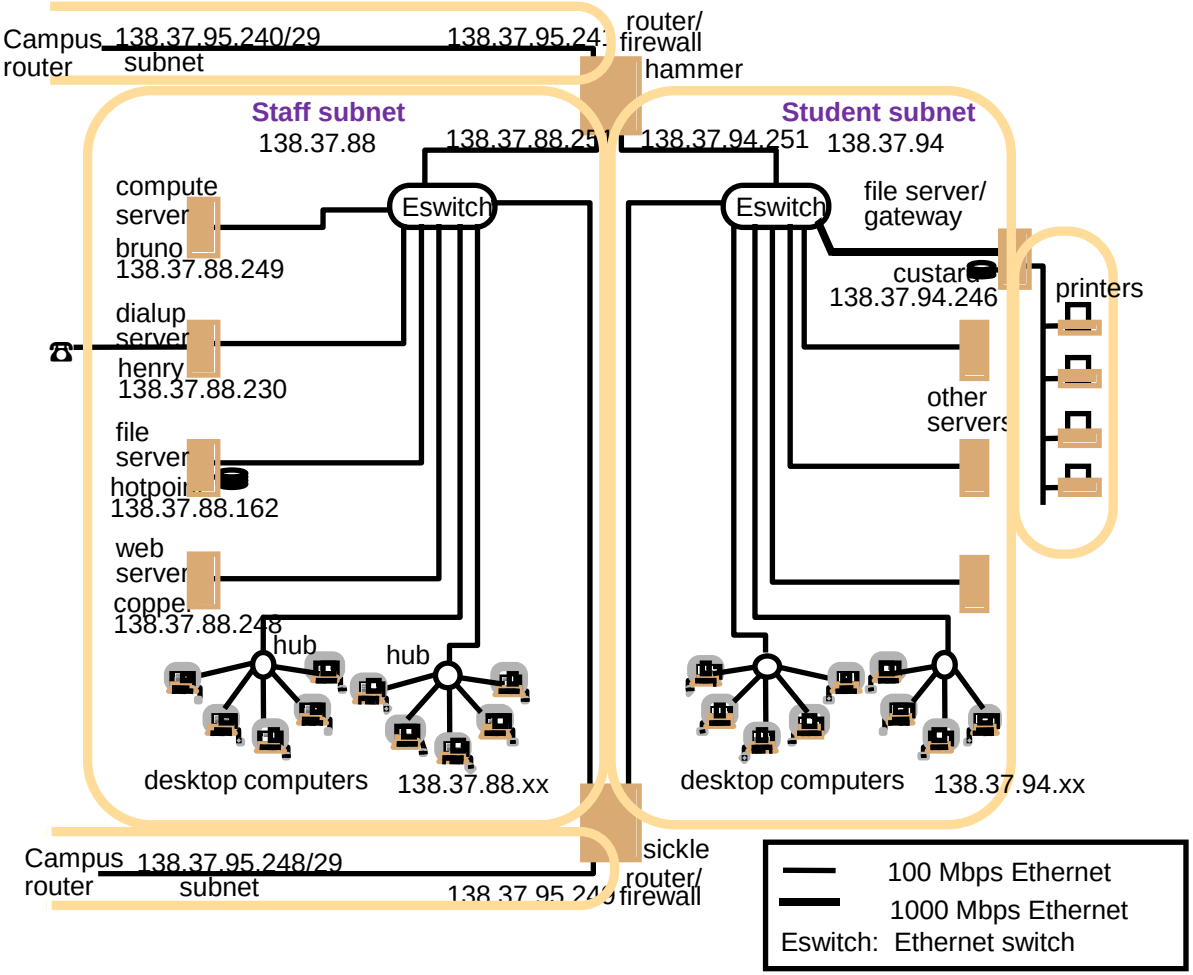
Routings from D

To	*Link*	*Cost*
A	3	1
B	3	2
C	6	2
D	local	0
E	6	1

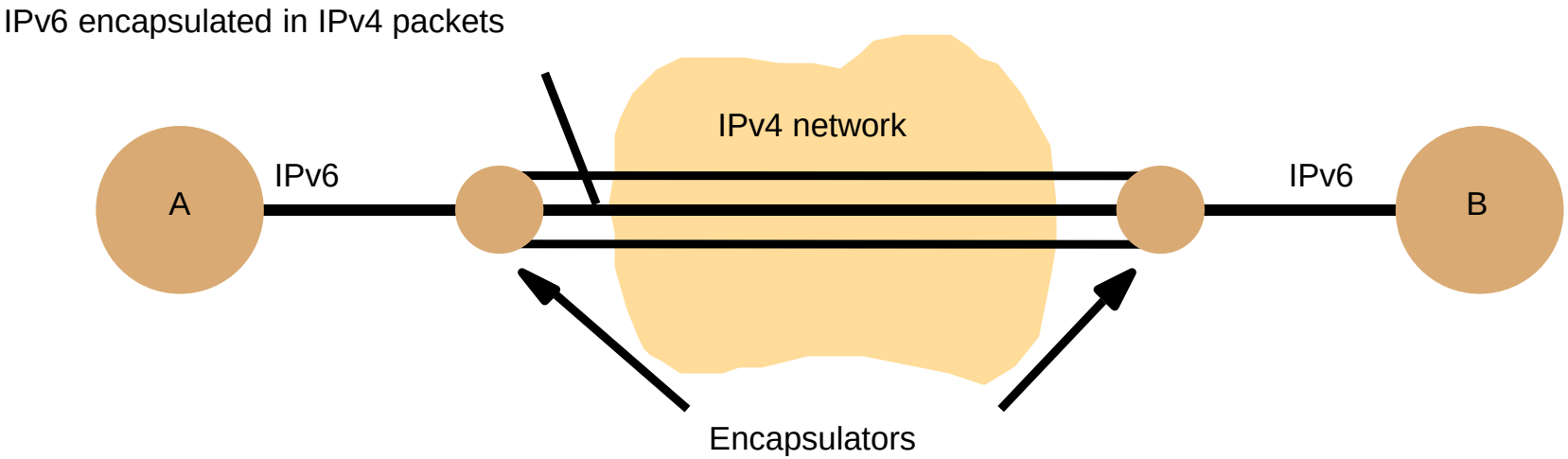
Routings from E

To	*Link*	*Cost*
A	4	2
B	4	1
C	5	1
D	6	1
E	local	0

Simplified View of Part of a University Campus Network

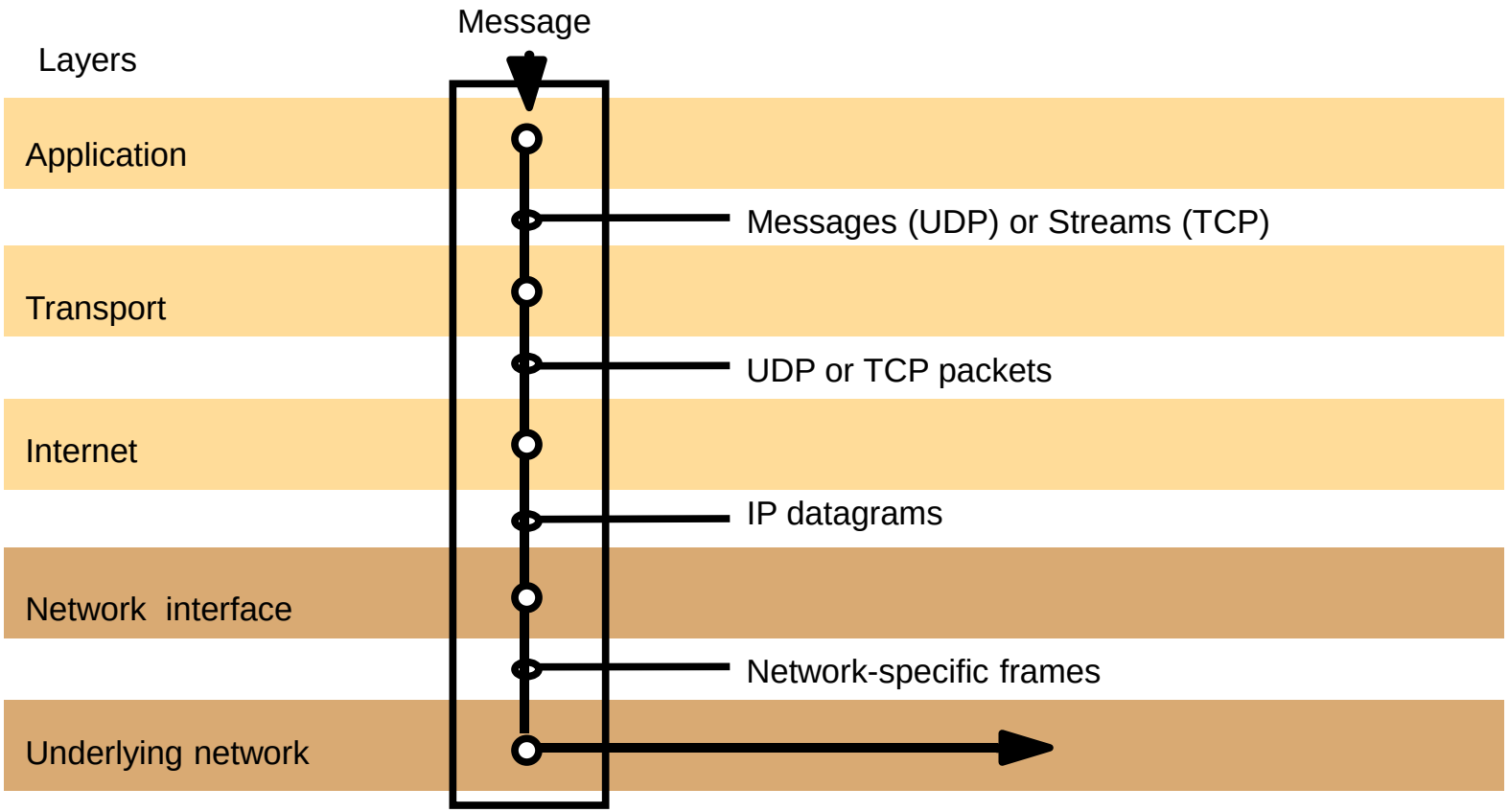


Tunnelling for IPv6 Migration

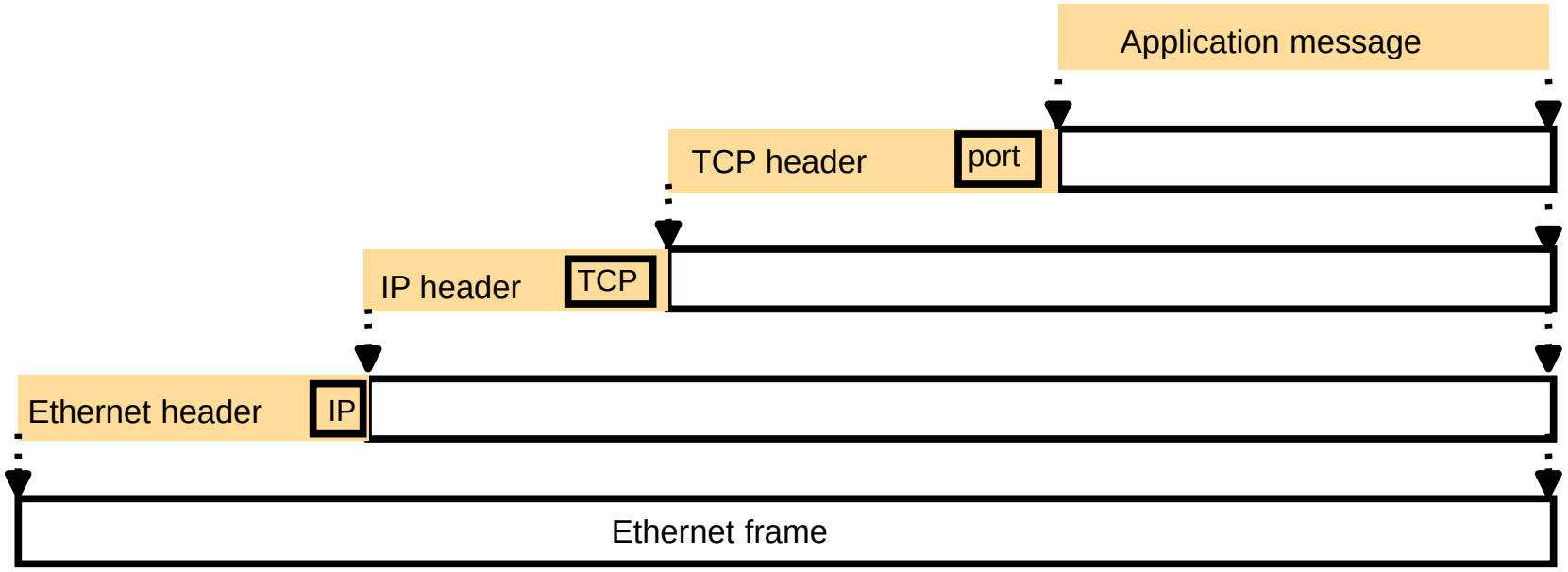


Tunneling IPv6 packets as the body of an IPv4 packets

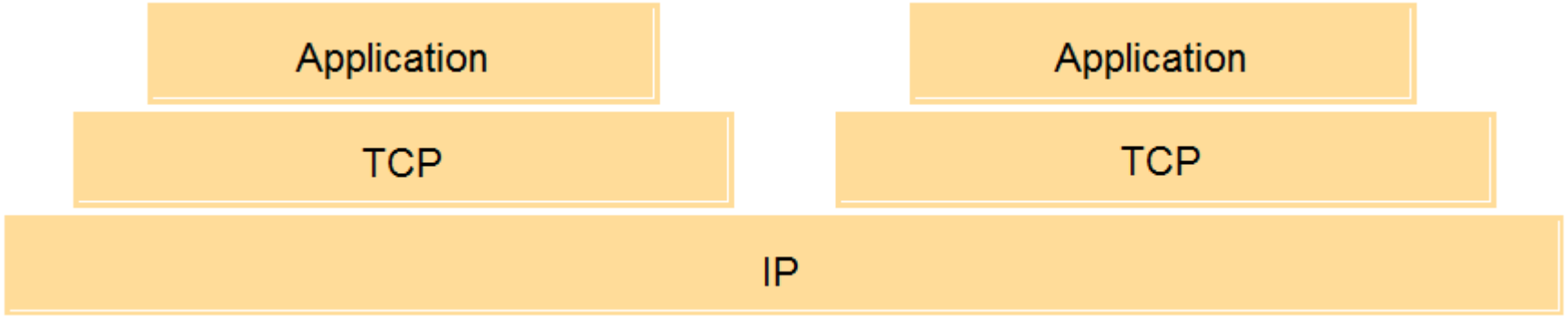
TCP/IP Layers



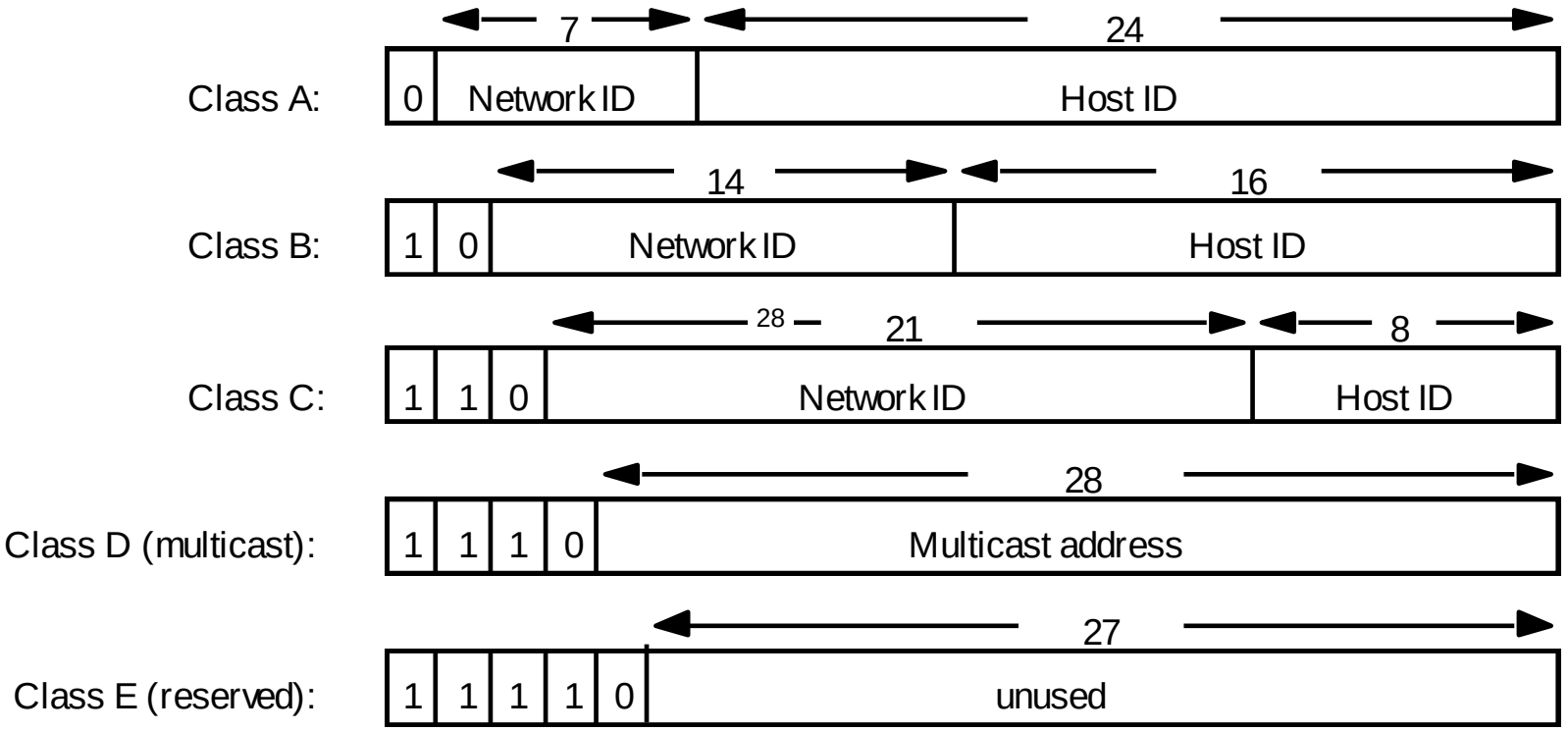
Encapsulation in a Message Transmitted via TCP over Ethernet



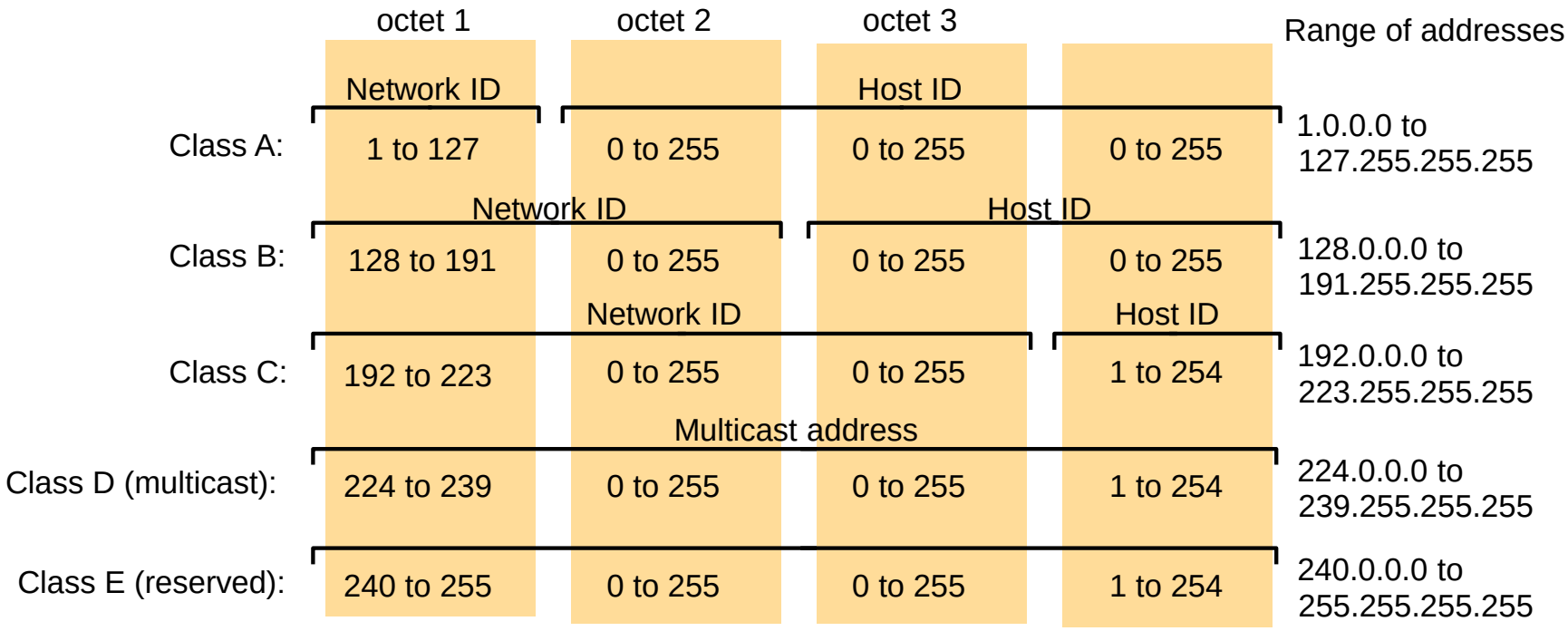
The Programmer's Conceptual View of a TCP/IP Internet



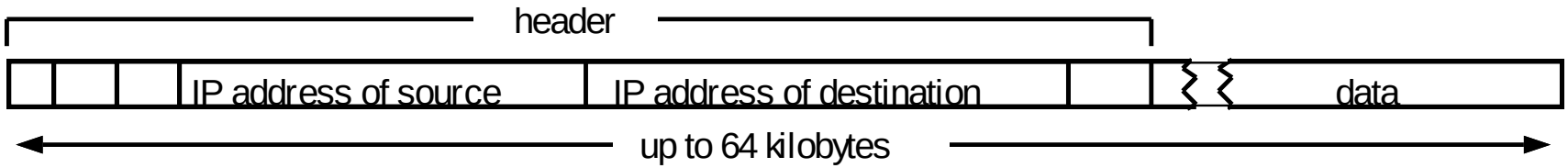
Internet Address Structure (5 classes), Showing Field Sizes in Bits



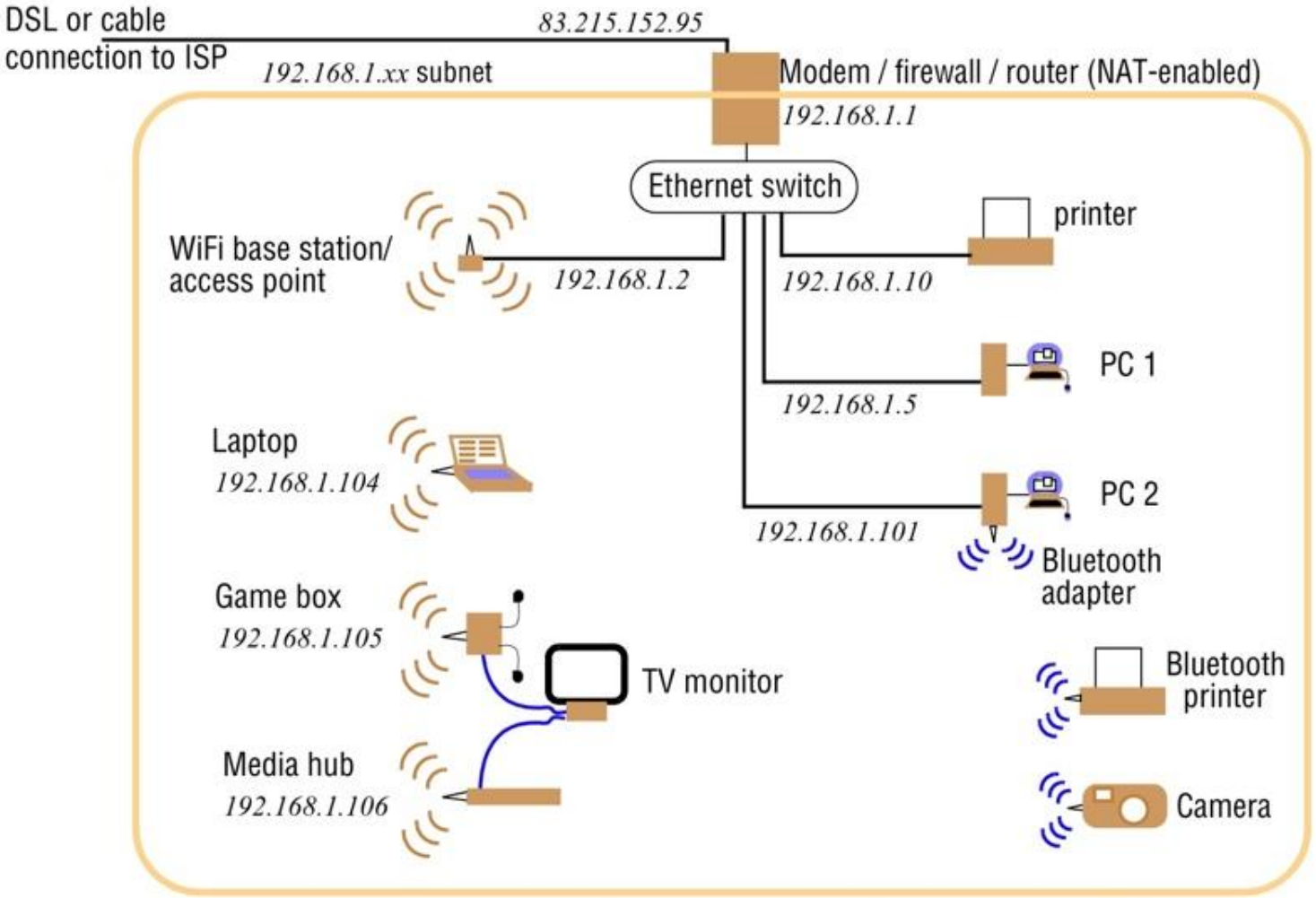
Permissible Values for each class of Network Address: x.y.w.z



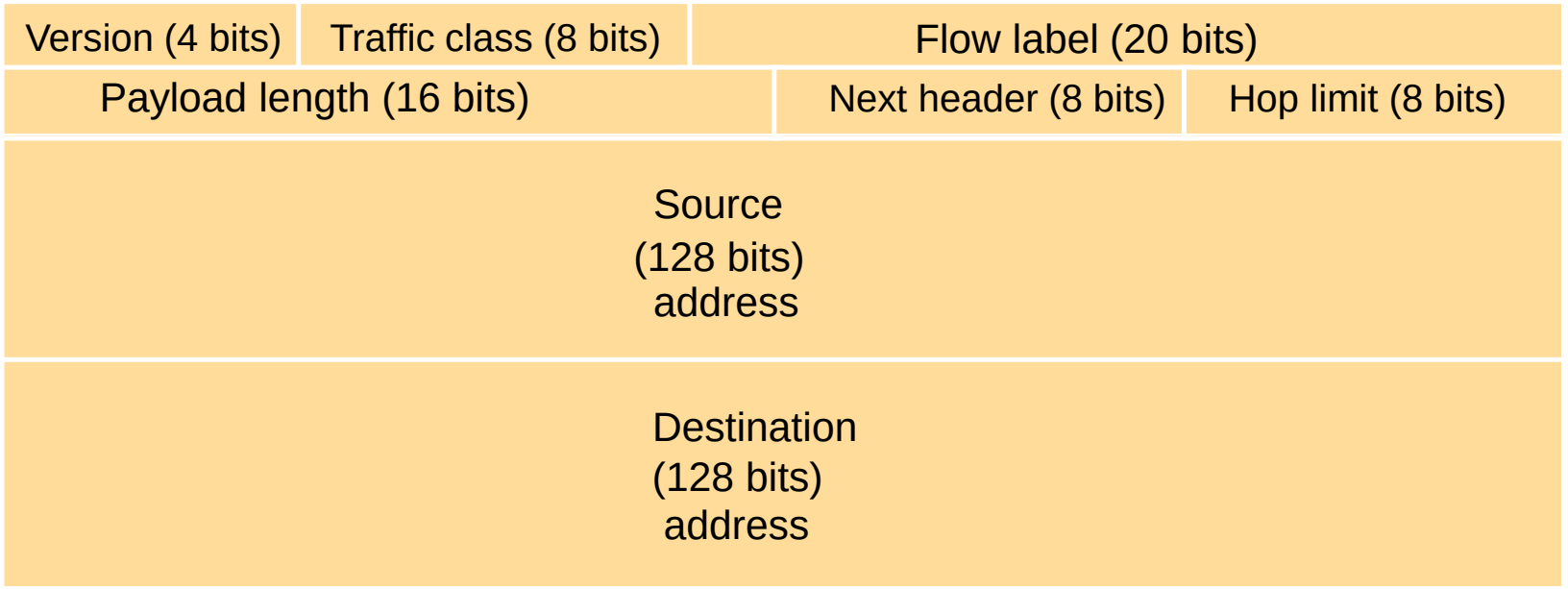
IP Packet Layout



A Typical NAT-based Home Network

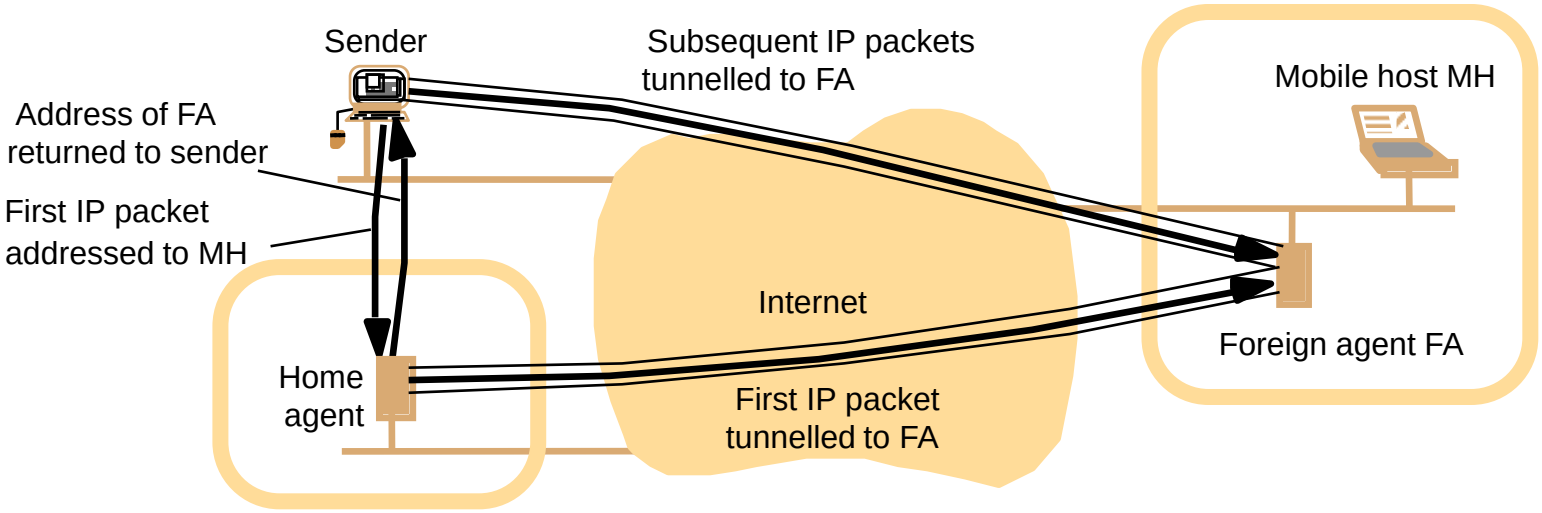


IPv6 Header Layout



IPv4 = 4-bytes = 32-bits
IPv6 = 16-bytes = 128-bits

The MobileIP Routing Mechanism



HA: Home Agent FA: Foreign Agent.

- **Problem statement:** Mobile device (Mobile Host) needs to continue to communicate with the sender independent of where they are; MobileIP is meant to solve this issue!
- Mobile Host has static normal IP address when they are connected to their home base.

The MobileIP Routing Mechanism

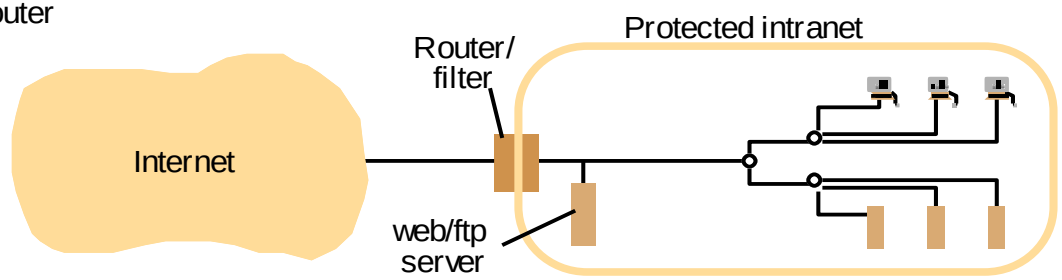
- Mobile Host has static normal IP address. Communication with senders is facilitated through the 2 agents (HA, FA) which are at fixed IP addresses.
- HA is responsible for holding up-to-date knowledge of the Mobile host's current location (IP address by which it can be reached). How? When the mobile host leaves its home site, it should inform the HA . During the interval where the mobile host is relocated, HA will behave as proxy – it tells the local router to cancel any cached info related to the mobile host's IP address. As a proxy, HA responds to ARP requests related to the Mobile Host's IP address by giving its own local network address as the network address of the mobile host.
- When the Mobile host arrives to new site, it inform the FA at that site. FA assigns new temporary IP address on its local subnet to the Mobile Host → FA contacts HA giving it <Host name IP address, the temporary assigned IP address to the Mobile Host>.

The MobileIP Routing Mechanism

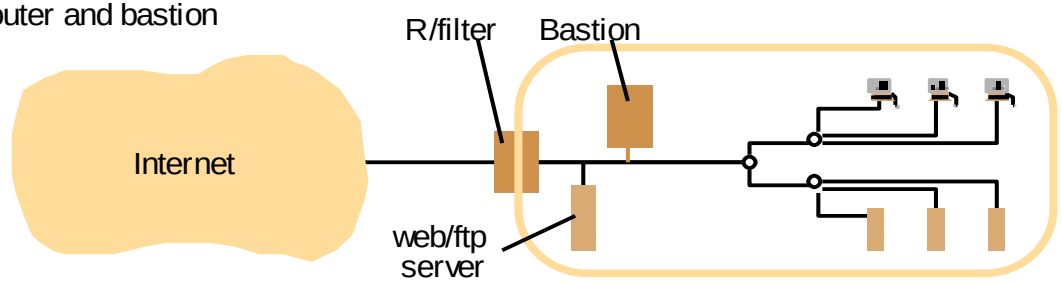
- **Scenario:** when an IP packet from the sender is addressed to the Mobile Host home address, it is re-routed to the HA. HA encapsulates the IP packet in a MobileIP packet and sends it to the FA. FA unpacks the original IP packet and deliver it to the Mobile host temporary IP address within the FA local subnet ← this is tunneling.

Firewall Configuration

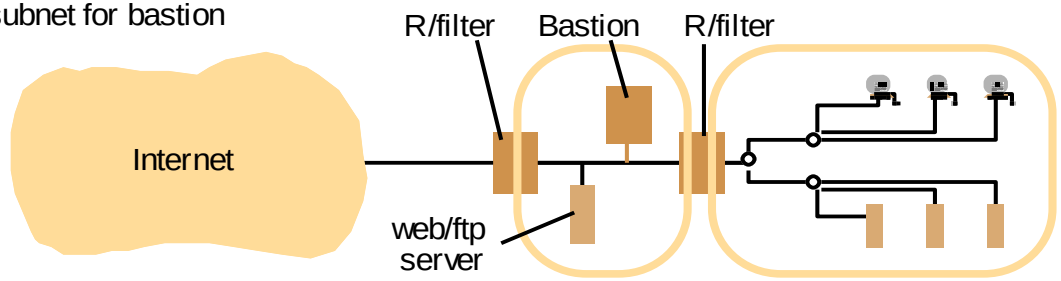
a) Filtering router



b) Filtering router and bastion



c) Screened subnet for bastion



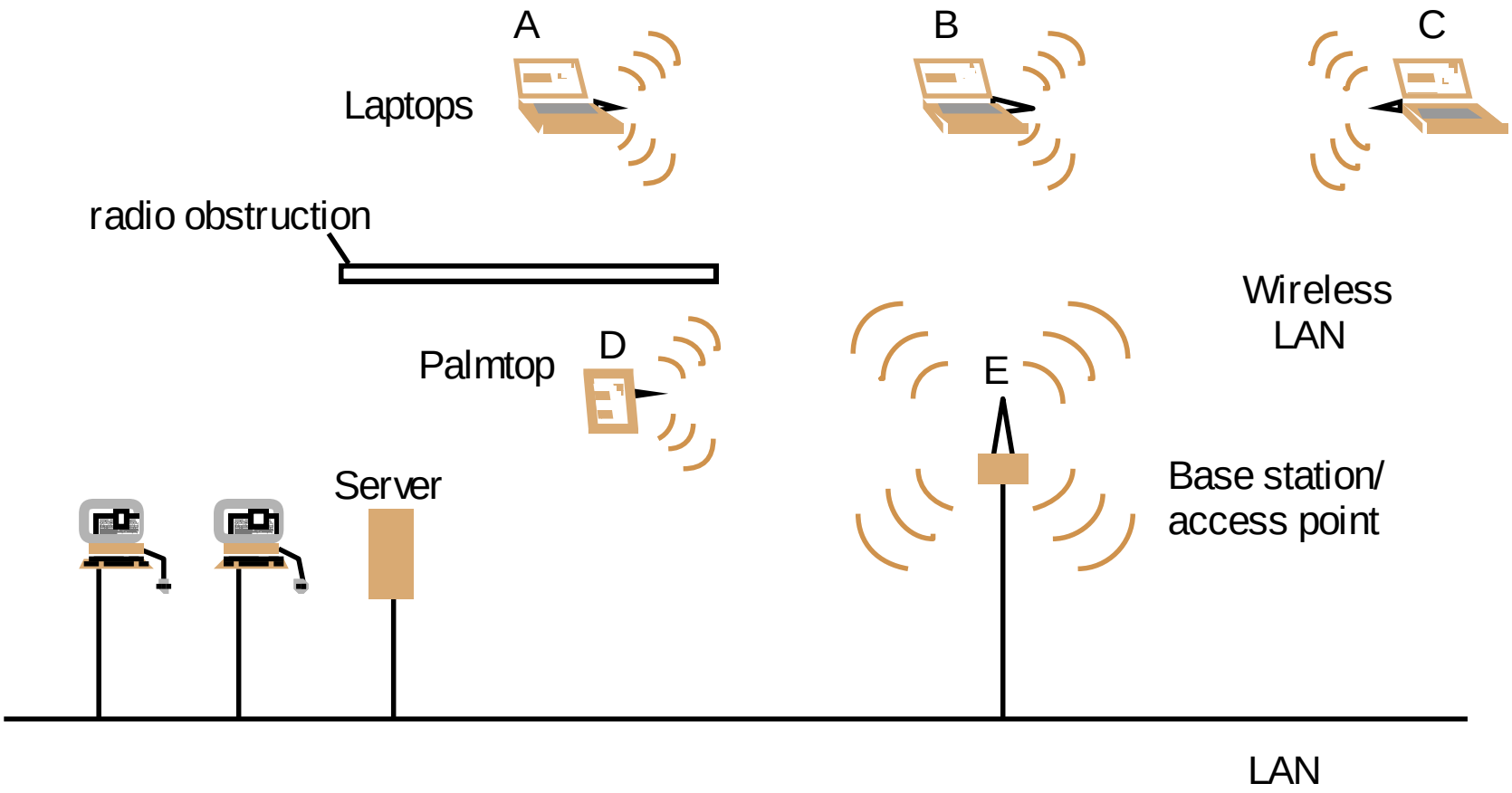
IEEE 802 Network Standards

IEEE No.	*Name*	*Title*	*Reference*
802.3	Ethernet	CSMA/CD Networks (Ethernet)	[IEEE 1985a]
802.4		Token Bus Networks	[IEEE 1985b]
802.5		Token Ring Networks	[IEEE 1985c]
802.6		Metropolitan Area Networks	[IEEE 1994]
802.11	WiFi	Wireless Local Area Networks	[IEEE 1999]
802.15.1	Bluetooth	Wireless Personal Area Networks	[IEEE 2002]
802.15.4	ZigBee	Wireless Sensor Networks	[IEEE 2003]
802.16	WiMAX	Wireless Metropolitan Area Networks	[IEEE 2004a]

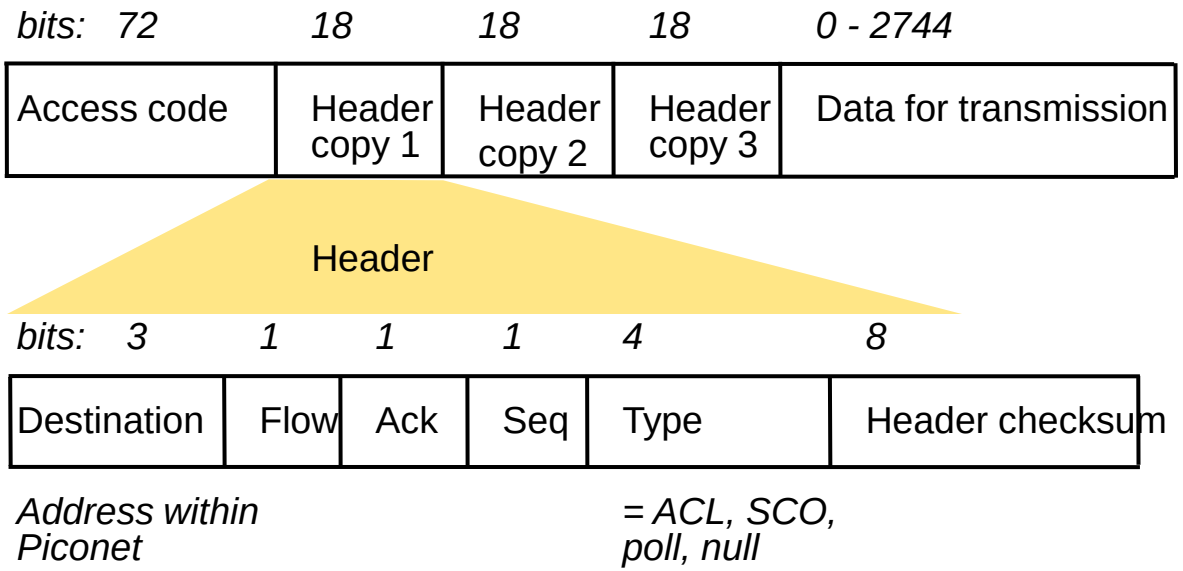
Ethernet Ranges and Speeds

	10Base5	*10BaseT*	*100BaseT*	*1000BaseT*
Data rate	10 Mbps	10 Mbps	100 Mbps	1000 Mbps
Max. segment lengths:				
Twisted wire (UTP)	100 m	100 m	100 m	25 m
Coaxial cable (STP)	500 m	500 m	500 m	25 m
Multi-mode fibre	2000 m	2000 m	500 m	500 m
Mono-mode fibre	25000 m	25000 m	20000 m	2000 m

Wireless LAN Configuration



Bluetooth Frame Structure



SCO packets (e.g. for voice data) have a 240-bit payload containing 80 bits of data triplicated, filling exactly one timeslot.

END