MASTER IN COMPUTER SCIENCE

UNIVERSITÉ DE NEUCHÂTEL

DISTRIBUTED SYSTEMS

Chapter 3 - Networking and Inter-networking

Dr Valerio Schiavoni and Prof Peter Kropf
University of Neuchâtel – Fall 2020

SUMMARY

NETWORKING AND INTERNETWORKING
Introduction

Types of network

Network principles

Internet protocols

ROADMAP

UNIVERSITÉ DE NEUCHÂTEL

Applications and Services

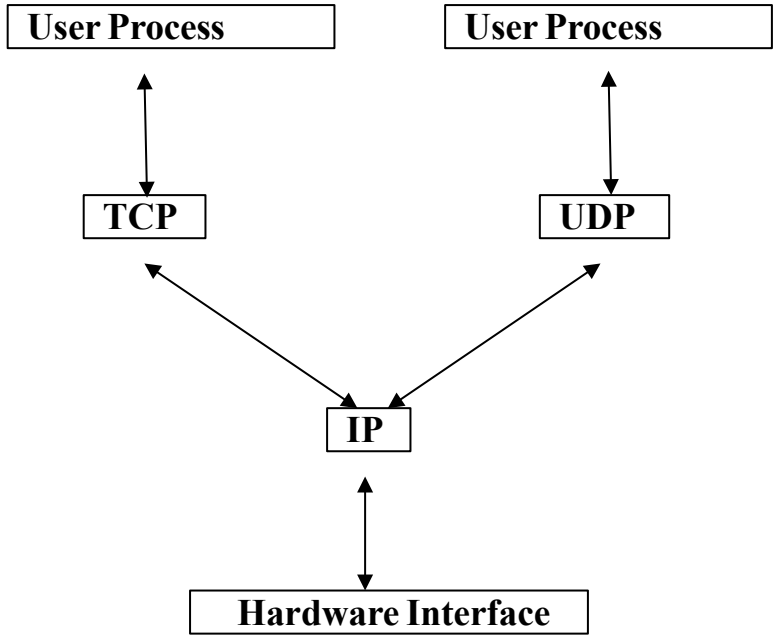
Remote invocation, indirect communication

Underlying interprocess communication
sockets, messages, multicast, overlays

TCP & UDP

THE LAYERED INTERNET COMMUNICATION SYSTEM

The Internet-Protocol-Suite relies on a 4-layer model.



Process Layer

Transport Layer

Network Layer

Communication Layer

INTRODUCTION

- Networks are the communication subsystem
  - Media: wire, cable, fiber, air
  - Devices: routers, switches, bridges, hubs, repeaters, interfaces
  - Software: protocol stacks, communication handlers and drivers
  - All affect functionality and performance

- Computers are hosts
  - Nodes are computers and other switching devices
- Internet: the network of all networks
  - It has many subnets
    - Collection of nodes
    - Same physical network
    - Routing units
- In this chapter
  - Introductory overview of computer networking
  - Communication requirements of distributed systems

NETWORKING ISSUES FOR DISTRIBUTED SYSTEMS

transmission time = *latency* + *message length／bandwidth*

- Scalability
- Reliability
- Security
- Mobility
- Quality of service
- Multicasting

COMMUNICATION CHANNELS

UNIVERSITÉ DE NEUCHÂTEL

- Performance characteristics of communication channels
  - **Latency:** delay (time) between sending and receiving a message including:
    - Network access time
    - Time for the first bit of the message to travel from the sender's network interface to the receiver's one.
    - Processing time within the sending and receiving processes
  - **Bandwidth**: amount of information (usually bits) transmitted per time unit
  - **Throughput:** number of transmission units transmitted per time unit
    - **Bandwidth-Delay Product** to affect the maximum throughput of channels
  - **Delay jitter:** variation in delay between different messages of the same type (e.g. video frames).

TYPES OF NETWORKS

- Local area networks (LANs)
  - Ethernet
- Wide area networks (WANs)
- Metropolitan area networks (MANs)
- Personal area networks (WPANs)
  - Bluetooth
- Wireless local area networks (WLANs)
  - WiFi
- Wireless metropolitan area networks (WMANs)
  - WiMax
- Wireless wide area networks (WWANs)
  - GSM
- Internetworks

TYPES OF NETWORKS
PERFORMANCE COMPARISON

	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
	4G phone	50km	20	50
	5G phone	200–300m.	10-20 (Gbps)	1

NETWORK PRINCIPLES

- Packet transmission
- Data streaming
- Switching
  - Broadcasting
  - Circuit switching
  - Packet switching
  - Frame relay
- Protocols
  - Layers

SWITCHES

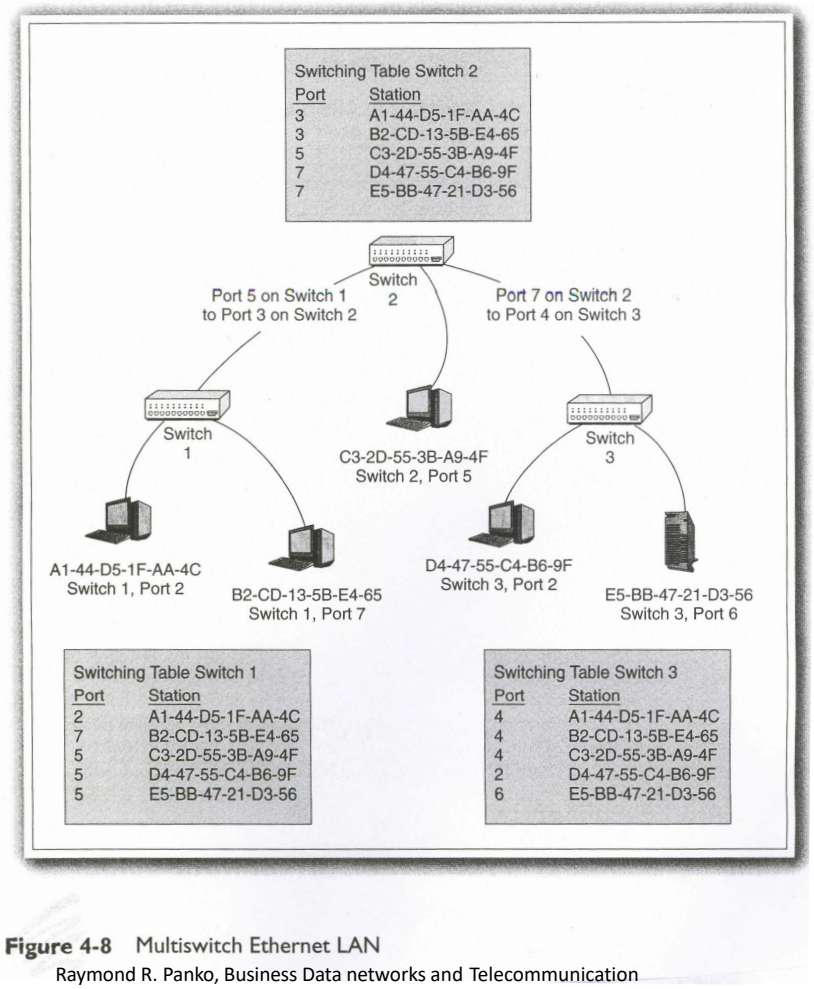
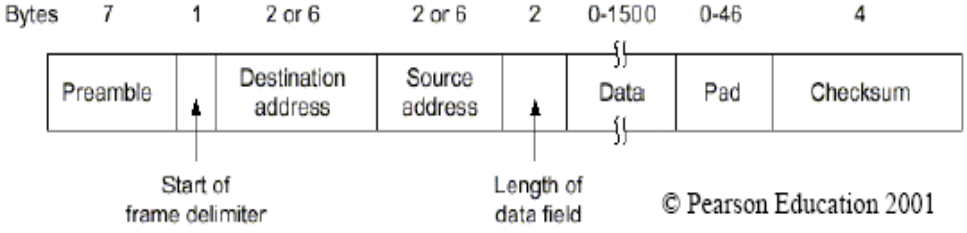


Figure 4-8 Multiswitch Ethernet LAN

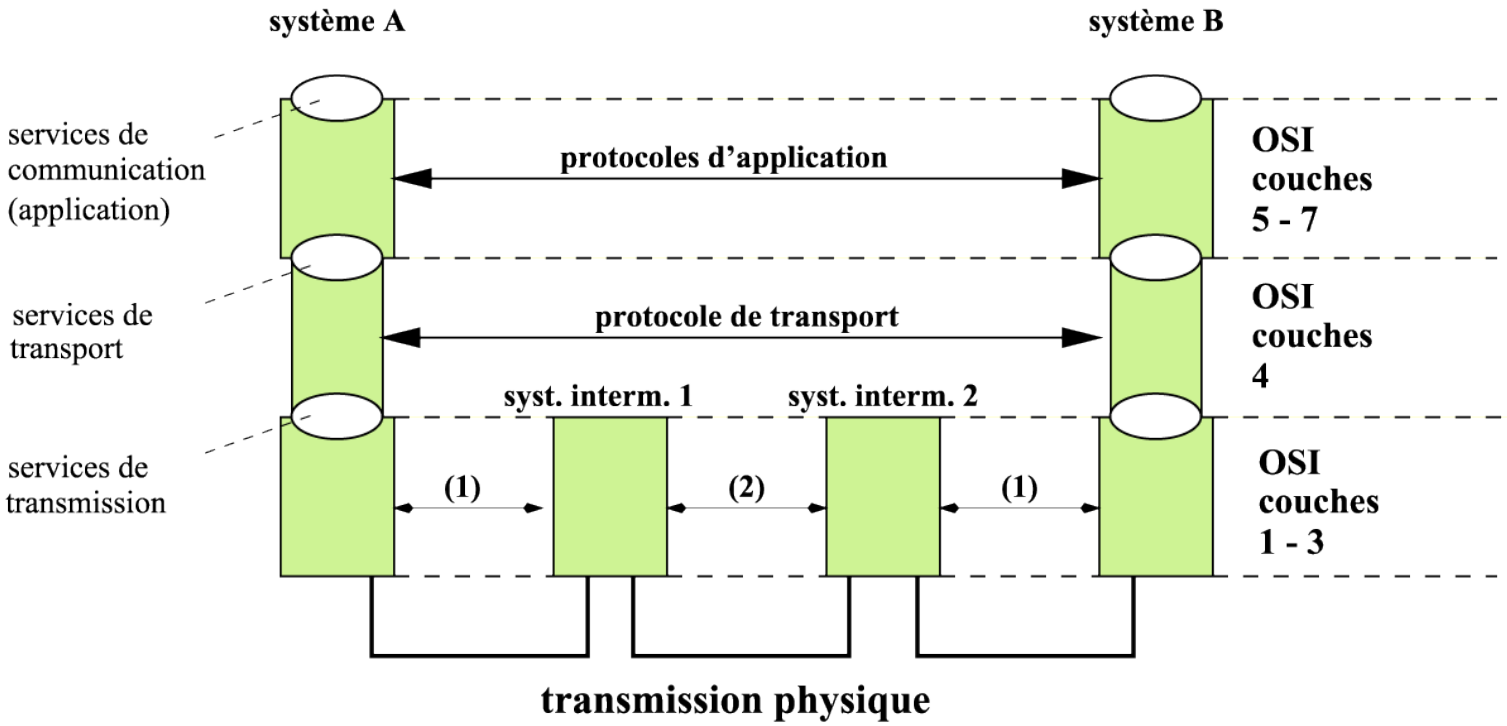
Raymond R. Panko, Business Data networks and Telecommunication

Connect machines/hosts switching frames

Ethernet frame



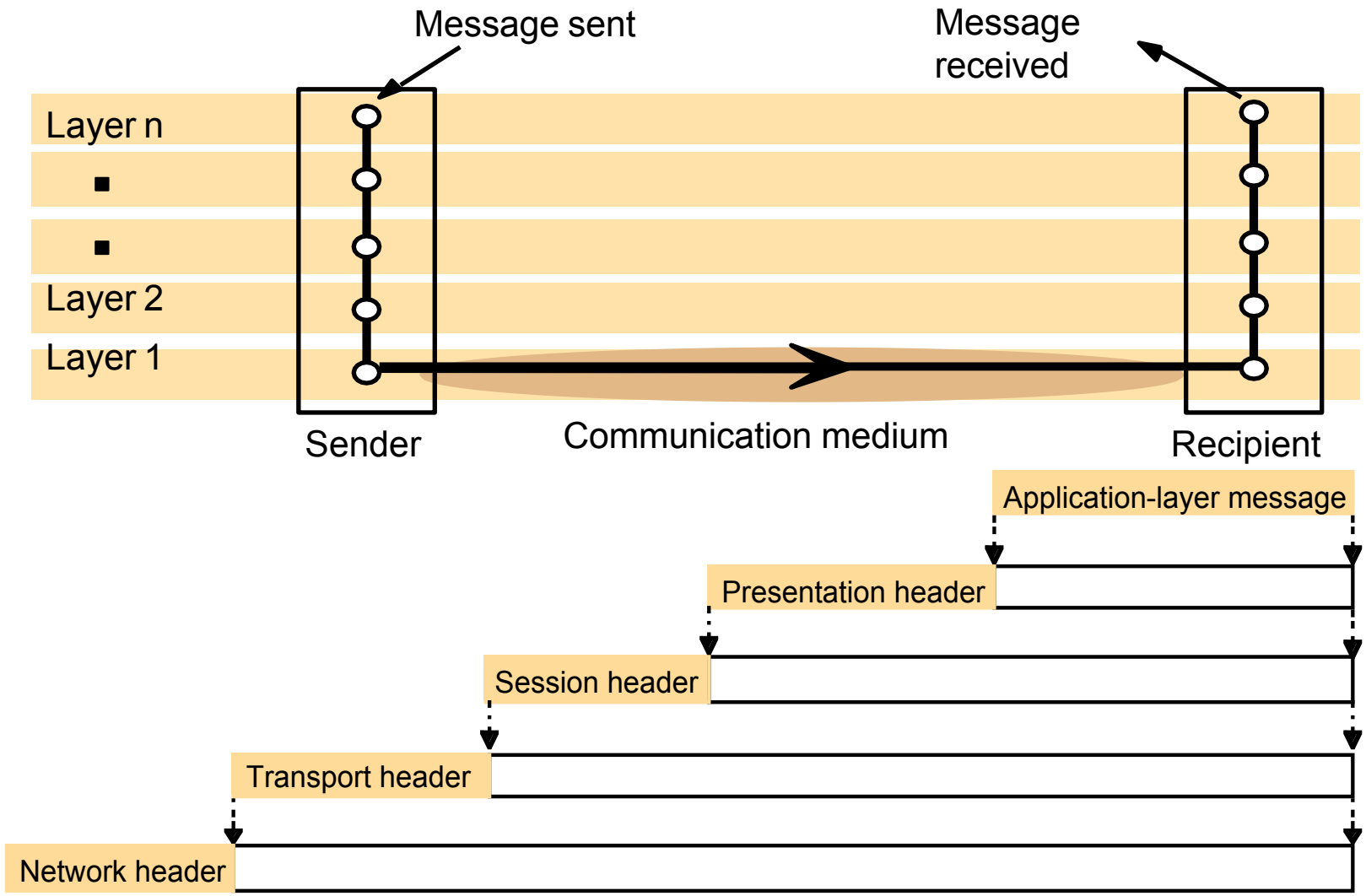
LAYERED NETWORK MODEL



OSI LAYERS

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

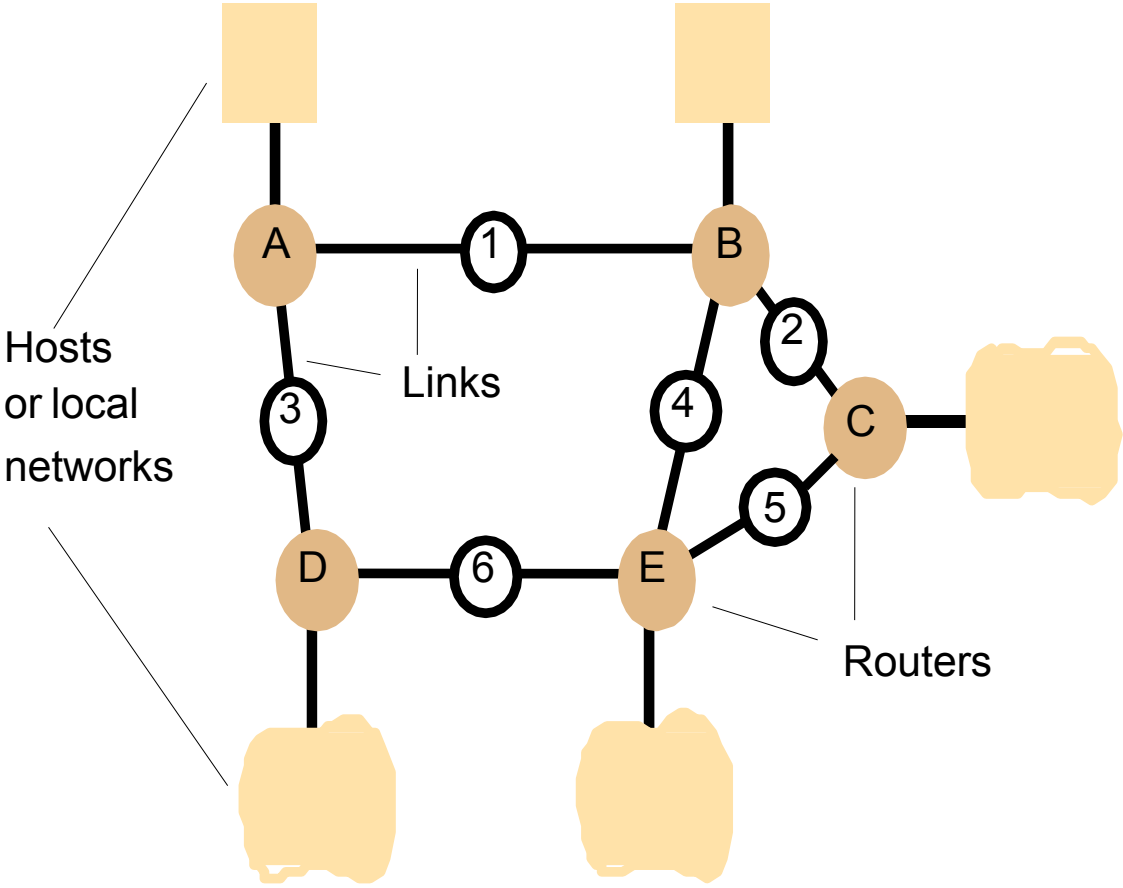
PROTOCOL LAYERS : ENCAPSULATION



Instructor's Guide for Coulouris, Dollimore, Kindberg and Blair, Distributed Systems: Concepts and Design Edn. 5

PROTOCOLS

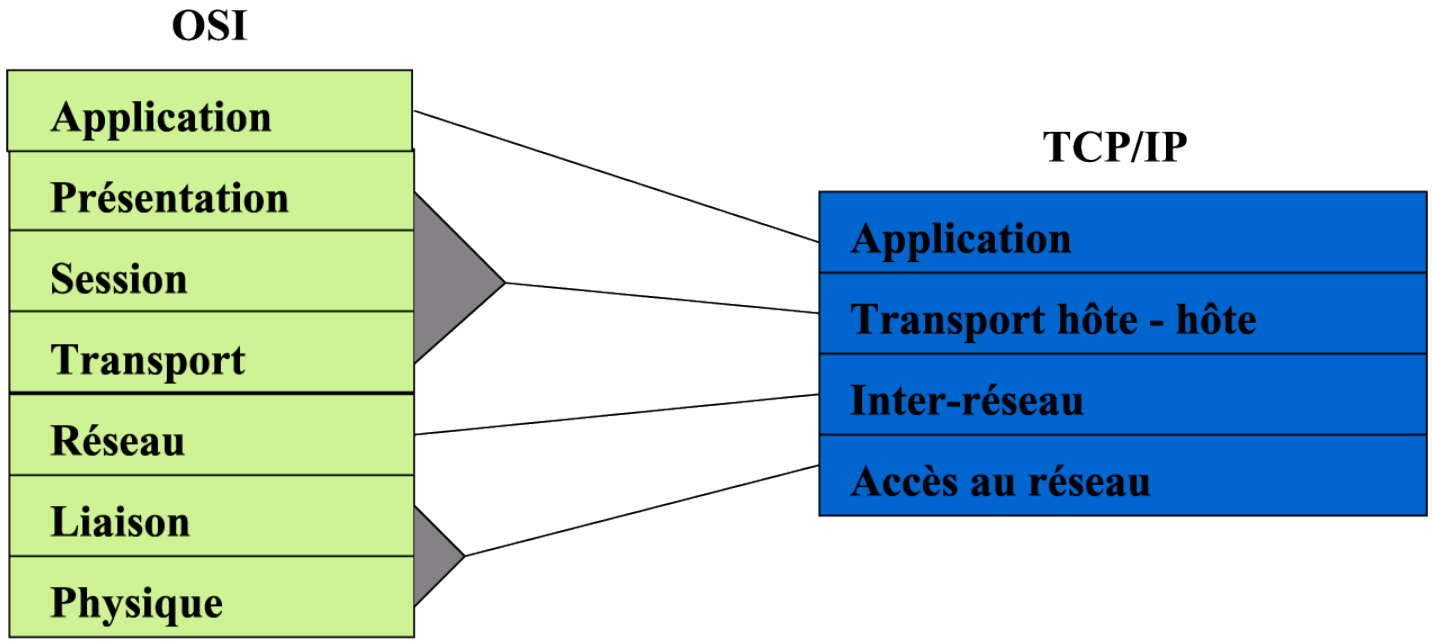
- Packet assembly
- Addressing, ports
- Packet delivery
  - Datagram
  - Virtual circuit
- Routing
  - Routers
  - Bridges
  - Hubs
  - Switches
  - Tunneling
- Congestion control



Instructor's Guide for Coulouris, Dollimore, Kindberg and Blair, Distributed Systems: Concepts and Design Edn. 5
© Pearson Education 2012

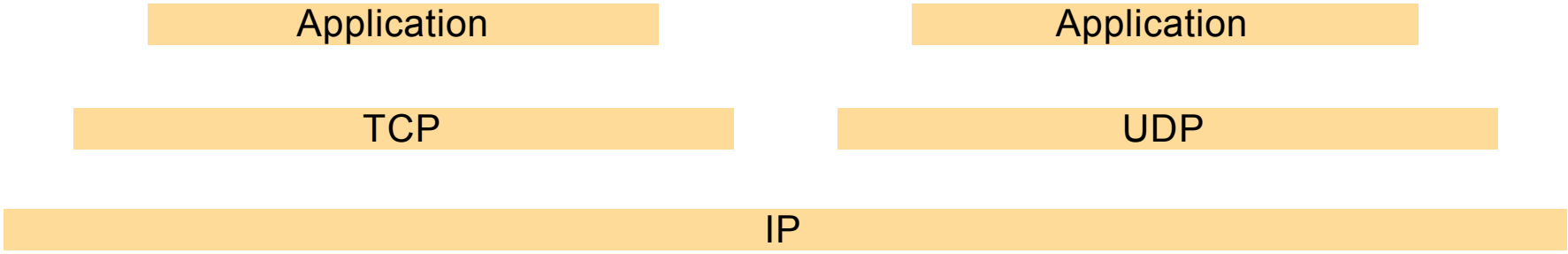
TCP/IP MODEL

- The 4 layer TCP/IP model



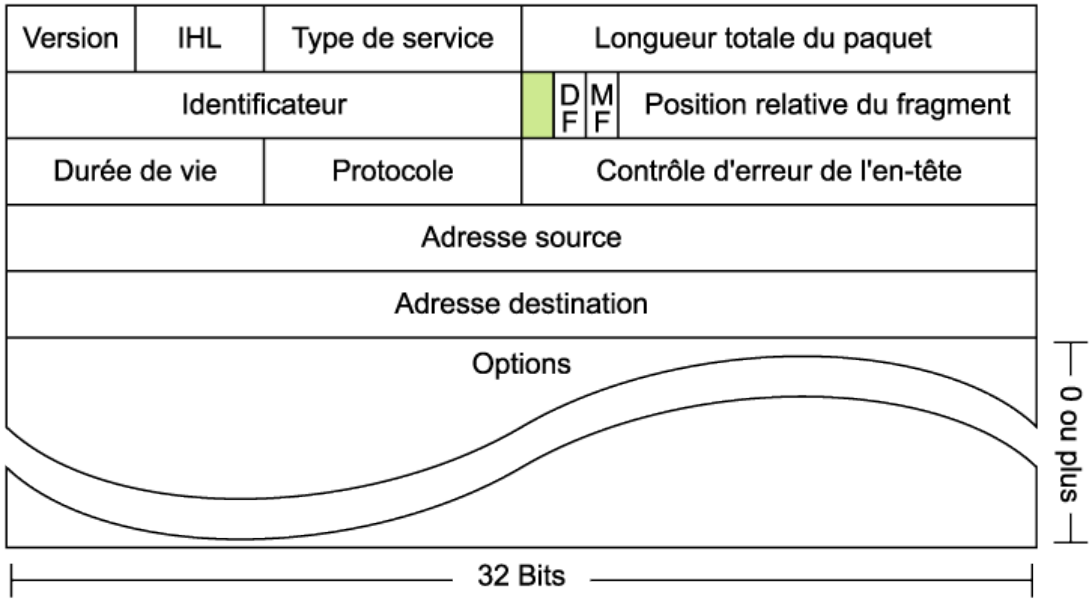
INTERNET PROTOCOLS

BGP DHCP DNS FTP HTTP IMAP LDAP NTP POP RSVP RIP
SIP SMTP SNMP SSH TLS/SSL …



IP PACKET

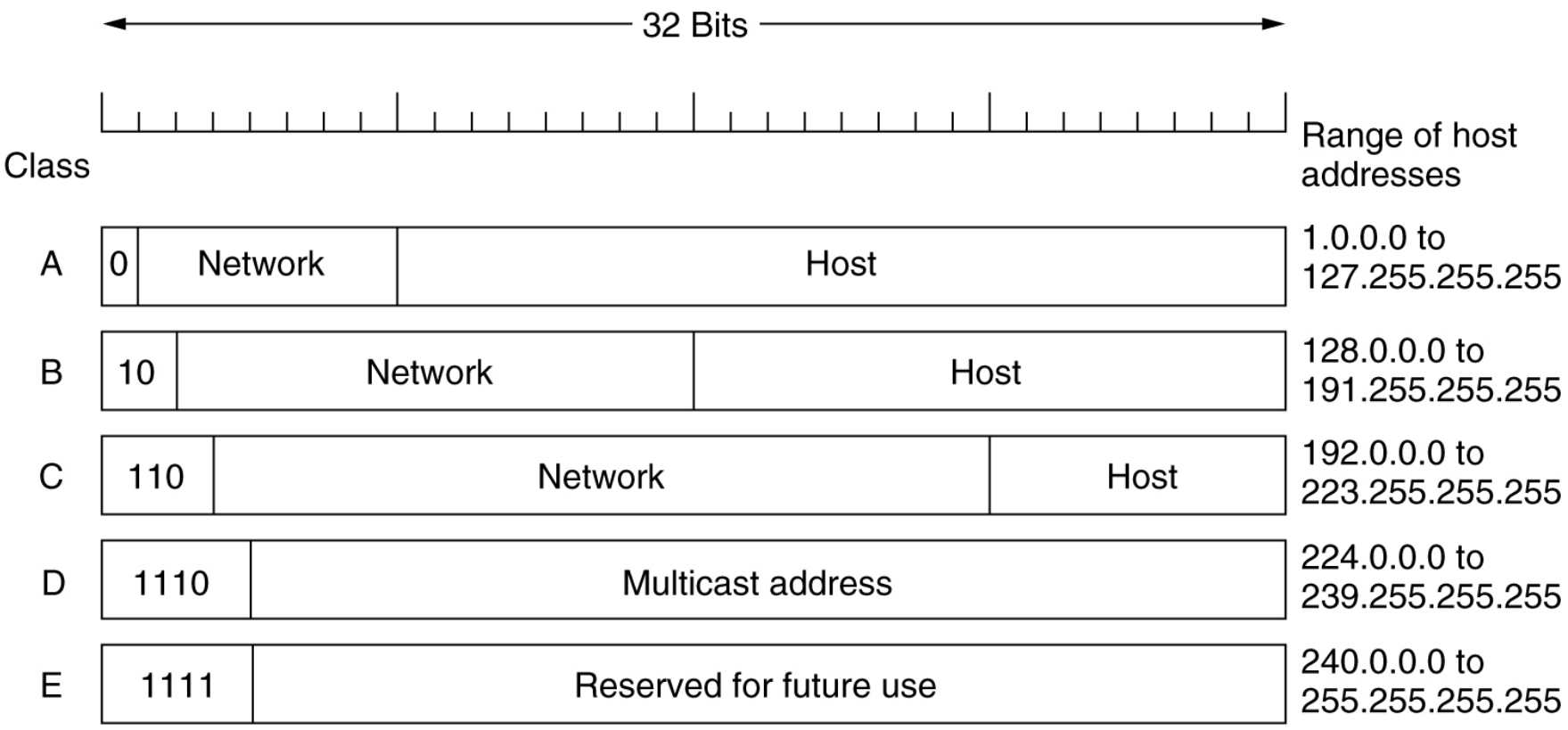
- The IP packet is composed of :
  - Header (20 Bytes)
  - Options (0 or more (max 10) words of 4 Bytes)
  - Data
- The length of an IP packet is limited to 64KBytes



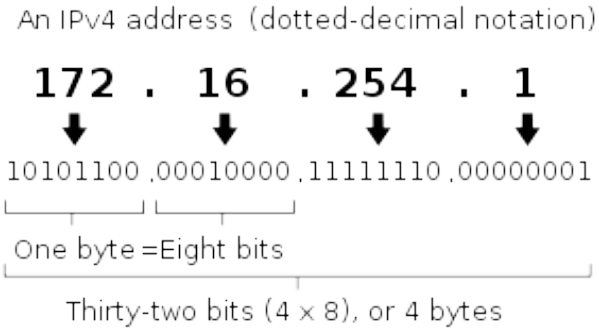
ADDRESSING IN THE INTERNET PROTOCOL

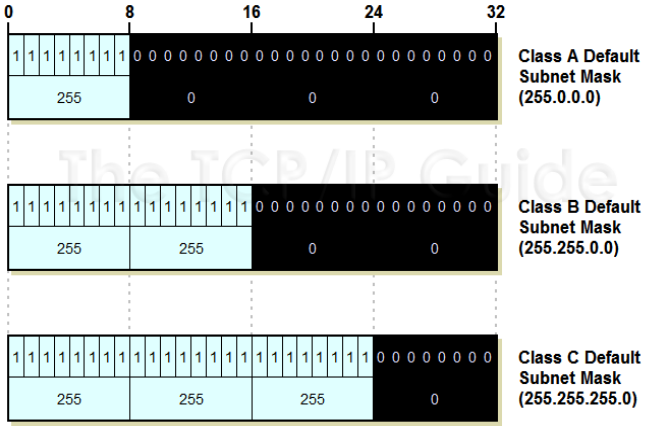
- Addresses used in source and destination fields of the protocol header
- Requirements:
  - Define *unique* addresses for any machine on the Internet: no two nodes may use the same address
  - Define a sufficiently large address space
    - **IPv4** : 32-bit addresses. This allows for 2^{32} (approx. 4 billion) addresses
      - Insufficient : unexpected growth of the Internet; inefficient use of address space
    - **IPv6** : 128-bit addresses. This allows for 2^{128} (approx. 3×10^{38}) addresses.
      - Corresponds to about 7×10^{23} addresses per m^2
  - Support a flexible and efficient *routing* scheme without that the addresses themselves contain routing information

IP ADDRESSES



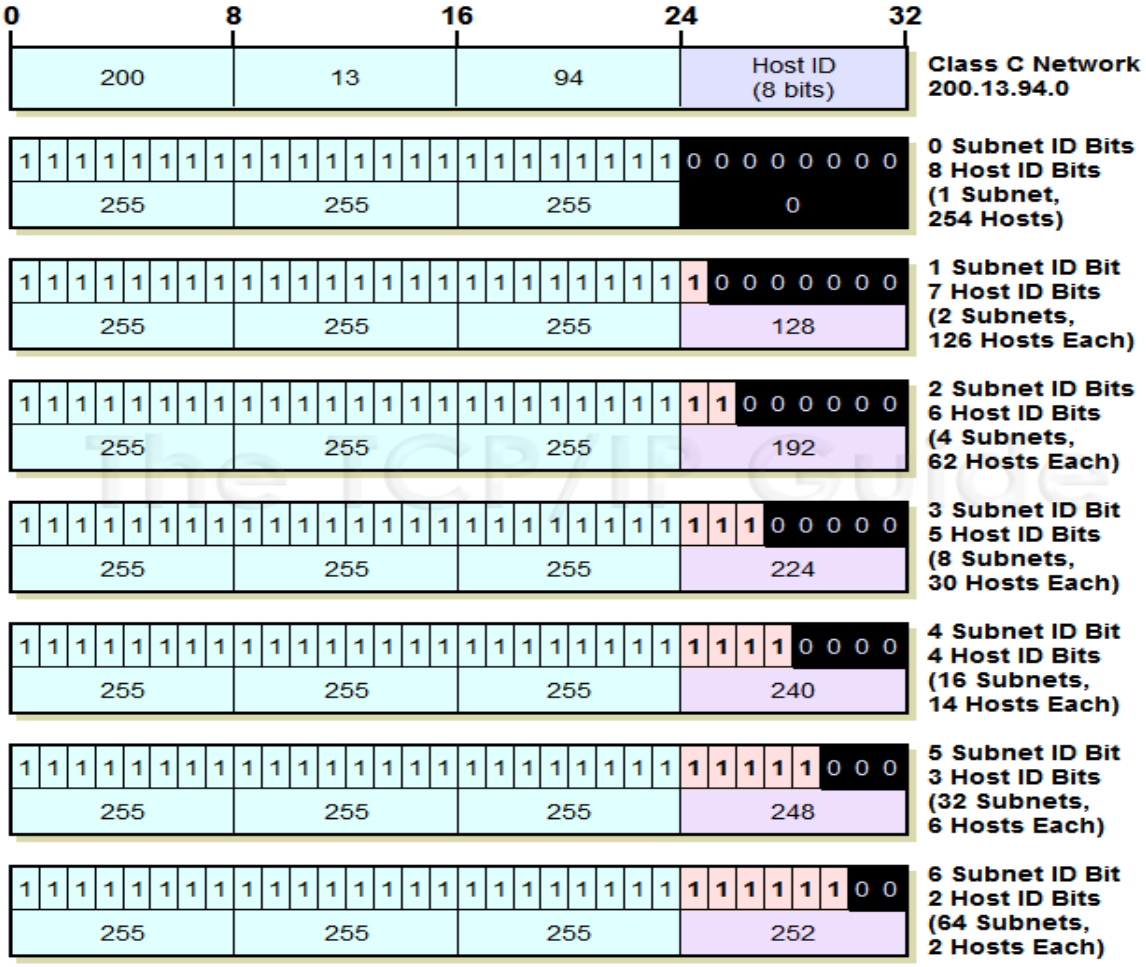
NETWORK ADDRESSES, CLASSES, MASKS





Class Leading bits

A 0
B 10
C 110



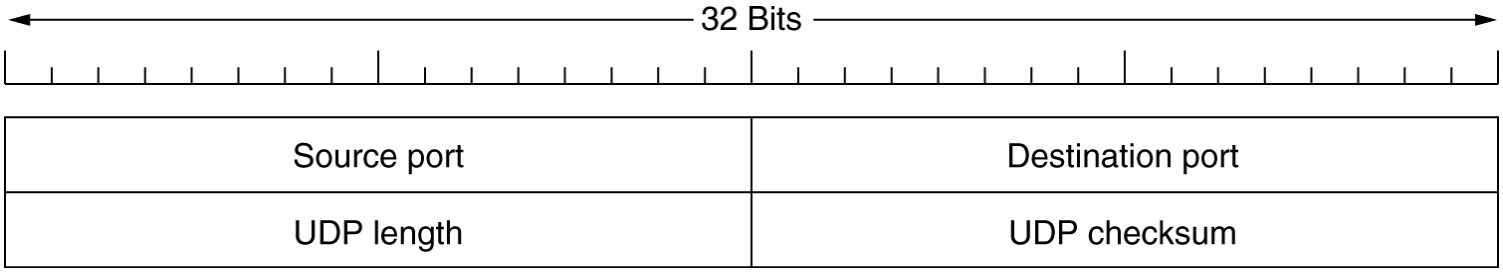
IP PROTOCOL

- Address resolution
- Routing
  - Protocols
  - Spoofing
  - Classless InterDomain Routing (CIDR)
  - Network Address Translation (NAT)

TCP/IP

- Transport Layer
  - Two protocols
    - TCP and UDP
  - TCP: « Transmission Control Protocol »
    - Offers a connection oriented and duplex communication
  - UDP: « User Datagram Protocol »
    - Offers a non reliable connectionless communication

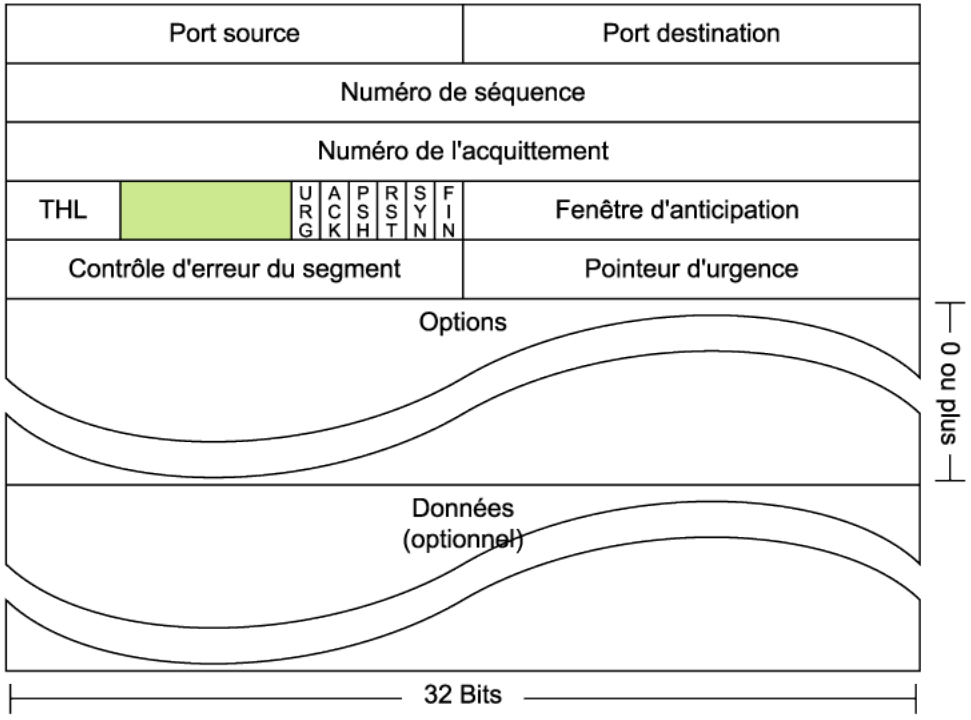
UDP PROTOCOL HEADER



Port	Protocol	Use
21	FTP	File transfer
23	Telnet	Remote login
25	SMTP	E-mail
69	TFTP	Trivial File Transfer Protocol
79	Finger	Lookup info about a user
80	HTTP	World Wide Web
110	POP-3	Remote e-mail access
119	NNTP	USENET news

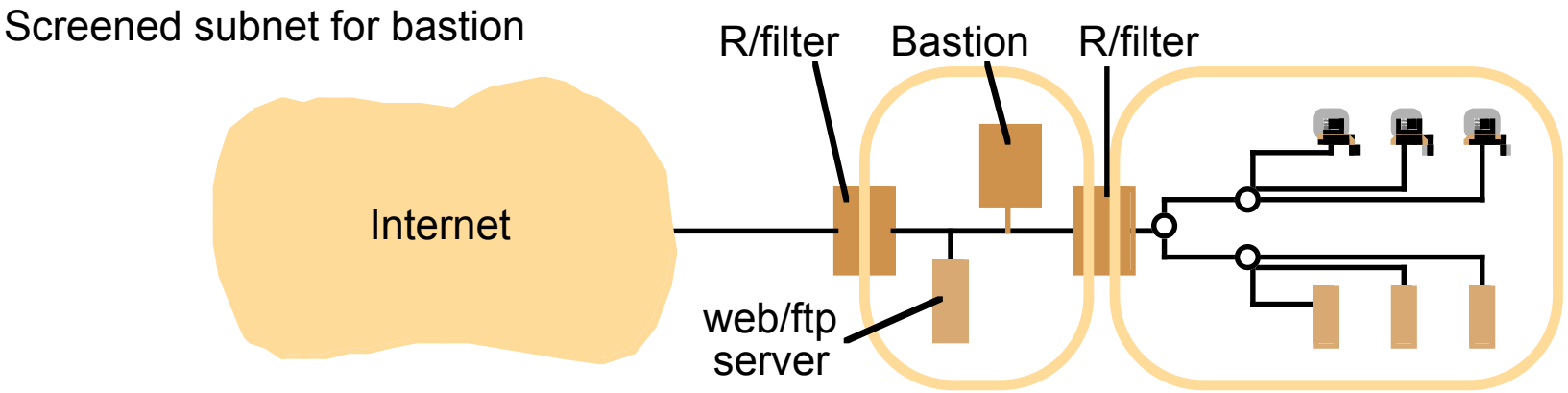
SEGMENT TCP

- The TCP segment is composed of a header (20 Bytes), options (0 or more) and data (optional). The maximum length of a segment is 65515 Bytes
- TCP applies a *go-back-N* strategy with sliding window for error control and dataflow



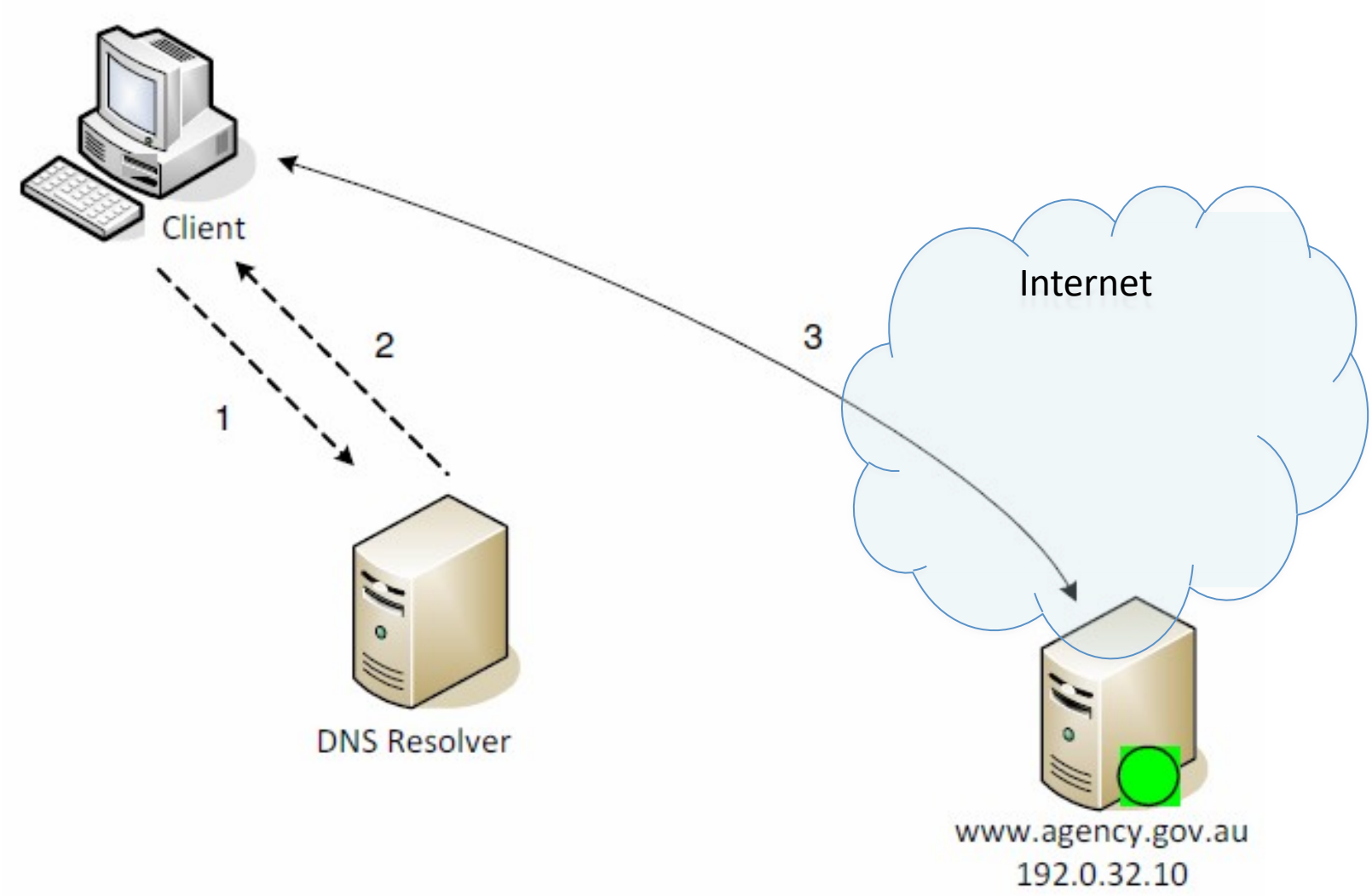
TCP, UDP, NAMES, FIREWALL

- UDP
- TCP
  - Sequencing
  - Flow control
  - Retransmission
  - Buffering
- Domain names
  - Name <-> address
- Virtual Private Network
- Firewall



Instructor's Guide for Coulouris, Dollimore, Kindberg and Blair, Distributed Systems: Concepts and Design Edn. 5

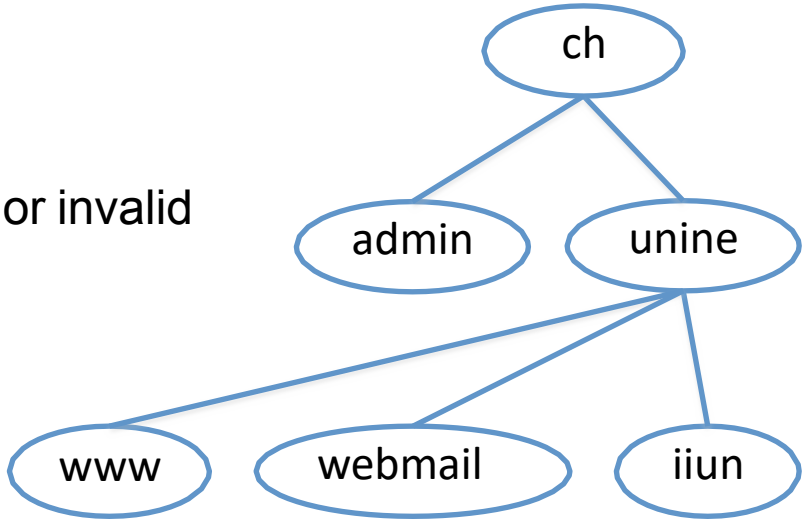
NAME SERVICES AND THE DNS NAME RESOLUTION



http://www.asd.gov.au/publications/protect/dns_security.htm

NAME SERVICES AND THE DNS NAME SPACE

- Name space
  - Collection of all names
  - Names may be unbound (no binding) or invalid
  - Often hierarchical
    - Separate contexts
    - Potentially infinite, easy to split
- DNS names
  - Domain names (hosts or domains)
  - Hierarchical / www.unine.ch webmail.unine.ch iiun.unine.ch
- Naming domain
  - Name space with an administrative authority
- Merging: add higher-level name
- Alias: name that denotes another name



NAME SERVICES AND THE DNS

- Domain names
  - Top-level domains
    - com, edu, gov, mil, net, org, int
    - ch, uk, fr, de, it, us, tu, cc, …
    - biz, mobi, museum, …
- Queries
  - Name resolution (address, alias, authority)
  - Attributes (mail host, host info, arbitrary data)

DNS RESOURCE RECORD

Resource record format :

```
<name> [<ttl>] [<class>] <type> <rdlength> <rdata>
```

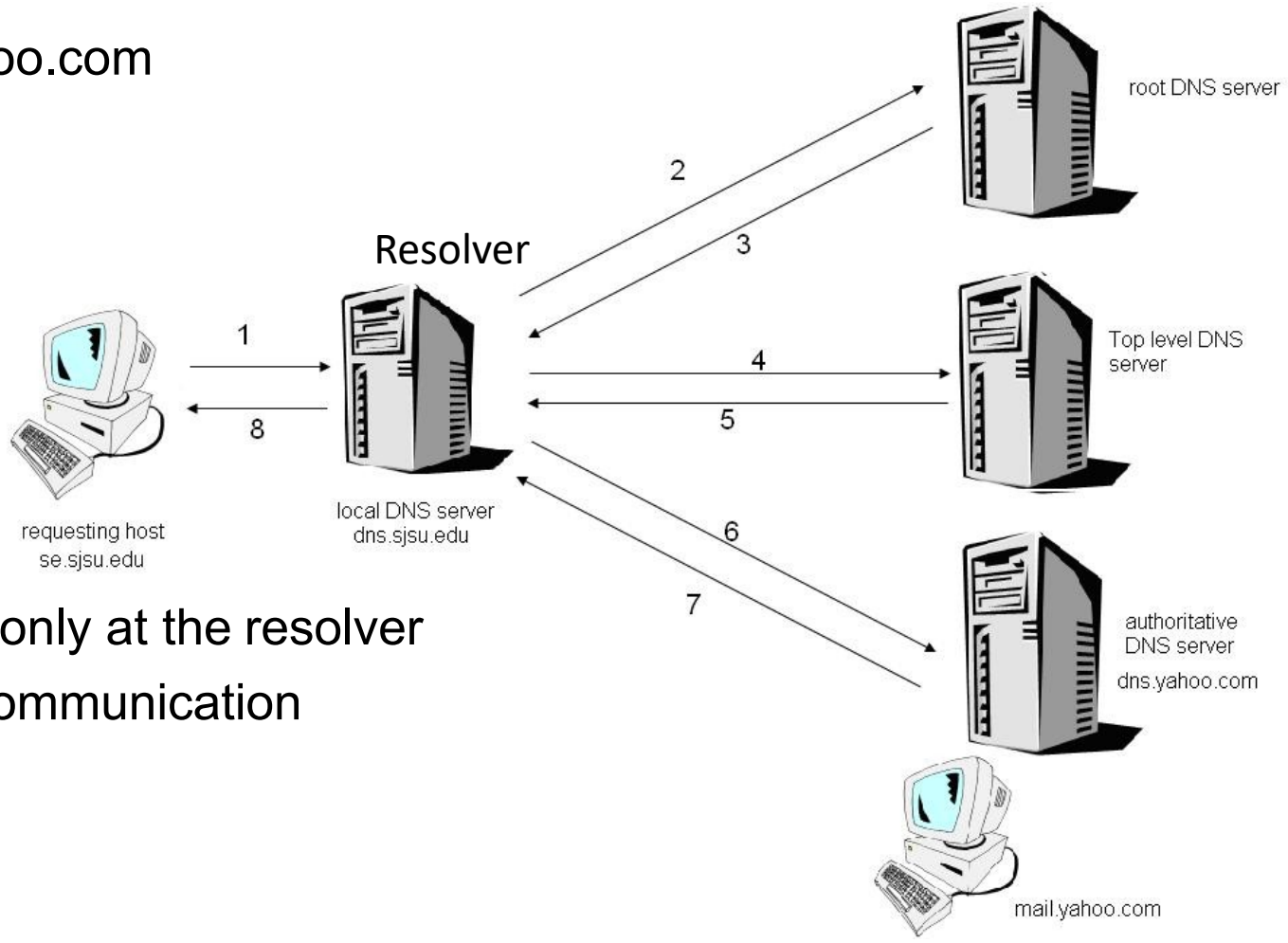
Record type	*Meaning*	*Main contents*
A	A computer address (IPv4) A	IPv4 number
AAAA	computer address (IPv6) An	IPv6 number
NS	authoritative name server	Domain name for server
CNAME	The canonical name for an alias	Domain name for alias
SOA	Marks the start of data for a zone	Parameters governing the zone
PTR	Domain name pointer (reverse lookups)	Domain name
HINFO	Host information	Machine architecture and operating
MX	Mail exchange	List of <*preference, host*> pairs
TXT	Text string	Arbitrary text

DNS EXAMPLE ZONE

```
$ORIGIN example.com.     ; designates the start of this zone file in the namespace
$TTL 1h                  ; default expiration time of all resource records without their own TTL value
example.com.    IN  SOA   ns.example.com. username.example.com. ( 2007120710 1d 2h 4w 1h )
example.com.    IN  NS    ns                    ; ns.example.com is a nameserver for example.com
example.com.    IN  NS    ns.somewhere.example. ; ns.somewhere.example is a backup nameserver for example.com
example.com.    IN  MX    10 mail.example.com.  ; mail.example.com is the mailserver for example.com
@               IN  MX    20 mail2.example.com. ; equivalent to above line, "@" represents zone origin
@               IN  MX    50 mail3              ; equivalent to above line, but using a relative host name
example.com.    IN  A     192.0.2.1             ; IPv4 address for example.com
                IN  AAAA  2001:db8:10::1        ; IPv6 address for example.com
ns              IN  A     192.0.2.2             ; IPv4 address for ns.example.com
                IN  AAAA  2001:db8:10::2        ; IPv6 address for ns.example.com
www             IN  CNAME example.com.          ; www.example.com is an alias for example.com
wwwtest         IN  CNAME www                   ; wwwtest.example.com is another alias for www.example.com
mail            IN  A     192.0.2.3             ; IPv4 address for mail.example.com
mail2           IN  A     192.0.2.4             ; IPv4 address for mail2.example.com
mail3                     192.0.2.5             ; IPv4 address for mail3.example.com
```

ITERATIVE RESOLUTION

- Resolve:
  mail.yahoo.com

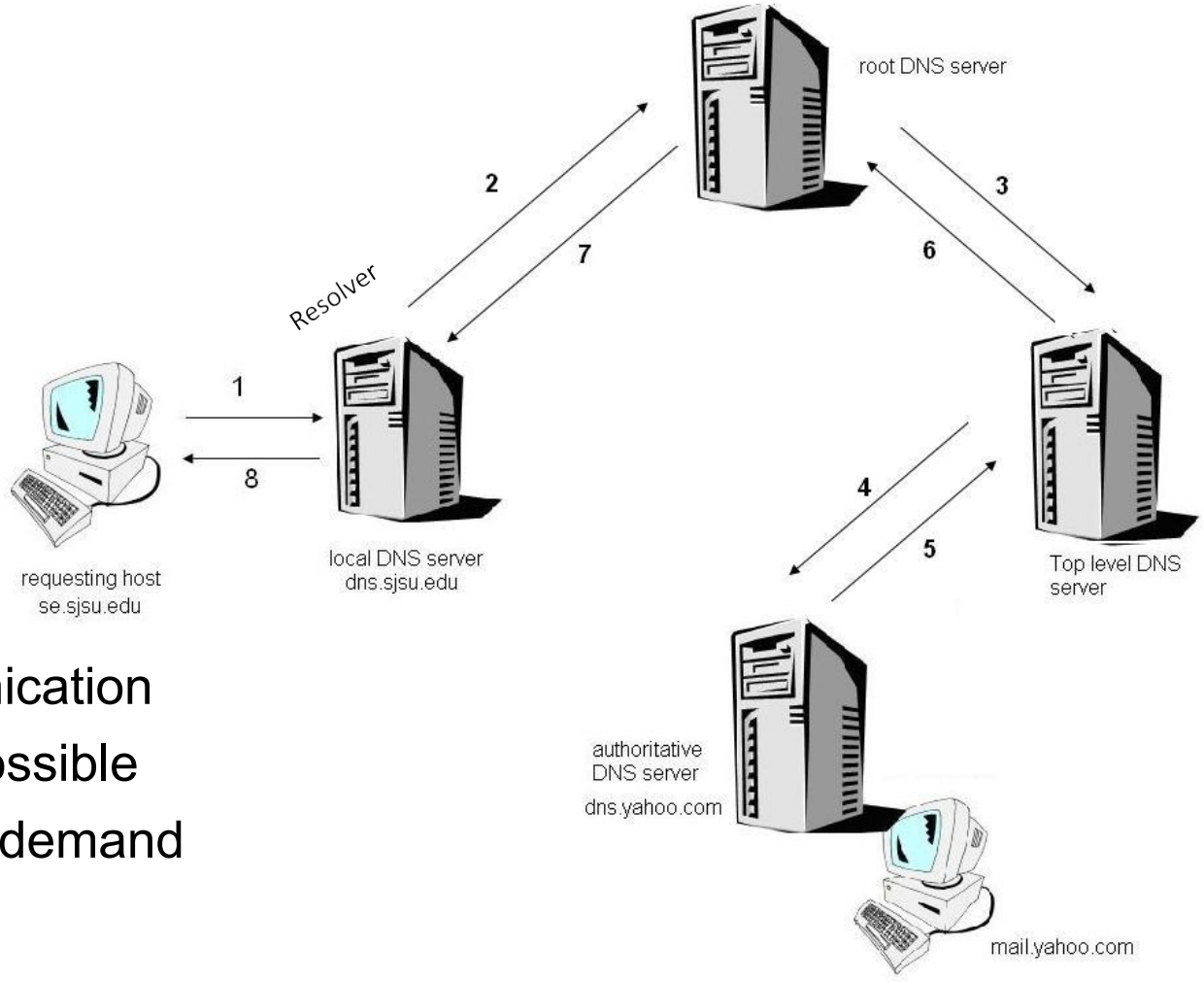


hbps://en.wikibooks.org/wiki/Communica<on_Networks/DNS

- Caching only at the resolver
- Lots of communication

RECURSIVE RESOLUTION

- Resolve:
  mail.yahoo.com



- Caching everywhere
- Reduce the communication
- Access protection possible
- Higher performance demand

hbps://en.wikibooks.org/wiki/Communica<on_Networks/DNS

DNS IMPLEMENTATION

- Partitioning
  - Each zone implemented by a name server
- Replication
  - Each zone replicated on at least two servers
  - Updates performed on primary
  - Contents transferred to secondary using zone transfer
  - Higher levels have many more replicas (12 root replicas)
- Caching
  - Servers cache results of queries
  - Original entries have time-to-live field (TTL)
  - Cached data is non-authoritative, provided until TTL expires
- Name Resolution
  - Query sent to local server
  - If cannot resolve locally then sent to root
  - Resolved recursively or iteratively

TRANSMISSION EXAMPLE

- First frame of an FTP transmission:
  - **Ethernet**:

    00 a0 24 0f 29 f5 00 e0 16 05 b6 87 08 00
  - **IP**:

    45 00 00 2c 05 0f 00 00 3a 06 e5 d3 84 cb 1a 07 84 cb 72 4c
  - **TCP**:

    05 ac 00 15 1f fc 7a 00 00 00 00 00 60 02 10 00 52 7f 00 00
    02 04
    05 b4 10 00

TRANSMISSION EXAMPLE

Description	Valeurs	Format hexadécimal
Trame Ethernet		
Adresse Ethernet source	00:E0:16:05:B6:87	00e0 1605 b687
Adresse Ethernet destination	00:A0:24:0F:29:F5	00a0 240f 29f5
Protocole encapsulé	IP	0800
En-tête IP		
Version	4	4
Longueur en-tête	20 octets	5
Type de service	0x00	00
Longueur totale	44 octets	002c
Identification	1295	050f
TTL	58	3a
Protocole encapsulé	TCP	06
« Checksum » en-tête		e5d3
Adresse IP source	132.203.26.7	84cb 1a07
Adresse IP destination	132.203.114.76	84cb 724c

TRANSMISSION EXAMPLE

Description	Valeurs	Format hexadécimal
En-tête TCP		
Port source	1452	05ac
Port destination	21 (ftp)	0015
Numéro de séquence	0536640000	1ffc 7a00
Numéro d'acquittement	0000000000	0000 0000
Longueur en-tête	24 octets	6002
« Flags »	URG=off, ACK=off, PSH=off, RT=off, SYN=on, FIN=off	011000000000010
Grandeur fenêtre	4096 octets	
« Checksum »		527f
Pointeur urgence	0	0000
Options	Type=2, len=4, MSS=1460	0204 5b4
Ethernet «padding »		1000
Données TCP	< pas de données >	

HOMEWORK

How does a newly installed personal computer connected to an Ethernet discover the IP addresses of local servers?
How does it translate them to Ethernet addresses?

What are the disadvantages of using broadcasting to locate resources:

i) in a single Ethernet?

ii) in the Internet?