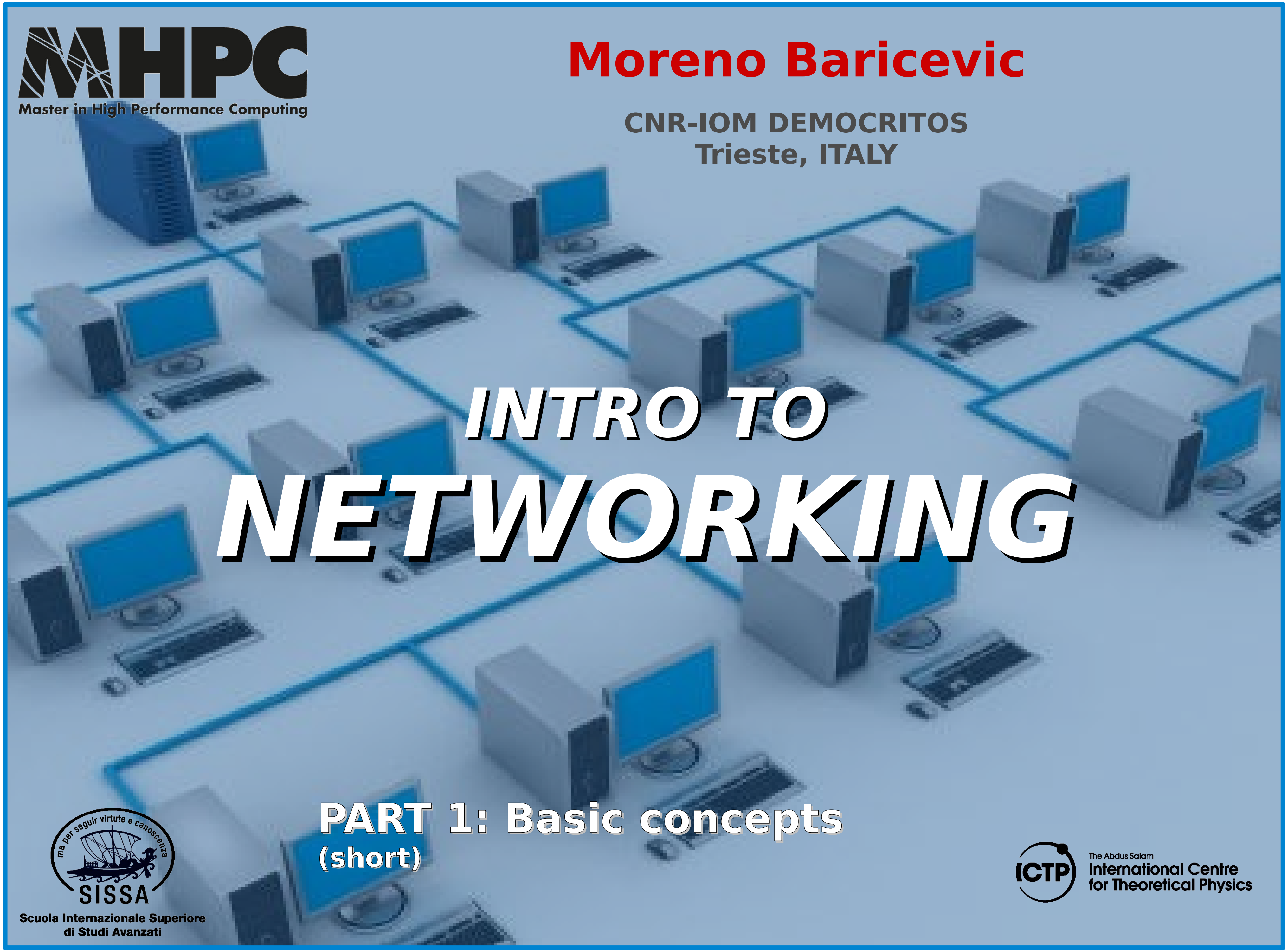
MHPC
Master in High Performance Computing

Moreno Baricevic

CNR-IOM DEMOCRITOS
Trieste, ITALY



INTRO TO NETWORKING

PART 1: Basic concepts
(short)

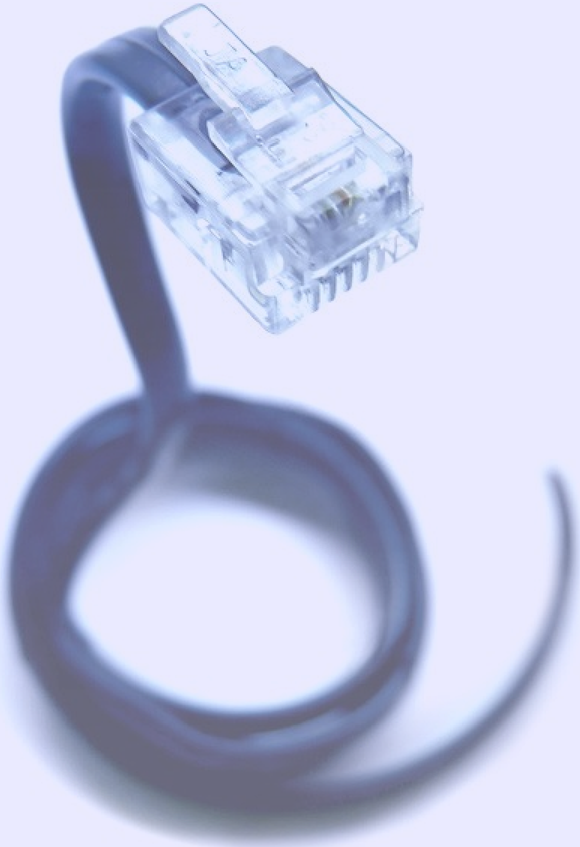
ma per seguir virtute e canoscenza
SISSA
Scuola Internazionale Superiore di Studi Avanzati

ICTP
The Abdus Salam International Centre for Theoretical Physics

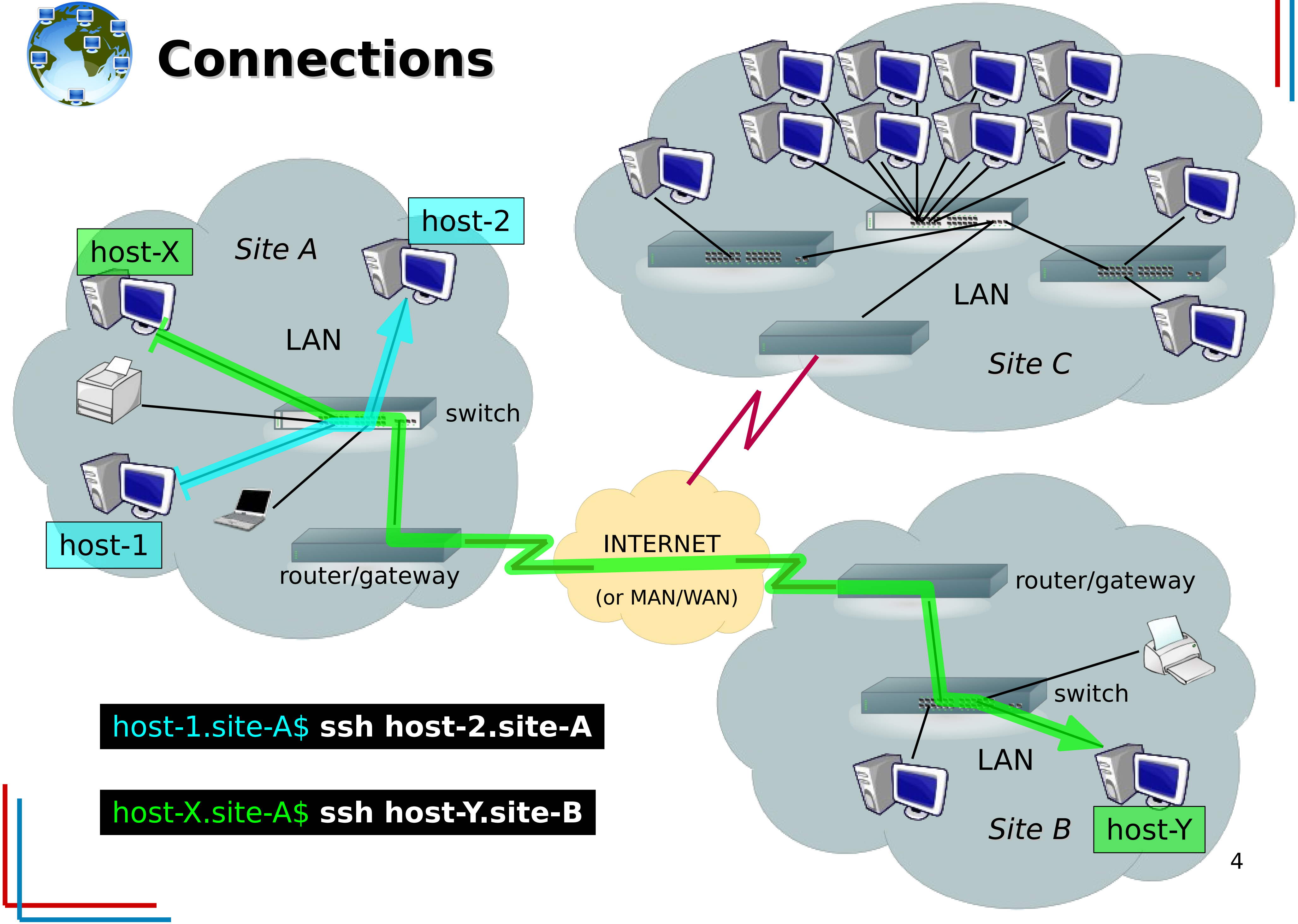
Agenda

- Connections
- Concept of Packet
- Network Stack Models (TCP/IP - ISO/OSI)
- Internet Protocol and IP Address Space
- Ethernet and Physical Address
- LINUX commands (configuration and diagnostic)

Connections



Connections

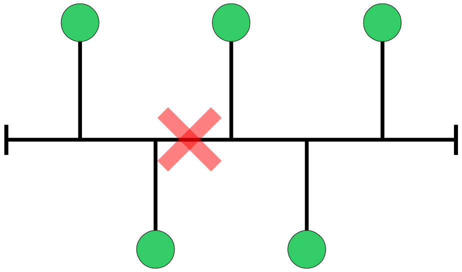


```
host-1.site-A$ ssh host-2.site-A
```

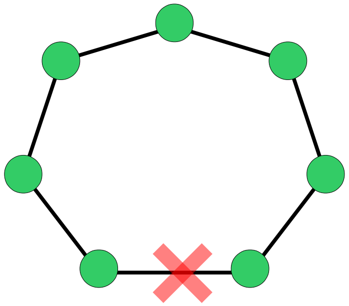
```
host-X.site-A$ ssh host-Y.site-B
```

Physical Network Topologies

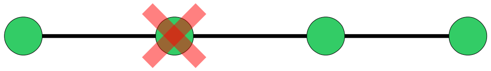
BUS



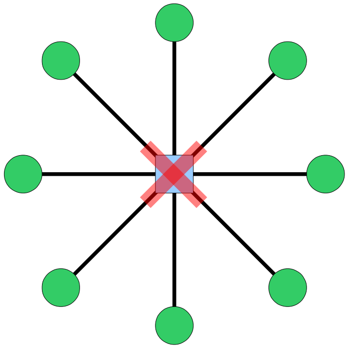
RING



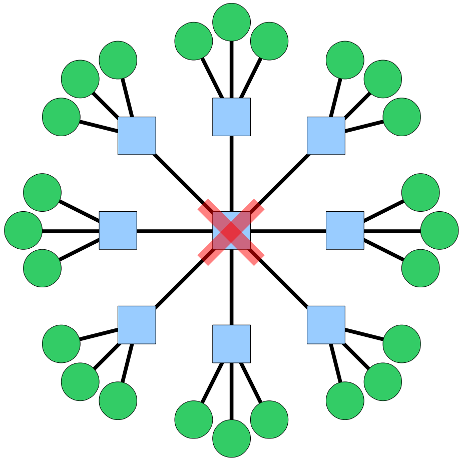
LINEAR



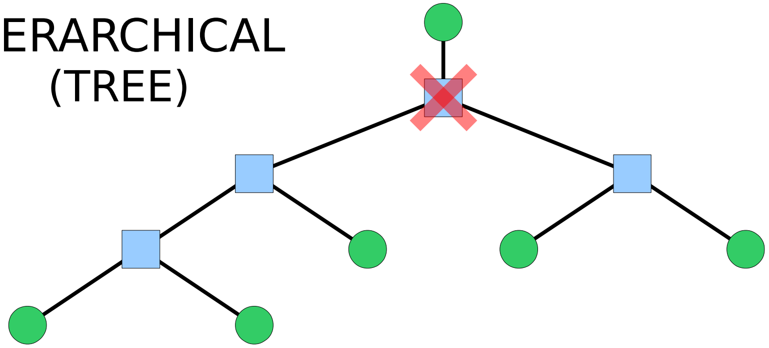
STAR



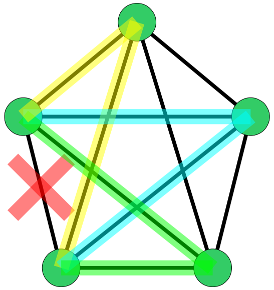
EXTENDED STAR



HIERARCHICAL (TREE)

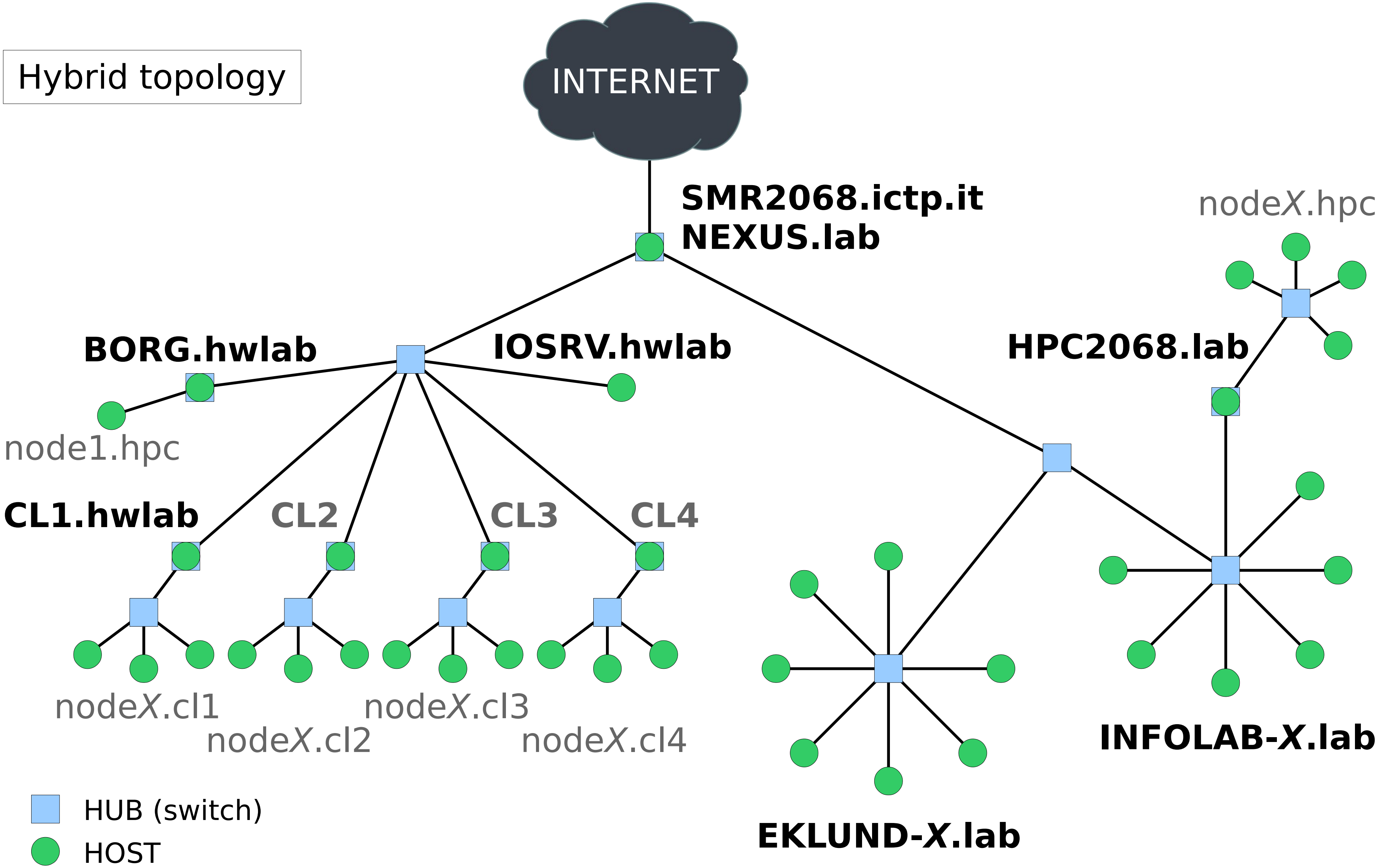


MESH (PARTIAL or FULLY CONNECTED)



Example: the lab network

Hybrid topology

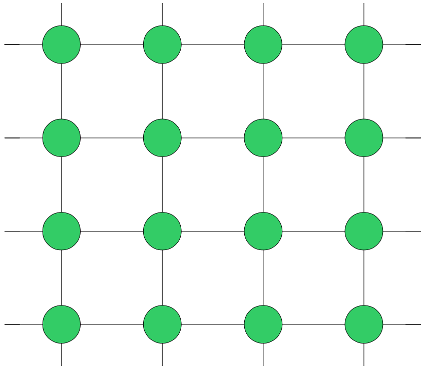


■ HUB (switch)

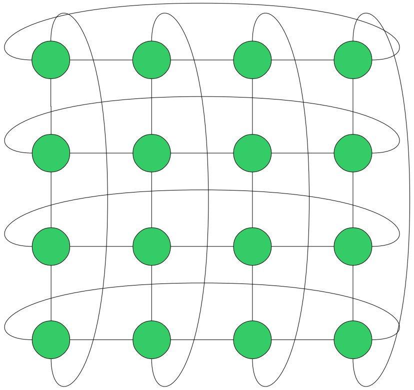
● HOST

◙ SERVER/GATEWAY

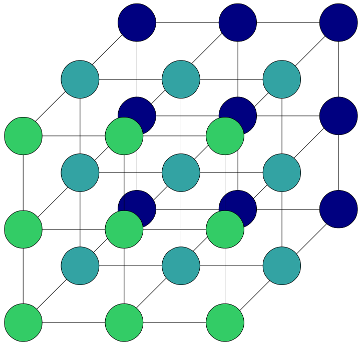
Clustering topologies (HPC)



2D Mesh



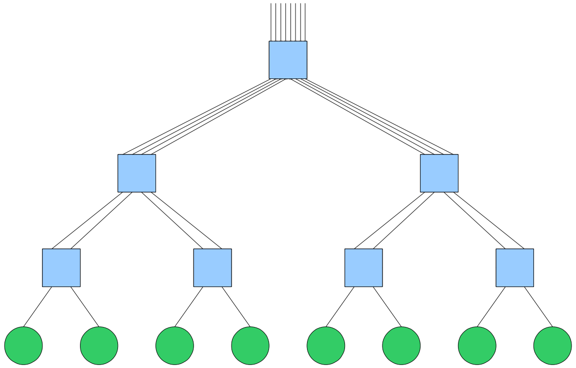
2D Torus



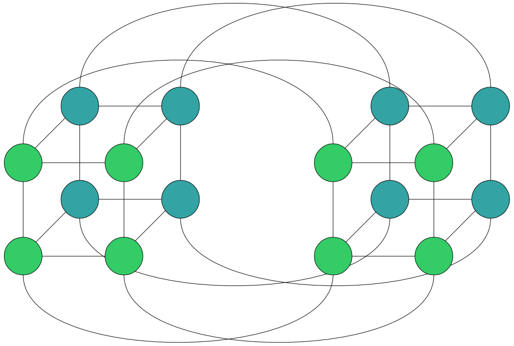
3D Mesh

Imagine a 3D mesh with the ends connected. Thanks ;)

3D Torus



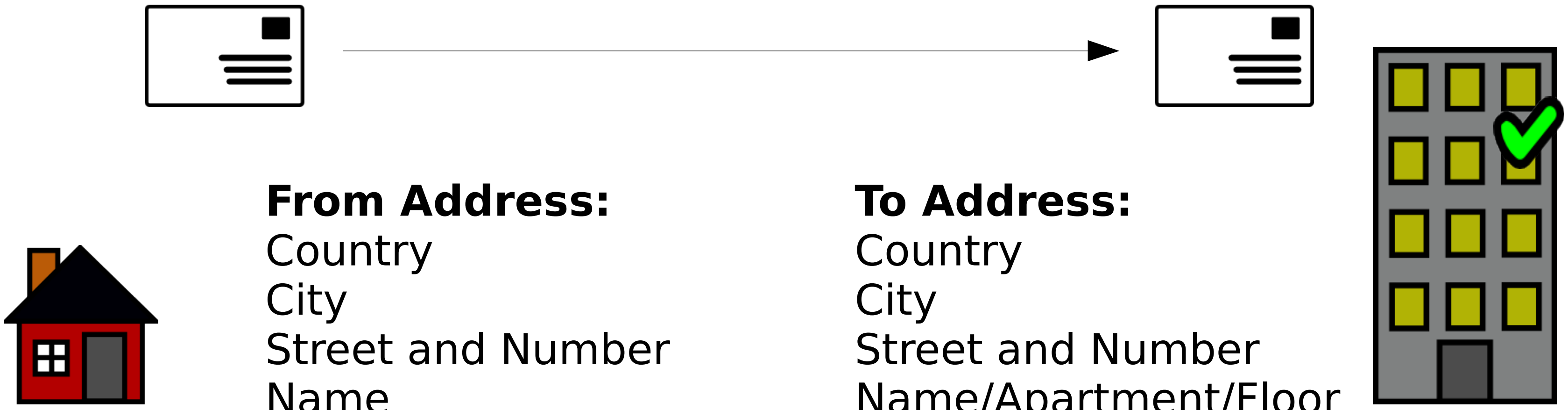
FAT TREE



Hypercube (4-cube)

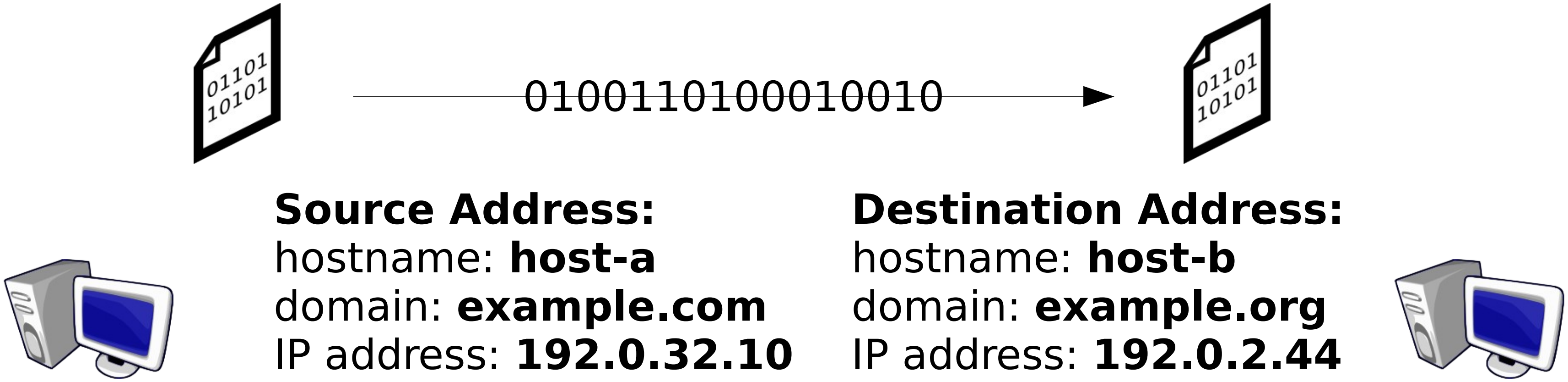
Concept of Packet

Addressing and Multiplexing



From Address:
Country
City
Street and Number
Name

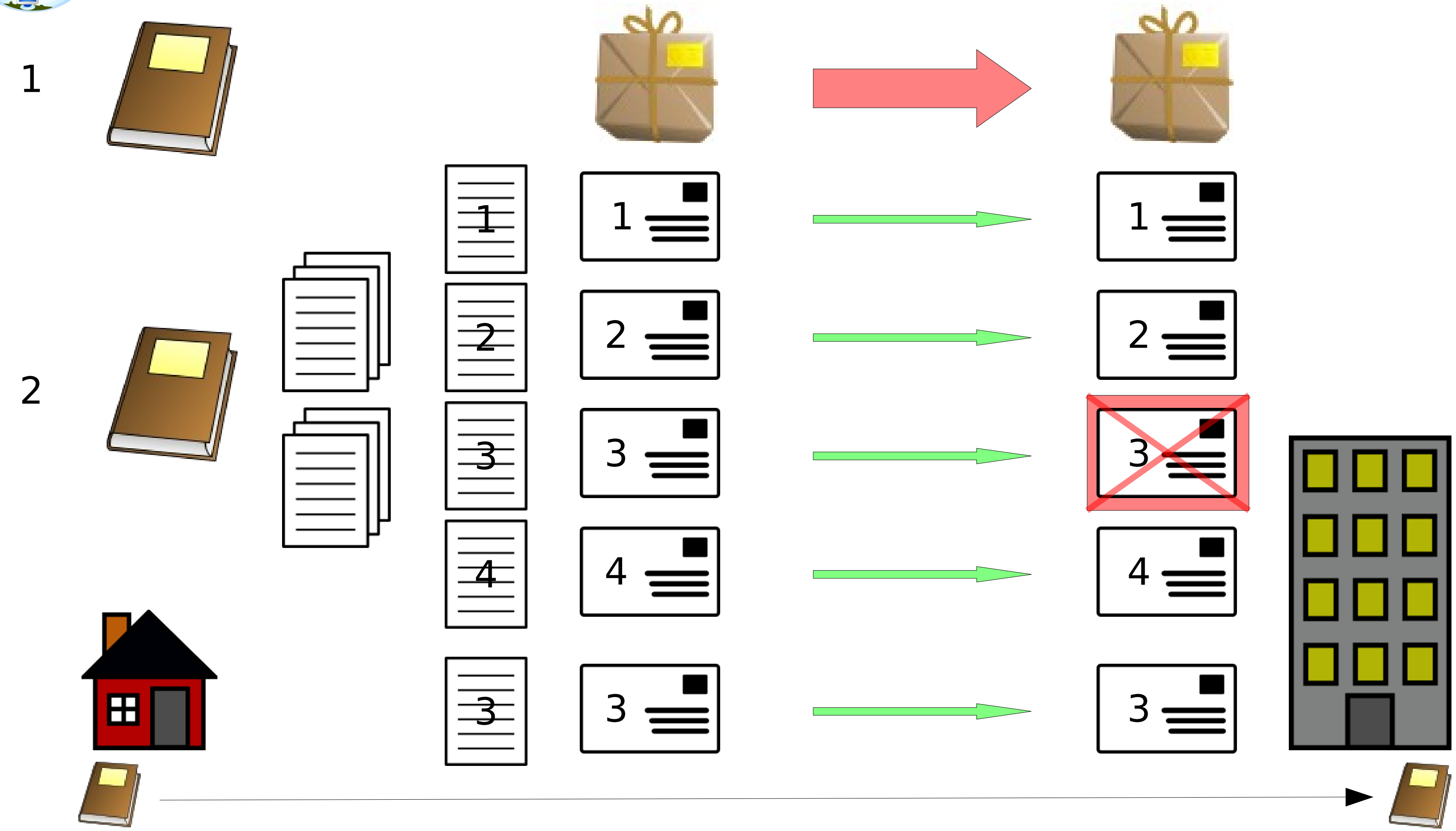
To Address:
Country
City
Street and Number
Name/Apartment/Floor



Source Address:
hostname: **host-a**
domain: **example.com**
IP address: **192.0.32.10**
protocol: **TCP**
port: **35432**

Destination Address:
hostname: **host-b**
domain: **example.org**
IP address: **192.0.2.44**
protocol: **TCP**
port: **25 (SMTP)**

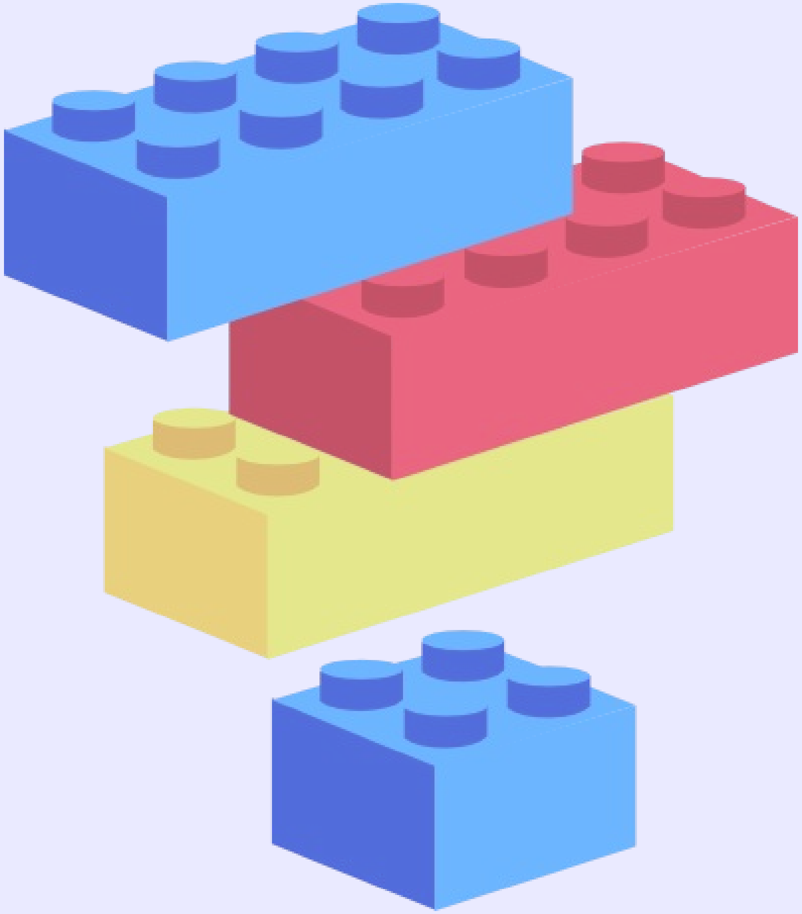
Fragmentation and Windowing



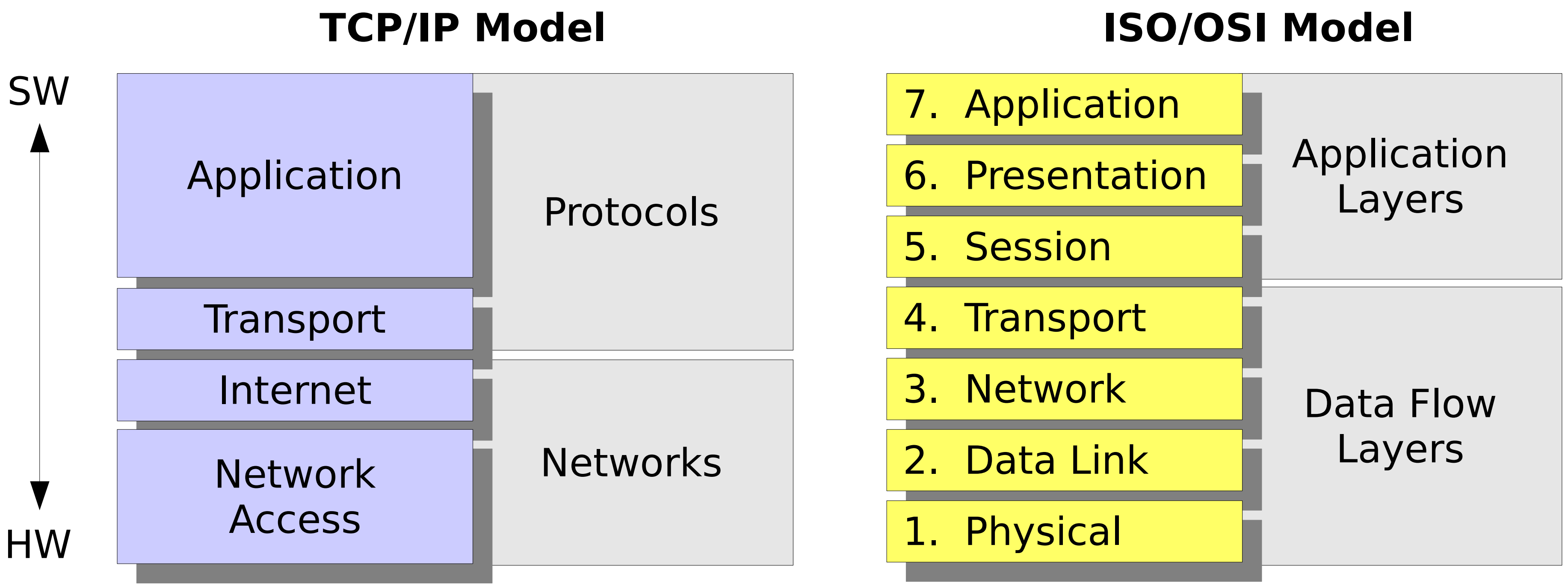
NETWORK CONNECTIONS ARE (OFTEN) NOT RELYABLE
BANDWIDTH IS NOT FREE AND IS NOT UNLIMITED

In case of failure, sending twice a large amount of data has a cost, both in terms of money and time. Network protocols splits and fragments the data stream, TCP uses sequence numbers to reassemble the data in case they reach the destination out of order (retransmission, timeout, different routes,...).

Network Stack



Network Stack Models



TCP/IP Model

SW

HW

Application

Transport

Internet

Network Access

Protocols

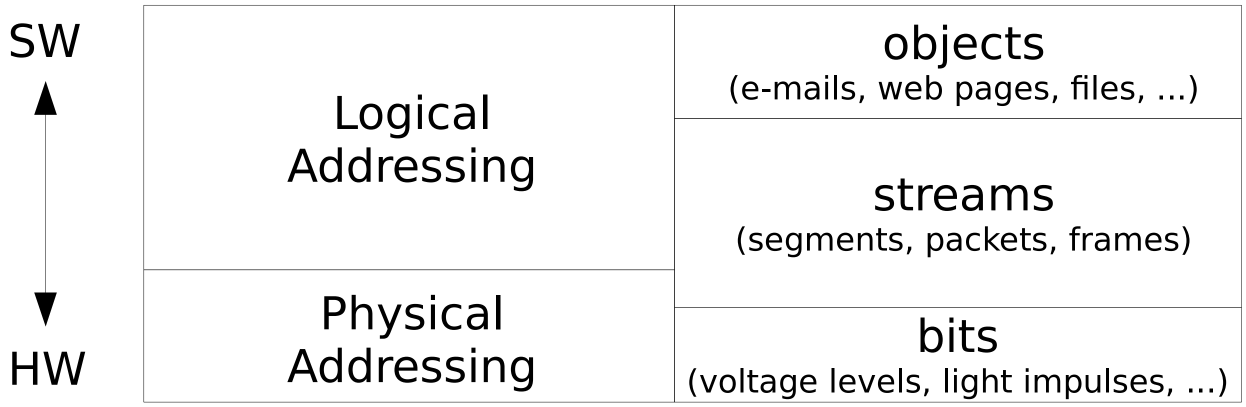
Networks

ISO/OSI Model

7. Application
6. Presentation
5. Session
4. Transport
3. Network
2. Data Link
1. Physical

Application Layers

Data Flow Layers



SW

HW

Logical Addressing

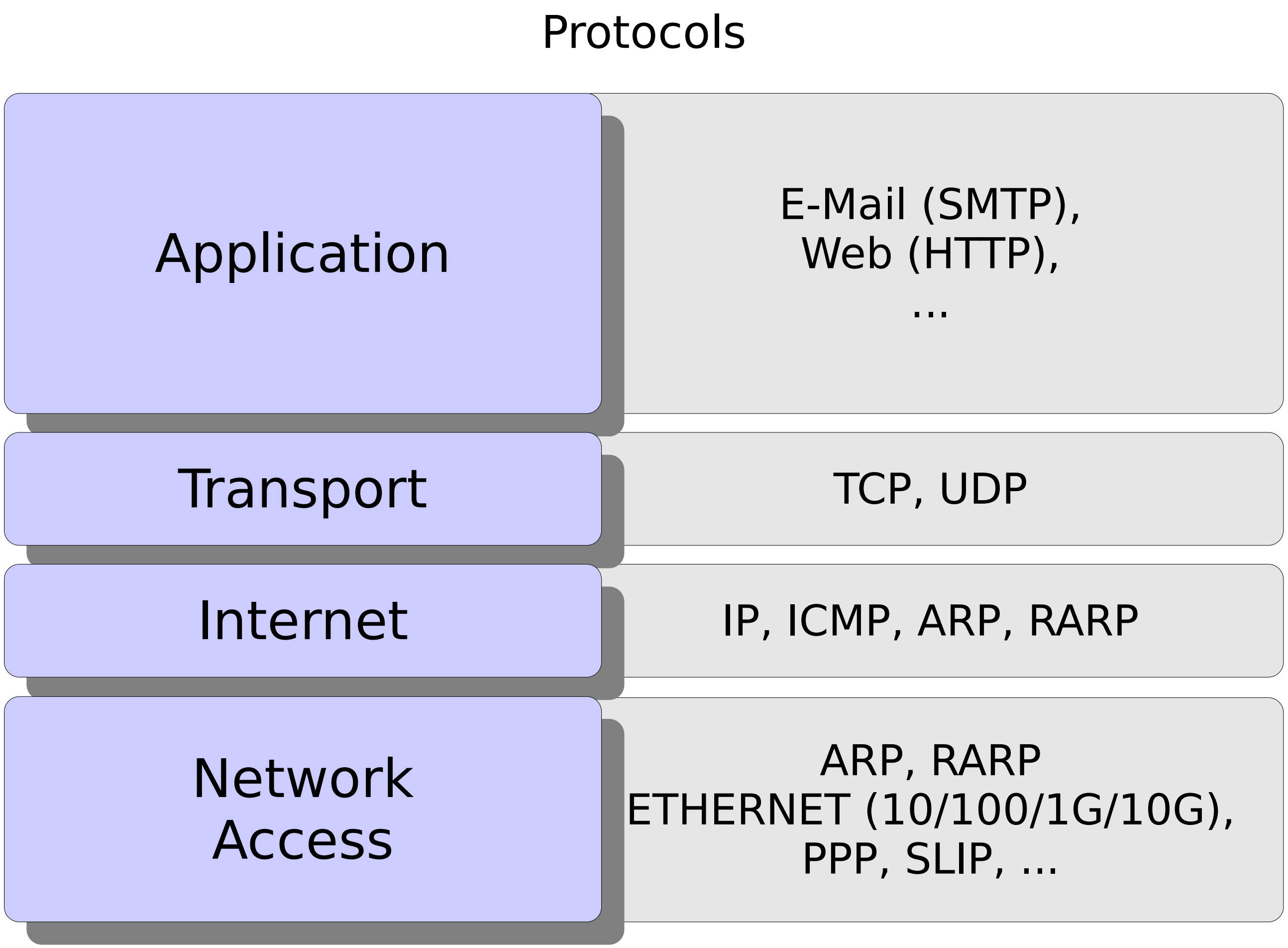
Physical Addressing

objects
(e-mails, web pages, files, ...)

streams
(segments, packets, frames)

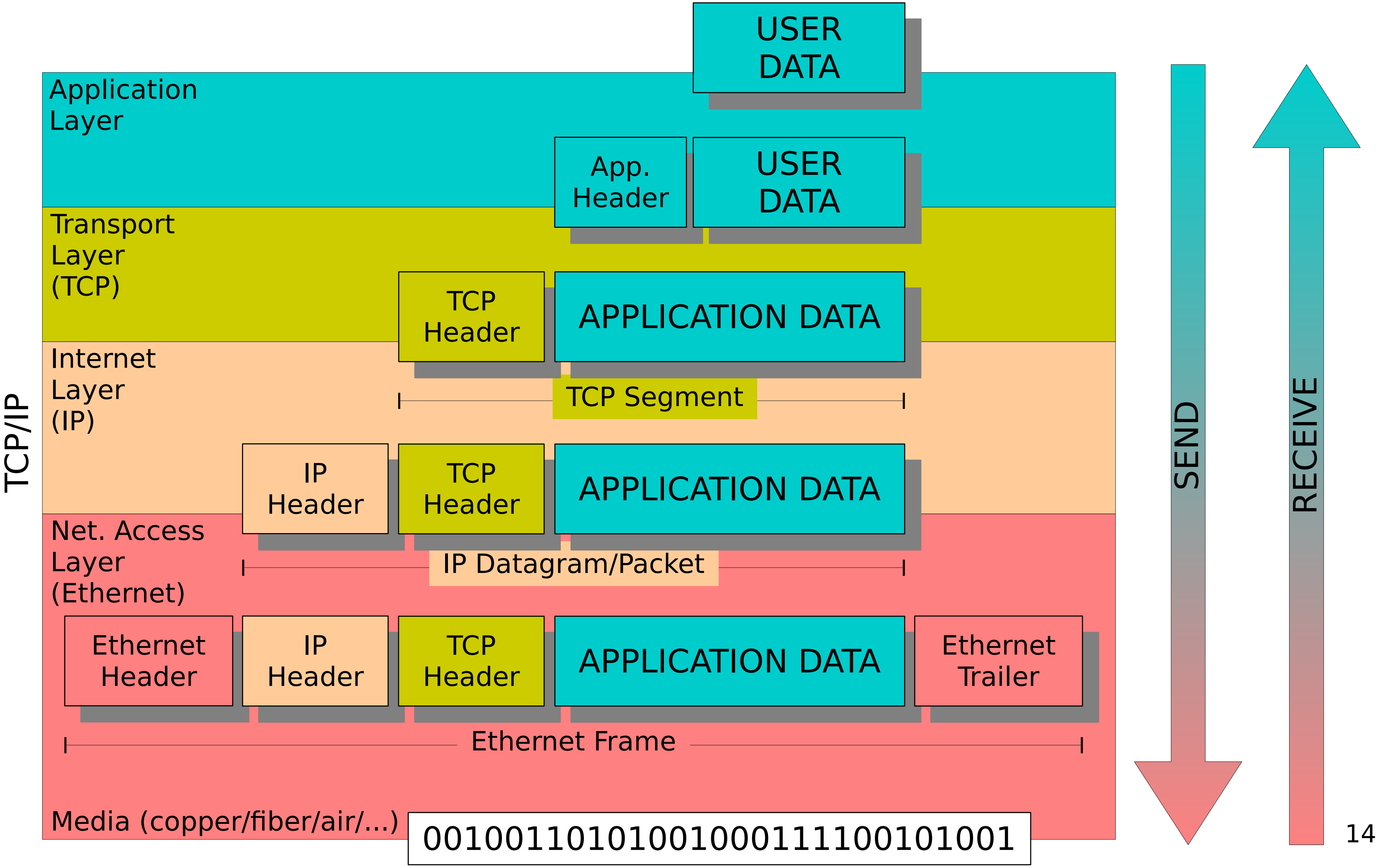
bits
(voltage levels, light impulses, ...)

TCP/IP Model

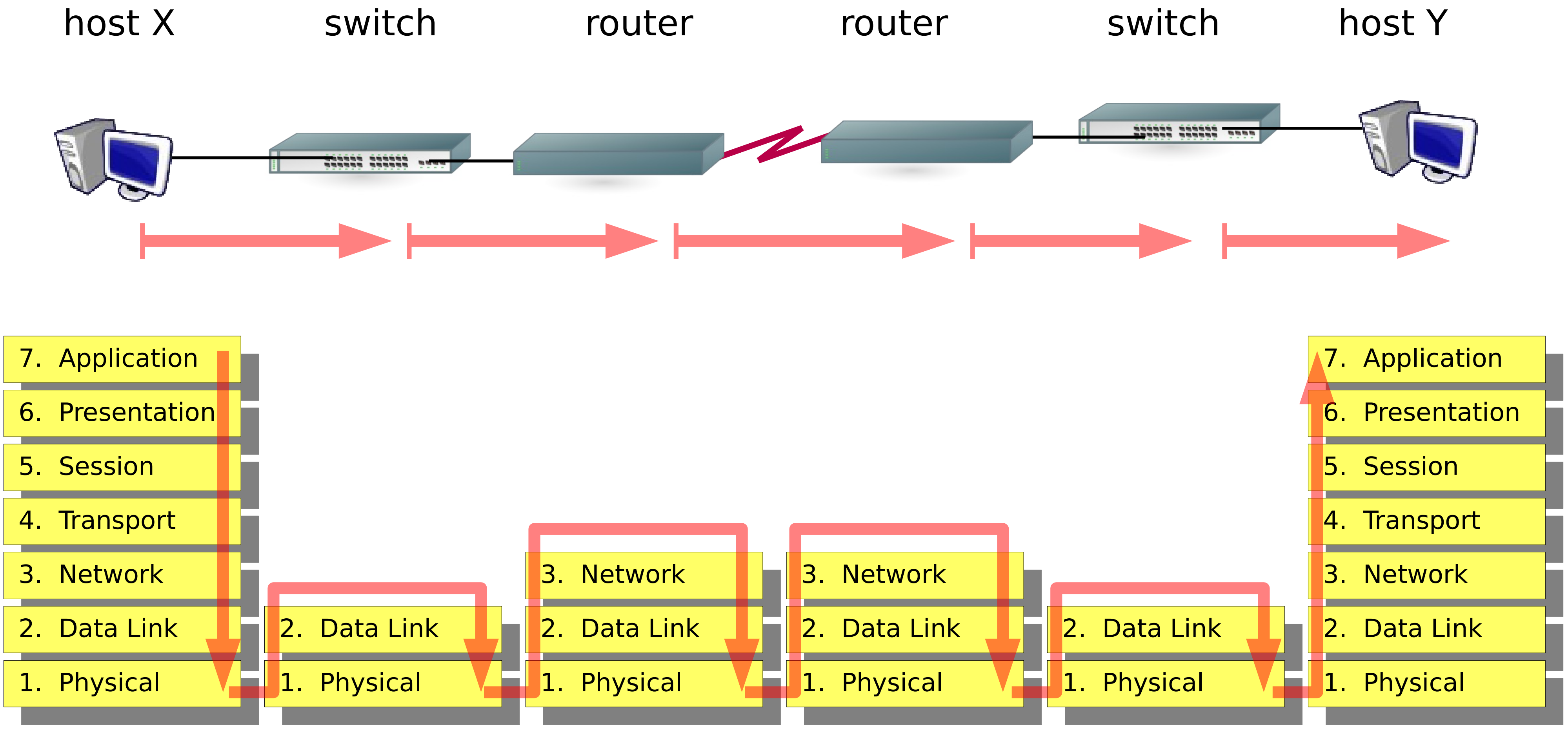


	Protocols
Application	E-Mail (SMTP), Web (HTTP), ...
Transport	TCP, UDP
Internet	IP, ICMP, ARP, RARP
Network Access	ARP, RARP ETHERNET (10/100/1G/10G), PPP, SLIP, ...

Encapsulation/De-encapsulation



Data flow

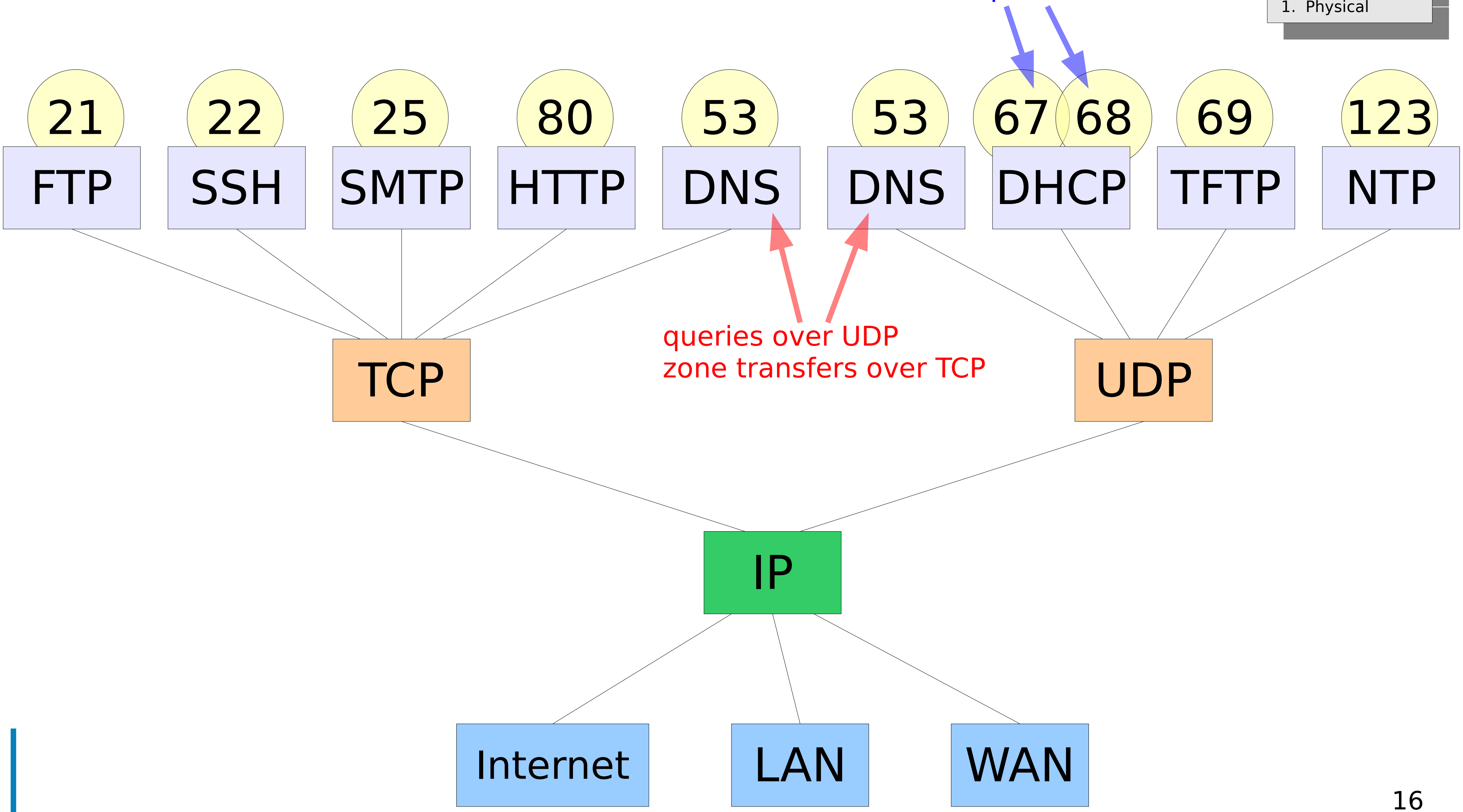


host X	switch	router	router	switch	host Y
7. Application					7. Application
6. Presentation					6. Presentation
5. Session					5. Session
4. Transport					4. Transport
3. Network		3. Network	3. Network		3. Network
2. Data Link	2. Data Link	2. Data Link	2. Data Link	2. Data Link	2. Data Link
1. Physical	1. Physical	1. Physical	1. Physical	1. Physical	1. Physical

- Switches inspect the traffic for layer 2 info (MAC)
- Routers inspect the traffic for layer 3 info (IP)

Protocols, Ports and Services

server listens on port 67
clients use port 68



queries over UDP
zone transfers over TCP

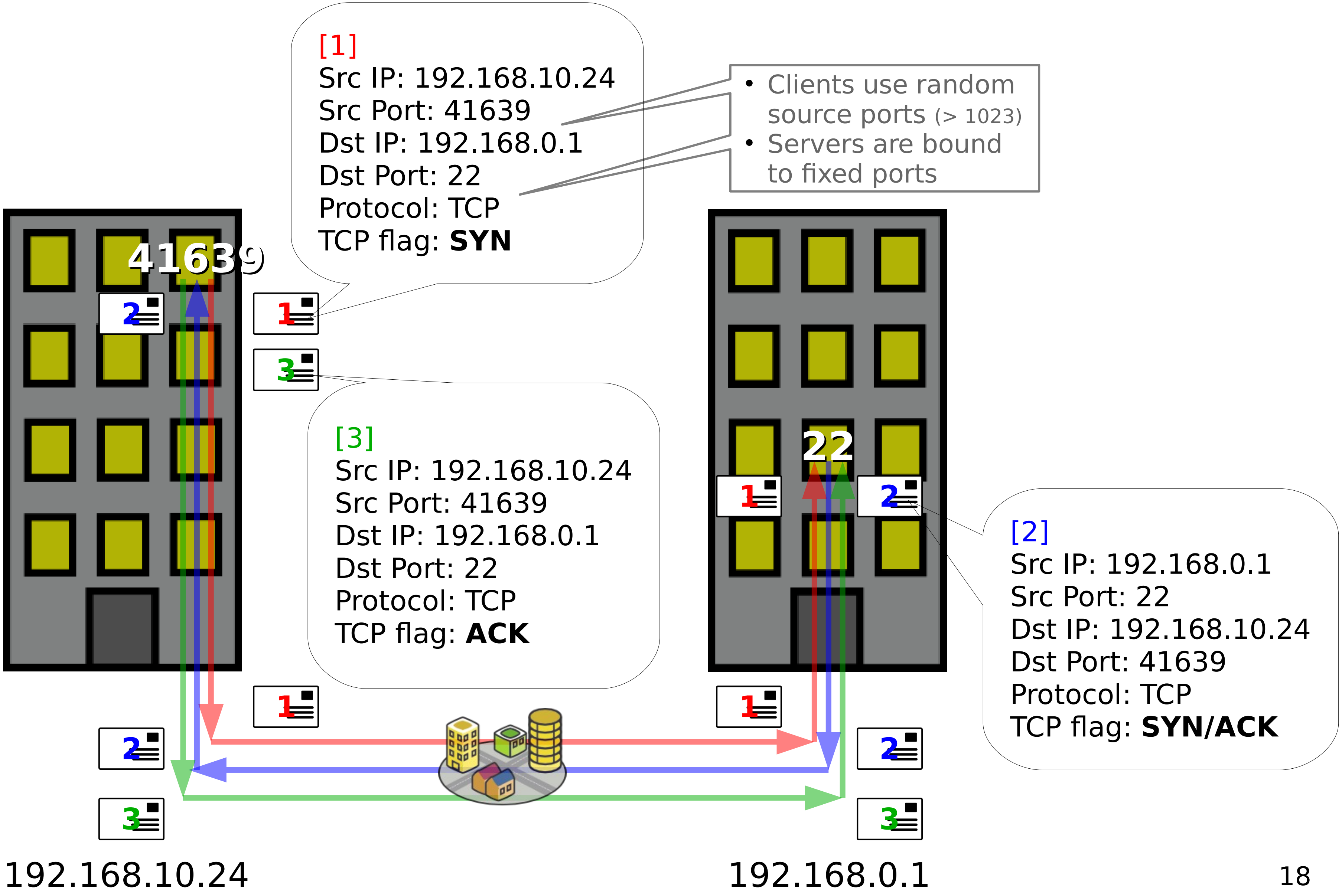
Ports

- Privileged Ports: 1-1023
  - main network services (SSH, SMTP, FTP, TFTP, DHCP, HTTP, HTTPS, ...)
  - need superuser's privileges

- Unprivileged Ports: 1024-65535
  - clients and unprivileged/no-suid services (Squid, NFS, X11, MySQL, ...)
  - any user can bind to any unprivileged port

Opening a connection

TCP 3-way Handshake



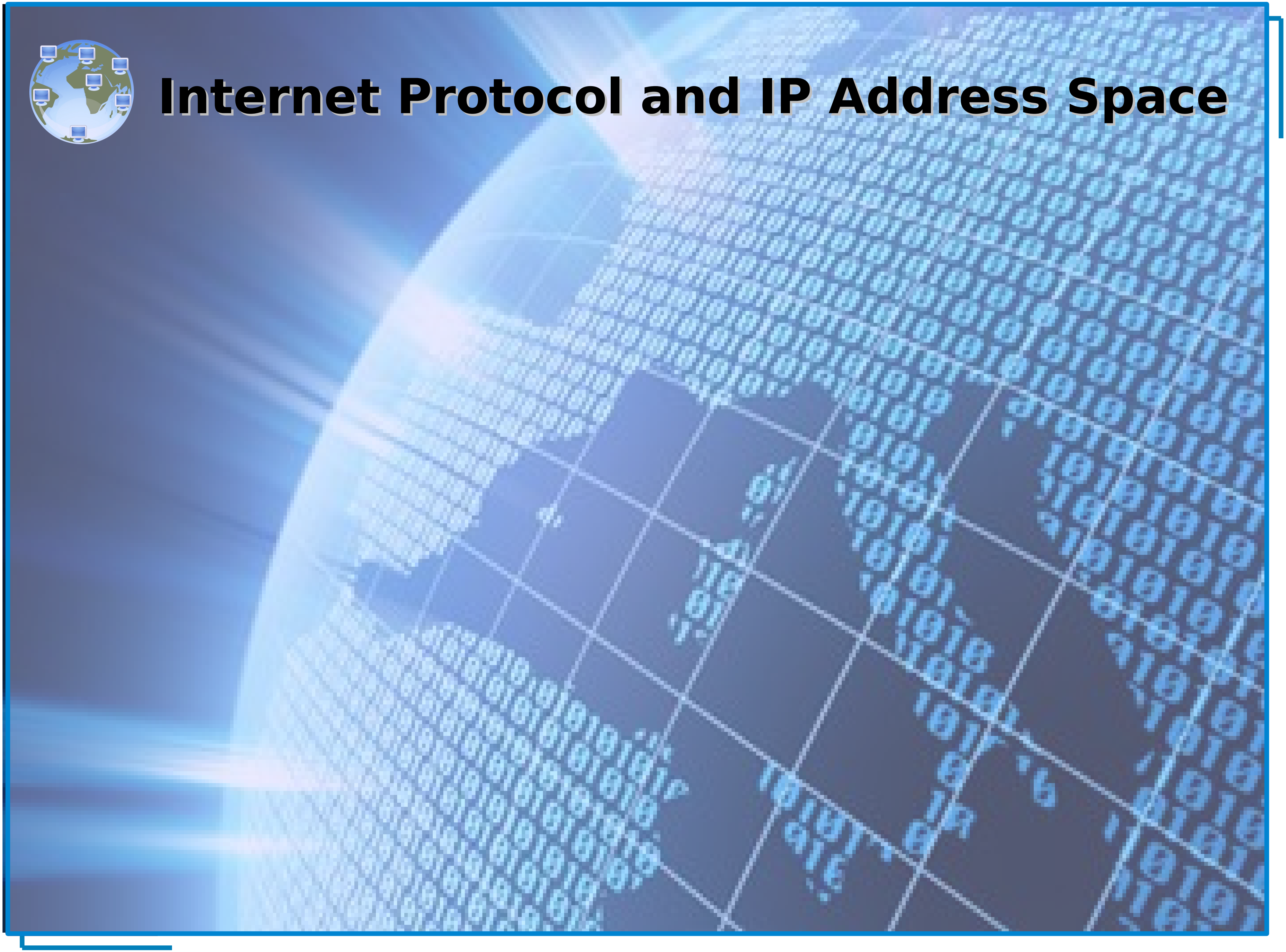
[1]
Src IP: 192.168.10.24
Src Port: 41639
Dst IP: 192.168.0.1
Dst Port: 22
Protocol: TCP
TCP flag: **SYN**

- Clients use random source ports (> 1023)
- Servers are bound to fixed ports

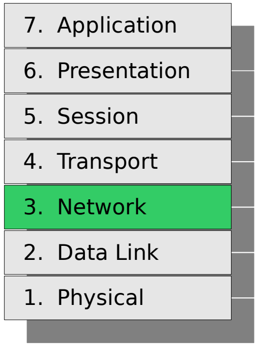
[3]
Src IP: 192.168.10.24
Src Port: 41639
Dst IP: 192.168.0.1
Dst Port: 22
Protocol: TCP
TCP flag: **ACK**

[2]
Src IP: 192.168.0.1
Src Port: 22
Dst IP: 192.168.10.24
Dst Port: 41639
Protocol: TCP
TCP flag: **SYN/ACK**

Internet Protocol and IP Address Space



Internet Protocol



7. Application
6. Presentation
5. Session
4. Transport
3. Network
2. Data Link
1. Physical

The **Internet Protocol (IP)**:

- provides network connectivity at **layer 3**
- it's a **hierarchical network-addressing scheme**
- **addresses are used to route packets** from a source to a destination through the **best available path**
- is a **connectionless, unreliable, best-effort delivery protocol** (verification handled by upper protocols)

IP(v4) addresses

The **IP address** is:

something like this: **10.1.2.3**

- a **numerical label** which **uniquely identify each host on a network**
- logically divided in two parts, the *network* portion and the *host* portion
- obtained by the ISP (public IPs) or the system/network administrator (private IPs)
- **assigned** to a host **statically or dynamically** (BOOTP/DHCP)
- a 32 bits / 4 bytes unsigned integer number, usually **represented in a dotted-decimal notation**, as four 8bits/1byte numbers (0-255), called “octets”, separated by a dot '.'

Netmask, Network and Broadcast

The **network address**:

- **identifies the network itself**
- **defines the group of IP addresses that belongs to the same broadcast domain**, hosts that can communicate with each other without the need of a layer 3 device
- is an IP address with the **host portion filled by 0s** (**10.1.2.0**)

The **netmask address** is:

- **a bit-mask of contiguous 1s** (starting from the MSB) that **separates the host portion from the network portion** of an IP address (1s on the network portion, 0s on the host portion)
- often represented in the “slash format” as the total number of bits used for the network and subnetwork portion of the mask (/8, /16, /24, /32, ...)
- something like this: **255.255.255.0**

The **broadcast address** is:

- a network address that **allows information to be sent to all nodes on a network**, rather than to a specific network host (unicast)
- an IP address with the **host portion filled by 1s** (**10.1.2.255**)

IP Address Notation

- *Dotted Quad Notation* (*four-octet dotted-decimal, numbers-and-dots*)
  - 10.240.27.73 / 255.255.255.0 (10.240.27.73/24)
- Hexadecimal Notation
  - 0AF01B49 / FFFFFF00
- Binary Notation
  - 00001010 11110000 00011011 01001001 / 11111111 11111111 11111111 00000000

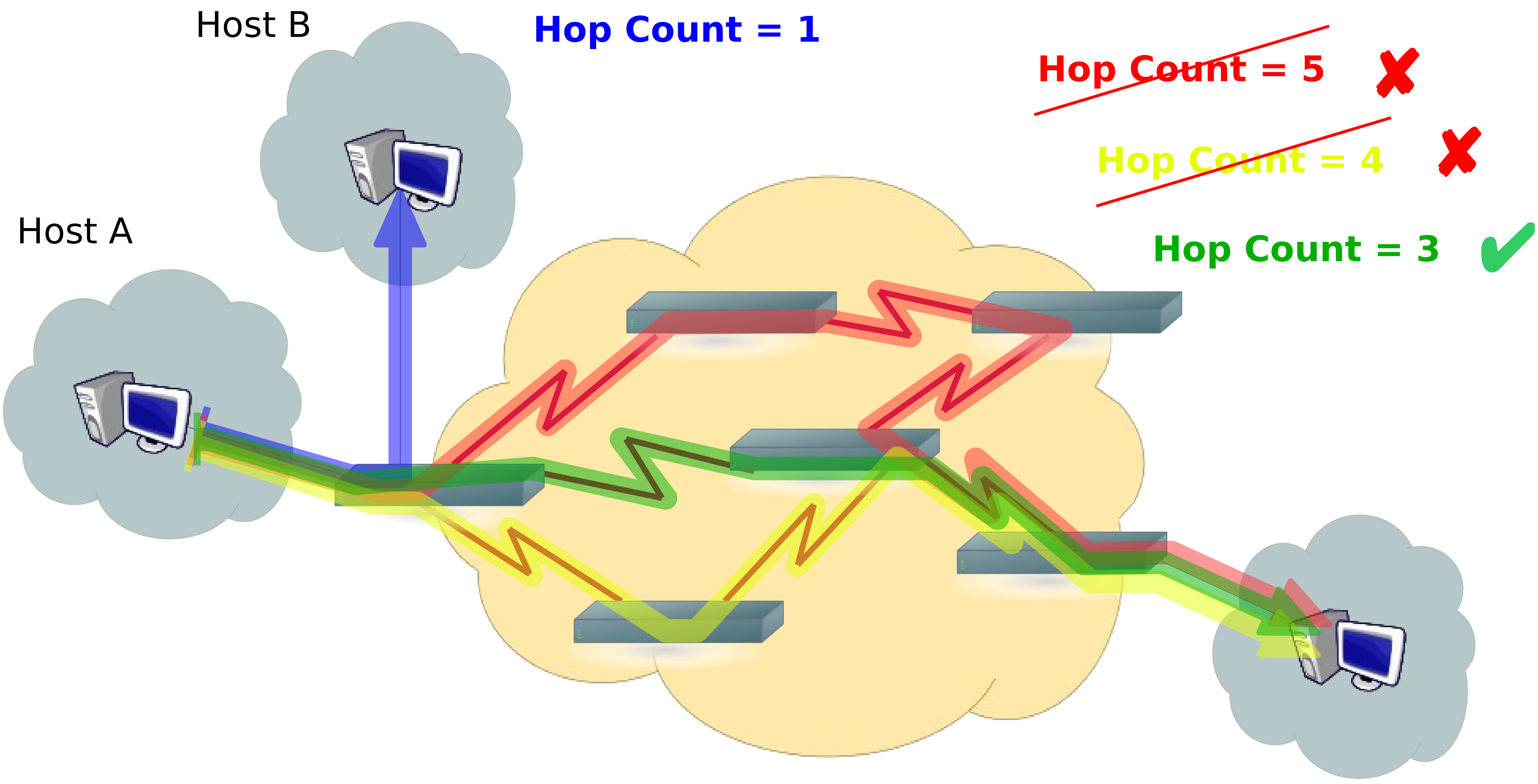
11111111 11111111 11111111	**00000000**	**FFFFFF**	**00**	**255.255.255.**	**0**	Netmask
00001010 11110000 00011011	01001001	0AF01B	49	10.240. 27.	73	IP Addr.
00001010 11110000 00011011	00000000	0AF01B	00	10.240. 27.	0	Network Addr.
00001010 11110000 00011011	11111111	0AF01B	FF	10.240. 27.	255	Broadcast Addr.

NETWORK PORTION **HOST PORTION**

Routing

- **routers** are layer 3 devices that **use the IP address to move data packets between networks**
- when packets arrive at an interface, the router uses the **routing table** to determine where to send them
- each router that the packet encounters along the way is called a **hop**, the **hop count** is the distance traveled
- routing **metrics** are used to determine the **best path (hop count, load, bandwidth, delay, cost, and reliability of a network link)**

Best path determination



Host B

Hop Count = 1

~~Hop Count = 5~~ ✘

~~Hop Count = 4~~ ✘

Hop Count = 3 ✔

Host A

Host C

Host A -> Host B	cost = 1
Host A -> Host C	cost = 3

Reserved IP Addresses

RFC 3330
RFC 1918
RFC 2606

“This” network:	0.0.0.0/8
Loopback:	127.0.0.0/8
Private addresses: 10.0.0.0 172.16.0.0 192.168.0.0 10.255.255.255 172.31.255.255 192.168.255.255	10.0.0.0/8 172.16.0.0/12 192.168.0.0/16
“TEST-NET” (example.com, org, net):	192.0.2.0/24
6to4 Relay:	192.88.99.0/24
“Link local” (zeroconf):	169.254.0.0/16
Multicast:	224.0.0.0/4

Host names, Domain names and DNS

- **hostname**
  - **cerbero**.hpc.sissa.it
- **first level domain**
  - cerbero.hpc.sissa.**it**
- **second level domain**
  - cerbero.hpc.**sissa**.it
- **third level domain**
  - cerbero.**hpc**.sissa.it
- ***Fully Qualified Domain Name*** **(FQDN)**
  - **cerbero.hpc.sissa.it**
- **DNS**
  - cerbero.hpc.sissa.it **-->** 147.122.17.62
  - 147.122.17.62 **-->** cerbero.hpc.sissa.it

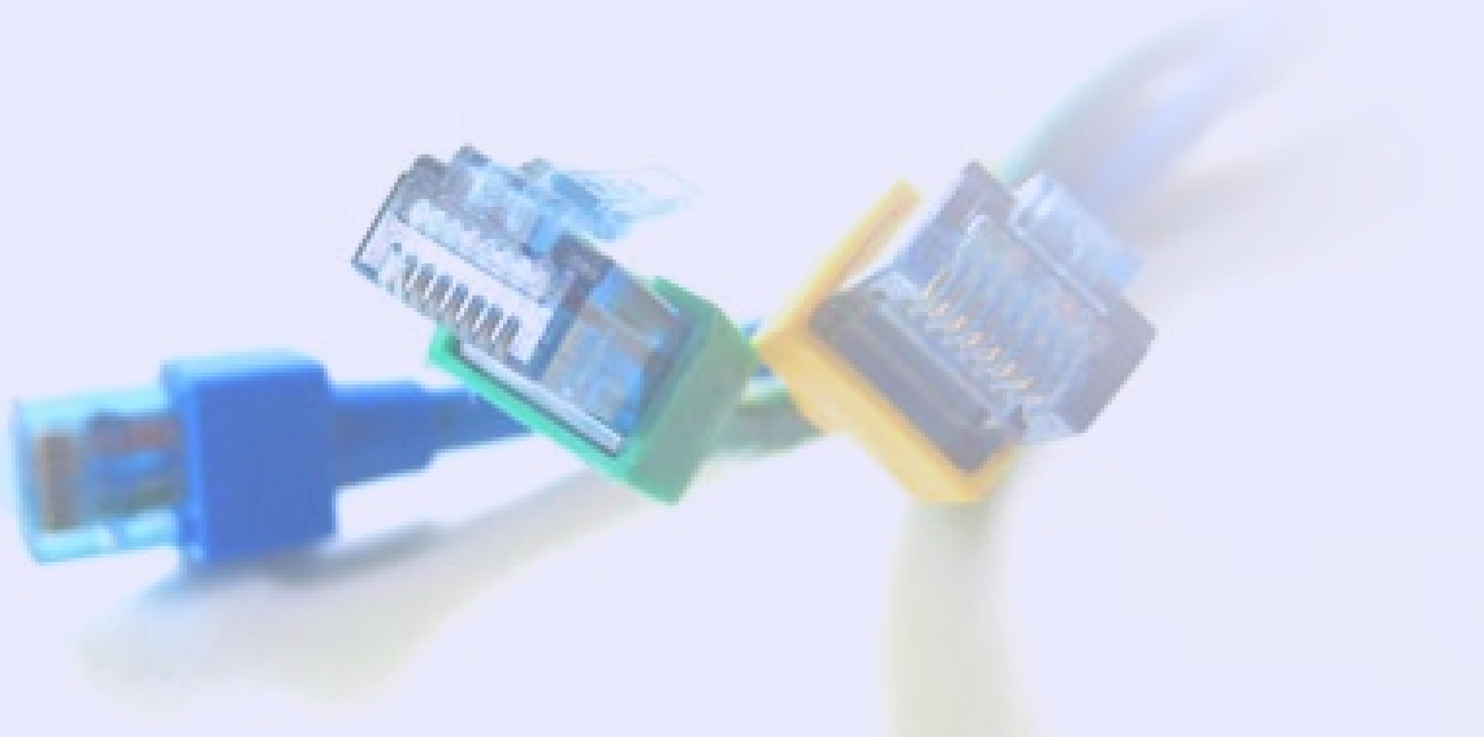
Static vs. Dynamic IP assignment

- **static**: manual configuration (servers, network devices, workstations)

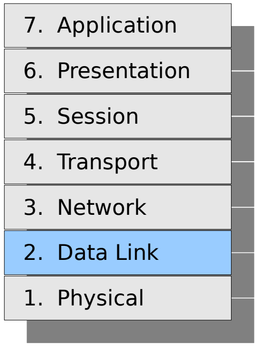
- **dynamic**: the DHCP server assigns an IP address to each DHCP client, associating the MAC address to an IP.
  The IP address can be:
  - **randomly assigned from a pool of IPs** (laptops on a wireless network or a LAN)
  - **sticky**, as above but the lease time is set to long periods (ISP)
  - **fixed** (workstations, network devices, cluster nodes, any device that must be always reachable at the same address), **requires individual profile for each device** (maps MAC-IP, providing Network Settings and, optionally, hostnames)

- **autoconfiguration (*link-local*)***:* communication between hosts on a single link (LAN segment) or a point-to-point connection

Ethernet and Physical Address



MAC Address

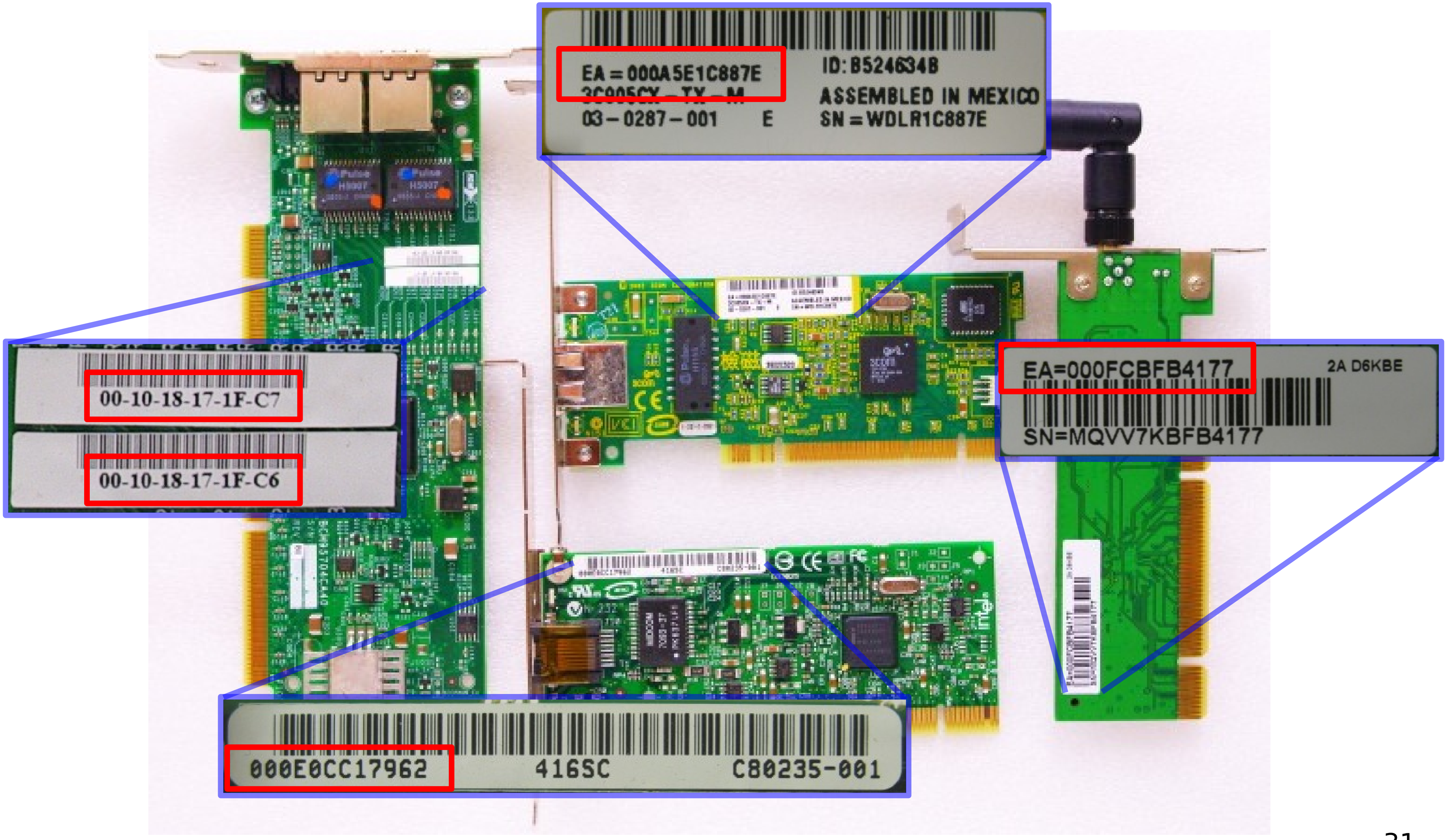

7. Application
6. Presentation
5. Session
4. Transport
3. Network
2. Data Link
1. Physical

The ***Media Access Control*** **Address** is:

- a **physical address**, **globally unique**
- **assigned by the manufacturer** of the NIC and **burned-in into the PROM of the NIC** (in some cases, can be administratively assigned)
- part of the Ethernet protocol and **operates at Layer 2**
- **used by DHCP to dynamically assign IP Addresses**
- a 48bits number represented as a 6 groups of two hexadecimal digits (6 bytes) separated by ':', made of two parts, 3 bytes each:
  - the OUI (Organizationally Unique Identifier)
  - the production number

`00:0e:0c:d7:3b:25`

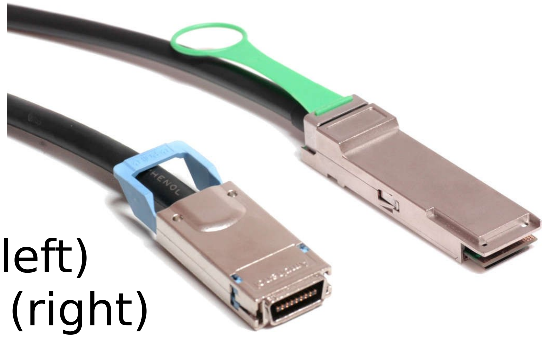
MAC Address


EA = 000A5E1C887E
3C905CX - TX - M
03 - 0287 - 001 E
ID: B524634B
ASSEMBLED IN MEXICO
SN = WDLR1C887E
00-10-18-17-1F-C7
00-10-18-17-1F-C6
EA=000FCBFB4177
2A D6KBE
SN=MQVV7KBFB4177
000E0CC17962
416SC
C80235-001

Cables and connectors

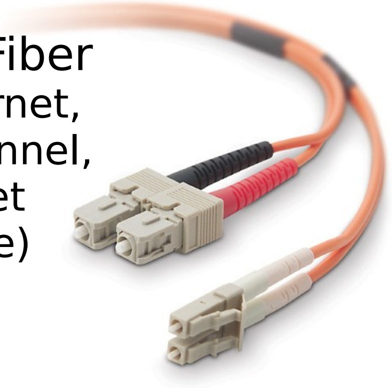
- **bandwidth varies depending upon the type of media as well as the technologies used,** the physics of the media account for some of the difference
- signals travel through twisted-pair copper wire, coaxial cable, optical fiber, and air
- **the physical differences in the ways signals travel result in fundamental limitations on the information-carrying capacity of a given medium**
- **actual bandwidth of a network is determined by a combination of the physical media and the technologies chosen for signaling and detecting network signals**.

Ethernet RJ45
(10/100/1000)



CX4 (left)
QSFP (right)
(Infiniband & 10GB Ethernet)

SC / LC Fiber
(*G Ethernet,
Fiber Channel,
Myrinet
& more)



Wrap up

- network topologies
- fragmentation
- network stacks and protocols
- (de)encapsulation
- ports (multiplexing) and services
- IP address space, DNS, routing
- physical address (MAC) and hardware

That's All Folks!

“Network is down.”

```
( questions ; comments ) | mail -s uheilaaa baro@democritos.it

( complaints ; insults ) &>/dev/null
```

REFERENCES AND USEFUL LINKS

SOFTWARE:

▪ Linux Kernel	http://www.kernel.org
▪ Netfilter	http://www.netfilter.org
▪ nmap	http://www.insecure.org/nmap/
▪ hping	http://www.hping.org/
▪ netcat	http://netcat.sourceforge.net/
▪ iptstate	http://www.phildev.net/iptstate/
▪ ss	http://linux-net.osdl.org/index.php/Iproute2
▪ lsof	ftp://lsof.itap.purdue.edu/pub/tools/unix/lsof/
▪ netstat	http://www.tazenda.demon.co.uk/phil/net-tools/
▪ tcpdump	http://www.tcpdump.org
▪ wireshark	http://www.wireshark.org
▪ ethereal	http://www.ethereal.com (see wireshark)
▪ iptraf	http://iptraf.seul.org/
▪ ettercap	http://ettercap.sourceforge.net
▪ dsniff	http://www.monkey.org/~dugsong/dsniff/
▪ tcptraceroute	http://michael.toren.net/code/tcptraceroute/
▪ (telnet, traceroute, ping, ...)	

DOC:

- IPTables HOWTO http://www.netfilter.org/documentation/HOWTO/
- IPTables tutorial http://iptables-tutorial.frozentux.net/
- Having fun with IPTables
  http://www.ex-parrot.com/~pete/upside-down-ternet.html
- Denial of Service http://www.cert.org/tech_tips/denial_of_service.html
- IPv4 Address space
  - http://www.cymru.com/Documents/bogon-bn.html
  - http://www.iana.org/assignments/ipv4-address-space
  - http://www.oav.net/mirrors/cidr.html
  - http://en.wikipedia.org/wiki/IPv4
  - IANA http://www.iana.org
  - RIPE http://www.ripe.net
  - RFC 3330 http://www.rfc.net/rfc3330.html
- SANS: http://www.sans.org/reading_room/whitepapers/firewalls/
  http://www.sans.org/reading_room/

RFC: (http://www.rfc.net)

- RFC 791 – Internet Protocol (IPv4)
  http://www.rfc.net/rfc791.html
- RFC 793 – Transmission Control Protocol (TCP)
  http://www.rfc.net/rfc793.html
- RFC 768 – User Datagram Protocol (UDP)
  http://www.rfc.net/rfc768.html
- RFC 792 – Internet Control Message Protocol (ICMP)
  http://www.rfc.net/rfc792.html
- RFC 1180 – A TCP/IP Tutorial
  http://www.rfc.net/rfc1180.html
- RFC 1700 / IANA db – Assigned Numbers
  http://www.rfc.net/rfc1700.html
  http://www.iana.org/numbers.html
- RFC 3330 – Special-Use IPv4 Addresses
  http://www.rfc.net/rfc3330.html
- RFC 1918 – Address Allocation for Private Internets
  http://www.rfc.net/rfc1918.html
- RFC 2196 – Site Security Handbook
  http://www.rfc.net/rfc2196.html
- RFC 2827 – Network Ingress Filtering
  http://www.rfc.net/rfc2827.html
- RFC 2828 – Internet Security Glossary
  http://www.rfc.net/rfc2828.html
- RFC 1149 – Transmission of IP Datagrams on Avian Carriers
  http://www.rfc.net/rfc1149.html
- Unofficial CPIP WG
  http://www.blug.linux.no/rfc1149/
- RFC 2549 – IP over Avian Carriers with Quality of Service
  http://www.rfc.net/rfc2549.html
- Firewalling the CPIP
  http://www.tibonia.net/
  http://www.hotink.com/wacky/dastrdly/

Some acronyms...

IP – Internet Protocol
TCP – Transmission Control Protocol
UDP – User Datagram Protocol
ICMP – Internet Control Message Protocol
ARP – Address Resolution Protocol
MAC – Media Access Control

OS – Operating System
NOS – Network Operating System
LINUX – LINUX is not UNIX

PING – Packet Internet Groper

FTP – File Transfer Protocol – (TCP/21,20)
SSH – Secure SHell – (TCP/22)
TELNET – Telnet – (TCP/23)
SMTP – Simple Mail Transfer Protocol – (TCP/25)
DNS – Domain Name System – (UDP/53)
NTP – Network Time Protocol – (UDP/123)
BOOTPS – Bootstrap Protocol Server (**DHCP**) – (UDP/67)
BOOTPC – Bootstrap Protocol Server (**DHCP**) – (UDP/68)
TFTP – Trivial File Transfer Protocol – (UDP/69)
HTTP – HyperText Transfer Protocol – (TCP/80)
NTP – Network Time Protocol – (UDP/123)
SNMP – Simple Network Management Protocol – (UDP/161)
HTTPS – HyperText Transfer Protocol over TLS/SSL – (TCP/443)
RSH – Remote Shell – (TCP/514,544)

ISO – International Organization for Standardization
OSI – Open System Interconnection

TLS – Transport Layer Security
SSL – Secure Sockets Layer

RFC – Request For Comments

ACL – Access Control List

PDU – Protocol Data Unit

TCP flags:

- **URG**: Urgent Pointer field significant
- **ACK**: Acknowledgment field significant
- **PSH**: Push Function
- **RST**: Reset the connection
- **SYN**: Synchronize sequence numbers
- **FIN**: No more data from sender

RFC 3168 TCP flags:

- **ECN**: Explicit Congestion Notification
- (**ECE**: ECN Echo)
- **CWR**: Congestion Window Reduced

ISN – Initial Sequence Number