Redes de Computadores

LEIC-A

5 – Data Link Layer

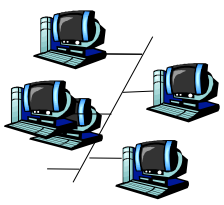
Prof. Paulo Lobato Correia

IST, DEEC – Área Científica de Telecomunicações

1

Outline

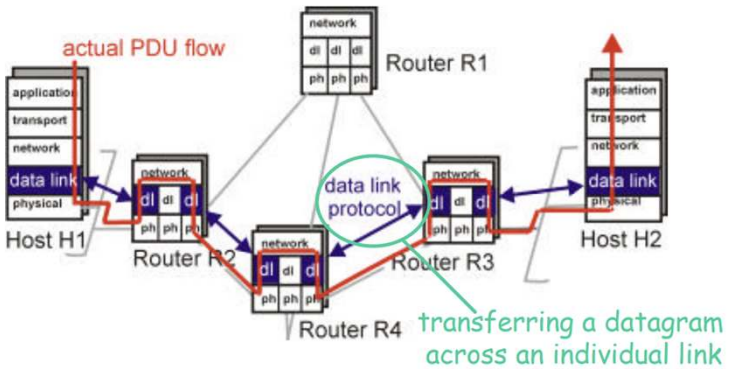
- Introduction and services
- Link-layer addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing



2

Data Link Layer: Context

- Datagram transferred using **different link protocols** on **different links**:
  - e.g., Ethernet on first link, optical fiber intermediate link, WiFi on last link;
- Each link protocol provides different services:
  - e.g., may or may not provide reliable data transfer over link.



3

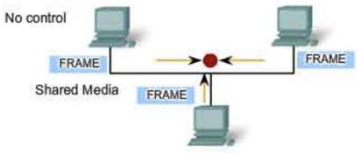
Data Link Layer

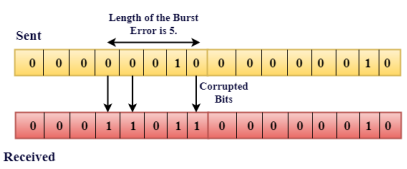
Application
Transport
Network
Data link
Physical

The link layer deals with data transfer between devices belonging to the **same subnet**.

Data link layer tasks:

- Framing, Link Access / Media Access Control
- Reliable Delivery, Error Detection and Correction





4

Data Link Layer Services

- *Framing, Link access :*
  - Encapsulate datagram into frame, adding header + trailer;
  - **Medium access control** (MAC) protocol - to access shared medium;
  - "MAC" addresses used in frame headers to identify source, destination:
    - Different from IP address!

- *Reliable delivery between adjacent nodes:*
  - Similar solutions to those adopted in transport layer;
  - Seldom used on low bit-error links (fiber, some twisted pair);
  - Wireless links: high error rates.

Q: Why both link-level and end-end reliability?

Data Link Layer Services

- *Error detection*:
  - Errors caused by signal attenuation and noise;
  - Receiver may detect presence of errors:
    - Signals sender for *retransmission* or *drop*s frame.

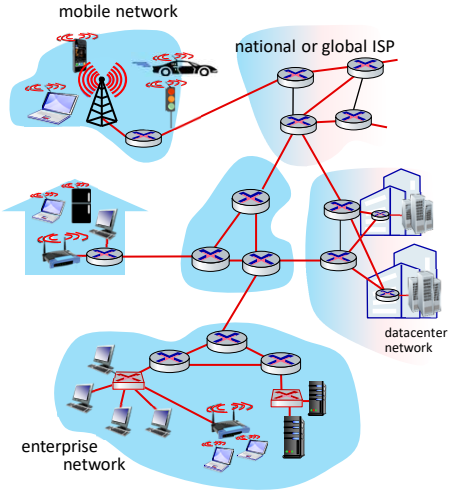
- *Error correction*:
  - Receiver may identify *and correct* bit error(s) without resorting to retransmission.
- *Flow control:*
  - To adjust pace between adjacent sending and receiving nodes;
- *Half-duplex and full-duplex links:*
  - With half duplex, nodes at both ends of link can transmit, but not at same time.

TÉCNICO LISBOA

Data Link Layer: Introduction

Some terminology:

- Hosts and routers are **nodes;**
- Communication channels that connect adjacent nodes along communication path are **links:**
  - Wired links;
  - Wireless links;
  - LANs.
- Layer-2 packet is a **frame.** It encapsulates a datagram.

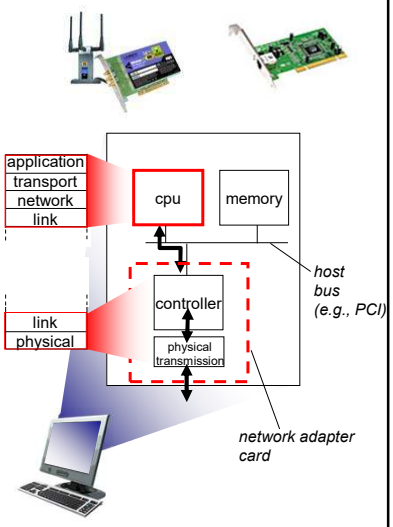


*Data link layer has the responsibility of **transferring datagrams from one node to the physically adjacent node** over a link*

TÉCNICO LISBOA

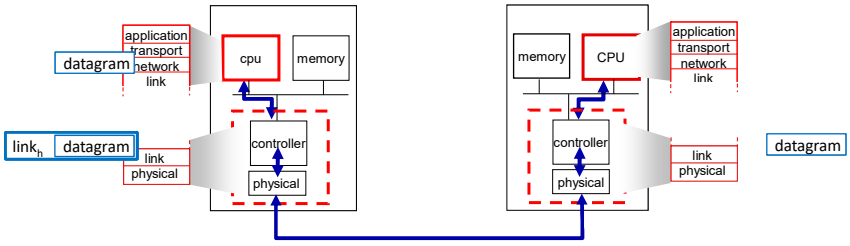
Where is the Link Layer Implemented?

- In each and every host;
- Link layer implemented in the "adaptor" (aka *network interface card* - NIC):
  - Ethernet card, PCMCI card, 802.11 card;
  - Implements link, physical layer.
- Attaches into host's system buses;
- Combination of hardware, software, firmware.



TÉCNICO LISBOA

Interfaces Communicating



Sending side:

- encapsulates datagram in frame
- adds error checking bits, reliable data transfer, flow control, etc.

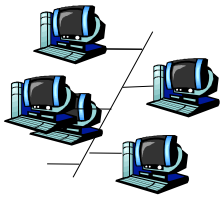
Receiving side:

- looks for errors, reliable data transfer, flow control, etc.
- extracts datagram, passes to upper layer at receiving side

TÉCNICO LISBOA

Outline

- Introduction and services
- Link-layer Addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- IEEE 802.11 Wireless LANs
- Link-layer switches
- Framing

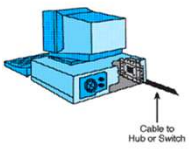


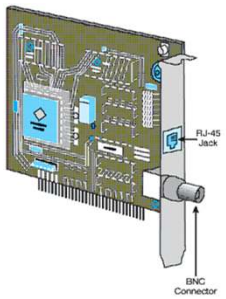
Link-Layer Addressing

- Network Layer: IP address
  - Used to get datagram to the destination IP subnet;
  - 32-bit IPv4 address;
  - **Hierarchical**;
  - Not portable.
- Data Link Layer: MAC (or LAN or **physical**) address
  - Function: *get frame from one interface to another physically-connected interface (within the same network);*
  - 48 bit MAC address (for most LANs):
    - Burned in NIC ROM; Sometimes adjustable by software;
  - **Portable**;

MAC Addresses

- Analogy:
  - (a) *MAC* address is like the "*Cartão de Cidadão*" Number;
  - (b) *IP* address is like the *postal address*;
- MAC flat address → portability:
  - Can move network interface card from one LAN to another;
- IP hierarchical address is NOT portable:
  - Address depends on IP subnet to which the node is currently attached.

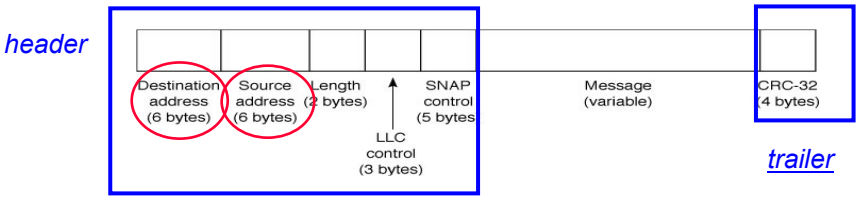




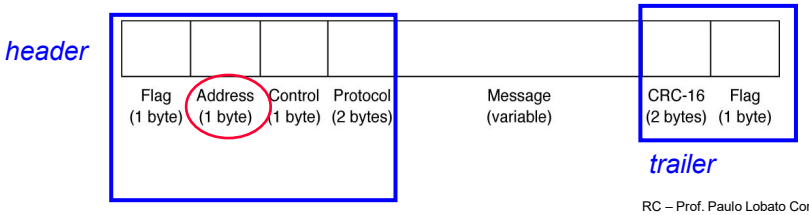
Link-Layer Addressing

Examples:

- A **Ethernet** LAN uses 48 bit MAC addresses:

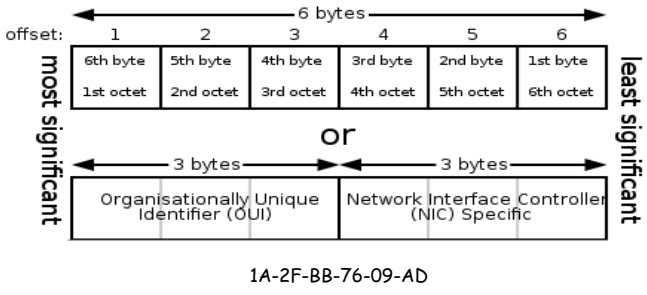


- A **PPP** connection (typical of dial-up accesses) uses 8 bit addresses:



MAC Addresses

- MAC address allocation is administered by IEEE Registration Authority;
- A manufacturer buys portion of MAC address space (to assure uniqueness);
- Number of available addresses: 2^{48} = 281 474 976 710 656
- MAC Address (Ethernet) – 48 bits: **MM-MM-MM-SS-SS-SS**

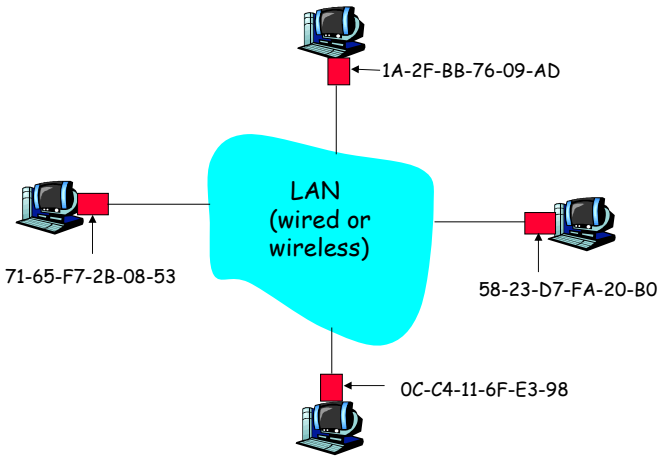


1A-2F-BB-76-09-AD

TÉCNICO LISBOA

MAC Addresses

Each adapter on a LAN has a unique MAC address



15

TÉCNICO LISBOA

Addressing

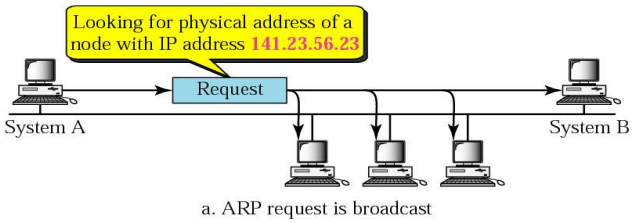
```
> ipconfig /all
```

```
C:\>ipconfig /all

Windows IP Configuration

        Host Name . . . . . . . . . . . . : magalhaes
        Primary Dns Suffix  . . . . . . . :
        Node Type . . . . . . . . . . . . : Unknown
        IP Routing Enabled. . . . . . . . : No
        WINS Proxy Enabled. . . . . . . . : No

Ethernet adapter Wireless Network Connection:

        Media State . . . . . . . . . . . : Media disconnected
        Description . . . . . . . . . . . : Intel(R) PRO/Wireless LAN 2100 3B Mini PCI Adapter
        Physical Address. . . . . . . . . : 00-04-23-5F-CA-08

Ethernet adapter Local Area Connection:

        Media State . . . . . . . . . . . : Media disconnected
        Description . . . . . . . . . . . : Realtek RTL8139/810x Family Fast Ethernet NIC
        Physical Address. . . . . . . . . : 00-0C-6E-76-8B-D0
```

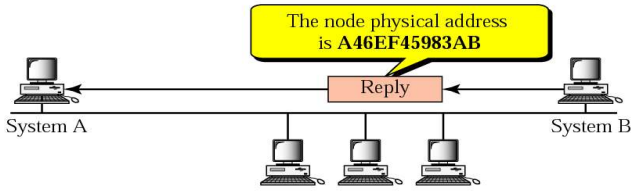

16

Address Resolution Protocol (ARP)

To deliver frames the physical (MAC) address of the host needs to be known.
ARP allows to get the MAC address from the IP address.



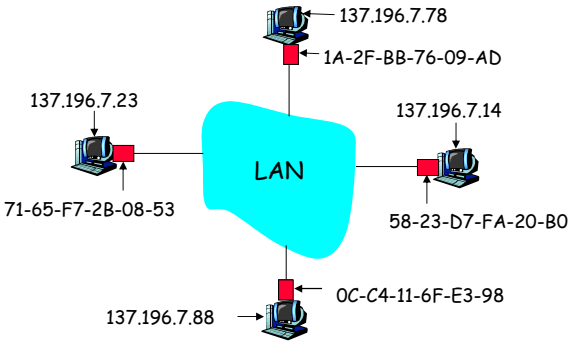
a. ARP request is broadcast



b. ARP reply is unicast

17

Address Resolution Protocol (ARP)



- Each IP node (host, router) on a LAN keeps an ARP table;

ARP table:

- IP/MAC address mappings for some LAN nodes:

  < IP address; MAC address; TTL>

  - TTL (*Time To Live*): time after which address mapping will be forgotten (typically 20 min).

18

TÉCNICO LISBOA

ARP Protocol (used in the same LAN)

Example:

- A wants to send datagram to B, and B's MAC address not in A's ARP table.
- A broadcasts **ARP query packet**, containing B's IP address:
  - Dest. MAC address = FF-FF-FF-FF-FF-FF
  - All machines on LAN receive the ARP query.
- **B** receives ARP packet and **replies to A** with its (B's) MAC address:
  - Frame sent to A's MAC address (unicast).
- A caches (saves) IP-to-MAC address pair in its ARP table until information becomes old (times out):
  - Soft state: information that times out (goes away) unless refreshed;

ARP is "plug-and-play":

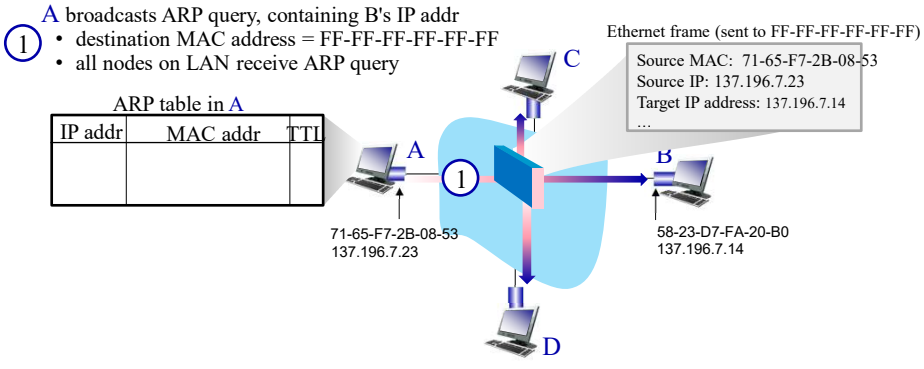
- Nodes create their ARP tables *without intervention from net administrator.*

TÉCNICO LISBOA

ARP protocol in action

example: A wants to send datagram to B

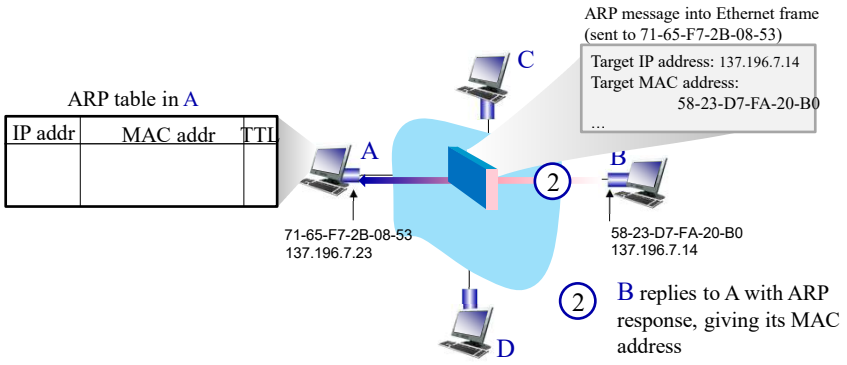
- B's MAC address not in A's ARP table, so A uses ARP to find B's MAC address



ARP protocol in action

example: A wants to send datagram to B

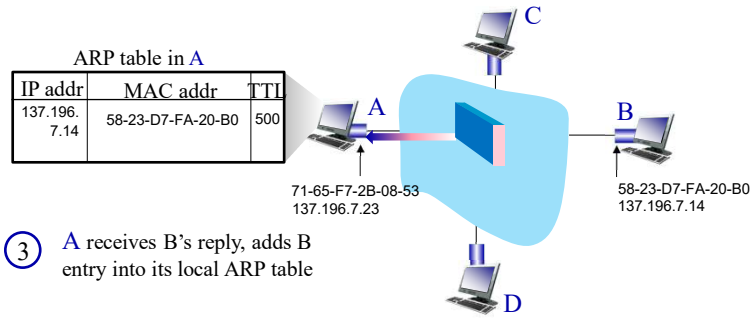
- B's MAC address not in A's ARP table, so A uses ARP to find B's MAC address



ARP protocol in action

example: A wants to send datagram to B

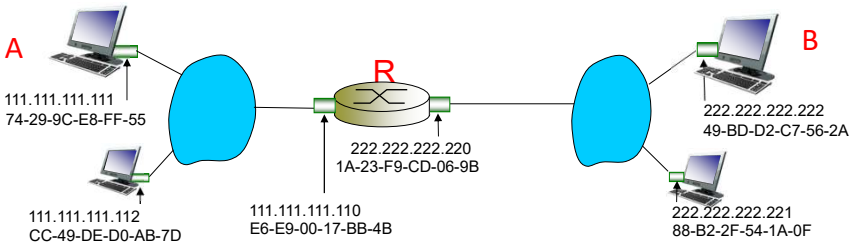
- B's MAC address not in A's ARP table, so A uses ARP to find B's MAC address



Addressing: Routing to Another LAN

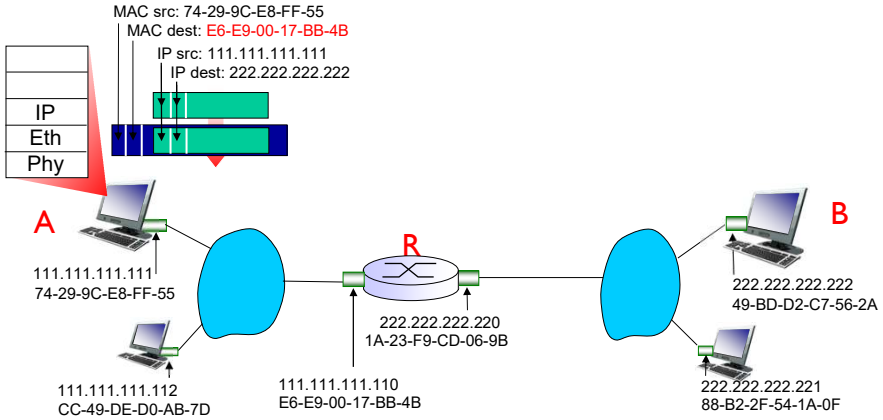
Send datagram from **A** to **B** via **R**:

- assume A knows B's IP address
- assume A knows IP address of first hop router, R (how?)
- assume A knows R's MAC address (how?)



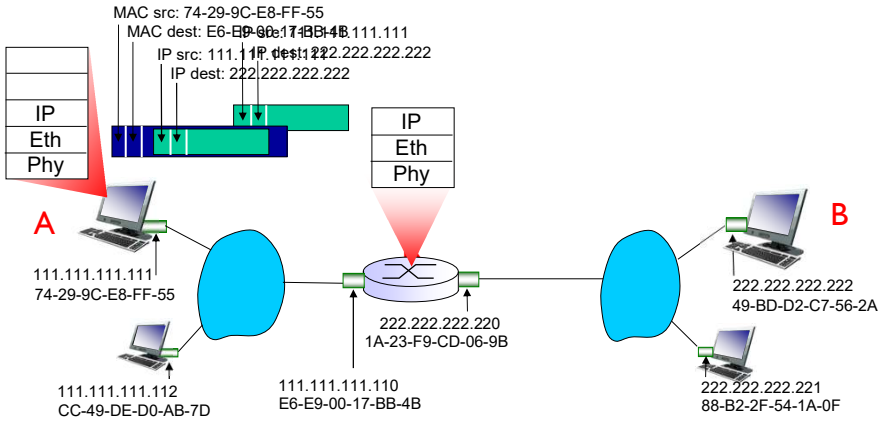
Addressing: Routing to Another LAN

- A – creates IP datagram with: *source IP A*, *destination IP B*
- A – creates **link-layer frame** with: destination *R's MAC address*, data field: *A-to-B IP datagram*



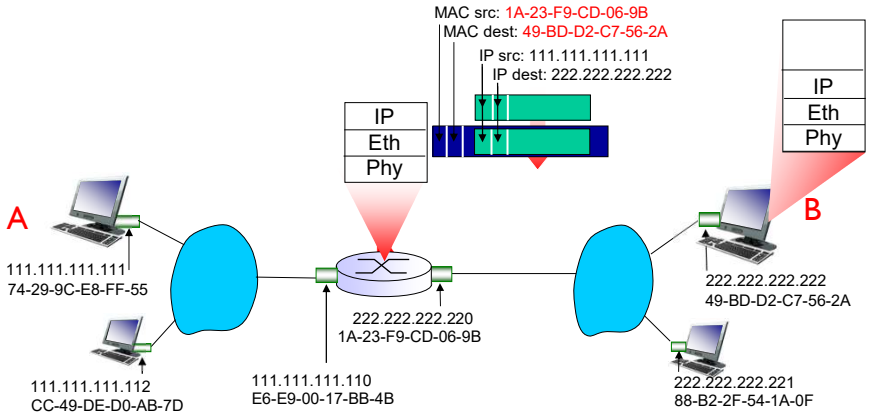
Addressing: Routing to Another LAN

- frame sent from A to R
- frame received at R, datagram removed, passed up to IP



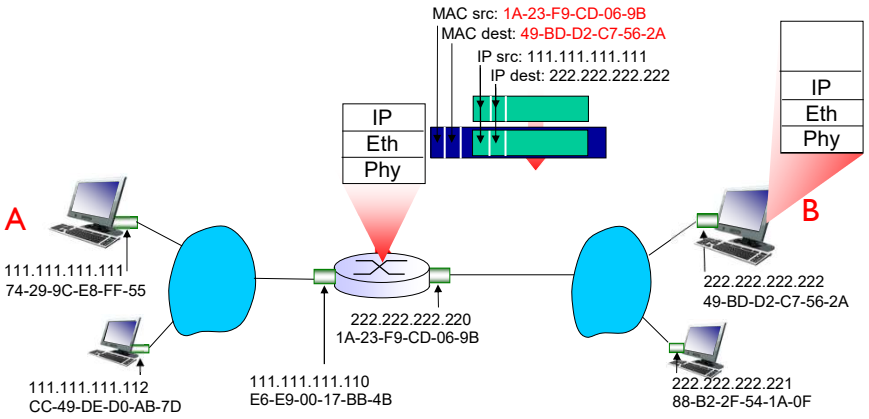
Addressing: Routing to Another LAN

- R forwards datagram with IP source A, destination B
- R creates link-layer frame with B's MAC address as dest, frame contains A-to-B IP datagram



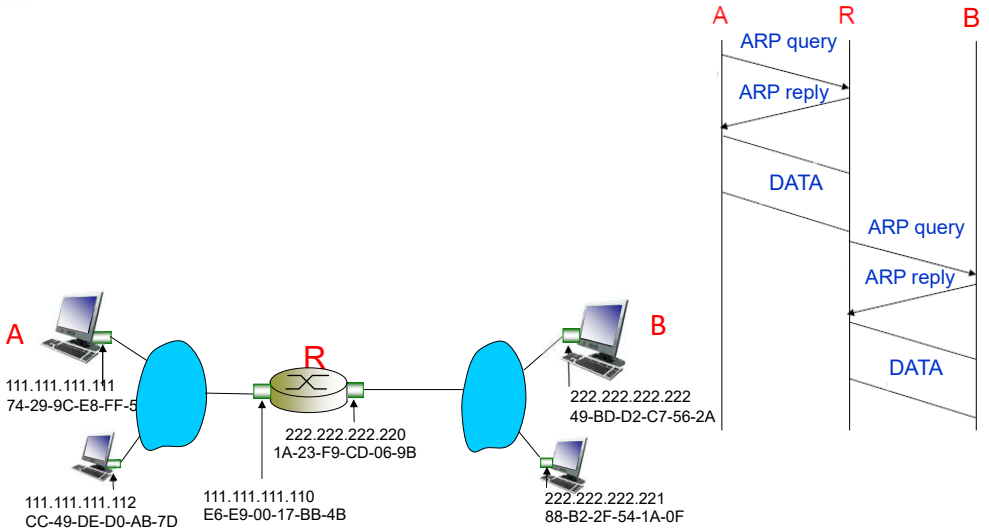
Addressing: Routing to Another LAN

- R forwards datagram with IP source A, destination B
- R creates link-layer frame with B's MAC address as dest, frame contains A-to-B IP datagram



30

Addressing: Routing to Another LAN

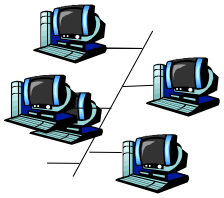


31

TÉCNICO LISBOA

Outline

- Introduction and services
- Link-layer Addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing



32

TÉCNICO LISBOA

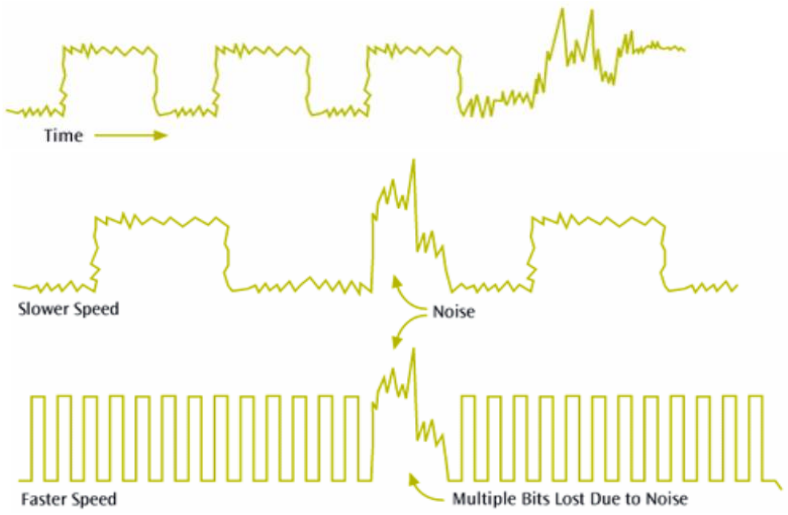
Why are there Transmission Errors?

Origin	Cause	Prevention
Line break	Storms, accidents	
White noise	Electron motion	Raise signal level
Impulsive noise	Lightening, voltage changes, car ignition, ...	Isolate or move wires
Cross-talk	Guard bands too small, Wires too close	Increase guard band or isolate wires
Eco	Bad quality connections	Fix or adjust equipment
Loss/attenuation	Signal intensity decreases with distance	Use repeaters or regenerators
Intermodulation noise	Combination of signals with different origins	Isolate or move wires
Jitter	Phase changes in the signals	Adjust the equipments
Harmonic distortion	Non-linear amplification in frequency	Adjust the equipments

33

TÉCNICO LISBOA

Error Examples



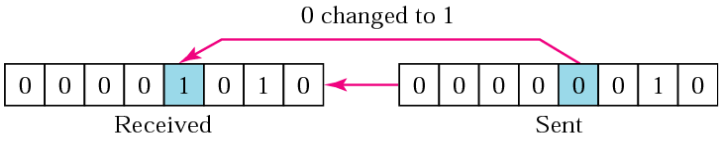
34

TÉCNICO LISBOA

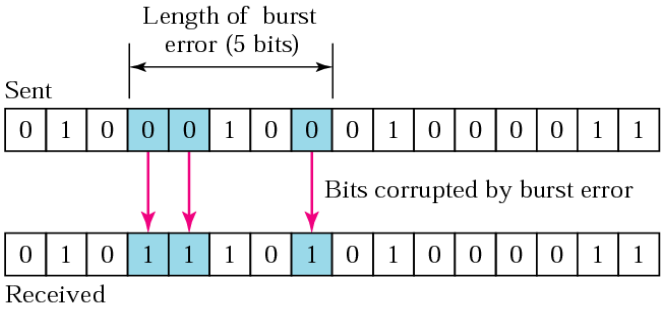
Error Types

Error types:

- Isolated:



- Burst:

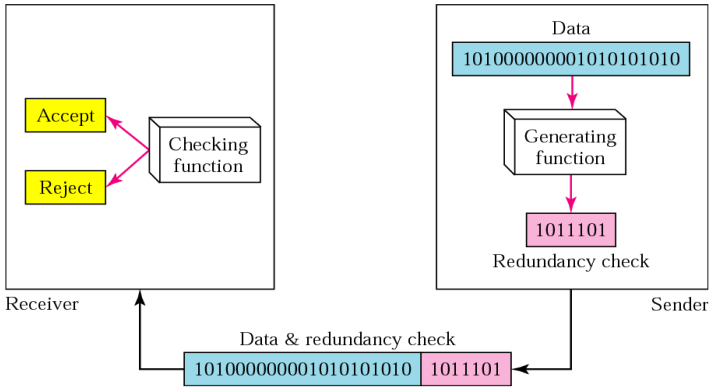


35

Error Detection

Error detection:

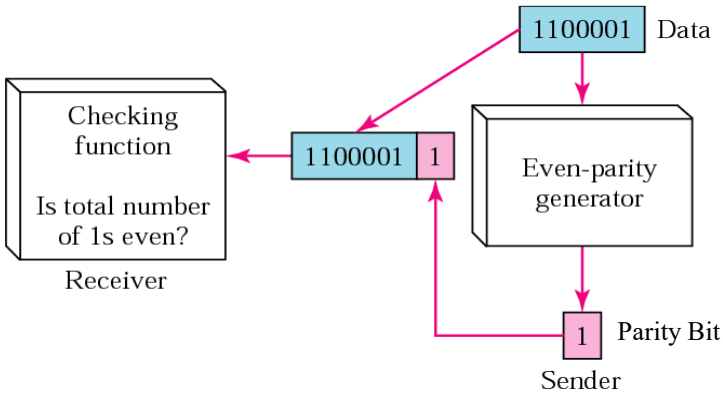
- Uses redundancy – bits are added to allow error detection in the receptor;



36

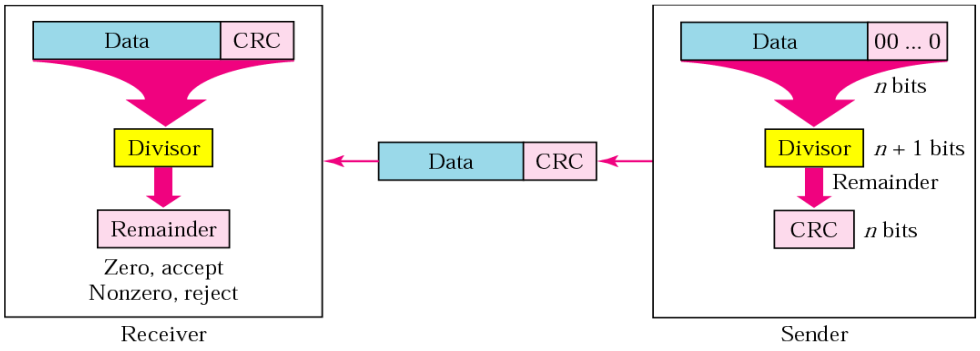
Error Detection: Parity Bit

Does not detect an even number of bits in error;
A substantial overhead is introduced; example: 1/8 = 12.5%.



38

Cyclic Redundancy Check (CRC)

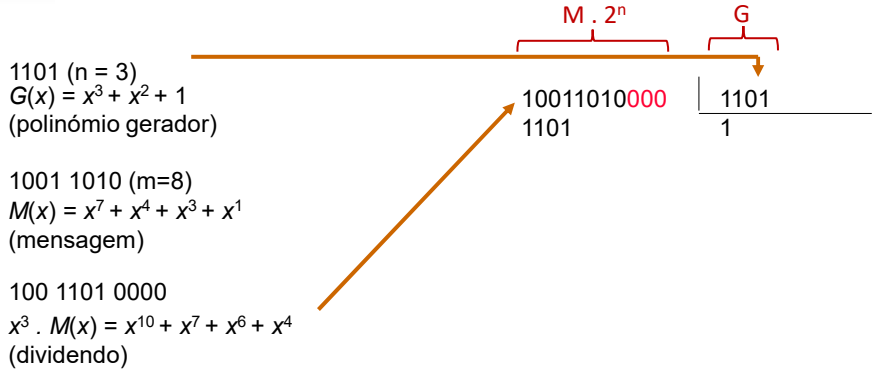


Widely used in practice (e.g., Ethernet, 802.11 WiFi).

Error Detection: CRC

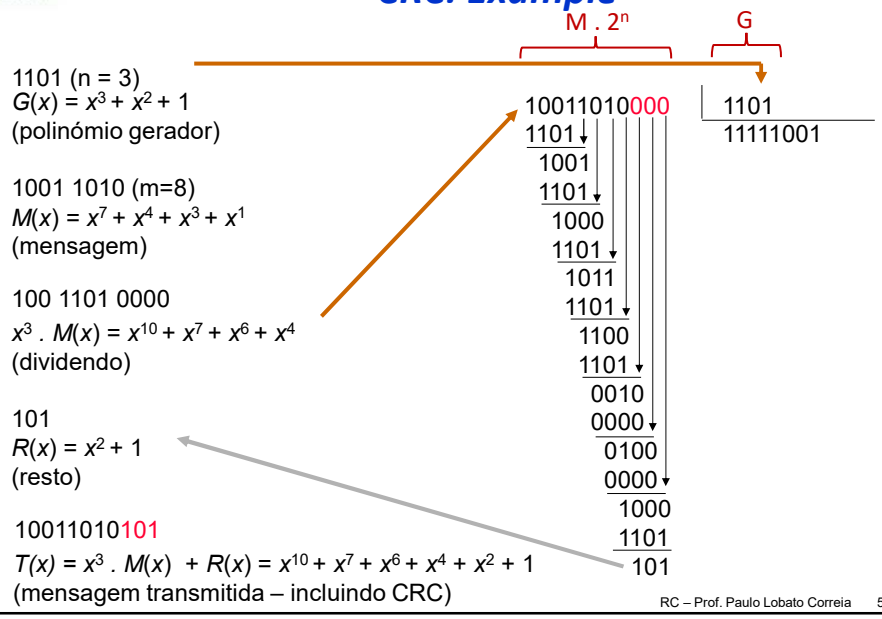
- Uses a generator polynomial $G(x)$, of degree n;
- A message with m bits is viewed as a polynomial $M(x)$ of degree lower than m;
- CRC computation:
  - Dividend is $x^n.M(x)$;
  - Divisor is $G(x)$;
  - Modulo 2 division (*exclusive or*) produces a remainder $R(x)$, of degree lower than n;
  - The **transmitted message** is $T(x) = x^n.M(x) + R(x)$;
  - Note that $T(x)$ is divisible by $G(x)$;
- Error detection:
  - If the received message is divisible by $G(x)$ then no error are detected.

CRC: Example



45

CRC: Example



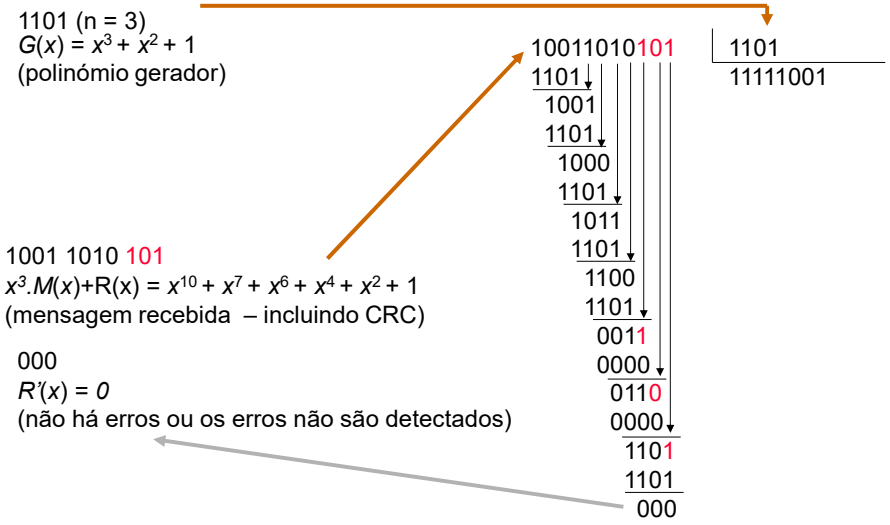
53

TÉCNICO LISBOA

TPC: Prob. 14

CRC: Example (2)

1101 (n = 3)
$G(x) = x^3 + x^2 + 1$
(polinómio gerador)



1001 1010 101
$x^3.M(x)+R(x) = x^{10} + x^7 + x^6 + x^4 + x^2 + 1$
(mensagem recebida – incluindo CRC)

000
$R'(x) = 0$
(não há erros ou os erros não são detectados)

TÉCNICO LISBOA

CRC: Properties

Let $T(x) + E(x)$, be the received message, with $E(x)$ being the error pattern:

- The error pattern is detected if and only if $E(x)$ is not divisible by $G(x)$;
- Single errors are detected if $G(x)$ has more than one term;
- Odd number of errors are detected if $G(x)$ has $x + 1$ as a factor;
- If $G(x)$ has degree n and includes the term 1 (i.e., x^0):
  - Burst errors of length up to n are detected;
  - Burst errors of length $n + 1$ are detected with probability $1 / 2^{n-1}$.

Ethernet, WiFi:

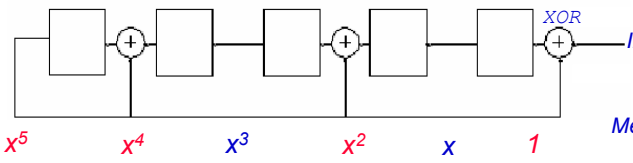
$G_{CRC-32} = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x^1 + x^0$

G_{CRC-32} =10000 0100 1100 0001 0001 1101 1011 0111 = 0x04C11DB7

CRC: Hardware Implementation

CRC is easily implemented in hardware, by using shift registers:

$G(x) = x^5 + x^4 + x^2 + 1$



Message M=1010001101

Input when calculating CRC:
1010001101 00000

CRC =01110 (to be calculated)

(Input when calculating CRC:
1010001101 01110)

Forward Error Correction

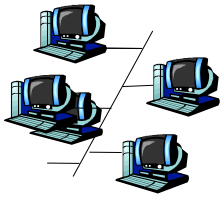
Forward error correction (FEC) uses codes with enough redundancy to allow the detection and correction of errors on the receptor, without requiring the message retransmission.

Examples:

- Hamming code: detects and corrects isolated bit errors.
- There are more sophisticated techniques often used, such as Reed-Solomon codes.
- FEC is often used in environments with long propagation delays (example: satellite transmissions).
- There are hardware implementations of FEC techniques (example: V.34 modem).

Outline

- Introduction and services
- Link-layer Addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing



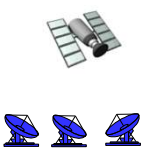
Multiple Access Links and Protocols

Two types of "links":

- Point-to-point:
  - PPP for dial-up access;
  - Point-to-point link between Ethernet switch and host.
- Broadcast (shared wire or medium):
  - Old-fashioned Ethernet;
  - Upstream HFC in cable networks;
  - 802.11 wireless LAN, 4G, 5G, satelitte.

Shared wire (e.g., cabled Ethernet)

Shared RF (e.g., 802.11 WiFi)



Shared RF (satellite)

Humans at a cocktail party (shared air, acoustical)

Multiple Access Control (MAC) Protocols

- Single shared broadcast channel;
- Two or more simultaneous transmissions by nodes – interference:
  - Collision if node receives two or more signals at the same time.

Multiple access control protocol:

- Distributed algorithm that determines how nodes share channel, i.e., determine when node can transmit;
- Communication about channel sharing must use the channel itself!
  - No out-of-band channel for coordination.

Ideal Multiple Access Protocol

Broadcast channel of rate R bps

1. When one node wants to transmit, it can send at rate R.
2. When M nodes want to transmit, each can send at average rate R/M.
3. Fully decentralized:
   - No special node to coordinate transmissions;
   - No synchronization of clocks, slots.
4. Simple.

MAC Protocols: a Taxonomy

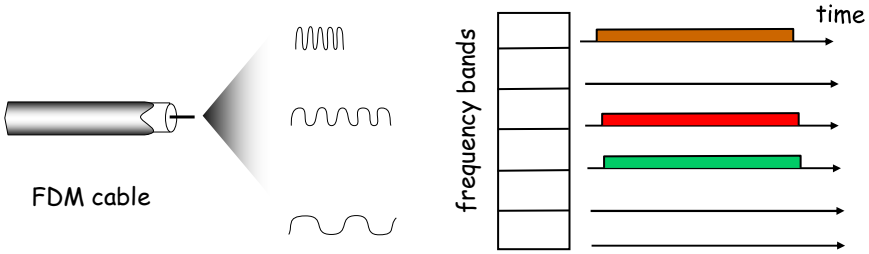
Three broad classes:

- Fixed Channel Partitioning
  - Divide channel into smaller “pieces” (time slots, frequencies, codes);
  - Allocate piece to node for exclusive use.
- Dynamic Allocation (“Taking turns”)
  - Nodes take turns
    (but, nodes with more to send can take longer turns);
  - Poll/Select; Token passing.
- Random Access
  - Channel not divided, allow collisions;
  - “Recover” from collisions.

Fixed Channel Partitioning: FDMA

FDMA – Frequency Division Multiple Access:

- Channel spectrum divided into frequency bands;
- Each station assigned a fixed frequency band;
- Unused transmission time in frequency bands go idle;

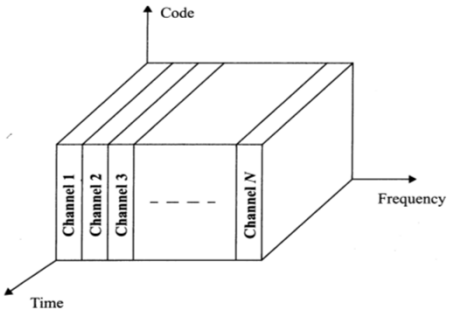


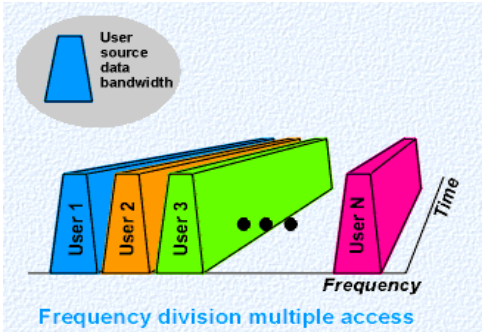
Example of 6-station LAN:
1,3,4 have packets;
Frequency bands 2,5,6 are idle.

TÉCNICO LISBOA

FDMA

FDMA – *Frequency Division Multiple Access*



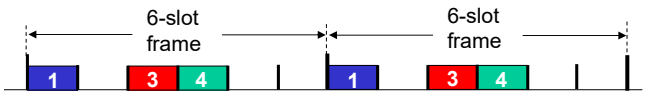


TÉCNICO LISBOA

Fixed Channel Partitioning: TDMA

TDMA – Time Division Multiple Access:

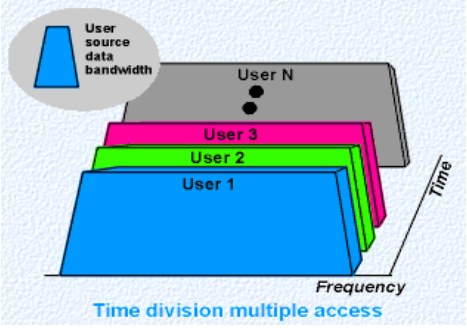
- Access to channel in "rounds";
- Each station gets fixed length slot (length = packet transmission time) in each round;
- Unused slots go idle.

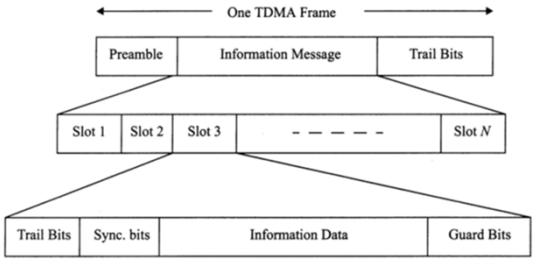


TÉCNICO LISBOA

TDMA

TDMA – *Time Division Multiple Access*



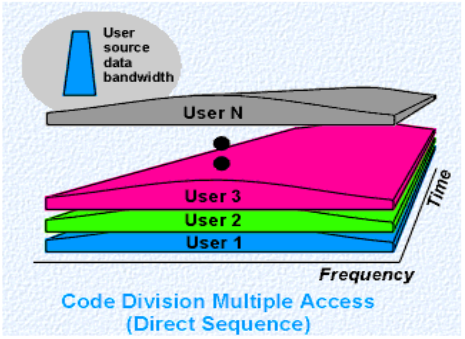


TÉCNICO LISBOA

Fixed Channel Partitioning: CDMA

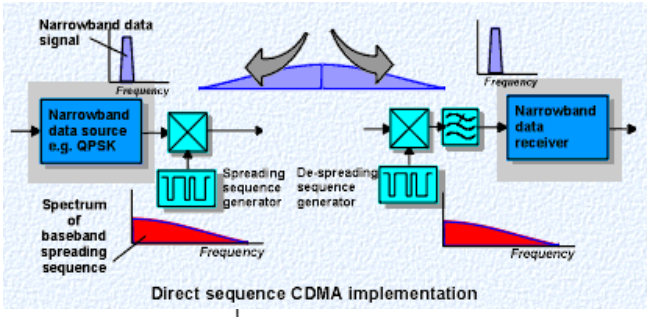
CDMA – *Code Division Multiple Access*

- Each user has access to the complete frequency band, during all the time.
- Users are distinguished by using different codes.

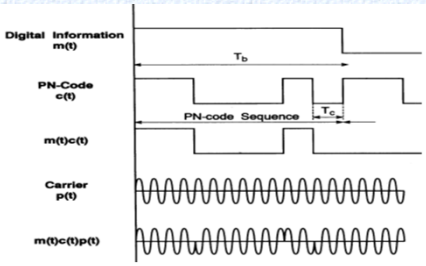


TÉCNICO LISBOA

CDMA



Direct sequence CDMA implementation



TÉCNICO LISBOA

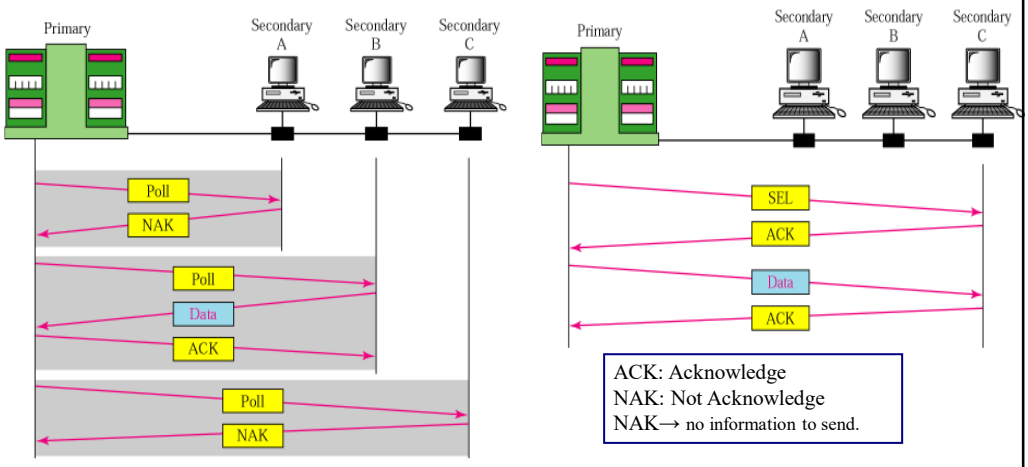
Dynamic Allocation Protocols

Three broad classes:

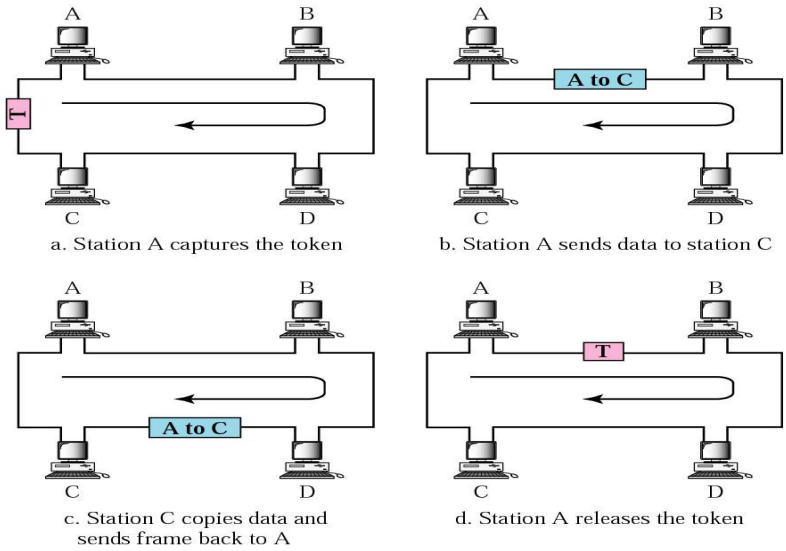
- Fixed Channel Partitioning
  - Divide channel into smaller “pieces” (time slots, frequencies, codes);
  - Allocate piece to node for exclusive use.
- **Dynamic Allocation (“Taking turns”)**
  - Nodes take turns
    (but, nodes with more data to send can take transmit more often);
  - Poll/Select; Token passing.
- Random Access
  - Channel not divided; collisions may occur;
  - “Need to recover” from collisions.

Dynamic Allocation: Poll / Select

A central (*primary*) computer controls the activity of the others.



Dynamic Allocation: Token Passing



a. Station A captures the token

b. Station A sends data to station C

c. Station C copies data and sends frame back to A

d. Station A releases the token

Random Access Protocols

Three broad classes:

- Fixed Channel Partitioning
  - Divide channel into smaller “pieces” (time slots, frequencies, codes);
  - Allocate piece to node for exclusive use.
- Dynamic Allocation (“Taking turns”)
  - Nodes take turns
    (but, nodes with more to send can take longer turns);
  - Poll/Select; Token passing.
- **Random Access**
  - Channel not divided, allow collisions;
  - “Recover” from collisions.

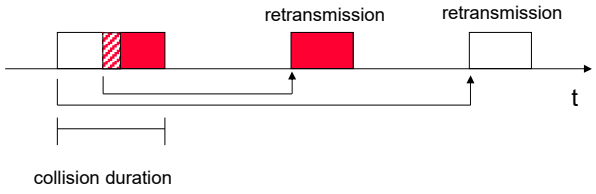
Random Access Protocols

- When node has packet to send:
  - Transmit at full channel data rate R;
  - No *a priori* coordination among nodes.
- Two or more transmitting nodes → “collision”.
- Random access MAC protocol specifies:
  - How to detect collisions;
  - How to recover from collisions (e.g., via delayed retransmissions).
- Examples of random access MAC protocols:
  - ALOHA;
  - Slotted ALOHA;
  - CSMA, CSMA/CD, CSMA/CA.

TÉCNICO LISBOA

ALOHA

- Created at the University of Hawaii in 1970;
- Aloha → simple, **no synchronization**;
- **Hosts transmit at channel rate**;
- When frame is ready: transmit immediately;
- **Collision** if two or more frames overlap (frames are lost);
- If frame is lost, schedule **retransmission** for a future instant (randomly chosen).

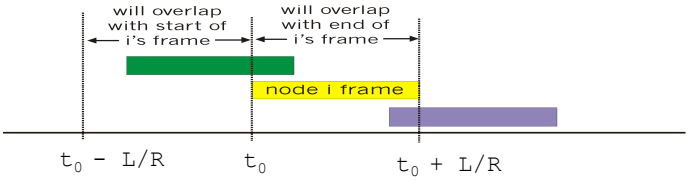


TÉCNICO LISBOA

ALOHA Efficiency

Collisions:

- Frame sent at t_0 collides with any other frames sent in $[t_0\text{-}L/R, t_0\text{+}L/R]$.



P (success by given node) = P (node transmits) ·
· P (no other transmission in $[t_0\text{-}L/R, t_0\text{+}L/R]$)

Vulnerability period

$$P\{success\} = P(k=0) = e^{-2 \cdot G}$$

G = g.L/R Poisson process: $P\{k\ arrivals\} = \frac{(g \cdot t)^k}{k!} \cdot e^{-g \cdot t}$

Slotted ALOHA

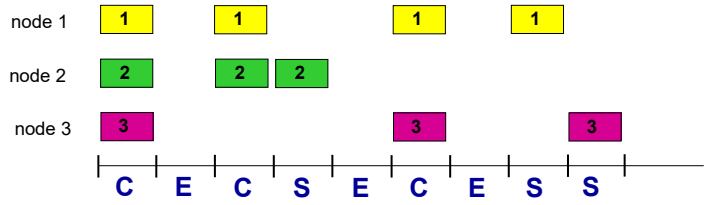
Assumptions:

- All frames of same size **L**;
- Time is divided into equal size slots (time to transmit 1 frame – **L/R**);
- Nodes start to transmit only at slot beginning;
- Nodes are synchronized;
- If 2 or more nodes transmit in a slot, all nodes detect the collision.

Operation:

- When node obtains fresh frame, transmits in next slot:
  - *If no collision:* node can send new frame in next slot;
  - *If collision:* node retransmits frame in each subsequent slot with probability *p* until success.

Slotted ALOHA



Pros:

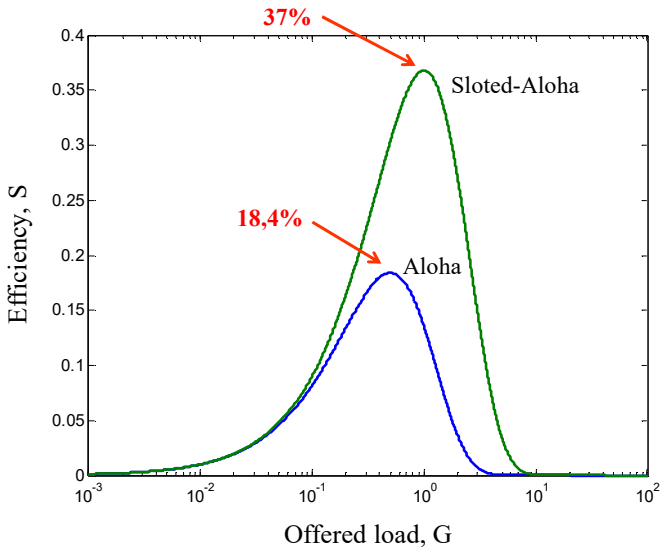
- Single active node can continuously transmit at full channel rate;
- Highly decentralized: only slot durations need to be in sync in every node;
- Simple.

Cons:

- Collisions, wasting slots;
- Idle slots;
- Nodes may be able to detect collision in less than time needed to transmit packet;
- Clock synchronization.

TÉCNICO LISBOA

Aloha Efficiency



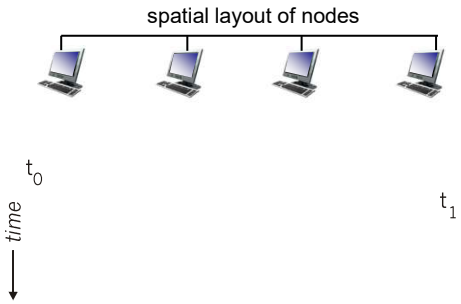
TÉCNICO LISBOA

Carrier Sense Multiple Access (CSMA)

CSMA:

- Listen before transmit:
  - If channel sensed idle → transmit entire frame;
  - If channel sensed busy → defer transmission.
- Human analogy: don't interrupt others!

CSMA Collisions



Collisions can still occur:

- Propagation delay means two nodes may not hear each other's transmission;

Collision:

- Entire packet transmission time wasted;

Note:

- Role of distance & propagation delay in determining collision probability.

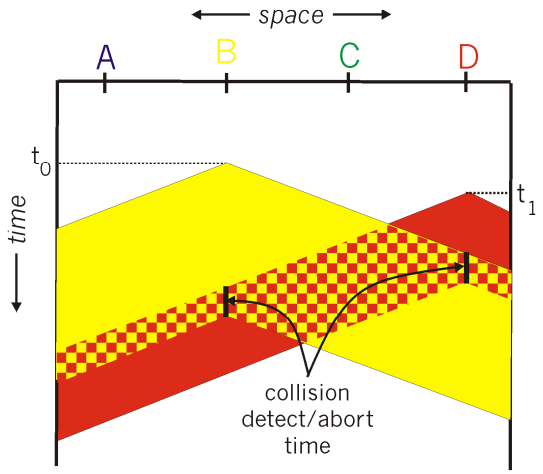
CSMA/CD (Collision Detection)

CSMA/CD:

- Carrier sensing, deferral as in CSMA:
  - **Collisions *detected*** within short time;
  - Colliding transmissions aborted, reducing channel useless occupation;
- Collision detection:
  - Easy in wired LANs: measure signal strengths, compare transmitted, received signals;
  - Difficult in wireless LANs: received signal strength overwhelmed by local transmission strength ;
- Human analogy: the polite conversationalist ☺

TÉCNICO LISBOA

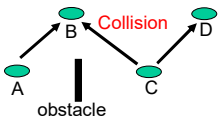
CSMA/CD Collision Detection



TÉCNICO LISBOA

Carrier Sense Multiple Access (CSMA)

- Listen to the channel before transmission:
  - If channel is busy, postpone frame transmission;
  - If channel is free, immediately start frame transmission;
- Persistent CSMA:
  - When channel is busy, the station transmits as soon as it becomes idle;
- Non-persistent CSMA:
  - When channel is busy, the station schedules the frame transmission for a future moment, randomly chosen;
- CSMA/CD:
  - Stations involved in a collision stop their transmission as soon as the collision is detected;
- CSMA/CA:
  - There may be “hidden“ stations…



Summary of MAC Protocols

Fixed channel partitioning, by time, frequency or code:

- Frequency Division, Time Division, Code Division;

Dynamic allocation ("taking turns"):

- Polling from central site, token passing;
- Bluetooth, FDDI, IBM Token Ring.

Random access (dynamic):

- ALOHA, S-ALOHA, CSMA, CSMA/CD;
- Carrier sensing: easy in some technologies (wire), hard in others (wireless);
- CSMA/CD used in Ethernet (IEEE 802.3);
- CSMA/CA used in Wi-Fi (IEEE 802.11).

Summary of MAC Protocols

Fixed channel partitioning MAC protocols:

- Efficient with steady high load from all nodes;
- Inefficient at low or unbalanced loads:
  - 1/N bandwidth allocated even if only 1 active node!

Dynamic allocation ("taking turns") protocols:

- Share channel *efficiently* and *fairly* at high loads;
- Low load: inefficient as active nodes have to wait for idle nodes to "pass their turn";

Random access MAC protocols:

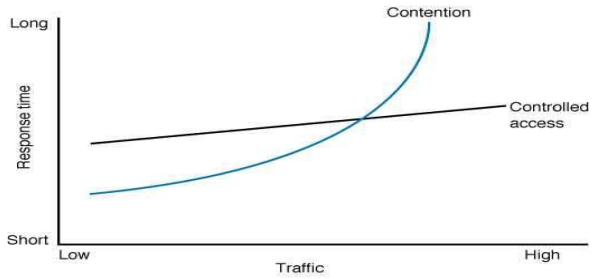
- Efficient at low load: single node can fully utilize channel;
- High load: collision overhead.

TÉCNICO LISBOA

MAC Protocols

Comparing dynamic allocation with random access protocols:

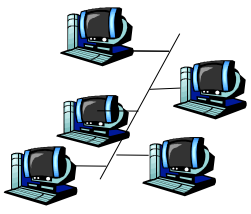
- Dynamic allocation protocols allow a better channel usage at high loads;
- Random access protocols impose lower delays at low loads;
- Dynamic allocation protocols require central management or token management;
- Random access protocols need to be controlled against unstable behavior.



TÉCNICO LISBOA

Outline

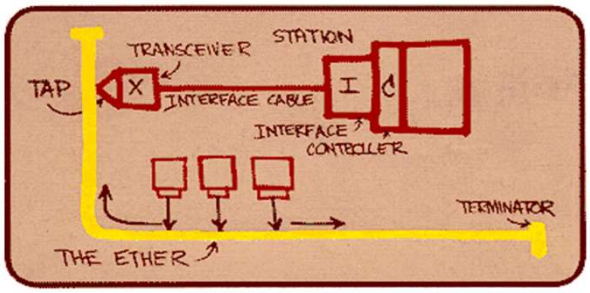
- Introduction and services
- Link-layer Addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing



IEEE 802.3 – Ethernet

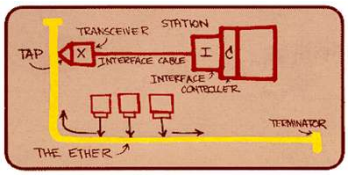
"Dominant" wired LAN technology:

- Cheap: < $10 for network interface card (NIC);
- First widely used LAN technology;
- Simpler, cheaper than token LANs and ATM;
- Kept up with speed race: 10 Mbps … 100 Gbps, 400Gbps, ...



Metcalfe's Ethernet sketch

IEEE 802.3 – Ethernet



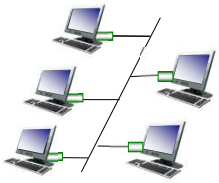
Dismantled vampire tap (10BASE5):

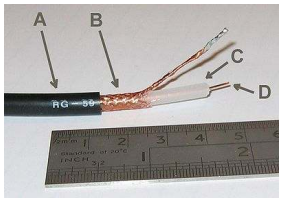
- Central metal-tipped insulated spike contacts cable core;
- smaller spikes contact cable shield.

Note black mark on cable sheath indicating suitable location for transceiver.

Bus Topology

- Bus topology popular through mid 1990s:
  - All nodes in same collision domain (can collide with each other);

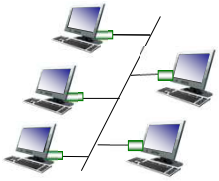




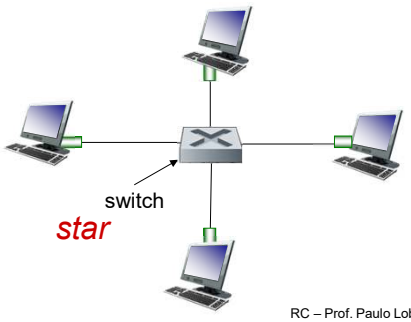
bus: coaxial cable

Star Topology

- Bus topology popular through mid 90s:
  - All nodes in same collision domain (can collide with each other);
- Today: star topology prevails:
  - Active *switch* in center;
  - Each "spoke" runs a (separate) Ethernet protocol (nodes do not collide with each other).



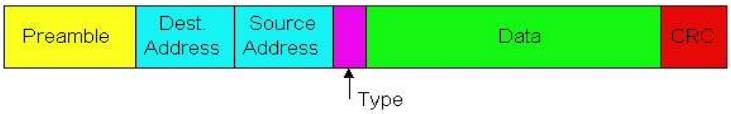
bus: coaxial cable



star

Ethernet Frame Structure

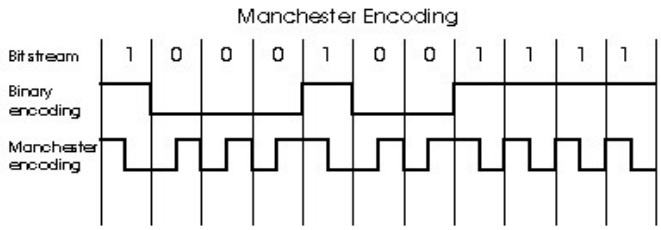
Sending adapter encapsulates IP datagram (or other network layer protocol packet) in Ethernet frame:



Preamble:

- 7 bytes with pattern 10101010 followed by one byte with pattern 10101011;
- Used to synchronize receiver and sender clock rates.

Manchester Encoding

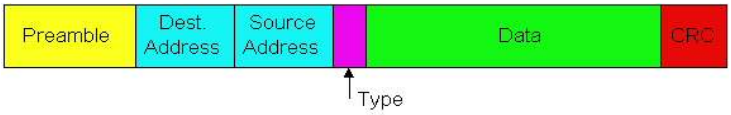


- Manchester encoding is used in 10BaseT;
- Each bit has a transition;
- Allows clocks in sending and receiving nodes to synchronize to each other:
  - No need for a centralized, global clock among nodes!

(This is physical-layer material)

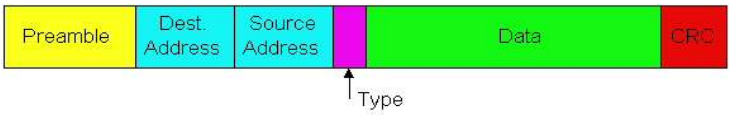
Ethernet Frame Structure

- Addresses: 6 bytes
  - If adapter receives frame with matching destination MAC address (or broadcast address – e.g. ARP packet), it passes data in frame to the network layer;
  - Otherwise, adapter discards frame.
- Type: indicates higher layer protocol:
  - Mostly IP but others possible, e.g., Novell IPX, AppleTalk;



Ethernet Frame Structure

- Data:
  - MTU: 46 to 1500 bytes.
- CRC:
  - $x^{32}+ x^{26}+ x^{23}+ x^{22}+ x^{16}+ x^{12}+ x^{11}+ x^{10}+ x^{8}+ x^{7}+ x^{5}+ x^{4}+ x^{2}+ x+1$
  - Checked at receiver – if error is detected the frame is dropped.



Ethernet: Unreliable, Connectionless

- Connectionless:
  No handshaking between sending and receiving NICs.
- Unreliable:
  Receiving NIC doesn't send ACKs or NACKs to sending NIC:
  - Stream of datagrams passed to network layer can have gaps (missing datagrams);
  - Gaps will be filled in if application is using TCP;
  - Otherwise, application will see the gaps.
- Ethernet's MAC protocol:
  Unslotted CSMA/CD with exponential binary backoff.

Ethernet CSMA/CD Algorithm

1. NIC receives datagram from network layer → creates frame;
2. If **channel idle** (96 bit times), starts frame transmission;
   If **channel busy**, waits until channel idle, then transmits
3. If NIC transmits entire frame without detecting another transmission:
   → success;
4. If NIC detects another transmission while transmitting - collision:
   → aborts and sends jam signal (reinforce collision);
5. After aborting, NIC enters **exponential backoff**:
   → After m^{th} collision, NIC chooses K at random from $\{0,1,2,\ldots,2^m-1\}$;
   → NIC waits K x 512 bit times; then returns to Step 2.

TÉCNICO LISBOA

Ethernet's CSMA/CD

TPC: Prob. 15

Jam Signal:

Make sure all other transmitters are aware of collision → 48 bits;

Bit time:

1 ns for 1 Gbps Ethernet;

Exponential Backoff:

Goal: adapt retransmission attempts to estimated load:

- Heavy load: random wait will be longer.

1st collision: choose K from {0,1}; delay is K x 512 bit transmission times;

After 2nd collision: choose K from {0,1,2,3}…

After 10 collisions: choose K from {0,1,2,3,4,…,1023}

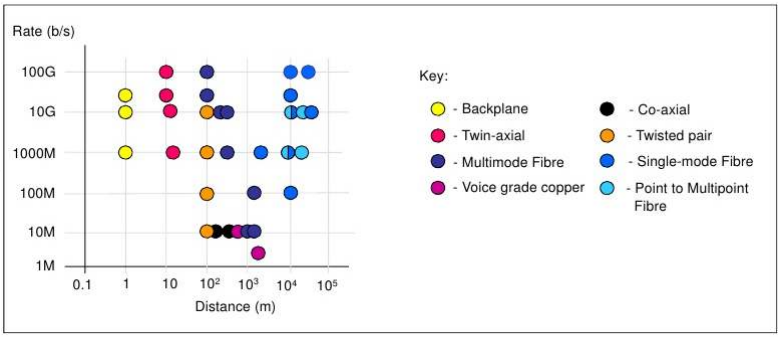
(K=1023 → wait time is about 0.5 msec, with 1 Gbps Ethernet)

https://media.pearsoncmg.com/aw/ecs_kurose_compnetwork_7/cw/content/interactiveanimations/csma-cd/index.html

TÉCNICO LISBOA

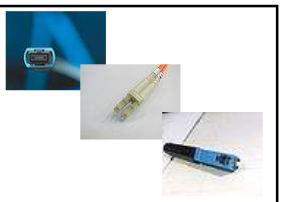
802.3 Ethernet Standards: Link & Physical Layers

- *Many* different Ethernet standards:
  - Common MAC protocol and frame format;
  - Different speeds: 2 Mbps, 10 Mbps, 100 Mbps, 1Gbps, 10 Gbps, 100 Gbps, 400 Gbps;
  - Different physical layer media: fiber, cable;



TÉCNICO LISBOA

802.3 Ethernet Standards: Link & Physical Layers



400 Gbit/s [edit]

Main article: Terabit Ethernet

The Institute of Electrical and Electronics Engineers (IEEE) has defined a new Ethernet standard capable of 200 and 400 Gbit/s in IEEE 802.3bs-2017.[25] 1 Tbit/s may be a further goal.[26]

In May 2018, IEEE 802.3 started the 802.3ck Task Force to develop standards for 100, 200, and 400 Gbit/s PHYs and attachment unit interfaces (AUI) using 100 Gbit/s lanes.[24]

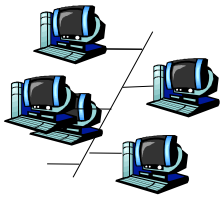
In 2008, Robert Metcalfe, one of the co-inventors of Ethernet, said he believed commercial applications using Terabit Ethernet may occur by 2015, though it might require new Ethernet standards.[27] It was predicted this would be followed rapidly by a scaling to 100 Terabit, possibly as early as 2020. It is worth noting that these were theoretical predictions of technological ability, rather than estimates of when such speeds would actually become available at a practical price point.[28]

Name	Standard (Clause)	Common connectors	Description
Fiber-optical cable			
400GBASE-SR16	802.3bs-2017 (123)	MPO	sixteen lanes (26.5625 Gbit/s) using individual strands of OM4/OM5 multi-mode fiber with 100 m reach or 70 m over OM3
400GBASE-DR4	802.3bs-2017 (124)	MPO	four PAM-4 lanes (53.125 GBd) using individual strands of single-mode fiber with 500 m reach (1310 nm)
400GBASE-FR8	802.3bs-2017 (122)	SC, LC	eight PAM-4 lanes (26.5625 GBd) using eight wavelengths (CWDM) over single-mode fiber with 2 km reach
400GBASE-LR8	802.3bs-2017 (122)	SC, LC	eight PAM-4 lanes (26.5625 GBd) using eight wavelengths (DWDM) over single-mode fiber with 10 km reach
400GBASE-FR4	*802.3cu*	SC, LC	four lanes/wavelengths (CWDM, 1271/1291/1311/1331 nm) over single-mode fiber with 2 km reach
400GBASE-LR4			four lanes over single-mode fiber with 10 km reach
400GBASE-SR8	*802.3cm*	SC, LC	eight-lane using individual strands of multi-mode fiber with 100 m reach
400GBASE-SR4.2			four-lane using individual strands of multi-mode fiber with 100 m reach
400GBASE-ER8	*802.3cn*	SC, LC	eight-lane using eight wavelengths over single-mode fiber with 40 km reach
Other			
400GBASE-KR4	*802.3ck* (tbd)		four-lane over electrical backplanes supporting an insertion loss of up to 28 dB at 26.56 GBd
400GBASE-CR4			four-lane over twin-axial copper with at least 2 m reach

TÉCNICO LISBOA

Outline

- Link-layer Addressing
- Introduction and services
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing

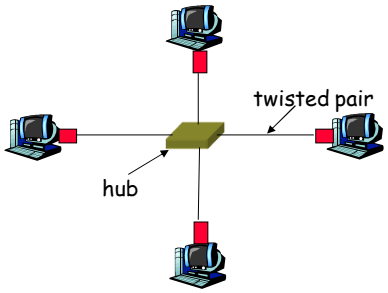


TÉCNICO LISBOA

Hubs

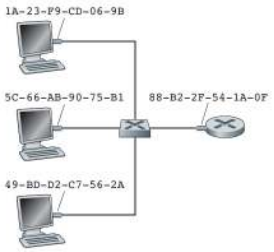
… physical-layer (“dumb”) repeaters:

- Bits coming in one link go out in *all* other links at same rate;
- All nodes connected to hub can collide with one another;
- No frame buffering;
- No CSMA/CD at hub: hosts detect collisions.



TÉCNICO LISBOA

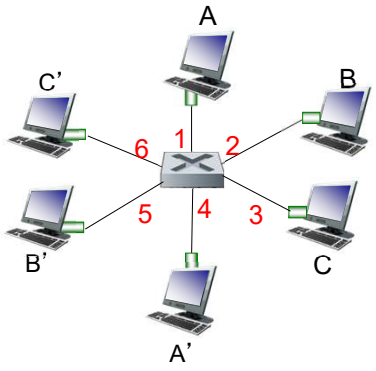
Switch



- Link-layer device;
  Smarter than hubs, takes *active* role:
  - Stores, forwards Ethernet frames;
  - Examines incoming frame’s MAC address:
    - Selectively forwards frame to one or more outgoing links when frame is to be forwarded on segment;
    - Uses CSMA/CD to access segment:
- *Transparent*
  - Hosts are unaware of presence of switches.
- *Plug-and-play, self-learning*
  - Switches do not need to be configured.

Switch
Allows Multiple *Simultaneous Transmissions*

- Hosts have dedicated, direct connection to switch.
- Switches buffer packets.
- Ethernet protocol used on *each* incoming link, but no collisions; full duplex:
  - Each link is its own collision domain.
- *Switching:* allows simultaneous communications A-to-A' and B-to-B' without collisions:
  - Not possible with dumb **hub**.



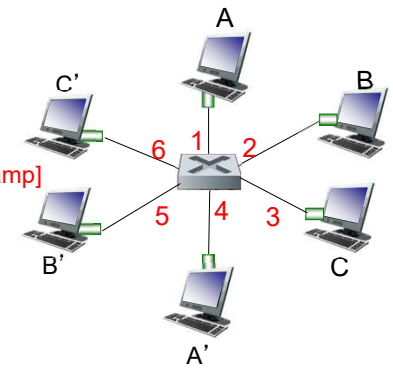
switch with six interfaces (1,2,3,4,5,6)

Switch Table

- *Q:* How does switch know that A' is reachable via interface 4, and B' is reachable via interface 5?

  A: Each switch has a **switch table**;
  Entries contain:
  [MAC address of host; interface to reach host; time stamp]

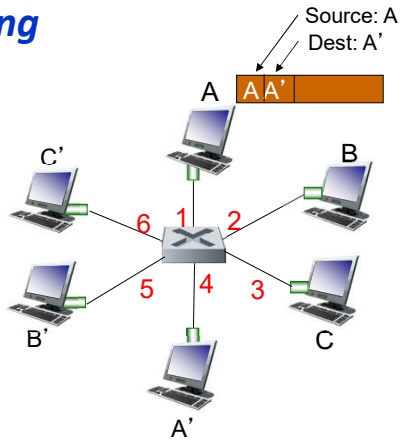
  Looks like a routing table!
  Q: How are entries created and maintained in switch table?
  Something like a routing protocol?



switch with six interfaces (1,2,3,4,5,6)

Switch: Self-Learning

- Switch *learns* which hosts can be reached through which interfaces:
  - When frame is received, switch "learns" location of sender → it is the incoming LAN segment
  - Records sender/location pair in the switch table.



MAC addr	interface	TTL
A	*1*	*60*

Switch table (initially empty)

Switch: Frame Filtering/Forwarding

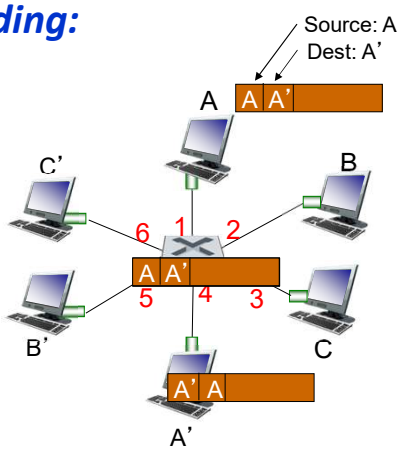
When frame received:

1. Record link associated with sending host;
2. Index switch table using MAC destination address;
3. **if** entry **found** for destination
   then {
   if destination on segment from which frame arrived
   then drop the frame;
   else forward the frame on interface indicated
   }
   else flood.

forward on all but the interface on which the frame arrived

Self-learning, Forwarding: Example

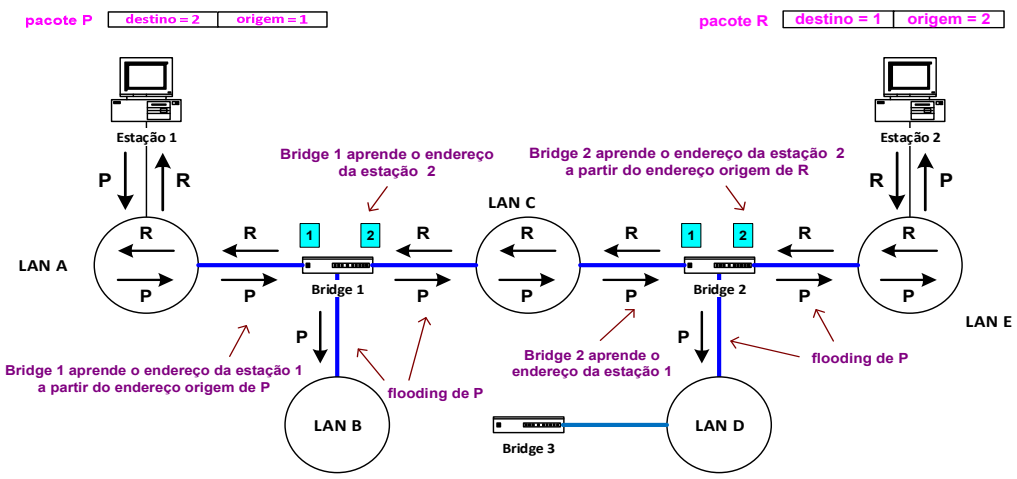
- Frame destination unknown: *flood*
- Destination is a known location: *selective send*



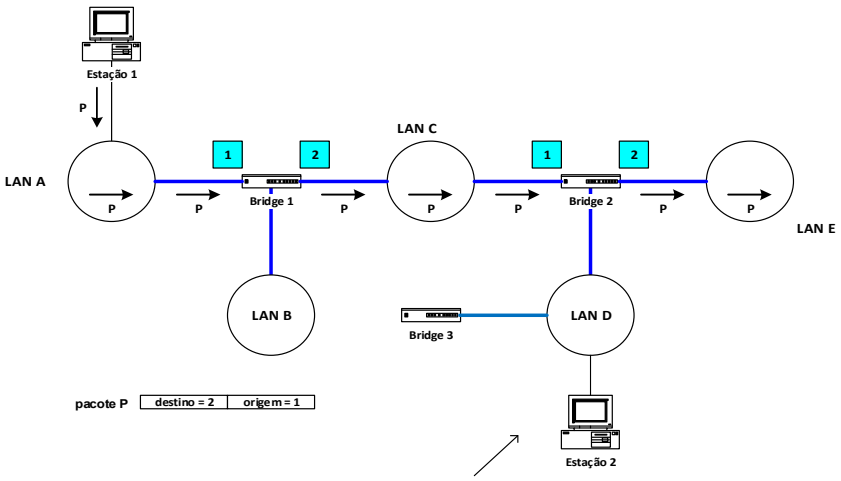
MAC addr	interface	TTL
A	*1*	*60*
A'	*4*	*60*

switch table (initially empty)

Self-learning: Example



TÉCNICO LISBOA



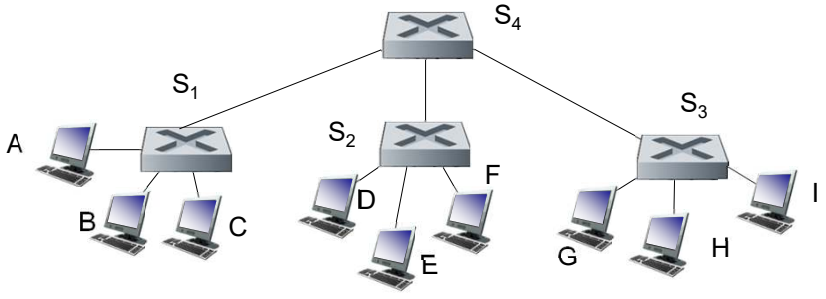
Estação 2 muda de LAN: não recebe pacotes

- até transmitir ou até
- expirar tempo de vida da sua entrada na tabela de encaminhamento

TÉCNICO LISBOA

Interconnecting Switches

- Switches can be interconnected:

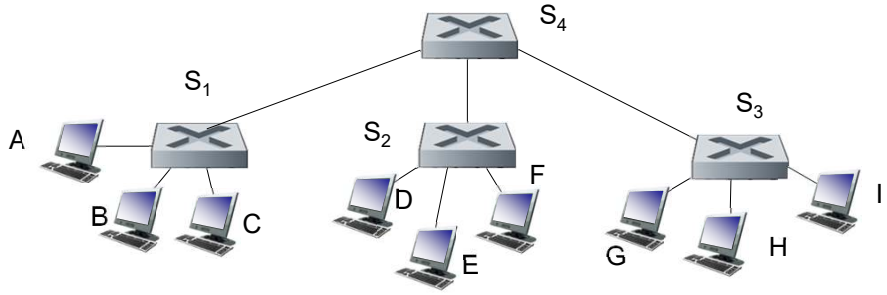


Q: Sending from A to G:
How does S_1 know to forward frame destined to G via S_4 and S_3?

A: Self learning!
(Works exactly as in single-switch case!)

Self-learning Multi-switch Example

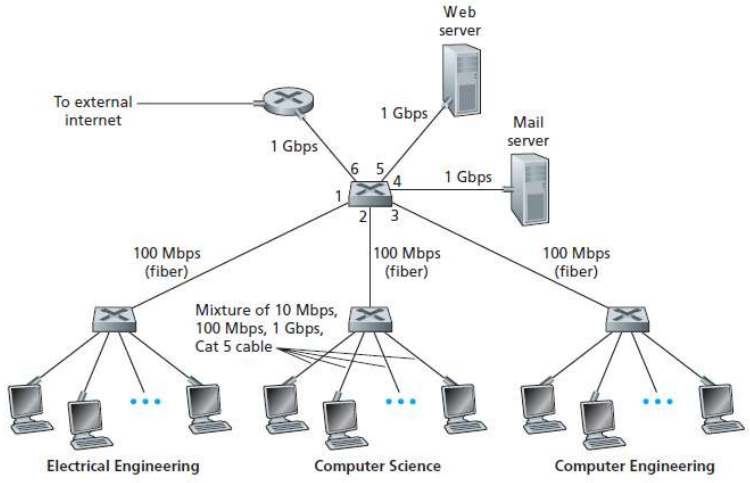
Suppose C sends frame to I; then I responds to C:



Q: Show switch tables and packet forwarding in S_1, S_2, S_3, S_4

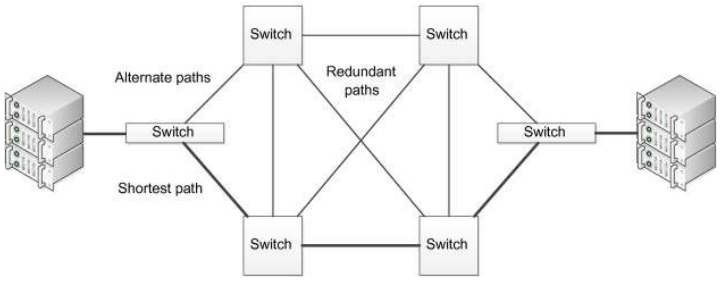
Example

An institutional network connected (one single subnet) using four switches:



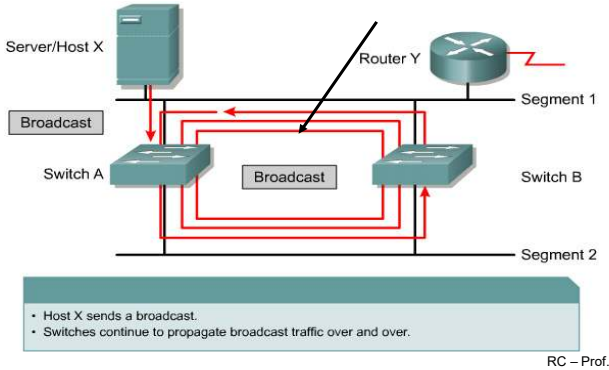
Redundant Topology

- Networks introduce **redundant links** between switches or bridges to overcome the failure of links.
- These connections introduce **physical loops** into the network!
- Bridging loops are created so **if one link fails another can take over** the function of traffic forwarding!



Redundant Topology

- Switches flood traffic out all ports when the traffic is sent to a destination not yet known.
- Broadcast and multicast traffic is forwarded out on every port (except the port on which the traffic arrived).
- This **traffic can be caught in a loop**.



TÉCNICO LISBOA

Redundant Topology

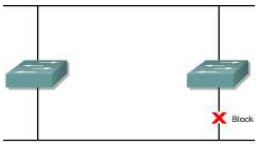
In **Layer 2** header there is **no Time To Live** (TTL):

- A frame sent into a Layer 2 redundant topology of switches, ***can loop forever…***

  (In Layer 3 the TTL is decremented and the packet is discarded when TTL reaches 0).

This creates a dilemma:

- **Loops are needed** in the physical topology **for reliability**
- But, **a switched network cannot have loops**.

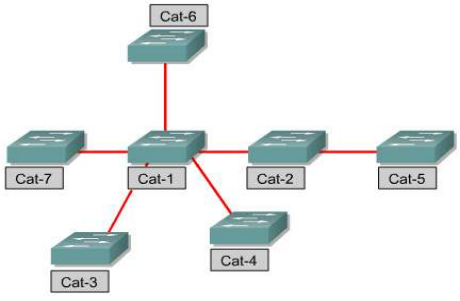


TÉCNICO LISBOA

Redundant Topology and Spanning Tree

Solution:

- Allow physical loops, but **create a loop free logical topology**.
- Traffic destined to Cat-5 from any host attached to Cat-4 will travel through Cat-1 and Cat-2.
- This happens even if there is a direct physical connection between Cat-5 and Cat-4.

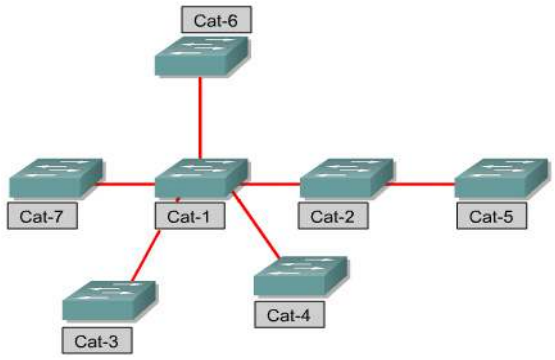


The loop free logical topology created is a **tree** and has a star or extended star logical topology: the **spanning tree** of the network.

TÉCNICO LISBOA

Spanning-Tree Protocol

- Ethernet bridges and switches can implement the **IEEE 802.1D Spanning-Tree Protocol** and use the spanning-tree algorithm to construct a loop free shortest path network.



133

TÉCNICO LISBOA

Spanning-Tree Protocol

- The Spanning-Tree Protocol establishes a root node, called the **root bridge**.
- The Spanning-Tree Protocol constructs a **topology** that has **one path for reaching every network node**.
- The **resulting tree originates from the root bridge**.
- **Redundant links** that are not part of the shortest path tree **are blocked**.

134

Basic Spanning Tree Concepts

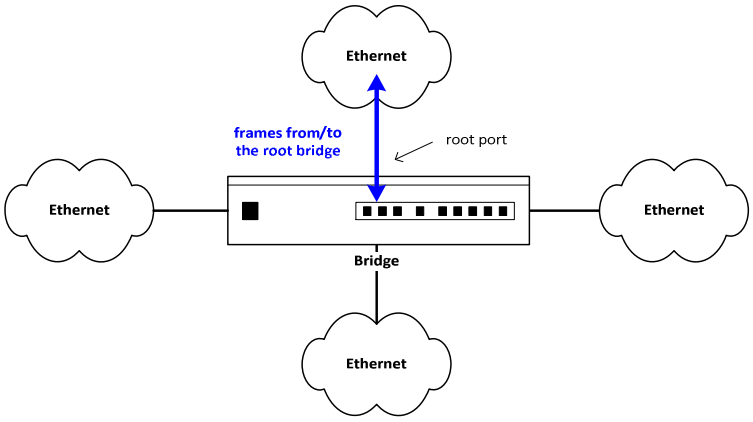
- Bridge ID – each bridge is identified by an address containing:
  - 2 priority bytes – configurable by the network manager
  - 6 fixed bytes (one of the bridge ports MAC addresses, or any other 48 bit unique address)

  The priority part has precedence over the fixed bytes.

- *Root Bridge* – bridge in the root of the spanning tree; bridge with the **lower** Bridge ID

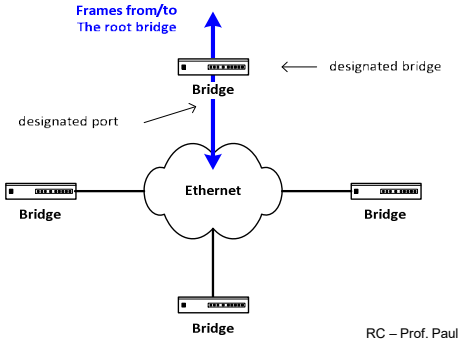
Basic Spanning Tree Concepts

- *Root Port* – port which, in a bridge, is responsible for the reception/transmission of frames from/to the root bridge



Basic Spanning Tree Concepts

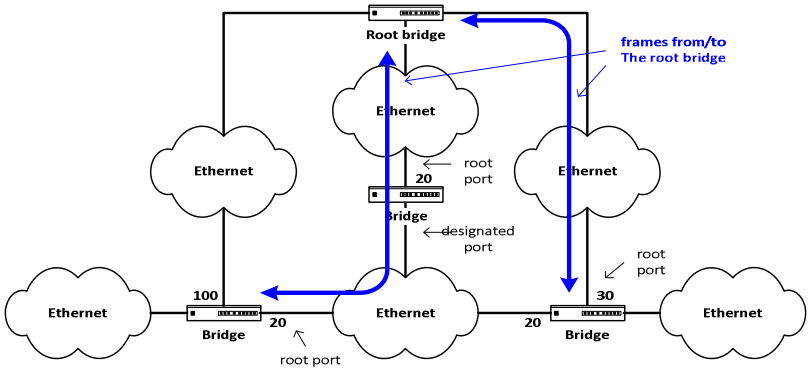
- *Designated Bridge* – bridge which, for a given LAN, is responsible for sending frames from that LAN to the root bridge and vice-versa; (the root bridge is the designated bridge in all the LANs to which it is connected);
- *Designated Port* – port which, for a given LAN, is responsible for sending frames from that LAN to the root bridge and vice-versa (it is the ports of the designated bridge connected to the LAN)



137

Basic Spanning Tree Concepts

- Each bridge has an associated cost of the path to the root bridge (*Root Path Cost*), equal to the sum of the costs from the ports that transmit frames towards the root bridge (root ports), in the least cost path to the root bridge



138

TÉCNICO LISBOA

Spanning-Tree Protocol

- Shortest path is based on cumulative link costs.
- Link costs are based on the speed of the link:

Table 17-3—Port Path Cost values

Link Speed	Recommended value	Recommended range	Range
<=100 Kb/s	200 000 000*	20 000 000–200 000 000	1–200 000 000
1 Mb/s	20 000 000[a]	2 000 000–200 000 000	1–200 000 000
10 Mb/s	2 000 000[a]	200 000–20 000 000	1–200 000 000
100 Mb/s	200 000[a]	20 000–2 000 000	1–200 000 000
1 Gb/s	20 000	2 000–200 000	1–200 000 000
10 Gb/s	2 000	200–20 000	1–200 000 000
100 Gb/s	200	20–2 000	1–200 000 000
1 Tb/s	20	2–200	1–200 000 000
10 Tb/s	2	1–20	1–200 000 000

*Bridges conformant to IEEE Std 802.1D, 1998 Edition, i.e., that support only 16-bit values for Path Cost, should use 65 535 as the Path Cost for these link speeds when used in conjunction with Bridges that support 32-bit Path Cost values.

TÉCNICO LISBOA

Basic Spanning Tree Concepts

- The root port is, in each bridge, the port which provides the best path (lower cost) to the root
- The designated port is, in each LAN, the port which provides the best path (lower cost) to the root

The **active ports** in each bridge are:
root port + the designated ports

The remaining ports stay **inactive (blocked)**

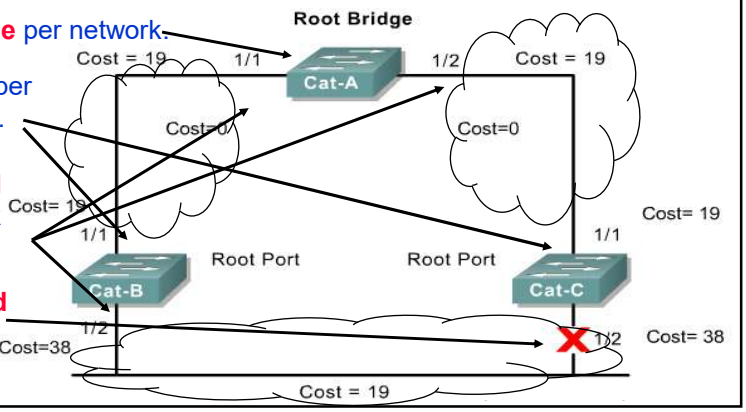
TÉCNICO LISBOA

Spanning-Tree Operation

When the network has stabilized, it has converged and there is one spanning tree per network.

As a result, for every switched network the following elements exist:

- One **root bridge** per network.
- One **root port** per non-root bridge.
- One **designated port** per network (LAN) segment.
- Unused, **blocked ports**.

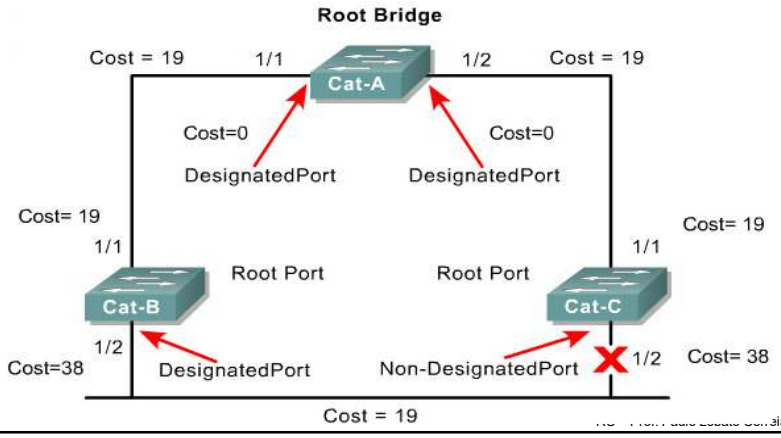


141

TÉCNICO LISBOA

Spanning-Tree Protocol

- The Spanning-Tree Protocol requires network devices to exchange messages to detect bridging loops.
- Links that will cause a loop are put into a blocking state.



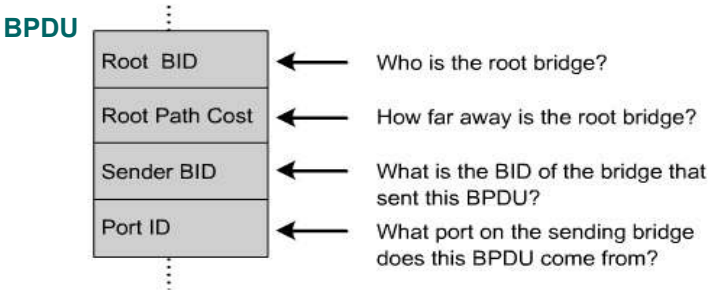
142

TÉCNICO LISBOA

Spanning-Tree Protocol

The message that a switch/bridge sends, allowing the formation of a loop free logical topology, is called a **Bridge Protocol Data Unit** (**BPDU**).

- BPDUs continue to be received on blocked ports.
- This ensures that if an active path or device fails, a new spanning tree can be calculated.

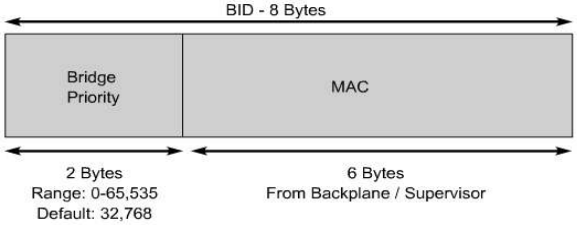


143

TÉCNICO LISBOA

Selecting the Root Bridge

- When a switch (or bridge) is turned on, the spanning-tree algorithm is used to identify the root bridge.
- BPDUs are sent out with the Sender **Bridge ID** (**BID**).
- The BID consists of a bridge priority that defaults to 32768 and the switch base MAC address (lowest MAC address of bridge ports).



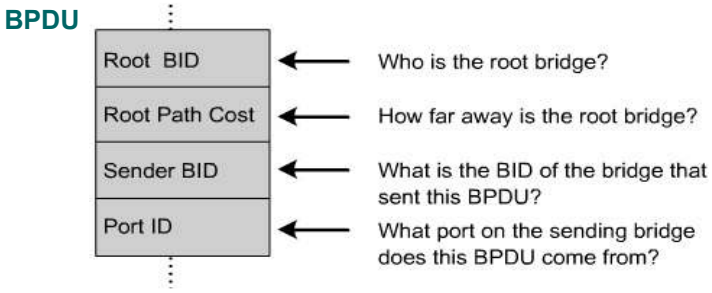
- Bridge ID (BID) is used to identify each bridge/switch.
- The BID is used in determining the center of the network, in respect to STP, known as the root bridge.

144

TÉCNICO LISBOA

Selecting the Root Bridge

- When a switch first starts up, it assumes it is the root switch and sends “inferior” BPDUs.
- These BPDUs contain the switch MAC address in both the root and sender BID.
- **By default BPDUs are sent every two seconds**.

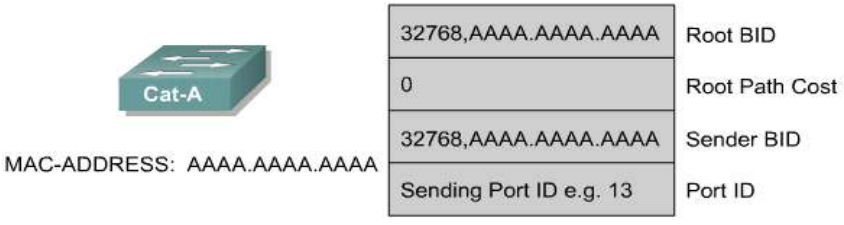


145

TÉCNICO LISBOA

Selecting the Root Bridge

- All switches see the BIDs sent.
- As a switch receives a BPDU with a lower root BID it replaces that in the subsequent BPDUs that are sent out.
- All bridges see these and decide that **the bridge with the smallest BID value will be the root bridge**.

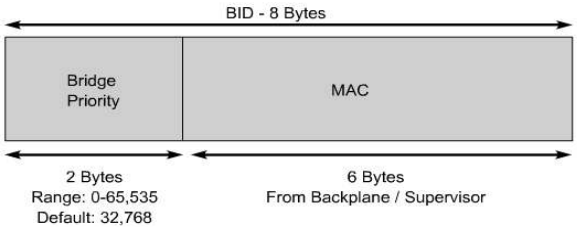


146

TÉCNICO LISBOA

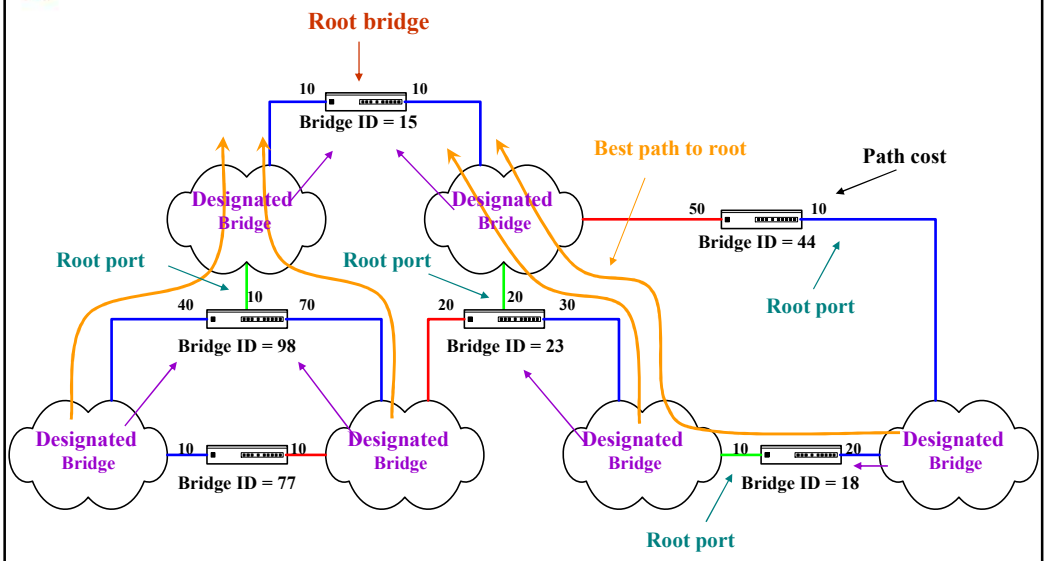
Selecting the Root Bridge

- A network administrator may want to influence the decision by setting the switch priority to a smaller value than the default, which will make the BID smaller.
- This should only be implemented when the traffic flow on the network is well understood.



TÉCNICO LISBOA

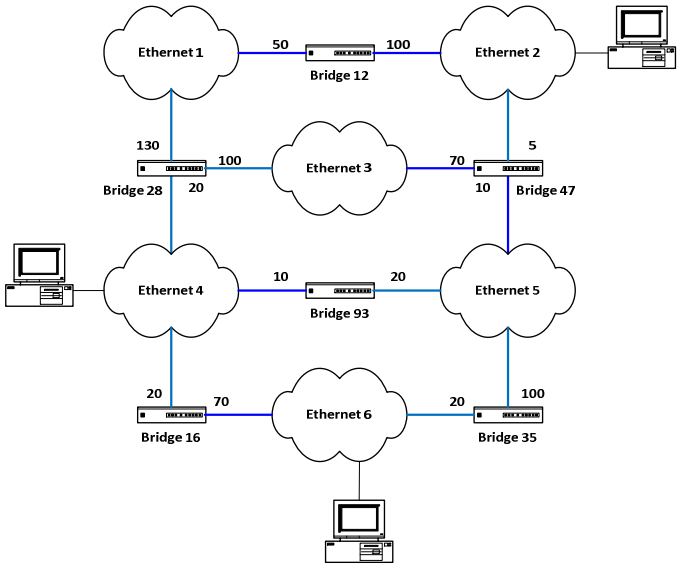
Basic Spanning Tree Concepts



Spanning Tree Example

Designated bridges

LAN	Bridge
Eth1	
Eth 2	
Eth 3	
Eth 4	
Eth 5	
Eth 6	

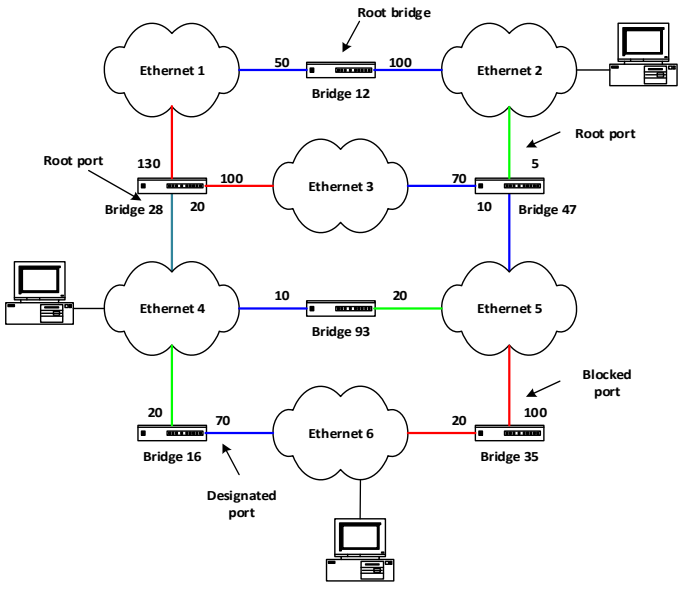


156

Spanning Tree Example

Designated bridges

LAN	Bridge
Eth1	12
Eth 2	12
Eth 3	47
Eth 4	93
Eth 5	47
Eth 6	16



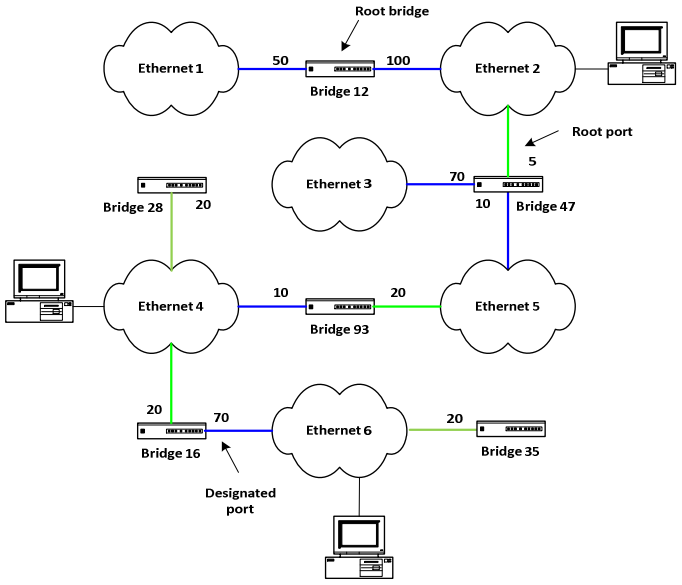
157

Spanning Tree Example

Designated bridges

LAN Bridge

Eth1	12
Eth 2	12
Eth 3	47
Eth 4	93
Eth 5	47
Eth 6	16



BPDUs (Bridge Protocol Data Units)

- To build and maintain the spanning tree, the bridges exchange special messages among them, known as BPDUs
- There exist two such types of messages:
  - ***Configuration*** and ***Topology Change Notification***

```
802.2: Address: 00-E0-B0-64-48-77 --->01-80-C2-00-00-00
LLC: Sap 0x42 ---> 0x42 (Command)
Bridge Protocol Data Unit
  Protocol ID: 0
  Version: 0
  Type: 0 (Configuration)
  Flags: 0x00
  Root ID: 40000800021EA1F1
    Priority: 0x4000
    MAC Address: 08-00-02-1E-A1-F1
  Root Path Cost: 100
  Bridge ID: 800000E0B0644876
    Priority: 0x8000
    MAC Address: 00-E0-B0-64-48-76
  Port ID: 0x8003
  Message Age: 1
  Max Age: 20
  Hello Time: 2
  Forward Delay: 15
  Frame Padding : (8 bytes)
Calculate CRC: 0xb209dfee
```

destination	source	DSAP	SSAP	BPDU

- Destination – **multicast address assigned to every bridge**
- Source – MAC address of the port sending the BPDU
- DSAP = SSAP = 01000010 = 42 hex

Configuration BPDUs

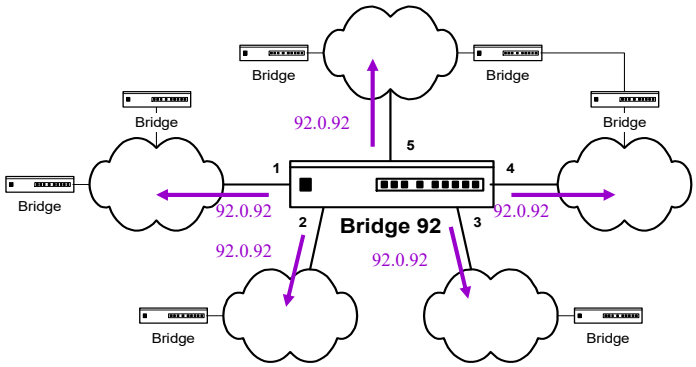
- The spanning tree configuration is done using Conf - BPDUs (configuration messages)

```
802.2: Address: 00-E0-B0-64-48-77 --->01-80-C2-00-00-00
LLC: Sap 0x42 ---> 0x42 (Command)
Bridge Protocol Data Unit
  Protocol ID: 0
  Version: 0
  Type: 0 (Configuration)
  Flags: 0x00
  Root ID: 40000800021EA1F1
    Priority: 0x4000
    MAC Address: 08-00-02-1E-A1-F1
  Root Path Cost: 100
  Bridge ID: 800000E0B0644876
    Priority: 0x8000
    MAC Address: 00-E0-B0-64-48-76
  Port ID: 0x8003
  Message Age: 1
  Max Age: 20
  Hello Time: 2
  Forward Delay: 15
  Frame Padding : (8 bytes)
Calculate CRC: 0xb209dfee
```

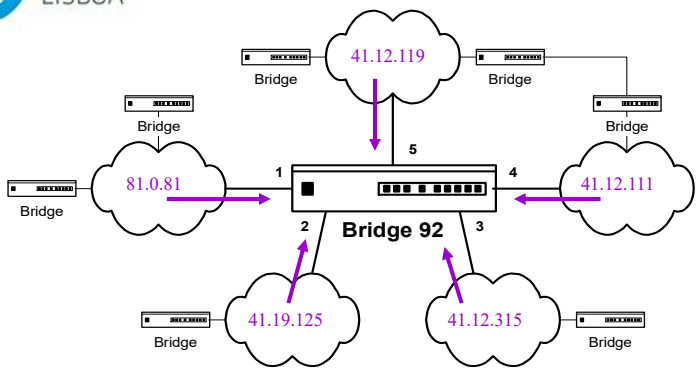
- More important fields:
  - Root ID: present estimate of the root bridge address
  - Root Path Cost: present estimate of the path cost to the root bridge
  - Bridge ID: address of the bridge sending the configuration message
  - Port ID: address of the port sending the configuration message

Spanning Tree Construction

- Each bridge initially assumes it is the root bridge (making Root Path Cost = 0); it sends configuration messages in all its ports



Spanning Tree Construction



Best received messages in bridge 92 so far

Bridge 92 estimates:

- Root bridge = 41
- Root port = 4
- Root path cost = 12 + 1

Configuration Message Order

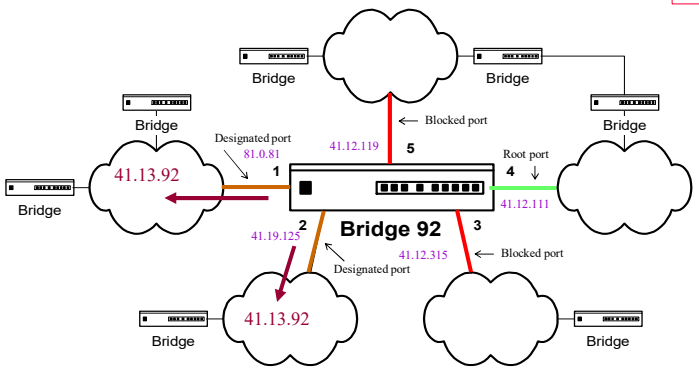
- A configuration message C_1 is said to be "better" than C_2 if:
  - C_1 Root ID is lower than that of C_2
  - if the Root IDs are identical, the Root Path Cost of C_1 is lower than that of C_2
  - being identical the Root ID and Root Path Cost, the Bridge ID of C_1 is lower than that of C_2
  - being identical the Root ID, Root Path Cost and Bridge ID, the Port ID of C_1 is lower than that of C_2

Root ID	Root Path Cost	Bridge ID	Port ID
41	12	111	2
41	12	111	4
41	12	119	1
41	19	125	3
81	0	81	2

TÉCNICO LISBOA

Spanning Tree Construction

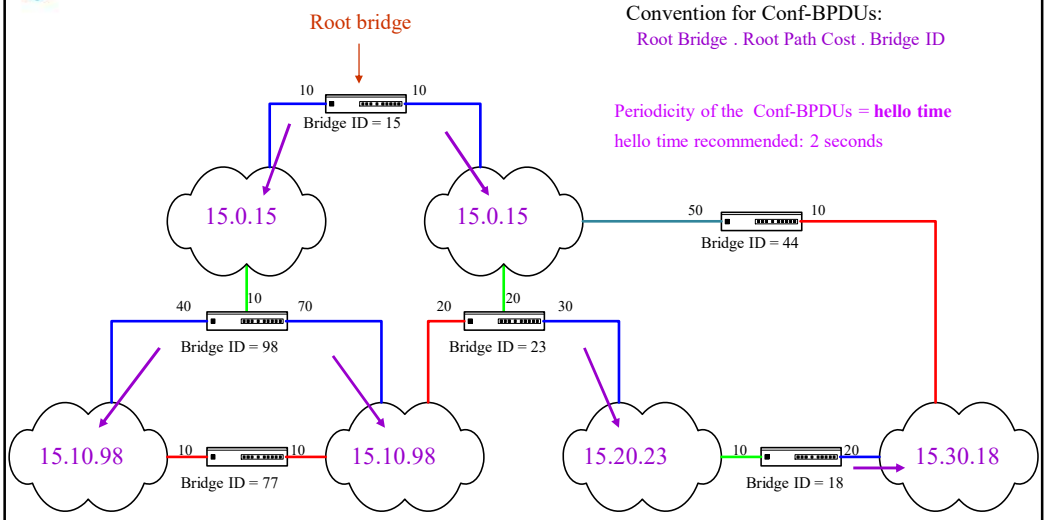
TPC: Prob. 16

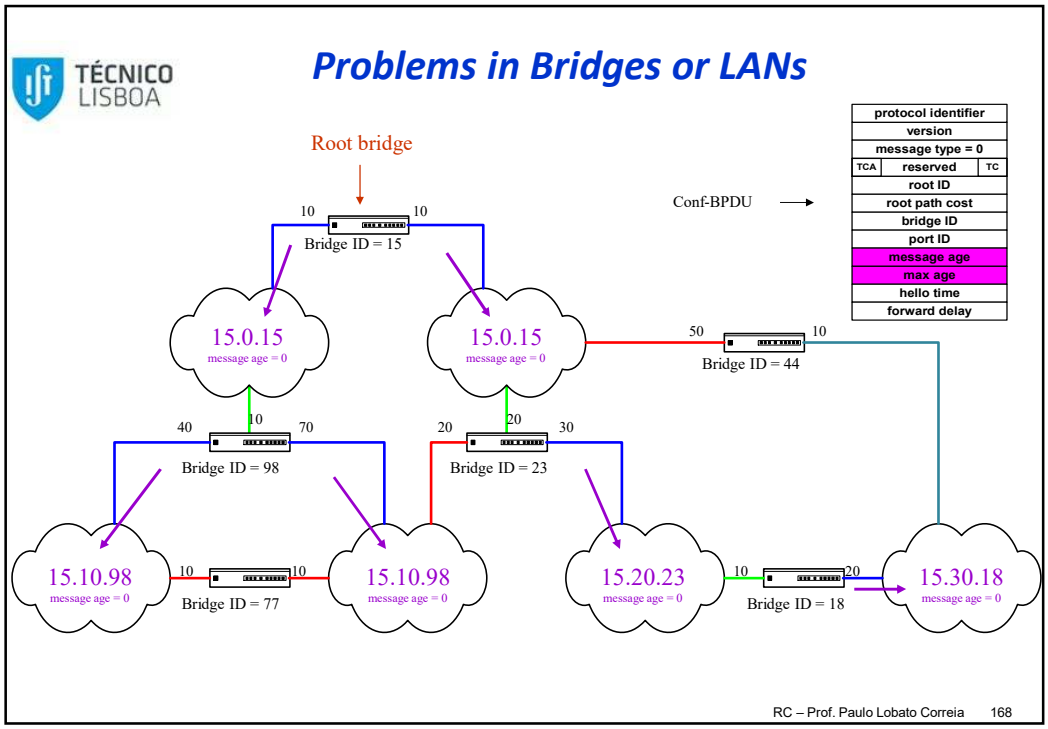


messages sent by Bridge 92 - 41.13.92

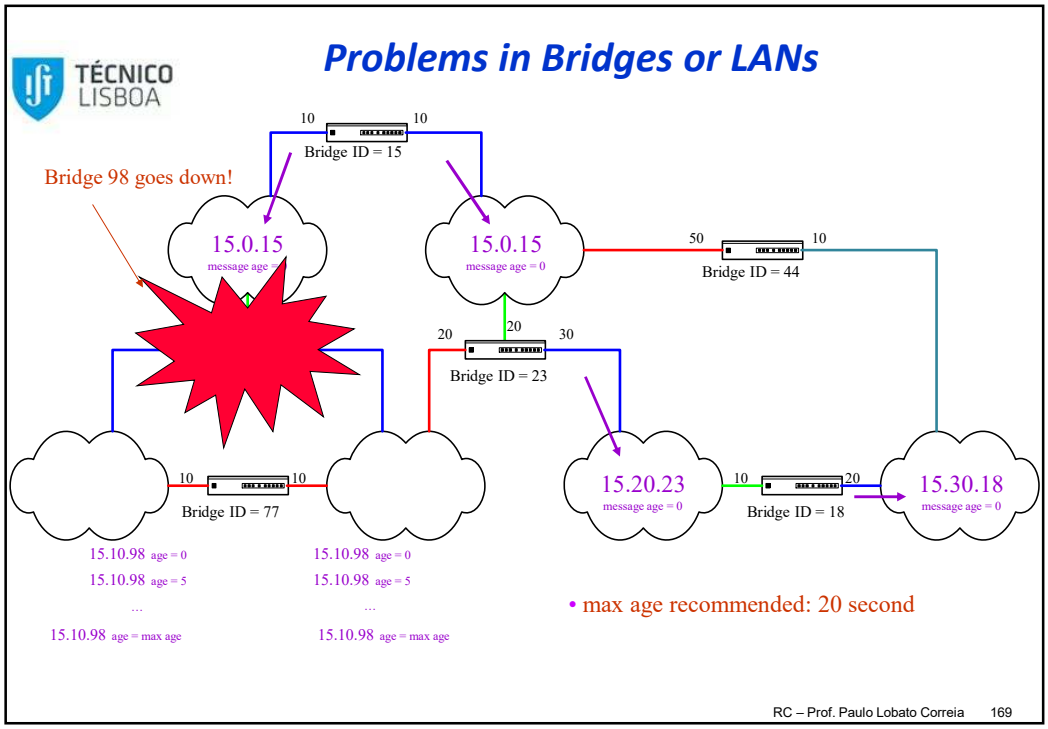
TÉCNICO LISBOA

Spanning Tree Maintenance

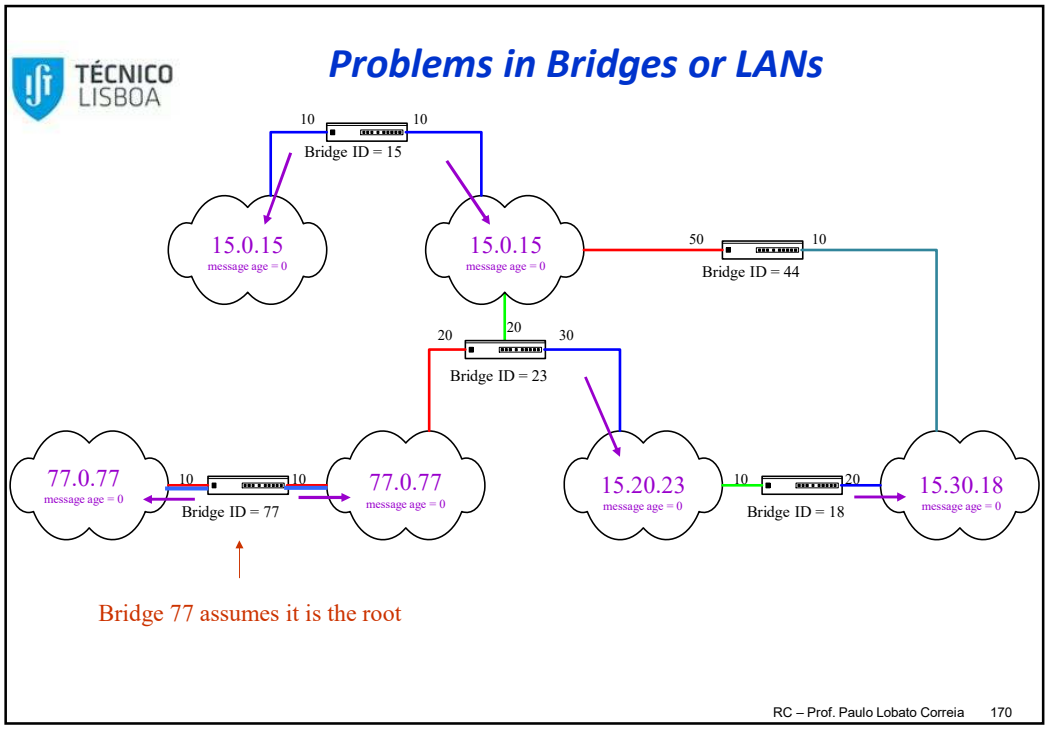



TÉCNICO LISBOA
Problems in Bridges or LANs
Root bridge
Conf-BPDU
protocol identifier
version
message type = 0
TCA
reserved
TC
root ID
root path cost
bridge ID
port ID
message age
max age
hello time
forward delay
Bridge ID = 15
Bridge ID = 44
Bridge ID = 98
Bridge ID = 23
Bridge ID = 77
Bridge ID = 18
15.0.15
message age = 0
15.10.98
message age = 0
15.20.23
message age = 0
15.30.18
message age = 0
RC – Prof. Paulo Lobato Correia 168

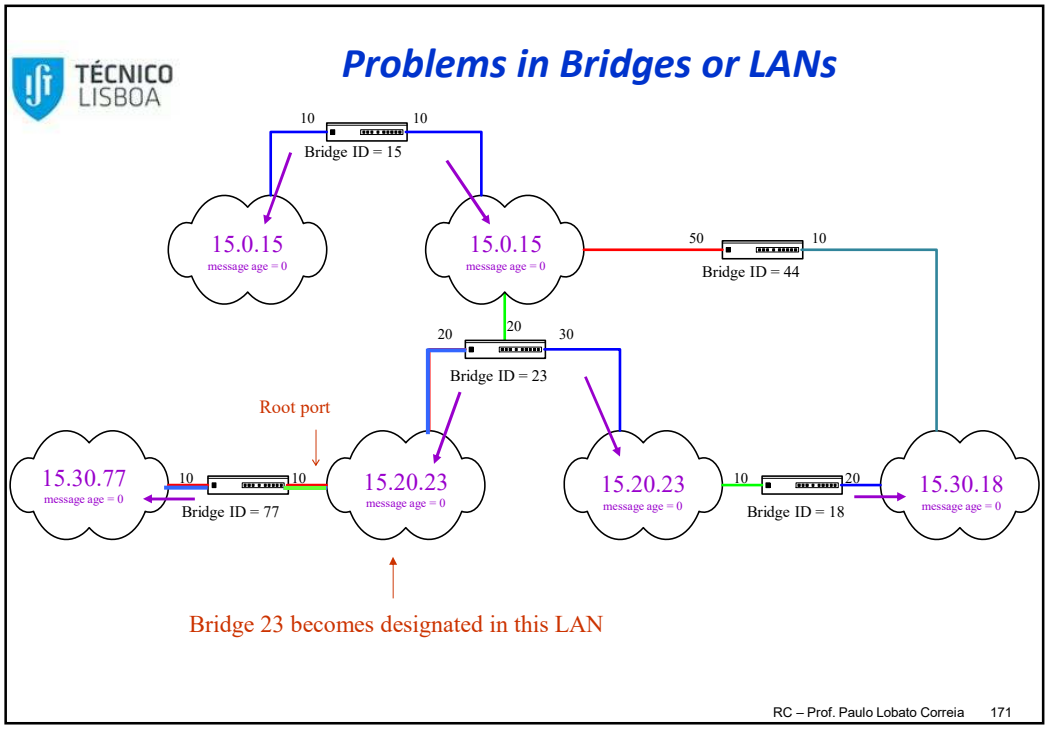
168


TÉCNICO LISBOA
Problems in Bridges or LANs
Bridge 98 goes down!
Bridge ID = 15
Bridge ID = 44
Bridge ID = 23
Bridge ID = 77
Bridge ID = 18
15.0.15
message age = 0
15.20.23
message age = 0
15.30.18
message age = 0
15.10.98 age = 0
15.10.98 age = 5
...
15.10.98 age = max age
15.10.98 age = 0
15.10.98 age = 5
...
15.10.98 age = max age
• max age recommended: 20 second
RC – Prof. Paulo Lobato Correia 169

169


TÉCNICO LISBOA
Problems in Bridges or LANs
10
10
Bridge ID = 15
15.0.15
message age = 0
15.0.15
message age = 0
50
10
Bridge ID = 44
20
20
30
Bridge ID = 23
77.0.77
message age = 0
10
10
Bridge ID = 77
77.0.77
message age = 0
15.20.23
message age = 0
10
20
Bridge ID = 18
15.30.18
message age = 0
Bridge 77 assumes it is the root
RC – Prof. Paulo Lobato Correia 170

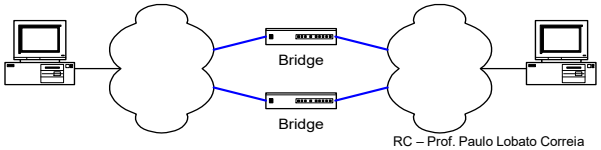
170


TÉCNICO LISBOA
Problems in Bridges or LANs
10
10
Bridge ID = 15
15.0.15
message age = 0
15.0.15
message age = 0
50
10
Bridge ID = 44
20
20
30
Bridge ID = 23
Root port
15.30.77
message age = 0
10
10
Bridge ID = 77
15.20.23
message age = 0
15.20.23
message age = 0
10
20
Bridge ID = 18
15.30.18
message age = 0
Bridge 23 becomes designated in this LAN
RC – Prof. Paulo Lobato Correia 171

171

Temporary Cycles/Connectivity Loss

- After the network topology change:
  - There may be a temporary loss of connectivity if one previously blocked port has not yet "realized" that it should become active in the new topology
  - There may be temporary cycles if a previously active port has not yet "realized" that it should become blocked in the new topology
- To minimize the probability of temporary cycles appearing, the bridges must wait some time before allowing one of their ports moving from the blocked state to an active state; the waiting time is controlled by the parameter: forward delay



Configuration Messages Format

# bytes			
2	protocol identifier		
1	version		
1	message type		
1	TCA	reserved	TC
8	root ID		
4	cost of path to root		
8	bridge ID		
2	port ID		
2	message age		
2	max age		
2	hello time		
2	forward delay		

TÉCNICO LISBOA

An Algorithm for Distributed
Computation of a Spanning Tree
in an Extended LAN

Radia Perlman

Digital Equipment Corporation
1925 Andover St., Tewksbury MA 01876

Abstract

A protocol and algorithm are given in which bridges in an extended Local Area Network of arbitrary topology compute, in a distributed fashion, an acyclic spanning subset of the network.

The algorithm converges in time proportional to the diameter of the extended LAN, and requires a very small amount of memory per bridge, and communications bandwidth per LAN, independent of the total number of bridges or the total number of links in the network.

Algorhyme

I think that I shall never see
A graph more lovely than a tree.

A tree whose crucial property
Is loop-free connectivity.

A tree which must be sure to span
So packets can reach every LAN.

First the Root must be selected.
By ID it is elected.

Least cost paths from Root are traced.
In the tree these paths are placed.

A mesh is made by folks like me
Then bridges find a spanning tree.

Introduction

Local area networks are limited in geography, traffic, and number of stations. A single local area network will often not meet the needs of an organization for these reasons. Conventional LAN interconnection mechanisms, for instance XNS [1], Sytek[2], IBM1[3], IBM2[4], DNA[5] require cooperation from the stations with compatible protocols layered above the protocol necessary to connect to a single LAN.

An approach that is transparent to stations, and thus allows a station to participate in an extended LAN with no modification, is presented in [6],[7], and [8]. In this approach, a bridge connected to two or more links will listen "promiscuously" to all packets transmitted on each of its links, and forward packets received on one of the links onto the others. A bridge also learns of the location of stations relative to itself, so that it will not forward traffic for a station onto a link unnecessarily.

This approach assumes that the topology is a tree (loop-free). However, requiring a topology to be loop-free means there are no backup paths in the case of bridge or LAN failures. Also, because the technology allows network growth so easily, it might be difficult to prevent someone from adding a bridge and creating a loop. A loop in the topology might cause severe performance degradation in the entire extended network due to congestion caused by infinitely circulating packets. It is undesirable to have a network that can be brought down so easily, merely by plugging a cable into the wrong place.

183

TÉCNICO LISBOA

An Algorithm for Distributed
Computation of a Spanning Tree
in an Extended LAN

Radia Perlman

Algorhyme

I think that I shall never see
A graph more lovely than a tree.

A tree whose crucial property
Is **loop-free connectivity**.

A tree that must be sure to span
So packets can reach every LAN.

First the **root must be selected**.
By ID, it is elected.

Least-cost paths from root are traced.
In the tree, these paths are placed.

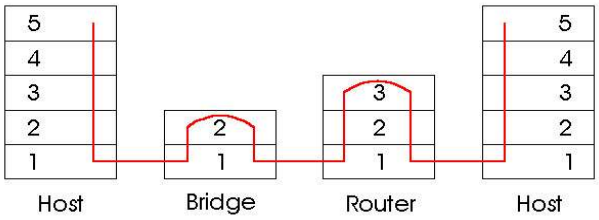
A mesh is made by folks like me,
Then the bridge finds a **spanning tree**.

184

TÉCNICO LISBOA

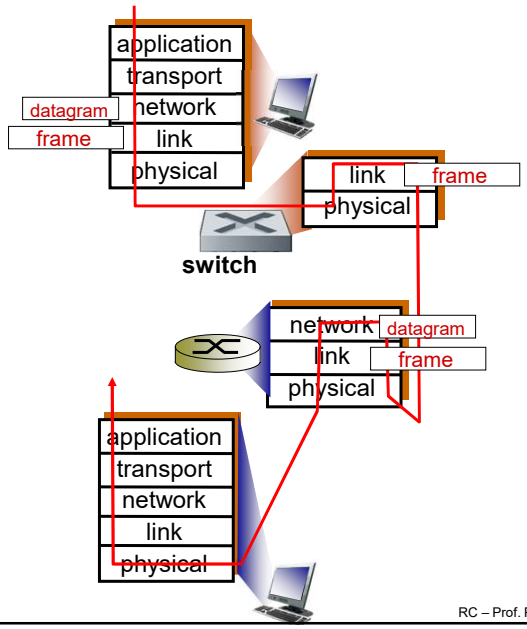
Switches vs. Routers

- Both store-and-forward devices:
  - Routers: network layer devices (examine network layer headers);
  - Switches are link layer devices.
- Routers maintain routing tables, implement routing algorithms;
- Switches maintain switch tables, implement filtering, learning algorithms.



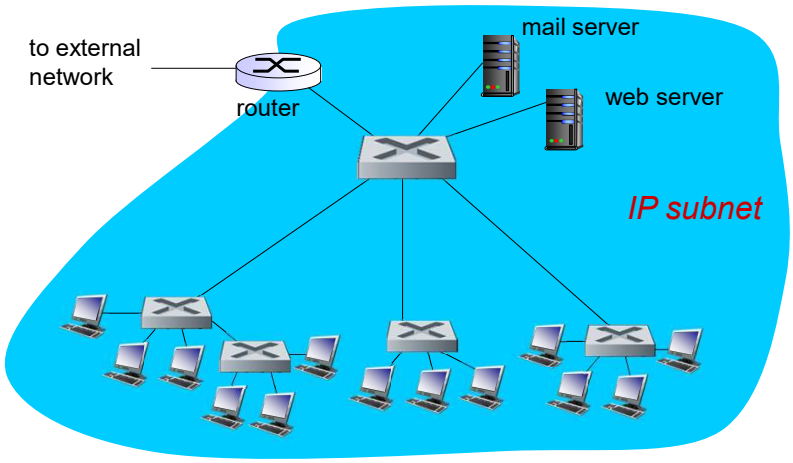
TÉCNICO LISBOA

Switches vs. Routers



TÉCNICO LISBOA

Institutional Network



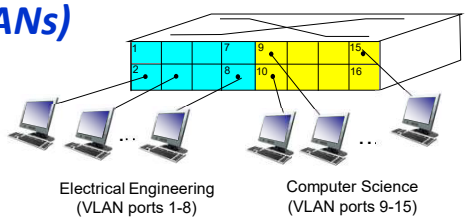
187

TÉCNICO LISBOA

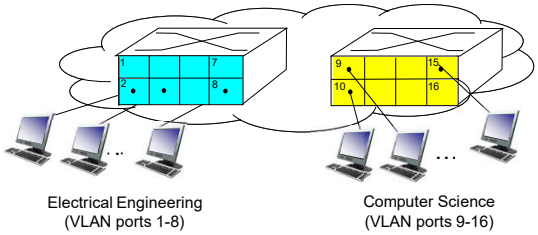
Virtual LANs (VLANs)

Virtual Local Area Network

*switch(es) supporting VLAN capabilities can be configured to define multiple **virtual** LANS over a single physical LAN infrastructure.*



Port-based VLAN: switch ports grouped (by switch management software) so that one single physical switch operates as multiple virtual switches

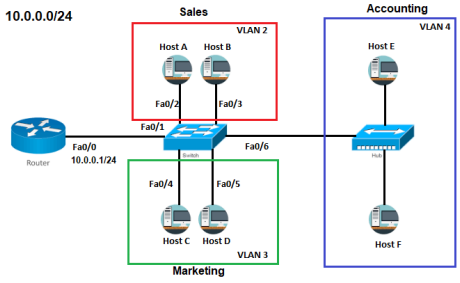


188

TÉCNICO LISBOA

Virtual LANs (VLANs)

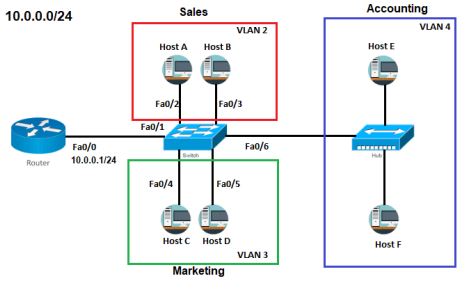
- A VLAN is a type of local area network that does not have its own dedicated physical infrastructure.
  Instead, it uses another LAN physical infrastructure to carry its traffic.
- The traffic is encapsulated so that **a number of logically separate VLANs can be carried by the same physical LAN**.



TÉCNICO LISBOA

Virtual LANs (VLANs)

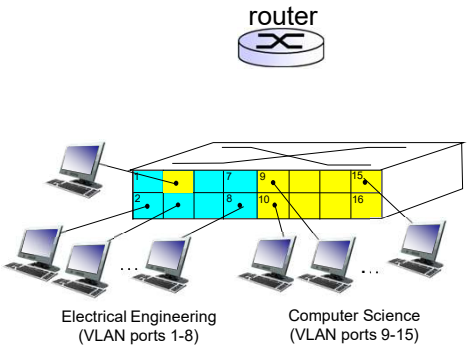
- A VLAN divides a physical LAN into parts, and breaks the network into **different broadcast domains** (with a limited, reasonable range).
- Limiting the broadcast can prevent switches from wasting bandwidth to forward messages to unnecessary ports.
- Since **hosts in different VLANs cannot communicate with each other** directly, VLAN can also provide increased security.



TÉCNICO LISBOA

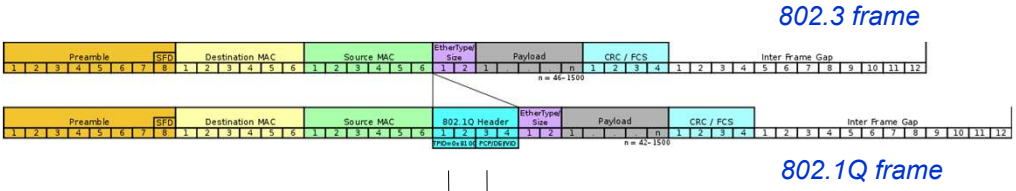
Port-based VLAN

- **Traffic isolation**: frames to/from ports 1-8 can only reach ports 1-8
  - can also define VLAN based on MAC addresses of endpoints, rather than switch port
- **Dynamic membership**: ports can be dynamically assigned among VLANs
- **Forwarding between VLANS**: done via routing (just as with separate switches)
  - in practice vendors sell combined switches plus routers



TÉCNICO LISBOA

802.1Q VLAN Frame Format



2-byte Tag Protocol Identifier (TPID) with value 0x8100

Tag Control Information (TCI):
- Priority code point (PCP) – 3 bit priority field (like IP TOS)
- Drop eligible indicator (DEI) – 1 bit: eligible to drop if congestion
- VLAN identifier (VID) – **12 bit VLAN ID field**

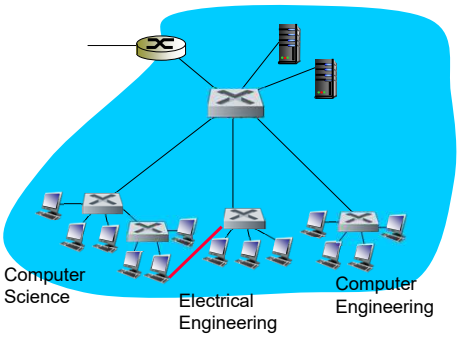
VID = 0x000 indicates that the frame does not carry a VLAN ID

Virtual LANs (VLANs)

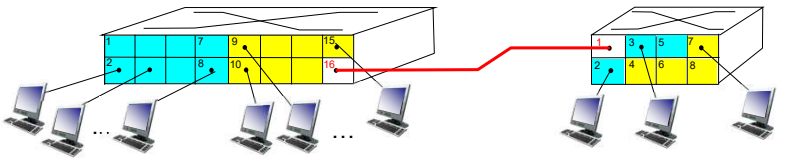
What if CS user moves office to EE, but wants connect to CS switch?

In fact, there should be a single broadcast domain:

- all layer-2 broadcast traffic (ARP, DHCP, broadcast to unknown destinations) must cross the entire LAN
- security/privacy, efficiency issues



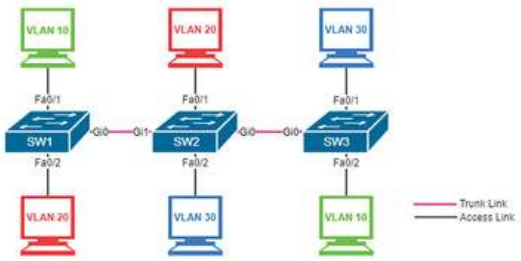
VLANS Spanning Multiple Switches



- **Trunk port**: carries frames between VLANS defined over multiple physical switches
  - frames forwarded within VLAN between switches can't be regular 802.3 frames (must carry VLAN ID info)
  - 802.1q protocol adds/removes additional header fields for frames forwarded between trunk ports

VLANS Spanning Multiple Switches

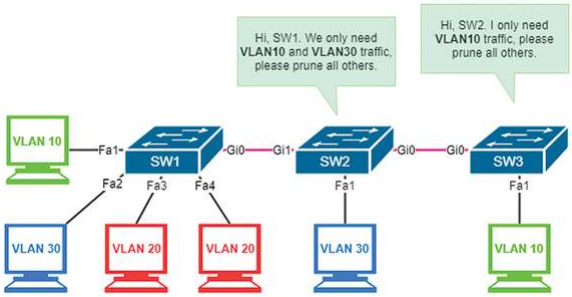
- A host can be assigned to any VLAN regardless of geographic location, as long as the ***switches are connected by trunk links***.
- VLANs need to be created on all switches on the path through which VLAN traffic passes.
  - Figure: VLAN20 created on SW1 and SW2 (for red hosts to communicate), VLAN 30 on SW2 and SW3 (for blue hosts to communicate) and VLAN10 on the three switches (for green hosts to communicate).



VLANS Spanning Multiple Switches

Pruning:

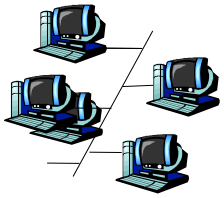
- The pruning feature can automatically block unnecessary traffic that passes through trunk links, to avoid wasting bandwidth.
  - Figure: SW3 only needs VLAN10 traffic → SW3 notifies SW2 to prune all other VLAN traffic.
    SW2 only needs VLAN30 traffic → it combines with SW3 request and tells SW1 to prune all VLANs other than VLAN10 and VLAN30.



TÉCNICO LISBOA

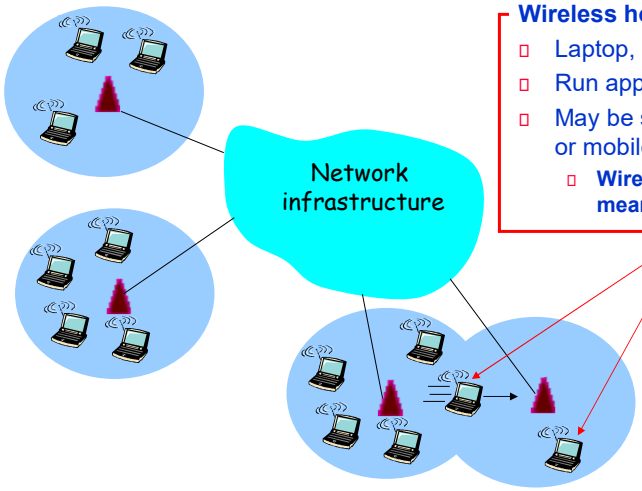
Outline

- Link-layer Addressing
- Introduction and services
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing



TÉCNICO LISBOA

Elements of a Wireless Network

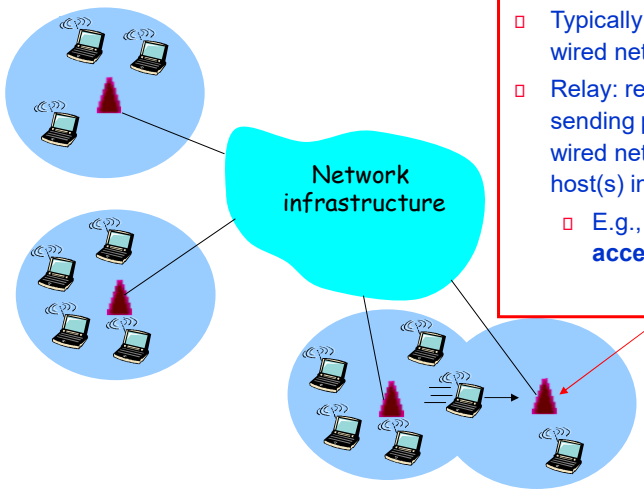


Wireless hosts:

- Laptop, Smartphone, Tablet, …;
- Run applications;
- May be stationary (non-mobile) or mobile:
  - **Wireless does *not* always mean mobility!**

TÉCNICO LISBOA

Elements of a Wireless Network



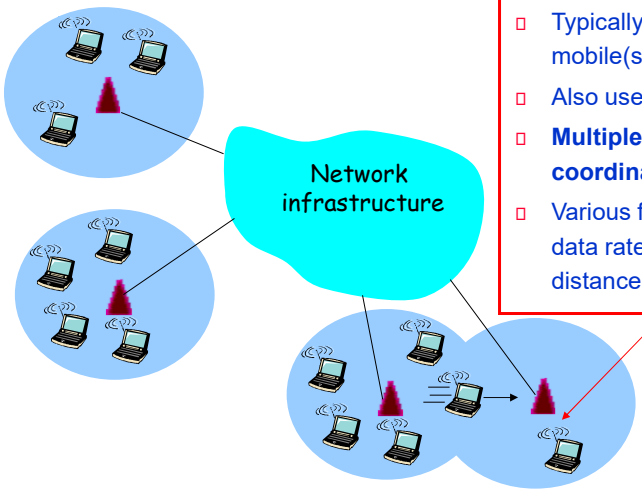
Base station:

- Typically connected to the wired network;
- Relay: responsible for sending packets between wired network and wireless host(s) in its "area":
  - E.g., cell towers, 802.11 **access points**.

200

TÉCNICO LISBOA

Elements of a Wireless Network



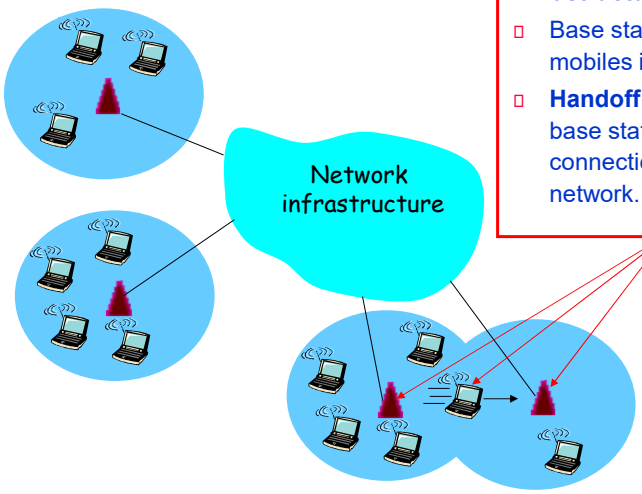
Wireless link:

- Typically used to connect mobile(s) to base station;
- Also used as backbone link;
- **Multiple access protocol coordinates link access**;
- Various frequency bands, data rates, transmission distance.

201

TÉCNICO LISBOA

Elements of a Wireless Network



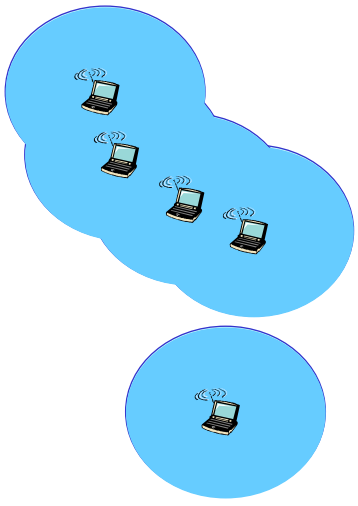
Infrastructure mode:

- Base station connects mobiles into wired network;
- **Handoff**: mobile changes base station providing connection into wired network.

202

TÉCNICO LISBOA

Elements of a Wireless Network



Ad hoc mode:

- **No base stations**;
- Nodes can only transmit to other nodes within link coverage;
- Nodes organize themselves into a network: route among themselves.

203

TÉCNICO LISBOA

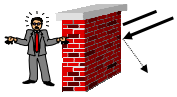
Wireless Communications

Wireless communications are needed to:

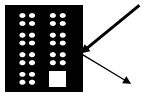
- Allow communications while moving;
- Allow communications in places where it is difficult, or impossible, to implement a cabled infrastructure;
- Allow broadcasting;
- Allow the fast implementation, with low initial cost, of a communications system;

However:

- Less controlled operation environment, more subject to interference, noise, unauthorized detection;
- Often provides lower transmission rates;
- Frequency bands are easier to reuse in guided media.



Shadow areas



Reflexions

Dispersion

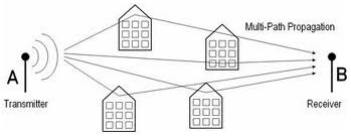
Difraction

TÉCNICO LISBOA

Wireless Link Characteristics

Differences from wired link:

- Interference from other sources:
  - Standardized wireless network frequencies (e.g., 2.4 GHz) shared by other devices (e.g., phone);
  - Devices (motors) interfere as well;
- Multipath propagation:
  Radio signal reflects off objects and ground, arriving at destination at slightly different times;



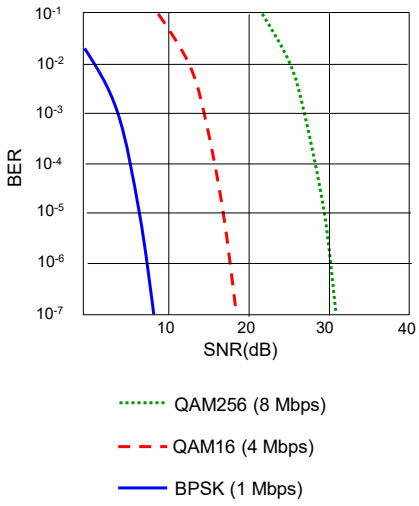
- Decreased signal strength:
  Radio signal attenuates as it propagates through matter (path loss).

…. make communication across (even a point to point) wireless link much more “difficult”!

TÉCNICO LISBOA

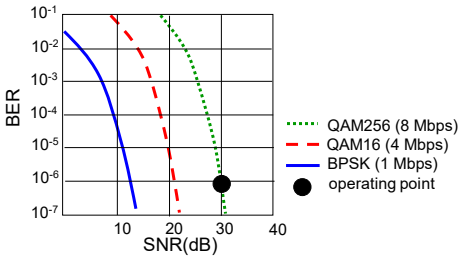
Wireless Link Characteristics

- Signal-to-noise ratio (SNR):
  - Larger SNR – easier to extract signal from noise (a "good thing");
- ***SNR versus BER** tradeoffs:*
  - *Given physical layer:* increase power → increase SNR → decrease bit error rate (BER);
  - *Given SNR:* choose physical layer that meets BER requirement, giving highest throughput:
    - SNR may change with mobility: **dynamically adapt physical layer (modulation technique, rate)**.



TÉCNICO LISBOA

802.11: Advanced Capabilities



Rate Adaptation:

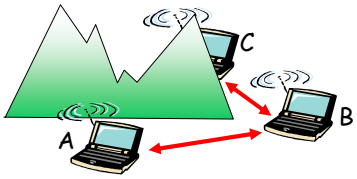
- Base station, mobile:
  - Dynamically change transmission rate (physical layer modulation technique) as mobile moves and SNR varies.

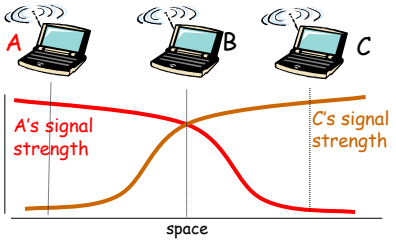
1. SNR decreases, BER increase as node moves away from base station
2. When BER becomes too high, switch to lower transmission rate but with lower BER

TÉCNICO LISBOA

Wireless Network Characteristics

Multiple wireless senders and receivers create additional problems (beyond multiple access):





Hidden terminal problem:
- B, A hear each other;
- B, C hear each other;
- A, C can not hear each other;

→ **A, C unaware of their interference at B.**

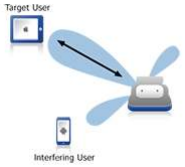
Signal attenuation:
- B, A hear each other;
- B, C hear each other;
- A, C can not hear each other interfering at B.

TÉCNICO LISBOA

IEEE 802.11 Wireless LAN

Standard	Maximum Speed	Frequency	Backwards Compatible
802.11a	54 Mbps	5 GHz	No
802.11b	11 Mbps	2.4 GHz	No
802.11g	54 Mbps	2.4 GHz	802.11b
802.11n	600 Mbps	2.4 GHz or 5 GHz	802.11b/g
802.11ac	1.3 Gbps (1300 Mbps)	2.4 GHz and 5.5 GHz	802.11b/g/n
802.11ad	7 Gbps (7000 Mbps)	2.4 GHz, 5 GHz and 60 GHz	802.11b/g/n/ac

- 802.11b
  - 2.4 GHz unlicensed spectrum; up to 11 Mbps.
- 802.11a
  - 5 GHz band; up to 54 Mbps.
- 802.11g
  - 2.4 GHz band; up to 54 Mbps.
- 802.11n (multiple antennae)
  - 2.4, 5 GHz band (40 MHz channel); up to 600 Mbps.
- 802.11ac (multiple antennae)
  - 5 GHz band (160 MHz channel); up to 1.3 Gbps; up to 3.5 Gbps (one user).
- 802.11ad – WiGig (60 GHz – up to 10 m)
  - 2.4, 5, 60 GHz (80 or 160 MHz channel); up to 6.7 Gbps.



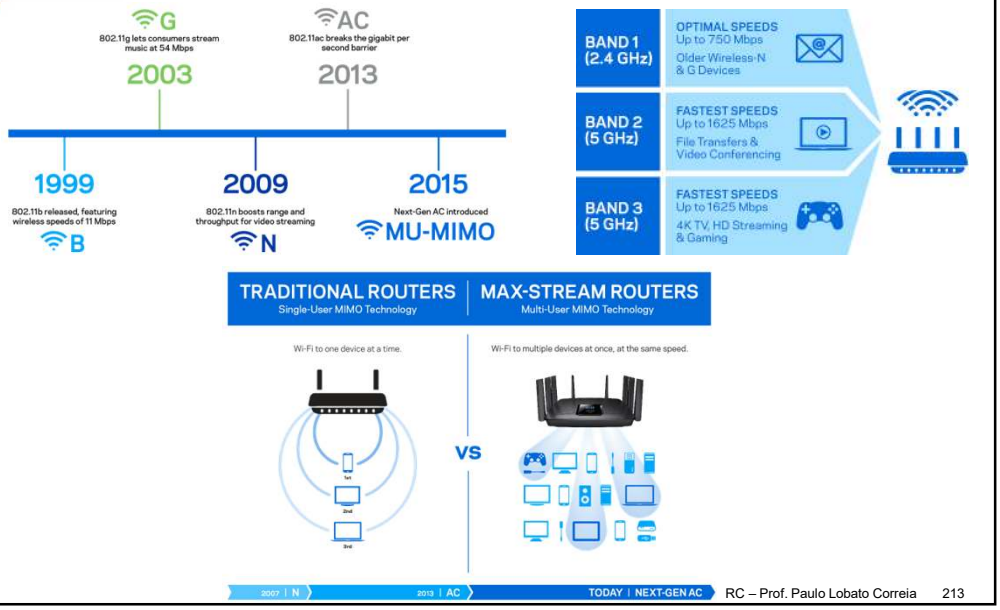
- All use CSMA/CA for multiple access;
- All have base-station and ad-hoc network versions.

IEEE 802.11 Wireless LAN

- 802.11ah (HaLow) – May 2017
  - <1, 2.4, 5 GHz; up to 347 Mbps.
- 802.11aj
  - To operate in China in bands 45, 59-64 GHz.
- 802.11ak
  - Emphasis on standardized security and quality-of-service improvements.
- 802.11ax (2019)
  - 2.4, 5 GHz band. High Efficiency WLAN; 4x faster than 802.11 n or ac.
- 802.11ay (2019)
  - Next Generation 60 GHz; > 20 Gbit/s.
- 802.11az (2021)
  - Next Generation Positioning (NGP) – emphasis on determining the absolute and relative position of stations.
- 802.11ba (2020)
  - Wake-Up Radio (WUR) – emphasis on extending battery life.

212

IEEE 802.11 Wireless LAN



213

TÉCNICO LISBOA

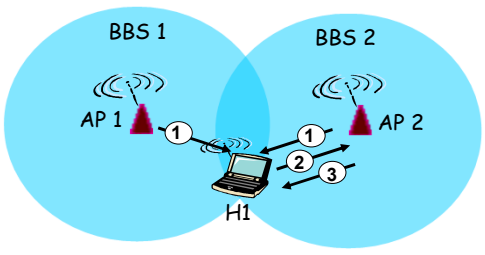
802.11: Channels, Association

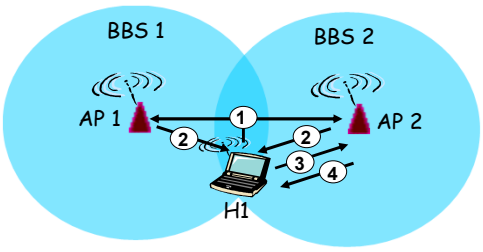
802.11b:

- 2.4GHz - 2.485GHz spectrum divided into 11 channels at different frequencies:
  - AP administrator chooses frequency for AP;
  - Interference possible:
    Channel can be same as that chosen by neighboring AP !
- Host – must *associate* with an AP:
  - Scans channels, listening for *beacon frames* containing APs names (SSID) and MAC addresses;
  - Selects AP to associate with;
  - May perform authentication;
  - Will typically run DHCP to get IP address in AP's subnet.

TÉCNICO LISBOA

802.11: Passive/Active Scanning





Passive Scanning:

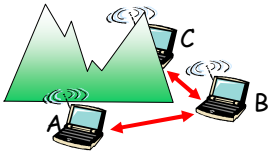
(1) Beacon frames sent from APs;
(2) Association Request frame sent: H1 to selected AP;
(3) Association Response frame sent: H1 to selected AP.

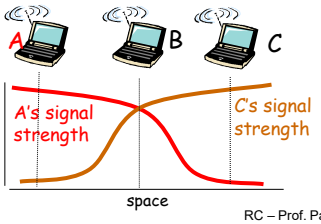
Active Scanning:

(1) Probe Request frame broadcast from H1;
(2) Probe Response frame sent from APs;
(3) Association Request frame sent: H1 to selected AP;
(4) Association Response frame sent: H1 to selected AP.

IEEE 802.11: Multiple Access

- 802.11: CSMA - sense channel before transmitting
  - Don't collide with ongoing transmission by other node.
- 802.11: no collision detection!
  - Difficult to receive (sense collisions) when transmitting due to weak received signals (fading);
  - Can't sense all collisions in any case: hidden terminal, fading;
  - **Goal: *avoid collisions:* CSMA/CA (Collision Avoidance).**





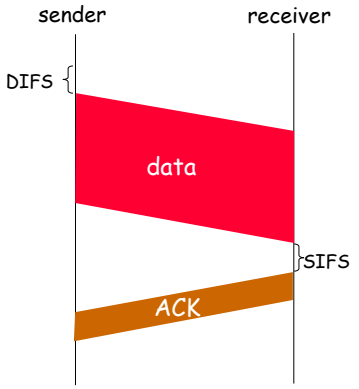
IEEE 802.3 (Ethernet) MAC Protocol: CSMA/CD

1. NIC receives datagram from network layer → creates frame;
2. If **channel idle** (96 bit times), starts frame transmission;
   If **channel busy**, waits until channel idle, then transmits
3. If NIC transmits entire frame without detecting another transmission:
   → success;
4. If NIC detects another transmission while transmitting - collision:
   → aborts and sends jam signal (reinforce collision);
5. After aborting, NIC enters **exponential backoff**:
   → After m^{th} collision, NIC chooses K at random from $\{0,1,2,\ldots,2^m-1\}$;
   → NIC waits K x 512 bit times; then returns to Step 2.

IEEE 802.11 MAC Protocol: CSMA/CA

802.11 sender

1. If sense **channel idle for DIFS** then:
   - Transmit entire frame (no CD);
2. If sense **channel busy** then:
   - Start random backoff time;
   - Timer counts down **only** when channel is idle;
   - Transmit entire frame when timer expires;
   - If no ACK, increase random backoff interval, then repeat from 2.



802.11 receiver

- If frame **received OK**:
  - Return **ACK after SIFS**

(*ACK is needed to confirm the success of a transmission*; *it also helps with the hidden terminal problem*).

Collision Avoidance Adding RTS/CTS

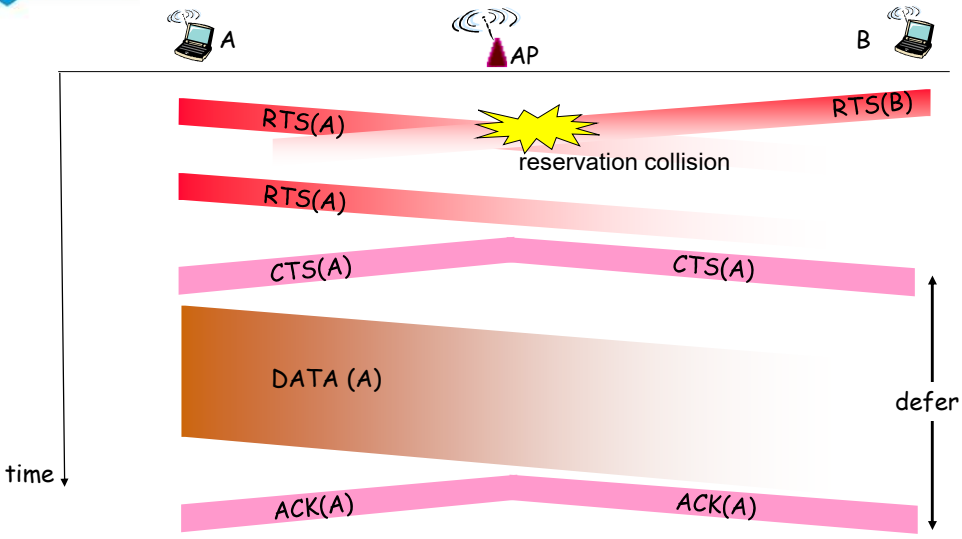
Idea:

Allow sender to **"reserve" channel** rather than random access of data frames → avoid collisions of long data frames;

- Sender first transmits *small* **request-to-send (RTS)** packets to BS using CSMA:
  - RTSs may still collide with each other (but they're short);
- BS broadcasts **clear-to-send CTS** in response to RTS;
- CTS heard by all nodes:
  - Sender transmits data frame;
  - Other stations defer transmissions.

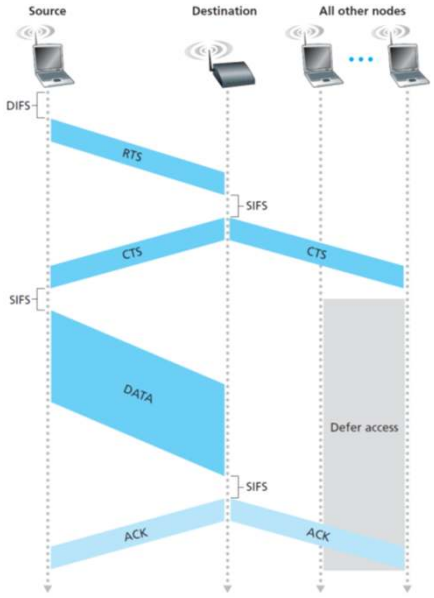
Avoid data frame collisions completely using small reservation packets!

Collision Avoidance: Adding RTS-CTS



222

Collision Avoidance: Adding RTS-CTS

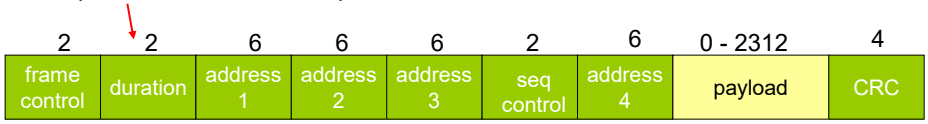


223

TÉCNICO LISBOA

802.11 Frame: Addressing

Duration: **to "reserve" channel**
(for transmission + ACK)



Address 1: MAC address of wireless host or AP to receive this frame (wireless destination addr)

Address 2: MAC address of wireless host or AP transmitting this frame (sender addr)

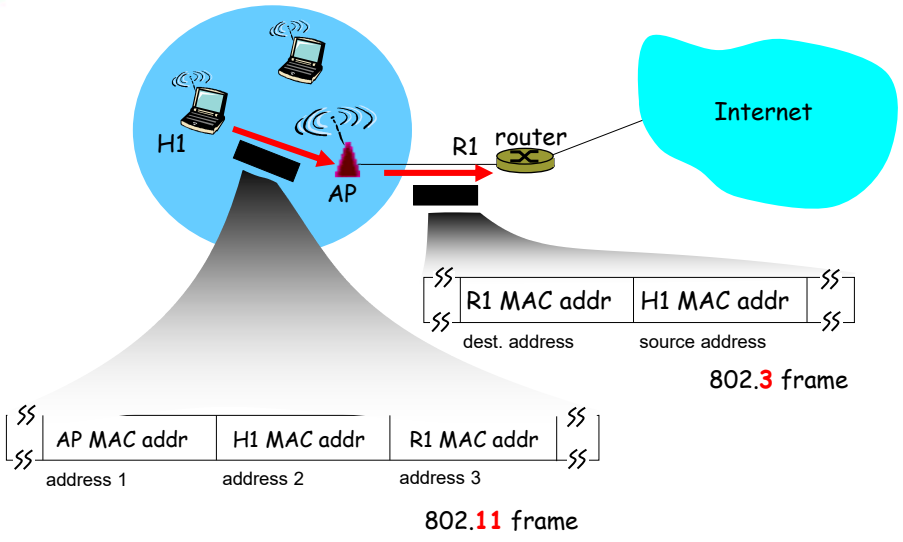
Address 3: MAC address of router interface to which AP is attached (router addr)

Address 4: used only in ad hoc mode

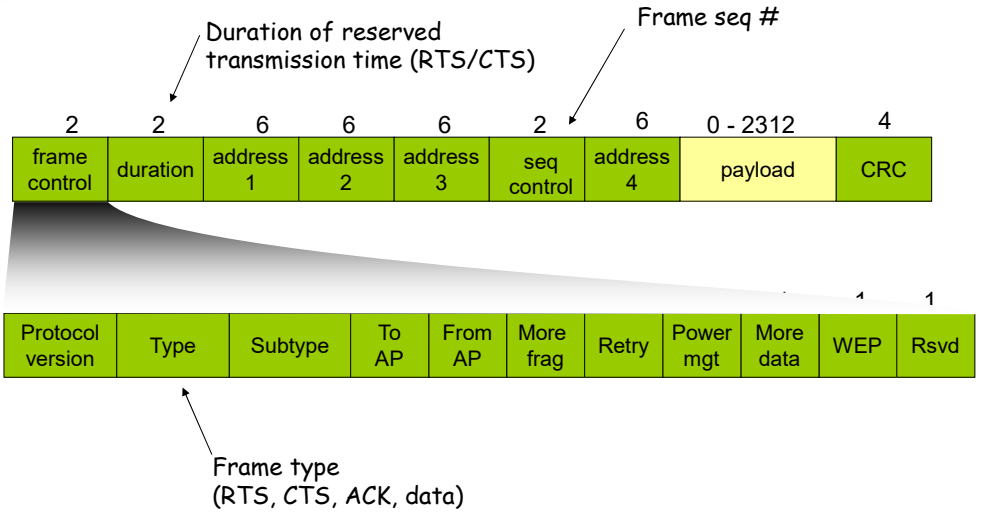
TÉCNICO LISBOA

802.11 Frame: Addressing

TPC: Prob. 17



802.11 Frame: more



802.11: Advanced Capabilities

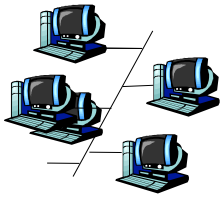
Power Management:

- Node-to-AP: "I am going to sleep until next beacon frame"
  - AP knows not to transmit frames to this node;
  - Node wakes up before next beacon frame;
- Beacon frame: contains list of mobiles with AP-to-mobile frames waiting to be sent:
  - Node will stay awake if AP-to-mobile frames are to be sent; Otherwise sleep again until next beacon frame.

TÉCNICO LISBOA

Outline

- Introduction and services
- Link-layer Addressing
- Error detection and correction
- Multiple access protocols
- Ethernet
- Link-layer switches
- IEEE 802.11 Wireless LANs
- Framing

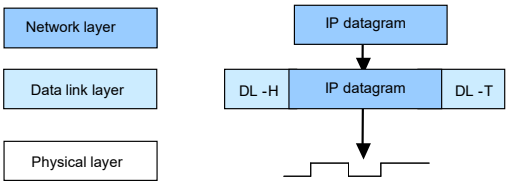


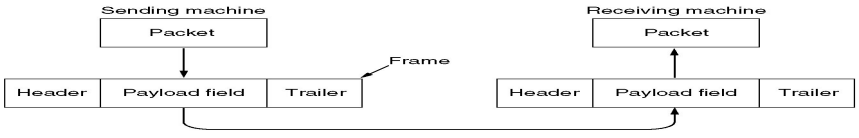
TÉCNICO LISBOA

Framing

Data Link layer:

- Frames include a ***header*** with synchronization, addressing and control (type and frame number) information;
- Frames also include a ***trailer*** with error control and synchronization information.

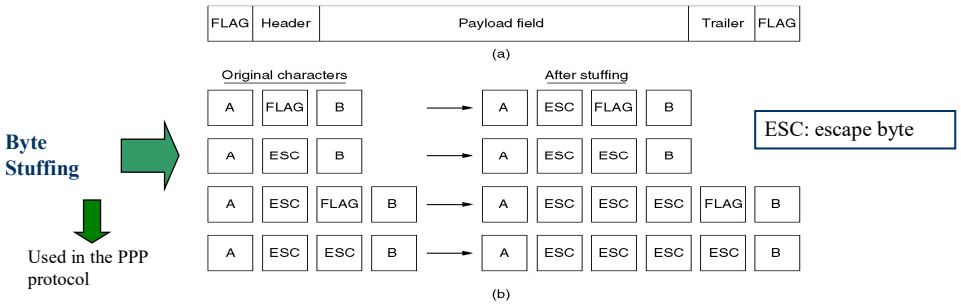




Framing: Byte "Stuffing"

Network layer data may contain binary patterns equal to those used for synchronization (*flags*) purposes in the data link layer:

- There is the need to avoid taking those IP data as a data link *flag*;
- To that purpose, if the *flag* pattern appears in the IP data it should be preceded, in the data link frame, by a ***stuffing*** pattern, to avoid any confusion.

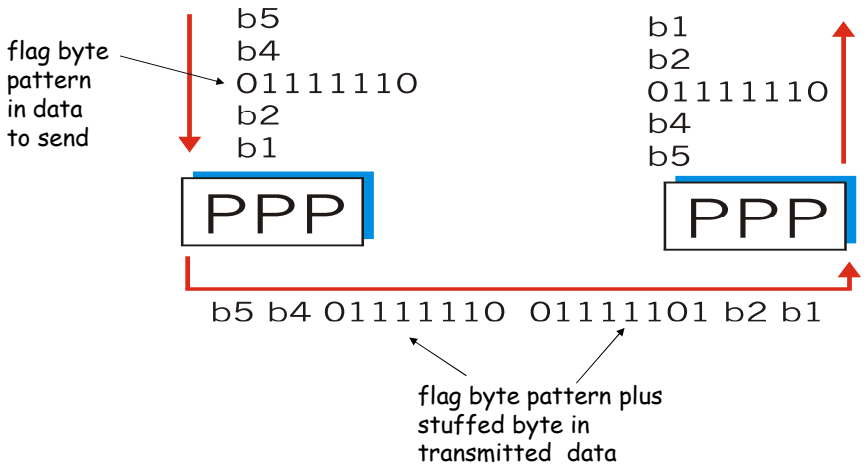


Byte Stuffing

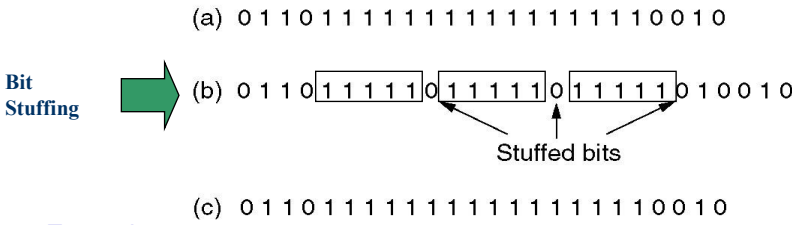
- "Data transparency" requirement: data field must be allowed to include flag pattern <01111110>.
  - Q: is received <01111110> data or flag?

- Sender: adds ("stuffs") extra < 01111110> byte after each <01111110> *data* byte.
- Receiver:
  - Two 01111110 bytes in a row: discard first byte, continue data reception;
  - Single 01111110: flag byte.

Byte Stuffing



Framing: Bit "Stuffing"



Example:

- Frame delimitation using the *flag:* ***0111 1110;***
- ***Bit stuffing*** consists in:
  - Add an extra *0* (b) after any sequence of five consecutive *1s* in the data field coming from the network layer (a), to avoid being taken for the *flag.*
- In the receptor the inverse operation takes place (c).

TÉCNICO LISBOA

Summary

- Principles behind data link layer services:
  - Link layer addressing;
  - Error detection, correction;
  - Sharing a broadcast channel: **multiple access**.
- Instantiation and implementation of various link layer technologies:
  - Ethernet;
  - IEEE 802.11 Wireless LANs;
  - Switched LANS;

TÉCNICO LISBOA

Summary

- Journey down protocol stack *complete* (except PHY);
- Solid understanding of networking principles and practice;
- ….. could stop here …. but *lots* of interesting topics!
  - Multimedia,
  - Security,
  - Network management, …