Mạng máy tính : Nguyên lý, Giao thức và luyện tập

Phần 5: Tầng liên kết dữ liệu và Mạng cục bộ

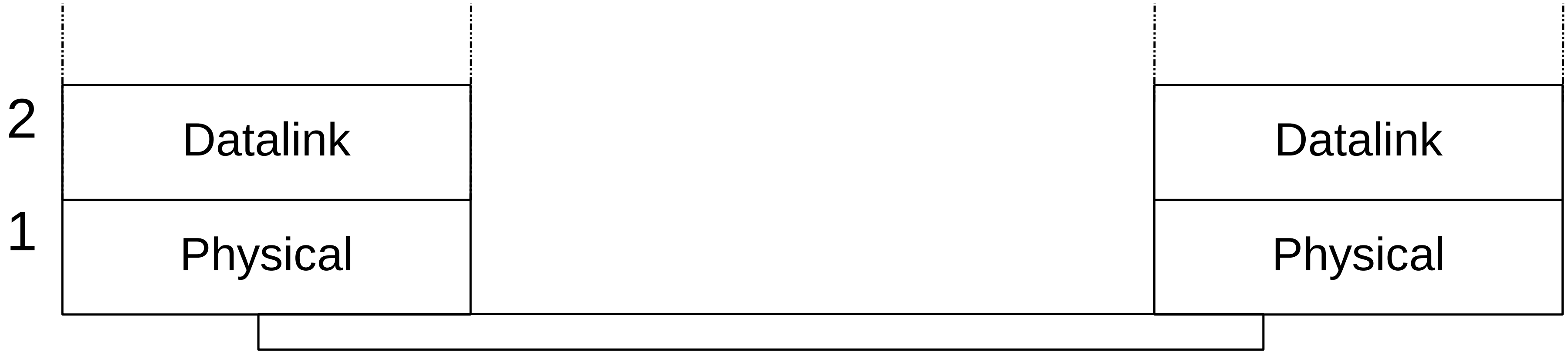
Datalink layer

- Point-to datalink layer
  - How to transmit and receive frames
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

Usage of the physical layer

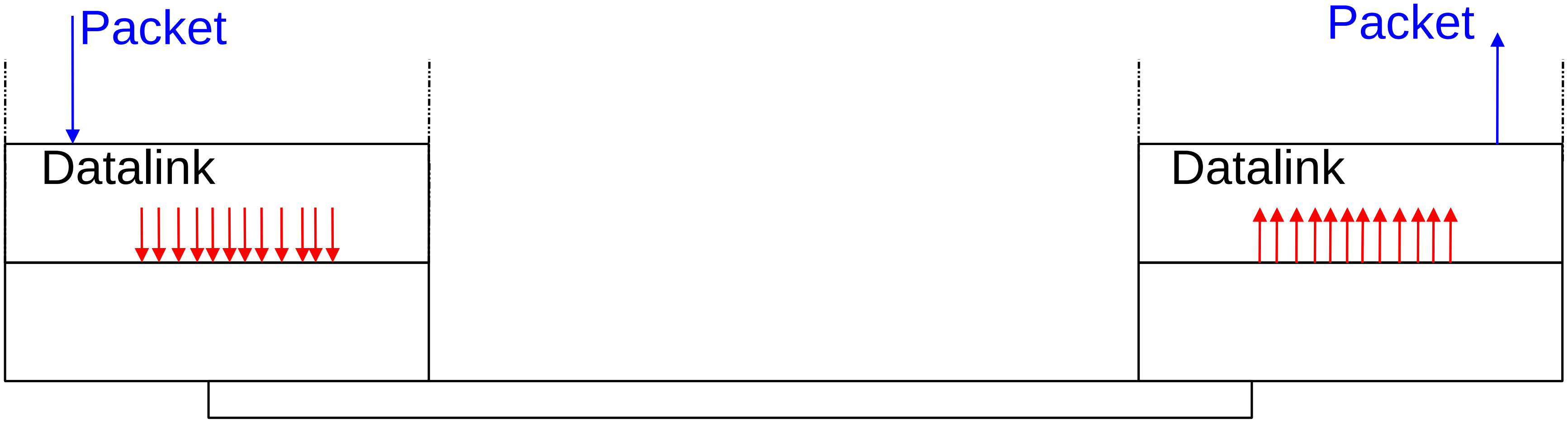
- Service provided by physical layer
  - Bit transmission between nodes attached to the same physical transmission channel
    - cable, radio, optical fiber, ...

- Better service for computers
  - Transmission/reception of short messages
  - Service provided by the datalink layer



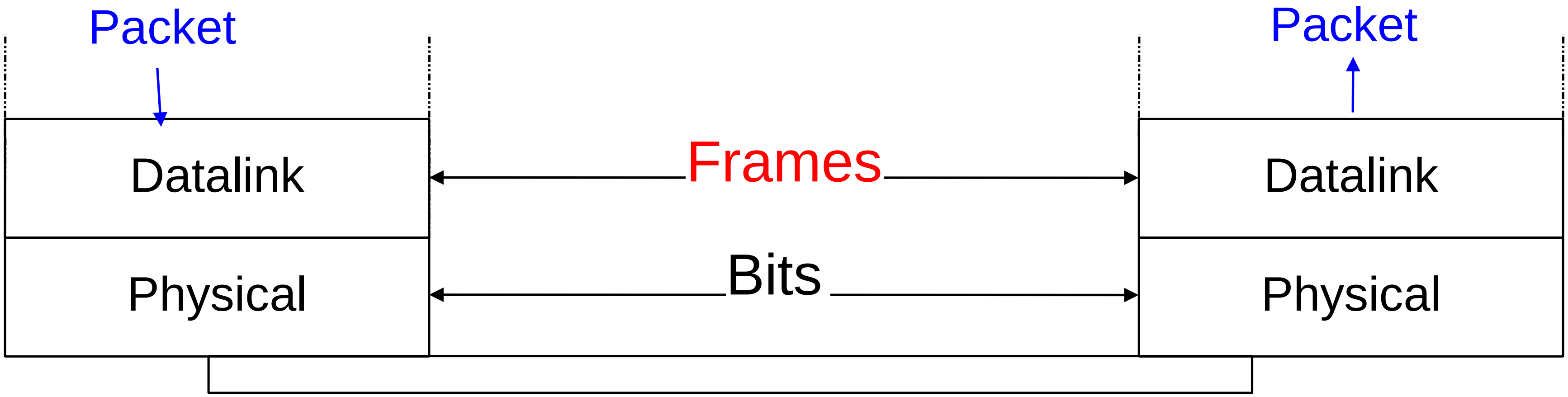
Datalink layer

- Goals
  - Provide a reliable transfert of packets although
    - Physical layer sends/receives bits and not packets
    - Physical layer service is imperfect
      - transmission errors
      - Losses of bits
      - Creation of bits



Frame delineation

- Frame
  - Unit of information transfer between two entities of the datalink layer
    - sequence of *N* bits
    - Datalink layer usually supports variable-length frames



- How can the receiver extract the frames from the received bit stream ?

Frame delineation

- Naïve solutions
  - Use frame size to delineate frames
    - Insert frame size in frame header
    - Issue
      - What happens when errors affect frame payload and frame header ?
  - Use special character/bitstring to mark beginning/end of frame
    - Example
      - all frames start with #
    - Issue
      - What happens when the special character/bitstring appears inside the frame payload ?

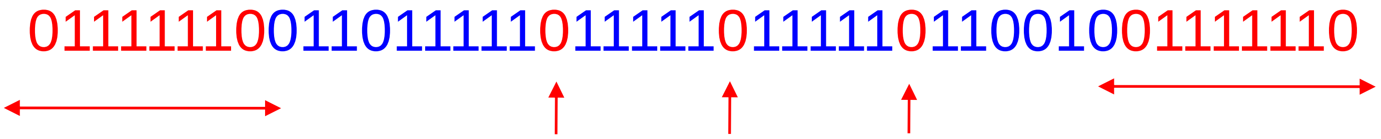
Character stuffing

- Character stuffing
  - Suitable for frames containing an integer number of bytes
  - 'DLE' 'STX' to indicate beginning of frame
  - 'DLE' 'ETX' to indicate end of frame
  - When transmitting frame, sender replaces 'DLE' by 'DLE' 'DLE' if 'DLE' appears inside the frame
  - Receiver removes 'DLE' if followed by 'DLE'

- Example
  - Packet : 1 2 3 'DLE' 4
  - Frame
    'DLE' 'STX' 1 2 3 'DLE' 'DLE' 4 'DLE' 'ETX'

Bit stuffing

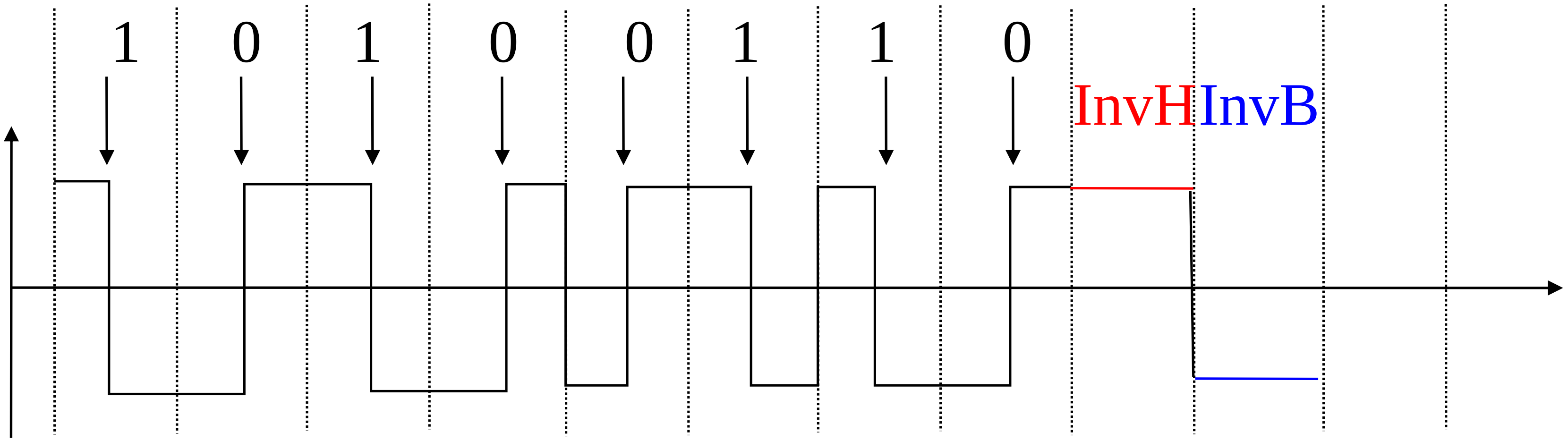
- Alternative to character stuffing
  - Suitable for frames composed of n bits
  - 01111110 used as marker at beginning and end of frame
  - Sender behaviour
    - If five bits set to ‘1’ must be sent, sender adds a bit set to ‘0’ immediately after the fifth bit set to ‘1’
  - Receiver behaviour
    - Counts the number of successive bits set to 1
      - 6 successive bits set to 1 followed by 0 : marker
      - 5 successive bits set to 1 followed by 0 : remove bit set to 0
    - Example
      - Packet : 0110111111111111111110010
      - Frame



01111110011011111011111011111011001001111110

Frame delineation

- Co-operation with physical layer
  - Some physical layers are able to transmit special physical codes that represent neither 0 nor 1
  - Example : Manchester coding



- invH (or N times invH) could be used to mark the beginning of a frame and invB (or N times invB) to mark the end of a frame

Frame delineation in practice

- Most datalink protocols use
  - Character stuffing or bit stuffing
    - Character stuffing is preferred by software implementations
  - A length field in the frame header
  - A checksum or CRC in the header or trailer to detect transmission errors

- A receiver frame is considered valid if
  - the correct delimiter appears at the beginning
  - the length is correct
  - the CRC/checksum is valid
  - the correct delimiter appears at the beginning

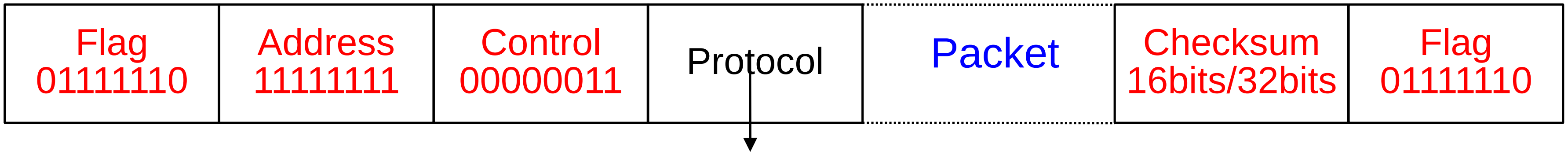
PPP : Point-to-Point Protocol

- Goal
  - Allow the transmission of network layer (IP but also other protocols) packets over serial lines
    - modems, leased lines, ISDN, ...
- Architecture
  - PPP is composed of three different protocols

  1. PPP
     - transmission of data frames (e.g. IP packets)
  2. LCP : Link Control Protocol
     - Negotiation of some options and authentication (username, password) and end of connection
  3. NCP : Network Control Protocol
     - Negotiation of options related to the network layer protocol used above PPP
       (ex: IP address, IP address of DNS resolver, ...)

PPP (2)

- PPP frame format



Flag	Address	Control	Protocol	Packet	Checksum	Flag
01111110	11111111	00000011			16bits/32bits	01111110

Identification of the network layer packet transported in the PPP frame

- Mechanisms used by PPP
  - character stuffing for asynchronous lines
  - bit stuffing for synchronous lines
  - CRC for error detection
    - 16 bits default but 32 bits CRC can be negotiated
  - No error correction by default
    - a reliable protocol can be negotiated
  - Data compression option
    - content of PPP frames can be compressed. To be negotiated at beginning of PPP connection

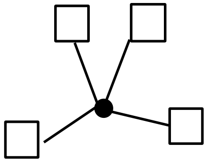
DataLink layer

- Point-to datalink layer
  - How to transmit and receive frames
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

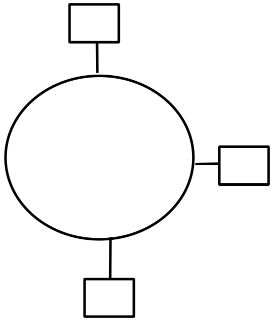
Local area networks

- How to efficiently connect N hosts together ?
  - Ideally we would like to have a single cable on each host while being able to reach all the others

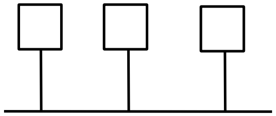
- Network topologies
  - Star-shaped network



  - Ring-shaped network

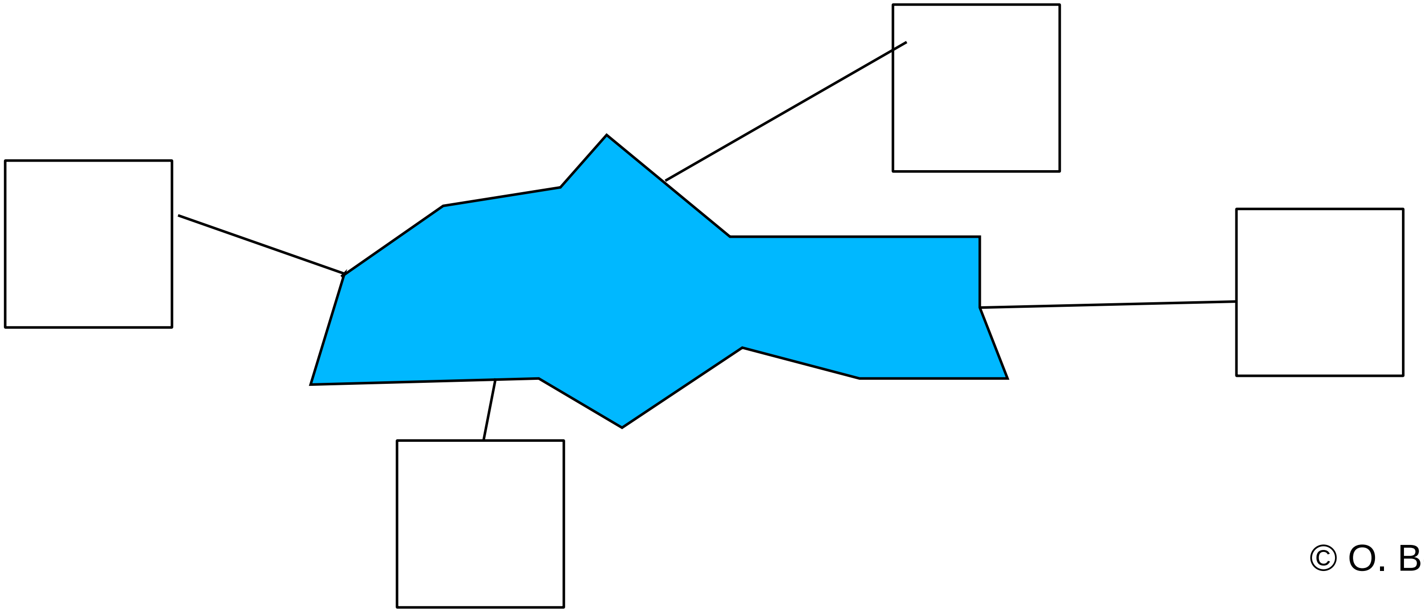


  - Bus-shaped network



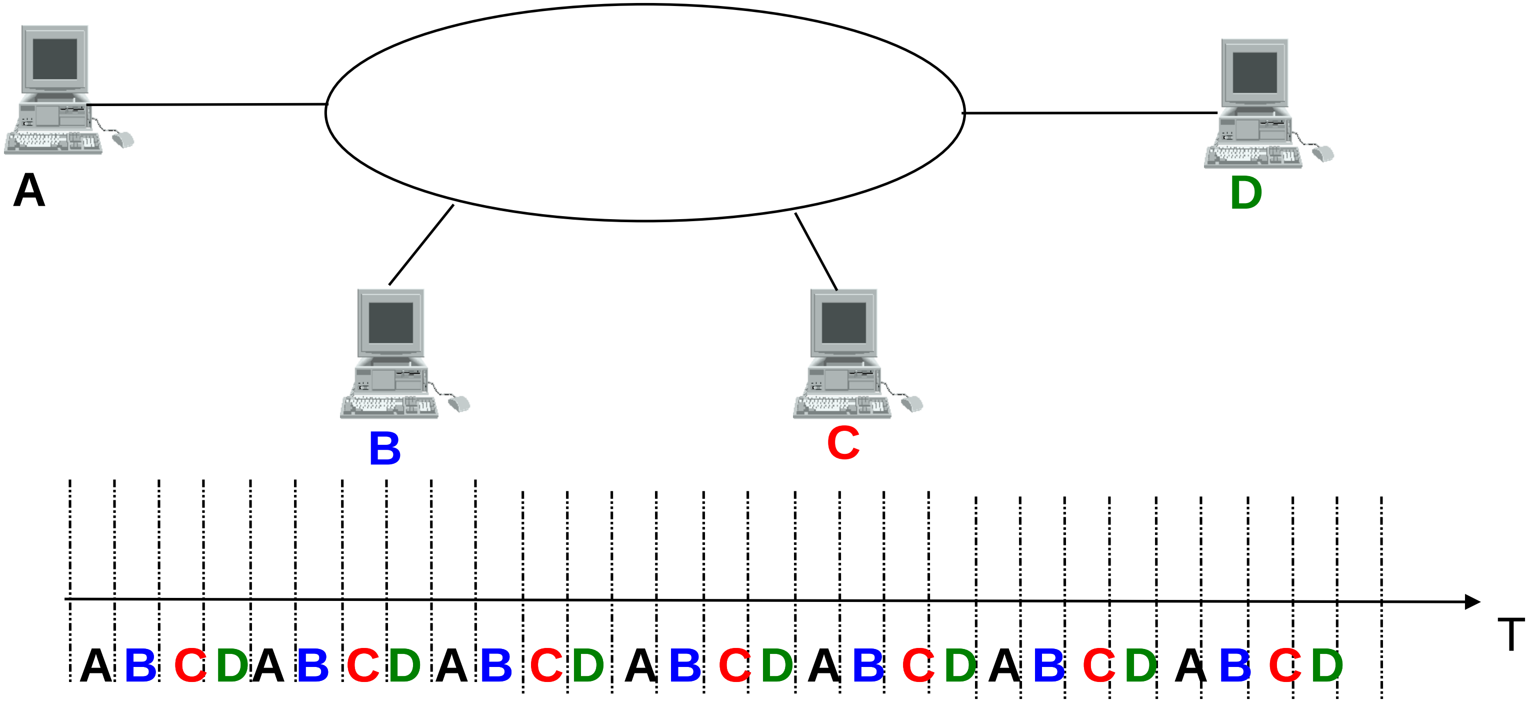
Local area networks

- Problems to be solved
  - How to identify the hosts attached to the LAN ?
  - The LAN is a shared resource
    - How to regulate access to this shared resource to provide :
      - fairness
        - All hosts should be able to use a fair fraction of the shared resource
      - performance
        - The shared resource should be used efficiently



Static allocation

- Time Division Multiplexing



- No suitable for a computer network
  - Leads to low link utilisation and high delays
  - Computers generate bursty trafic
- A more adaptive access control mechanism is required

Medium access control

- Hypotheses
  - N stations need to share the same transmission channel
    - A single transmission channel is available
  - Definition
    - Collision
      - If two stations transmit their frame at the same time, their electrical signal appears on the channel and causes a collision

- Options
  - Frame transmission
    - A station can transmit at any time
    - A station can only transmit at specific instants
  - Listening while transmitting
    - A station can listen while transmitting
    - A station cannot listen while transmitting

Medium access control

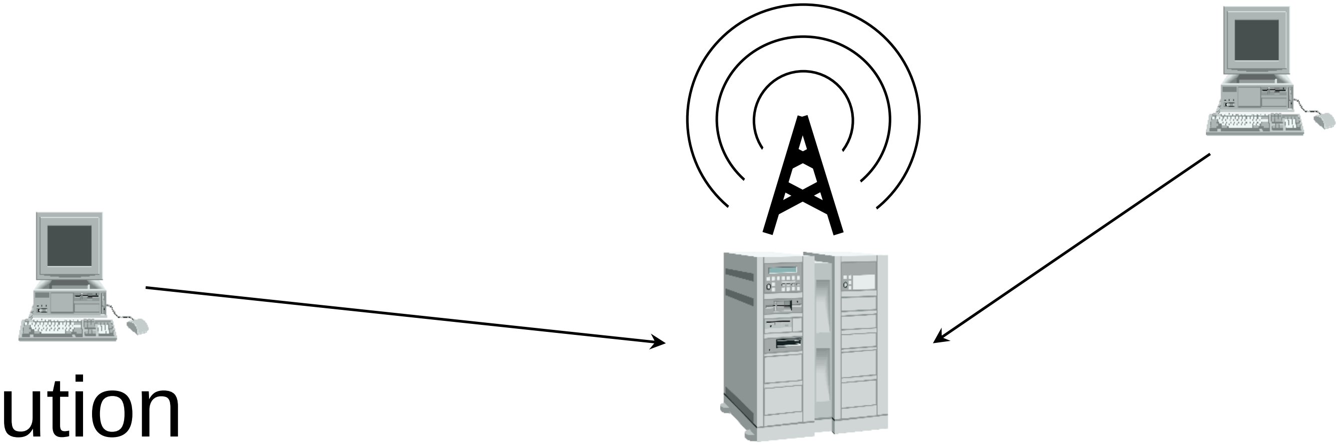
- How to regulate access to the shared medium ?
  - Statistical or optimistic solutions
    - hosts can transmit frames at almost any time
      - if the low is low, the frames will arrive correctly at destination
      - if the low is high, frames may collide
    - distributed algorithm allows to recover from the collisions
  - Deterministic or pessimistic solutions
    - Collisions are expensive and need to be avoided Distributed algorithm distributes authorisations to transmit to ensure that a single host is allowed to transmit at any time
      - avoids collisions when load is high, but may delay transmission when load is low

DataLink layer

- Point-to datalink layer
  - How to transmit and receive frames
- Local area networks
  - → Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

ALOHA

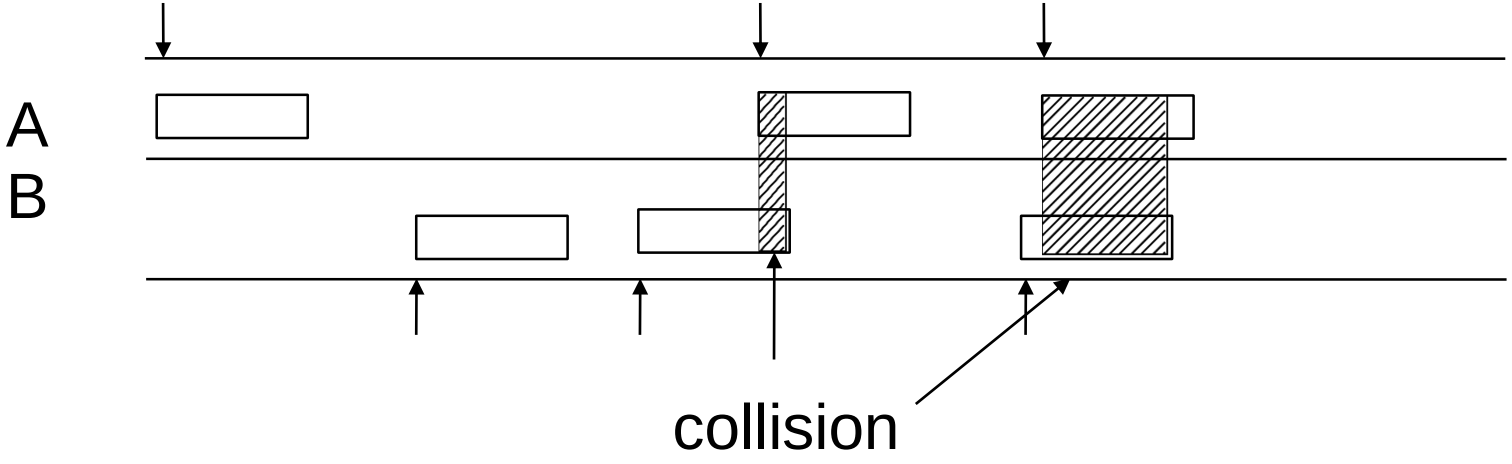
- Problem
  - terminals need to exchange data with computer but phone lines are costly



- Solution
- Wireless network
  - upstream frequency shared by all terminals
    - terminals do not hear each other
  - downlink frequency reserved for computer

ALOHA (3)

- How to organise frame transmission ?
  - If a host is alone, no problem
  - If two hosts transmit at the same time, a collision will occur and it will be impossible to decode their transmission



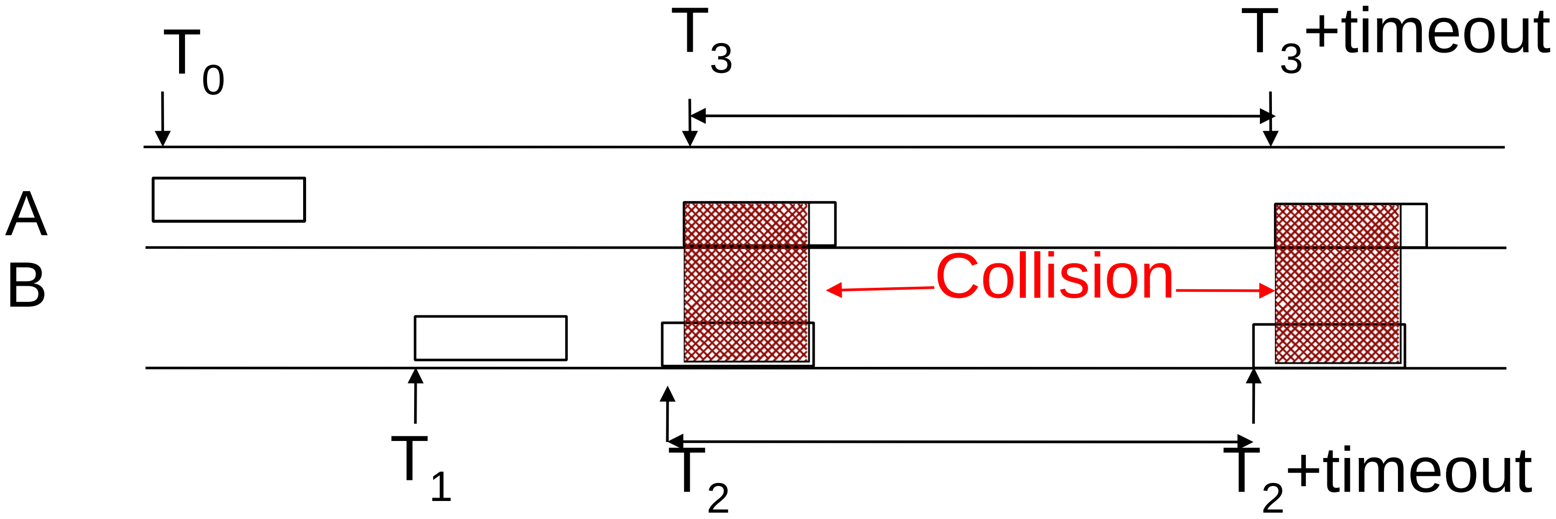
ALOHA (3)

- Medium access algorithm
  - First solution

```
N=1;
while ( N<= max) do
    send frame;
    wait for ack on return channel or timeout:
    if ack on return channel
        exit while;
    else
        /* timeout */
        /* retransmission is needed */
        N=N+1;
end do
/* too many attempts */
```

ALOHA (4)

- Drawback
  - When two stations enter in collision, they may continue to collide after



- How to avoid this synchronisation among stations ?

ALOHA (5)

- Improved algorithm

```
N=1;
while ( N<= max) do
    send frame;
    wait for ack on return channel or timeout:
    if ack on return channel
        exit while;
    else
        /* timeout */
        /* retransmission is needed */
        wait for random time;
        N=N+1;
end do
/* too many attempts */
```

Carrier Sense Multiple Access

- How to improve slotted Aloha ?

- Idea
  - Stations should be polite
    - Listens to the transmission channel before transmitting
    - Wait until the channel becomes free to transmit

  - Limitations
    - Politeness is only possible if all stations can listen to the transmission of all stations
      - true when all stations are attached to the same cable, but not in wireless networks

CSMA

- CSMA
  - Carrier Sense Multiple Access

```
N=1;
while ( N<= max) do
    wait until channel becomes free;
    send frame immediately;
    wait for ack or timeout:
    if ack received
        exit while;
    else
        /* timeout */
        /* retransmission is needed */
        N=N+1;
end do
/* too many attempts */
```

non-persistent CSMA

- Idea
  - Transmitting a frame immediately after the end of the previous one is a very aggressive behaviour
    - If the channel is free, transmit
    - Otherwise wait some random time before listening again

```
N=1;
while ( N<= max) do
    listen channel;
    if channel is empty
        send frame;
        wait for ack or timeout
        if ack received
            exit while;
        else /* retransmission is needed */
            N=N+1:
    else
        wait for random time;
end do
```

p-persistent CSMA

- Tradeoff between CSMA and non-persistent CSMA

```
N=1;
while ( N<= max) do
    listen channel;
    if channel is empty
        with probability p
            send frame;
            wait for ack or timeout
            if ack received
                exit while;
            else /* retransmission needed */
                N=N+1;
    else
        wait for random time;
end do
```

Improvements to CSMA

- Problems with CSMA
  - If one bit of a frame is affected by a collision, the entire frame is lost

- Solution
  - Stop the transmission of a frame as soon as a collision has been detected

- How to detect collisions ?
  - Station listens to channel while transmitting
    - If there is no collision, it will hear the signal it transmits
    - If there is a collision, is will hear an incorrect signal

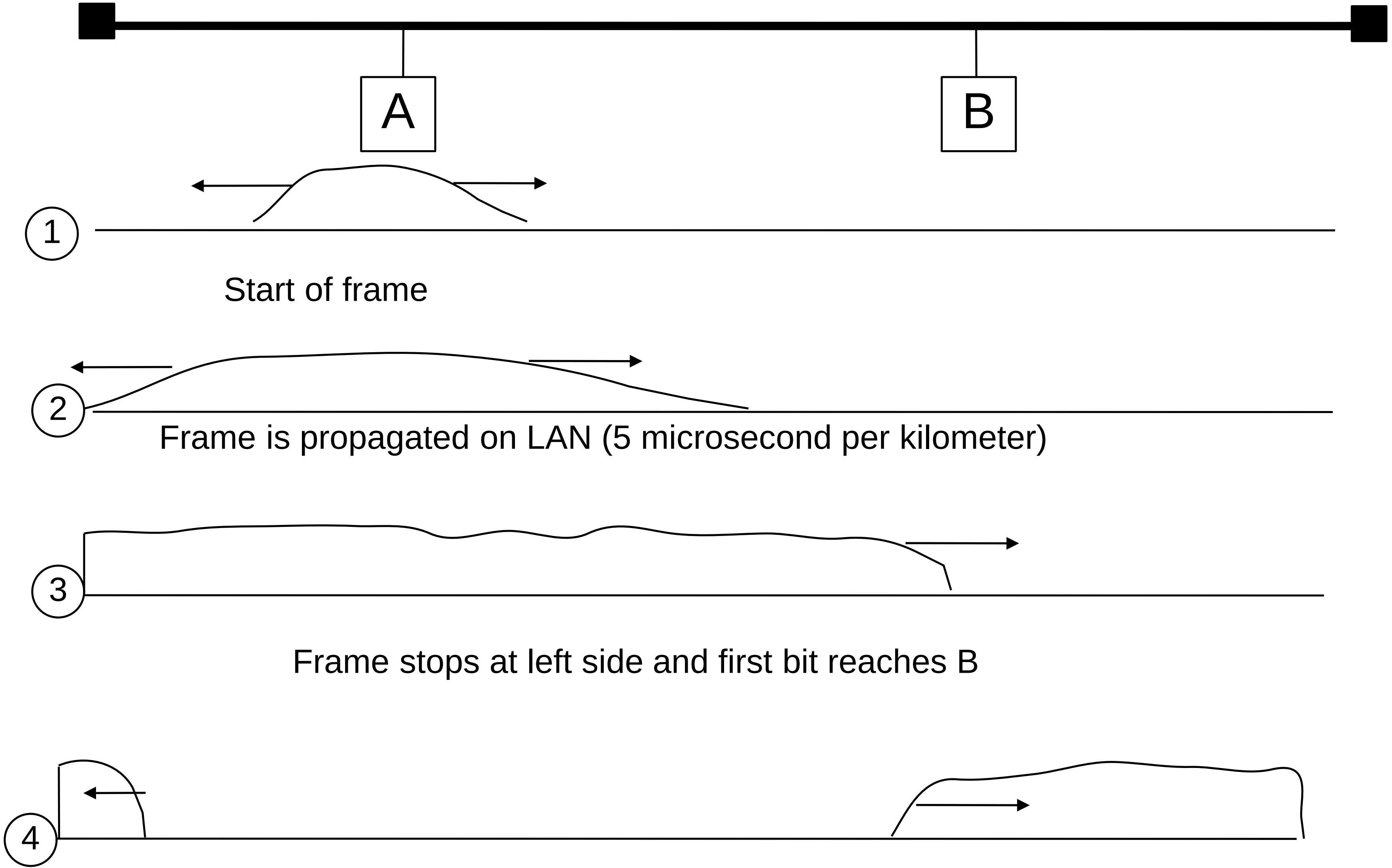
- CSMA/CD
  - Carrier Sense Multiple Access with Collision Detection
  - pas besoin d'ack puisque la station écoute le canal

CSMA/CD

- Medium access control

```
N=1;
while ( N<= max) do
     wait until channel becomes free;
     send frame and listen;
     wait until (end of frame) or (collision)
     if collision detected
          stop transmitting;
          /* after a special jam signal */
     else
          /* no collision detected */
          wait for interframe delay;
          exit while;
     N=N+1;
end do
/* too many attempts */
```

CSMA/CD : Example



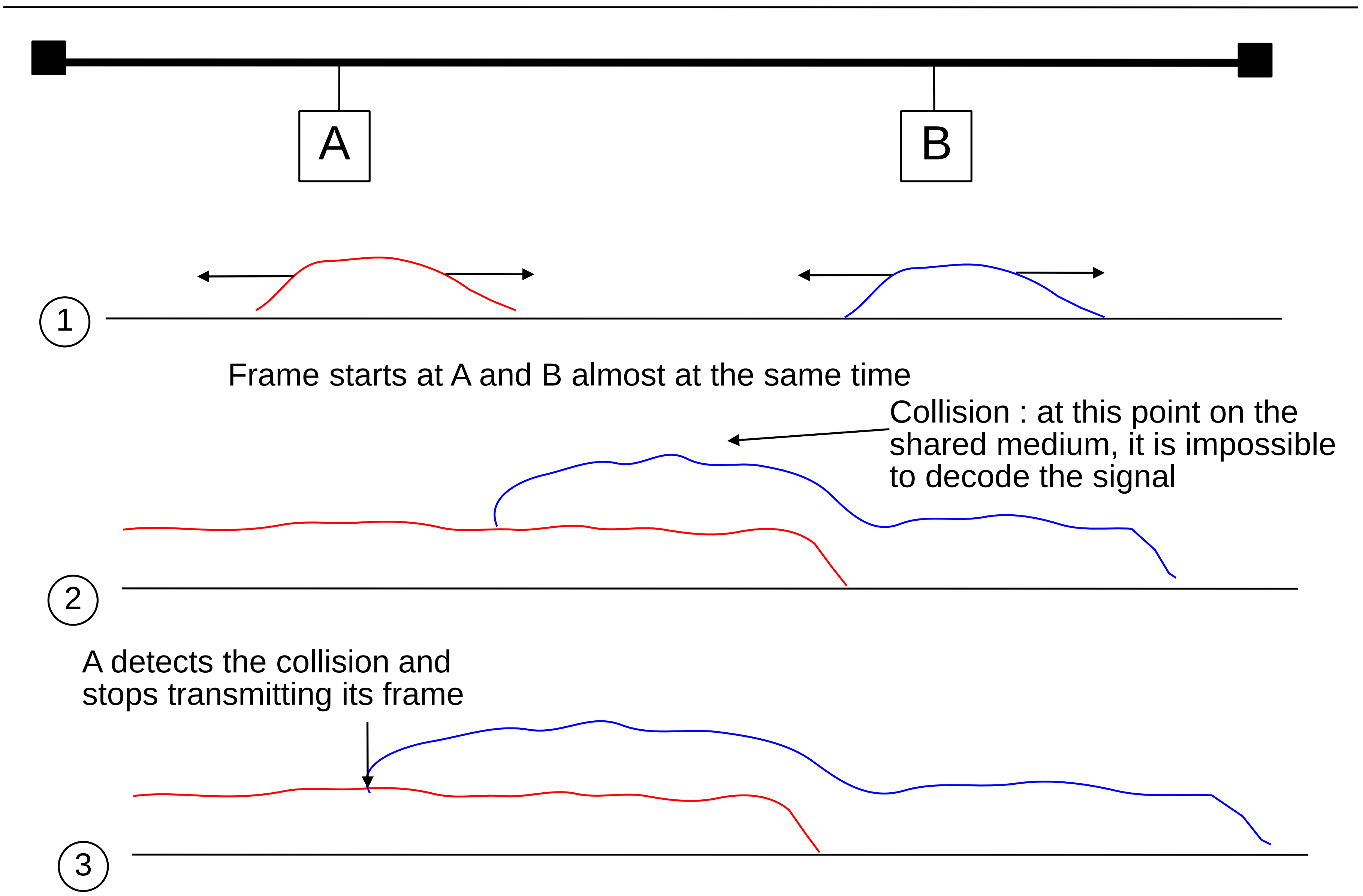
Start of frame

Frame is propagated on LAN (5 microsecond per kilometer)

Frame stops at left side and first bit reaches B

Frame leaves the LAN

CSMA/CD : Collisions



Frame starts at A and B almost at the same time

Collision : at this point on the shared medium, it is impossible to decode the signal

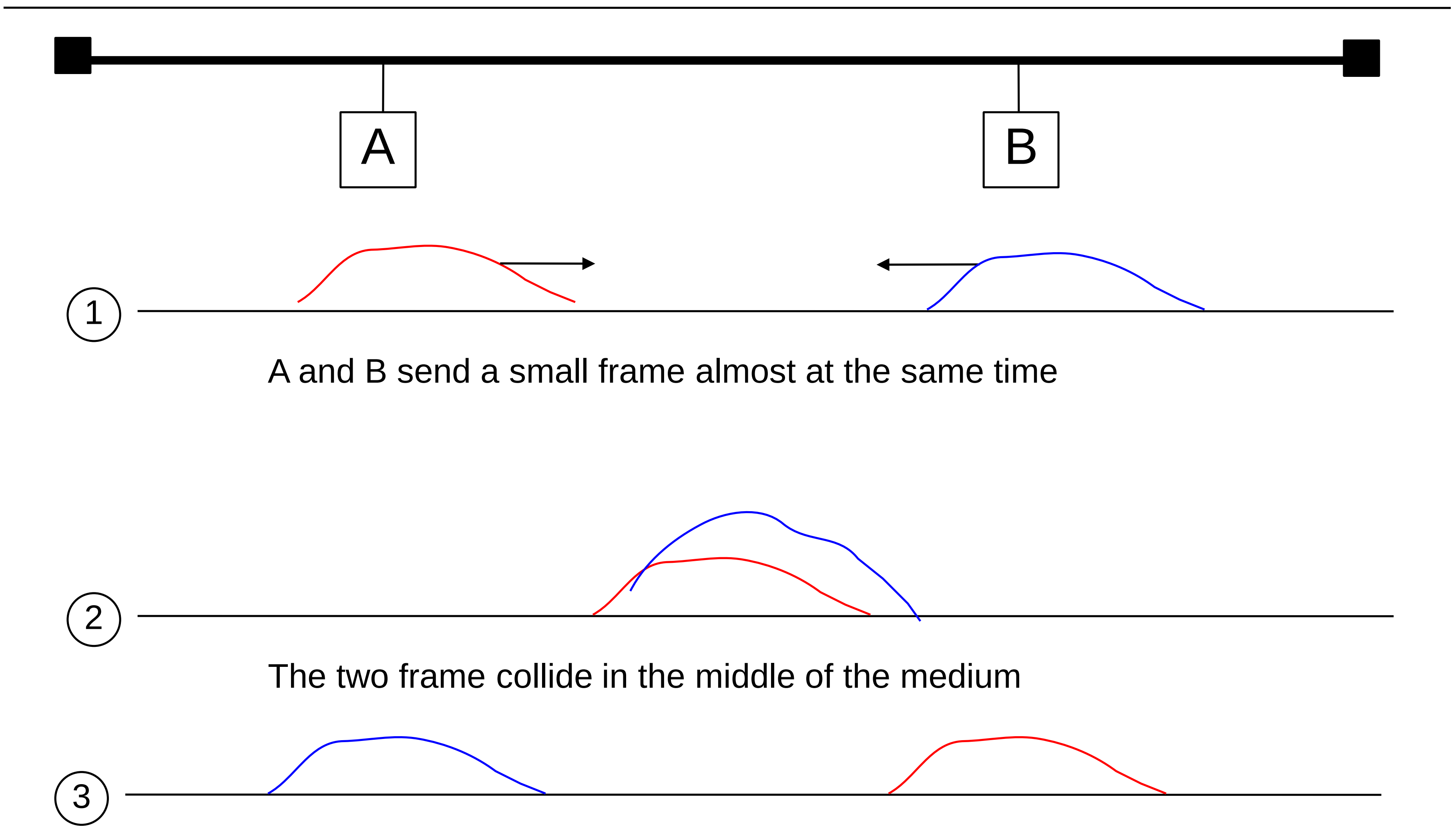
A detects the collision and stops transmitting its frame

CSMA/CD : Collisions (2)

- Advantages
  - Improves channel utilisation as stations do not transmit corrupted frames
  - a station can detect whether its frame was sent without collision
    - implicit acknowledgement if destination is up
    - when a collision is detected, automatic retransmission

- Is it possible for a station to detect all collisions on all its frames ?

CSMA/CD : Collisions (3)



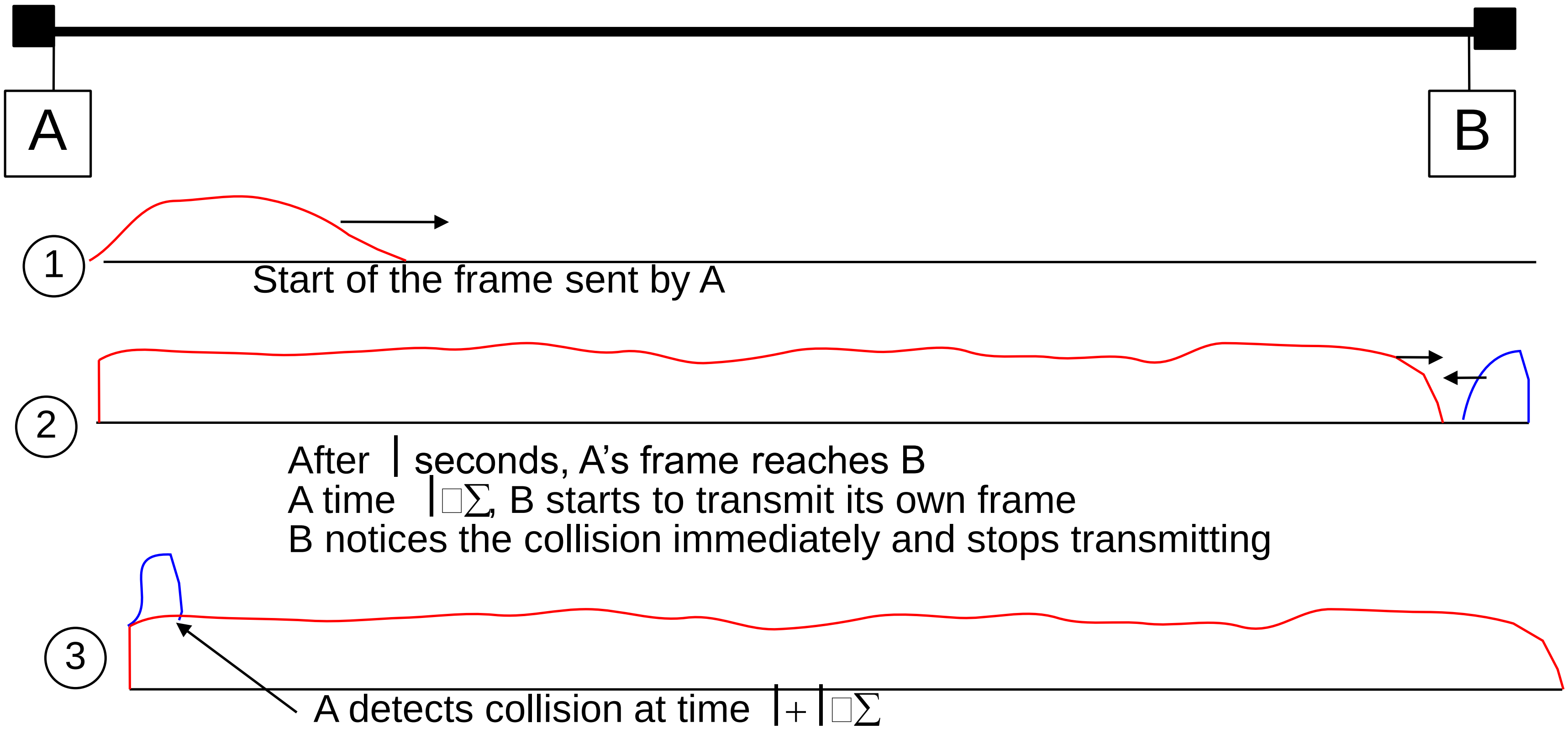
A and B send a small frame almost at the same time

The two frame collide in the middle of the medium

A did not notice any collision
B did not notice any collision
They both consider that their frames were received correctly

CSMA/CD : Collisions (4)

- How to ensure that all collisions are detected ?
  - Worst case scenario



Start of the frame sent by A

After | seconds, A’s frame reaches B
A time |□∑, B starts to transmit its own frame
B notices the collision immediately and stops transmitting

A detects collision at time |+|□∑

CSMA/CD : Collisions (5)

- How can a station ensure that it will be able to detect all the collisions affecting its frames ?
  - Each frame must be transmitted for at least a duration equal to the two way delay (2□|)
    - As the throughput on a bus is fixed, if the two way delay is fixed, then all frames must be larger than a minimum frame size
    - Improvement
      - To ensure that all stations detect collisions, a station that notices a collision should send a jamming signal

Exponential backoff

- How to deal with collisions ?
  - If the stations that collide retransmit together, a new collision will happen

- Solution
  - Wait some random time after the collision
  - After collision, time is divided in slots
    - a slot = time required to send a minimum sized frame
      - After first collision, wait 0 or 1 slot before retransmitting
      - After first collision, wait 0, 1,2 or 3 slots before retransmitting
      - After first collision, wait 0..2^i-1 slots before retransmitting

CSMA/CD with exponential backoff

- Medium access control

```
N=1;
while ( N<= max) do
     wait until channel becomes free;
     send frame and listen;
     wait until (end of frame) or (collision)
     if collision detected
          stop transmitting;
          /* after a special jam signal */
          k = min (10, N);
          r = random(0, $2^k$ - 1) * slotTime;
          wait for r time slots;
     else
          /* no collision detected */
          wait for interframe delay;
          exit while;
     N=N+1;
end do
/* too many attempts */
```

CSMA with Collision Avoidance

- Goal
  - Design a medium access control method suitable for wireless networks
    - on a wireless network, a sender cannot usually listen to its transmission (and thus CSMA/CD cannot be used)
- Improvements to CSMA
  - Initial delay before transmitting if channel is empty
    - Extended Inter Frame Space (EIFS)
  - Minimum delay between two successive frames
    - Distributed Coordination Function Inter Frame Space (DIFS)
  - Delay between frame reception and ack transmission
    - Short Inter Frame Spacing (SIFS, SIFS< DIFS < EIFS)

CSMA/CA (1)

- Sender

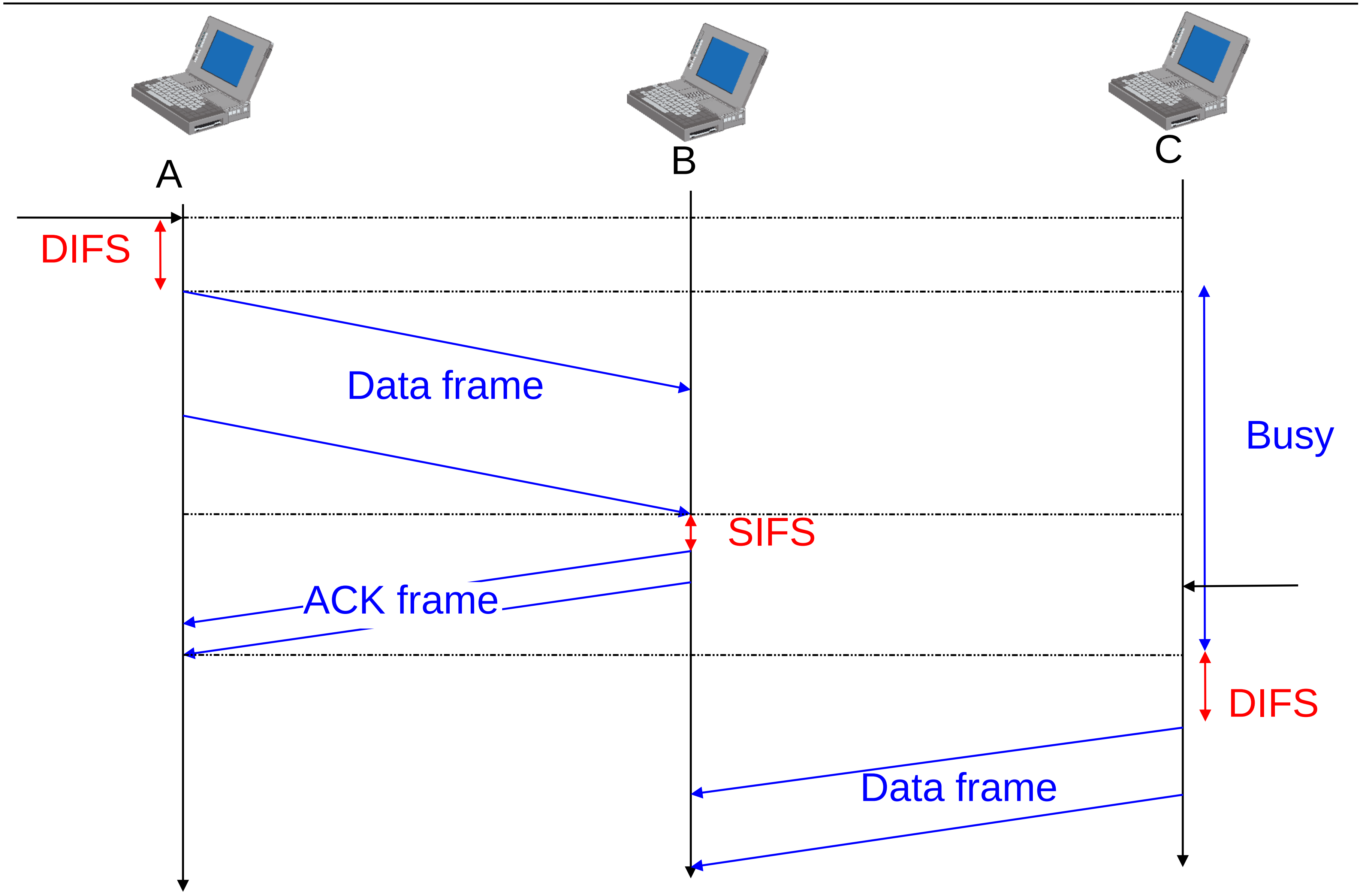
```
N=1;
while ( N<= max) do
     if (previous frame corrupted)
     { wait until channel free during t>=EIFS; }
     else
     { wait until endofframe;
       wait until channel free during t>=DIFS; }
     send data  frame ;
     wait for ack or timeout:
     if ack received
          exit while;
     else
          /* timeout retransmission is needed */
          N=N+1;
end do
/* too many attempts */
```

CSMA/CA (2)

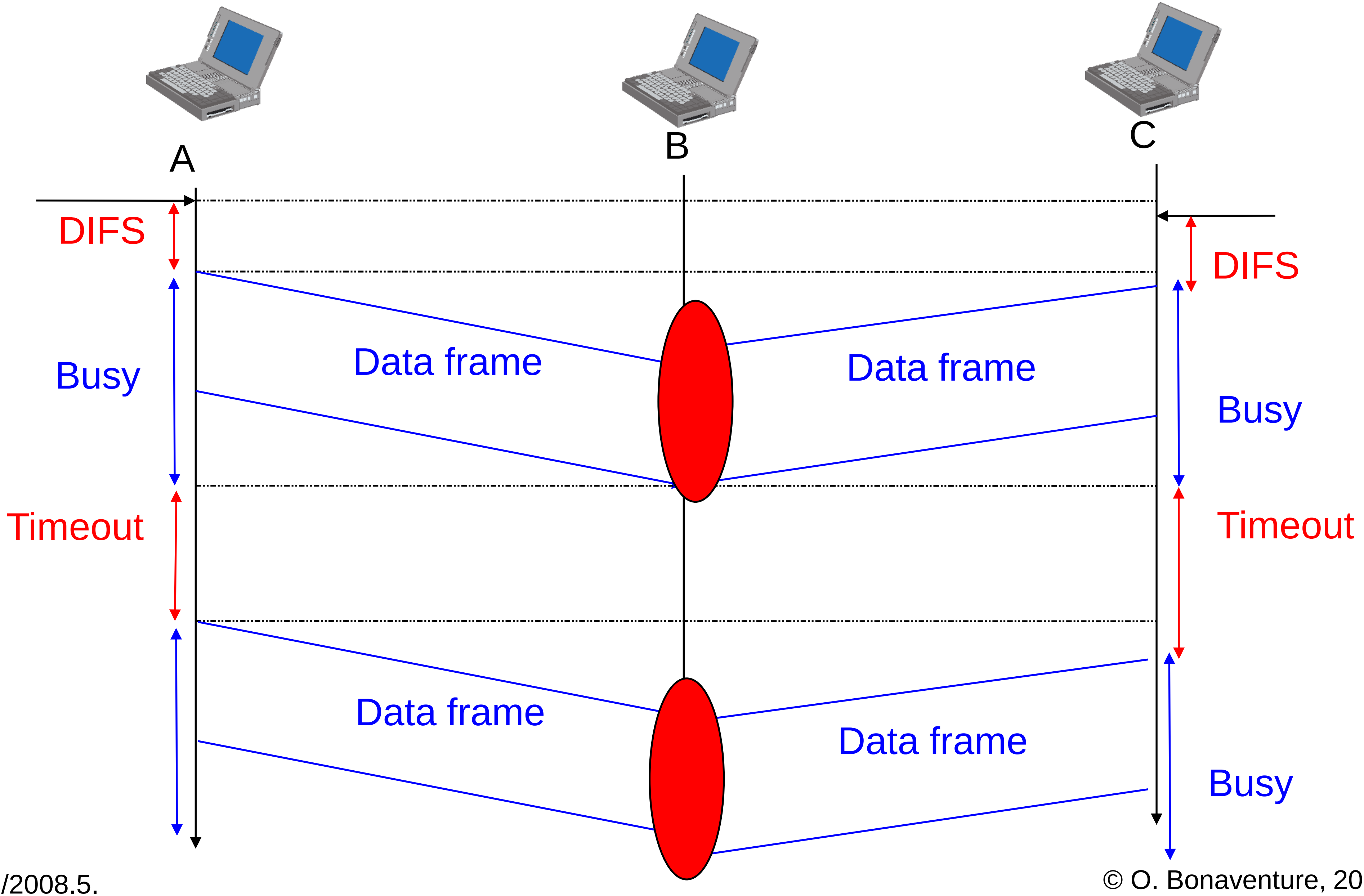
- Receiver

```
While (true)
{
 Wait for data frame;
    if not(duplicate)
        { deliver (frame) }
 wait during SIFS;
 send ack (frame) ;
}
```

CSMA/CA : Example



CSMA/CA : Problem

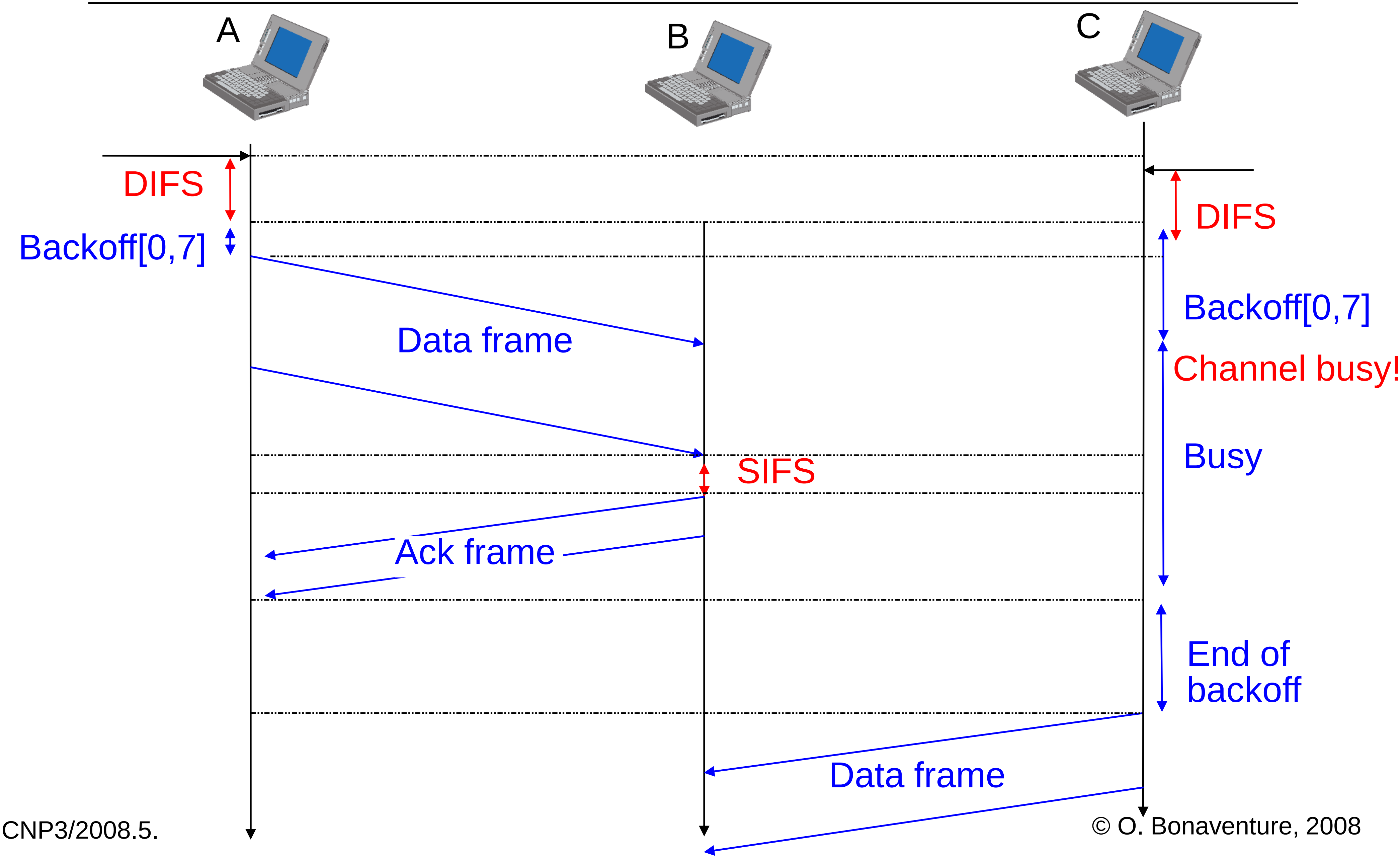


CSMA/CA
First improvement (2)

- Sender

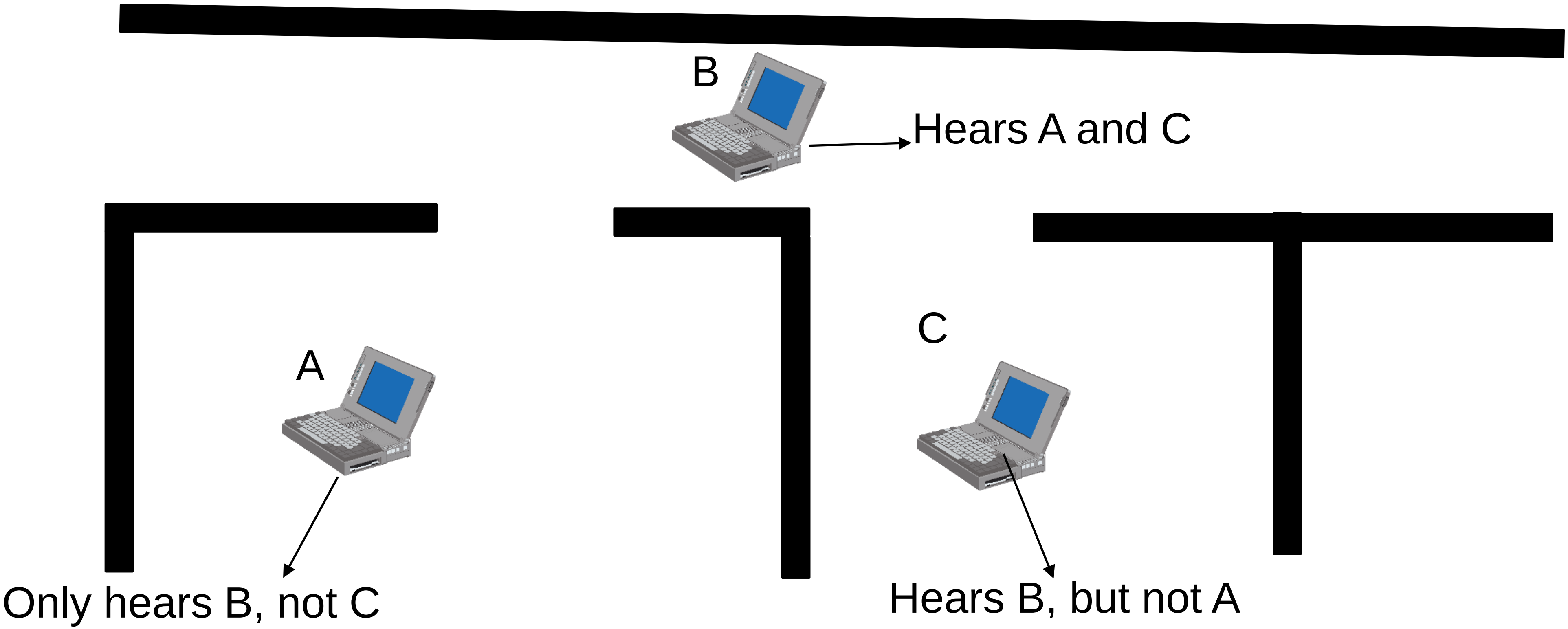
```
N=1;
while ( N<= max) do
      if (previous frame corruped)
      { wait until channel free during t>=EIFS; }
      else
      { wait until endofframe;
        wait until channel free during t>=DIFS; }
      backoff_time = int(random[0,min(255,7*$2^{N-1}$)])*T
      wait(backoff_time)
      if (channel still free)
      { send data  frame ;
        wait for ack or timeout:
       if ack received
           exit while;
       else /* timeout retransmission is needed */
          N=N+1; }
end do
```

CSMA/CA : Example 2



CSMA/CA
Hidden station problem

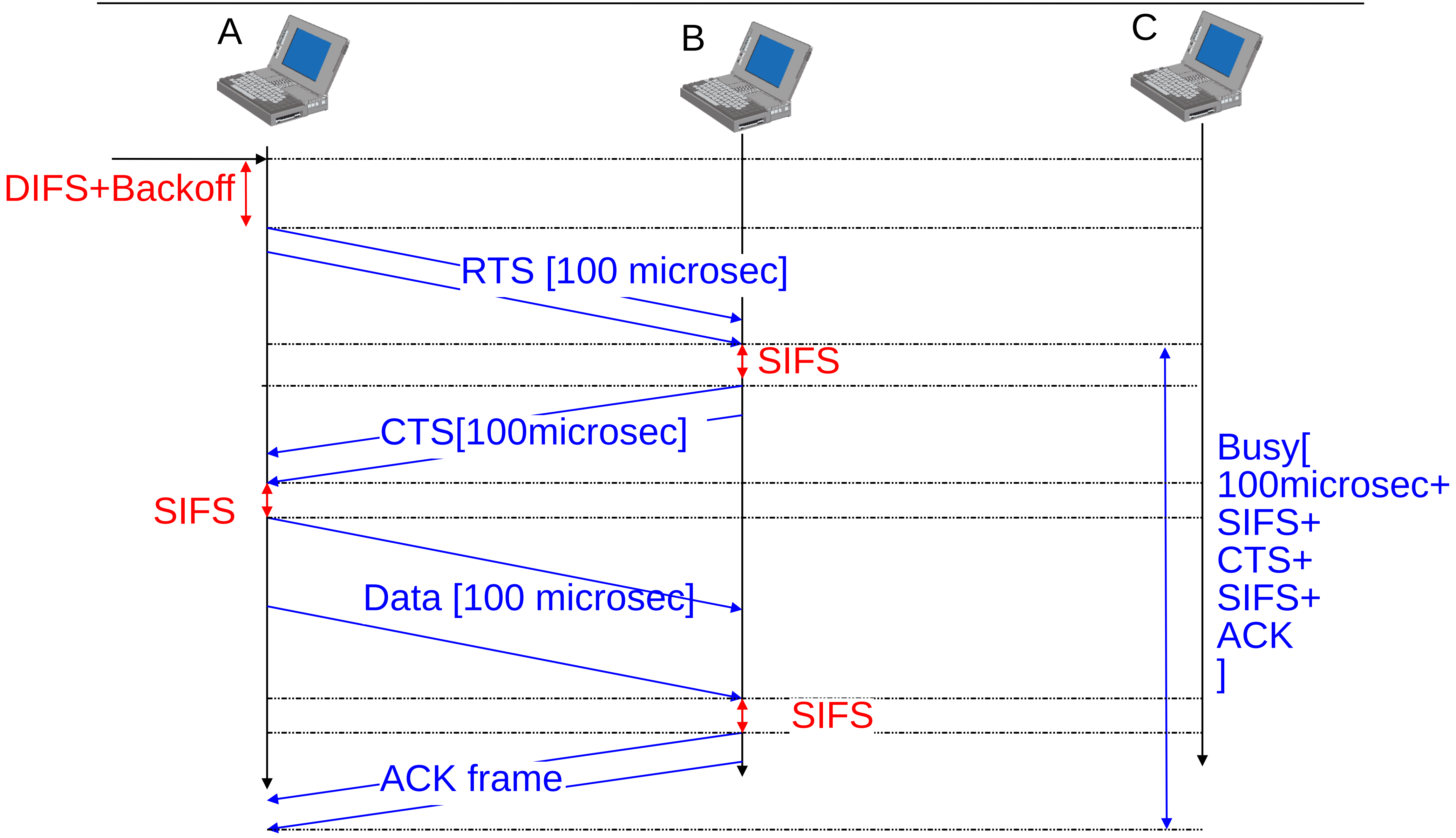
- Often occurs in wireless networks



CSMA/CA
Second improvement

- Principle
  - Allow the sender to “reserve” some air time
    - Special (short) RTS frame indicates duration
      - Using a short RTS frame reduces the risk of collisions while transmitting this frame
  - Allow the receiver to confirm the reservation
    - Special (short) CTS frame indicates reservation
      - Using a short CTS frame reduces the risk of collisions while transmitting this frame
  - The stations that could collide with the transmission will hear at least CTS
  - Frame contains an indication of transmission time

CSMA/CA : Example 3



Datalink layer

- Point-to datalink layer
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
    - Basics of Ethernet
    - IP over Ethernet
    - Interconnection of Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

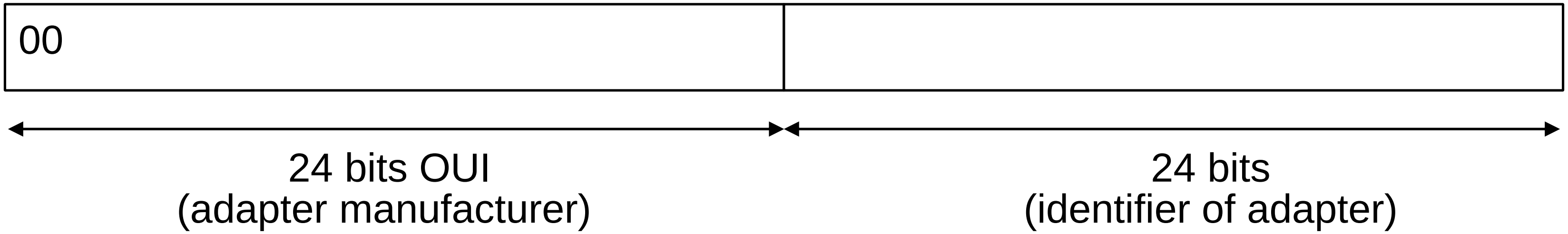
Ethernet/802.3

- Most widely used LAN
  - First developed by Digital, Intel and Xerox
  - Standardised by IEEE and ISO

- Medium Access Control
  - CSMA/CD with exponential backoff
  - Characteristics
    - Bandwidth: 10 Mbps
    - Two ways delay
      - 51.2 microsec on Ethernet/802.3
        - => minimum frame size : 512 bits
    - Cabling
      - 10Base5 : (thick) coaxial cable maximum 500 m,100 stations
      - 10Base2 : (thin) coaxial 200 m maximum and 30 stations

Ethernet Addresses (2)

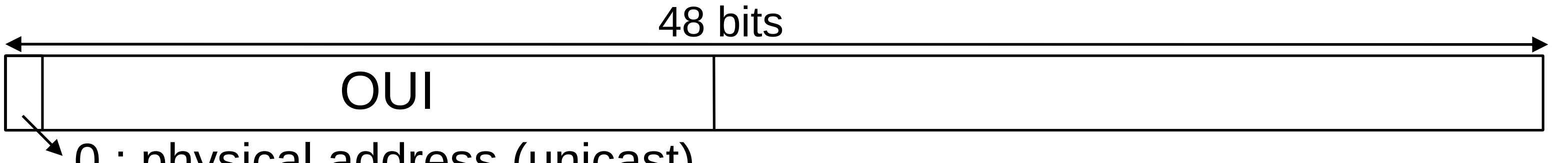
- Each Ethernet adapter has a unique Ethernet address
  - ensures that two hosts on the same LAN will not use the same Ethernet address
- Ethernet Addressing format
  - 48 bits addresses
    - Source Address



    - Destination address
      - If high order bit is 0, host unicast address
      - If high order bit is 1, host multicast address
        - broadcast address = 111111..111

LAN-level multicast

- Principle
  - Two types of destination Ethernet addresses
    - Physical addresses
      - identifies one Ethernet adapter
    - Logical addresses
      - identifies a logical group of Ethernet destinations

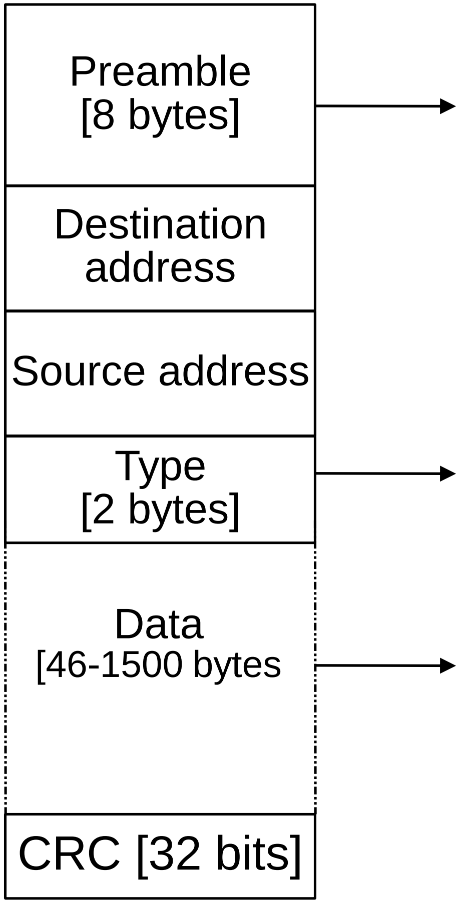


0 : physical address (unicast)
1 : logical address (multicast)

- Transmission of multicast frame
  - sender transmits frame with multicast destination addr.
- Reception of multicast frames
  - Ethernet adapters can be configured to capture frames whose destination address is
    - Their unicast address
    - One of a set of multicast addresses

Ethernet Frames

- DIX Format
  - proposed by Digital, Intel and Xerox



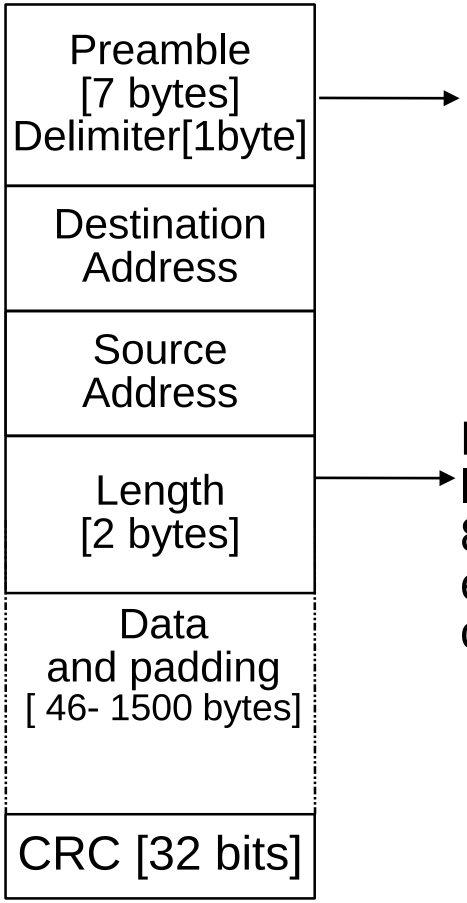
Used to mark the beginning of the frame
Allows the receiver to synchronise its
clock to the sender's clock

Indication of the type of packet contained
inside the frame

Upper layer protocol must ensure that
the payload of the Ethernet frame is
at least 46 bytes and at most 1500 bytes

802.3 Frames

- Ethernet 802.3
  - standardised by IEEE



Used to mark the beginning of the frame
Allows the receiver to synchronise its clock to the sender’s clock

Provides the real length of the network layer packet placed inside the payload. 802.3 adds padding (bytes set to 0) to ensure that the data field of the frame contains at least 46 bytes

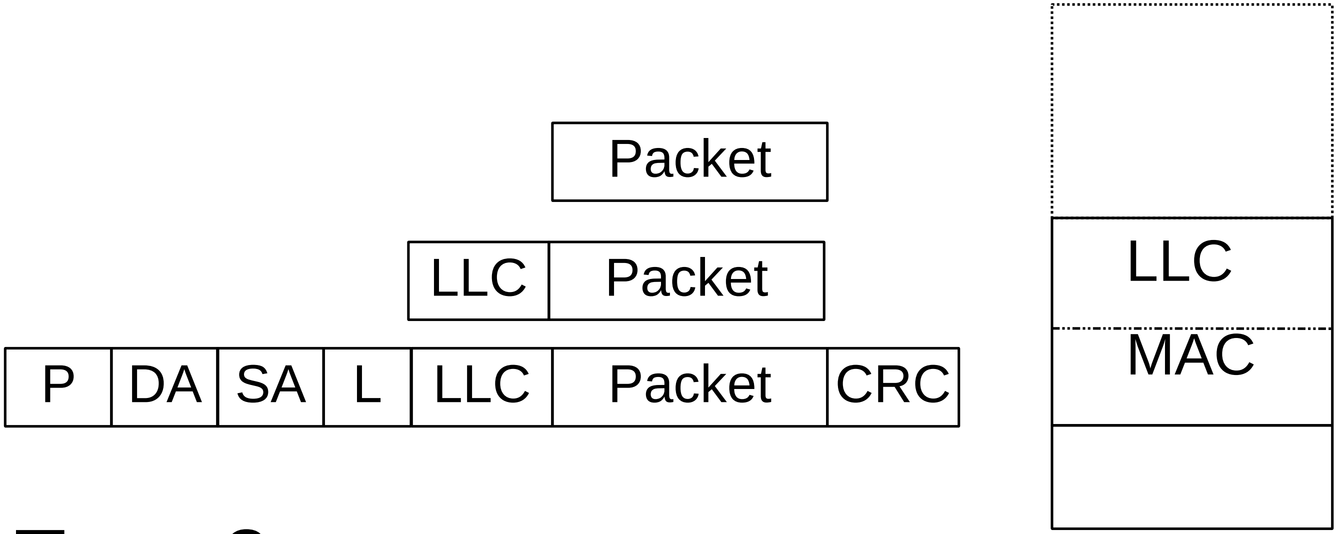
Ethernet and 802.3 : details

- How can the receiver identify the type of network protocol packet inside the frame ?
  - Ethernet : thanks to Type field
  - 802.3 : no Type field !

- IEEE standard
  - Divide datalink layer in two sublayers
    - Medium Access Control (MAC)
      - lower sublayer responsible for the frame transmission and medium access control (CSMA/CD)
      - interacts with but does not depend from the physical layer
      - example : 802.3
    - Logical Link Control (LLC)
      - higher sublayer responsible for the exchange of frames with the higher layers
      - interacts with the higher layer
      - does not depend from the MAC layer
      - several variants of LLC exist

802.2 : LLC

- LLC Type 1
  - Unreliable connectionless service
  - Addition to 802.3
    - New LLC header allows to identify upper layer protocol
      - similar to Type field of Ethernet DIX



- LLC Type 2
  - Reliable transmission with acknowledgements
    - An example of a protocol developed by a standardisation body but used by nobody...

Ethernet Service

- An Ethernet network provides a connectionless unreliable service
- Transmission modes
  - unicast
  - multicast
  - broadcast

- Even if in theory the Ethernet service is unreliable, a good Ethernet network should
  - deliver frames to their destination with a very hig probability of delivery
  - not reorder the transmitted frames
    - reordering is obviously impossible on a bus

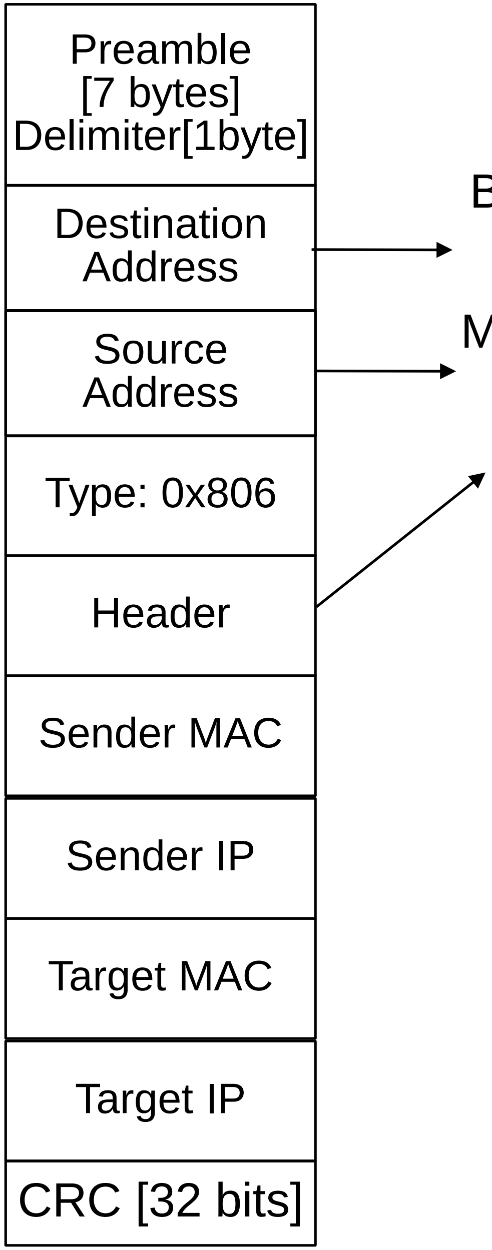
Datalink layer

- Point-to datalink layer
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
    - Basics of Ethernet
    - IP over Ethernet
    - Interconnection of Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

IP on LANs

- Problems to be solved
  - How to encapsulate IP packets in frames ?
  - How to find the LAN address of the IP destination ?
- LAN efficiently supports broadcast/multicast transmission
  - When a host needs to find the LAN address of another IP host, it broadcasts a request
    - The owner of the destination IP address will reply and provided its LAN address
- LAN doesn’t efficiently support broadcast/multicast
    - Maintain a server storing *IP address:MAC address pairs*
    - Each host knows server’s MAC address and registers its address pair
    - Each host sends request to server to map IP addresses

ARP frame format



Preamble [7 bytes] Delimiter[1byte]

Destination Address

Source Address

Type: 0x806

Header

Sender MAC

Sender IP

Target MAC

Target IP

CRC [32 bits]

Broadcast : 111...111

MAC address of the sender

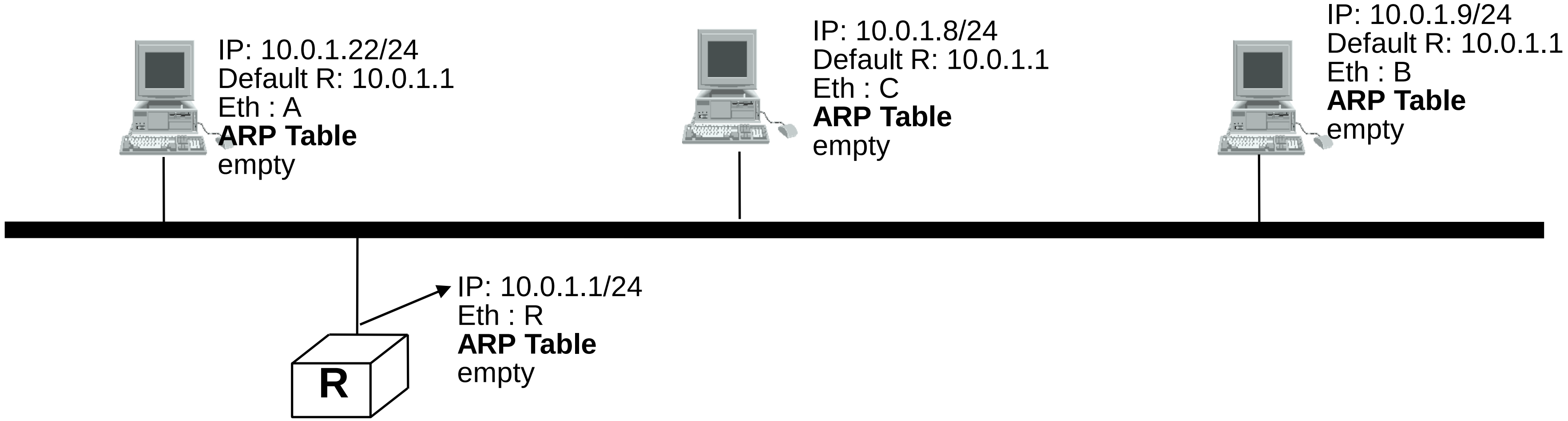
Common header for all ARP frames
- Hardware type Ethernet is 1
- Protocol type , IP is 0x0800.- Hardware length : length of MAC address - Protocol length : length of network layer address - Operation : 1 for request, 2 for reply, 3 for RARP request, and 4 for RARP reply.

Optimisations

- When should a host send ARP requests ?
  - Before sending each IP packet ?
    - No, each host/router maintains an ARP table that contains the mapping between IP addresses and Ethernet addresses. An ARP request is only sent when the ARP table is empty

- How to deal with hosts that change their addresses ?
  - Expiration timer is associated to each entry in the ARP table
    - Line of ARP table is removed upon timer expiration.
    - Some implementations send an ARP request to revalidate it before removing the line
    - Some implementations remember when ARP lines have been used to avoid removing an important entry

IP over Ethernet : Example



IP: 10.0.1.22/24
Default R: 10.0.1.1
Eth : A
ARP Table
empty

IP: 10.0.1.8/24
Default R: 10.0.1.1
Eth : C
ARP Table
empty

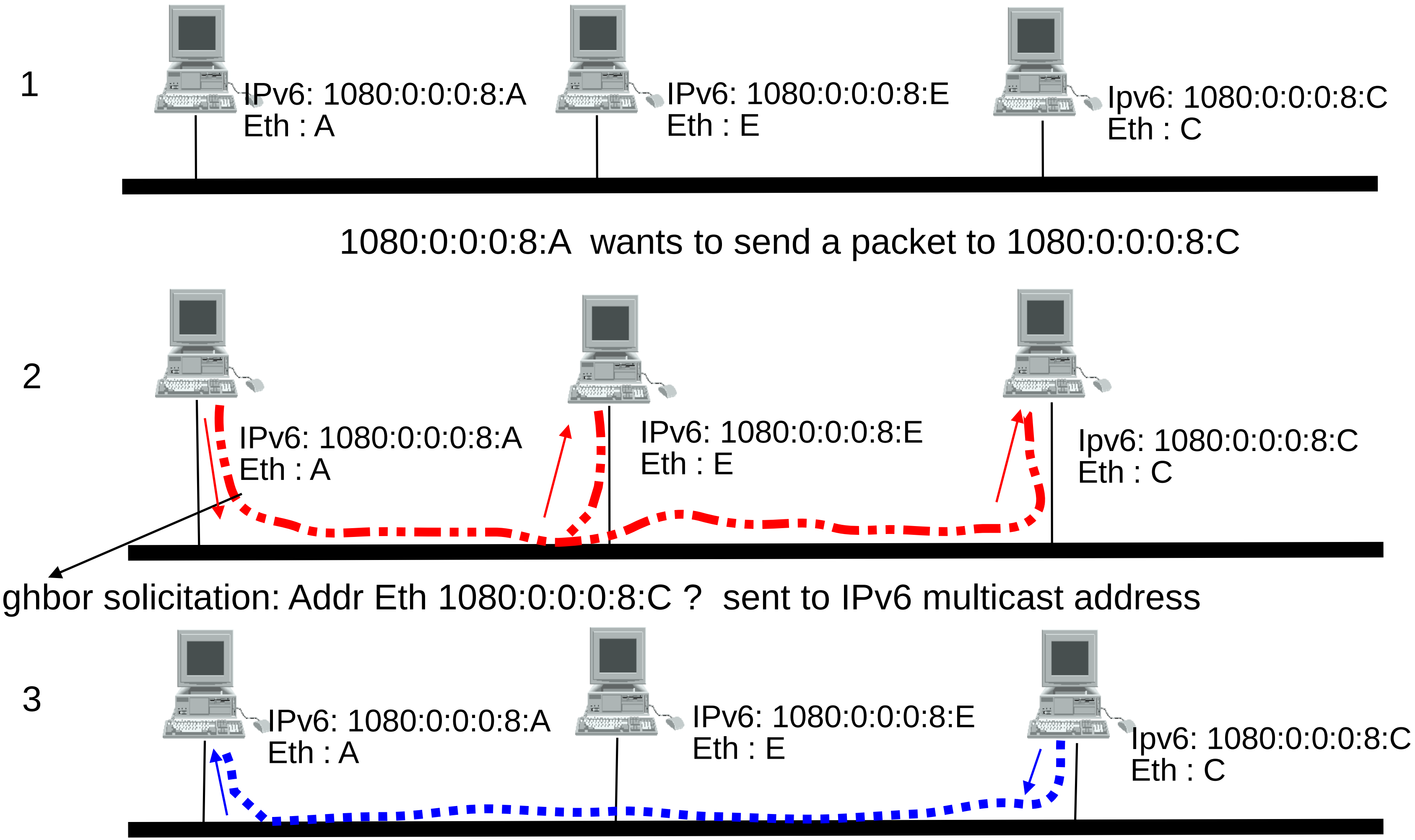
IP: 10.0.1.9/24
Default R: 10.0.1.1
Eth : B
ARP Table
empty

IP: 10.0.1.1/24
Eth : R
ARP Table
empty

- Transmission of an IP packet from 10.0.1.22 to 10.0.1.9
- Transmission of an IP packet from 10.0.1.22 to 10.0.2.9

IPv6 over Ethernet

- Neighbour discovery / address resolution



1080:0:0:0:8:A wants to send a packet to 1080:0:0:0:8:C

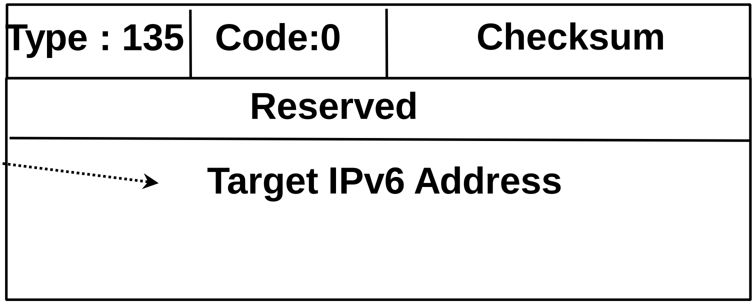
Neighbor solicitation: Addr Eth 1080:0:0:0:8:C ? sent to IPv6 multicast address

Neighbor advertisement: 1080:0:0:0:8:C is reachable via Ethernet Add : C

ICMPv6 Neighbour Discovery

- Replacement for IPv4’s ARP
- Neighbour solicitation
  - Sent to

The IPv6 address for which the link-layer
(e.g. Ethernet) address is needed.
May also contain an optional field with the link-layer
(e.g. Ethernet) address of the sender.

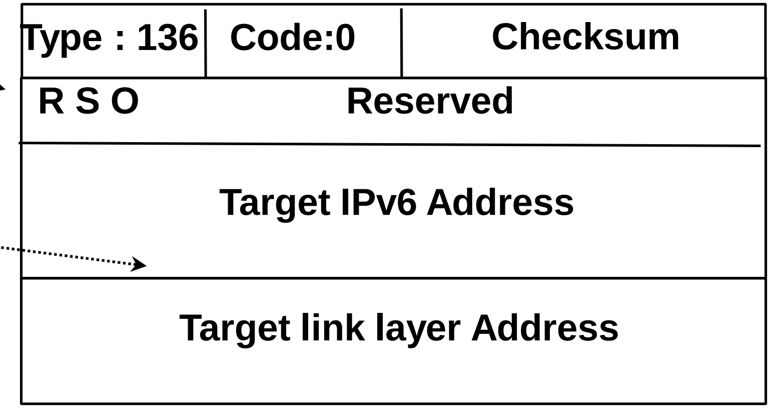


Type : 135	Code:0	Checksum
Reserved		
Target IPv6 Address		

- Neighbour advertisement

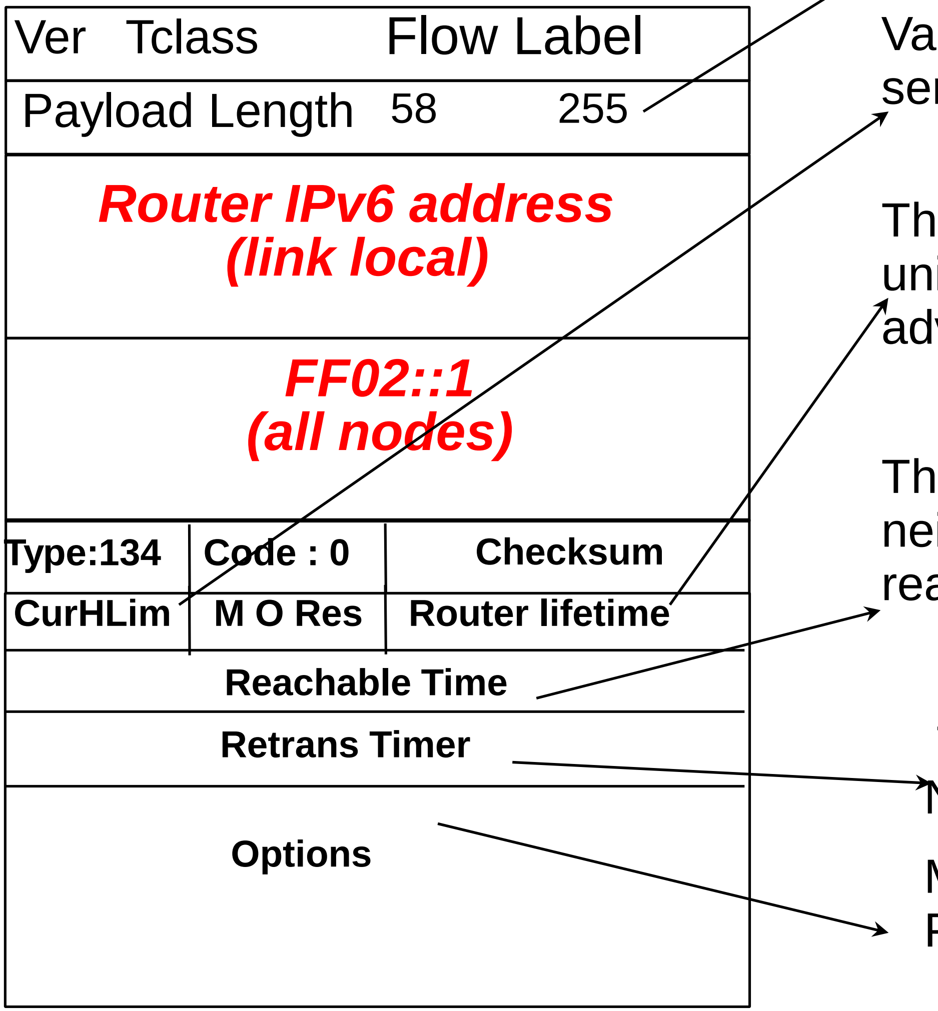
R : true if node is a router
S : true if answers to a neighbour solicitation

The IPv6 and link-layer addresses



Type : 136	Code:0	Checksum
R S O	Reserved	
Target IPv6 Address		
Target link layer Address		

Router advertisements



Ver Tclass Flow Label

Payload Length 58 255

Router IPv6 address
(link local)

FF02::1
(all nodes)

Type:134 **Code : 0** **Checksum**

CurHLim **M O Res** **Router lifetime**

Reachable Time

Retrans Timer

Options

Maximum hop limit to avoid spoofed packets from outside LAN

Value of hop limit to be used by hosts when sending IPv6 packets

The lifetime associated with the default router in units of seconds. 0 is the router sending the advertisement is not a default router.

The time, in milliseconds, that a node assumes a neighbour is reachable after having received a reachability confirmation.

The time, in milliseconds, between retransmitted Neighbor Solicitation messages.

MTU to be used on the LAN
Prefixes to be used on the LAN

Router advertisements options

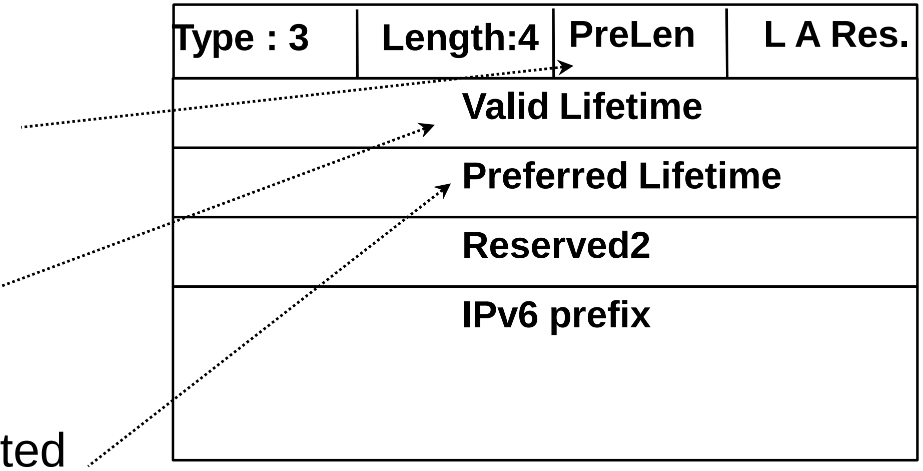
- Format of the options

Type	Length	Options
Options (cont.)		

- MTU option

Type : 5	Length:1	Reserved
MTU		

- Prefix option



Type : 3	Length:4	PreLen	L A Res.
Valid Lifetime			
Preferred Lifetime			
Reserved2			
IPv6 prefix			

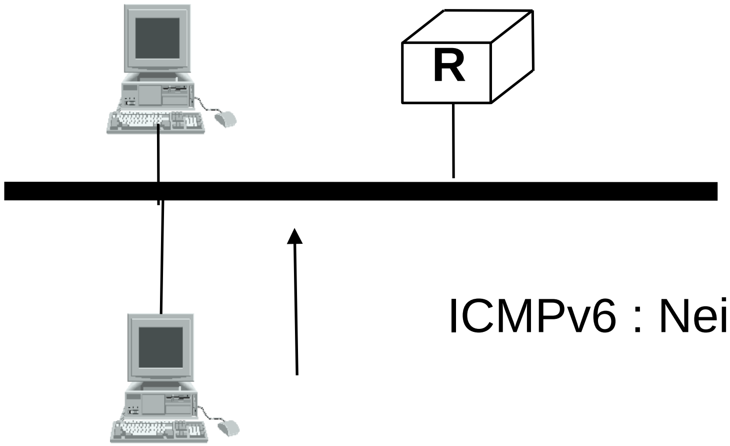
Number of bits in IPv6 prefix that identify subnet

The validity period of the prefix in seconds

The duration in seconds that addresses generated from the prefix via stateless address autoconfiguration remain preferred.

IPv6 autoconfiguration (2)

- What happens when an endsystem boots ?
  - It knows nothing about its current network
    - but needs an IPv6 address to send ICMPv6 messages



ICMPv6 : Neighbour SolicitationSent to multic

Ethernet : 0800:200C:417A

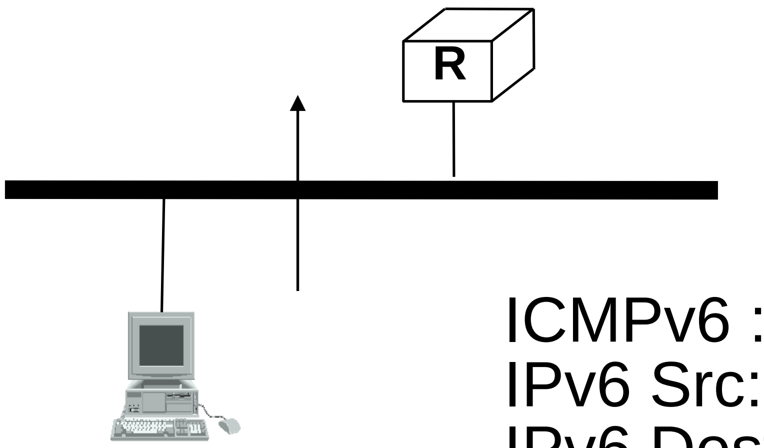
FE80::M_{64}(800:200C:417A))

Address is valid if nobody answers

- Use Link-local IPv6 address (FE80::/64)
  - Each host, when it boots, has a link-local IPv6 address
  - But another node might have chosen the same address !

IPv6 autoconfiguration (2)

- How to obtain the IPv6 prefix of the subnet ?
  - Wait for router advertisements (e.g. 30 seconds)
  - Solicit router advertisement

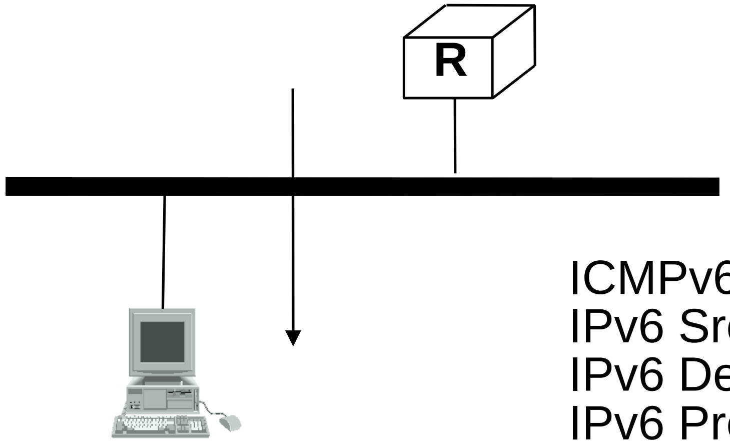


ICMPv6 : Router Solicitation
IPv6 Src: FE80::M_{64}(800:200C:417A)
IPv6 Dest: FF02::2

Ethernet : 0800:200C:417A
FE80::M_{64}(800:200C:417A)

IPv6 autoconfiguration (3)

- Router will re-advertise prefix



ICMPv6 : Router Advertisement
IPv6 Src: FE80::M_{64}(EthernetR)
IPv6 Dest: FF02::1
IPv6 Prefix = 2001:6a8:1100::/48
Prefix lifetime

Ethernet : 0800:200C:417A
FE80::M_{64}(800:200C:417A)

- IPv6 addresses can be allocated for limited lifetime
  - This allows IPv6 to easily support renumbering

Privacy issues with IPv6 address autoconfiguration

- Issue
  - Autoconfigured IPv6 addresses contain the MAC address of the hosts
    - MAC addresses are fixed and unique
    - A laptop/user could be identified by tracking the lower 64 bits of its IPv6 addresses
- How to maintain privacy with IPv6 ?
  - Use DHCPv6 and configure server to never reallocate the same IPv6 address
  - Allow hosts to use random host ids in lower 64 bits of their IPv6 address
    - algorithms have been implemented to generate such random host ids on nodes with and without stable storage

Datalink layer

- Point-to datalink layer
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
    - Basics of Ethernet
    - IP over Ethernet
    - Interconnection of Ethernet networks
  - WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

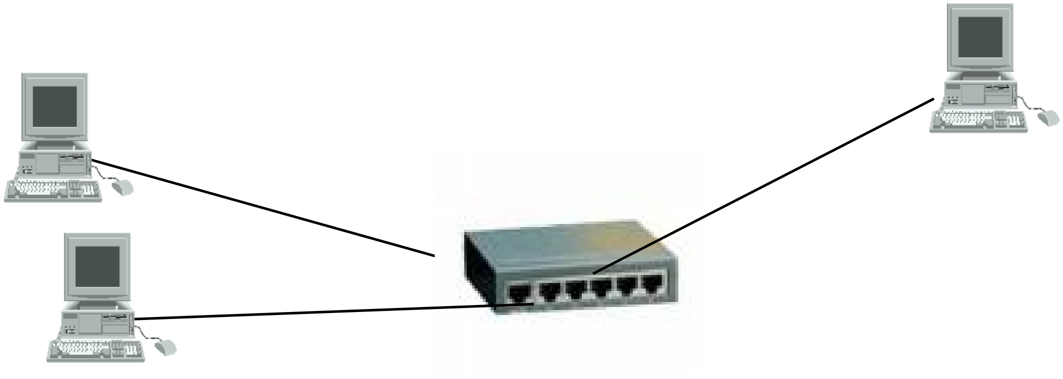
Ethernet today

- The coaxial cable is not used anymore

- Ethernet cabling today
  - Structured twisted pair cabling
  - Optical fiber for some point-to-point links
- Ethernet organisation
  - Not anymore a bus
  - Ethernet is now a star-shaped network !

Ethernet with structured cabling

- How to perform CSMA/CD in a star-shaped network ?

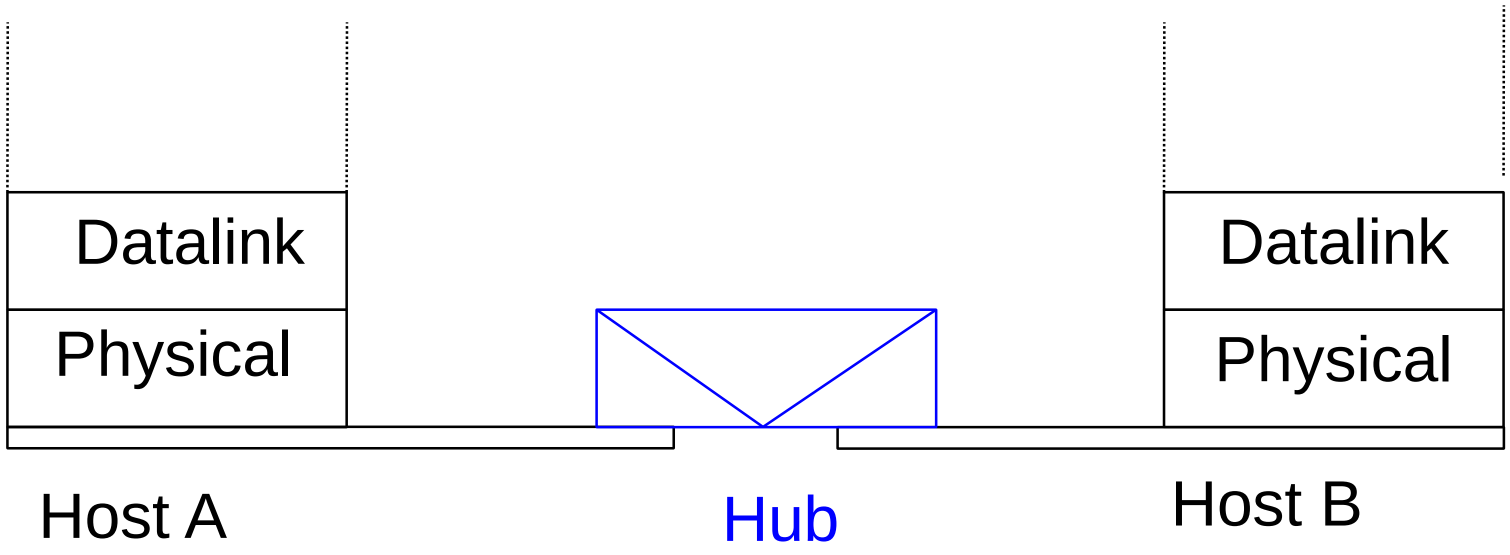


Hub :
receives electrical signal on one port,
regenerates this signal and forwards it over all
other ports besides the port from which it
received it

Collision domain : set of stations that could be in collision

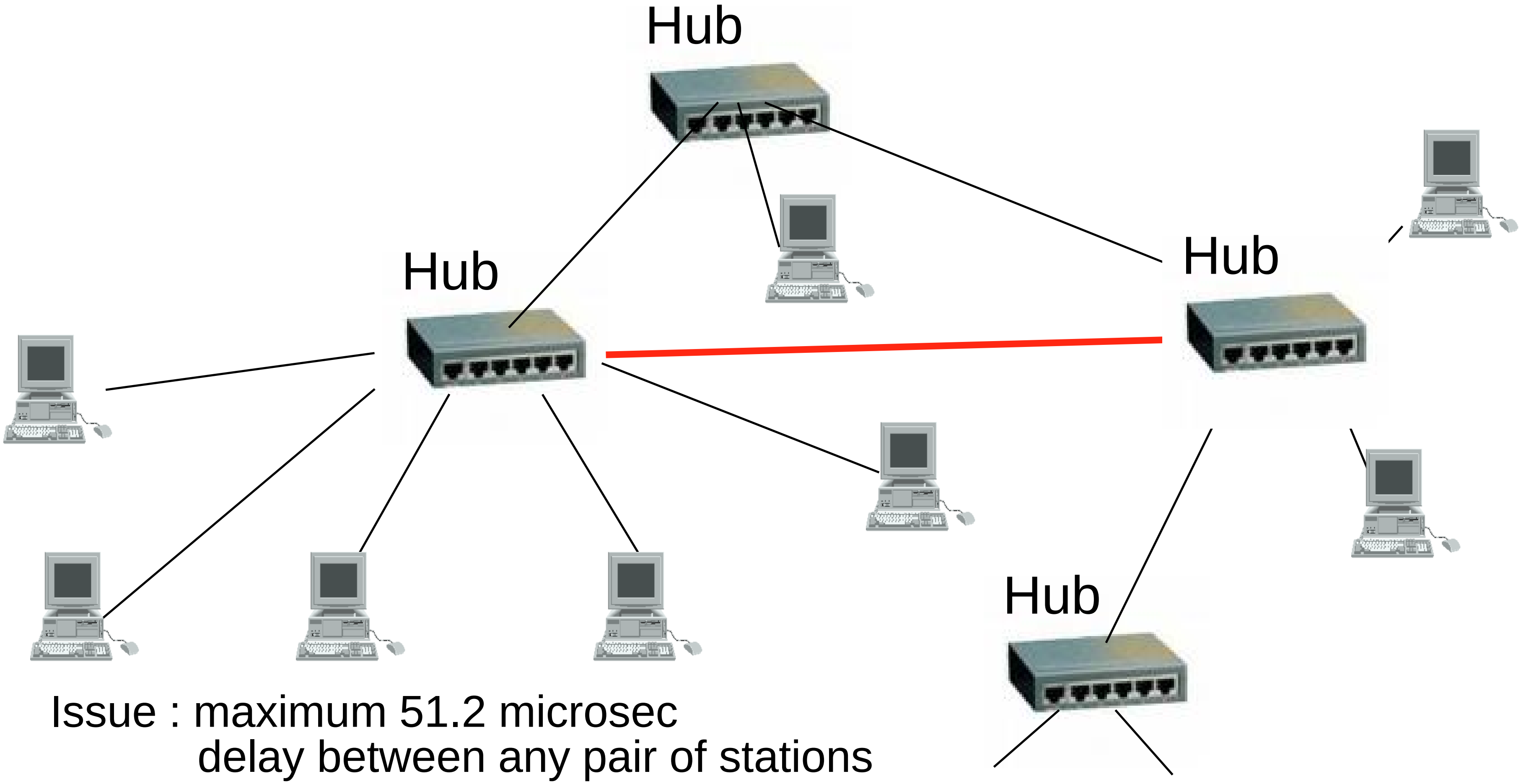
Hub and the reference model

- A hub is a relay operating the physical layer



Ethernet with structured cabling (2)

- How to build a larger Ethernet network ?
  - Interconnect hubs together

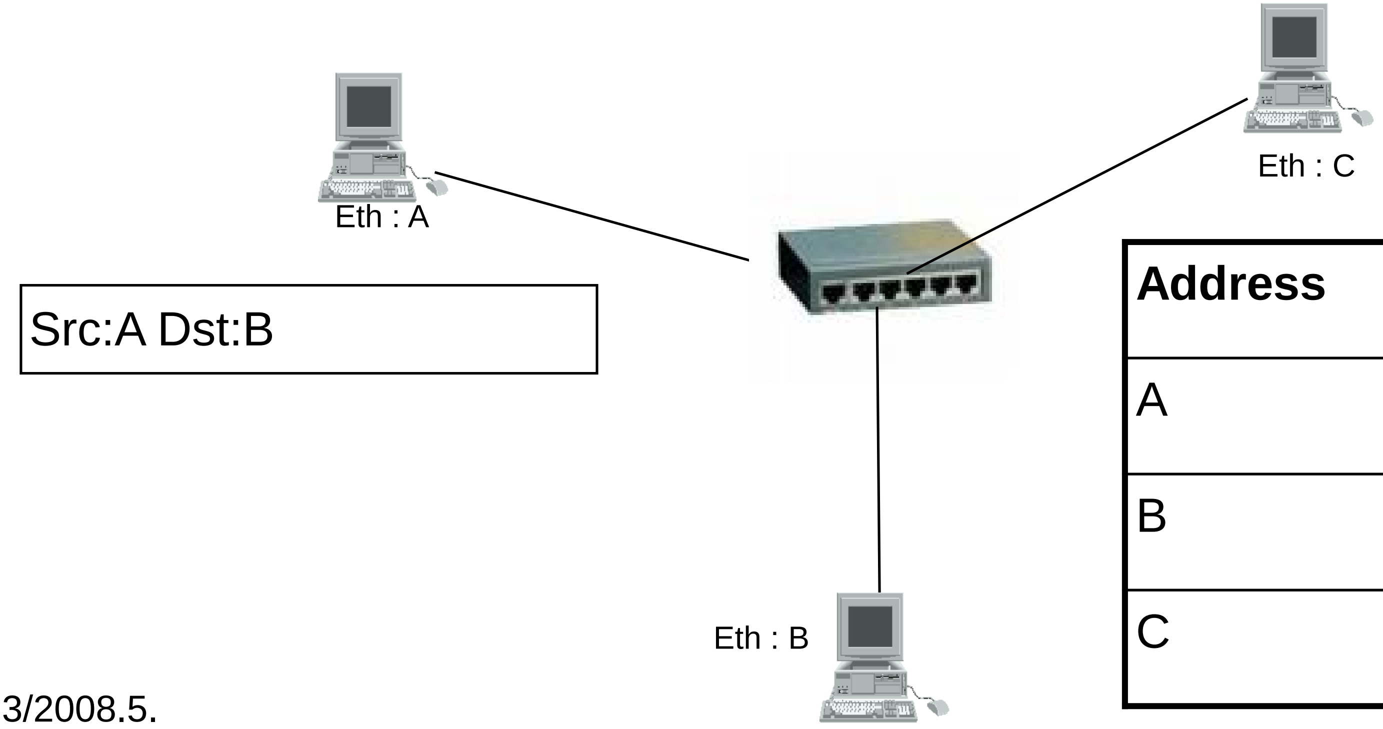


Issue : maximum 51.2 microsec
delay between any pair of stations

Collision domain : entire network

Ethernet Switch

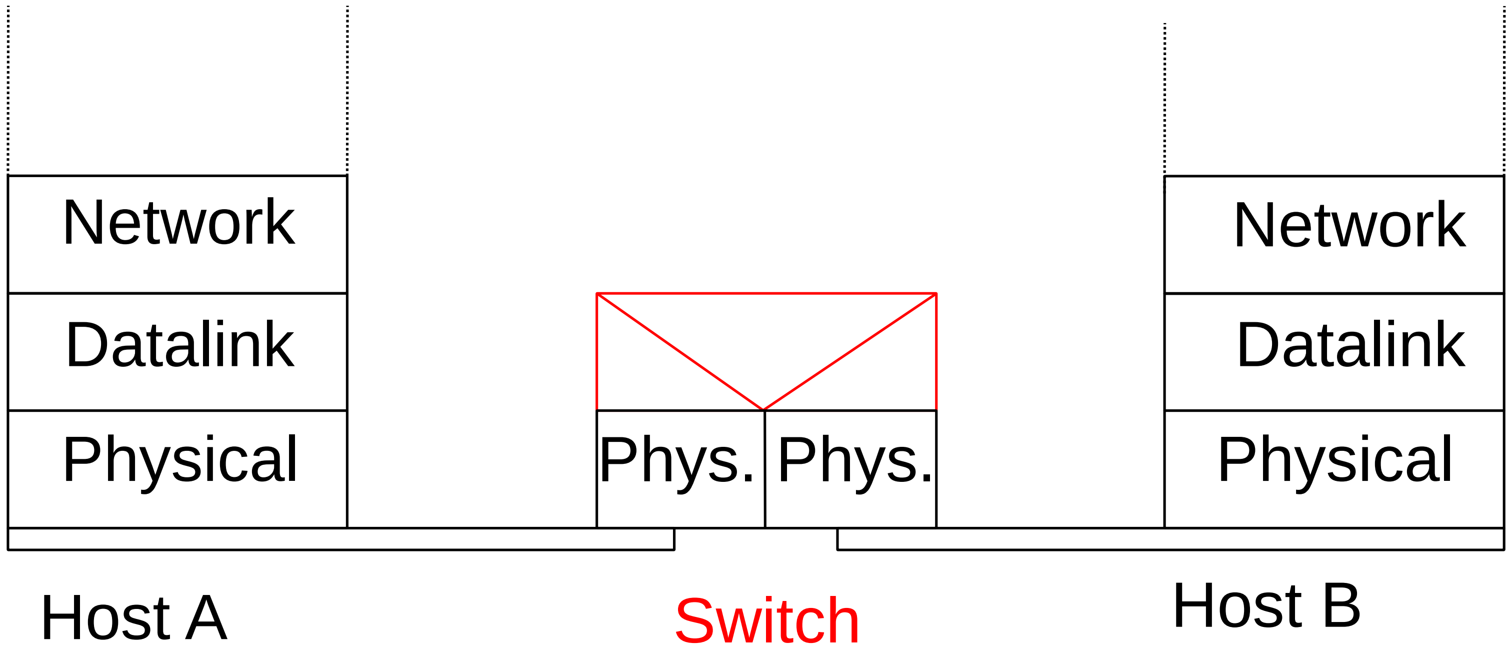
- Can we improve the performance of hubs ?
- Ethernet switch
  - Operates in the datalink layer
  - understands MAC address and filters frames based on their addresses



Address	Port
A	West
B	South
C	East

Switch in the reference model

- A switch is a relay that operates in the datalink layer



Port-address table

- How to build the port-address table used by Ethernet switches ?
- Manually
  - Works in a lab, but Ethernet must be plug and play
- Automatically
  - Frame source address allows switch to learn the location of hosts
  - What happens when a destination address cannot be found in the port-address table ?
  - But be careful to age the information inside tables as some hosts move from one port to another

- How to forward broadcast frames ?
- How to forward multicast frames ?

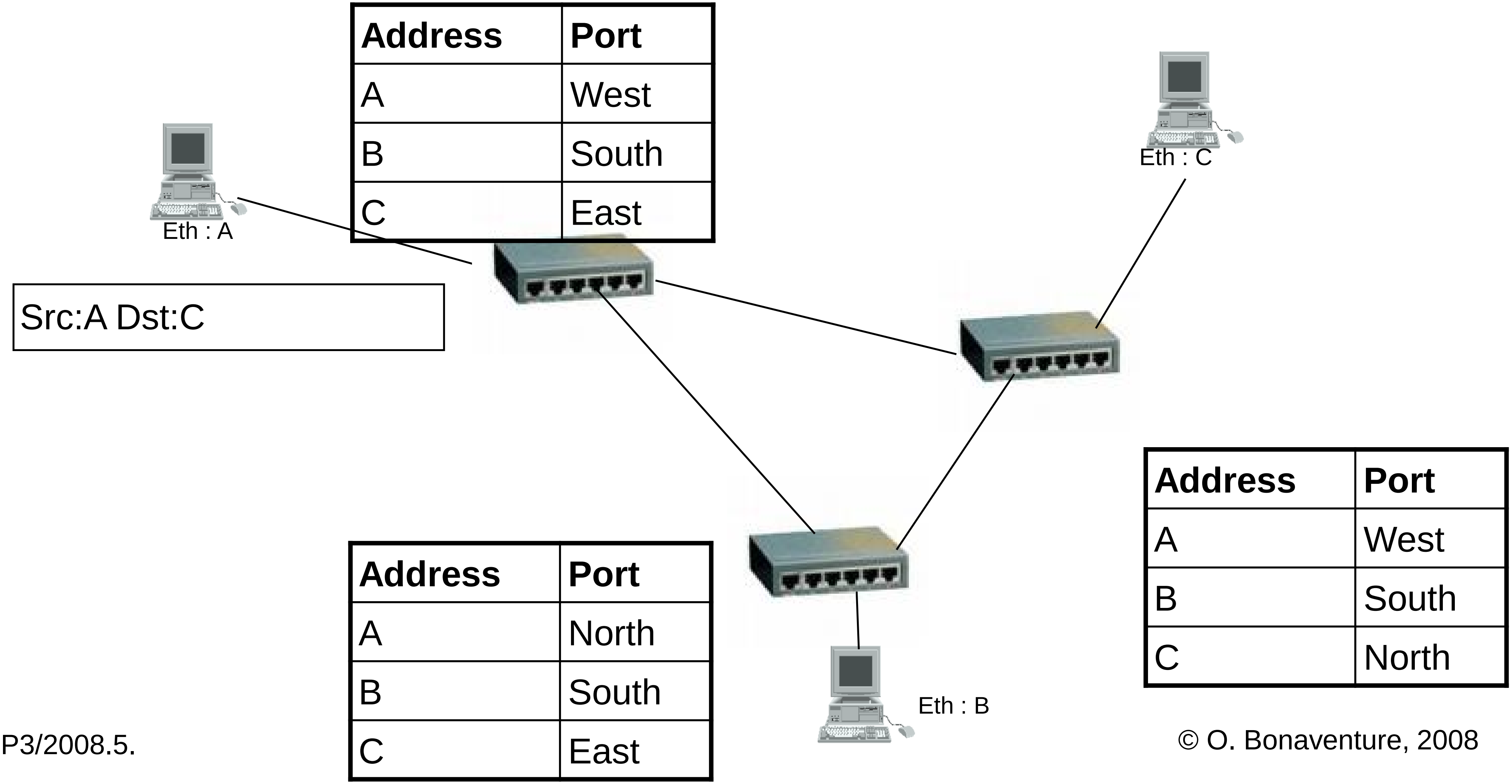
Frame processing

- Basic operation of an Ethernet switch

```
Arrival of frame F on port P
src=F.Source_Address;
dst=F.Destination_Address;
UpdateTable(src, P); // src heard on port P
if (dst==broadcast) || (dst is multicast)
{
for(Port p!=P)       // forward all ports
    ForwardFrame(F,p);
}
else
{
 if(dst isin AddressPortTable) {   ForwardFrame(F,AddressPortTab
}
```

Network redundancy

- How to design networks that survive link and node failures ?
  - Add redundant switches



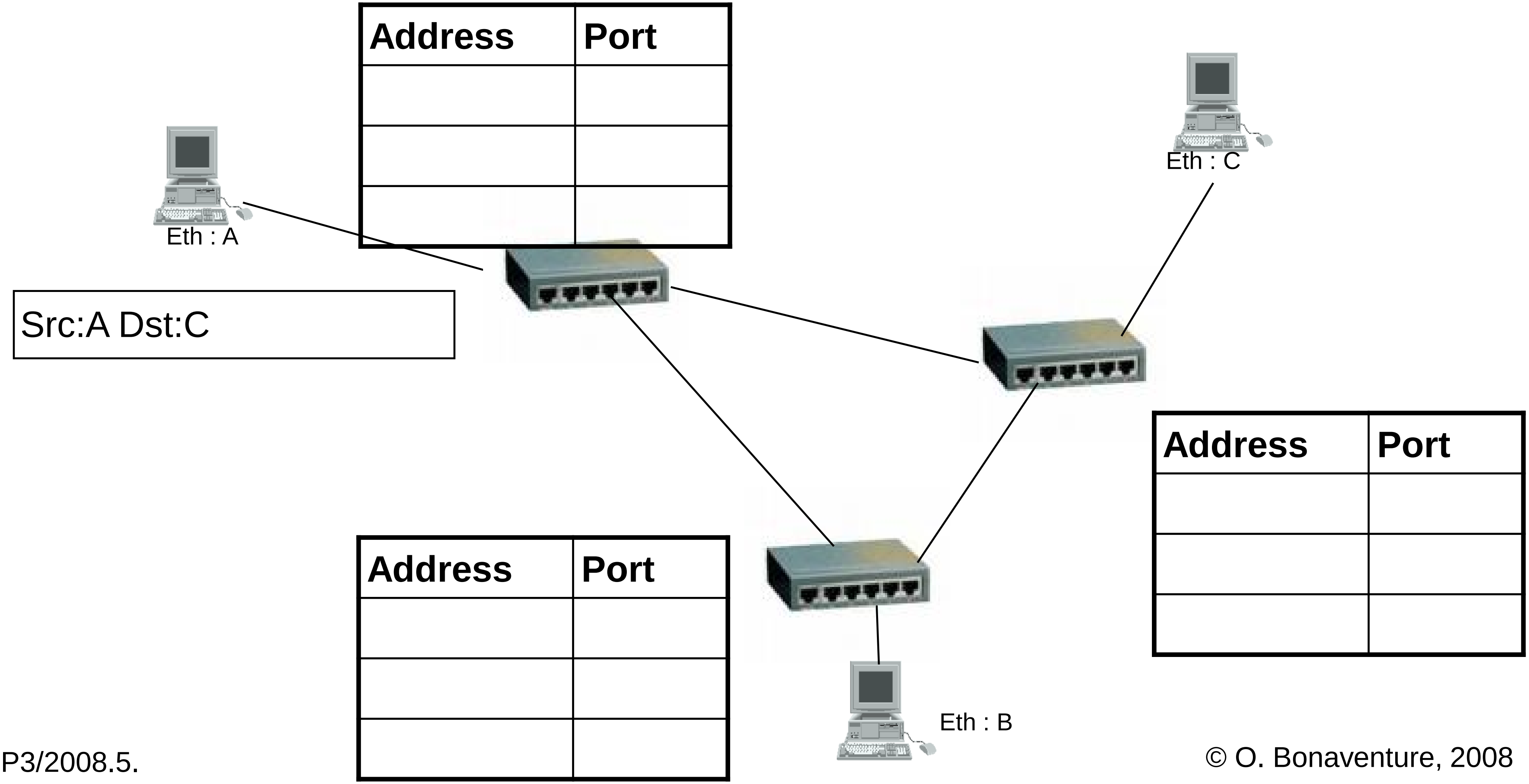
Address	Port
A	West
B	South
C	East

Address	Port
A	North
B	South
C	East

Address	Port
A	West
B	South
C	North

Network redundancy (2)

- Does this always work ?
  - Assume all switches have rebooted



Address	Port

Src:A Dst:C

Address	Port

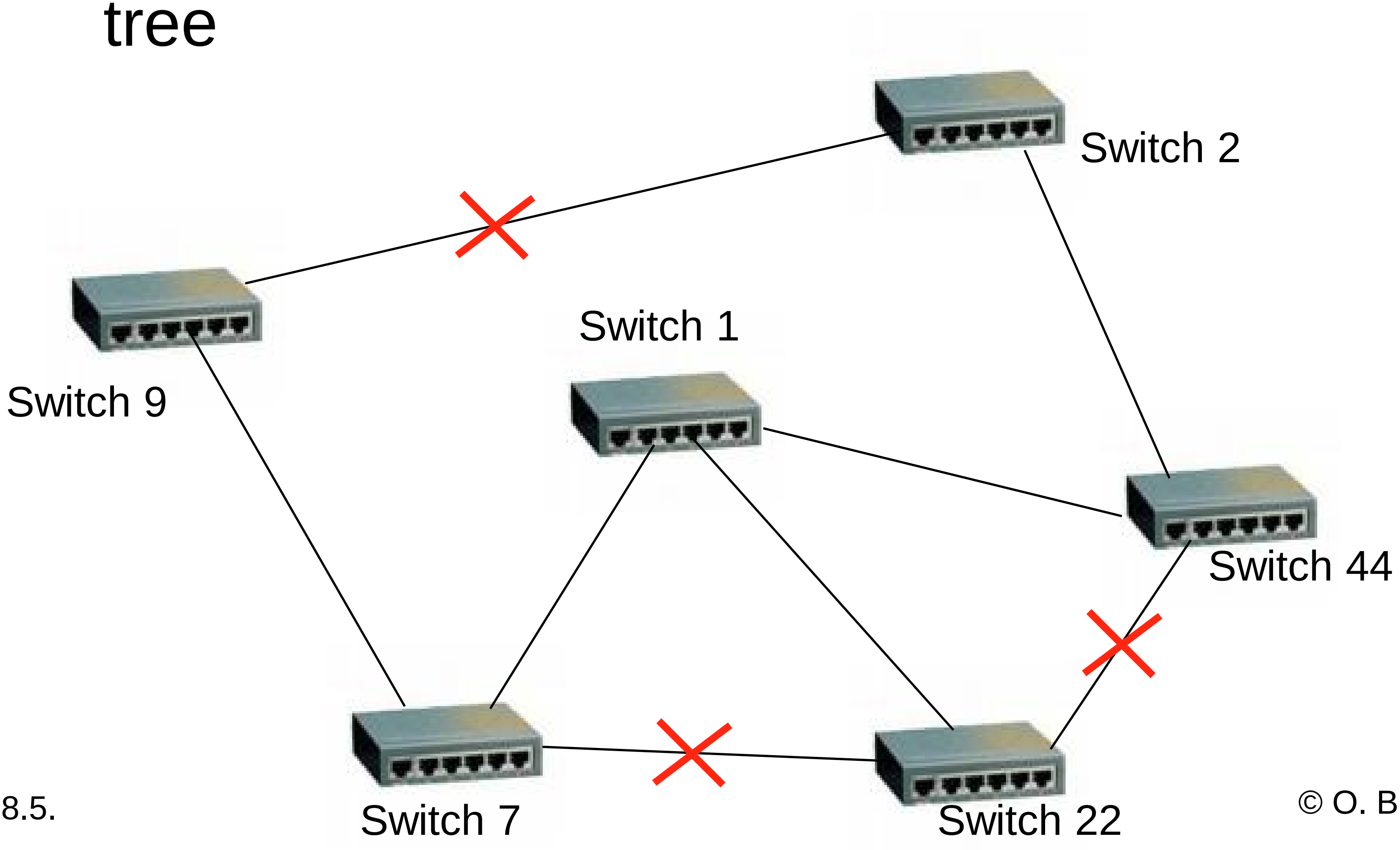
Address	Port

How to solve this problem ?

- The lawyer's way
  - Add a sticker on all switches to indicate that they must only be used in tree shaped networks and should never ever be interconnected with loops
- The computer scientist's way
  - Define a distributed algorithm that allows switches to automatically discover the links causing loops and remove them from the topology

Principle of the solution

- Build a spanning tree inside network
  - Each switch has a unique identifier
  - The switch with the lowest id is the root
  - Disable all links that do not belong to spanning tree



How to build the spanning tree

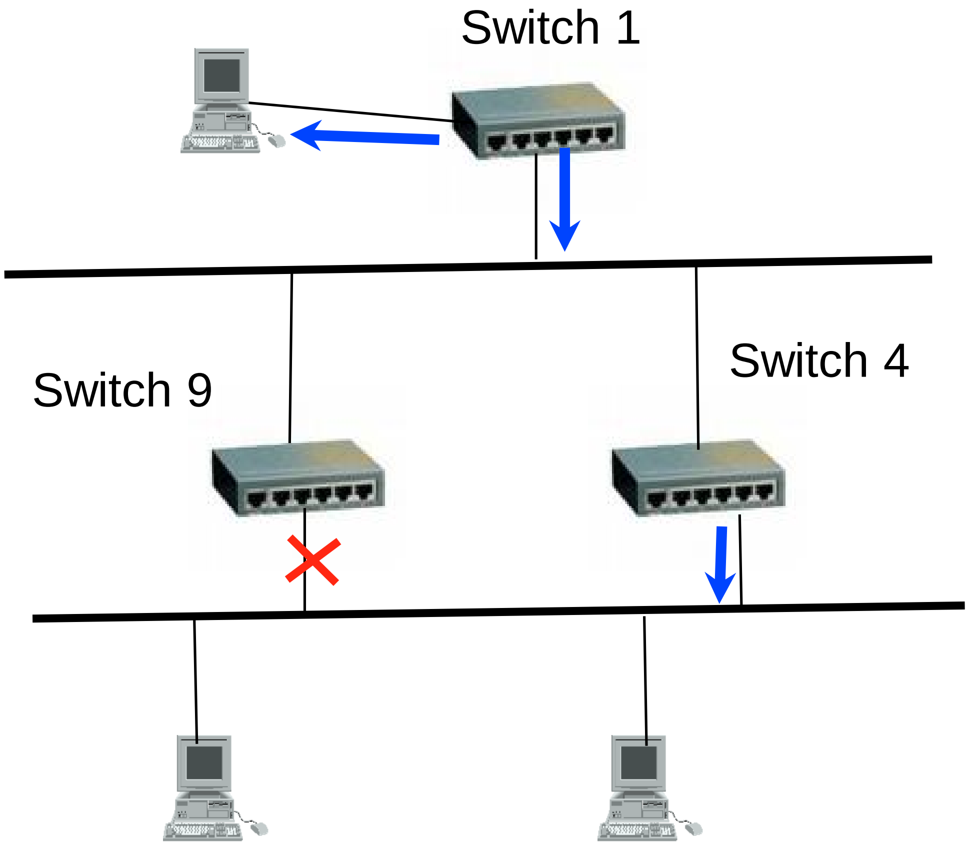
- Distributed algorithm run by switches

- Goals of the spanning tree protocol
  - Elect the root of the spanning tree
    - In practice, this will be the switch with the lowest id
  - Compute the distance between each switch and the root
  - When several switches are attached to the same LAN elect one forwarder and disable the others
  - determine which ports/links should belong to the spanning

Root and Designated Switches

- Root switch
  - The Root Switch is the root of the spanning tree
  - The Root switch may change upon the arrival of new switches in the network

- Designated switch
  - to avoid loops, only one switch should be responsible for forwarding frames from the root on any link
  - Root switch is always designated switch for all its links



The switch identifiers

- Switch identifiers must be unique
  - The easiest solution is to ask each manufacturer to embed a unique Ethernet address on each switch
- But since the switch with the lowest identifier is the network root, network operators need to influence the selection of the root switch
- 64 bits switch identifier
  - Upper 16 bits
    - Priority defined by operator (default value : 32768)
  - Lower 48 bits
    - Unique Ethernet address assigned by manufacturer

The link costs

- Each switch port is attached to a link
- The costs of the links can be configured on each link by the network operator
  - Common guideline : Cost = 1000 / bandwidth
- Recommended values of link costs

Bandwidth	Recommended link cost range	Recommended link cost value
10 Mbps	50-600	100
100 Mbps	10-60	19
1000 Mbps	3-10	4

Building the spanning tree

- 802.1d protocol
  - 802.1d uses Bridge PDUs (BPDUs) containing
    - Root ID : identifier of the current root switch
    - Cost : Cost of the shortest path between the switch transmitting the BPDU and the root switch
    - Transmitting ID : identifier of the switch that transmits the BPDU
  - The BPDUs are sent by switches over their attached LANs as multicast frames but they are never forwarded
    - switches that implement 802.1d listen to a special Ethernet multicast group

Ordering of BPDUs

- BPDUs can be strictly ordered
  - BPDU11[R=R1,C=C1, T=T1] is better than BPDU2 [R=R2,C=C2, T=T2] if
    - R1<R2
    - R1=R2 and C1<C2
    - R1=R2 and C1=C2 and T1<T2
  - Example

BPDU1			BPDU2		
R1	C1	T1	R2	C2	T2
29	15	35	31	12	32
35	80	39	35	80	40
35	15	80	35	18	38

Building the spanning tree (2)

- Behaviour of 802.1d protocol
  - The root switch sends regularly BPDUs on all its ports
    - R=Root switch id, C=0, T= Root switch id
    - Bootstrap
      - If a switch does not receive BPDUs, it considers itself as root and sends BPDUs
  - On each port, a switch parses all the received BPDUs and stores the best BPDU received on each port
    - Each switch can easily determiner the current root by analysing all the BPDUs stored in its tables
  - A switch stops sending BPDUs on a port if it received a better BPDU on this port
  - 802.1d stabilises when a single switch sends a BPDU over each LAN

802.1d port states

- 802.1d port state based on received BPDUs
  - Root port
    - port on which the best 802.1d BPDU was received
    - port used to receive the BPDUs sent by the root form the shortest path
    - A root port does not transmit BPDUs
    - Only one root port on each switch
  - Designated port
    - port(s) used to send switch's BPDU upon reception of a BPDU from the root via the Root port
    - Switch's BPDU is
      - current root, cost to reach root, switch identifier
    - 0, one or more designated ports on each switch
    - a port is designated if the switch's BPDU is better than the best BPDU received on this port
  - Blocked port (only receives 802.1d BPDUs)

802.1d port states (2)

- Example
  - BPDUs received by switch 18

	Root	Cost	Transmitter
port1	12	93	51
port2	**12**	**85**	**47**
port3	81	0	81
port4	15	31	27

  - Root : switch 12
  - port2 is the root port
  - Switch’s BPDU
    - R=12, C=86, T=18
  - This BPDU is better than the BPDUs received on the other ports. They are thus designated

802.1d port states (3)

- Example
  - BPDUs received by switch 92

	Root	Cost	Transmitter
port1	81	0	81
port2	41	19	125
port3	41	12	315
port4	41	12	111
port5	41	13	90

  - root : 41
  - root port : port4
  - Switch’s BPDU
    - R=41,C=13, T=92
  - Port state
    - port1 and port 2 : designated
    - port 3 and port 5 : blocked

Port activity

- A port can be either active or inactive for data frames
  - Active port
    - The switch captures Ethernet frames on its active ports and forwards them over other ports (based on its own port/address tables)
    - The switch updates its port/address table based on the frames received on this port
  - Inactive port
    - The switch does not listen to frames neither forward frames on this port
- The port activity is fixed once the spanning tree has converged
  - Root and designated ports become active
  - Blocked ports become inactive
  - Duration spanning tree computation, all ports are inactive

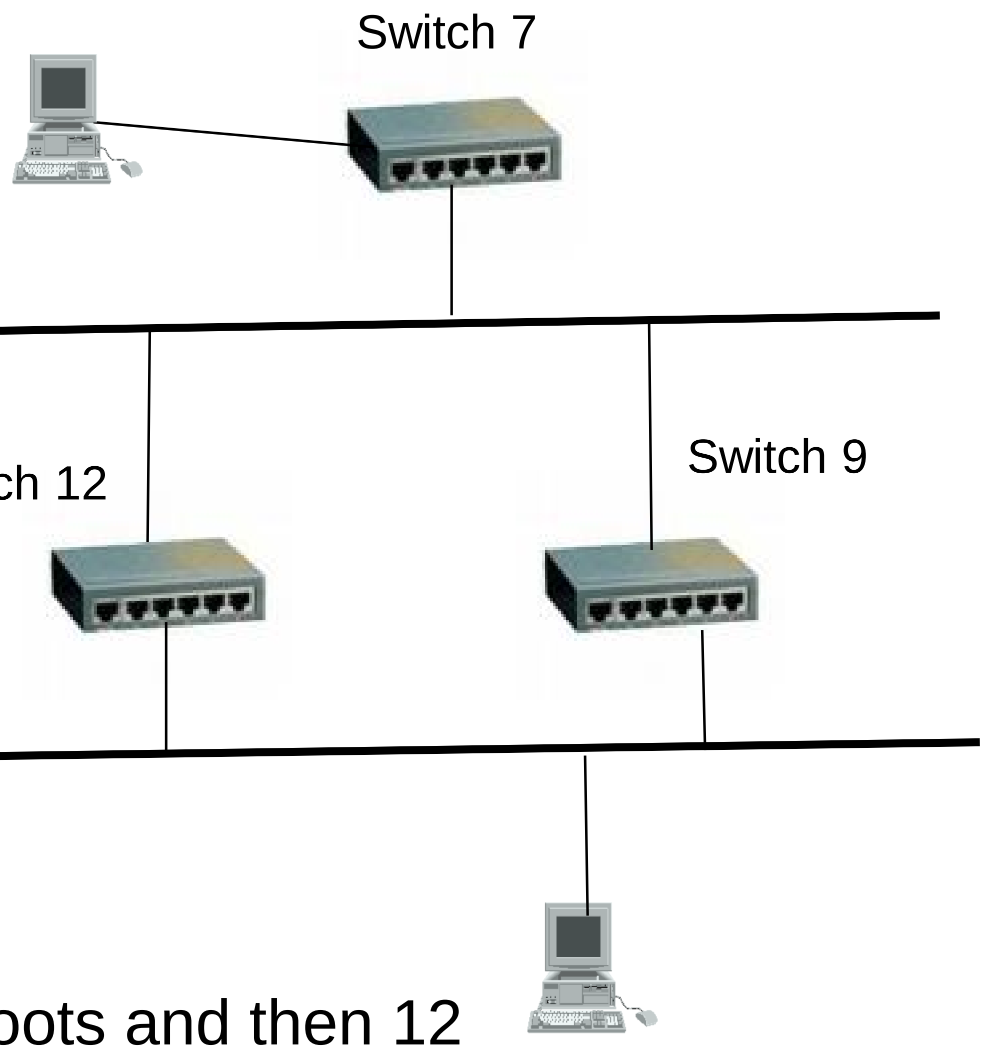
Port states and activity

	Receive BPDUs	Transmit BPDUs
Blocked	yes	no
Root	yes	no
Designated	yes	yes

	Learn Addresses	Forward Data Frames
Inactive	no	no
Active	yes	yes

Example network

- Compute the spanning tree in this network by using 802.1d



  - Assume that 9 boots and then 12 and eventually 7 boots

Impact of failures

- What kind of failures should be considered ?
  - Failure (power-off) of the root switch
    - A new root needs to be elected
  - Failure of a designated switch
    - Another switch should replace the designated one
  - Failure of a link
    - If the network is redundant, a disabled link should be enabled to cope with the failure
  - Failure of a link that disconnects the network
    - We now have two different networks and a root switch must be elected in each network

How to deal with failures ?

- Failure detection mechanisms
  - Root switch sends its BPDU every Hello timer and designated switches generate their own BPDUs upon reception of this BPDU
    - Default Hello timer is two seconds
  - BPDUs stored in the switches age and are removed when they timeout
- Failure notification mechanism
  - When a switch detects an important failure, it sends a topology change (TC) BPDU to the Root
  - Upon reception of a TC BPDU all switches stop forwarding data frames and recompute spanning tree

Ethernet Evolution

- Networks require higher bandwidth

- Fast Ethernet
  - Physical layer
    - bandwidth : 100 Mbps
    - twisted pair or optical fiber
    - No coaxial cable anymore

  - MAC sublayer
    - CSMA/CD unchanged
      - minimum frame size : 512 bits
      - slot time : 5.12 micro seconds
    - Maximum distance : shorter than Ethernet 10 Mbps
    - Same frame format as 10 Mbps Ethernet

Ethernet Evolution (2)

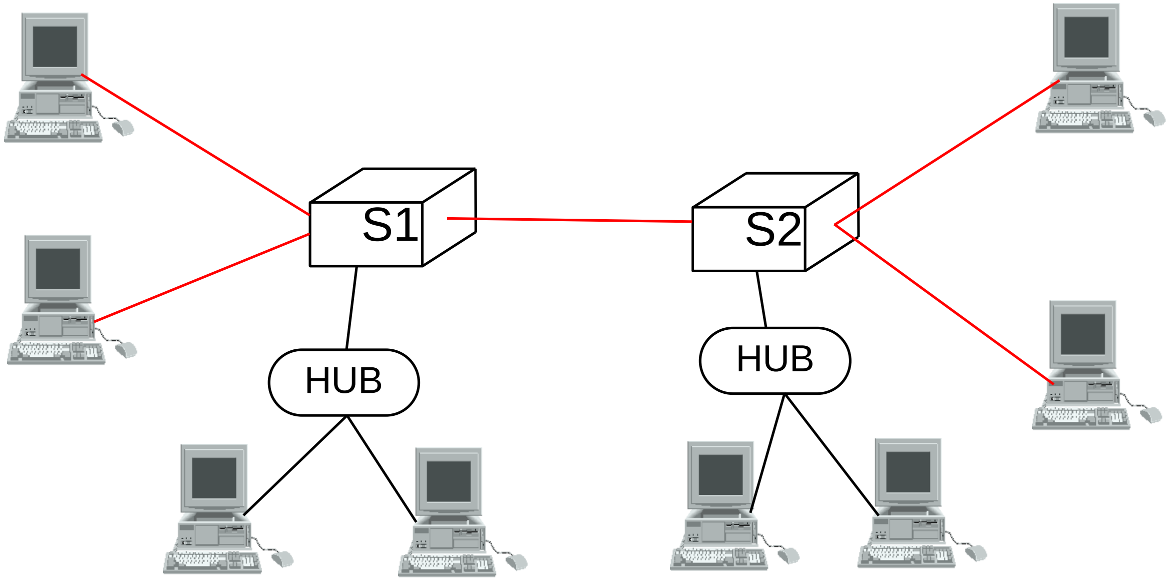
- Gigabit Ethernet
  - Physical layer
    - Bandwidth 1 Gbps
    - Optical fiber or twisted pair
  - MAC sublayer
    - CSMA/CD still supported
      - How was this achieved ?
      - Two options
        - Increase minimum frame size : not backward compatible with Ethernet
        - Reduce the maximum distance as for FastEthernet : but then networks would have a diameter of 10 m
      - Gigabit CSMA/CD hack
        - minimum frame size is still 512 bits but the sender must continue to send an electrical signal during the equivalent o 4096 bits
    - same frame format as Ethernet
      - but extensions allow to transmit Jumbo frames of up to 9KBytes

The Ethernet zoo

10BASE5	Thick coaxial cable, 500m
10BASE2	Thin coaxial cable, 185m
10BASE-T	Two pairs of category 3+ UTP
10BASE-F	10 Mb/s over optical fiber
100BASE-TX	Category 5 UTP or STP, 100 m maximum
100BASE-FX	Two multimode optical fiber, 2 km maximum
1000BASE-CX	Two pairs shielded twisted pair, 25m maximum
1000BASE-SX	Two multimode or single mode optical fibers with lasers
10 Gbps	optical fiber but also cat 6 twisted pair
40-100 Gbps	being developed, standard expected in 2010, 40Gbps one meter long for switch backplanes, 10 meters for copper cable and 100 meters for fiber optics

Full duplex Ethernet

- Observations
  - In many networks, Ethernet is a often a point-to-point technology
    - host-to-switch
    - switch to switch



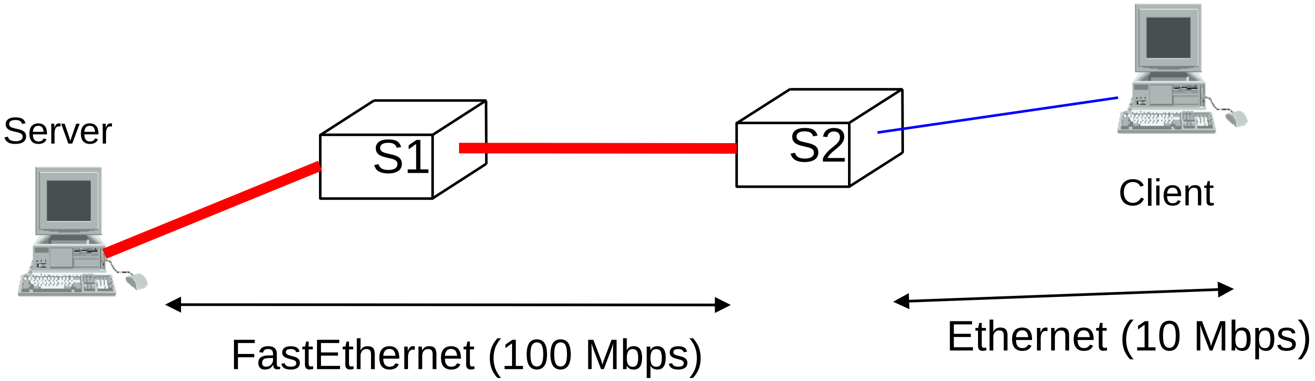
  - Twisted-pairs and fiber-based physical layers allow to send and receive at the same time

Ethernet full duplex (2)

- No collision is possible on a full duplex Ethernet/FastEthernet/GigabitEthernet link
  - Disable CSMA/CD on such links
- Advantages
  - Improves bandwidth
    - Both endpoints can transmit frames at the same time
  - CSMA/CD is disabled
    - No constraint on propagation delay anymore
      - Ethernet network can be as large as we want !
    - No constraint on minimum frame size anymore
      - We do not need the frame extension hack for Gigabit Ethernet!

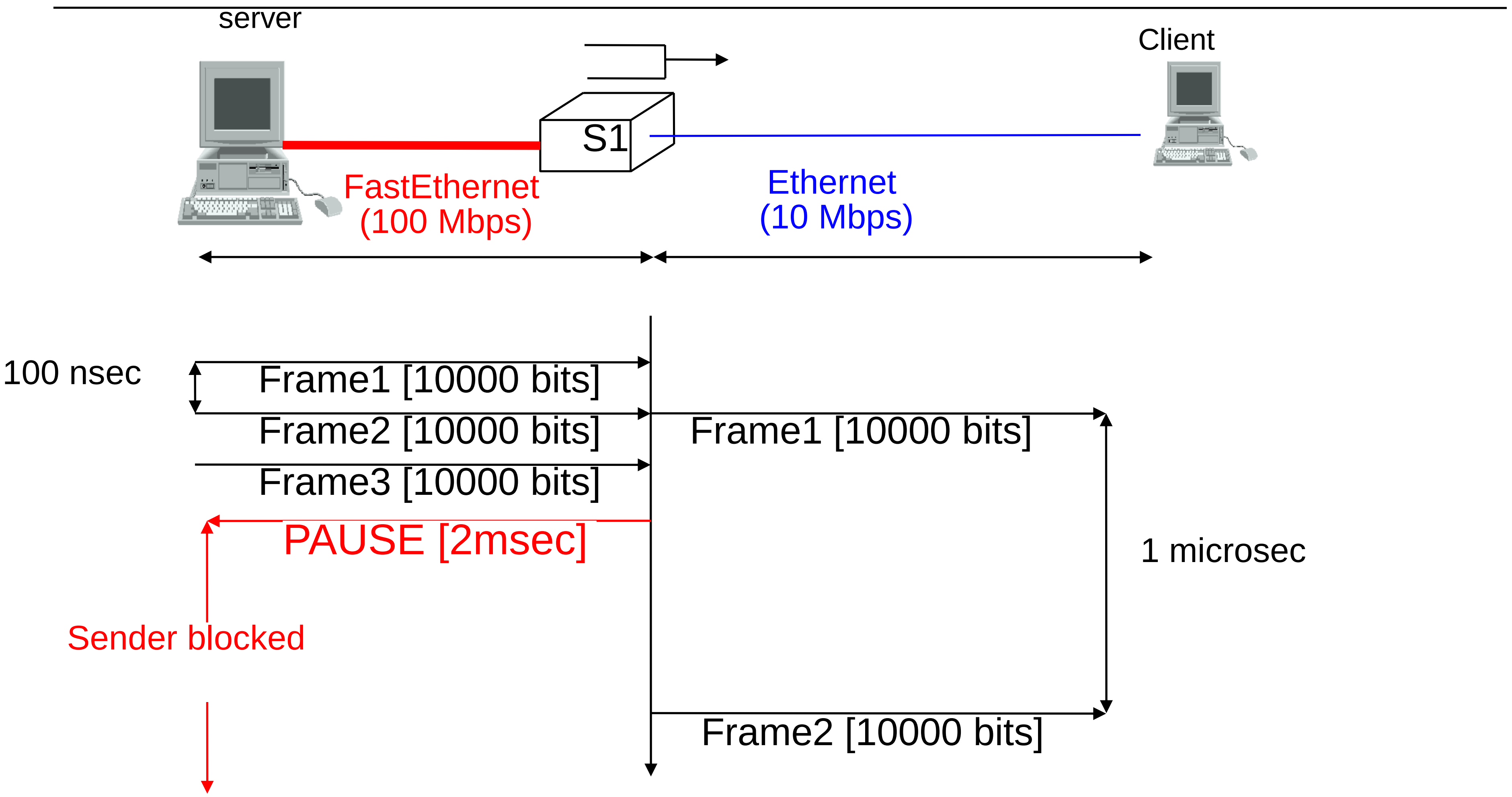
Full duplex Ethernet (3)

- Drawback
  - If CSMA/CD is disabled, access control is disabled and congestion can occur



- How to solve this problem inside Ethernet ?
  - Add buffers to switches
    - but infinite buffers are impossible and useless anyway
  - Cause collisions (e.g. jamming) to force collisions on the inter-switch link and uplink is server is too fast
    - Drawback : interswitch link could be entirely blocked
  - Develop a new flow control mechanism inside MAC layer
    - Pause frame to slowdown transmission

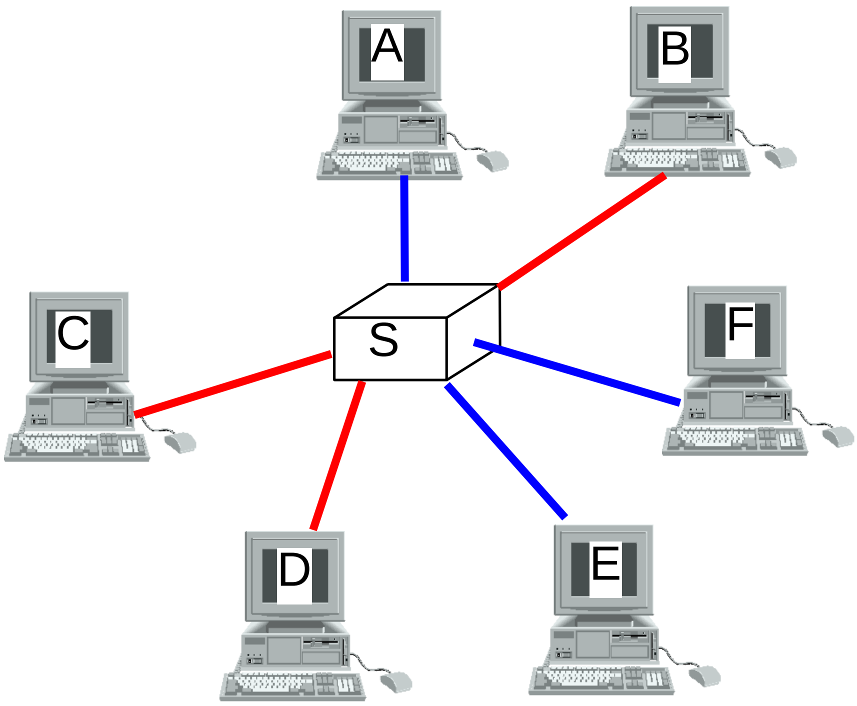
Ethernet flow control



- PAUSE frame indicates how much time the upstream should wait before transmitting next frame

Virtual LANs

l Allows to build several logical networks on top of a single physical network



VLAN1 : A,E,F
VLAN2 : B,C,D

l Each port on each switch is associated to a particular VLAN

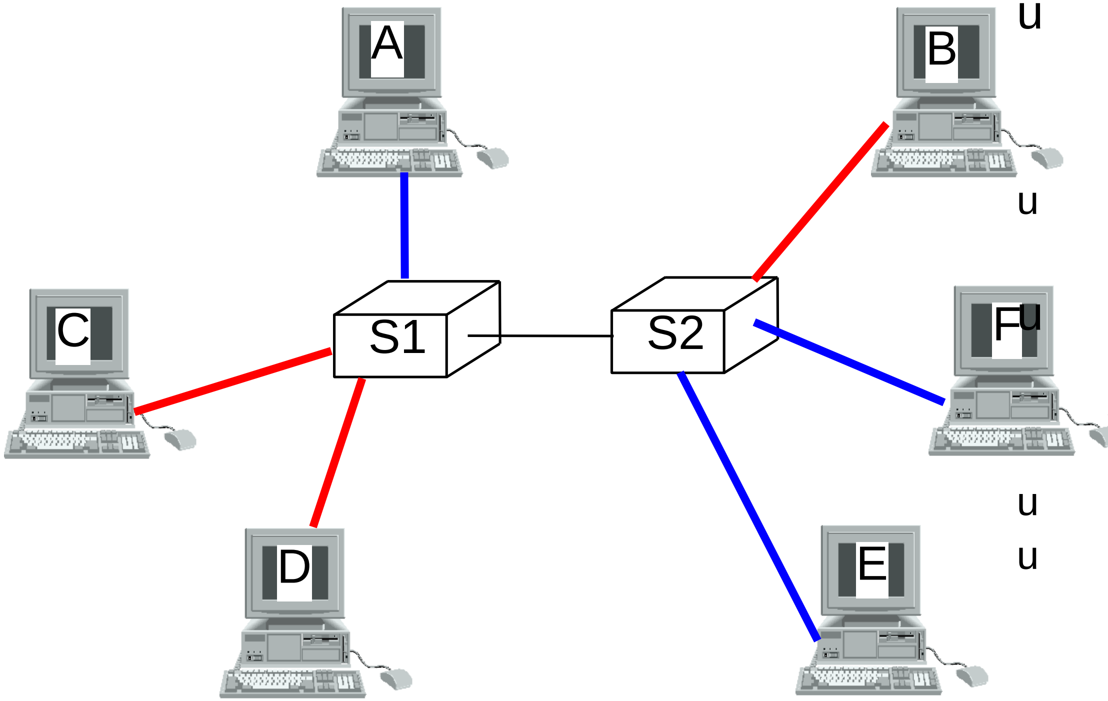
u All the hosts that reside on the same VLAN can exchange Ethernet frames

u A host on VLAN1 cannot send an Ethernet frame towards another host that belongs to VLAN2

u Broadcast and multicast frames are only sent to the members of the VLAN

VLANs in campus networks

l How to support VLANs in a campus network



VLAN1 : A,E,F
VLAN2 : B,C,D

l Possible solutions

u Place on each switch a table that maps each MAC address on a VLAN id

u difficult to manage this table

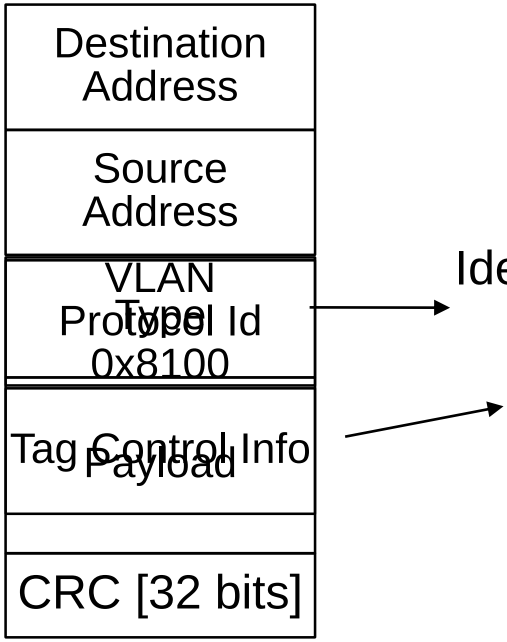
u Change frame format used on inter-switch links to include a VLAN identifier

u new header added by first switch

u new header removed by last switch

VLAN frame format

- Used on inter-switch links



Identifies the frame as containing VLANtag

Tag control information contains two types of information :
- VLAN identifier (12 bits) : up to 4094 different VLANs can be defined
- Priority (3 bits) : indicates the importance of the frame and can be used by switches to provide a better service for some frames (e.g. Voice)

- Can also be used by trusted hosts (e.g. servers) or routers

Datalink layer

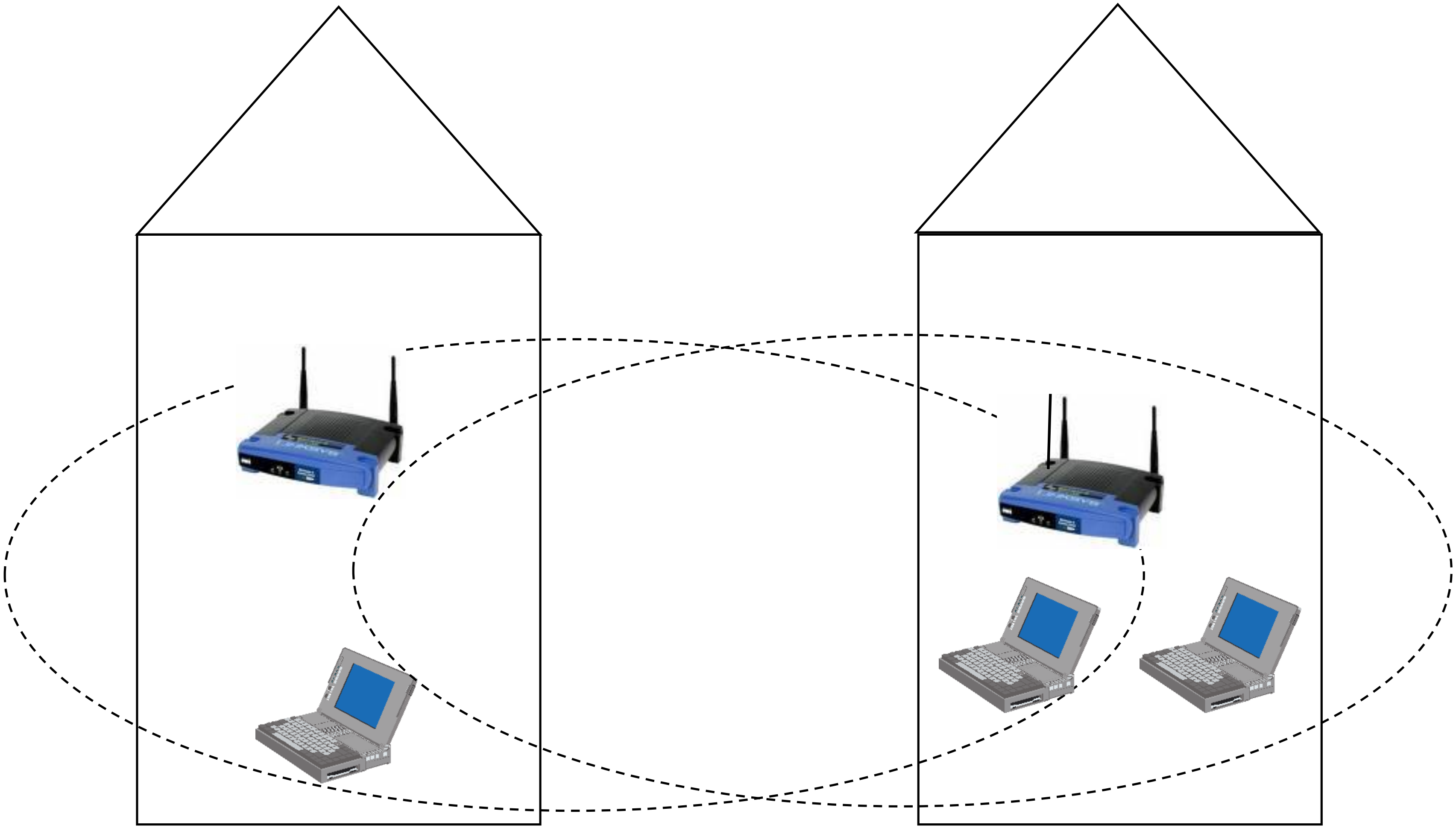
- Point-to datalink layer
- Local area networks
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA
  - Ethernet networks
    - Basics of Ethernet
    - IP over Ethernet
    - Interconnection of Ethernet networks
  - → WiFi networks
  - Deterministic Medium access control
    - Token Ring, FDDI

The WiFi zoo

Standard	Frequency	Typical throughput	Raw bandwidth	Range in/out (m)
802 .11	2.4 GHz	0.9 Mbps	2 Mbps	20 / 100
802 .11a	5 GHz	23 Mbps	54 Mbps	35 / 120
802 .11b	2.4 GHz	4.3 Mbps	11 Mbps	38 / 140
802 .11g	2.4 GHz	19 Mbps	54 Mbps	38 / 140
802 .11n	2.4 / 5 GHz	74 Mbps	up to 600 Mbps	70 / 250

Practical issues with WLAN deployments

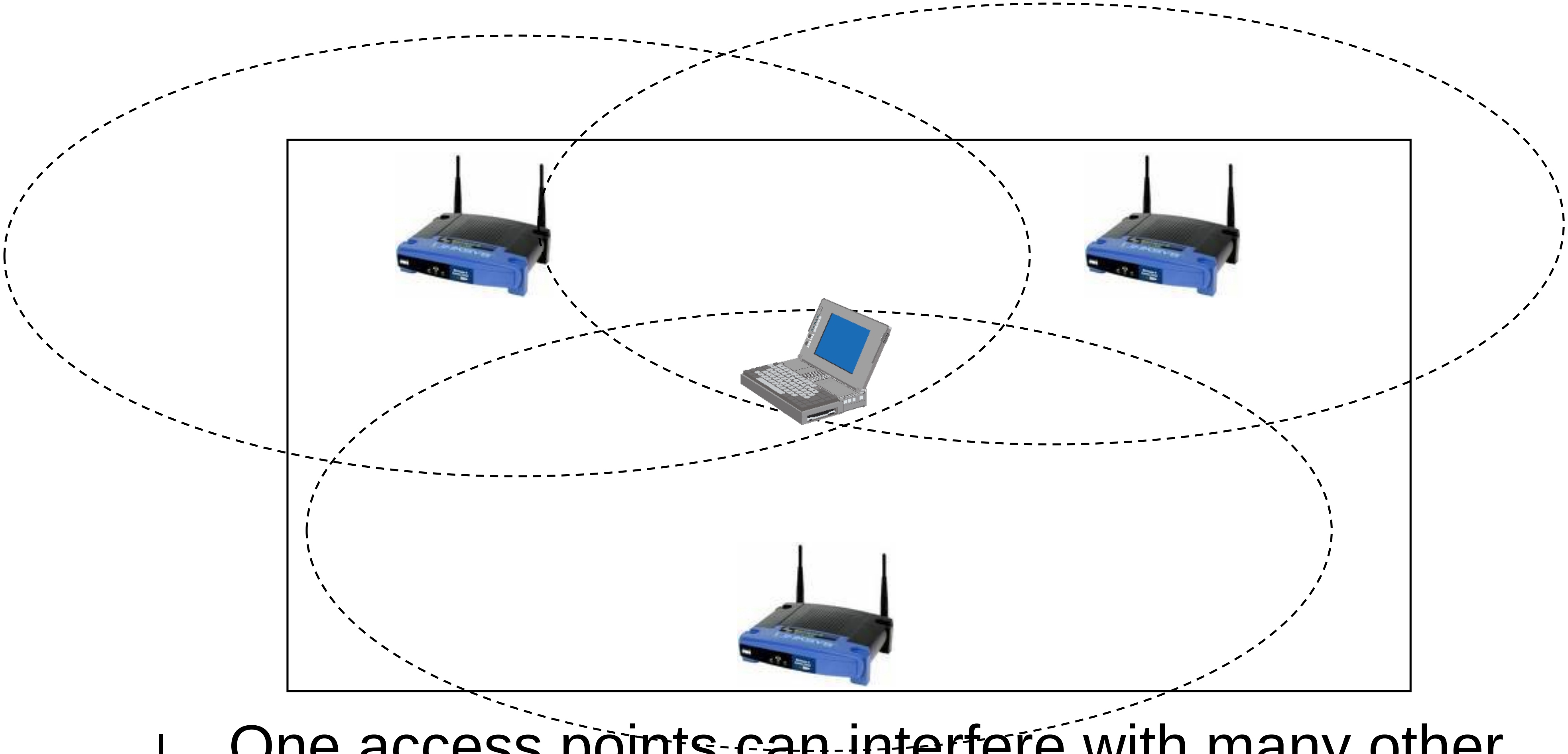
- Home environment



- A WLAN can interfere with the neighbour’s WLAN

Practical issues with WLAN deployments

- Enterprise networks



- One access points can interfere with many other access points
  - reduces significantly overall available bandwidth

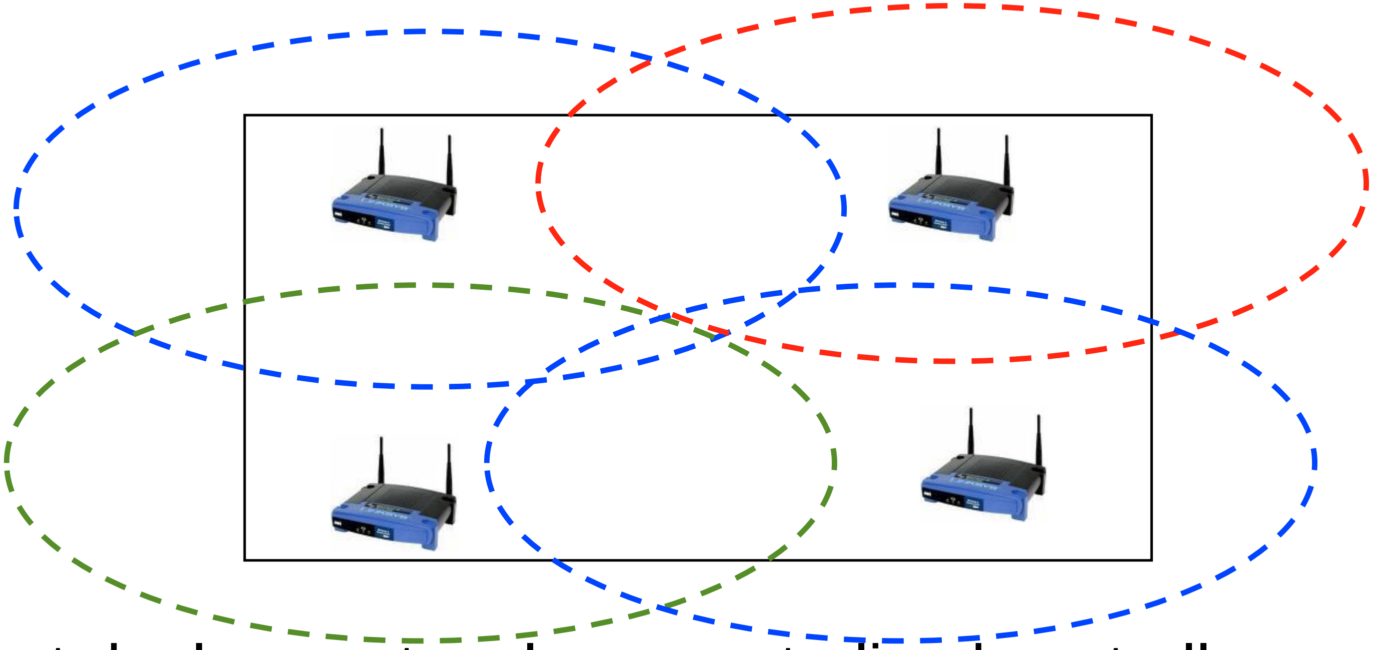
The WiFi channel frequencies

- WiFi standards operate on several frequencies called channels
  - Usually about a dozen channels

- Why multiple channels ?
  - Some channels my be affected by interference and have a lower performance
  - Some frequencies are reserved for specific usage in some countries
  - Allows frequency reuse when there are multiple WiFi networks in the same area
    - Unfortunately, many home access points operate by default on the same factory set channel which causes interference and reduced bandwidth

WLAN in enterprise environments

- What could be done to improve the performance of WLANs ?
  - Reduce interference as much as possible
    - Tune channel frequencies
    - Reduce transmission power
    - Similar to techniques used in GSM networks



- Recent deployments rely on centralised controllers and thin access points

Datalink layer

- Point-to datalink layer

- Local area networks

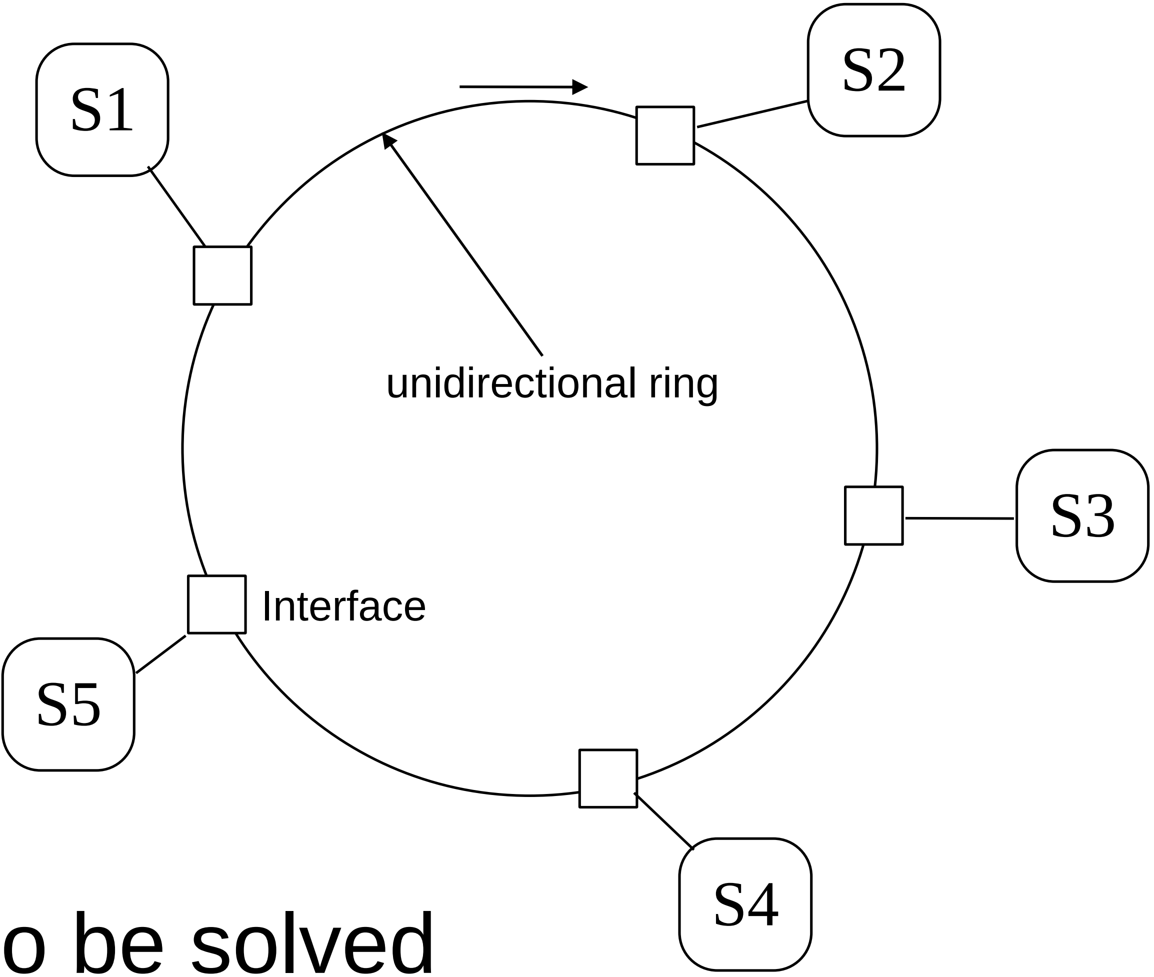
  - Optimistic Medium access control
    - ALOHA, CSMA, CSMA/CD, CSMA/CA

  - Ethernet networks
    - Basics of Ethernet
    - IP over Ethernet
    - Interconnection of Ethernet networks

  - WiFi networks

  - Deterministic Medium access control
    - Token Ring, FDDI

Ring networks



- Problem to be solved
  - How to share fairly ring transmission capacity among all devices attached to the ring ?

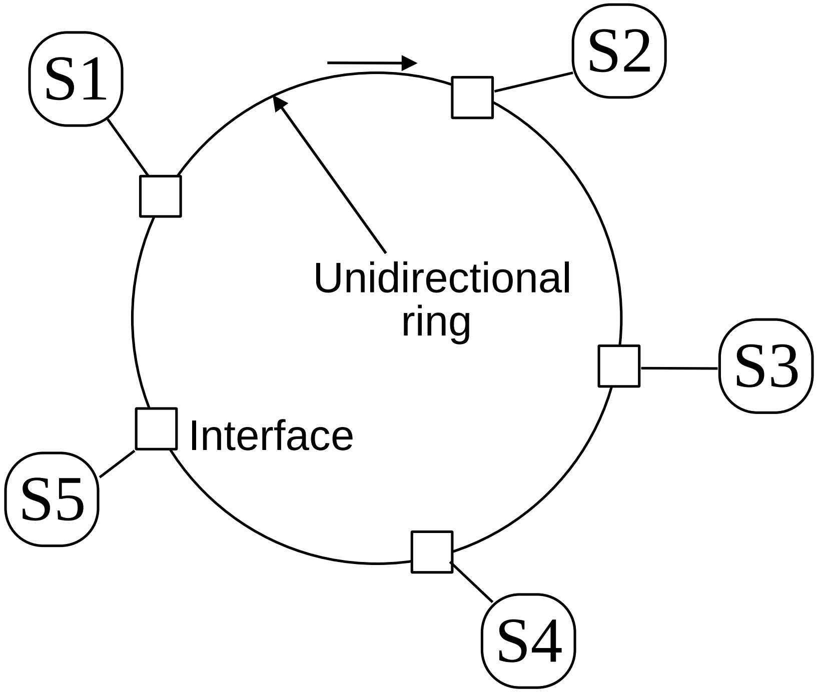
Ring networks (2)

- How to share transmission capacity ?
  - To avoid collisions, only one station should be able to transmit a frame at any time
  - The station that has the right to transmit must own a special frame called token
- How can stations exchange token ?
  - Token is a special frame that can be sent over the ring network
  - A station that needs to transmit a data frame can
    - capture the token and remove it from the ring
    - send one or more data frames
    - send the token back on the ring to allow other stations to capture it and transmit

Ring networks (3)

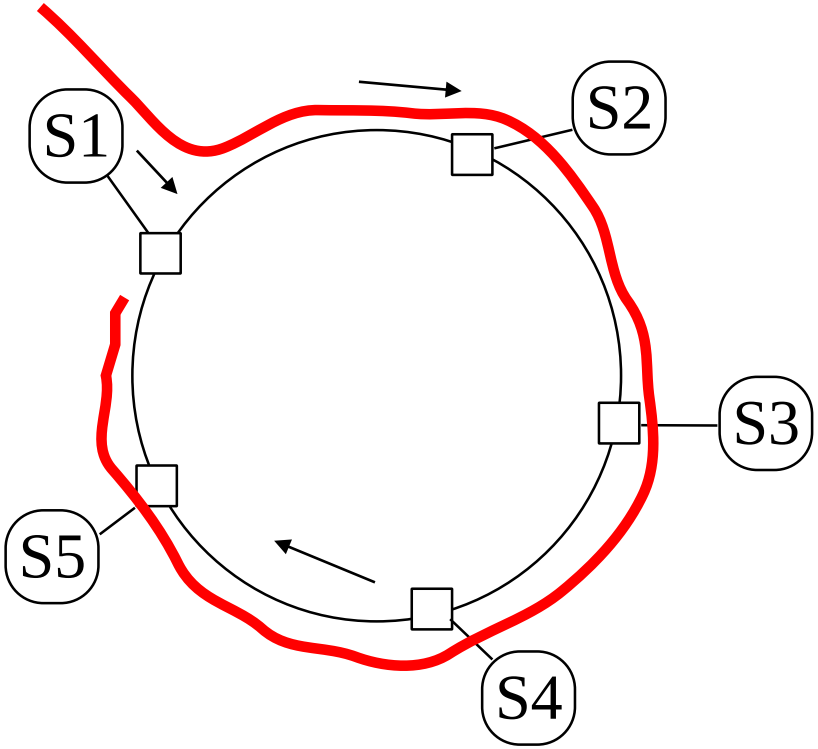
- Consequence
  - When there are no data frames sent, stations should continuously exchange the token

- How to achieve this ?
  - A station must relay the electrical signal it receives upstream when not transmitting
    - it introducing a delay of one bit transmission time
  - If all stations behave so, and token is small, token will travel permanently
    - If token is not small, increase the delay on the token ring network



Ring networks (4)

- Data frame transmission
  - A data frame requires a longer transmission time than the ring delay



- Sender behaviour
  - Captures token
  - Sends data frame
  - Removes data frame from ring
  - Sends token

Ring networks in practice

- Two types of ring LANs
  - Token Ring
    - Invented by IBM
    - Standardised by IEEE/ISO (802.5)
    - Ring build with point-to-point twisted pair links
      - 4 Mbps
      - 16 Mbps
      - Some work for 100 Mbps Token Ring
  - Fiber Distributed Data Interface (FDDI)
    - First data networks built with optical fiber
    - standardised by ANSI
    - 100 Mbps
    - up to 200 km and 1000 stations

- Other ring technologies exist and are used
  - SONET/SDH
  - DPT

Token Ring (1)

- Token
  - travels permanently on ring when stations are idle
  - Size 24 bits
  - Minimum delay on ring
    - 24 bits transmission times
  - Actual ring delay
    - Each station introduces a one-bit transmission time delay
    - Physical links have a propagation delay
    - Each ring contains a monitor station that measures delay during ring initialisation and adds delay if needed
- Interfaces
  - Two modes of operation
    - Listen : interface adds a one bit transmission delay
    - Transmit : only if station owns the token

Token Ring (2)

- Frame format
  - Token (24 bits)
    - SD : starting delimiter
      - invalid physical layer symbol with Manchester coding
    - AC : Access control
    - ED : ending de fin
      - invalid physical layer symbol with Manchester coding

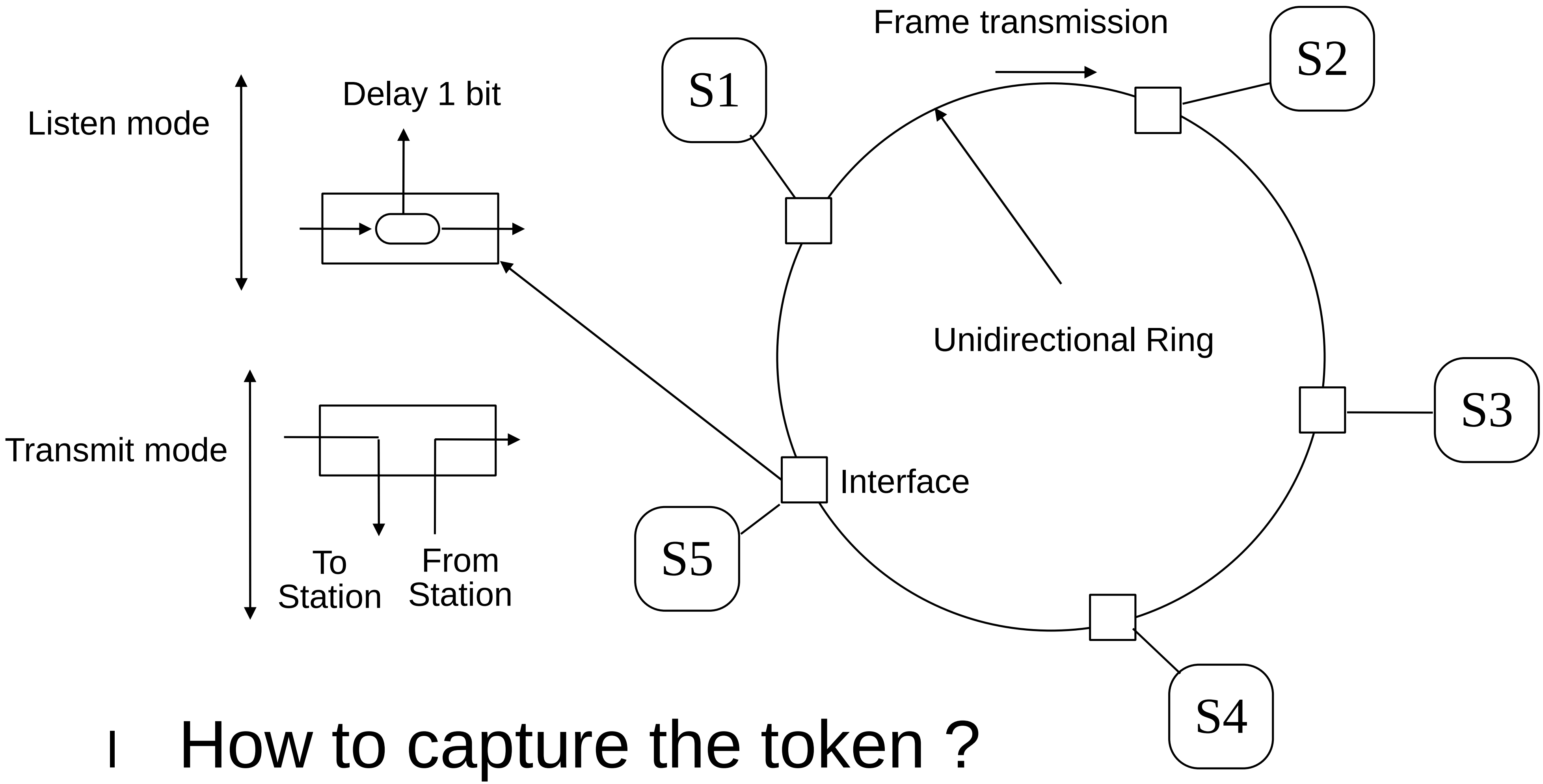
1	1	1
SD	AC	ED

  - Data frame

1	1	1	(2or)6	(2 or)6	0... (nolimit)	4	1	1
SD	AC	FC	Dest	Source	Data	CRC	ED	FS

    - FC : Frame control
      - Allows to distinguish between control frames and data frames
    - FS : Frame status

Token Ring (3)



- How to capture the token ?
  - Rely on Token bit of AC field and one bit delay

Token Ring (4)

- What’s special about Token Ring
  - Can efficiently support acknowledgements
  - Frame Status contains two bits : A and C
    - A and C are set to 0 when transmitting a frame
    - When a receiver sees one frame destined to itself, it sets A to 1
    - When a receiver copies one frame destined to itself inside its buffers, it sets C to 1
  - Data frame (and FS) return to sender. By checking A and C, it knows that :
    - if A=0 and C=0, destination is down
    - if A=1 and C=0, destination is up, but congested
    - if A=1 and C=1, frame was received by destination

Token Ring (5)

- Issues with Token Ring
  - How to ensure fairness ?
    - A station should not be allowed to transmit indefinitely
    - Token Holding Time
      - Maximum time during which a station can own the token and transmit data frames without releasing the token
      - Default : 10 milliseconds
  - How to bootstrap the Token Ring ?
    - Which station sends the first token ?
    - How to ensure that the Ring delay is long enough ?
    - What happens when a station fails ?
      - If it did not own the token, no issue
      - If it owned the token while failing, then
        - Which station will remove the current data frame from the ring ?
        - Which station will send the token on the ring ?

Token Ring (6)

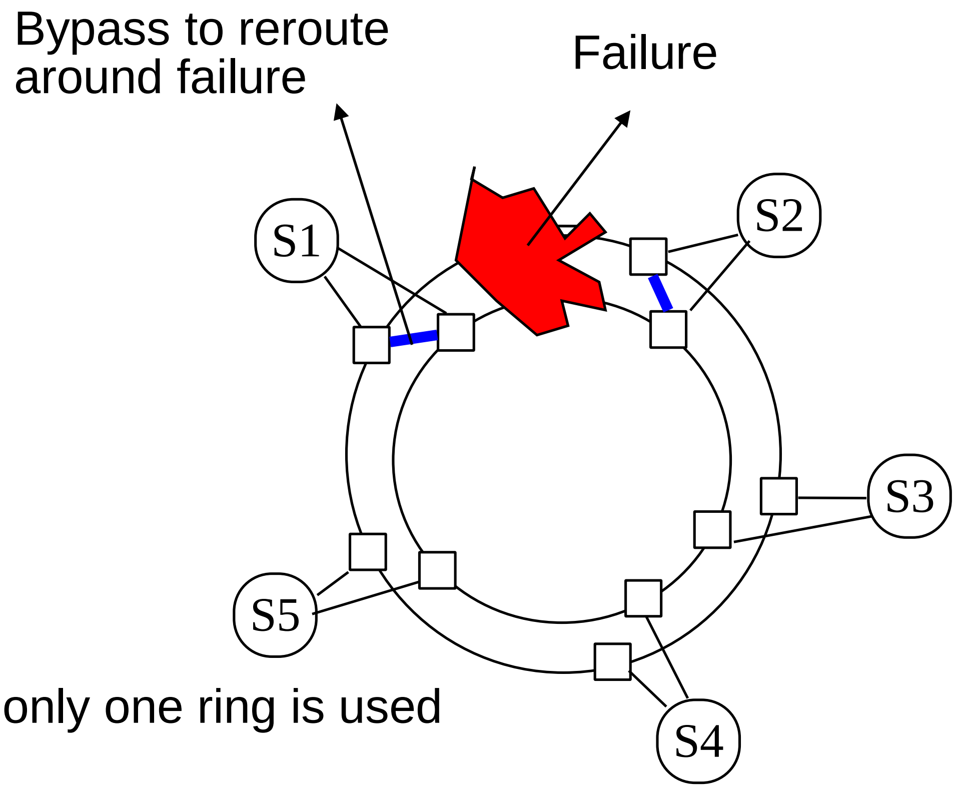
- How to bootstrap a Token Ring ?
  - Complex problem
  - Main idea
    - One station should send the token
    - The first station on the ring hears nothing and notices that there is a problem. It sends a special frame called CLAIM_TOKEN
    - If it receives the frame back, it becomes the monitor
      - Each station must be able to become monitor
- Monitor's responsibilities
  - Ensure that token is never lost or corrupted
  - Insert an artificial delay of 24 bit transmission times on the ring
  - Remove orphan and looping frames
  - If the monitor fails, the ring must be bootstrapped again

Token Ring (7)

- Token surveillance
  - Monitor checks how often its sees the token
    - If there are N stations on the ring, then the monitor should see the token at worst every N*THT seconds
    - If token is lost, monitor cuts ring, removes electrical signal and resend a new token
- Orphan frames
  - Frame with invalid coding or incomplete frame
    - monitor cuts ring, removes electrical signal and resend a new token
- Looping frames
  - Every time monitor sees a frame, it sets its *Monitor* bit of the AC field to 1
    - All stations send their frames with *Monitor=0*
    - If a frame is seen twice by the monitor, it cuts ring, removes electrical signal and resend a new token

FDDI

- Network topology FDDI
  - Single ring like Token Ring
  - Two counter rotating rings to deal with failures



Usually, only one ring is used

FDDI (2)

- Medium access control
  - Token based access control
    - A station can only transmit a data frame provided that it owns the token
  - Token Holding Time (THT)
    - maximum duration of transmission
    - Token Rotation Time (TRT)
      - maximal delay for a token to rotate around the entire ring
      - TRT δ Actives_Stations * THT + Ring_Latency
  - When should the Token be released
    - Immediately after removal of the data frame sent
      - as in Token Ring
    - Immediately after transmission of the data transfer, without waiting for it to come back
      - solution chosen for FDDI due to the high bandwidth and long latency of the FDDI ring

FDDI (3)

- Delay sensitive service
  - How to support two types of frames in FDDI ?
    - normal data frames (*asynchronous frames*)
      - example : file transfer, email, www
    - delay sensitive data frames (*synchronous frames*)
      - example : telephone, videoconference

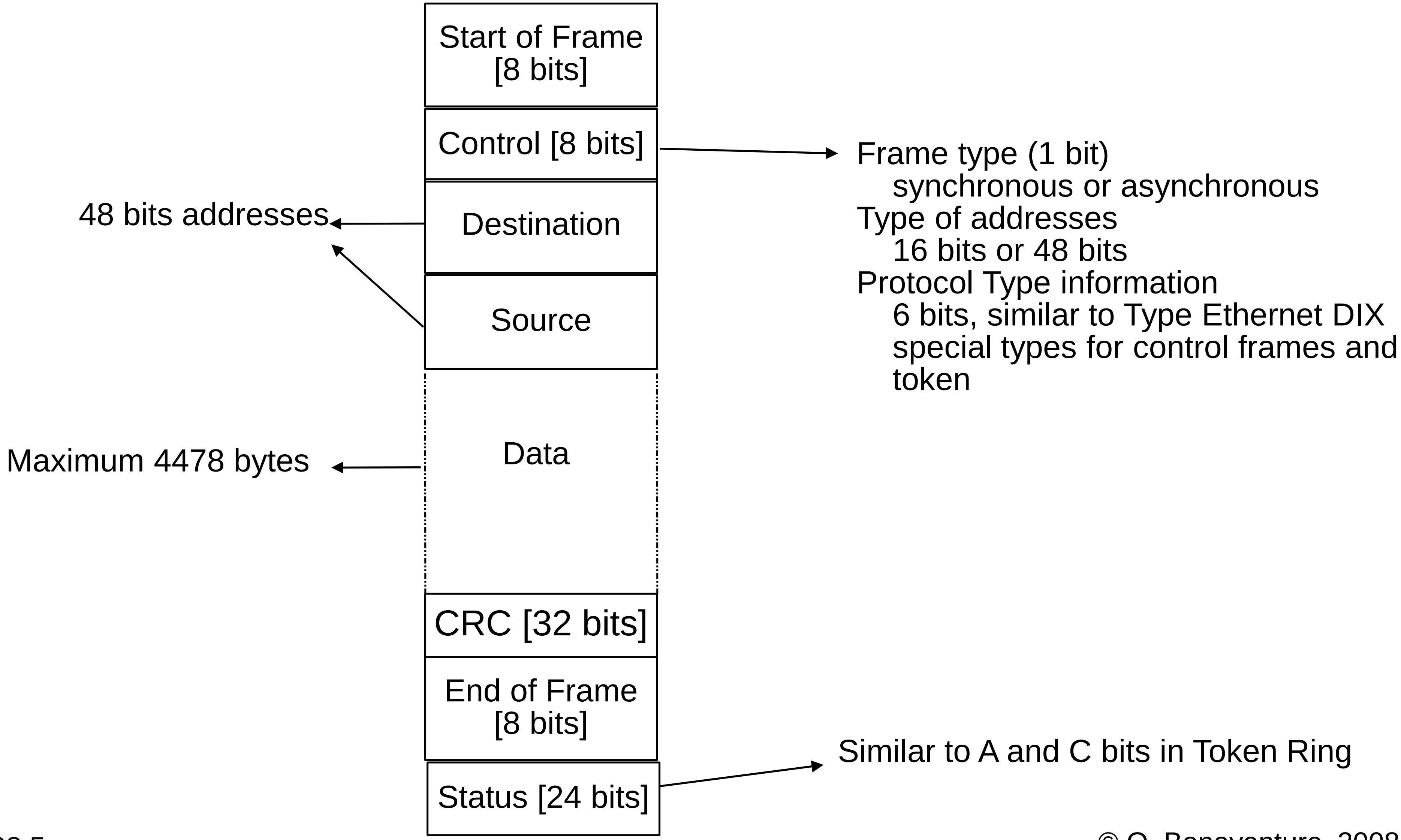
- Solution
  - Delay sensitive frames can be supported provided that a FDDI ring can bound the transmission delay of such a frame
    - synchronous frames should be transmitted earlier than normal frames on each station
    - Since a station can always transmit when it captures the token, a solution should bound the Token Rotation Time to provide strict guarantees to delay sensitive frames

FDDI (4)

- How to bound the TRT ?
  - Target Token Rotation Time (TTRT)
    - At ring initialisation, all stations propose their expected TTRT and the smallest proposed value is chosen
    - All stations must control their transmissions such that the token rotation time is always smaller than TTRT
    - each station measures the current TRT
      - When a station captures the token, it can send its synchronous frames
        - there is a maximum amount of synchronous frames that can be sent by each station. This maximum is negotiated by using control frames.
      - If after having sent synchronous frames TRT < TTRT, this means that the token is circulating quickly and the station can send asynchronous frames
      - Otherwise the token must be released

FDDI (5)

- Frame format



Start of Frame [8 bits]

Control [8 bits]

Destination

Source

Data

CRC [32 bits]

End of Frame [8 bits]

Status [24 bits]

48 bits addresses

Maximum 4478 bytes

Frame type (1 bit)
- synchronous or asynchronous

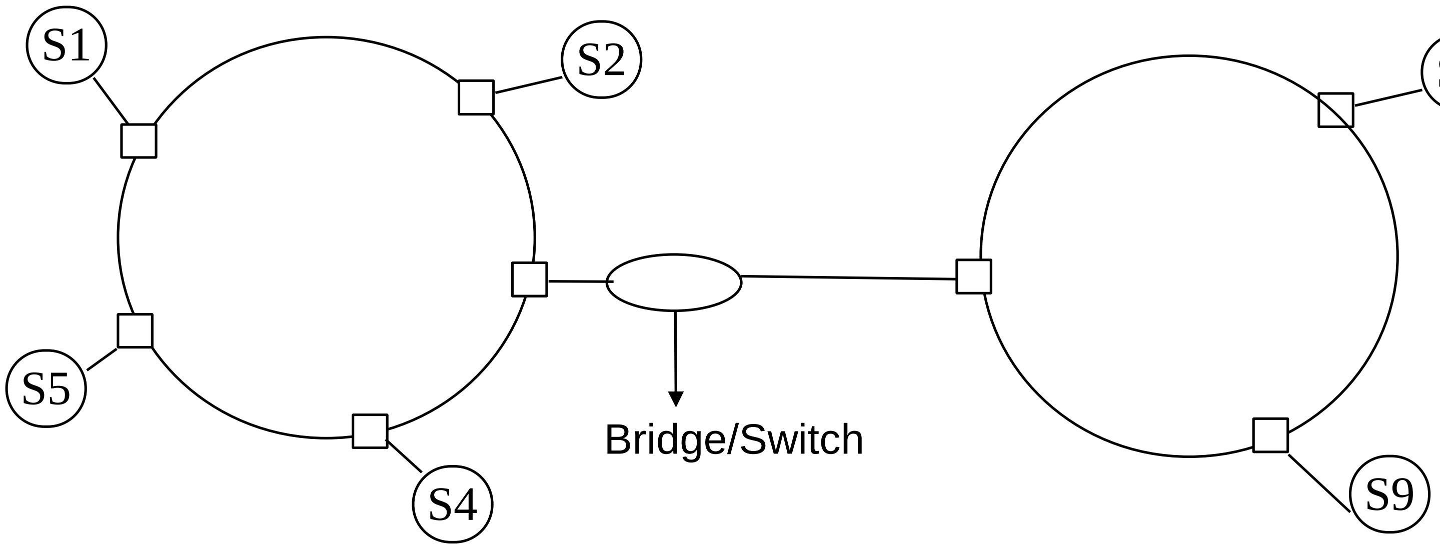
Type of addresses
- 16 bits or 48 bits

Protocol Type information
- 6 bits, similar to Type Ethernet DIX
- special types for control frames and token

Similar to A and C bits in Token Ring

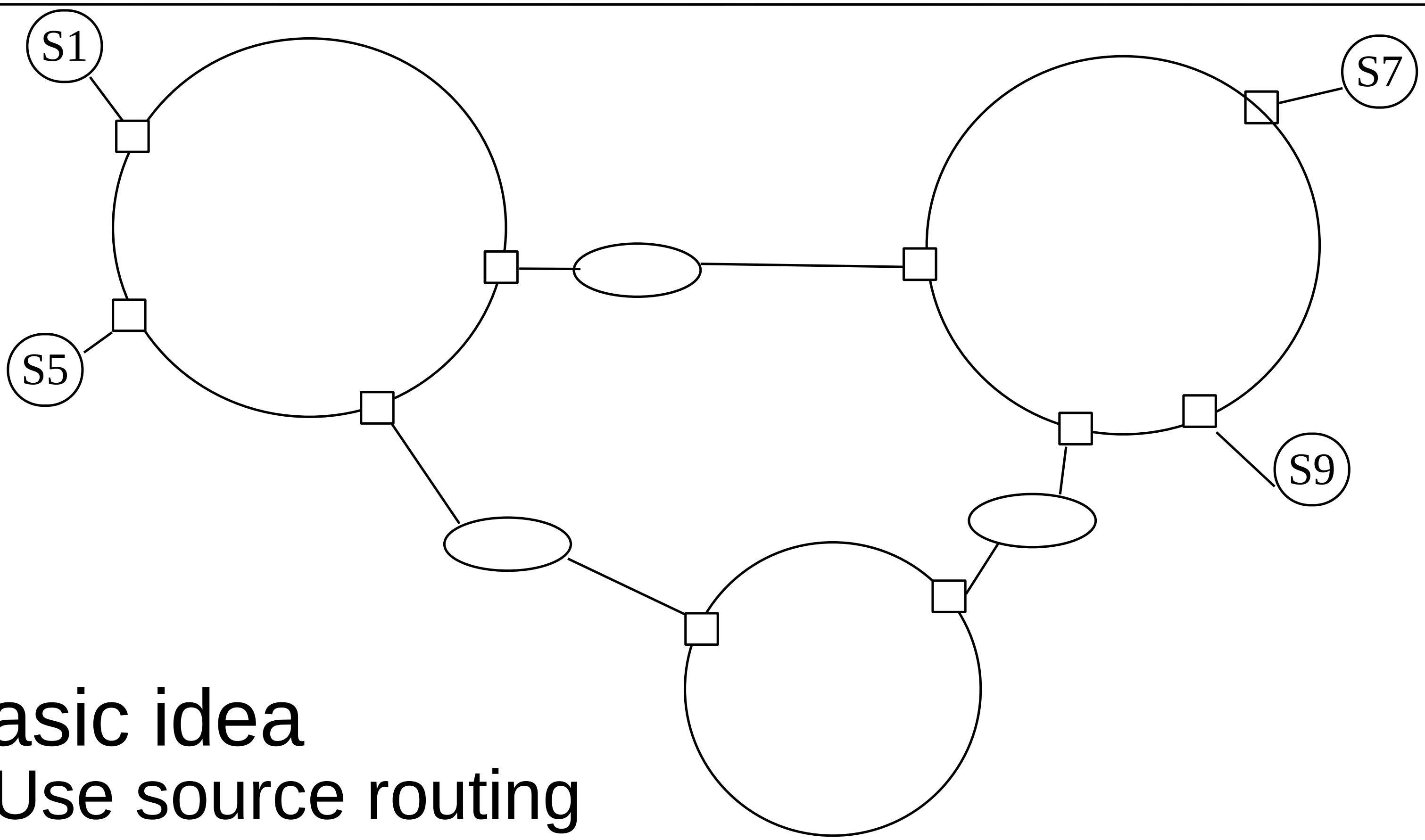
Interconnection of Token Rings

- How to interconnect Token Ring networks ?



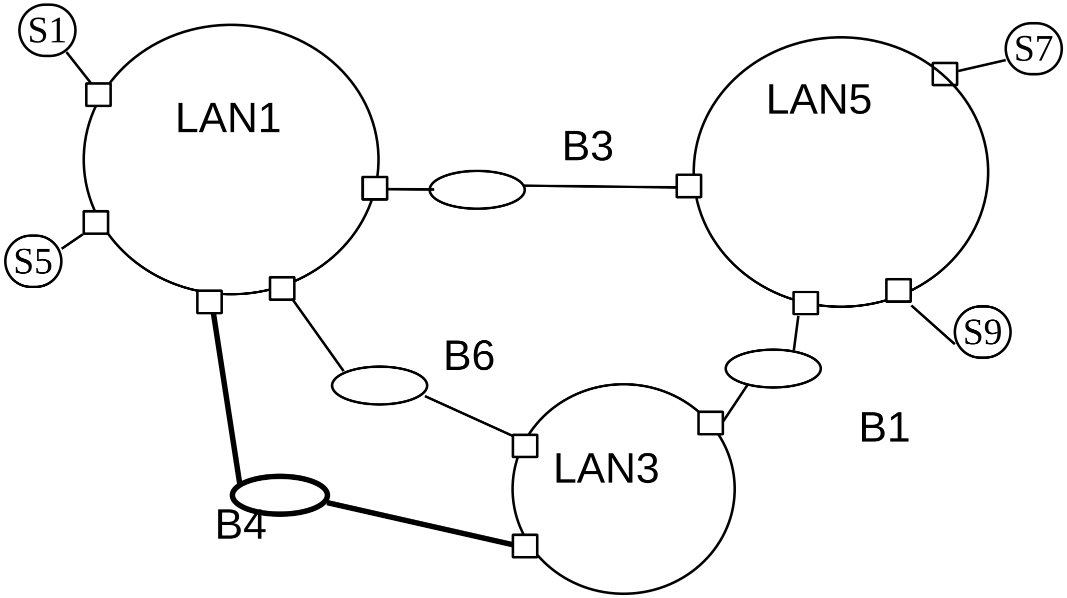
- Possible solutions
  - Use the spanning tree designed for Ethernet
  - Invent a new protocol
    - solution chosen by IBM for Token Ring

Interconnection of Token Rings (2)



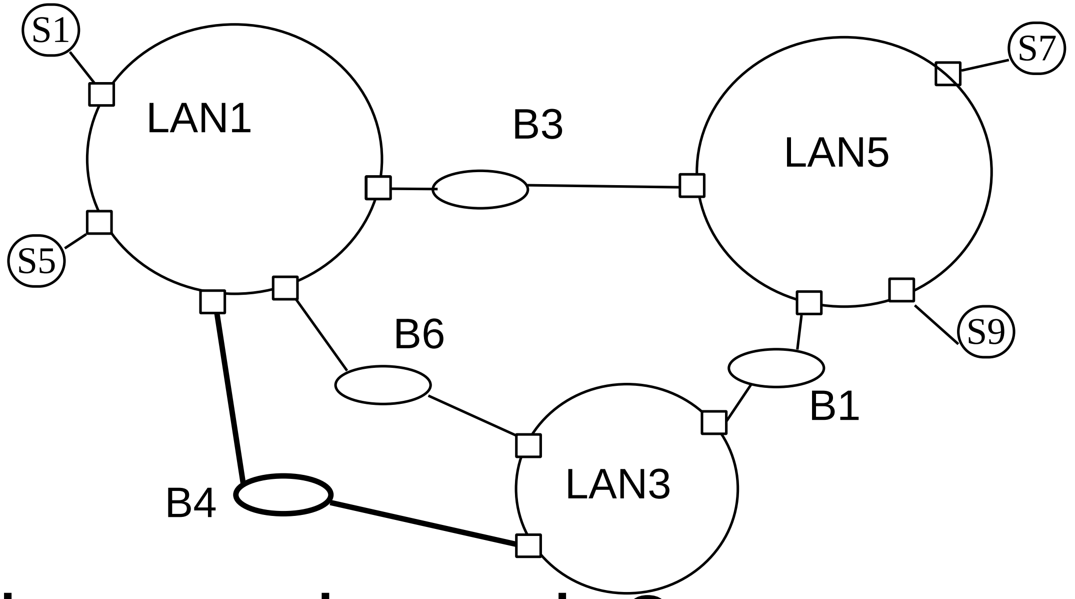
- Basic idea
  - Use source routing
- Problems
  - How to identify the paths
  - How to discover the paths ?

Interconnection of Token Rings (3)



- Identification of paths
  - Each LAN has one unique identifier
  - Each bridge has one identifier
  - Each path is a list of pairs LAN#,bridge#

Interconnection of Token Rings (4)



- How to discover the path ?
  - Control frame : all paths explorer
    - Sent by source towards destination
    - Forwarded by all bridges that add their identifier and LAN identifier
    - Destination sends back the ape frame to source by using reverse path
    - Each station caches the recent paths

Spanning Tree versus Source Routing

- Spanning tree
  - complexity in switches/bridges
  - only a subset of the network is used
  - entirely transparent
  - multicast natively supported
  - few control frames (802.1d)

- Source routing
  - complexity in all stations
  - the entire network is used
  - requires support on stations
  - spanning tree required for multicast
  - many control frames can be required