

Cours Réseaux Locaux

UE Réseaux Locaux et de Télécommunications
2SN Parcours A - R - T Année 2024-25

Katia Jaffrès-Runser
kjr@n7.fr



Objectifs du cours

SAVOIR :

- Décrire une architecture protocolaire de réseau local,
- Décrire les principaux mécanismes liaison de données et MAC du standard IEEE, et leurs interactions avec les protocoles IP et de transport,
- Décrire l'effet des mécanismes intervenant dans la commutation de trames Ethernet (apprentissage, VLAN, protocoles d'arbres couvrants, qualité de service, EEE),
- Configurer un réseau local Ethernet,
- Choisir une architecture et une topologie de réseau local pour un cas d'usage donné en l'argumentant,
- Définir un protocole de réseau local simple.



Plan

1. Contexte et rappels (Séance 1)
2. Architecture, Topologie, Standards (Séance 2)
3. Ethernet commuté
 1. Le lien Ethernet (Séance 3)
 2. La commutation de trames (Séance 3)
 3. Les Protocoles d'arbres couvrants (Séance 4 et 5)
 4. Les VLAN (Séance 6)
4. Mécanismes de qualité de service et d'économie d'énergie (EEE) (Séance 7)
5. Définir un protocole de réseau local simple. (Séance 8)

Au total -> 8CM, 4 TD et 2TP.

Un examen 1h30 - 1 feuille A4 recto verso manuscrite autorisée.



[CM1]

Contexte et rappels



Wooclap pour commencer

<https://app.wooclap.com/RLCM1>



Vocabulaire

- Réseau informatique
- Médium de communication
- Interface de communication
- Protocole
- Adressage
- Architecture / Topologie
- Empilement protocolaire
- Méthodes d'accès
- Standard / technologie
- Réseau opéré, réseau d'accès, réseau de cœur

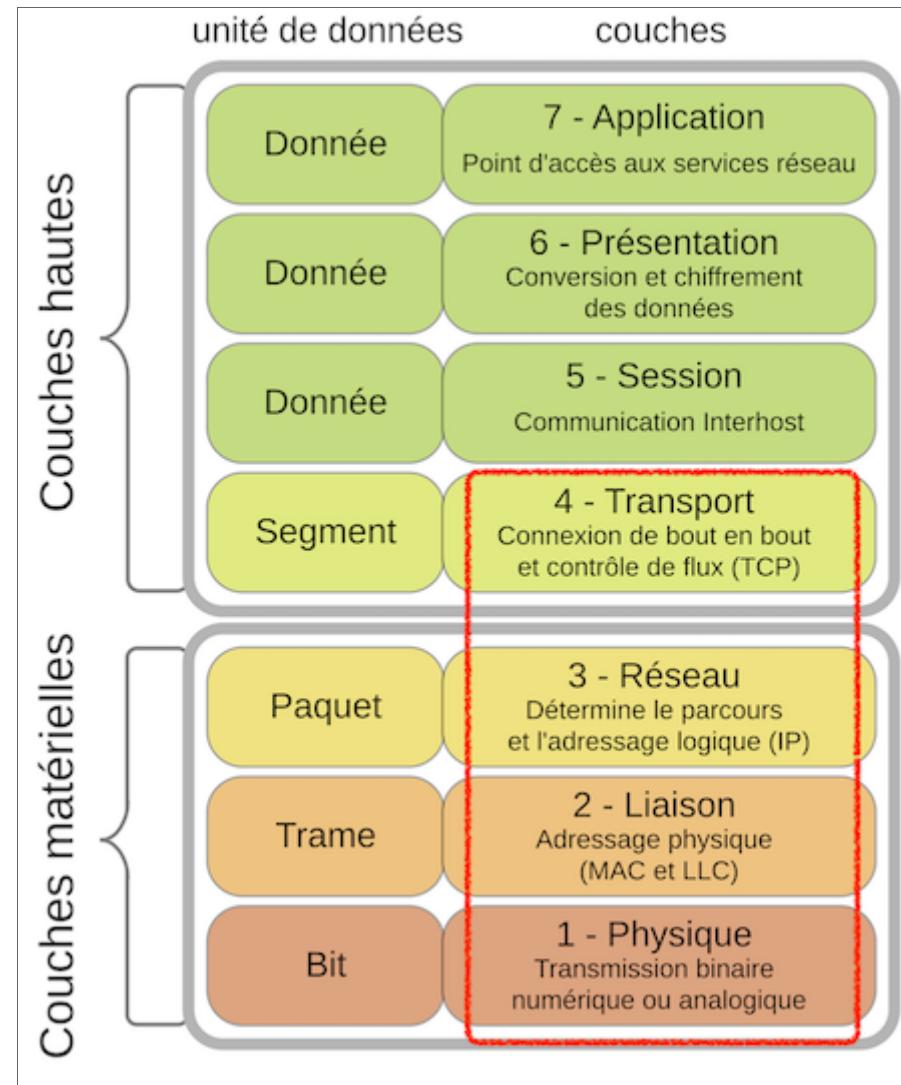


Classification des réseaux

- Couverture du service :
 - PAN / MAN / LAN / WAN
- Technologique :
 - Nature du médium : filaire / sans-fil
 - Caractéristiques du médium : débit d'émission, portée, taux d'erreur bit...
 - Nb d'utilisateurs max, débit utilisateur, latence, consommation énergie, déterminisme, taux de perte de messages...
- Type d'entités communicantes : des stations (PC), des choses (IoT), des robots (drones), des calculateurs embarqués ..



Modèle en couches



Couche 1 : physique

Rôle

Transmission bit-à-bit sur le médium de communication

Fonctions

- Définit les techniques de communication numériques utilisées (modulation, codage,...). Elles dépendent de la technologie
- Détermine le débit débit d'émission des bits.
- Est utilisé par la couche liaison pour accéder au support



Couche 2 : liaison

Rôle

Transformer la transmission physique en **une communication fiable** permettant l'échange d'**une trame** (une séquence de bits avec un sémantique connue) entre deux entités.



Couche 2 : liaison

Fonctions

- Définit l'**adresse physique** des entités communicantes
- Compose les trames pour l'émission et est capable de détecter leur début et leur fin en réception.
- Peut ajouter des mécanismes de détection des erreurs de transmission : **le contrôle d'erreur**.
- Si plusieurs utilisateurs doivent se partager le médium, la couche liaison détermine quel utilisateur peut émettre ses données à tout instant : **le contrôle d'accès** => on parle ici de **protocole MAC** : medium access control protocol.



Couche 2 : liaison

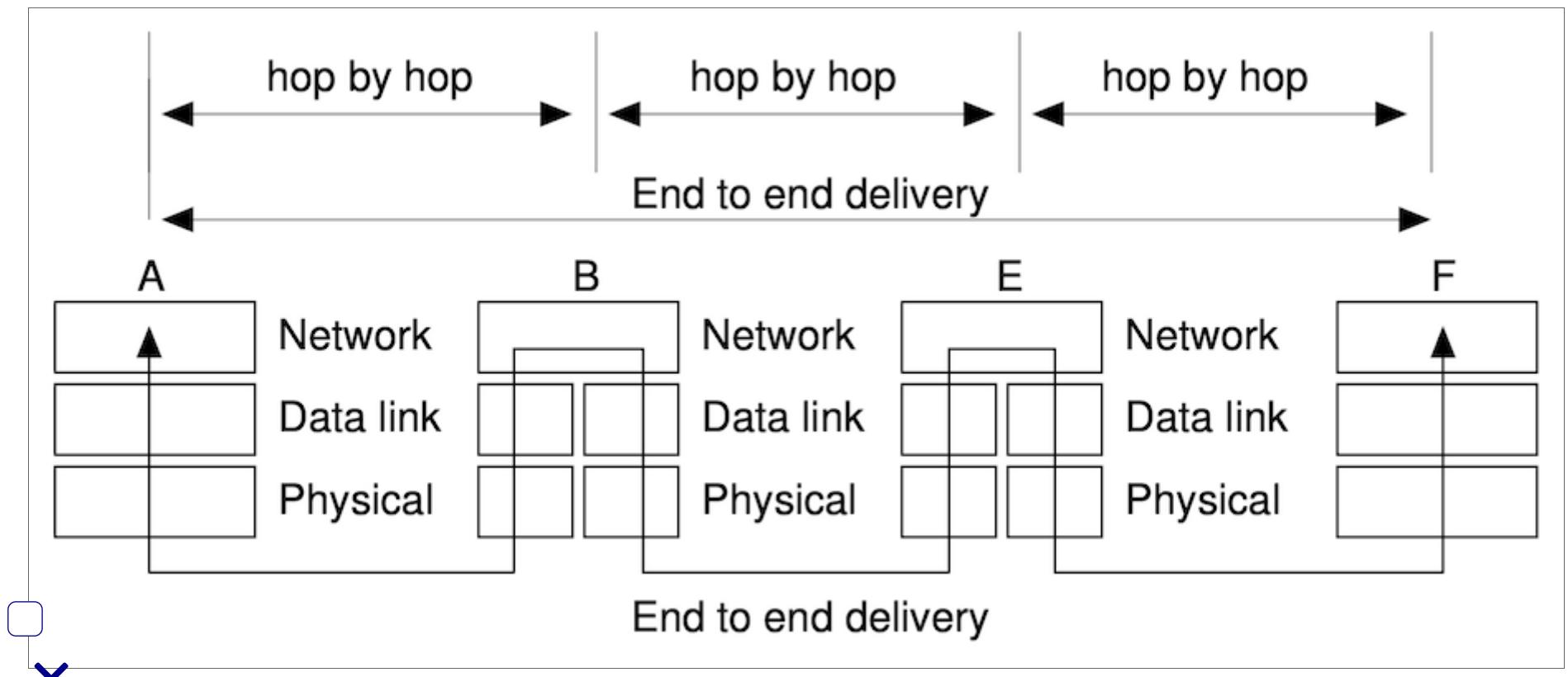
- **Note 1 :** Une interface de communication (eth0, if0, wlan0...) regroupe la couche physique et la couche liaison.
- **Note 2 :** Tous les terminaux d'un même réseau local utilisent le même protocole de couche physique et de liaison pour échanger des trames.



Couche 3 : routage

Rôle

Responsable de la transmission des paquets de la source à la destination à travers plusieurs réseaux. Chaque réseau peut utiliser une technologie (couche 1 et 2) différente.



Couche 3 : routage

Fonctions

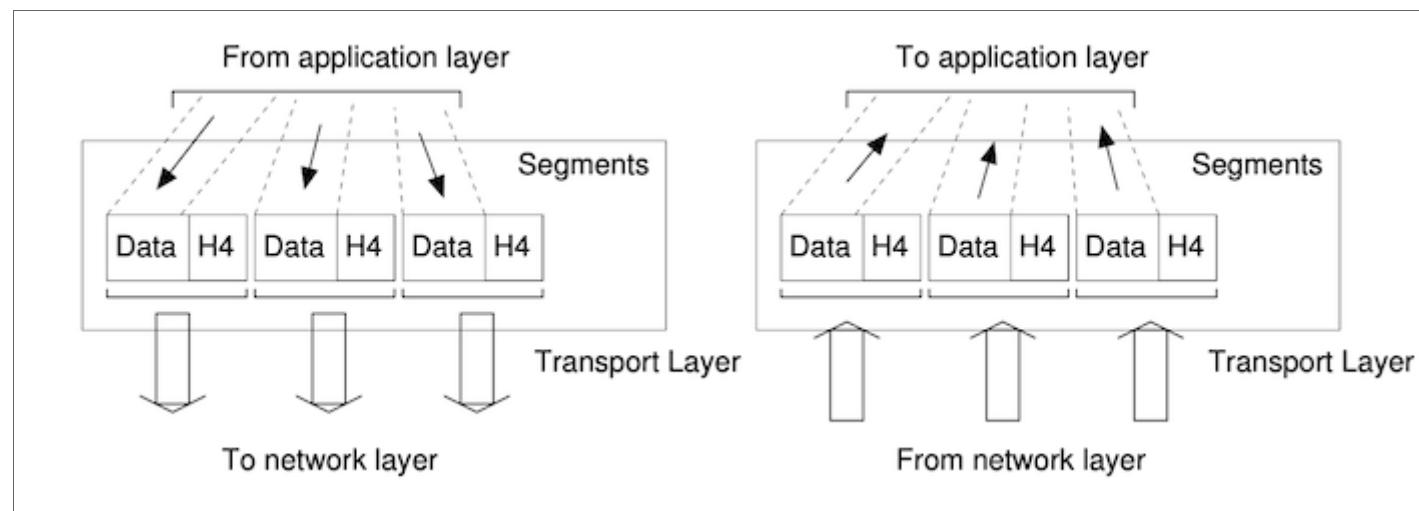
- Définit des adresses logiques. Un mécanisme doit permettre de faire correspondre les adresses logiques et physiques (ARP dans le monde Internet).
- Définit un mécanisme de routage des paquets en ajoutant un en-tête spécifique (IP dans le monde Internet).
- Des protocoles de routage permettent de configurer les équipements pour qu'ils soient en mesure de routes les paquets (RIP, OSPF...).



Couche 4 : transport

Rôle

Responsable de l'acheminement de bout-en-bout d'un message complet qui lui est confié par une application.



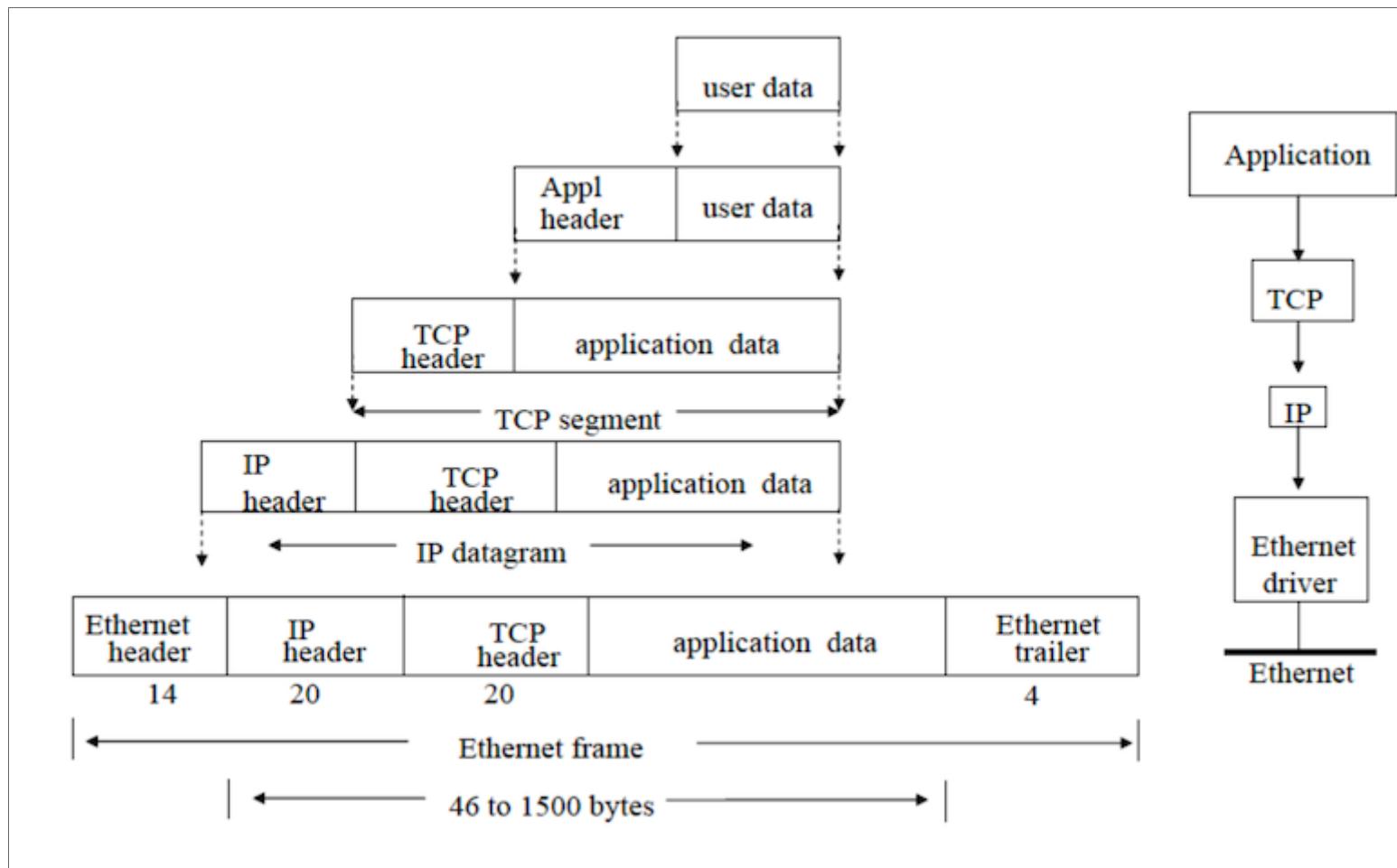
Couche 4 : transport

Fonctions

- Le message d'origine (fichier, vidéo, ..) est découpé en **segments** de taille compatible avec les couches 2 et 3. Chaque segment est numéroté.
- Chaque segment est identifié dans l'en-tête par le numéro de l'application qui l'a envoyé : **le port**
- Le récepteur acquitte les segments reçus et les ré-ordonne pour les rassembler
- Si des segments sont perdus, il sont ré-émis par la source
- Il existe des mécanismes de contrôle de flux pour adapter la vitesse d'émission à la congestion dans le réseau.



Encapsulation protocolaire

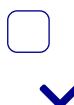
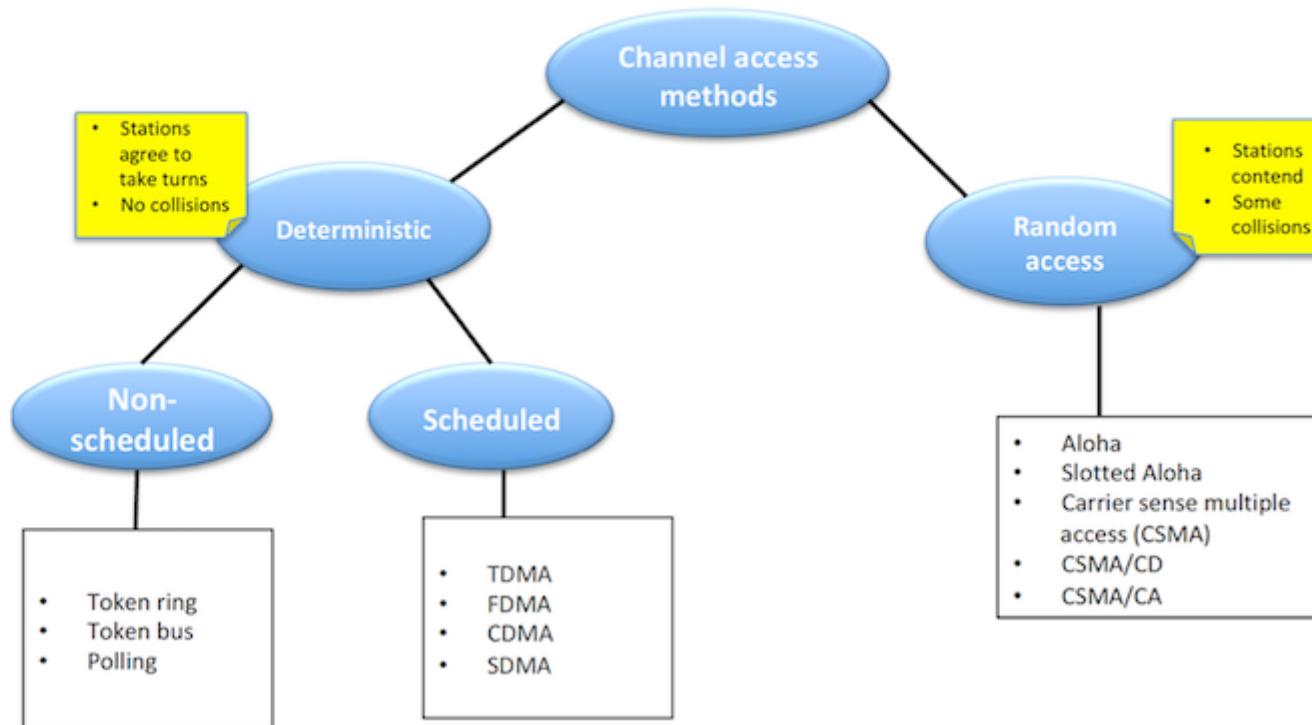


Méthodes d'accès

(rappel 1SN)

Channel access methods

MAC protocols follow different approaches for sharing the channel. Each type is called a *channel access method*.



Méthodes d'accès

| Méthode d'accès | Avantages | Inconvénients |
|-----------------|--|--|
| Aléatoire | Ajout / suppression de station naturel -> protocole d | Non déterministe, quelques collisions |
| Polling | Déterministe : accès garantit, mais il faut attendre son tour. | Ajout/suppression de station plus complexe. Un point de défaillance unique (le maître) |
| Token Ring | Déterministe : accès garantit, mais il faut attendre son tour. | Ajout/suppression de station complexe. Gestion de la perte du jeton. |



Méthodes d'accès

| Méthode d'accès | Avantages | Inconvénients |
|-----------------|---|--|
| FDMA | Déterministe : accès garantit. Pas de collision par définition. | Nombre de fréquences limitées, on perd des ressources si l'utilisateur n'émet pas constamment. |
| TDMA / FTDMA | Déterministe : accès garantit. Permet un bon taux d'utilisation du support. | Moins réactif que l'accès aléatoire. Il faut un mécanisme d'allocation de ressources des stations aux slots. Nécessite une synchronisation des noeuds. |



Dans ce cours de réseaux locaux

On traitera :

- principalement de couche 2,
- un peu moins de couche 1
- et un petit peu de couche 3 (surtout pour faire le lien avec certains mécanismes de couche 2).

On s'intéressera principalement aux **réseaux type Ethernet commuté**, et un petit peu aux réseaux sans-fil.



[CM2]

Architecture, Topologie, Standards



Définitions

L'architecture de réseau est l'organisation d'équipements de transmission, de logiciels, de protocoles de communication et d'infrastructure filaire ou radioélectrique permettant la transmission des données entre les différents composants.

(src : Wikipedia)

- Pour un **réseau local**, les équipements sont :
 - des stations fixes ou mobiles (si sans-fil)
 - des commutateurs
 - des points d'accès
 - une passerelle (gateway ~ routeur) pour sortir du réseau
- Les protocoles de communications sont choisis en fonction des technologies choisies (WiFi, Ethernet, Bluetooth...)



La topologie d'un réseau est une représentation logique des équipements et de leur interconnection qui se représente sur un schéma.

- Dans un **réseau local**, on trouve différentes topologies :
 - En bus,
 - En anneau,
 - En arbre (ou hiérarchique)
 - En étoile
 - Maillé



Les standards de réseaux locaux

Les principaux standards utilisés par les technologies grand public sont édités par IEEE

(Institute of Electrical and Electronics Engineers)

- C'est le comité 802 de l'IEEE qui normalise les couches physique et liaison des principales technologies de réseaux locaux :
 - Ethernet - standard IEEE802.3
 - WiFi - IEEE802.11
 - Token Bus - IEEE802.4 et Token Ring - IEEE802.5
 - Bluetooth, Zigbee, UWB (technos IoT) - IEEE802.15
 - ...



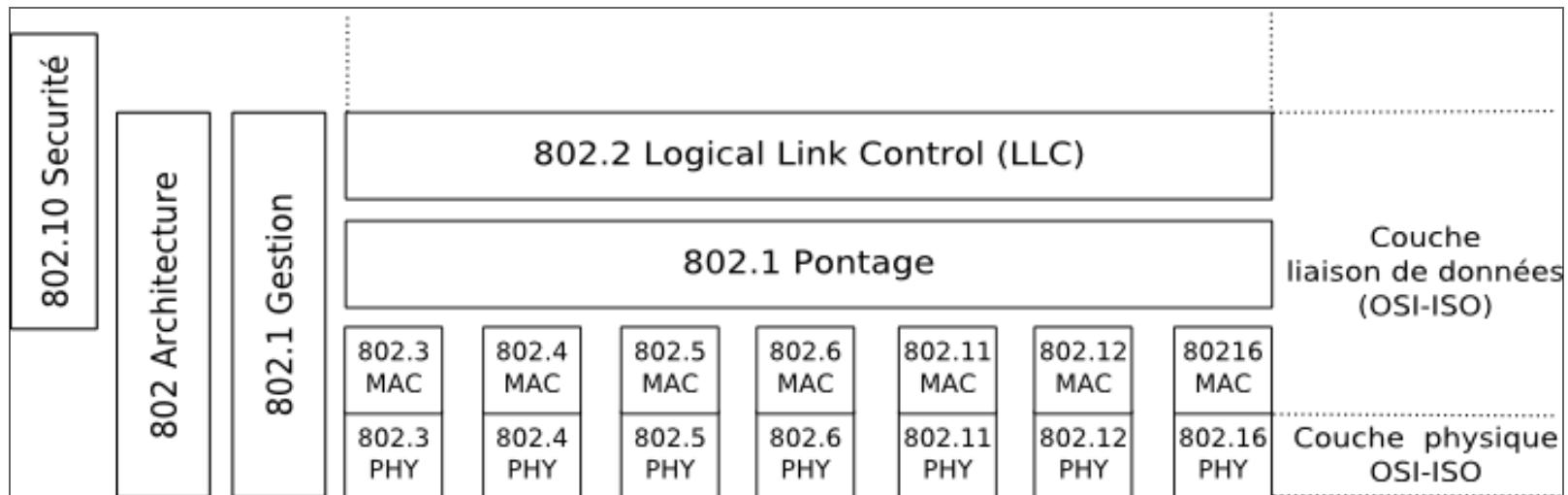
IEEE802.1 et IEEE802.2

Ces standards sont communs à toutes les technologies de réseaux locaux 802.x

- **IEEE802.1** chapeaute tous ces standards en décrivant les principales règles permettant
 - d'interconnecter les réseaux locaux par des **PONTS** (Bridge = commutation et routage de niveau 2),
 - de les gérer (créer des VLAN par exemple),
 - de proposer un service de synchronisation,
 - d'intégrer des politiques de qualité de service,
 - etc..
- **IEEE802.2** définit la **partie haute** de la couche liaison qui se nomme ici **LLC - Logical Link Control**.



Organisation des différents standards



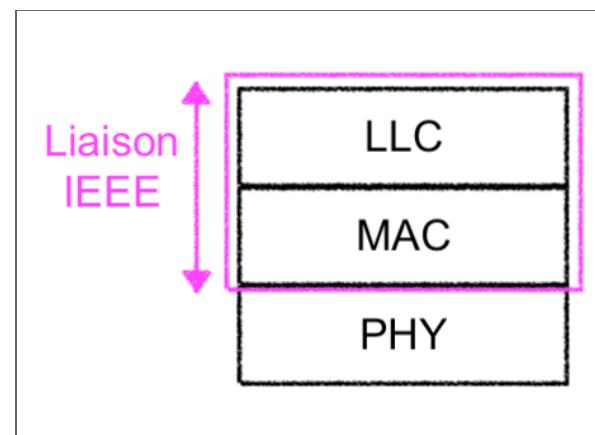
(src : Wikipedia)



Couche Liaison = MAC + LLC

La couche Liaison IEEE802 se compose de 2 parties :

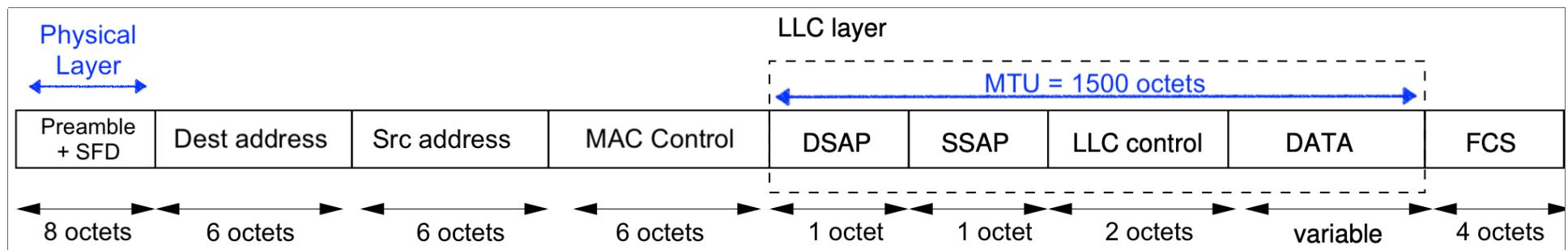
- **LLC** : Elle permet de fiabiliser la transmission par un contrôle d'erreur et un contrôle flux. **Elle est ++ optionnelle ++ !**
- **MAC** : Elle s'occupe des autres fonctions de la couche Liaison et dépend de la technologie pour ce qui est du contrôle d'accès. Ses fonctions sont :
 - Insérer les adresses MAC source et destination,
 - Filtrer les trames reçues (matching @dest et @mon_interface)
 - Assurer le contrôle d'accès



La trame Ethernet IEEE

https://en.wikipedia.org/wiki/Ethernet_frame

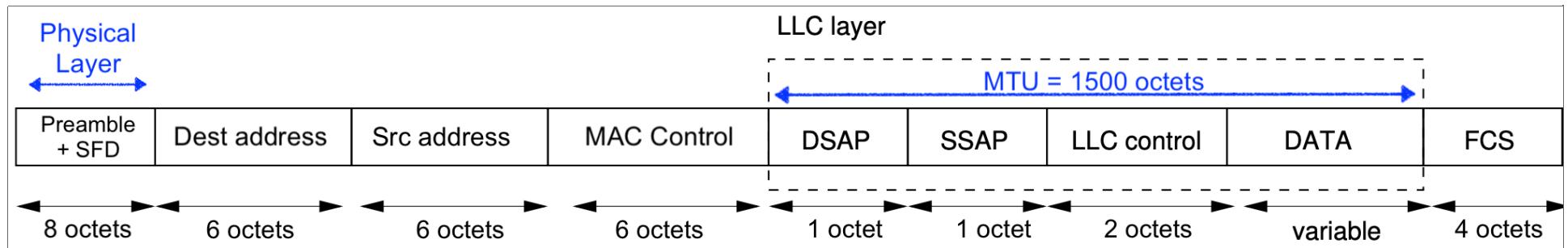
La trame Ethernet peut comporter OU NON l'en-tête LLC :



- Elle est suivie de 12 octets d'*Interpaquet gap* (IPG) de SILENCE :-)
- **Sa taille maximale de données utile (MTU) est 1500 octets**
=> Si LLC est utilisé, la MTU est de 1500-4 octets !



La trame Ethernet IEEE

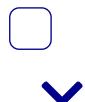


LE PRÉAMBULE (OU DRAPEAU / FLAG)

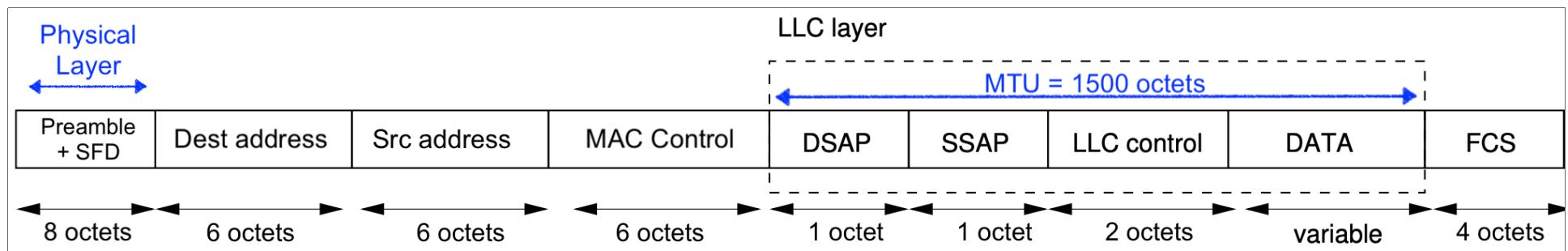
Permet de délimiter les trames. C'est un motif sur 7 octets qui permet de synchroniser l'émetteur et le récepteur sur la durée du bit. Chaque octet a la forme :

10101010

Note : il existe d'autres types de préambules dans d'autres protocoles.



La trame Ethernet IEEE



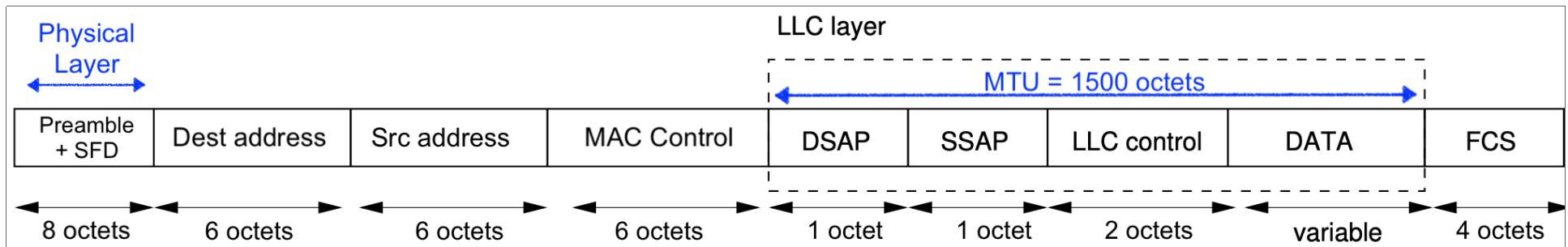
LE SFD (START FRAME DELIMETER)

Il permet d'indiquer le début de la trame. Cet octet a la forme :

101010 11



La trame Ethernet IEEE



LE CHAMP DE CONTRÔLE MAC

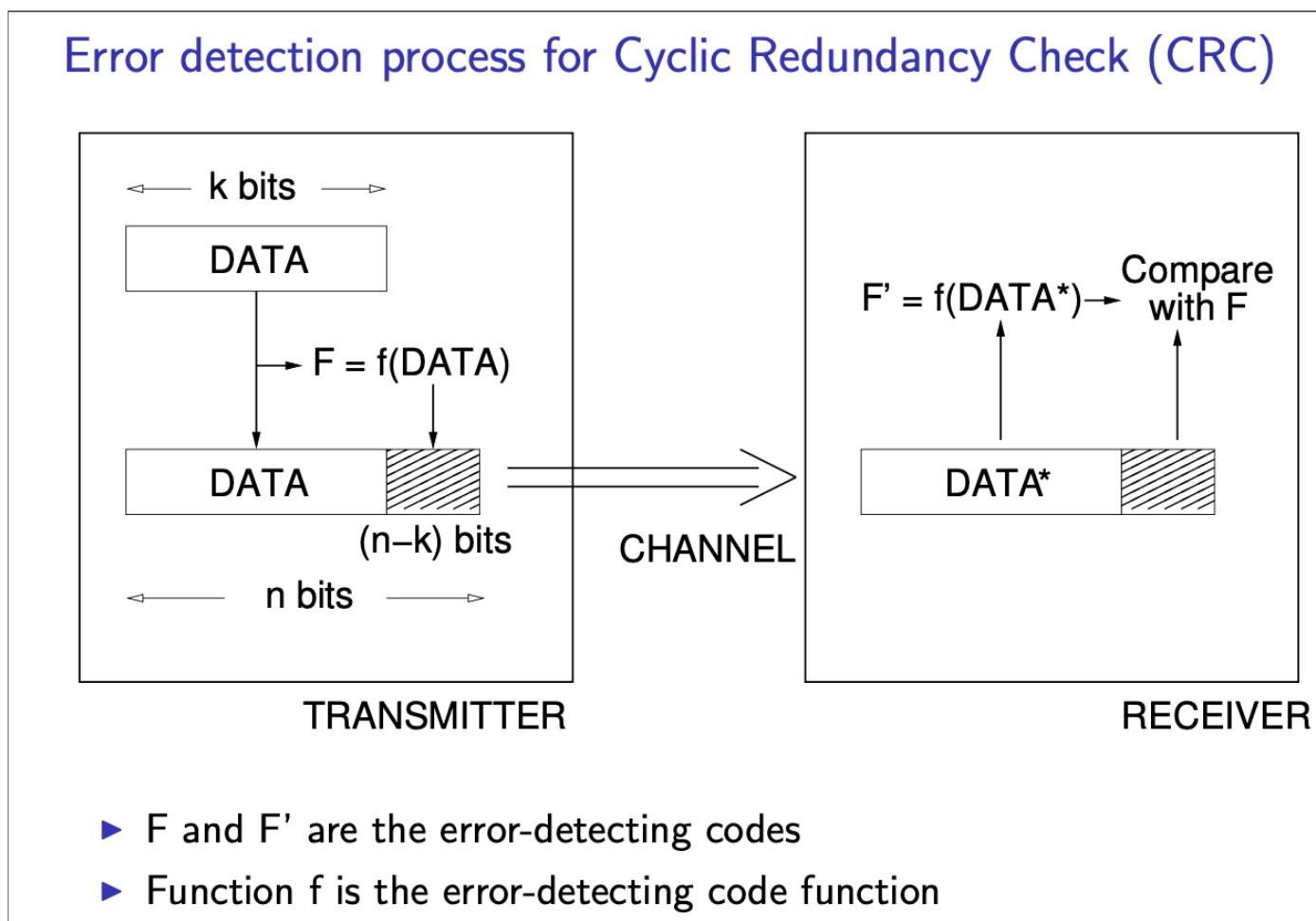
- Il comporte :
 - 4 octets utilisés par **IEEE802.1Q (optionnel)** => *On le verra dans le cours sur les VLAN*
 - 2 octets 'Ethertype' ou 'Length' => *j'y reviendrai plus tard*



La trame Ethernet IEEE

LE FCS

Il permet de détecter des erreurs dans la transmission avec un code cyclique sur 4 octets.



Efficacité maximale du protocole

=> Envoi en continu de trames de taille max :

$$\eta = \frac{Payload}{FrameSize} = \frac{1500}{1542} = 97,28\%$$

avec $FrameSize = 1500 + 8 + 2*6 + 6 + 4 + 12$
si le champ optionnel IEEE802.1Q existe.



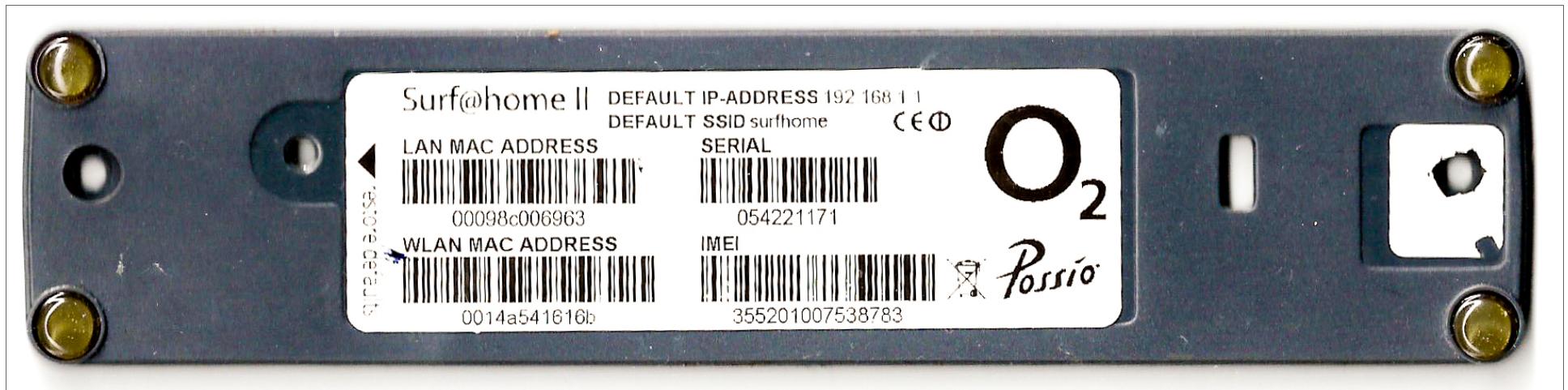
Debit utile maximal

On parle aussi de **throughput** en anglais : $D = \eta * R$ avec R le débit d'émission en bits/s.



L'adresse MAC

C'est une adresse unique, associée à une interface réseau (NIC - Network Interface Controller). Elle est codée sur 6 octets. Chaque octet est représenté par 2 chiffres en hexadécimal.



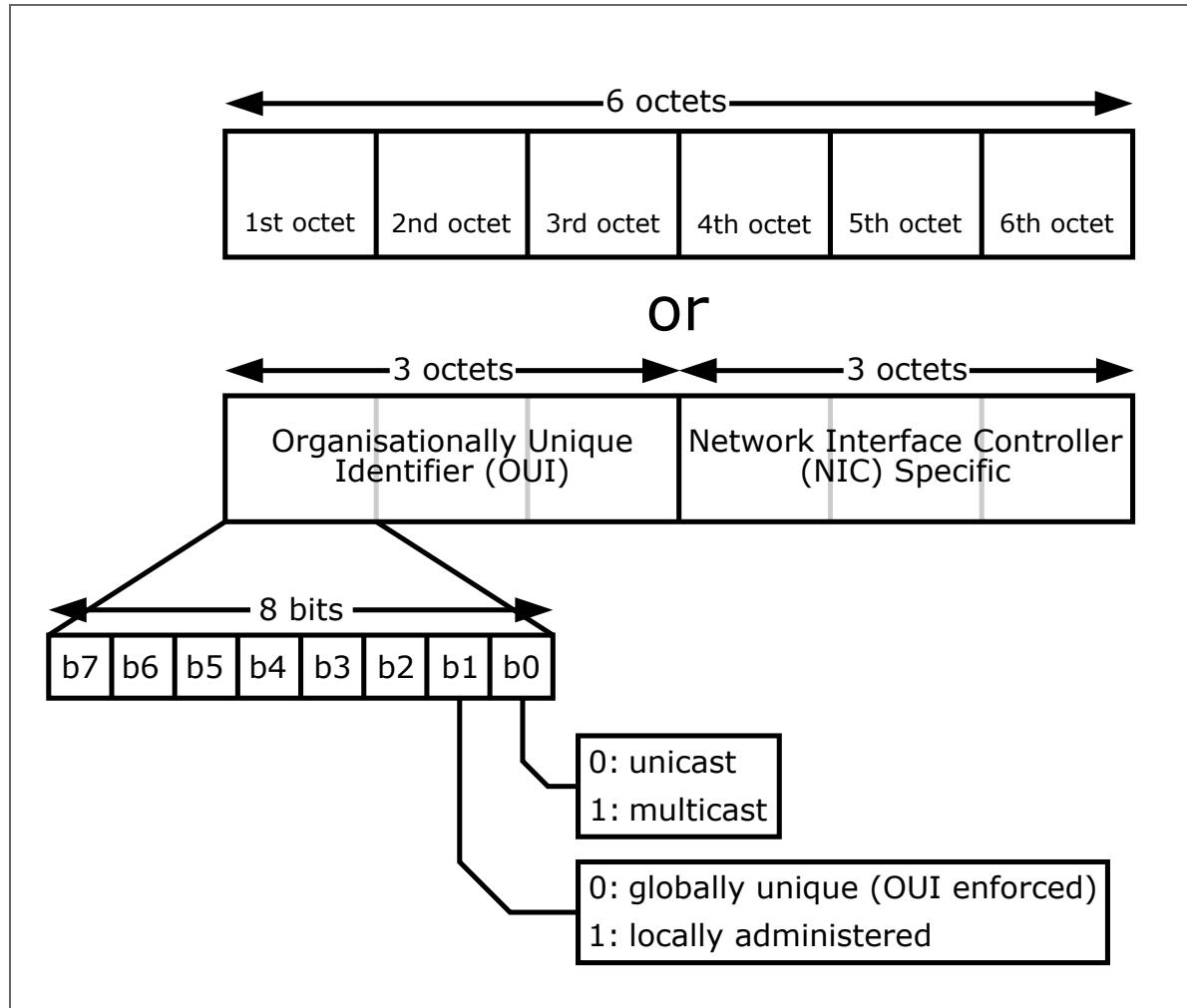
Ici on trouve 00-09-8c-00-69-63 pour l'interface Ethernet, et 00-14-a5-41-61-6b pour l'interface WiFi.



Ses autres petits noms : adresse physique ; adresse hardware.



L'adresse MAC : Format



Chaque octet est transmis de b₀ à b₁. Le 1er bit transmis est b₀ du 1er octet - le bit I/G.

By Inductiveload, modified/corrected by Kju - SVG drawing based on PNG

Unicité de l'adresse MAC : Bit U/L

- Une adresse *globally unique* est assignée par le constructeur, identifié par l'**OUI**
- Les interfaces des équipements peuvent changer d'adresses MAC si elles sont *locally administered* (adresses virtuelles)
 - Un équipement qui possède plusieurs interfaces (e.g. routeur) doit posséder des adresses MAC uniques pour les interfaces liées au même réseau.
 - Par contre, il peut posséder deux interfaces avec des adresses identiques si elles sont liées à deux réseaux différents



Réception des trames

Une interface sur un bus partagé ne traite que les trames qui présentent l'adresse MAC de destination identique à sa propre adresse MAC.

Il est possible de passer l'interface en **mode Promiscuous** pour qu'elle devienne un **espion** : elle traitera toutes les trames, sans distinction.



Unicast vs. Multicast : Bit I/G

Quand Bit I/G, le dernier du 1er octet (i.e. de poids le plus faible) mais le premier transmis, vaut 0, la transmission est **unicast**

Si le Bit I/G vaut 1, la communication est **multicast**. Plusieurs stations vont traiter le paquet. Ceci nécessite leur configuration pour qu'elles traitent les trames liées à l'adresse multicast.

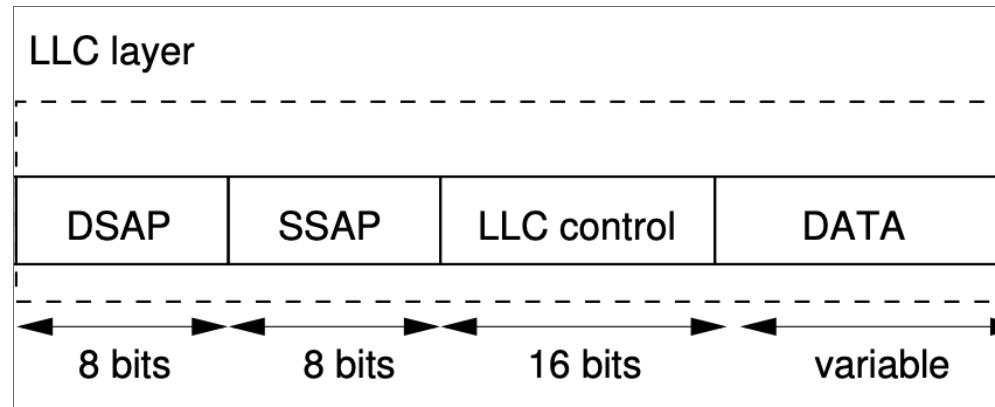
Adresse de broadcast

Tous les bits de cette adresse sont à 1 :

FF-FF-FF-FF-FF-FF



LLC



Permet le **multiplexage des protocoles de la couche 3** :

- DSAP (Destination Service Access Point) : indique le type de protocole de la couche 3 à la destination.
- SSAP (Source..) : idem à la source.
- Exemple de valeur pour TCP/IP : DSAP = 06.

Une doc détaillée mais assez claire



Contrôle d'erreur et de flux

LLC offre 3 modes de fonctionnement :

- Type 1 : aucun contrôle supplémentaire ; simple aiguillage des données vers les protocoles de couche 3. Mode non connecté, trame sans acquittement.
- Type 2 : type 1 + contrôle de séquence + contrôle de flux ; Mode connecté avec acquittement.
- Type 3 : type 1 + acquittement de trame.

LLC Control comporte les informations pour que les mécanismes de contrôle d'erreur et de flux puissent avoir lieu.

Note : On verra le contrôle de flux et d'erreur plus en détail dans le cours Réseaux Télécom.



Ethernet II vs. IEEE802.3

Ethernet II (DIX Frame) est une version alternative à IEEE802.3

Il y a deux façons d'envoyer des trames Ethernet (si on ne souhaite pas de contrôle d'erreur / de flux):

- En utilisant IEEE802.3 + LLC
- En utilisant Ethernet II sans LLC

La seconde est la plus utilisée car la plus simple à implanter.

| 802.3 Ethernet packet and frame structure | | | | | | | | | |
|---|-------------------------|-----------------------------|--------------------|------------|-----------------------|--|------------------------|-----------------------------------|-------------------------|
| Layer | Preamble | Start frame delimiter (SFD) | MAC destination | MAC source | 802.1Q tag (optional) | Ethertype (Ethernet II) or length (IEEE 802.3) | Payload | Frame check sequence (32-bit CRC) | Interpacket gap (IPG) |
| Length (octets) | 7 | 1 | 6 | 6 | (4) | 2 | 42–1500 ^[c] | 4 | 12 |
| Layer 2 Ethernet frame | (not part of the frame) | | ← 64–1522 octets → | | | | | | (not part of the frame) |
| Layer 1 Ethernet packet & IPG | ← 72–1530 octets → | | | | | | ← 12 octets → | | |

(src : Wikipedia)

Ethernet II ou IEEE802.3 ?

Tout se joue dans le champ **Ethertype** de 2 octets:

- Si le champ vaut une **valeur ≥ 1536** , il décrit le *type de protocole qu'il encapsule*.
 - 0x0800 indicates that the frame contains an **IPv4** datagram,
 - 0x0806 indicates an **ARP** datagram,
 - and 0x86DD indicates an **IPv6** datagram
- S'il vaut ≤ 1500 , il donne *la taille des données transportées dans la trame*. On est en IEEE 802.3 et il faut ensuite lire l'en-tête LLC pour connaître le type de protocole encapsulé (DSAP).

Ainsi les deux types de trames peuvent cohabiter sur un même support Ethernet. La longueur de la trame avec Ethernet II est obtenue en détectant le FCS et les 12 octets IPG qui séparent deux trames successives.



Pour finir

<https://app.wooclap.com/RLCM2>



[CM3]

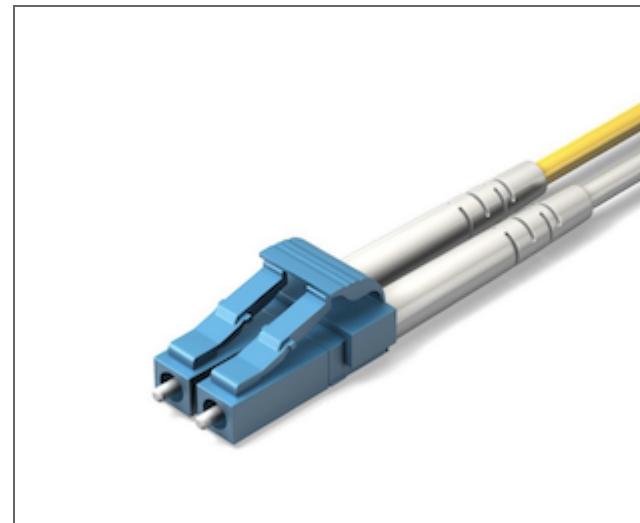
Ethernet commuté

LE LIEN ETHERNET



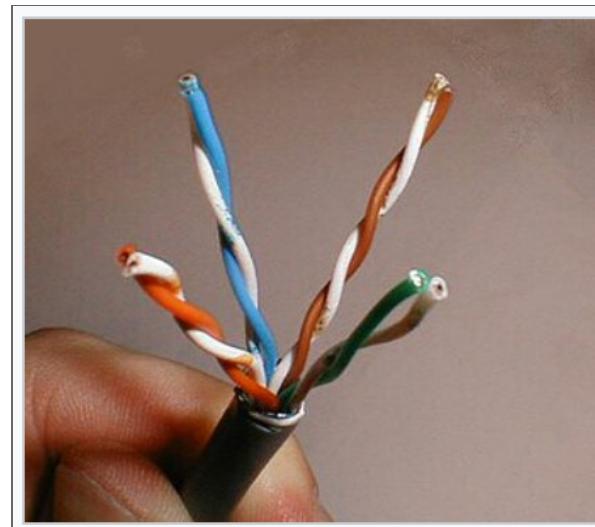
Le lien Ethernet

- Première version sur câble coaxial **10Base2**.
- Il existe plusieurs types de supports. Actuellement on a :
 - Les *paires torsadées* et le connecteur RJ45
 - La fibre optique, via différents connecteurs (LC, SC, MTRJ, ST, MTP).



La paire torsadée

La **paire torsadée** est une ligne symétrique formée de deux fils conducteurs enroulés en hélice l'un autour de l'autre. Un câble Ethernet se compose de 4 paires torsadées.



La paire torsadée

- Différentes normes de blindage pour limiter les interférences extérieures et réduire le taux d'erreur bit.
 - De la paire torsadée non blindée à la *super blindée* !
- Il y a différentes catégories de câbles capables d'atteindre des débits différents :
 - La catégorie 5 : une bande passante de 100 MHz et un débit allant jusqu'à 100 Mbit/s.
 - La catégorie 5e (en anglais *enhanced*) peut permettre une quantité d'information allant jusqu'à 1 Gbit/s. C'est un type de câblage testé à 100 MHz



Standard **X**Base**Y**

- Standard de câblage de l'IEEE pour la technologie Ethernet :
 - X = débit binaire du réseau, en Mbit/s
 - Y = type de connexion utilisée.
- Par exemple, pour la technologie Fast Ethernet à 100Mbit/s, on a 100Base-T car :
 - X = 100Mbit/s
 - Y = T pour *Twisted Pair*



NORME 100BASE-T

Le **100BASE-TX** est une norme de câblage réseau mise au point pour Fast Ethernet 100BASE-T.

NORME 100BASE-F

Le 100Base-FX est un exemple de lien Ethernet en fibre optique. Il fonctionne avec 2 brins, l'un pour la transmission des données et l'autre pour la réception.



Half et Full Duplex

- Les supports Ethernet historiques étaient **half duplex** : *bidirectionnel à l'alternat*. Il suppose que le média est partagé entre plusieurs stations et que les informations transitent dans les 2 sens.
- Le mode **full duplex** correspond à une communication point-à-point entre deux équipements => Dans ce contexte, le média n'est plus partagé entre plus de 2 stations et les informations transitent toujours dans les deux sens mais sur des canaux (paires torsadées ou fibres) distincts.
- En conséquence, en full duplex, l'algorithme d'accès au média est considérablement simplifié et **la bande passante utile est doublée en full duplex**.



Auto-négociation

- Quand 2 équipements activent leur interface de communication sur un même câble, ils doivent se mettre d'accord sur
 - Le débit binaire à utiliser,
 - Le mode half ou full duplex.
- La fonction de négociation utilise les signaux de contrôle d'état du lien physique. L'ordre des négociations va du plus rapide au plus lent :
 - 100BaseTX Full-Duplex,
 - 100BaseT4,
 - 100BaseTX,
 - 10BaseT Full-Duplex,
 - 10BaseT.



Gigabit Ethernet

- Les paires torsadées de catégorie 5 sont certifiés pour des fréquences allant jusqu'à 100MHz et ne permet pas d'atteindre 1Gbps
- La couche physique a été entièrement revue : c'est une « fusion » de deux technologies : l'Ethernet IEEE802.3 sur une interface physique de type Fiber Channel ANSI X3/T11.
- Standard de câblage : 1000BaseX tel que défini dans le standard GigabitEthernet IEEE802.3z (1998)
- En 2002 : 10 Gbit/s avec IEEE802.3ae, le 10GigabitEthernet.



Gigabit - 1000BaseX

- 1000BaseLX : laser grandes ondes sur fibre optique multimodes et monomode destiné aux artères de campus. Longueur max 3km
- 1000BaseSX : laser ondes courtes sur fibre optique multimodes destiné aux artères intra-muros. Longueur max 500m
- 1000BaseCX : câble en paires torsadées blindées 150 Ohms destiné aux connexions entre serveurs dans le même local. Longueur max 25m
- 1000BaseT : câble en paires torsadées non blindées de catégorie 5 avec IEEE802.3ab pour déployer le Gbit/s entre les stations et le premier commutateur. Longueur max 100m.



La prise RJ45

- C'est le *petit nom* de la broche 8P8C (8 positions et 8 contacts électriques) du standard TIA/EIA-568-B
- Il est utilisé avec tous les XBase-T.
- Câblage droit vs. croisé
 - Droit : entre une station et un équipement d'interconnexion
 - Croisé : entre deux équipements d'interconnexion.
- Aujourd'hui, la connexion entre deux appareils se fait par un câble droit car la carte réseau est généralement capable d'analyser si le câble utilisé est croisé ou non (cette détection est intégrée aux cartes gigabit).



[CM3]

Ethernet commuté

EQUIPEMENTS D'INTERCONNEXION

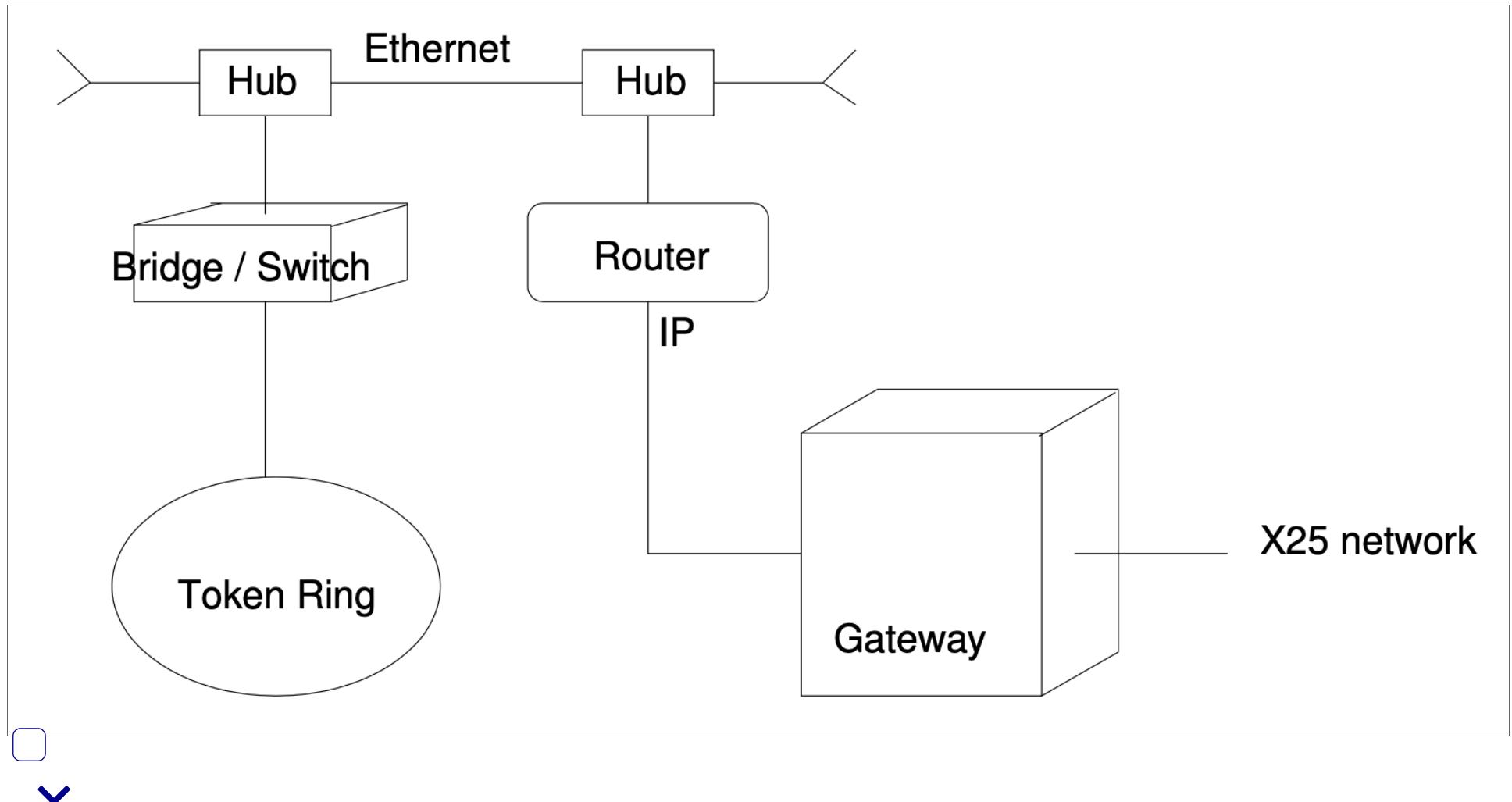


Equipements d'interconnexion

- Niveau Physique : le Hub / le Répéteur
- Niveau Liaison : le Pont / le Switch
- Niveau Routage : le Routeur / la Gateway : segmentation des domaines de diffusion



Equipements d'interconnexion



Le répéteur (cf. 1SN)

- Topologies en Bus, legacy Ethernet
- Ré-émet les bits reçus pour augmenter la portée
- A placer tous les 500m.
- Maximum 4 répéteurs sur 2500.

DEPRECATED...



Le hub (cf. 1SN)

- Topologie en étoile
- Ré-émet les bits reçus sur un câble vers les autres ports :
=> Extension du domaine de diffusion à tous les câbles connectés au hub.



Le pont

- Équipement d'interconnexion de deux réseaux locaux *sans décapsulation IP* :
 - Dissimilaires : e.g. Point d'accès WiFi
 - Similaires : e.g. Réseau Ethernet commuté.
En Ethernet commuté
- on décrit le pont également par le terme de **switch** ou **commutateur**

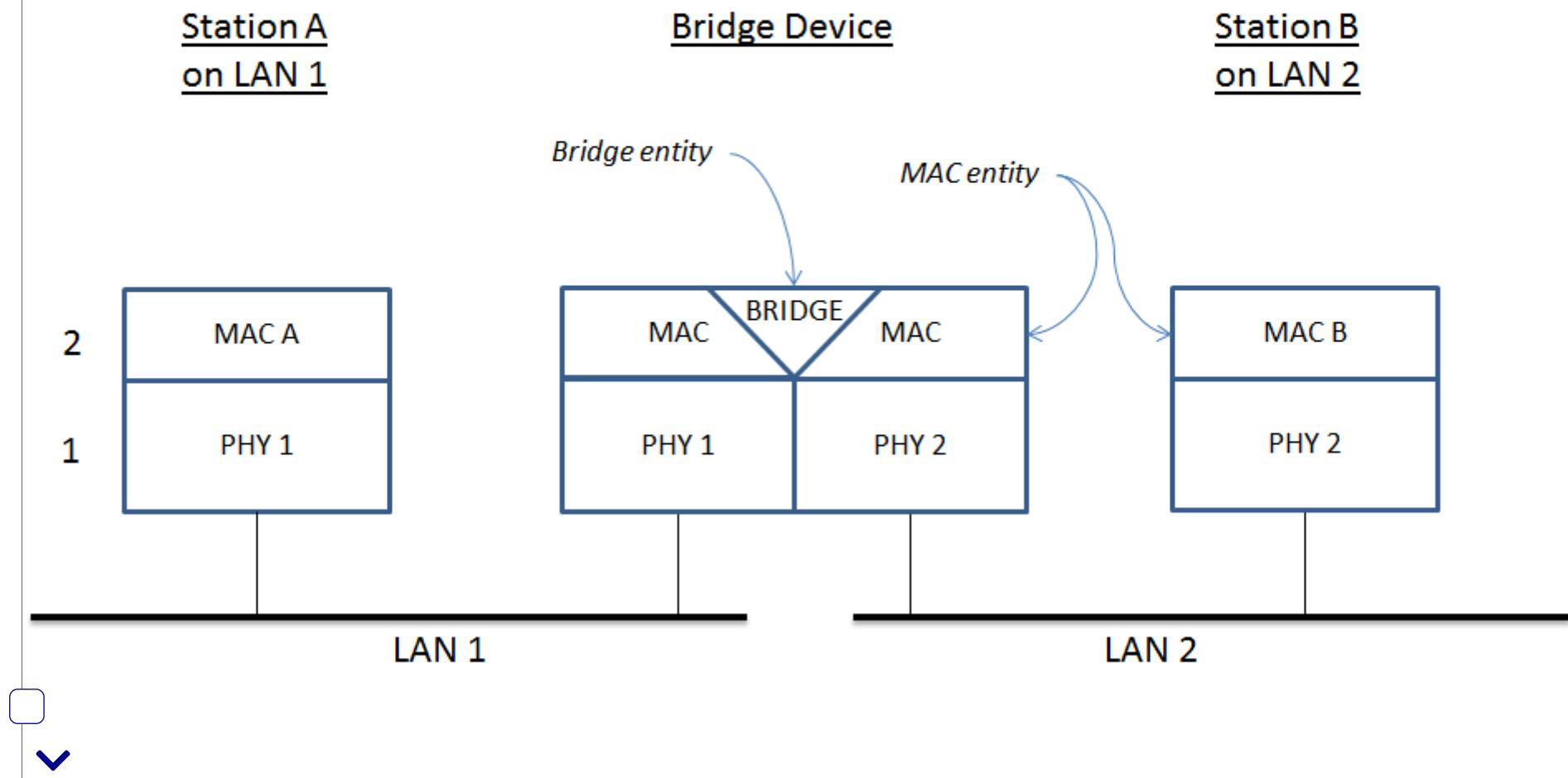


Le pont

- Relayage :
 - si les protocoles sont similaires, on commute simplement
 - s'ils sont dissimilaires, on doit traduire ou encapsuler les trames (cf. cours d'interconnexion en 2SN)
- Peut implanter un routage de niveau 2 :
 - Si nécessaire, détermine le chemin entre deux entités
 - Existe dans IEEE802.1Q pour les réseaux commutés.

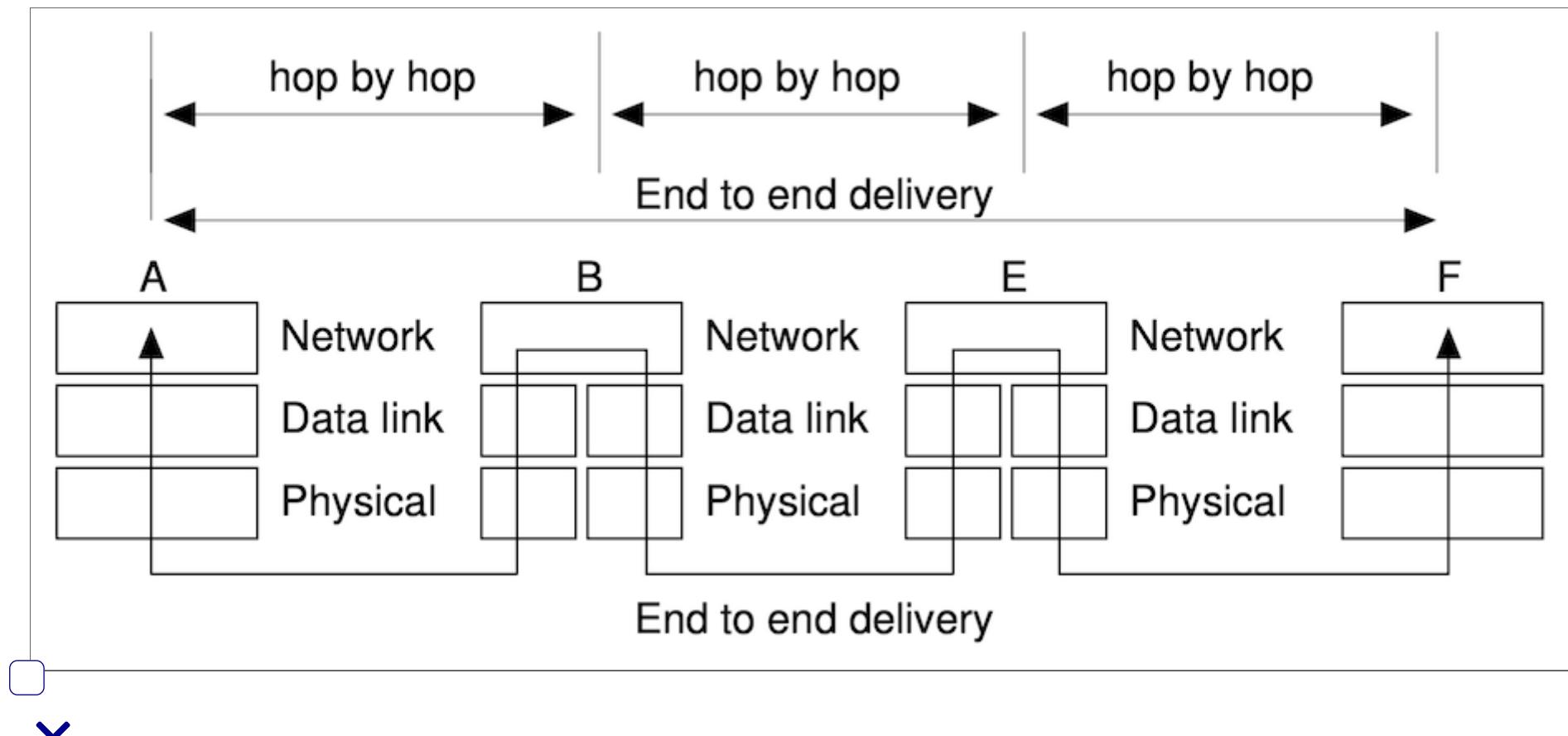


A bridge connecting two LAN segments



Routeur et Gateway

- Equipement interconnexion de deux réseaux de couche 3.
- Responsable du routage des messages entre les réseaux de couche 2.



Segmentation des domaines

- Les équipements niveau 2 permettent de segmenter les domaines de collision.
- Les équipements niveau 3 permettent de segmenter les domaines de collision et de diffusion.



[CM3]

Ethernet commuté

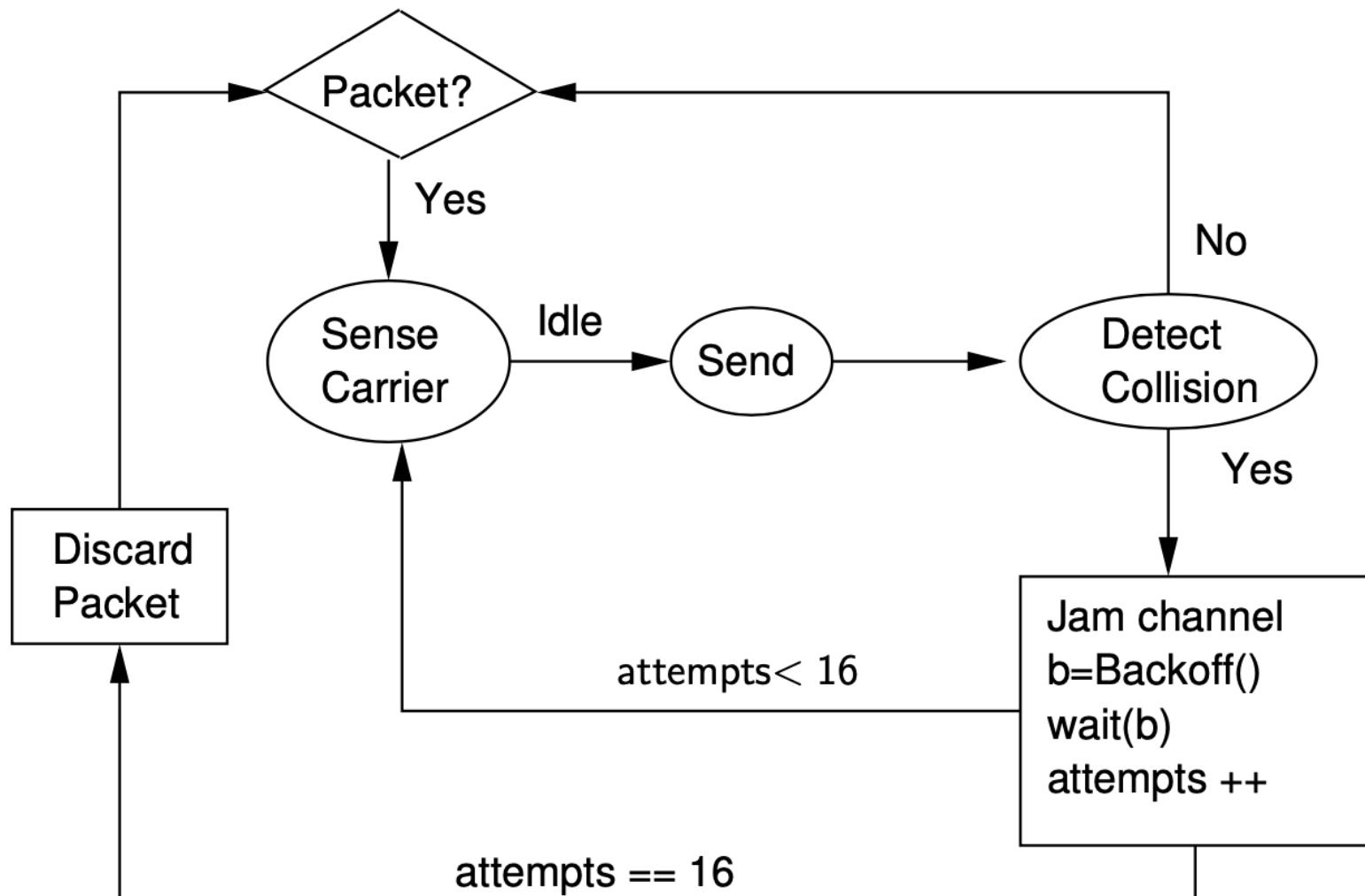
COMMUTATION



CSMA/CD

- ▶ In ALOHA, decisions to transmit are made without paying attention to what other nodes might be doing
- ▶ In CSMA/CD, the device listens to the line *before and during sending*
- ▶ If the line is idle (no carrier sensed)
 - ▶ Send packet immediately
 - ▶ Must wait 9.6 μs between any back-to-back frames (IFS: InterFrame Space)
- ▶ If the line is busy (carrier sensed)
 - ▶ Wait until idle and transmit packet immediately
 - ▶ Called I-persistent sending
- ▶ Collision detected (while sending)
 - ▶ Stop sending and jam signal
 - ▶ Try again later

CSMA/CD State Diagram



Exponential Backoff

The idea:

If a collision is detected, each station waits until its slot of $51.2\mu s$ is finished.

Then it waits a random time and tries again if bus is idle.

- ▶ Random waiting time is selected using binary exponential backoff
 - ▶ First collision: choose K from {0,1} and delay = $K * 51.2\mu s$
 - ▶ Second collision: choose K from {0,1,2,3} and delay = $K * 51.2\mu s$
 - ▶ n^{th} time: choose K from $\{0, \dots, 2^n - 1\}$ and delay = $K * 51.2\mu s$
Max value for K=1023 (i.e. $n = 10$)
 - ▶ Give up after several tries (usually 16) and report transmit error to upper layer
- ▶ If delays were not random, then there is a chance that sources would retransmit in each step
- ▶ Why not just choose from a small set for K?
 - ▼ ▶ This works for a low number of stations
 - ▶ Large number of nodes would result in more collisions

Rappels : Ethernet I (cf. 1SN)

- Cette première version d'Ethernet utilisait des câbles coaxiaux. Elle présentait les **avantages** suivants :
 - Protocole distribué : ajout/suppression de stations transparents
 - Meilleure résilience aux défaillance des équipements
 - Simplicité de l'implémentation et du déploiement, coût réduit
 - Bon débit binaire



Rappels : Ethernet I (cf. 1SN)

- Et les **désavantages** suivants :
 - Impact important de la gestion des collisions
 - Utilisation du support restreinte :
 - 30% est considéré élevé.
 - Passage à l'échelle au-delà de 50% d'utilisation : réseau très congestionné.
 - Distance maximale de câble à respecter



Ethernet commuté

OBJECTIFS

1. Maximiser l'utilisation de la bande passante
2. Garder la simplicité de déploiement, le coût réduit

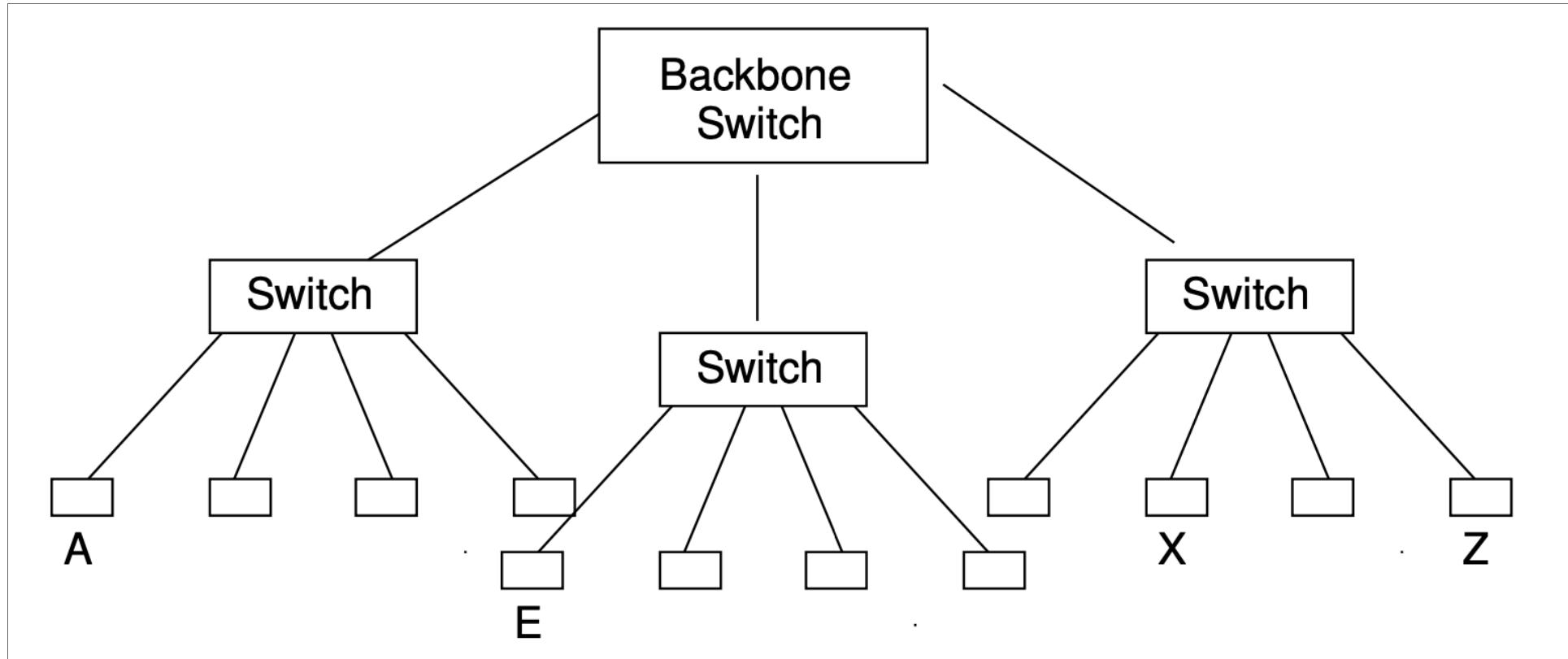
MOYENS

1. Changement de topologie : **Etoile**
2. Equipements actif : **Commutateur / Switch**
3. Communication sur un lien : **Full Duplex**
=> CSMA/CD est désactivé.

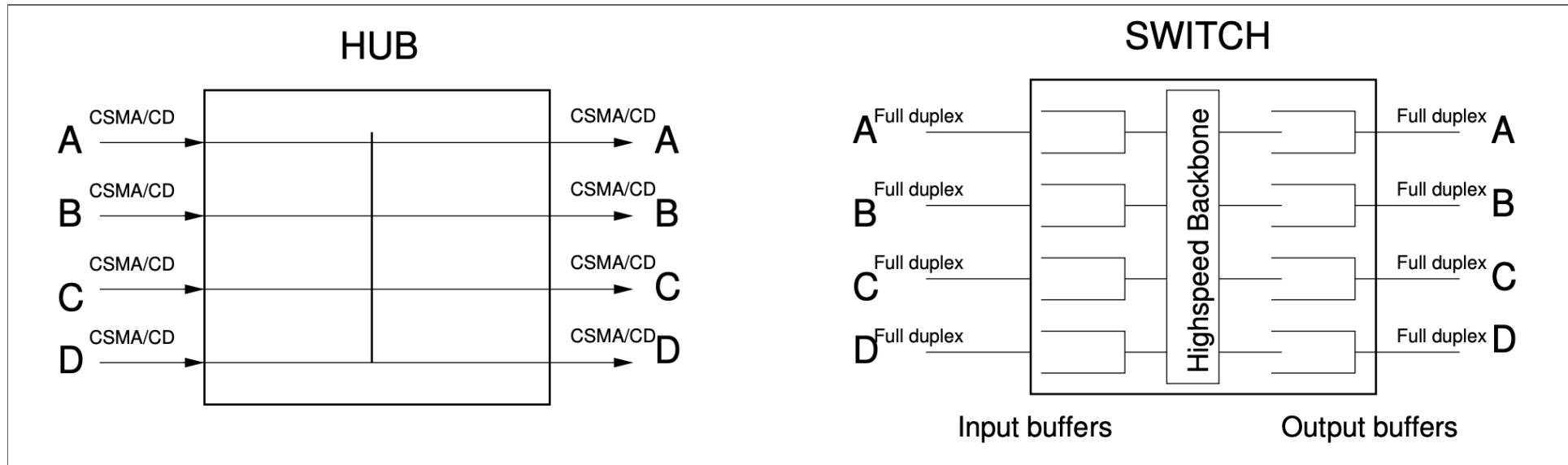
Distance maximale n'est plus limitée pour détecter les collisions. Seule la limite physique de la perte en puissance avec la distance existe.



Ethernet commuté



Switch vs. Hub

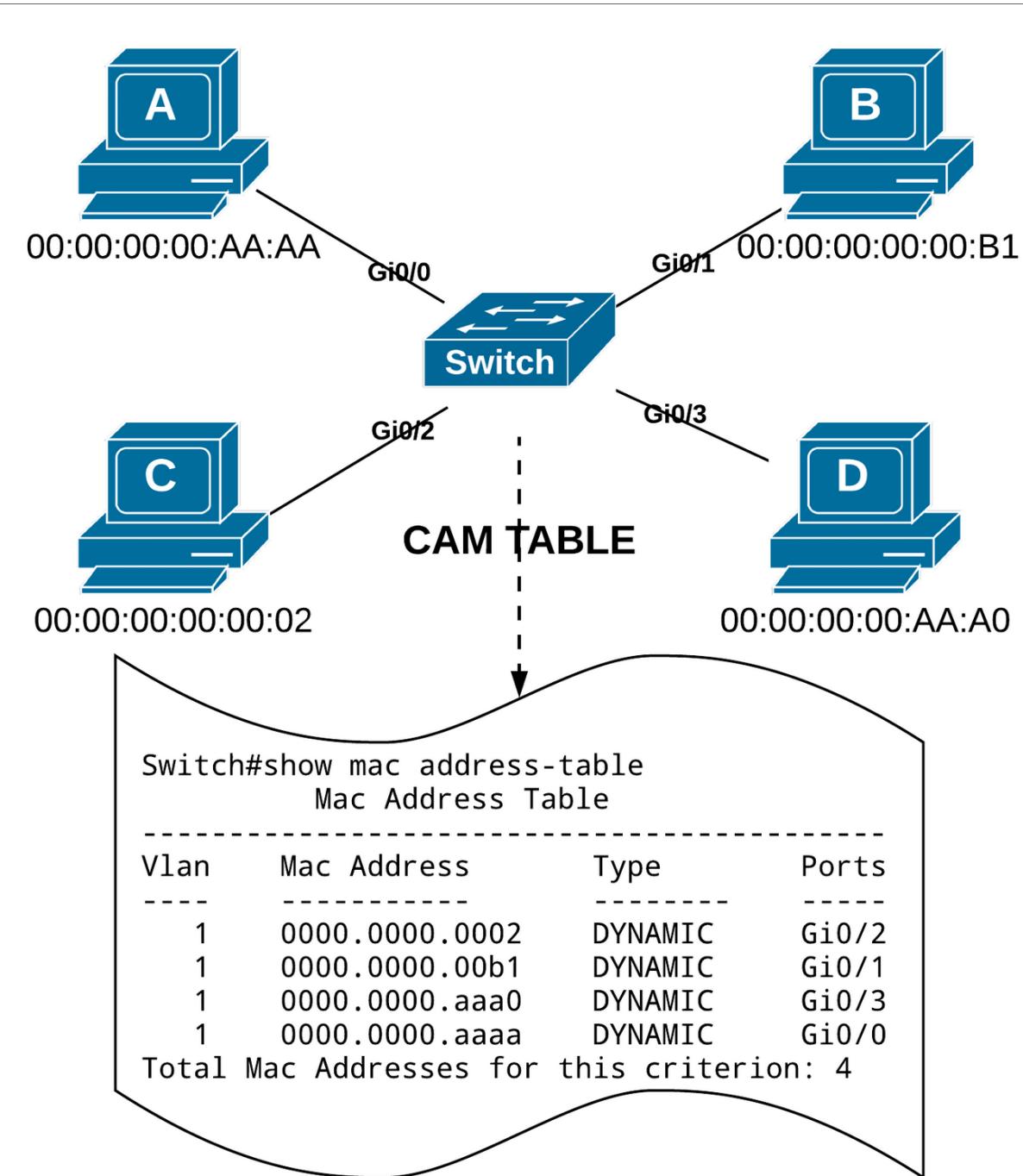


Chaque commutateur possède **une table de commutation** qui associe une adresse MAC à un numéro de port de sortie et à un age.

Une entrée a la forme :

{Adresse MAC - Port de sortie - Age}

- Transfert très rapide d'une trame en entrée vers son port de sortie : puces spécialisées (ASIC)
- Chaque entrée de la table présente un Age : estampille temporelle qui permet
✓ d'en déduire si l'entrée est obsolète ou non.



<https://cisco.goffinet.org/ccna/ethernet/commutation-ethernet-cisco/>

Modes de commutation

STORE AND FORWARD

- La trame est entièrement enregistrée dans la file d'attente du port d'entrée,
- La trame est vérifiée (FCS)
- Le port de sortie est déterminé
- Elle est placée dans un file d'attente FIFO du port de sortie

La trame attend que le support soit libre si des trames la précèdent dans le port de sortie.

VOCABULAIRE EN ANGLAIS

- **Ingress** port : port d'entrée
- **Egress** port : port de sortie



Modes de commutation

FAST FORWARD

Principe : La trame n'est pas entièrement reçue que les premiers bits sont déjà envoyés vers le port de sortie

- Le port de sortie est déterminé dès que l'adresse destination est lue
- La trame **n'est pas vérifiée**
- Si le port de sortie n'est pas disponible, le switch repasse en Store and Forward.

Il existe **deux variantes** :

- **Cut-Through** : dès que l'adresse destination est lue, les bits sont commutés vers le port de sortie
- **Fragment Free** : dès que les 64 premiers octets sont lus, les bits sont commutés.

Fragment free permettait de détecter une collision sur le lien en entrée s'il est en half duplex



=> Fragment free : Inutile en full duplex.



Apprentissage de la table de commutation

S'apparente à un routage de niveau 2 car on détermine les ports de sortie pour atteindre une adresse MAC destination dans le réseau.

On parle de **pont transparent** car il apprend la correspondance entre port de sortie et adresse MAC destinataire sans intervention humaine.



Apprentissage de la table de commutation

Transparent bridging can also operate over devices with more than two ports. As an example, consider a bridge connected to three hosts, A, B, and C. The bridge has three ports. A is connected to bridge port 1, B is connected to bridge port 2, C is connected to bridge port 3. A sends a frame addressed to B to the bridge. The bridge examines the source address of the frame and creates an address and port number entry for host A in its forwarding table. The bridge examines the destination address of the frame and does not find it in its forwarding table so it floods (broadcasts) it to all other ports: 2 and 3. The frame is received by hosts B and C. Host C examines the destination address and ignores the frame as it does not match with its address. Host B recognizes a destination address match and generates a response to A. On the return path, the bridge adds an address and port number entry for B to its forwarding table. The bridge already has A's address in its forwarding table so it forwards the response only to port 1. Host C or any other hosts on port 3 are not burdened with the response. Two-way communication is now possible between A and B without any further flooding to the network. Now, if A sends a frame addressed to C, the same procedure will be used, but this time the bridge will not create a new forwarding-table entry for A's address/port because it has already done so.



Texte : [Wikipedia](#)

Apprentissage de la table de commutation

Le protocole appliqué se décompose en 2 règles :

- 1. Apprentissage "Lazy"** : A chaque réception de trame sur un port X, je mets à jour ma table de commutation avec l'entrée qui fait correspondre l'adresse MAC source avec le port X. Si l'entrée existe, je mets à jour son *age*, sinon je créé une nouvelle ligne.
- 2. Si pas d'entrée dans la table** : Si j'ai une trame à commuter vers une adresse (MAC) qui n'apparaît pas dans ma table de commutation, je l'envoie sur tous les ports différents du port d'où elle vient ;



Standard IEEE802.1Q - version 2022

Tout ce que je vous raconte la commutation, et ce qui va suivre, est standardisé dans la norme IEEE802.1Q, édition (ou version) 2022.

On trouve ce standard en ligne ici : [https://ieeexplore.ieee.org/
document/10004498](https://ieeexplore.ieee.org/document/10004498)

"Bridges and Bridged networks", 1993 pages.

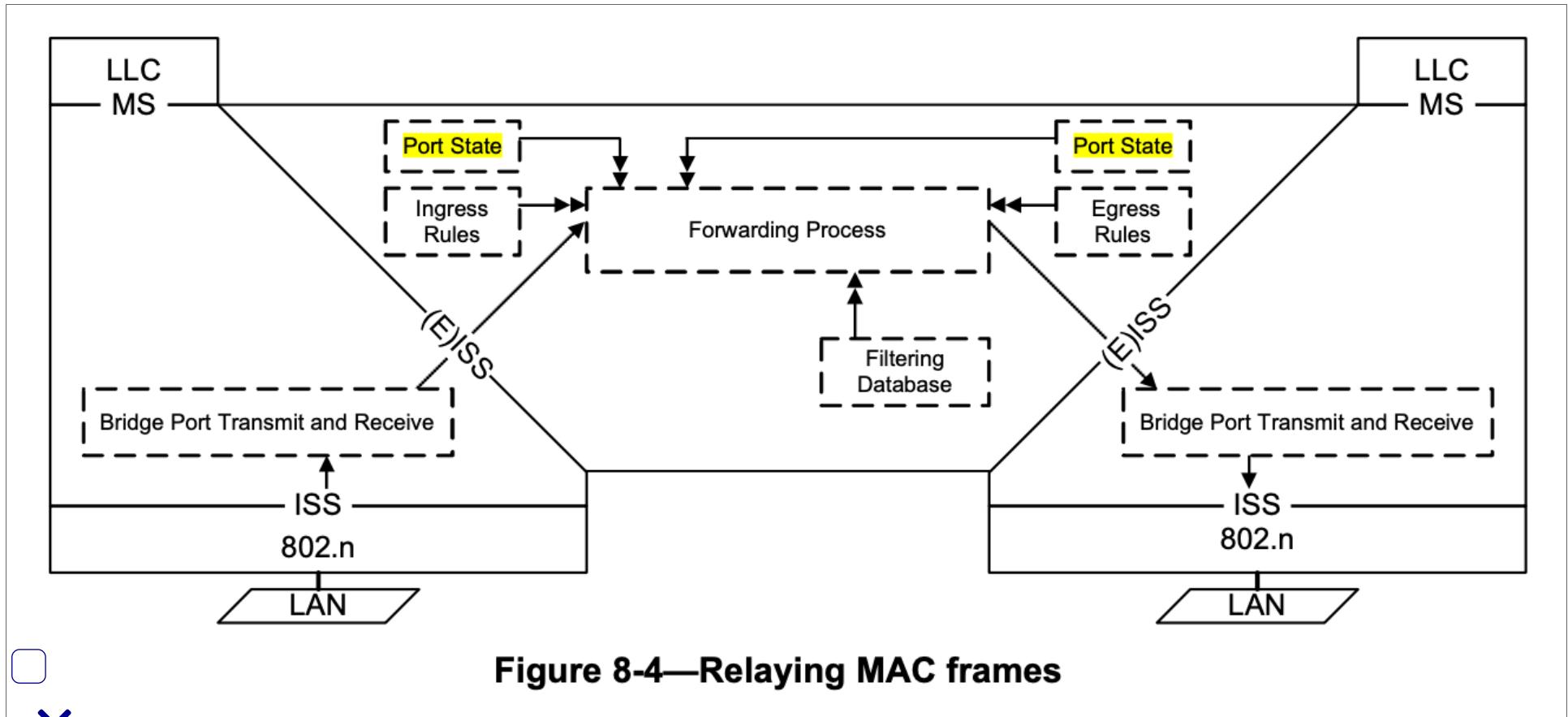
- Les standards évoluent ..
 - d'abord avec des amendements qui sortent au rythme des groupes de travail,
 - et régulièrement, une version complète qui concentre l'ancienne version complète avec ses amendements est produite.
- Ainsi on trouve 802.1Q - version 2022, version 2018, version 2014...



Standard IEEE802.1Q - version 2022

On trouve page 215, paragraphe 8.7 - The Learning Process :

- La table de commutation s'appelle **Filtering Database (FDB)**
- Elle a une taille finie, on a donc des mécanismes pour supprimer des entrées
 - Sur l'age par exemple (300s. de durée de vie par défaut)



- On notera que chaque port a un état : **Port State**

Avant d'aller plus loin ...

<https://app.wooclap.com/RLCM3>



[CM4]

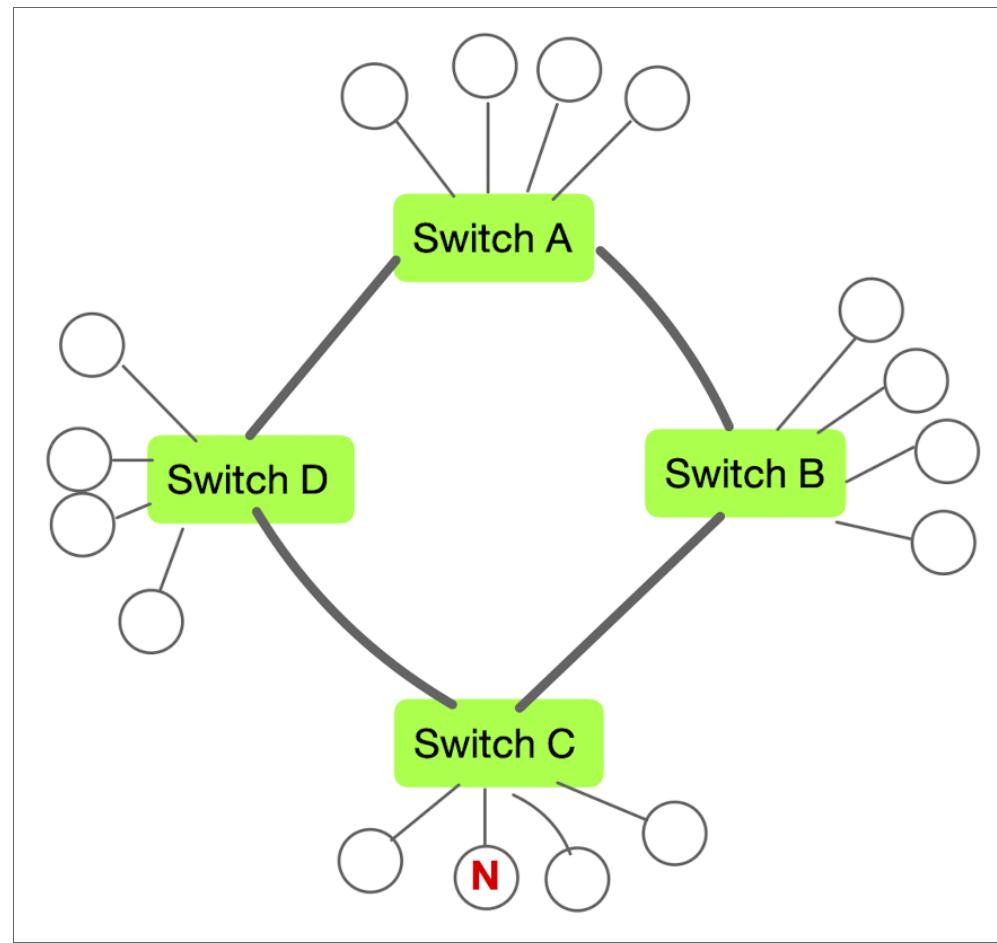
Le danger des boucles



On suppose que la **station N** émet une trame pour une destination **X**.
Dans le réseau suivant, **il n'existe pas de station d'adresse X ...**
C'est donc une destination inconnue du réseau ..



Que se passe-t-il ?



Un peu de poésie pour calmer cette tempête de diffusion ...



Algorhyme

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

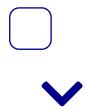
*A tree whose crucial property
Is loop-free connectivity.*

*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.*

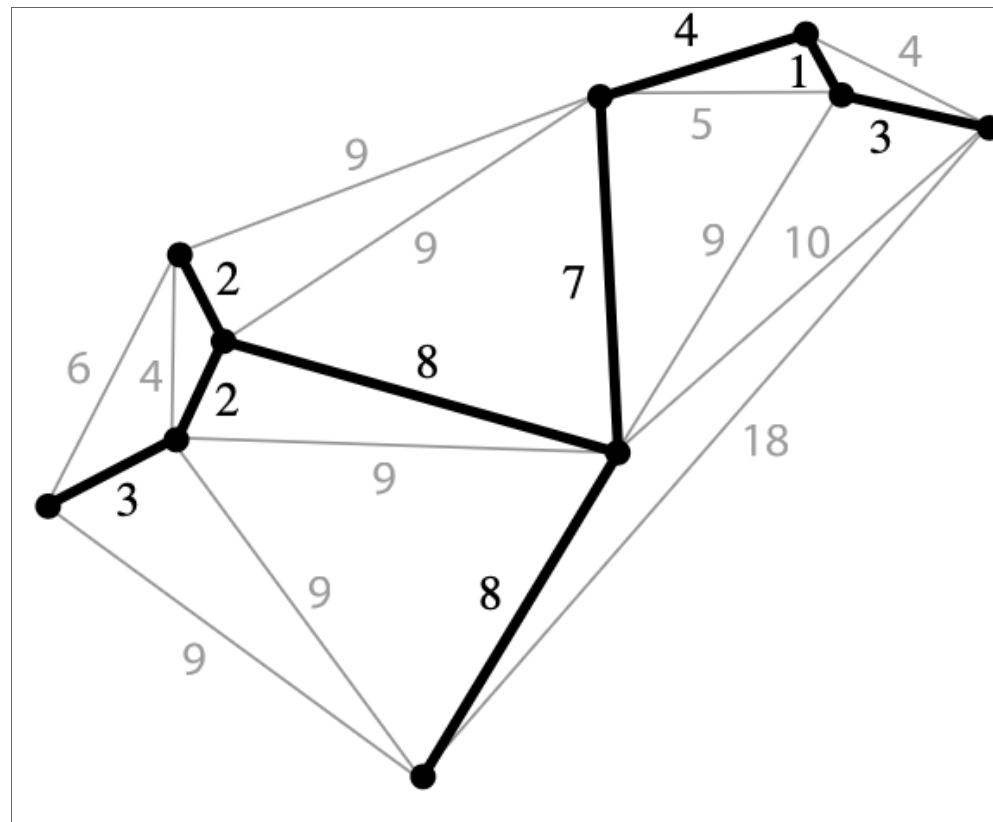
*A mesh is made by folks like me
Then bridges find a spanning tree.*



Radia Perlman, SIGCOMM 1985
<https://dl.acm.org/doi/10.1145/319056.319004>

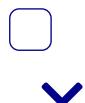
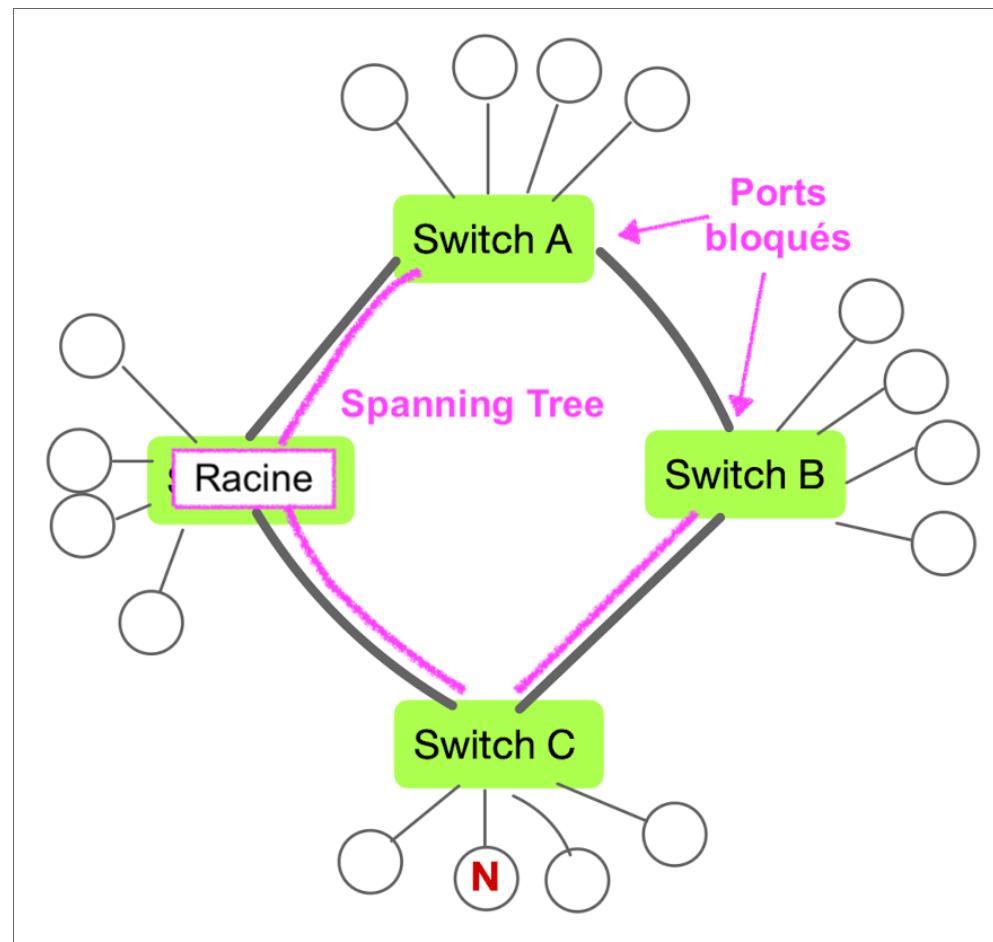
Spanning Tree

ou arbre couvrant



Spanning Tree Protocol (STP)

Construire une vue logique du réseau (i.e. **Active Topology**) pour éviter les tempêtes de diffusion (i.e. *broadcast storms*) sur la topologie physique.



Etat des ports

Un port est caractérisé par plusieurs booléens :

- **Learning :**
 - Si VRAI, observe les trames entrantes pour renseigner la Filtering Database.
- **Forwarding :**
 - Si VRAI, il peut émettre et recevoir des données
 - Sinon il est **bloqué**



Grandes étapes STP

1. Election de la racine : le plus petit identifiant (niveau de priorité + @MAC)
2. Sélection du port qui offre le plus court chemin à la racine comme **port racine** (root port ou **RP**) : Forwarding = VRAI.
3. Sélection du port opposé sur le même lien comme **port désigné** ou de **relayage** (designated port ou **DP**) : Forwarding = VRAI.
4. Blocage des ports qui ne sont ni RP ni DP : Forwarding = FAUX.

Attention :

C'est un **algorithme distribué**, de routage de niveau 2



Messages de contrôle

Les messages utilisés pour faire communiquer les switch entre eux sont d'un format particulier. Ils respectent le format des messages **BPDU**

Bridge Protocol Data Unit

- Un message BPDU est encapsulé dans une trame Ethernet.
- L'adresse MAC destination est un groupe multicast 01:80:C2:00:00:00 qui ne regroupe que les interfaces entre les switch.



Types de trames BPDU

Il y a 2 principaux types de trames BPDU :

- **Configuration messages (utilisés par STP)**
- Topology Change Notification (TCN)



Format BPDU

| Champ | Description |
|------------------------|---|
| type de protocole | fixé à 0 pour STP |
| version | 0 pour STP, 2 pour RSTP |
| type de message | Configuration ou TCN |
| des flags | pour diverses indications |
| Root Identifier | Identifiant du switch racine |
| Path cost | Coût sur le chemin entre le switch qui a émis ce message et la racine |
| Switch Id | Identifiant du switch émetteur du BPDU |
| Port Id | Identifie le port d'émission de ce BPDU |

Format BPDU (suite)

| Champ | Description |
|---------------|--|
| Message age | Temps écoulé depuis l'émission du messages |
| Max Age | Age maximal |
| Hello Time | Période entre l'envoi des messages par la racine |
| Forward Delay | Durée que met un port entre l'écoute des BPDU et son activation dans STP |



Les champs en gras sont importants à connaître.

Switch and Port identifiers

L'identifiant d'un switch est défini par :

Switch Id = { Priorité switch sur 2 octets + @MAC switch }

L'identifiant d'un switch est défini par :

Port Id = { Priorité port sur 1 octet + ID Port }

Table 13-3—Bridge and Port Priority values

| Parameter | Recommended or default value | Range |
|-----------------|------------------------------|---------------------------|
| Bridge Priority | 32 768 | 0–61 440 in steps of 4096 |
| Port Priority | 128 | 0–240 in steps of 16 |

Les Priorités sont toutes configurables !



Coût des liens

Table 13-4—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 | 2 000 000–200 000 000 | 1–200 000 000 |
| 10 Mb/s | 2 000 000 | 200 000–20 000 000 | 1–200 000 000 |
| 100 Mb/s | 200 000 | 20 000–2 000 000 | 1–200 000 000 |
| 1 Gb/s | 20 000 | 2000–200 000 | 1–200 000 000 |
| 10 Gb/s | 2000 | 200–20 000 | 1–200 000 000 |
| 100 Gb/s | 200 | 20–2000 | 1–200 000 000 |
| 1 Tb/s | 20 | 2–200 | 1–200 000 000 |
| 10 Tb/s | 2 | 1–20 | 1–200 000 000 |

Exemple de configuration STP

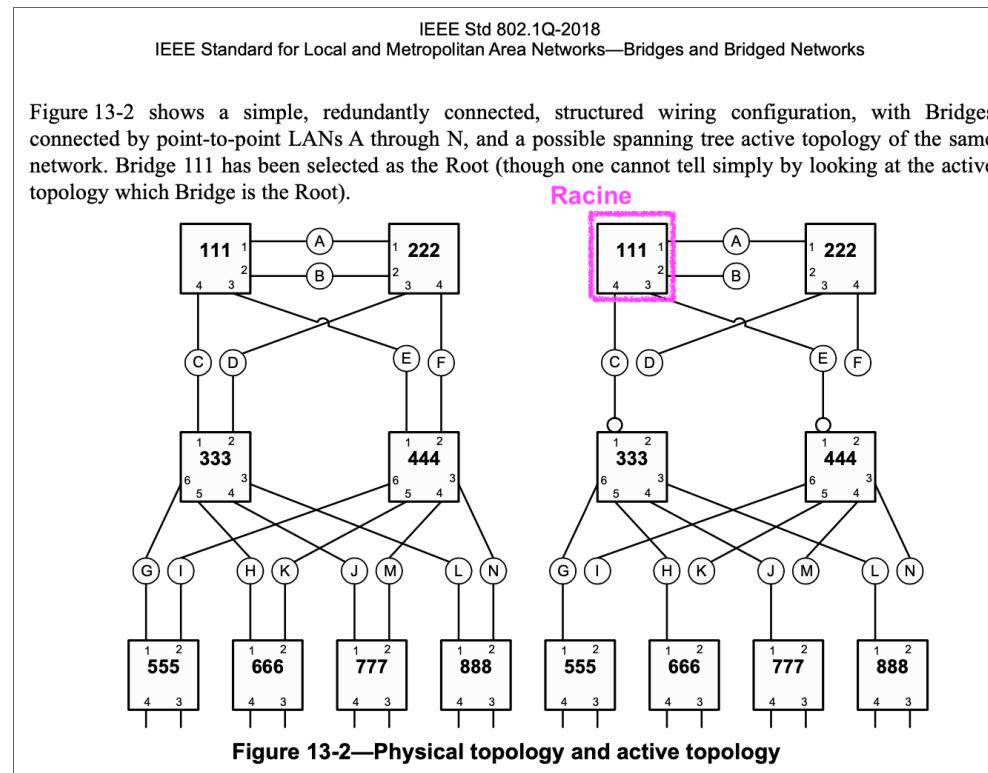
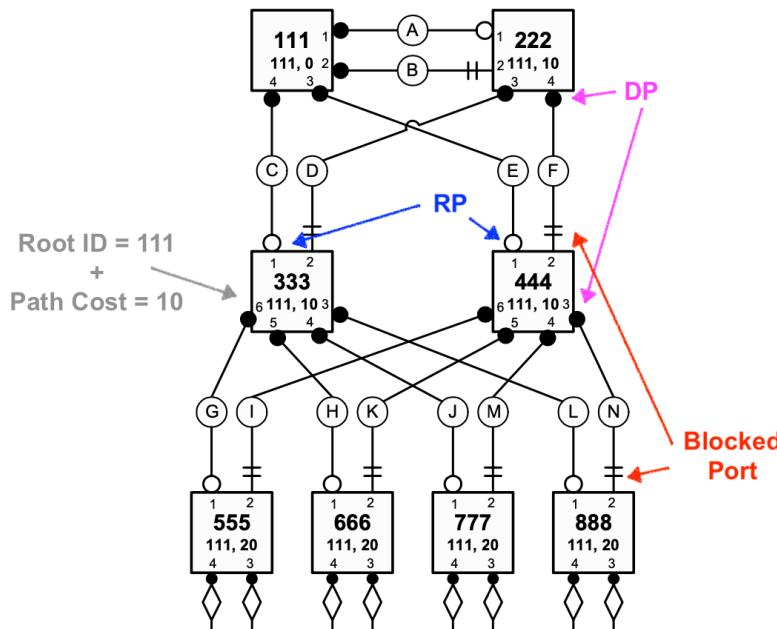


Figure 13-3 shows the Port Roles and Port States of each Bridge Port. It can be seen that Bridge 111 is the Root, as its Ports are all Designated Ports, each of the remaining Bridges have one Root Port.



Spanning Tree Protocol et consorts

Dans IEEE802.1Q - 2022, on trouve plusieurs **protocoles de routage de niveau 2** :

- **STP** : Spanning Tree Protocol (1990)
 - La proposition d'origine. Mets du temps à converger (30-50s.)
- **RSTP** : Rapid Spanning Tree Protocol (1998)
 - Amélioration du temps de convergence
- **MSTP** : Multiple Spanning Tree Protocol (début 2000)
 - Une extension de RSTP par groupe de VLAN (on verra plus tard ce qu'est un VLAN ...)
- **SPB** : Shortest Path Bridging (2012)
 - Permet à tous les liens d'être actifs, et de répartir la charge entre différents chemins (sans boucle). C'est un protocole de routage à état de lien (type OSPF).

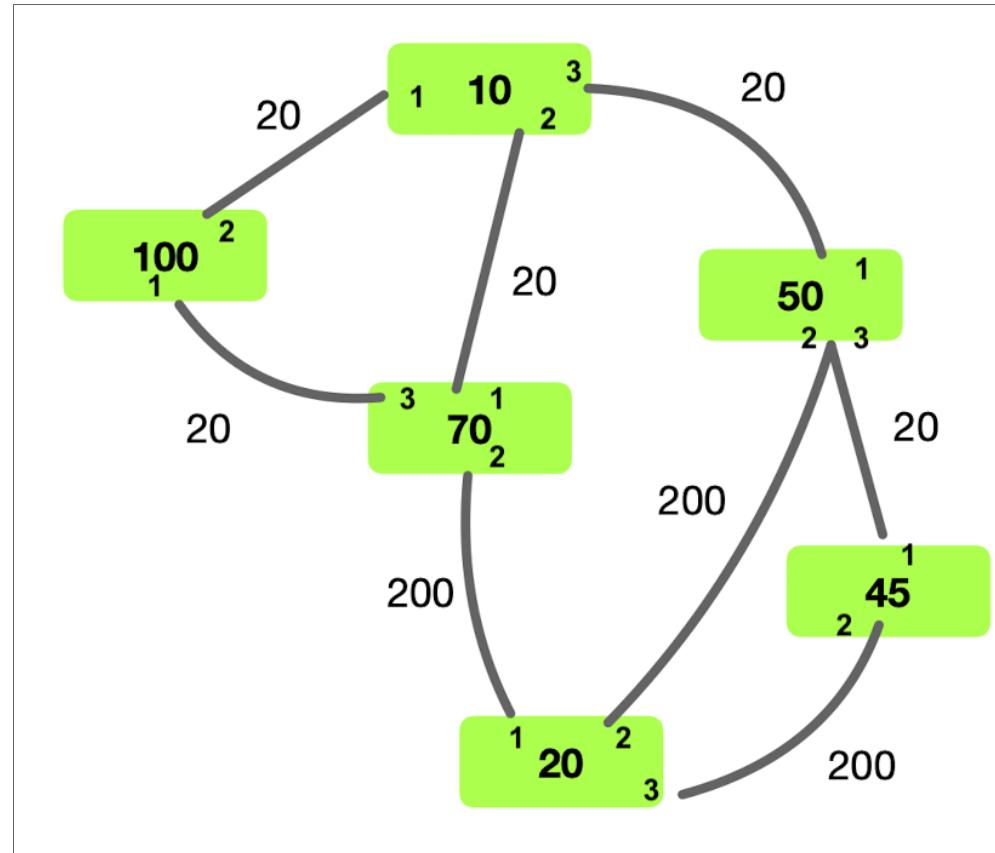


Avant d'aller plus loin ...

<https://app.wooclap.com/RLCM4>



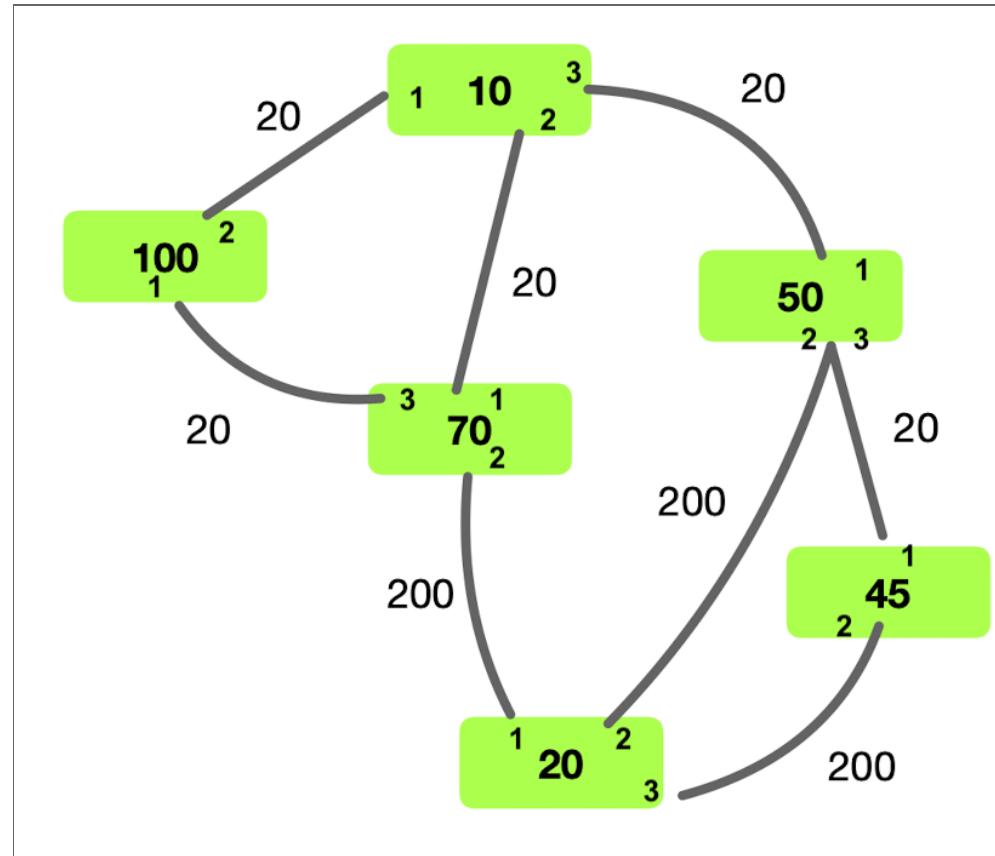
Et maintenant, comment ce protocole distribué fonctionne-t-il ?



Rappel

Les coûts des liens sont généralement inversement proportionnels au débit binaire.

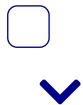
Et maintenant, comment ce protocole distribué fonctionne-t-il ?



Rappel

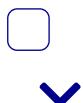
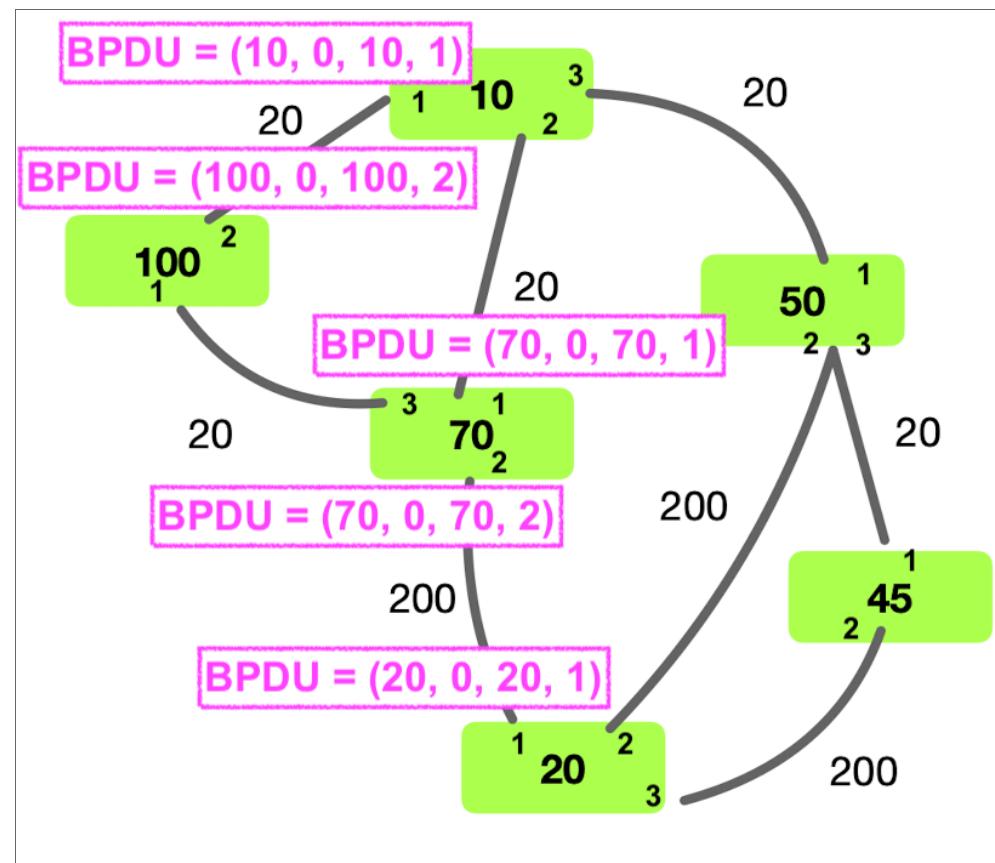
Les **identifiants des switch et des ports** sont issus de :

- Leurs adresses MAC et numéro de ports respectifs,
- Des priorités possiblement choisies par l'administrateur.



Configuration BPDU

Chaque switch envoie sur ses ports un BPDU pour s'annoncer comme racine avec les données :
(Root Id ; Path Cost = 0 ; Swith Id ; Port Id)



Quand 100 reçoit le message de 10, il observe que son identifiant est supérieur à celui de 10.

- Il considère que 10 est sa racine.
- Il mets à jour son Root Path Cost à 20.
- Il **choisit le port 2 comme RP**
- Il envoie le message suivant sur son port 1 :

(Root = 10 ; Path cost = 20 ; Switch ID = 100 ; Port = 1)

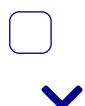
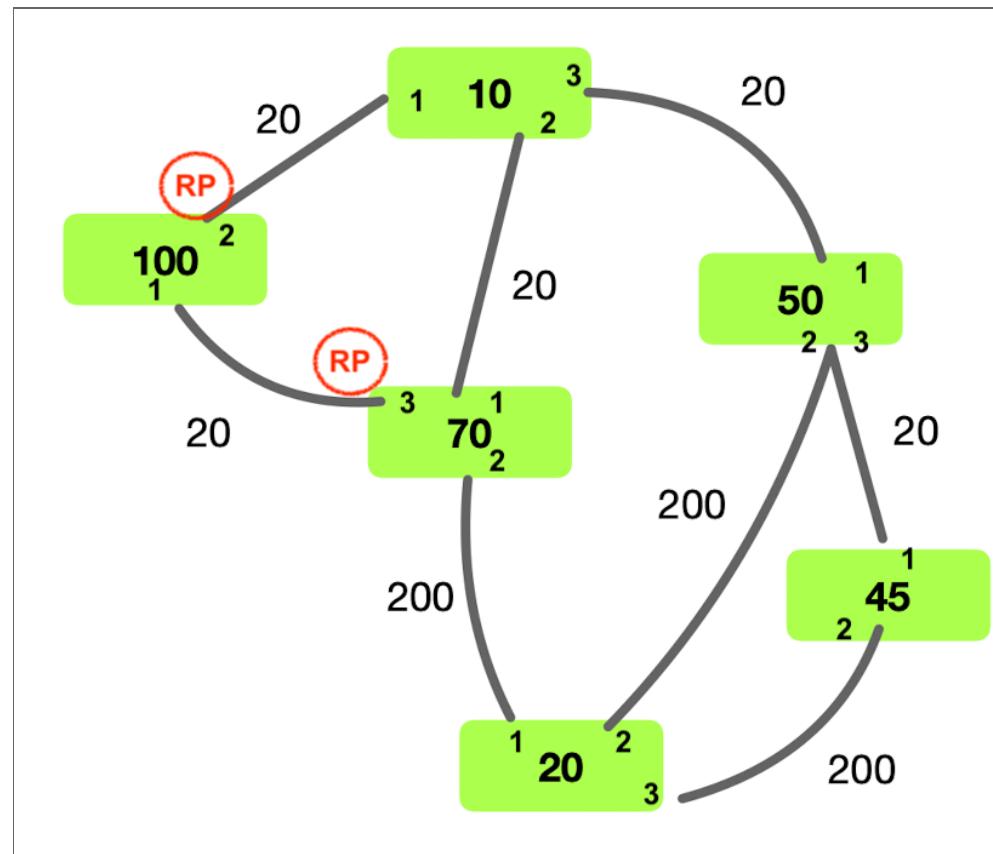
- et le message suivant sur son port 2 :

(Root = 10 ; Path cost = 20 ; Switch ID = 100 ; Port = 2)



Quand 70 reçoit ce message de 100 sur son port 3 :
(Root = 10 ; Path cost = 20 ; Switch ID = 100 ; Port = 1)

- Il considère que 10 est sa racine
- Il met à jour son Root Path Cost à 40
- Il choisit le **port 3 comme RP**
- Il annonce dans son BPDU qu'il a choisi 10 comme racine et qu'il est à distance 40

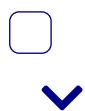
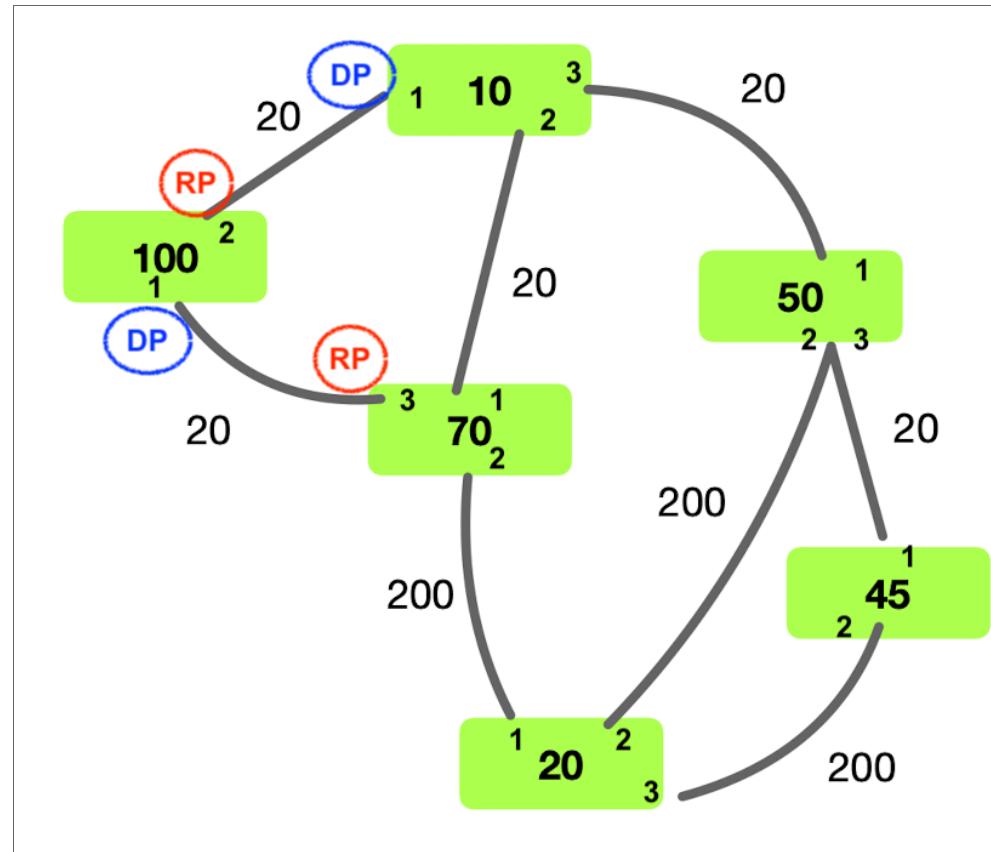


Quand 10 reçoit ce message de 100 sur son port 1 :

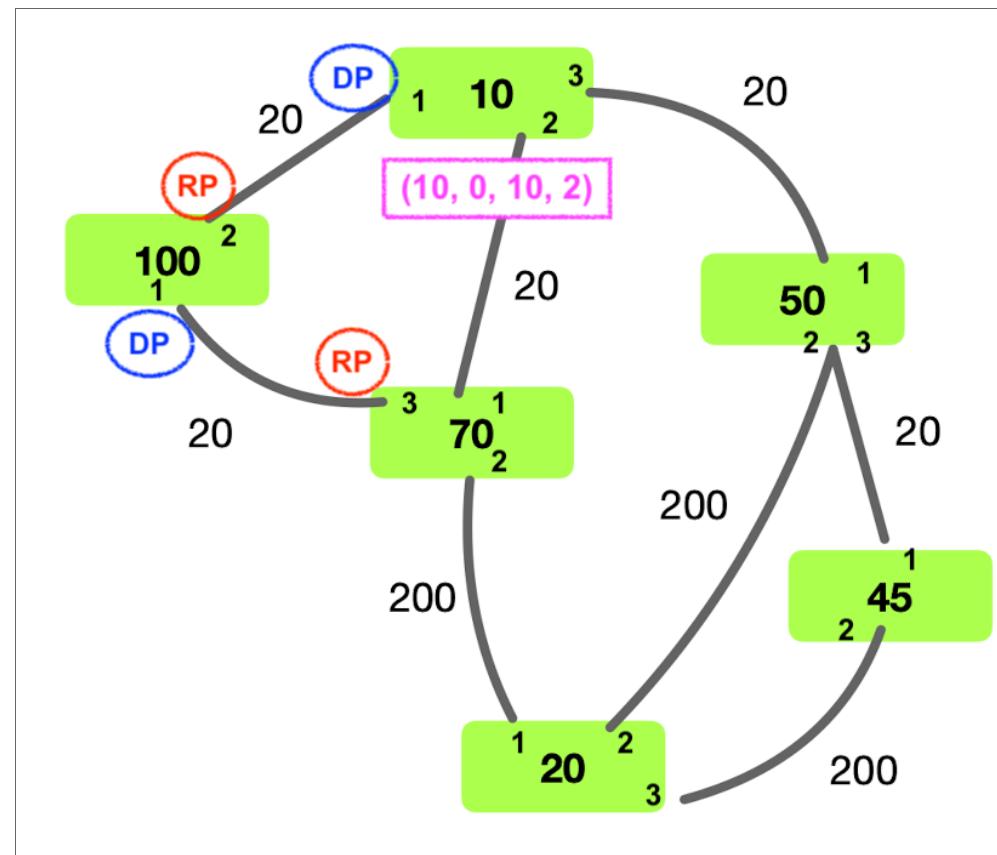
(Root = 10 ; Path cost = 20 ; Switch ID = 100 ; Port = 2)

- Il sait qu'il a été élu racine par 100
- Il met son **port 1 en DP**
- Il continue d'émettre le même message d'annonce qu'avant.

Quand 100 reçoit le BPDU de 70 qui annonce que 10 est sa racine, il met son port 1 en DP aussi



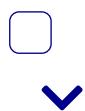
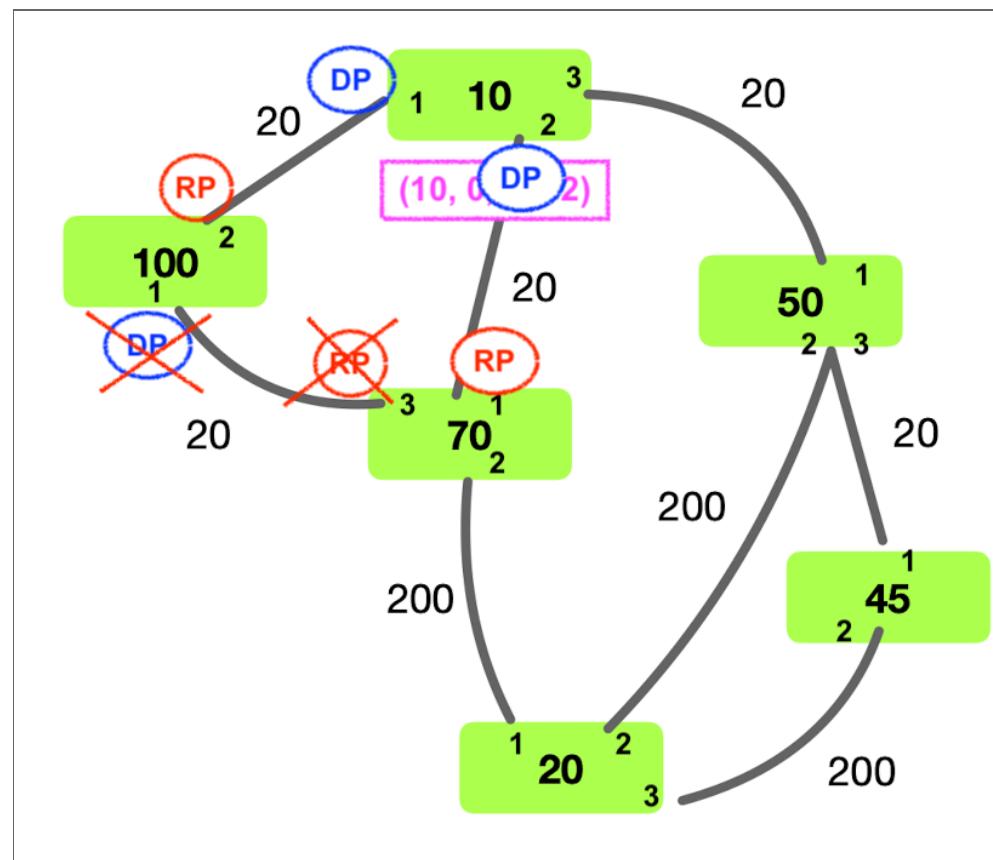
Mais 70 reçoit plus tard une annonce de 10 (mais qu'il aurait aussi pu recevoir avant le message de 100...), qui présente un Path Cost plus petit par le port 1



70 doit donc changer son port RP du port 3 au port 1 :

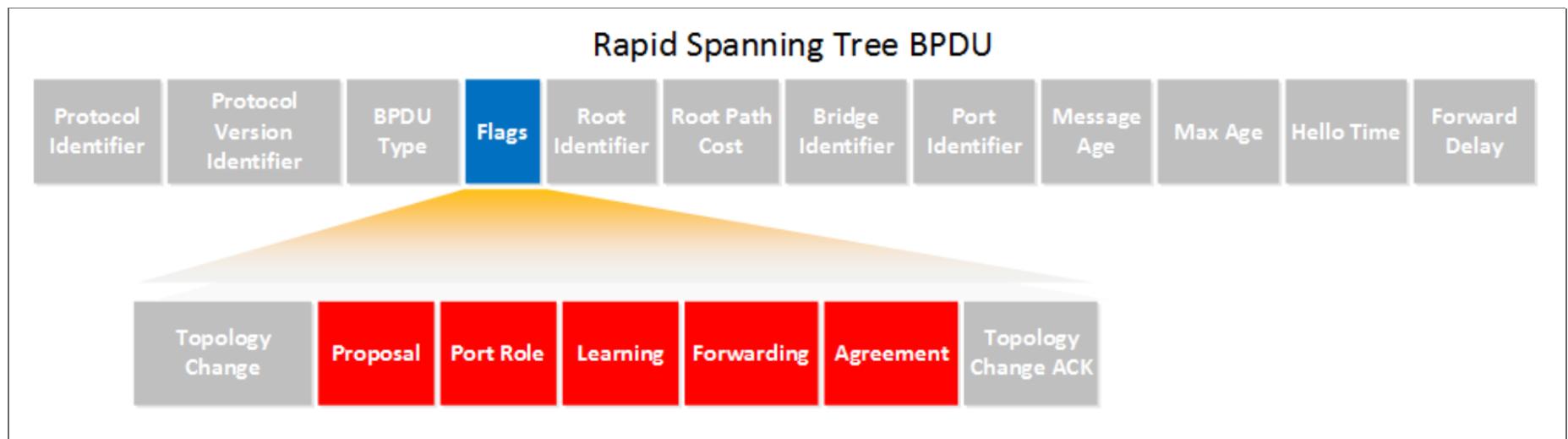
- Il sélectionne le port 1 comme RP
- Bloque le port 3
- N'envoie plus de BPDU sur le port 3

=> 100 va se rendre compte que son port 1 est en face d'un port bloqué, et supprimer la mention DP pour bloquer son port.



En RSTP : convergence grâce aux flags

- Quand un BPDU est émis sur un port, les flags présentent les informations importantes pour la convergence de cet algorithme

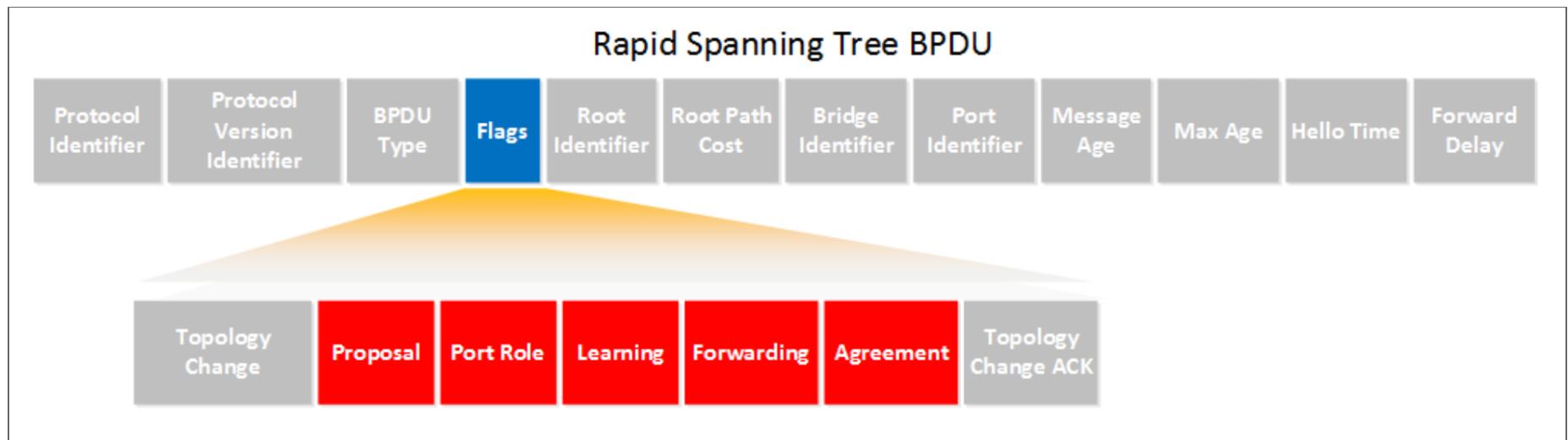


- Si **Learning** = 1 :
 - apprentissage de la table de commutation
- Si **Forwarding** = 1 :
 - relayage et broadcast possible via ce port

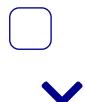


En RSTP : convergence grâce aux flags

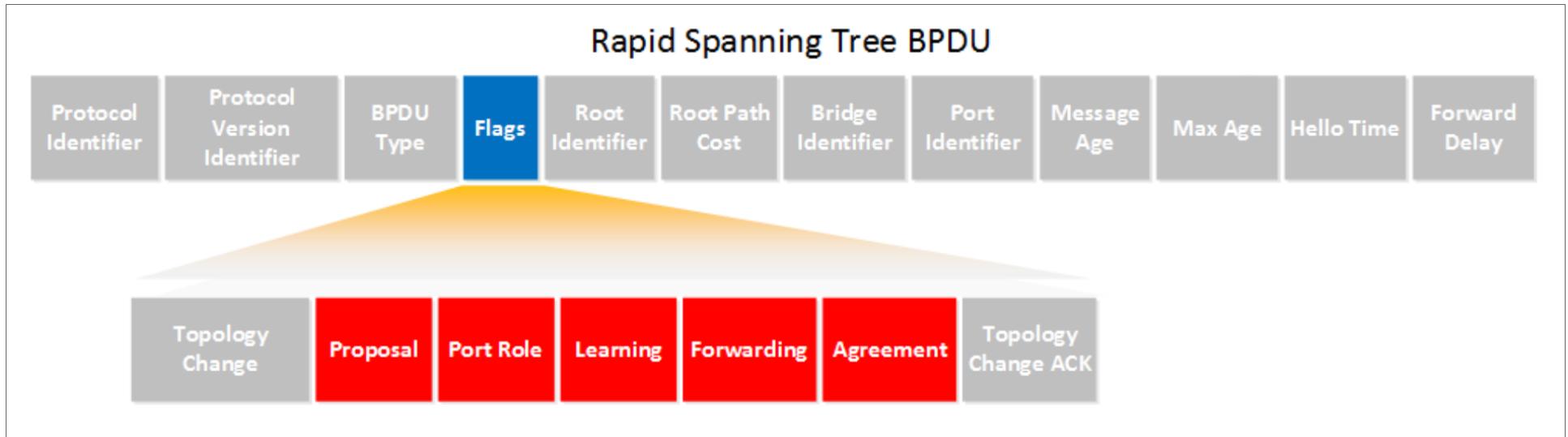
- Quand un BPDU est émis sur un port, les flags présentent les informations importantes pour la convergence de cet algorithme



- Le **Port Role** peut prendre 4 valeurs :
 - Unknown
 - Root port.
 - Designated port.
 - Alternate / Backup port.



En RSTP : convergence grâce aux flags



- Les flags **Proposal + Agreement**

- Négociation de l'état des ports (DP / RP) au moment de la construction de l'arbre.
- Un switch A qui rejoint le réseau (ou active STP sur un lien) envoie son BPDU avec **Proposal = 1** pour possiblement devenir la racine.
- Le switch B, qui le reçoit, lui renverra un BPDU avec **Agreement = 1** s'il le choisit comme racine.

<https://www.cisco.com/c/en/us/support/docs/lan-switching/spanning-tree-protocol/24062-146.html>

Une fois que la topologie active est connue

- En RSTP (attention, c'est différent avec STP) :
 - Chaque switch émet un BPDU toutes les 2 secondes sur son Root Port



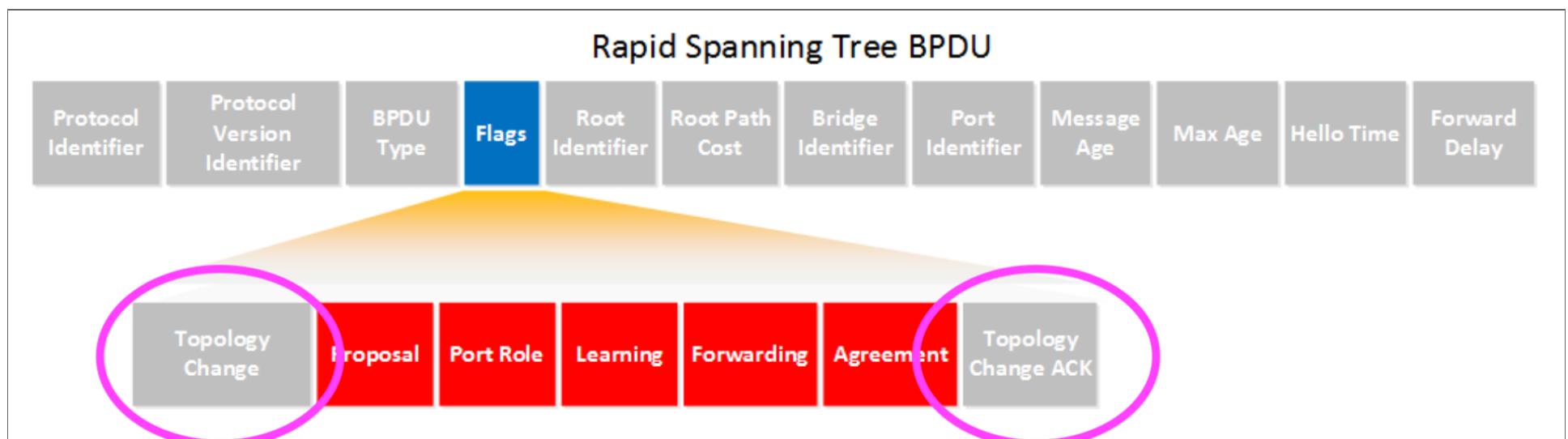
Changement de topologie

Qu'est-ce qu'un changement de topologie pour vous ?



Changement de topologie

Il existe, dans les flags de la BPDU, deux bits qui permettent d'indiquer que l'on a un **Topology Change Notification**



<https://networklessons.com/spanning-tree/rapid-spanning-tree-rstp>

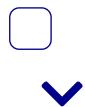
- Note : En STP, TC et TCack = 2 seuls bits utilisés.



Changement de topologie

Un switch qui détecte un changement de topologie émet un BPDU sur son RP avec
Topology Change = 1

Ce message est relayé à la racine via les RP des switch intermédiaires.



Changement de topologie

- Une fois que la racine l'a reçu, elle va envoyer des BPDU de configuration, avec le flag TC à 1, ce qui permet :
 - de reconstruire la topologie active,
 - de réduire temporairement l'age max de la table de commutation de 300s. à 15s.
- Topology Change Acknowledgement :
 - A chaque réception d'un BPDU avec un flag TC = 1, un switch répond avec un BPDU avec TCAck = 1 pour acquitter le message.



Chemins alternatifs avec RSTP

Une des principales différences entre RSTP et STP est l'identification du rôle du port dans les flags du BPDU.

Quand un des liens est en **half-duplex**

- Le switch sélectionne le port DP
- Et un second port vers ce réseau, le port alternatif.

En cas de défaillance du port DP, c'est ce port alternatif qui devient DP. Cela évite de reconstruire tout l'arbre couvrant.



Note sur le cas où les ponts interconnectent des réseaux diffusants (half duplex)

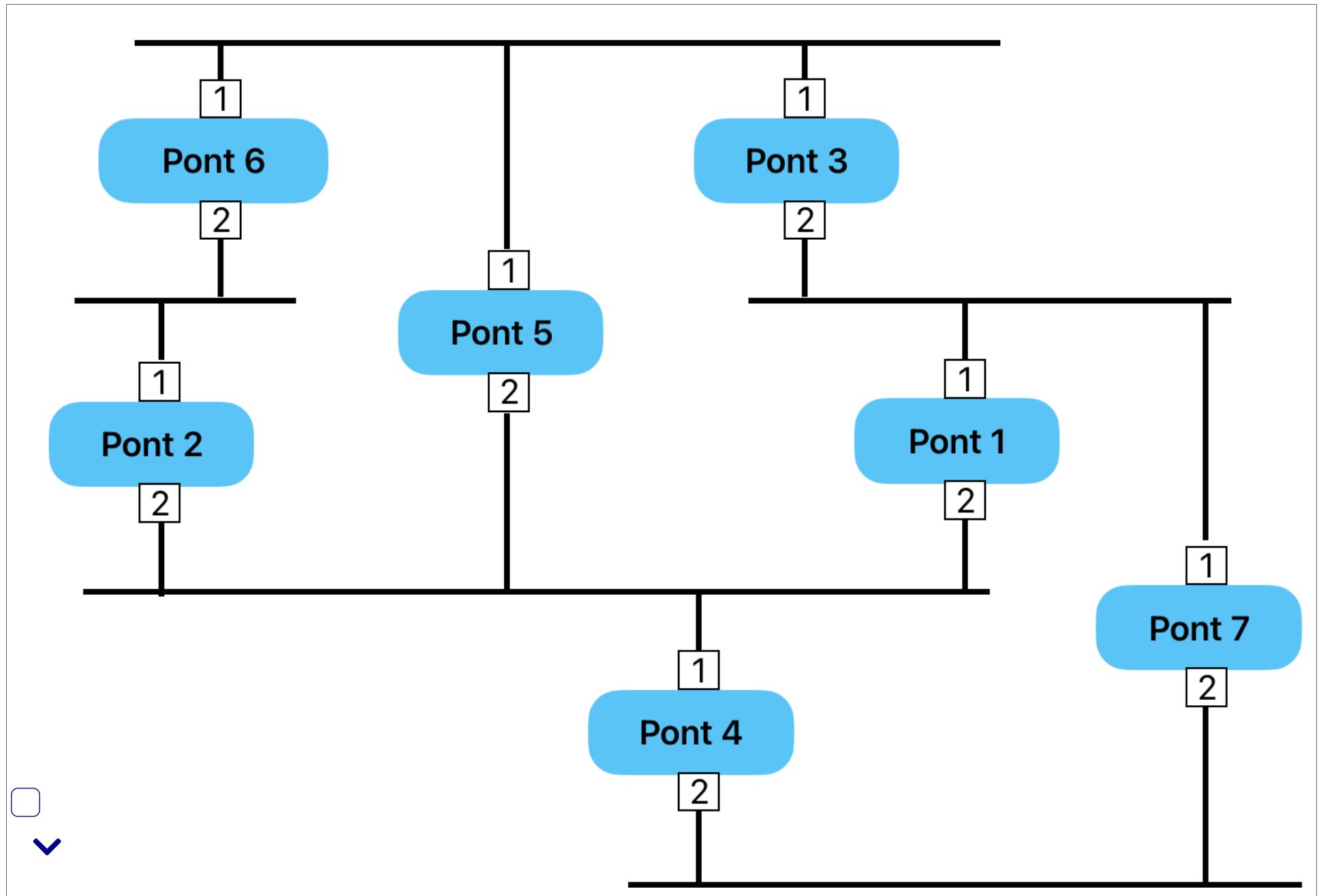


PORTS DÉSIGNÉS

Les topologies actives dans le cas où on a un réseau diffusant (segment Ethernet half duplex) derrière un port ou plusieurs ports d'un ou de plusieurs switch utilisent une définition du Port Désigné plus large que celle donnée dans ce cours.



exemple. D'autres stations sont possiblement connectées sur ce segment.



Dans ce schéma il y a 5 réseaux partagés.

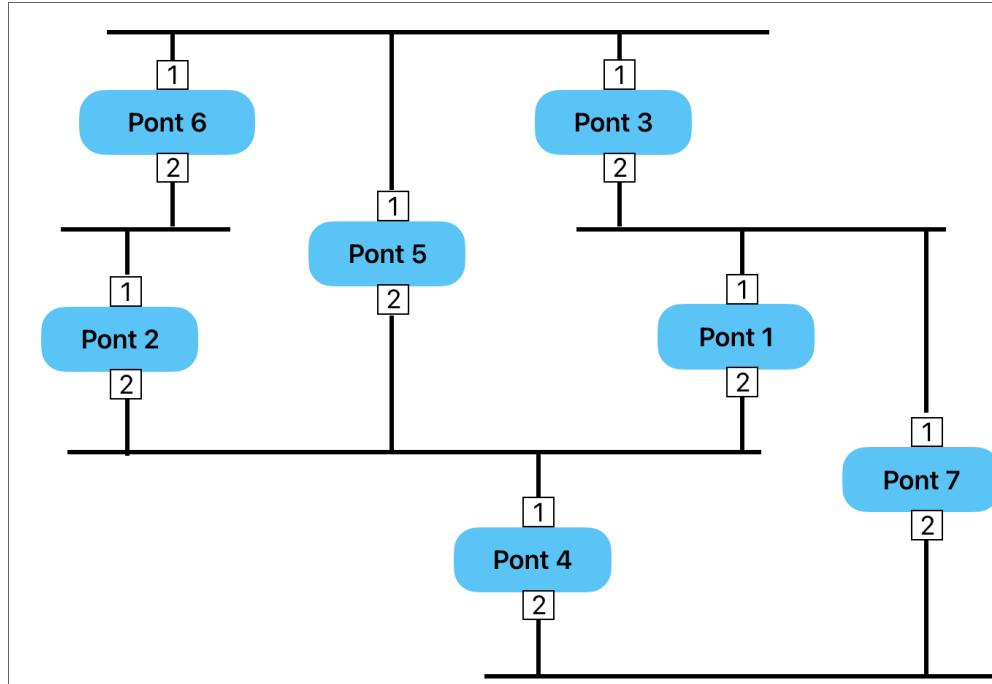
Ce qui ne change pas par rapport au cas full duplex :

- Le pont 1 est la racine (plus petit identifiant).
- Les ports racine sont choisis par le switch pour arriver à la racine par le plus court chemin (identique au cas full duplex)

Ce qui change :

Chaque réseau possède un port désigné unique qui lui permet d'atteindre la racine.





Dans cette figure, le segment du haut possède 3 ports pour atteindre la racine 1 : le port 1 du pont 6, du pont 3 et du pont 5. **Il faut en choisir un comme DP.**

Le port choisi comme DP appartient au pont de plus petit identifiant (ici le pont 3). Si ce pont est relié par deux ports au même segment, on utilise le port de plus petit identifiant.



[CM5]

Les VLANs

Virtual Local Area Networks



Rappel

Dans un réseau Ethernet commuté, **le domaine de diffusion est l'ensemble des ports du réseau quand STP n'est pas activé.**

Ce domaine de diffusion est utilisé pour :

- L'envoi d'une trame en broadcast (@MAC destination FF-FF-FF-FF-FF-FF).
- L'envoi d'une trame de données vers une destination d'adresse inconnue.

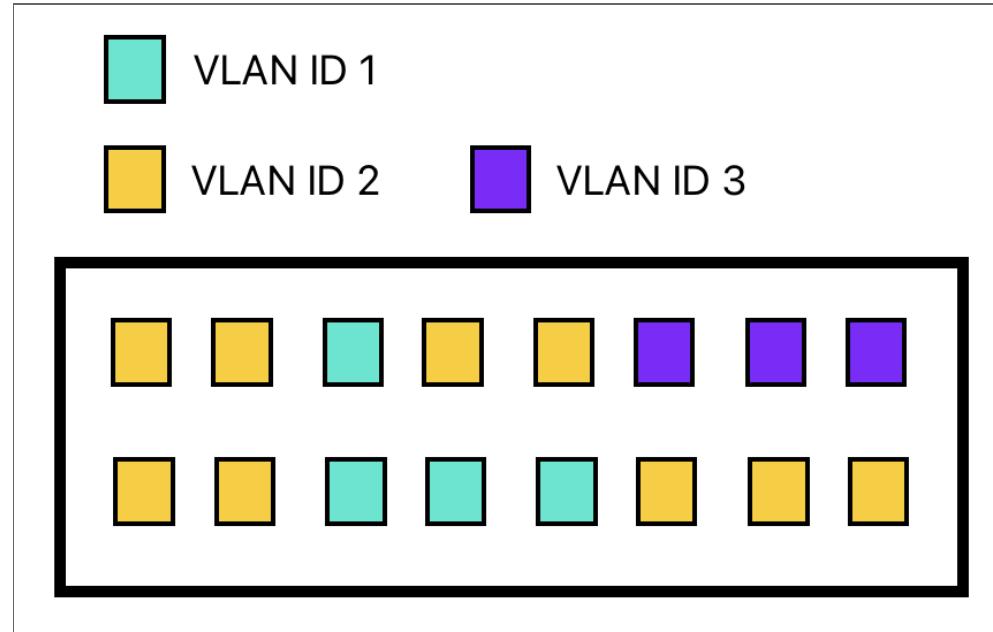


Un switch permet de restreindre le domaine de diffusion par la création d'un réseau local virtuel en associant à un sous-ensemble de ports d'un switch un identifiant

VLAN ID ou VID

Dans cette première définition, **un VID regroupe des ports** d'un même switch.





Dans ce switch on a 4 ports dans le VID 2, trois dans le VID 3 et le reste dans le VID 1.



Le mécanisme des VLANs permet, sur une même topologie physique, de définir des réseaux virtuels.

Ces réseaux virtuels suivent des règles qui restreignent le domaine de diffusion et de relayage des trames.



Est-ce que l'apprentissage de la table de commutation marche toujours ?



Revenons à la trame MAC

Il y a un **tag optionnel sur 4 octets**, après l'adresse MAC source : **tag 802.1Q ou VLAN TCI**.

| 802.3 Ethernet packet and frame structure | | | | | | | | | |
|---|-------------------------|-----------------------------|--------------------|------------|-----------------------|--|------------------------|-----------------------------------|-----------------------|
| Layer | Preamble | Start frame delimiter (SFD) | MAC destination | MAC source | 802.1Q tag (optional) | Ethertype (Ethernet II) or length (IEEE 802.3) | Payload | Frame check sequence (32-bit CRC) | Interpacket gap (IPG) |
| Length (octets) | 7 | 1 | 6 | 6 | (4) | 2 | 42–1500 ^[c] | 4 | 12 |
| Layer 2 Ethernet frame | (not part of the frame) | | ← 64–1522 octets → | | | | | | |
| Layer 1 Ethernet packet & IPG | ← 72–1530 octets → | | | | | | | | ← 12 octets → |

S'il est défini, il constitue, avec les 2 octets du champ Etherype/Length, le champ  MAC Control.



VLAN Tag Control Information

Si les deux octets qui suivent l'adresse MAC source valent 0x8100, le tag 802.Q existe et a la forme suivante :

9.6 VLAN Tag Control Information (TCI)

The VLAN TCI field (Figure 9-1) is two octets in length and encodes the `vlan_identifier`, `drop_eligible`, and priority parameters of the corresponding EISS M_UNITDATA.request as unsigned binary numbers.



Figure 9-1—VLAN TCI format

The VID is encoded in a 12-bit field. A VLAN Bridge may not support the full range of VID values but shall support the use of all VID values in the range 0 through a maximum N, less than or equal to 4094 and specified for that implementation. Table 9-2 identifies VID values that have specific meanings or uses.



VLAN Tag Control Information

La VLAN TCI permet de marquer chaque trame avec :

- Priority Code Point sur 3 bits pour donner un niveau de priorité
- DEI = Drop Eligible Indicator pour dire si on peut jeter la trame si besoin
- **VLAN ID sur 12 bits**



Valeur réservées de VLAN ID

PVID = Port VID

Table 9-2—Reserved VID values

| VID value (hexadecimal) | Meaning/Use |
|----------------------------|---|
| 0 | The null VID. Indicates that the tag header contains only priority information; no VID is present in the frame. This VID value shall not be configured as a PVID or a member of a VID Set, or configured in any FDB entry, or used in any Management operation. |
| 1 | The default PVID value used for classifying frames on ingress through a Bridge Port. The PVID value of a Port can be changed by management. |
| 2 | The default SR_PVID value used for SRP [35.2.1.4 item i)] Stream related traffic. The SR_PVID value of a Port can be changed by management. |
| FFF | Reserved for implementation use. This VID value shall not be configured as a PVID or a member of a VID Set, or transmitted in a tag header. This VID value may be used to indicate a wildcard match for the VID in management operations or FDB entries. |

Le VID 1 est la valeur par défaut d'un port qui n'a pas été configuré par l'administrateur à une autre valeur.

Qui renseigne le VID dans la trame?

La station ou le switch ?



Qui renseigne le VID dans la trame ?

Les deux sont possibles mais cela dépend de la configuration du port d'entrée du trafic.

- Si le port est en mode **Untagged**, c'est **le switch qui tag** les trames avec un VID avant de les relayer. Dans ce cas, le port est associé à un **UN SEUL VID**.
- Si le port est en mode **Tagged**, c'est à la station de renseigner le VID avant de l'envoyer. Ce port peut être associé à plusieurs VLAN et dans ce cas, si la station envoie une trame avec un VID non autorisé sur le port, la trame est jetée..



Peut-on associer autre chose que des ports à un VID ?



Peut-on associer autre chose que des ports à un VID ?

Oui, en mode Tagged, la station peut choisir de regrouper dans un VLAN :

- Des adresses MAC,
- Des applications (la VoIP, un flux vidéo...)



Avantages des VLANs

On va **ségréguer le trafic**, c'est-à-dire regrouper des flux de trames qui ont des points communs :

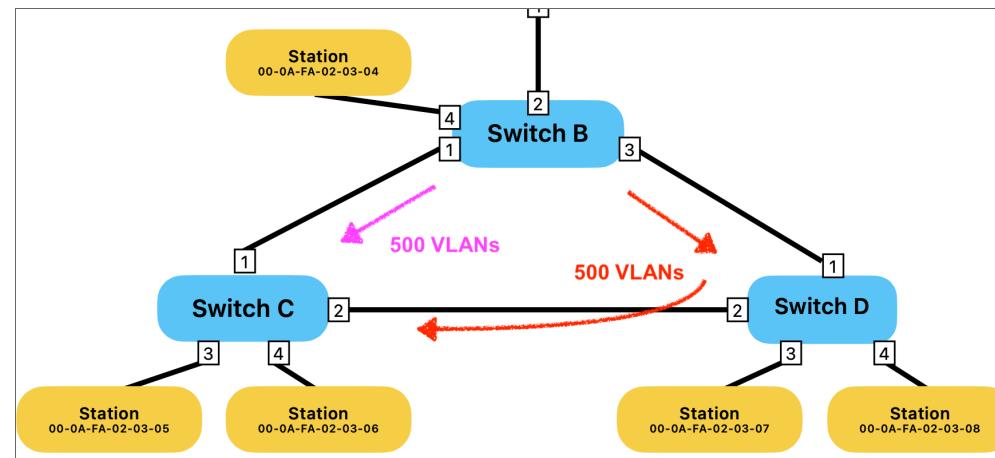
- Ils viennent des mêmes applications (VoIP, Streaming,..)
- Ils regroupent des stations qui appartiennent à une même entité (la comptabilité, l'usine, le service Marketing ...)



Avantages des VLANs

Cela permet d'introduire :

- De la qualité de service
- Des règles de sécurité plus fines
- De l'équilibrage de charge



Est-ce que l'apprentissage de la table de commutation marche toujours ?
(revenons à nos moutons ...)

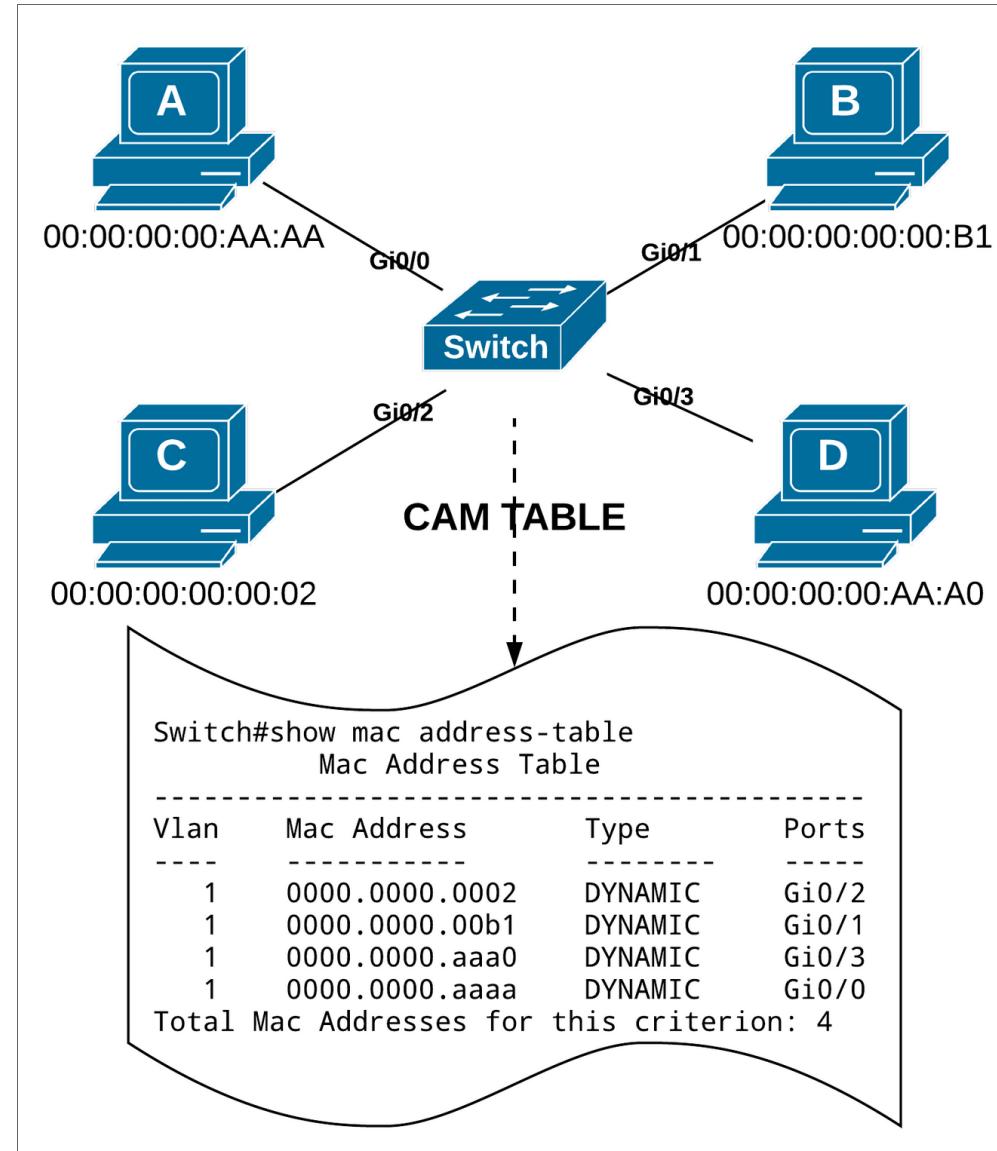


Est-ce que l'apprentissage de la table de commutation marche toujours ?

Oui, mais il faut aussi enregistrer la VLAN ID dans la table de commutation.



Dans ce schéma, on observe la colonne VLAN dans la sortie de la commande d'affichage de la table de commutation du switch.



Réception d'une trame dans un VLAN

A la réception d'une trame, le switch détermine son VID et parmi les entrées de la table de commutation associées à ce VID, il identifie le port de l'adresse destination de la trame.

Si pas d'entrée dans la table associé à (VID, @MAC dest), alors le message est diffusé à tous les ports associés au VID du switch.

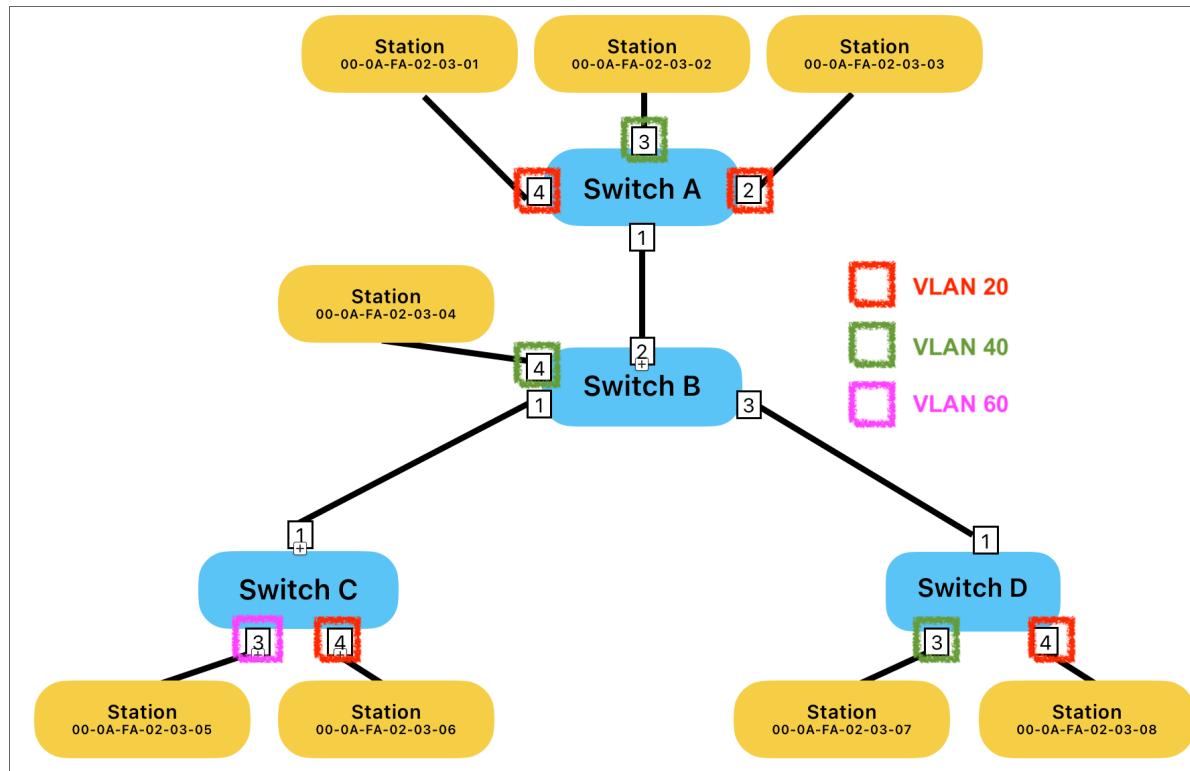


Identification du VID d'une trame en entrée

- Si la trame présente champ VLAN TCI dans lequel se trouve un VID, le switch utilise ce VID s'il est autorisé sur ce port. Le port est en mode **Tagged**.
- Si la trame arrive d'une station sans VID, le switch lui associe le VID du port d'entrée. Le port est en mode **Untagged**.
- Si aucun VID n'a été positionné sur ce port d'entrée, le VID 1 par défaut est utilisé.



Et si on a plusieurs switchs ...



Et si on a plusieurs switchs ...

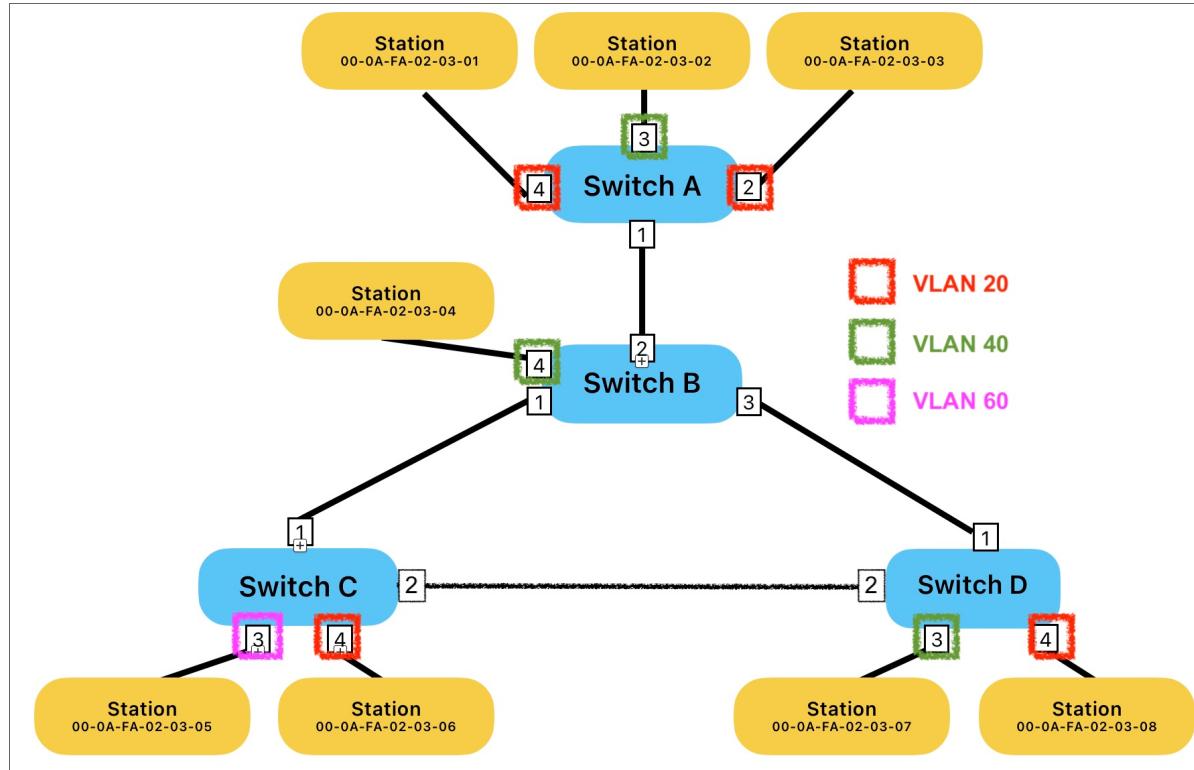
Ports Trunk ou Jonction ou MultiVLAN

Il faut configurer les liens entre les switchs en **trunk** pour qu'ils soient capables d'étendre les VLAN d'un switch à un autre.

Le switch D sera capable, sur son port configuré en trunk, de multiplexer sur un même lien des trames avec un VID 40 et un VID 20.



Quid des boucles si on a des VLAN ?



Spanning Tree et VLAN

On peut avoir une boucle dans un VLAN (VLAN 20)

:-)

Pour éviter cela, on peut :

1. Enraciner un **spanning tree par VLAN**
2. Utiliser **MSTP** : Multiple Spanning Tree Protocol



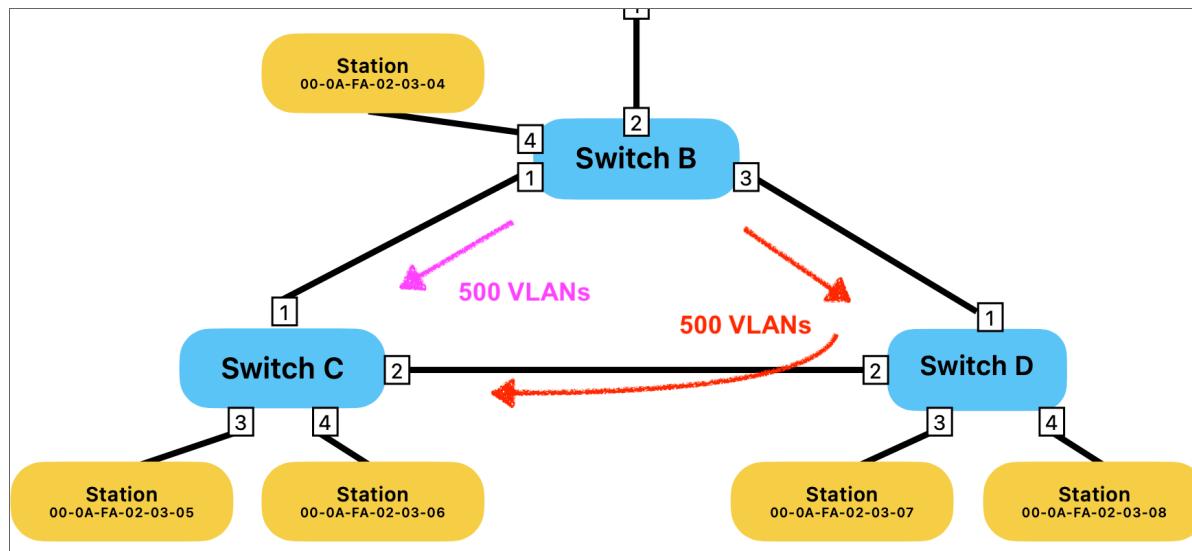
Spanning Tree et VLAN

Pourquoi utiliser le spanning tree créé par STP ou RTSP pour tous les VLAN n'est pas une très bonne idée ?



Spanning Tree et VLAN

Le Spanning Tree supprimera le lien entre C et D.



On ne pourra plus faire d'équilibrage de charge : les 1000 VLAN passeront par le lien entre B et C.



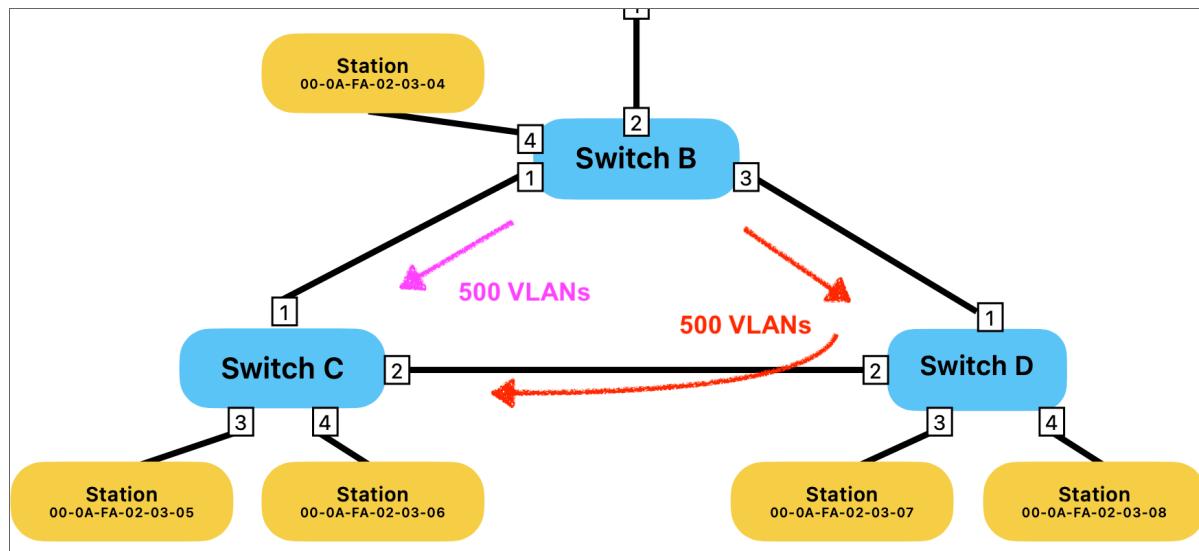
Spanning Tree et VLAN

Et si dans ce cas on génère un spanning tree par VLAN ? Est-ce une bonne idée ?



Spanning Tree et VLAN

Si on a un spanning tree par VLAN : on a 1000 Spanning Trees !!



Une forêt dense, qui émet plein de messages toutes les 2 secondes :-(
Et pourtant, on n'a besoin que de 2 topologies actives ici, l'une pour les 500 VLANs de gauche, et une pour les 500 VLANs de droite.



Multiple Spanning Tree Protocol

Ce protocole définit le nombre minimal de topologies actives capable de supporter l'ensemble des VLANs :

Par exemple, si on a 10 VIDs, MSTP définit

- Topologie active A qui convient à VID 1, 2, 4, 7, 9
- Topologie active B qui convient à VID 3, 5, 6, 8, 10.



La Qualité de Service

ou QoS

Un grand mot mais qui tient sur 3 bits.



La Qualité de Service

PCP - Priority Code Point

9.6 VLAN Tag Control Information (TCI)

The VLAN TCI field (Figure 9-1) is two octets in length and encodes the `vlan_identifier`, `drop_eligible`, and priority parameters of the corresponding EISS M_UNITDATA.request as unsigned binary numbers.



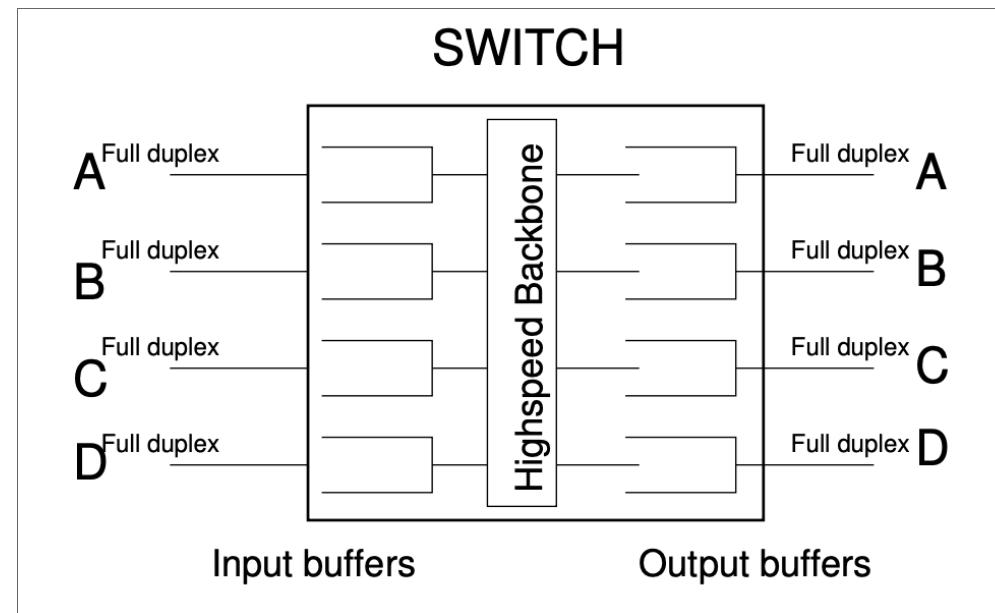
Figure 9-1—VLAN TCI format

The VID is encoded in a 12-bit field. A VLAN Bridge may not support the full range of VID values but shall support the use of all VID values in the range 0 through a maximum N, less than or equal to 4094 and specified for that implementation. Table 9-2 identifies VID values that have specific meanings or uses.



La Qualité de Service

Sur 3 bits, on peut coder 8 niveaux de priorité.



Dans les files d'attente de sortie du switch, on peut utiliser ces niveaux de priorité pour permettre aux trames d'un VLAN de passer avant les autres par exemple.



Priorité et types de trafic

Plus la priorité est haute, plus les flux qui y sont associés sont prioritaires.
A chaque priorité, on associe habituellement un **type de trafic** pour aider au déploiement. IEEE802.1Q définit les types (ou classes) de trafic et y associe des priorités.

Table I-2—Traffic type acronyms

| Priority | Acronym | Traffic type |
|-------------|---------|--------------------------------------|
| 1 | BK | Background |
| 0 (Default) | BE | Best Effort |
| 2 | EE | Excellent Effort |
| 3 | CA | Critical Applications |
| 4 | VI | “Video,” < 100 ms latency and jitter |
| 5 | VO | “Voice,” < 10 ms latency and jitter |
| 6 | IC | Internetwork Control |
| 7 | NC | Network Control |



La gestion d'une trame en 802.1Q

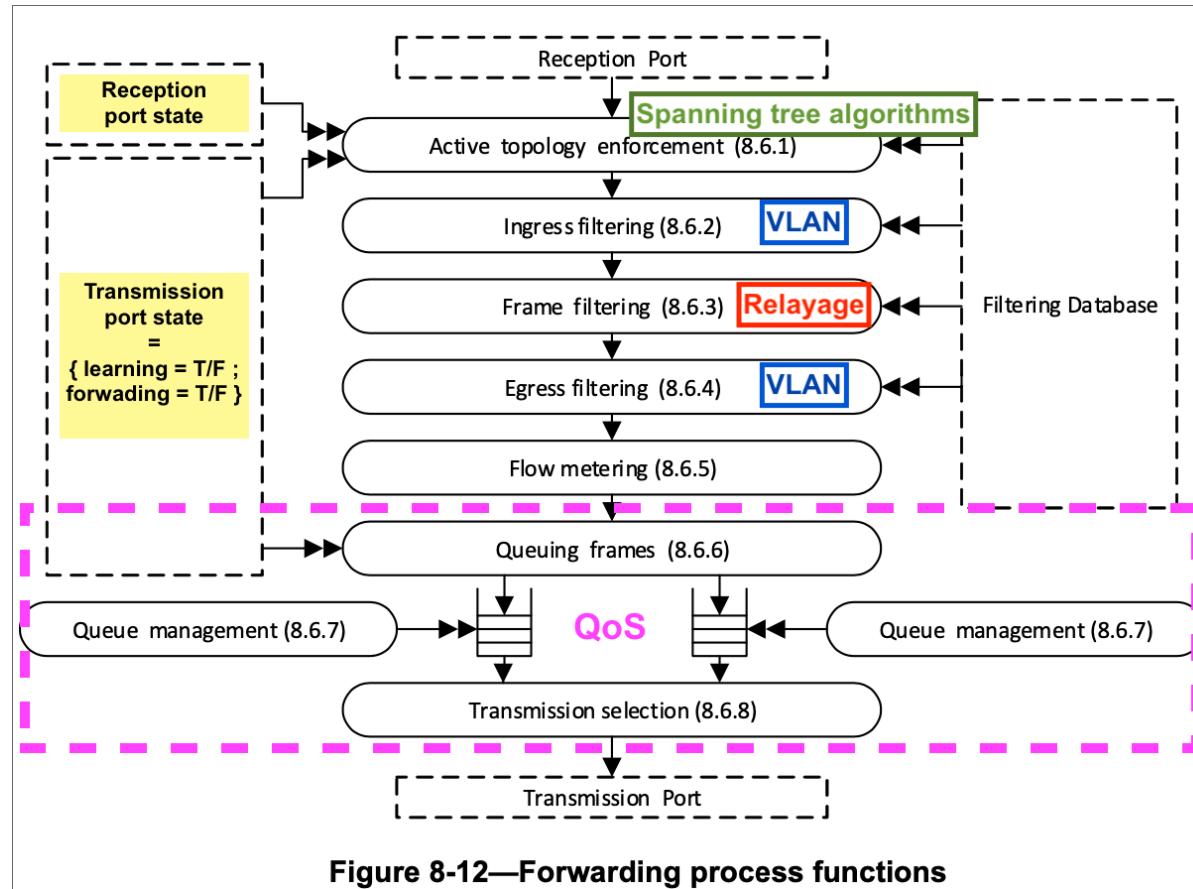


Figure 8-12—Forwarding process functions

- Les ports en entrée et en sortie sont décrits par 2 états :
 - Learning (T/F) et Forwarding (T/F).

Filtering

C'est la fonction qui implémente les VLAN dans la norme IEEE802.1Q.

- Ingress filtering : Si cette fonctionnalité est active, une trame qui entre avec un VLAN ID qui ne fait pas partie du *member set* associé au port d'entrée est détruite
- **Member set** : la liste des VLAN ID autorisés sur un port
- Egress filtering : La trame membre d'un VLAN ID ne peut être diffusé que vers un port qui présente ce même VLAN ID dans son member set.
- Frame filtering : le relayage d'une trame en associant VLAN ID et adresse de destination via l'information présente dans la Filtering Database (la table de commutation).



Flow metering (and classification)

C'est la fonction qui permet de

- classifier un ensemble de trames selon des critères complexes qui peuvent s'écrire à partir des propriétés suivantes :
 - Adresses MAC source et destination
 - VLAN ID
 - Priorité de la trame
- Cette classification permettra au *flow meter* de positionner un tag DEI de façon à pouvoir supprimer la trame les files d'attente en sortie sont vides.



Un petit questionnaire sur les VLAN

<https://app.wooclap.com/RLCM6>



[CM6]

Mécanismes de qualité de service et d'économie d'énergie



Les files d'attente d'un port

On peut avoir plusieurs files d'attente par port sur un switch. Chaque file a une discipline **First In First Out** généralement.

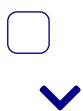
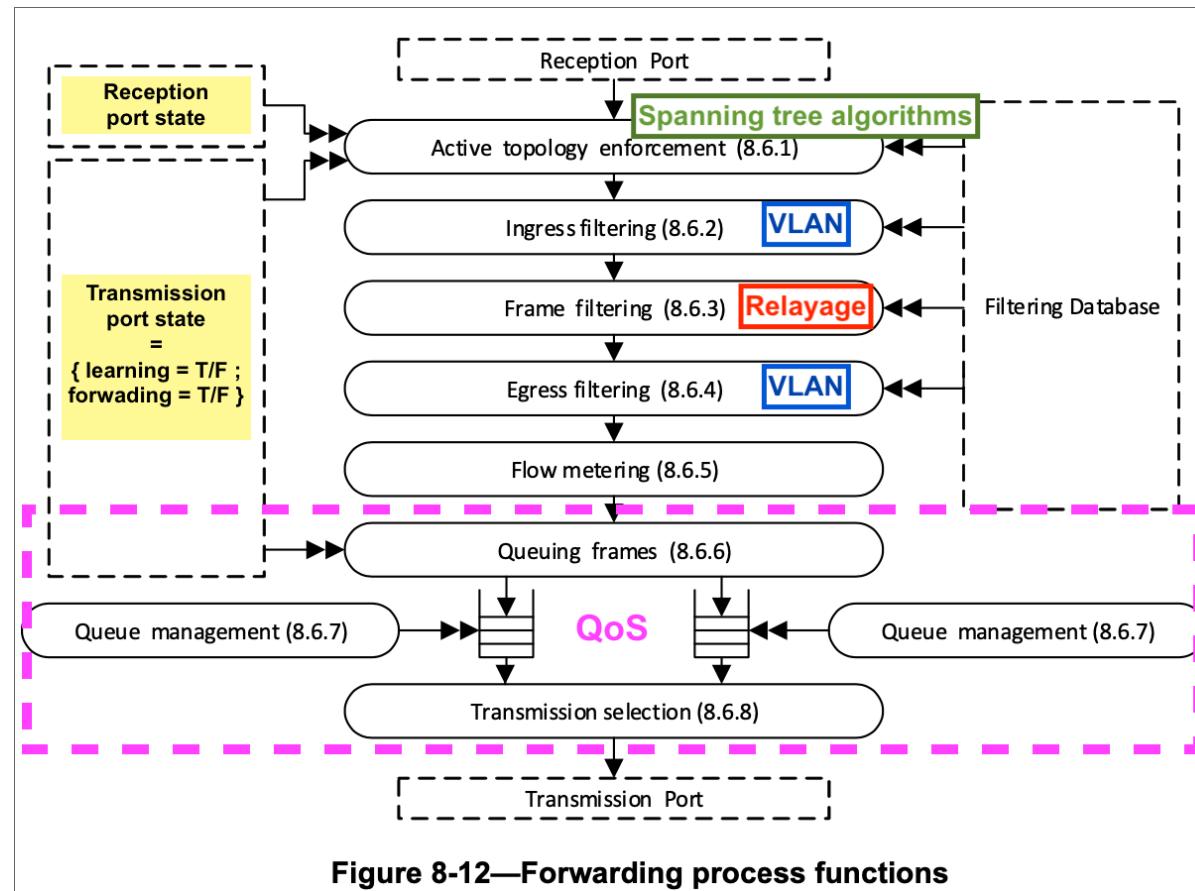
A chaque file d'attente on peut associer un(e) ou plusieurs types (ou classes) de trafic.

Si on a moins de classes de trafic que de priorités, on fait un **mapping entre priorités et classes de trafic**. Par exemple, si j'ai 2 classes de traffic, j'associe les valeurs de priorité entre 0 et 3 du PCP à la classe de trafic 0, et les valeurs entre 4 et 7 à la classe de trafic 1.



Queue management

- On extrait la trame de sa file d'attente une fois qu'elle est émise (même si on n'a pas reçu d'acquittement) -> Comportement FIFO
- On extrait la trame si le temps d'attente dans la file dépasse la durée de transit maximale dans le switch autorisée (~1s.) -> Comportement pas FIFO.



Transmission selection

C'est le serveur de nos files d'attente qui utilise un algorithme pour déterminer l'ordre d'envoi des trames sur le port de sortie en allant les lire dans les différentes files d'attente.



Transmission selection

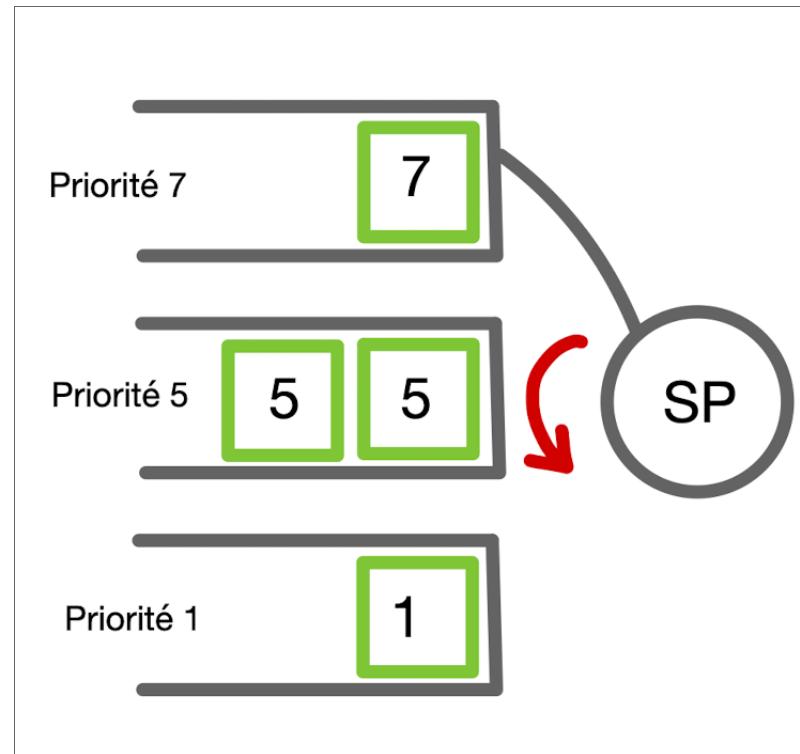
C'est ici qu'on peut introduire des mécanismes de QoS avancés.

Voici des algorithmes normalisés dans 802.1Q :

- Strict priority (SP),
- Credit-based shaper (CBS),
- Enhanced Transmission Selection (ETS) : Il est possible de déterminer son propre algorithme.



Strict priority



On émet dans l'ordre 7, puis 5, puis 5 et enfin 1.



Strict priority

:-)

Assure un accès privilégié aux priorités élevées

:-(|

Risque de famine des flux peu prioritaires si le trafic prioritaire est important.



Credit-based shaper

On associe N files d'attente au Credit-based shaper.

A chaque file d'attente, le CBS associe les paramètres suivants pour décider si une trame en attente est émise ou non :

- Le débit binaire du port de transmission
- Le **idle slope** : la vitesse d'augmentation du Credit associé à la file d'attente quand sa trame n'est pas en cours de transmission (`transmit = false`) et qu'une trame d'une autre file d'attente est en cours de transmission sur le support. Sa valeur max est celle du débit binaire du lien.
- Le **credit** d'émissions (en bits) courant. Si le crédit est positif
- Le **sendSlope** qui est la vitesse de consommation du credit quand la file transmet une trame. On a :

$$sendSlope = idleSlope - debitBinaire$$



Credit-based shaper

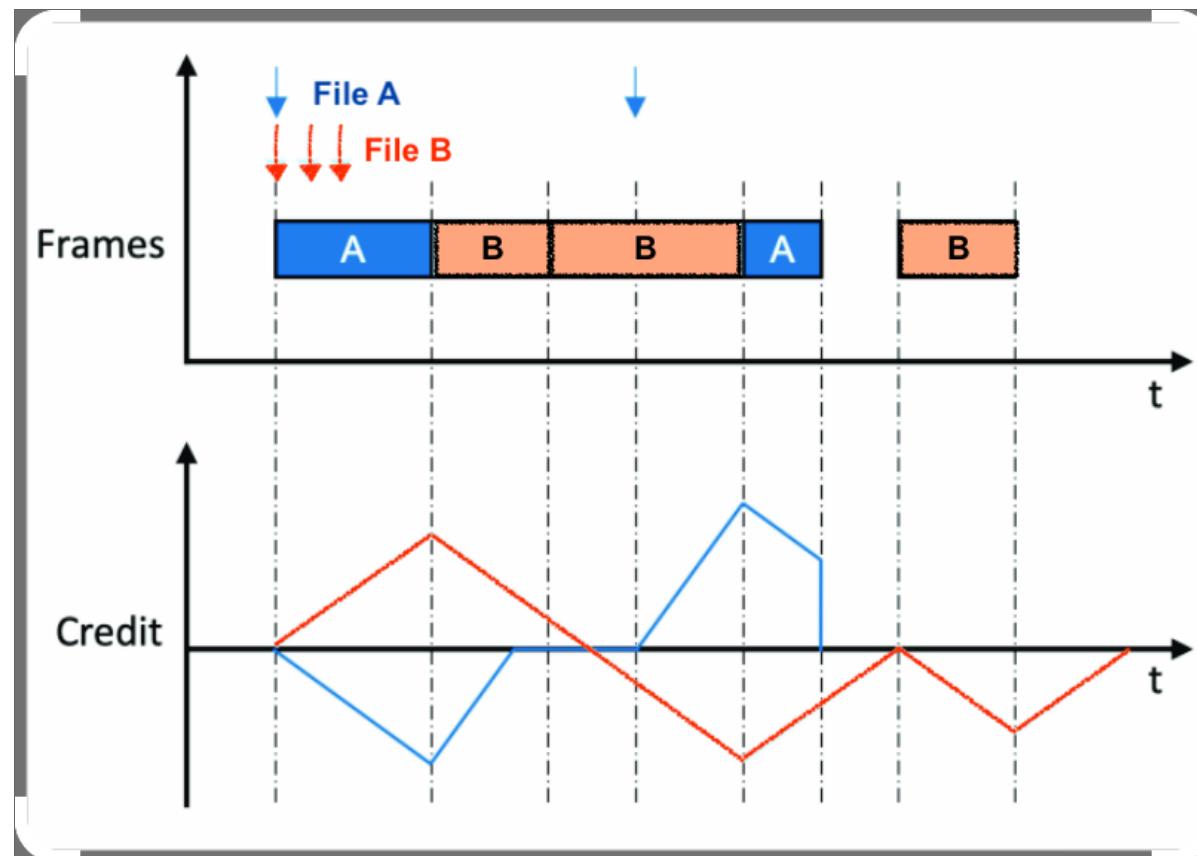
LE CRÉDIT

- est initialisé à 0.
- remis à 0 quand il n'y a plus de trame en attente dans la file.
- décroît à la vitesse sendSlope quand une trame est en cours d'émission
- croît avec la pente idleSlope quand une trame d'une autre file est en cours d'émission.

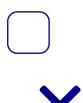


Credit-based shaper

Exemple avec 2 files d'attente



La file A est choisie en premier car plus prioritaire



Credit-based shaper

- On peut ajuster la valeur de *idleSlope* par file d'attente pour modifier la priorité (relative) des files.
- Permet d'éviter la famine des priorités faibles de Static priority
- En pratique, on utilise 2 files d'attente avec une discipline CBS.



Enhanced Transmission Selection

Avec ETS, on peut paramétriser son propre algorithme de sélection de transmission de façon à répartir le débit entre les files d'attente.

Pour cela on doit :

- Connaitre le nombre de classes de trafic (min 3 et max 8) et donc de files d'attentes,
- Pour chaque file, on doit associer un pourcentage de la bande passante



Enhanced Transmission Selection

Différents algorithmes existent pour cela :

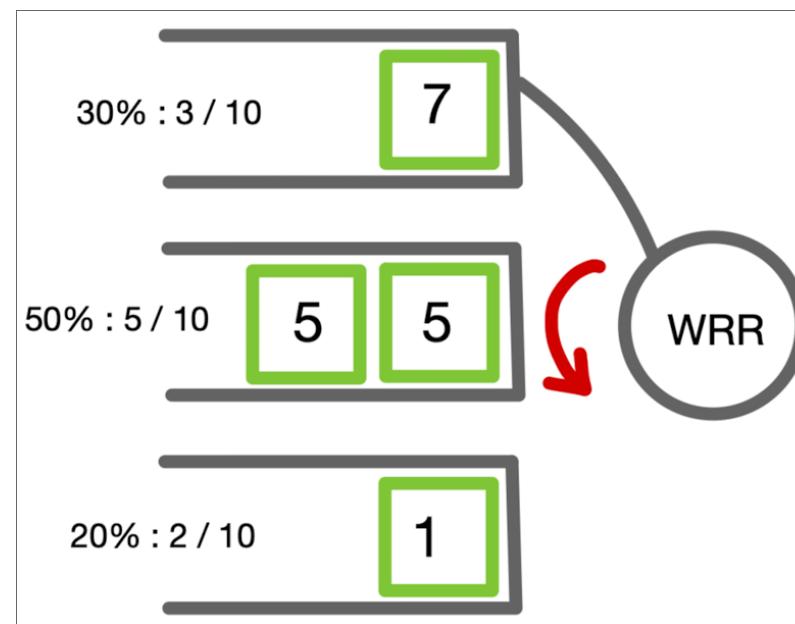
- Weighted round robin,
- Weighted fair queuing,
- Deficit round robin.



Weighted Round Robin

Sur 10 opportunités de transmission :

- 3 sont données à la file à 30% de bande passante,
- 5 à la file à 50%
- 2 à la file à 20%



Weighted Round Robin

L'ordonnateur WRR va consulter la file 1, et tant qu'il y a des trames à transmettre, il va en transmettre au max 3 d'affilée. Puis il passe à la file suivante.

- avantage : répartit les émissions selon les proportions attendues quand toutes les files sont pleines
- désavantage : si le trafic n'est pas équilibré entre les files, on observe dans les faits un déséquilibre dans les émission des classes de trafic.
- désavantage : on génère des trains de trames de même classe, et donc une attente plus grande entre certaines trames (gigue...)



Interleaved Weighted Round Robin

CWRR = Classical WRR

IWRR = Interleaved WRR

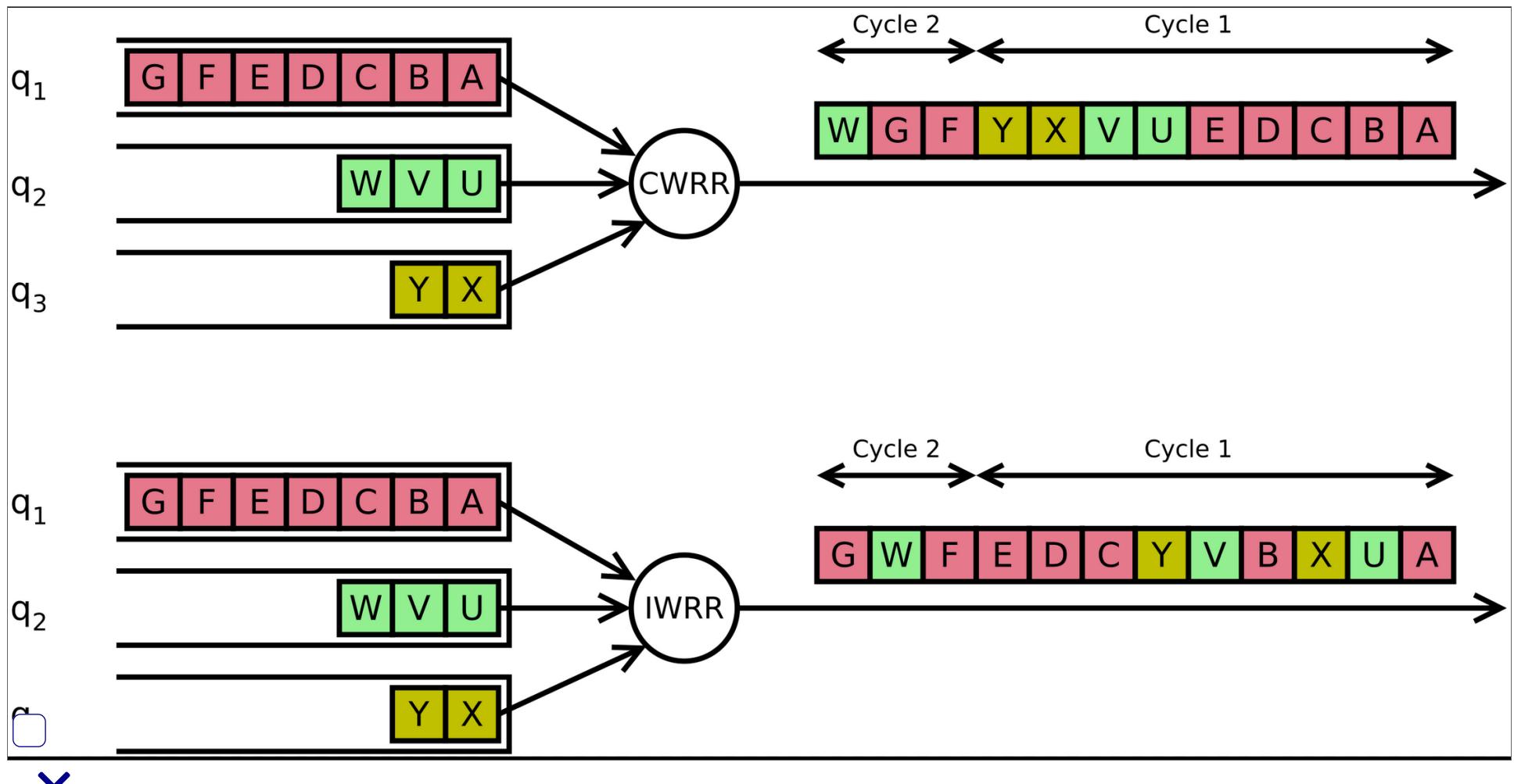


Figure by Marc Boyer, Onera

Interleaved Weighted Round Robin

- Permet de réduire la gigue = l'écart-type des durées de transmissions des messages d'un même flux.



