IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

F2B002C - Introduction aux Réseaux

Christophe Lohr

Automne 2018

Objectifs

- Appréhender les concepts fondamentaux des Réseaux
  - Multiplexage, paquetisation, délais, protocoles, modélisation, ...
- Montrer comment s'appliquent ces concepts à quelques exemples de réseaux fortement utilisés en entreprise
  - Les réseaux locaux et la technologie Ethernet
    - Concepts et réalité
  - Les réseaux IP
    - Concepts, principes de fonctionnement, routage...
    - Vous devriez savoir, en fin de cours, ce qu'est un routeur, un commutateur et connaître les principes de l'architecture d'un réseau d'entreprise

Plan

- Première partie : Concepts fondamentaux des réseaux
  - Notion de protocole et modélisation
  - Les couches 1, 2 et 3 dans le contexte connecté
- Deuxième partie : les réseaux locaux
  - La standardistaion
  - La technologie Ethernet
    - Topologie et supports physique, Adressage, Notion de VLAN
- Troisième partie : Les réseaux IP
  - IP :concepts, adressage, routage
  - TCP,UDP : concepts, notion de port, contrôle de flux
  - Quelques protocoles applicatifs

- A. Tanenbaum.
  Réseaux.
  Eyrolles, 2003
- G. Pujolle.
  Les réseaux.
  Eyrolles, 2003
- P. Rolin, G. Martineau, L. Toutain, A. Leroy.
  Les réseaux.
  Hermes, 1997

- P. Toutain.
  Réseaux locaux et Internet.
  Hermes, 2003
- `http://www.iec.org/online/tutorials`
- `http://www.techfest.com/networking/index.htm`

Et bien d'autres...
(Voir D. Comer, W.R. Stevens,...)

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Première partie

Introduction aux réseaux

Les Réseaux...

Partout et pour tout...

- L'épine dorsale de l'entreprise, les synapses de ses neurones...
- Les entreprises spécialisées
  - Opérateurs de télécommunication
    - L'offre aux particuliers et aux entreprises
    - Téléphone
    - Internet
  - Les équipementiers
  - Les fournisseurs de service
    - Les applications et les services à valeur ajoutée
- Enjeux économiques majeurs

- Le Système d'Information (le SI) : le cœur de l'entreprise
  - Des bases de données
  - Le réseau permet l'accès à ces bases
    - Par l'infrastructure : câbles, matériels, machines...
    - Par les applications
      . Ne pas oublier les applications
      . Les logiciels de groupware (travail en groupe)
      . Courrier, agenda partagé, workflow (suivi de documents électroniques), ...
- L'entreprise répartie
  - Agences dispersées, il faut les relier au siège
    - Dispersion à l'échelle mondiale parfois

Réseaux : le mariage des Télécommunications et de l'Informatique

- Le monde des télécommunications
  - Histoire plus que centenaire
  - Une culture : le téléphone
  - Des principes forts, une normalisation très structurée
- Le monde de l'informatique communicante
  - Histoire récente
  - Au départ pour l'entreprise, en interne, pas de facturation
  - Des acteurs divers
    - Fournisseurs de matériels et de logiciels, Utilisateurs, Chercheurs
  - Une réactivité et innovation fortes
  - Standardisation rapide et légère
- Les deux mondes convergent

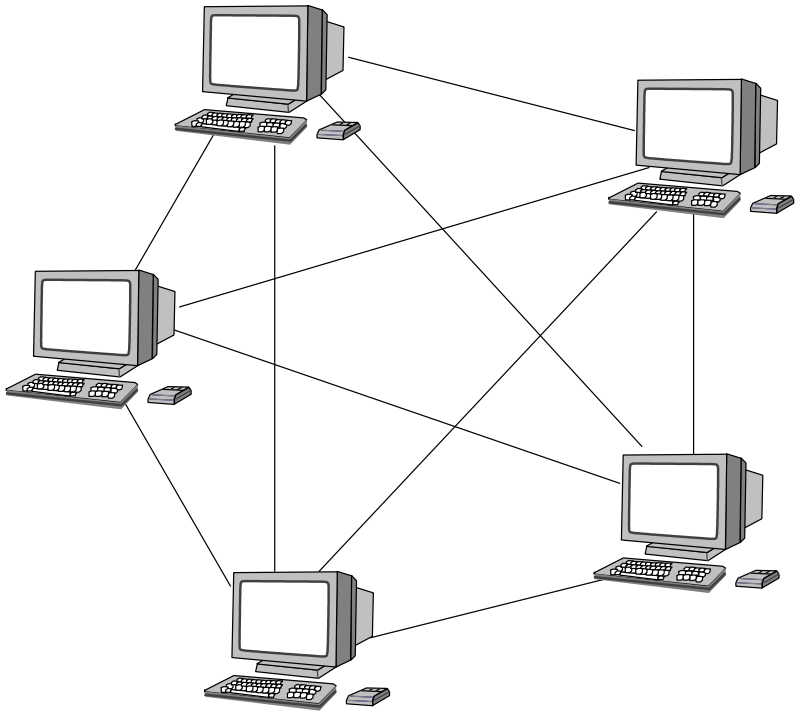
Réseaux : l'explosion du nombre de services

- Les services logiciels : ce sont les applications
  - Logiciels pour ordinateurs de bureau (portables ou non)
  - Logiciels embarqués
    - Sur les portables de type téléphone
    - Sur les portables de type PDA
    - La combinaison des appareils et des services
      . Le téléphone portable qui devient appareil photo et PDA,
      . média-center / média-renderer, domotique, M2M, etc.
- Les services matériels
  - La technologie va très vite

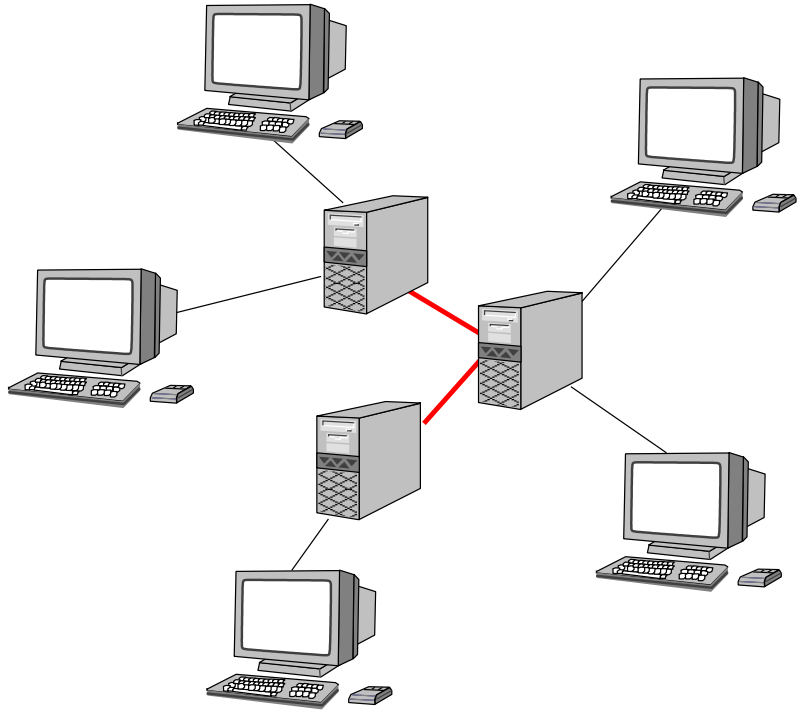
    (à peine avons nous le temps de nous habituer au modem V92 que l'ADSL est arrivé, et maintenant le FTTH pointe son nez)
  - Les technologies sans fil s'imposent rapidement

    (IEEE-802.11{abg}, Bluetooth, Zigbee)

Le maillage du réseau



Solution idéale utopique



Solution réelle
Mot clé : multiplexage

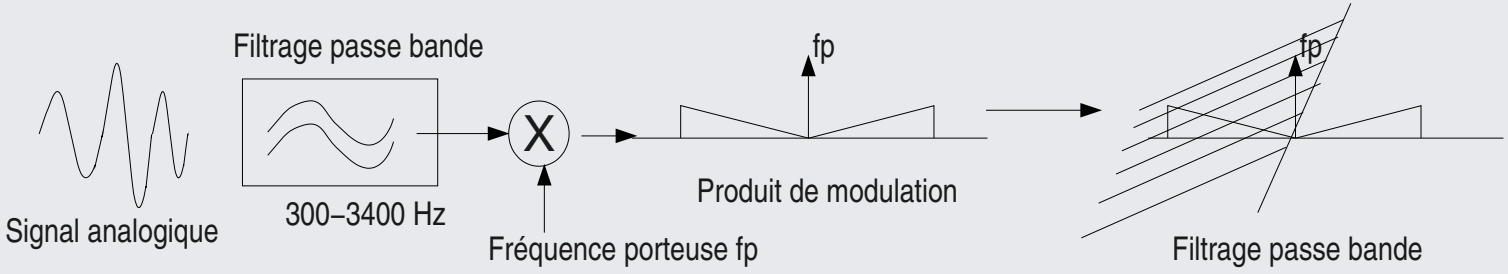
Différents types de multiplexes

- Le multiplexage en fréquence
- Le multiplexage temporel analogique
  - Par échantillonnage du signal origine
- Le multiplexage temporel numérique
  - Par échantillonnage et numérisation
- Le multiplexage statistique
  - Par acheminement sur canal commun de segments d'informations appartenant à diverses communications

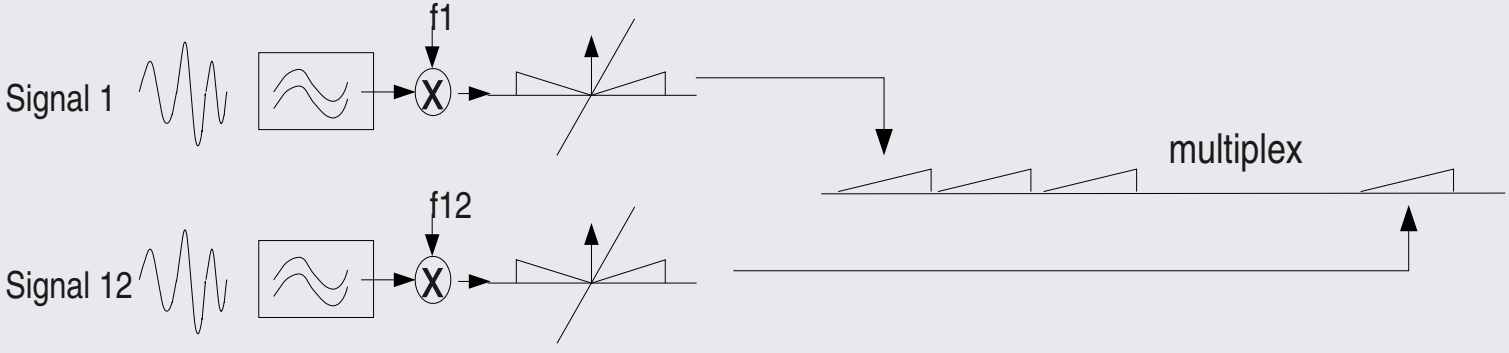
Multiplexage en fréquence

Exemple du multiplexage de canaux téléphoniques : Modulation en amplitude

Modulation d'un canal

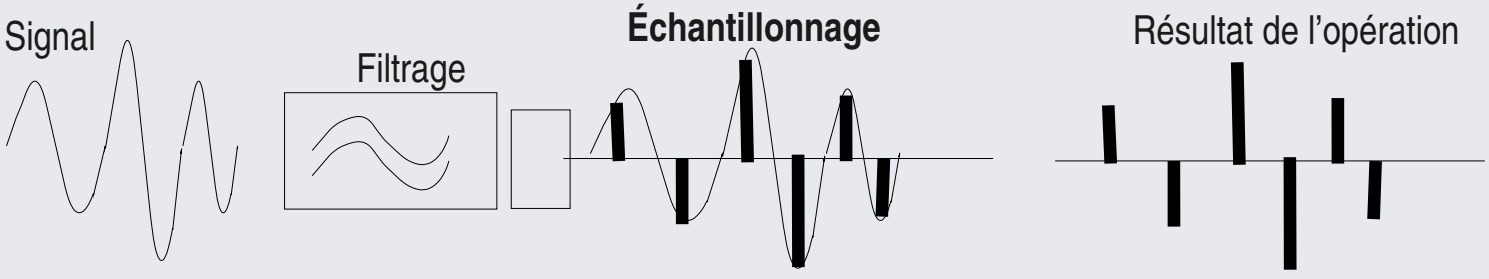


Multiplexage



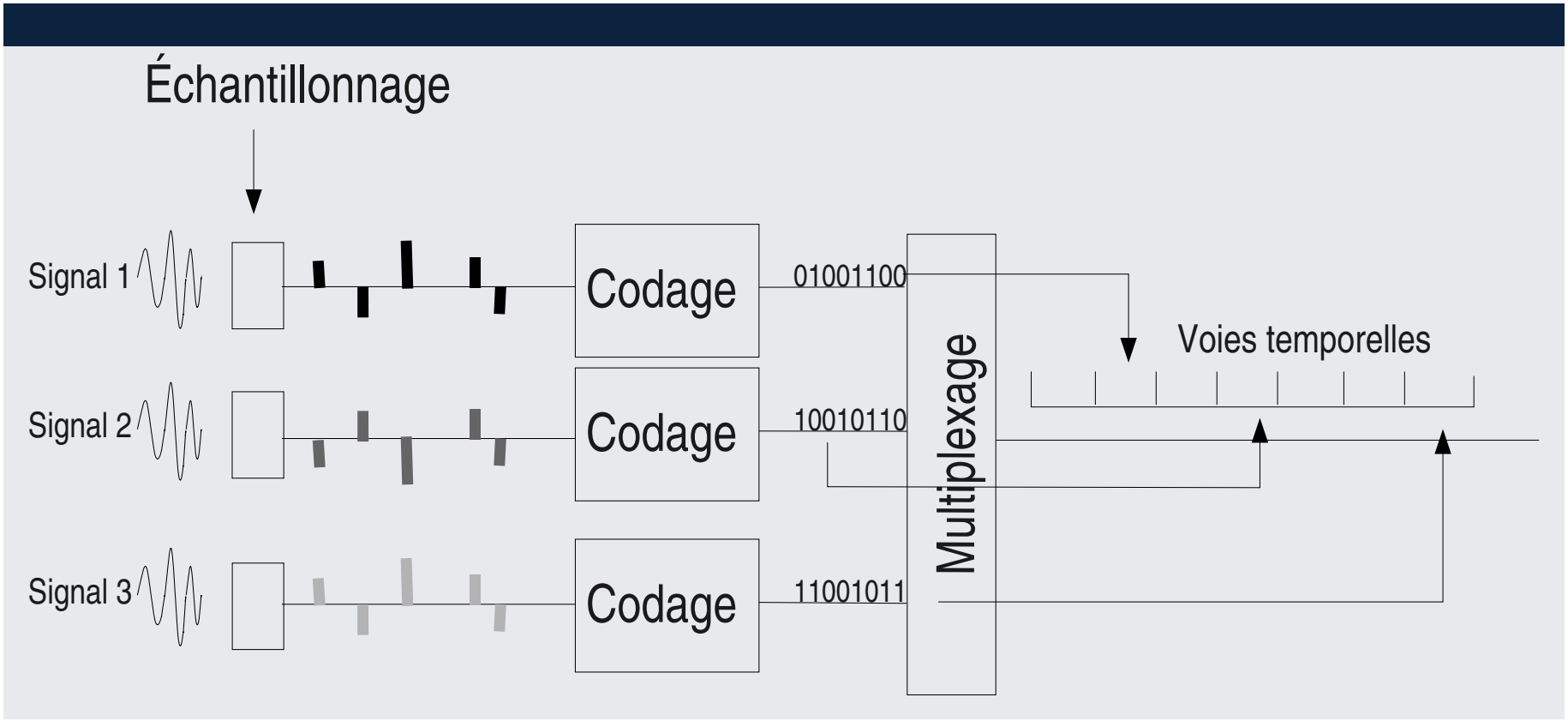
Multiplexage temporel analogique

Échantillonnage d'un canal

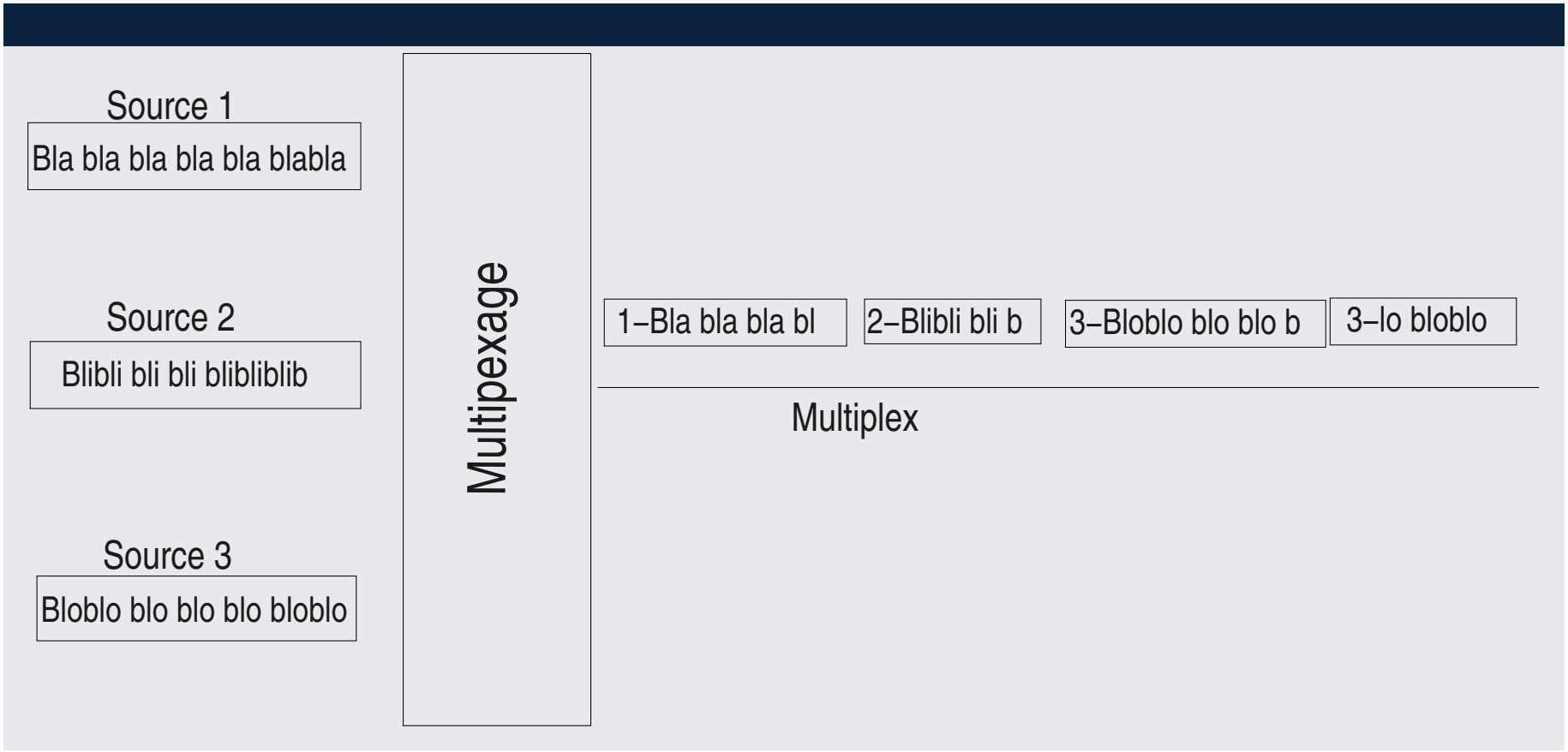


Multiplexage

Multiplexage temporel numérique



Multiplexage statistique



Analogique

Une variation d'un signal physique, la pression d'air générée par un son, l'intensité de lumière et de couleur générée par la lumière éclairant une scène, un paysage, est transformée en un signal électrique représentatif du signal physique. Le signal électrique varie avec les variations du signal physique. Ses variations sont «analogues» à celles du signal physique, c'est un signal analogique.

Numérique

Un signal électrique analogique est échantillonné (on considère de courts instants du signal analogique, des «échantillons»). Les échantillons sont alors mesurés sur une certaine échelle et à chaque pas de cette échelle correspond un nombre. L'échantillon est remplacé par le nombre, nombre lui même codé en base 2 c'est à dire une une suite de 0 et 1. L'échantillon est numérisé, il devient un élément d'un signal numérique.

Quelques remarques sur l'échantillonnage

- Nyquist (1889-1976) a montré que pour restituer correctement un signal après échantillonnage il faut que la fréquence de celui-ci soit au moins le double de la fréquence maximale à échantillonner
- Exemple en téléphonie :
  - Bande analogique : 300Hz-3400Hz
  - Fréquence d'échantillonnage minimale : 6800Hz
    - Par précaution on prend 8000Hz

Avantages et inconvénients des différents types de multiplexage

- Multiplexage temporel
  - ressource (la voie temporelle) réservée pour la durée de la communication, même si celle-ci est silencieuse : mauvaise utilisation de la ressource
  - bande passante et délais de transfert garantis
- Multiplexage en fréquence
  - mêmes avantages et inconvénients

Avantages et inconvénients des différents types de multiplexage II

- Multiplexage statistique
  - utilisation optimale du canal, si des sources sont silencieuses, le canal peut être utilisé par d'autres
  - bande passante globale partagée entre toutes les sources, pas de garantie de réservation (sauf Frame Relay et ATM au prix de complexité supplémentaire pour la réservation et le contrôle de l'utilisation)
  - pas de délai de transfert garanti
  - gigue (variation des délais) pouvant être importante
  - C'est actuellement le moyen le plus utilisé pour les données.
  - La voix et la vidéo sont mal adaptées à cette technique car les délais et la bande passante ne sont pas garantis

La paquetisation dans les réseaux de données

- Multiplexage statistique
  - Pendant qu'une source émet sur un multiplex, les autres sources doivent être silencieuses
  - Pour rendre équitable l'utilisation du multiplex, une source ne peut pas le monopoliser trop longtemps, il faut limiter sa durée d'émission
  - Les données sont segmentées en unités appelées «paquets»
  - Les paquets ont une taille maximale et parfois une taille minimale
  - Les paquets doivent être munis d'une entête, sorte d'étiquette, qui permet de les reconnaître et ainsi de savoir en réception vers quel destinataire acheminer le paquet

Communications avec ou sans connexions

- Avec connexion, comme le téléphone ?
  - Il faut chercher un chemin dans le réseau entre la source et la destination puis le réserver et l'établir
  - Lorsque la communication est terminée il faut libérer le chemin
  - Ces opération nécessitent des échanges d'informations spécifiques que l'on appelle la signalisation
  - Le chemin est appelé circuit
- Sans connexion, comme à la poste ?
  - On munit les données «d'enveloppes» contenant l'adresse de la destination et le réseau achemine ces données en les routant dans chaque nœud en fonction de cette adresse
  - Les unités de données véhiculées sont appelées des datagrammes

Avantages et inconvénients des modes avec et sans connexion I

- Mode orienté connexion
  - Avantages
    - Le chemin est toujours le même pour la durée de la connexion, les données sont reçues dans l'ordre ou elles ont été émises
    - La signalisation nécessaire à l'établissement de la connexion peut permettre de véhiculer des informations de demande de qualité de service
    - Les délais de traitement dans les nœuds sont généralement courts
  - Inconvénients
    - Il faut un certain délai d'établissement et de rupture de la connexion
    - Le chemin est préétabli et si une maille du réseau devient inutilisable (panne d'un nœud, rupture de la maille) la communication est rompue

Avantages et inconvénients des modes avec et sans connexion II

- Mode sans connexion
  - Avantages
    - Pas de nécessité de signalisation pour établir des chemins
    - Reroutage facilité des données en cas de rupture d'un lien dans le réseau
  - Inconvénients
    - Il n'y a pas de chemin préétabli, deux unités de données successives peuvent arriver dans le désordre si le chemin de la seconde a été plus court que le chemin de la première en cas de modification de parcours entre les deux unités
    - On peut envoyer des données vers des destinations inexistantes

Réseaux actuels avec et sans connexion

- Orientés connexion
  - X25 : le réseau de données des années 80 (né vers 1976)
  - Frame Relay, très utilisé aujourd'hui pour interconnecter des unités dispersées de mêmes entreprises
  - ATM : dans le cœur de réseau des opérateurs (mais aussi dans votre modem ADSL...)
  - MPLS (MultiProtocol Label Switching) : dans le cœur des réseaux d'opérateurs (au service des entreprises)
- Orientés sans connexion
  - Les réseaux locaux d'entreprises : Ethernet, Token-Ring
  - IP : Internet
  - Réseau Cyclades (l'ancètre de IP, abandonné en 1978 par choix politique...)

Caractéristiques des réseaux orientés connexion

- Les nœuds acheminent les unités de données (les paquets) entre les entrées et les sorties en effectuant des opération de commutation
  - Les nœuds sont des commutateurs
  - Les paquets sont munis d'étiquettes qui les identifient
  - Les étiquettes sont attribuées lors de la phase d'établissement de la connexion. Elles identifient les paquets sur chaque lien.
  - Elles identifient aussi la communication, le canal, le circuit.
  - La série d'étiquette réservées pour une communication sur chaque lien constitue un circuit virtuel

Connexion et circuit virtuel

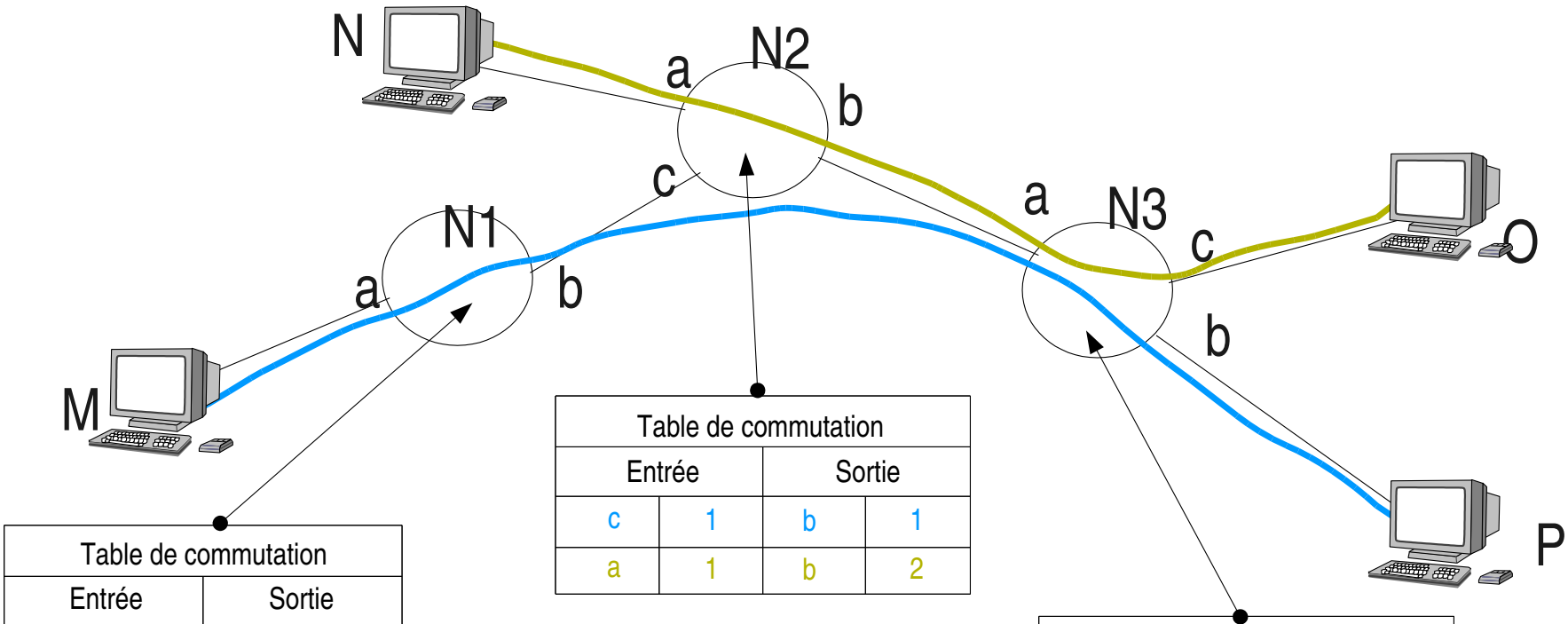
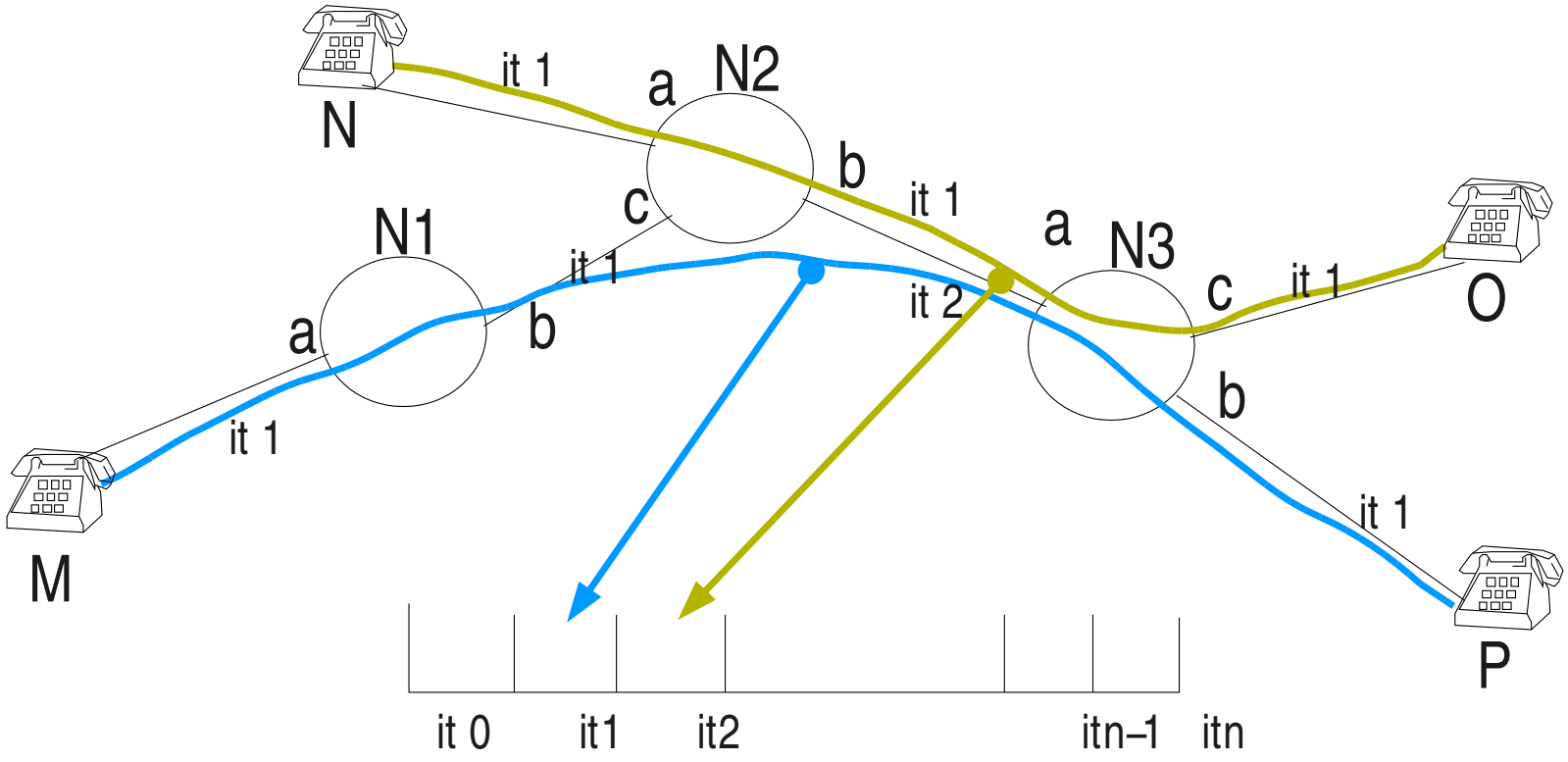


Table de commutation			
Entrée		Sortie	
a	1	b	1

Table de commutation			
Entrée		Sortie	
c	1	b	1
a	1	b	2

Table de commutation			
Entrée		Sortie	
a	1	b	1
a	2	c	1

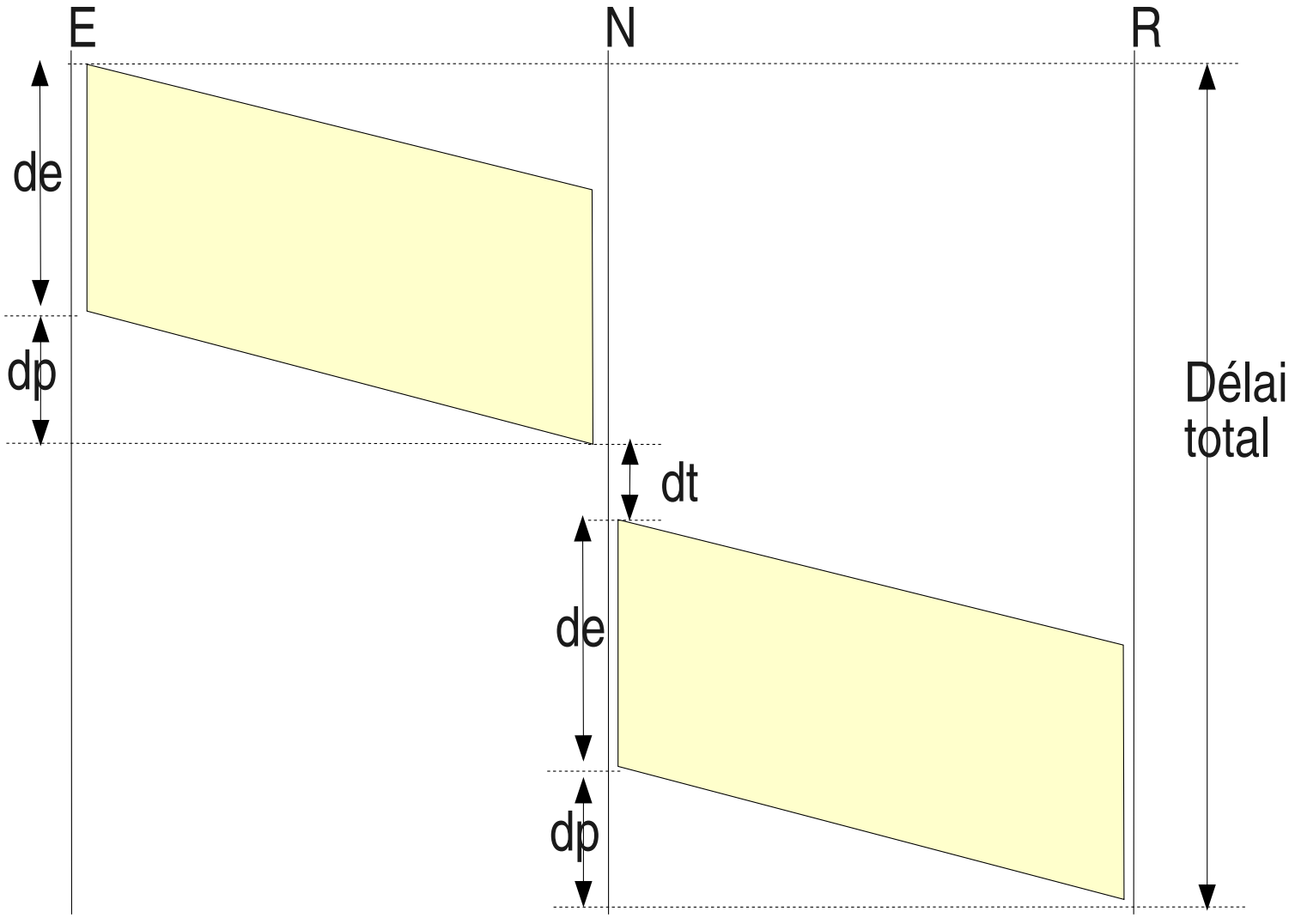
- Il n'y a pas d'étiquette associée aux données
- les données sont véhiculées dans des canaux spécifiques, réservé lors de la phase d'appel, établis lors de l'établissement de la connexion
- les canaux peuvent être des intervalles de temps, régulièrement espacés (en téléphonie numérique européenne : 32 intervalles de temps tous les 125μs)
- les noeuds du réseaux sont des commutateurs, ils commutent les données en fonction des informations contenues dans leurs tables de commutation
- les tables de commutation donnent la correspondance entre le canal entrant et le canal sortant



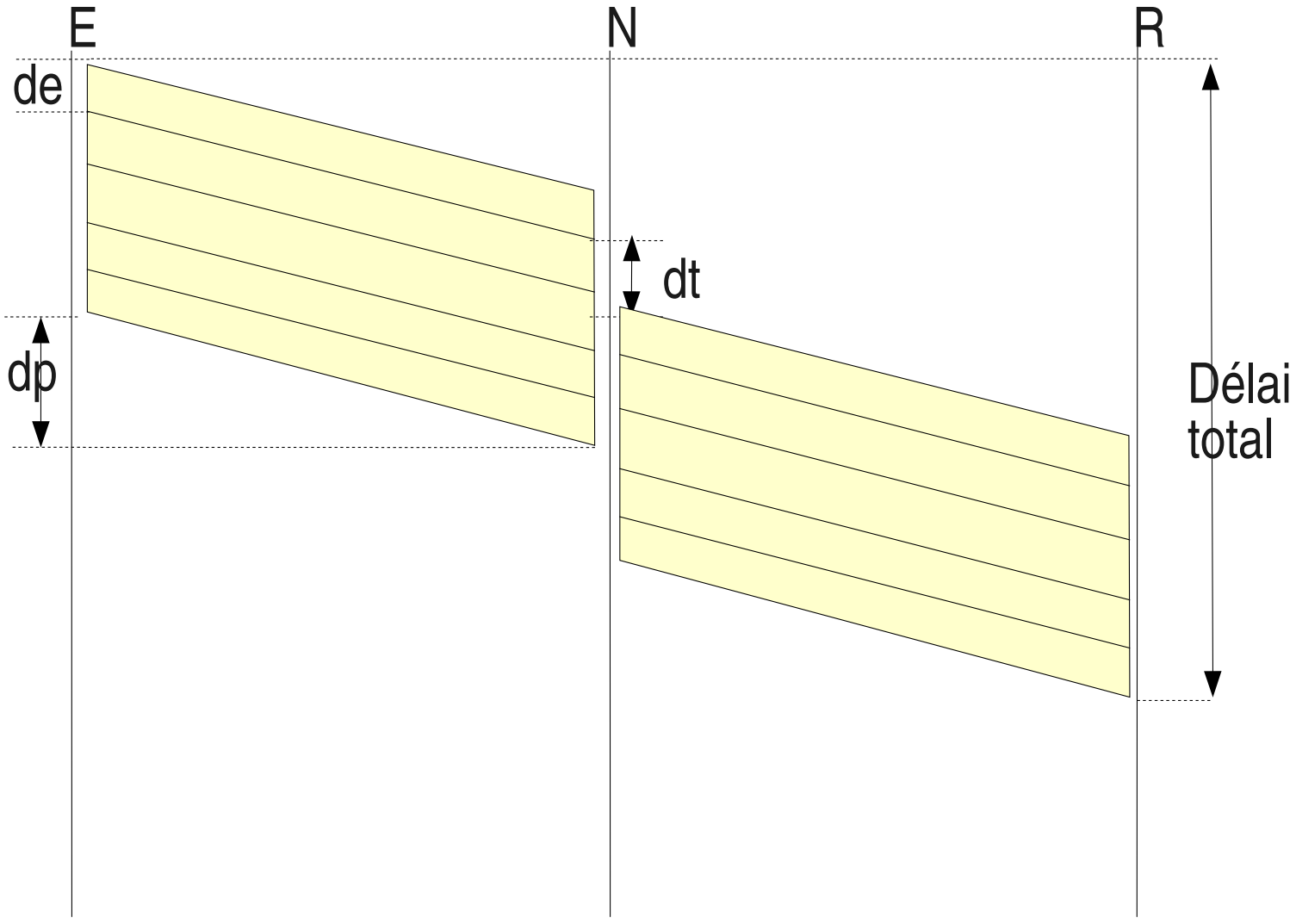
- Entre $N2$ et $N3$, les intervalles de temps 1 et 2 sont attribués aux communications $M - P$ et $N - O$, respectivement, et ce, pour la durée des communications
- Si la communication $M - P$ se termine, la ressource $it1$ est libérée, la communication $N - O$ n'en profite pas
- Si la communication $M - N$ est silencieuse, la bande passante non utilisée est perdue

Les délais

- Entre un émetteur et un récepteur, les données vont subir différents délais
  - Durée d'émission : inversement proportionnelle au débit, proportionnelle à la longueur des segments de donnée
  - Le délai de propagation : au minimum le délai de la lumière dans le vide ($c \approx 300000 Km/s$), dépend par ailleurs du support ($2/3c$ dans le verre et le cuivre).
    Exemple : 250μs pour une liaison par satellite géostationnaire (terre - terre)
  - Le délai de traitement dans les nœuds : délai de traitement réel plus temps passé dans les files d'attente en attente de traitement et en attente de ré-émission


E
N
R
de
dp
dt
de
dp
Délai
total

Réduisons les délais avec de petits paquets



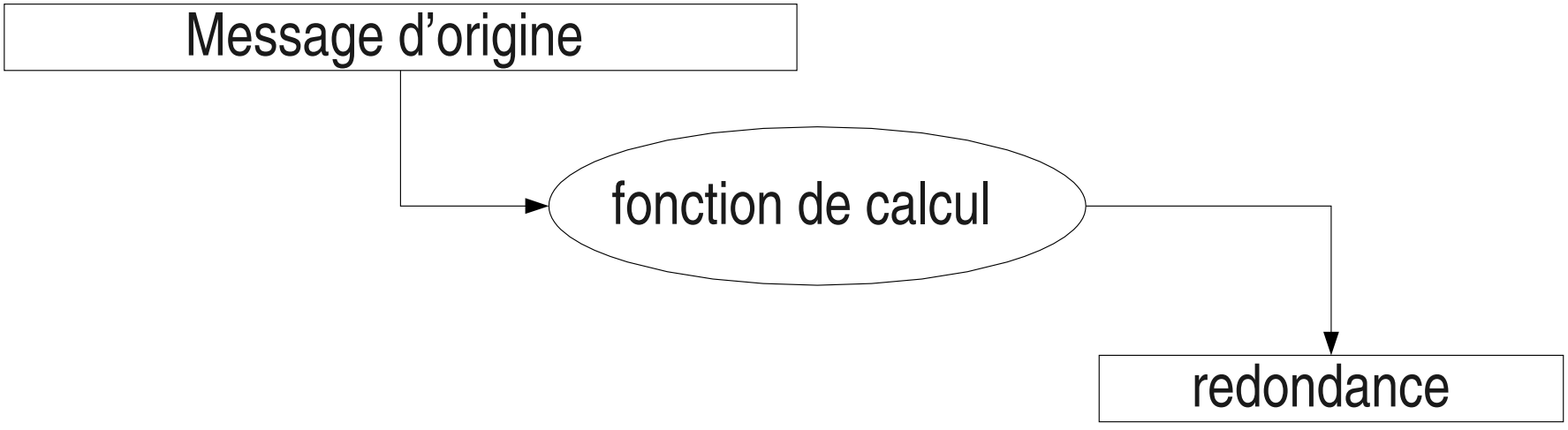
Notion de qualité de service liée aux délais

- Pour le téléphone
  - Délai de 50ms : bon confort de communication
  - Au delà de 200ms (communications par satellites geostationnaires) : qualité très moyene, nécessité d'annullation d'écho
  - 400ms : toute dernière extrémité à ne pas dépasser
- Pour les données
  - Dépend fortement du type d'application
- Variation des délais
  - Inhérente aux réseaux de transmission de données
  - Inadapté aux services de type téléphone ou vidéo (mais on tente quand même avec succès, p.ex. : VoIP)

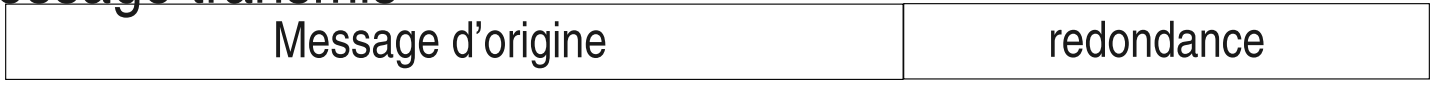
Détection et correction d'erreur

- Une transmission binaire ne se fait pas sans erreur : des «1» peuvent devenir des «0» et réciproquement
  - Comment détecter des erreurs ?
  - Comment les corriger ?
- Détection (principe général)
  - transmet l'information par bloc plus une séquence de redondance
  - la séquence de redondance est calculée par une fonction spécifique
  - le même calcul est fait à l'arrivée et le résultat est comparé
- Correction
  - directe si la redondance est suffisante et possède des propriétés de correction
  - par retransmission

Détection d'erreur : par séquence de redondance



Message transmis



À la réception l'opération est reconduite, le résultat est comparé à ce qui est reçu et le message est accepté ou non.

Détection d'erreur : la méthode du bit de parité

- Un bit de parité est rajouté à une séquence de bits
  - le nombre de bits à 1 résultant doit être pair (parité paire) ou impair (parité impaire)
  - Exemples :
    - parité paire

1	0	0	1	0	0	1	1
0	0	0	0	1	1	0	0

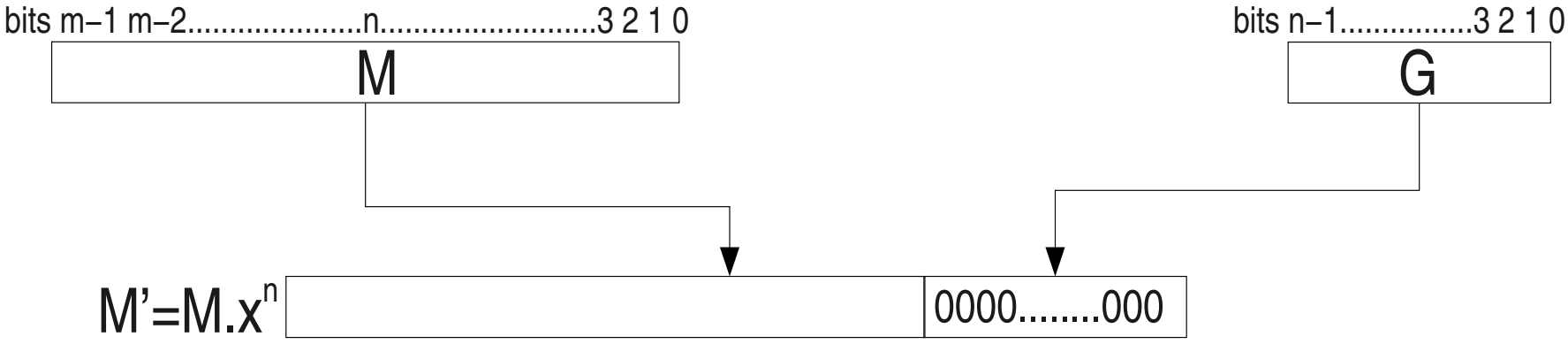
    - parité impaire

1	0	0	1	0	0	1	0
0	0	0	0	1	1	0	1

  - Ne permet pas de détecter des erreurs doubles
  - Méthode utilisée par exemple dans les interfaces séries de nos PC (interfaces de type ANSI RS232-C (ITU-T V24))

Détection d'erreur : la méthode de la division par polynôme I

- Un message M est une suite de bits, donc un nombre. On peut l'assimiler à un polynôme. Exemple : la suite de bits 101101 peut être représentée par le polynôme $x^5 + x^3 + x^2 + x^0$
- Le message M a au plus m bits, il est de degré $m - 1$
- Considérons un polynôme G de n bits (degré n-1) avec $n < m$
- On multiplie M par x^n (2^n). Cela revient à «décaler» M de n bits vers la gauche et à ménager ainsi n bits vides à droite



Détection d'erreur : la méthode de la division par polynôme II

- Le polynôme M' est divisé par G (division modulo 2)
- Alors : $M' = G \times Q + R$ (Q polynôme quotient, R reste)
- Le polynôme $T = M' - R$ est construit
  - le reste R vient se placer dans l'espace droit de M' (la soustraction binaire modulo 2 est équivalent à une addition)

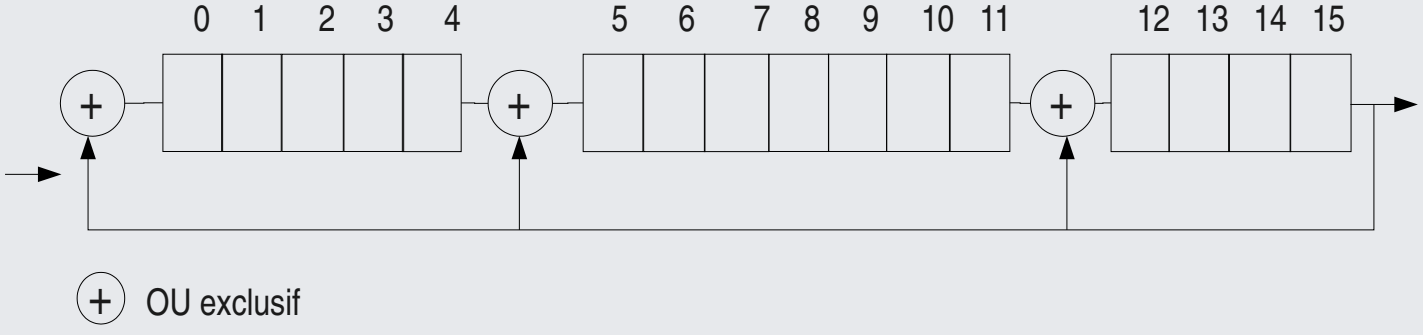
    $M' = M.x^n$ | 0000 . . 0000

    $T = M' - R$ | R
  - $T = M' - R = G \times Q$ donc T est divisible par G
- On transmet T, à la réception on divise le polynôme reçu par G, si on a reçu T, pas d'erreur, alors $R = 0$

Quelques CRC types et le calcul par registres à décalage

- CRC-12 $= x^{12} + x^{11} + x^3 + x^2 + x + 1$
- CRC-16 $= x^{16} + x^{15} + x^2 + 1$
- CRC-CCITT $= x^{16} + x^{12} + x^5 + 1$
- CRC-32 $= x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$

Codage du CRC-16 à l'aide d'un registre à décalage



Détection d'erreur : la méthode de la somme de contrôle

(ou *Checksum*)

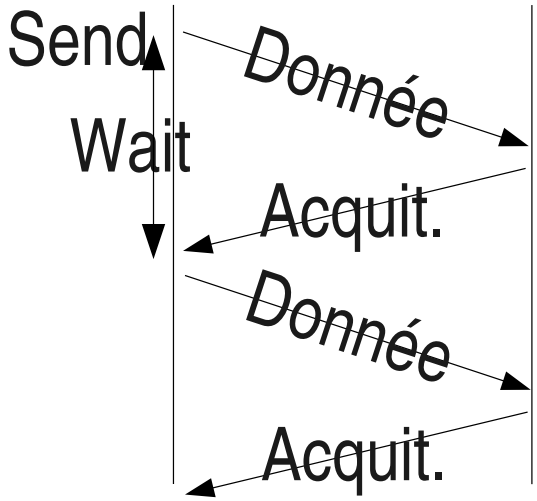
- Soit un message formé par la suite de mots A, B, C, D, ..., R, S, T
  - Les mots sont des ensemble de 8 ou 16 bits ou plus
  - Un mot contiendra la somme de contrôle dans le message, soit S ce mot. Il est mis à 0 pour le calcul
  - Calcul : $Z = A + B + C + D + \ldots + R + 0 + T$
  - Le résultat Z est inversé, on obtient $\bar{Z}$
  - On transmet $ABCD \ldots R\bar{Z}T$
  - À la réception on fait la somme $Z' = A + B + C + D + \ldots + R + \bar{Z} + T$
  - S'il n'y a pas d'erreur alors $Z' = Z + \bar{Z} = 1111 \ldots 1111$
  - le résultat est inversé comme à l'émission et donc $\bar{Z'} = 0$

- Si la redondance est suffisante et l'algorithme sufisamment puissant il est possible de détecter les erreurs et de les corriger.
  - On peut le montrer sur un exemple simple avec le mécanisme du bit de parité.
    Soit le bloc suivant, chaque ligne et chaque colonne est munie d'un bit de parité paire (dernier bit).

	1	0	1	0	1	0	1	0
	1	0	1	0	0	0	0	1
Cherchez l'erreur :	0	0	0	1	0	1	0	0
	1	1	1	1	0	0	0	0
	1	1	1	0	0	1	1	1

  - Autres techniques : codes de Reed-Solomon, Turbo-codes (origine ENST Bretagne), etc.

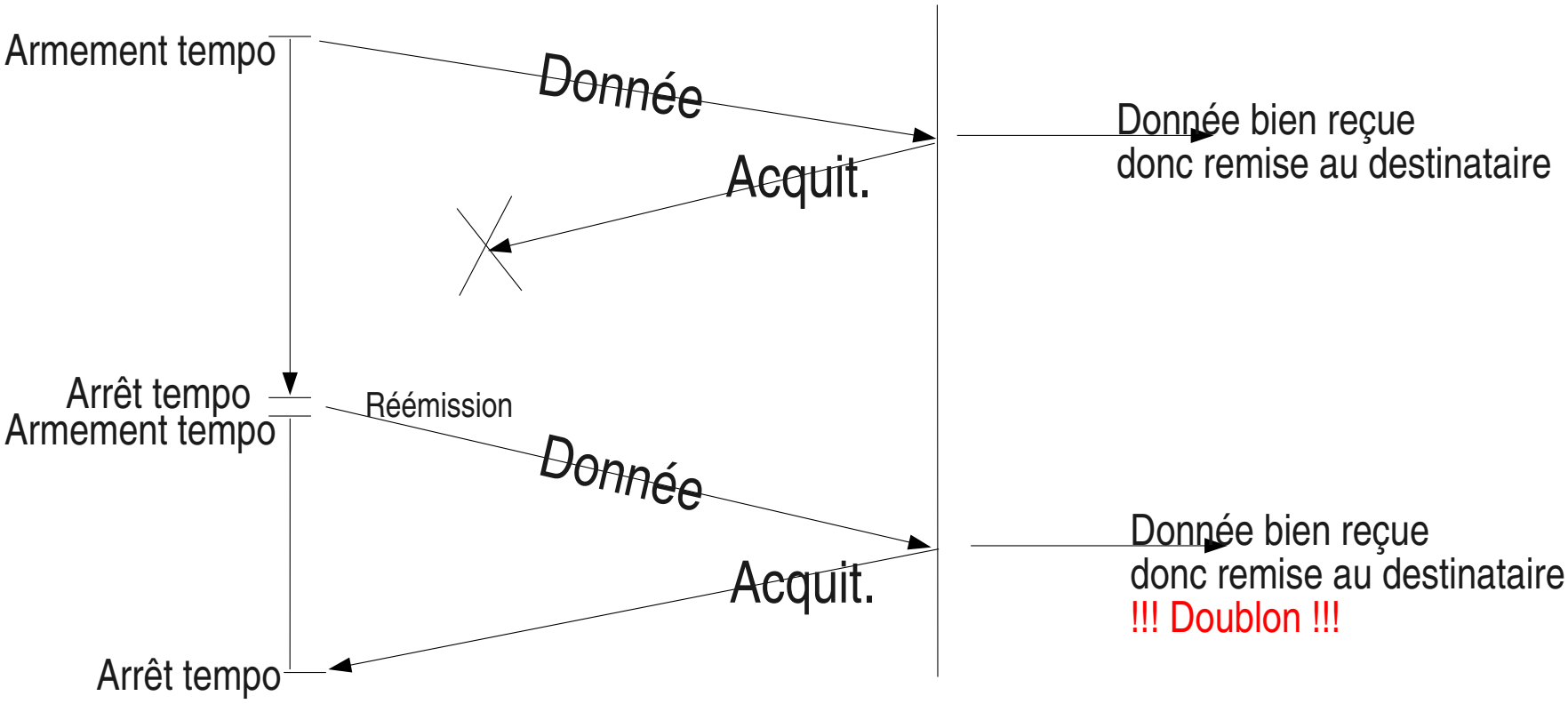
- Un récepteur détecte une erreur, il jette le segment de données erroné
- Comment l'émetteur peut-il détecter qu'il y a eu une erreur ?
  - Impossible sans l'utilisation d'un mécanisme d'acquittement



Ce mécanisme est appelé «Send & wait»

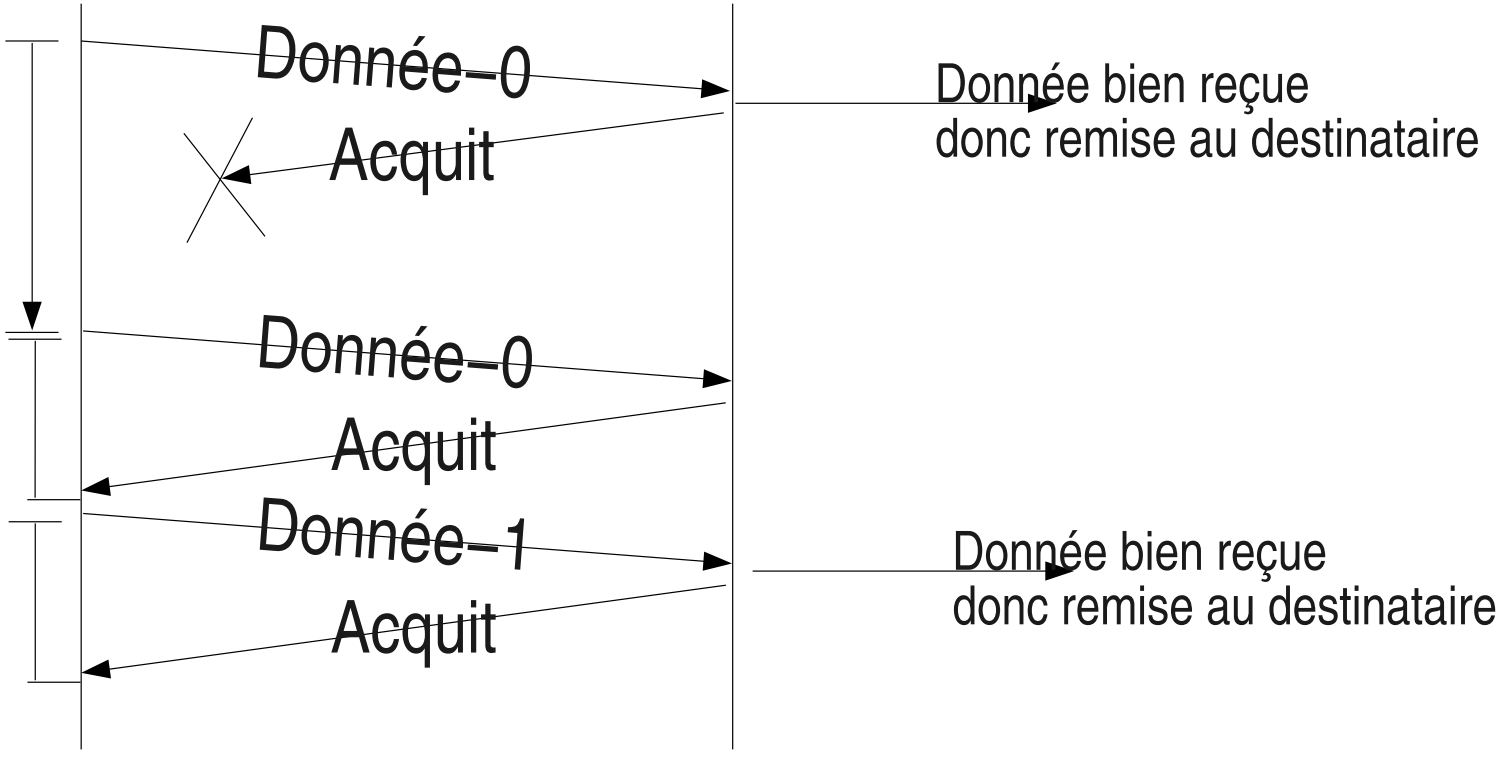
Que se passe-t'il si la donnée se perd ?
Si l'acquittement se perd ?

- Utilisation de temporisateur (timer)

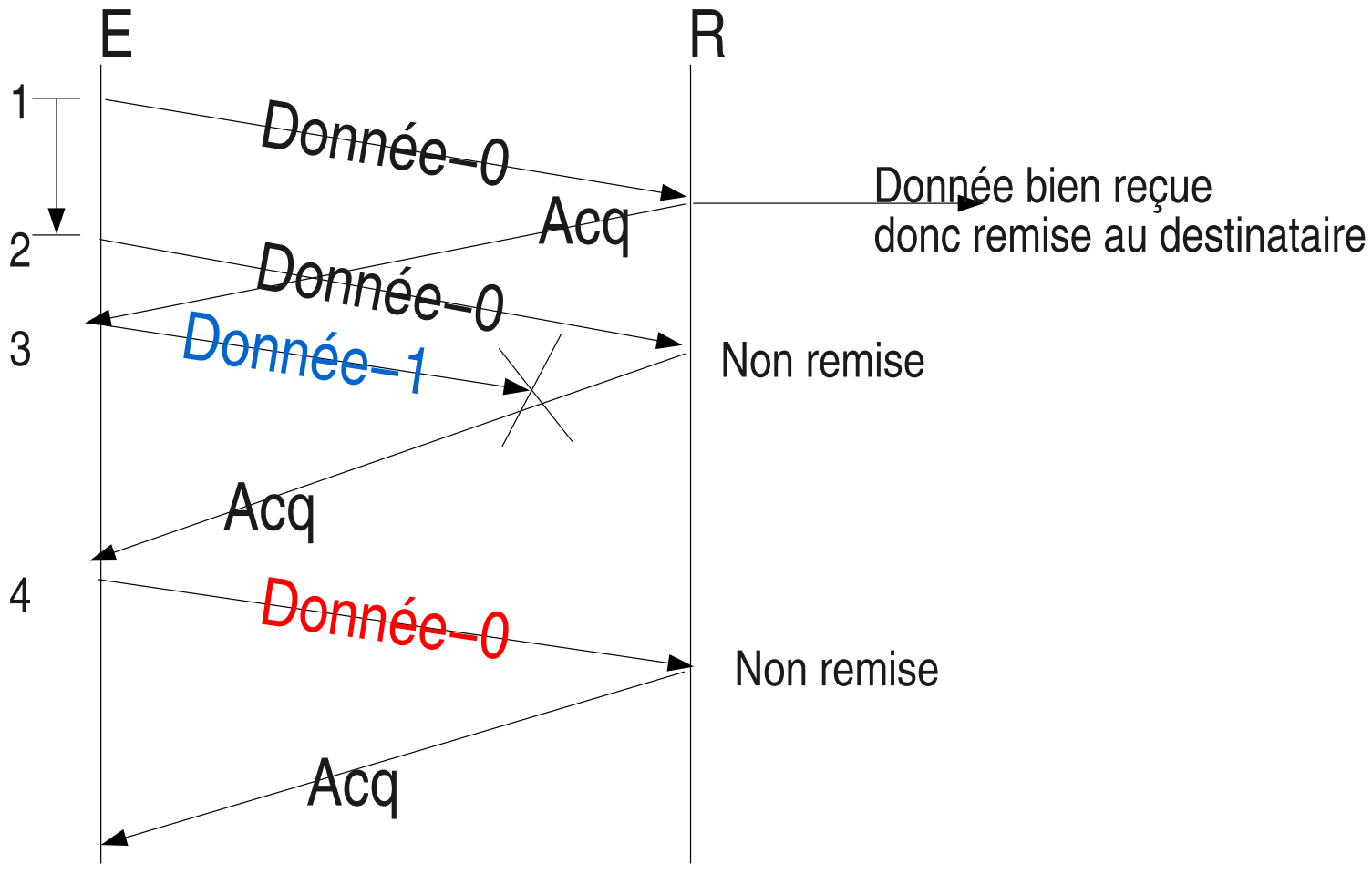


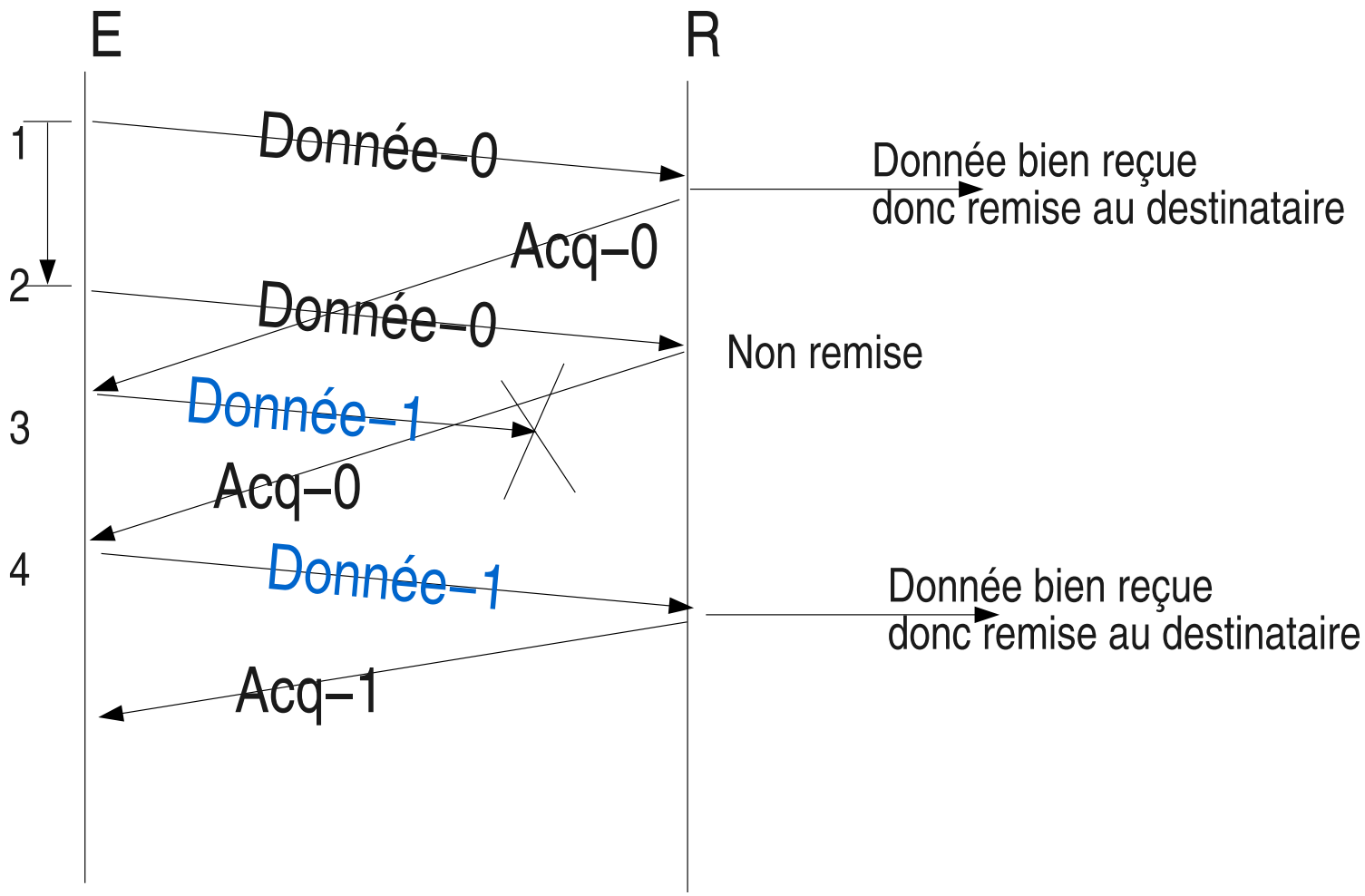
Retransmission : le problème des doublons

- On peut éviter les doublons en numérotant les données
  - Exemple : les données sont munies d'un numéro 0 et 1 alternativement



Timer trop court et malchance...

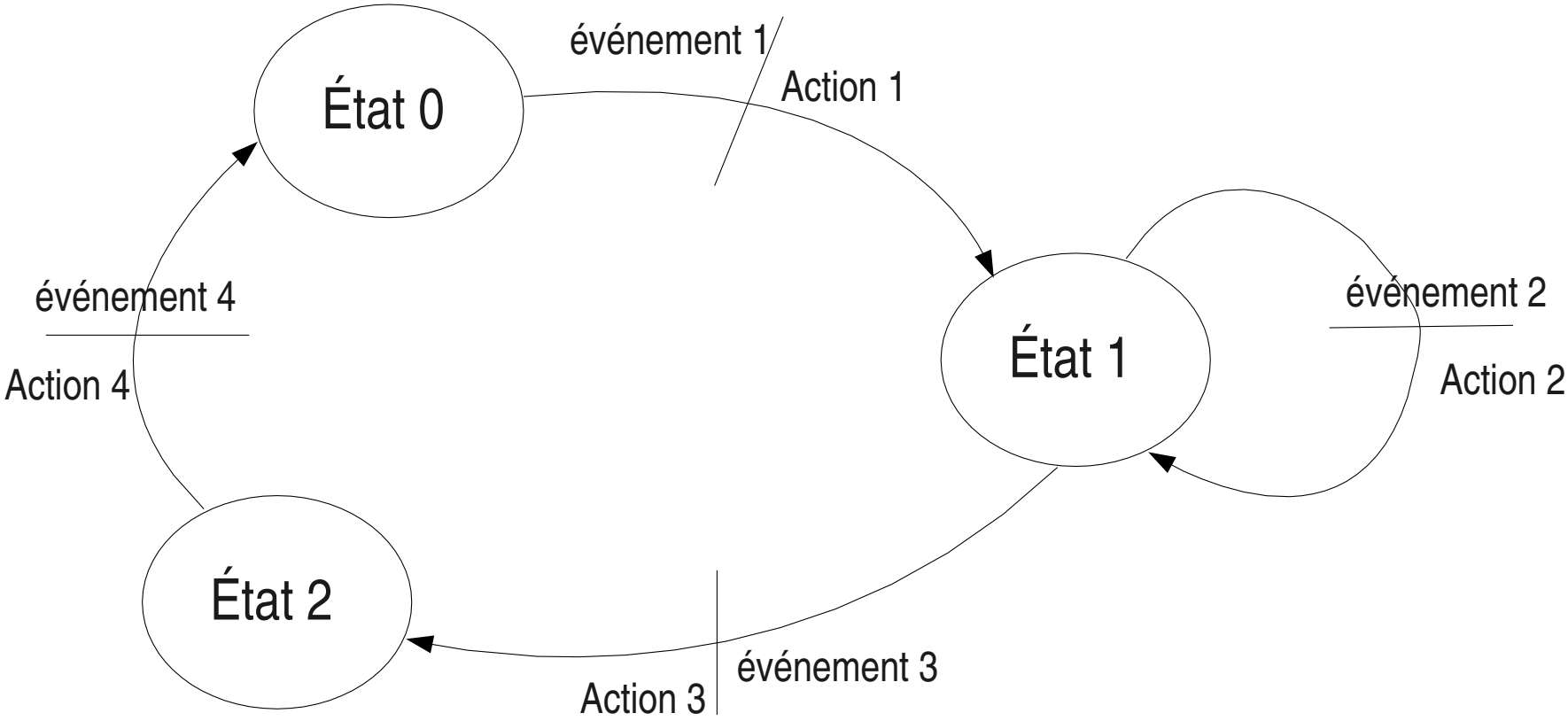



E
R
1
2
3
4
Donnée–0
Acq–0
Donnée–0
Donnée–1
Acq–0
Donnée–1
Acq–1
Donnée bien reçue
donc remise au destinataire
Non remise
Donnée bien reçue
donc remise au destinataire

- Dans ce qui précède nous avons développé des moyens de transmettre des données sans erreurs en rajoutant de l'information et des messages spécifiques
  - À chaque ajout de mécanisme on corrige un problème, on en rajout un autre
  - Il faut que la série correction/nouveau problème converge

- En rajoutant un protocole au dessus d'une voie de communication on modifie les propriétés de cette voie de telle manière que l'on obtient une nouvelle voie de comunication
- Pour mettre en œuvre et gérer les mécanimes il faut construire un automate logiciel comportant des états et des transitions d'états sur événements

Exemple d'automate d'états finis

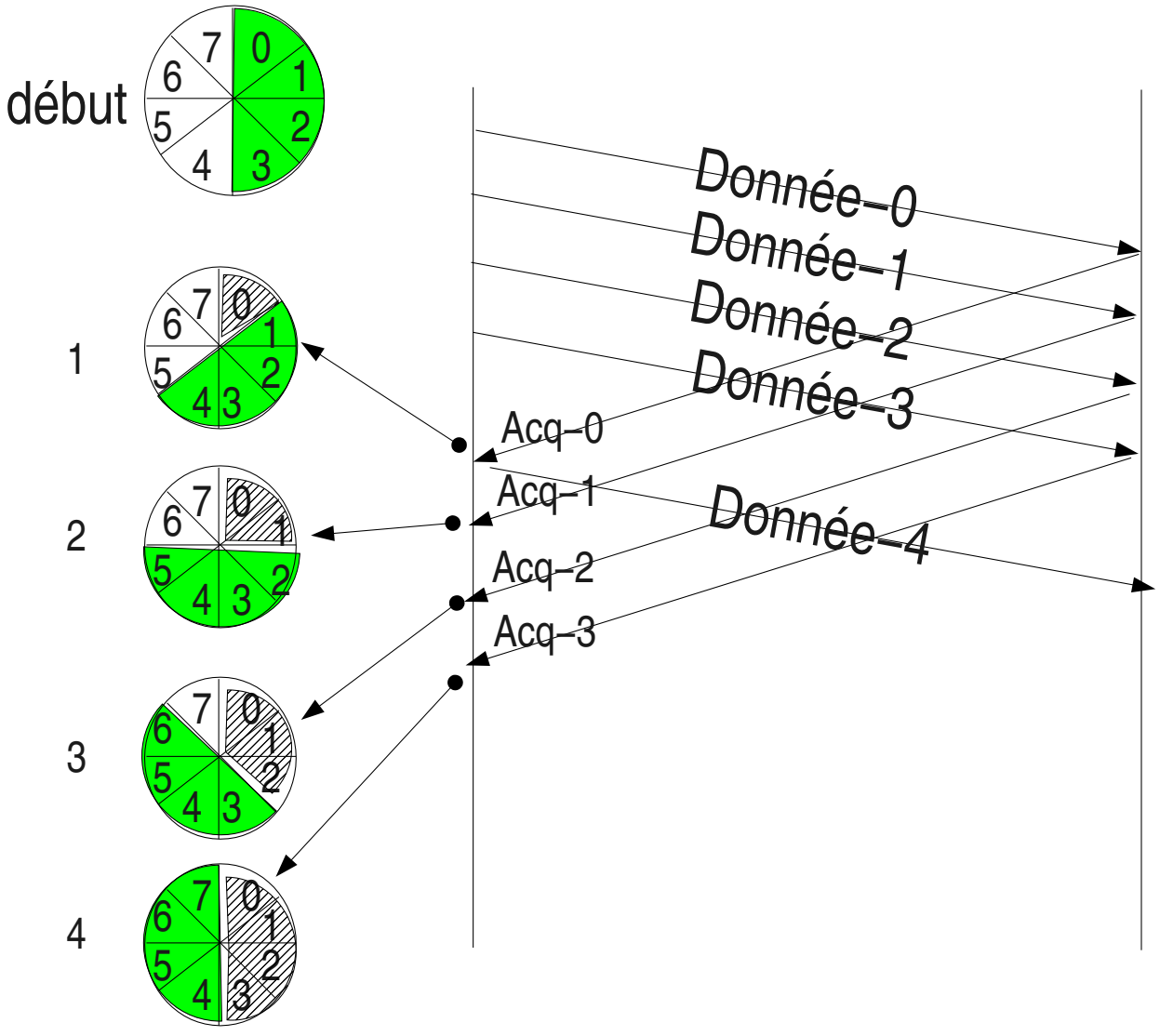


L'automate du protocole du bit alterné

Événement État	Demande d'envoi de donnée	Arrivée Ack 0	Arrivée Ack 1	Expiration Timer
État 0 nouvel état	Envoyer Data-0 Armer Timer Attente Ack 0			
Attente Ack 0 nouvel état		Arrêt timer État 1		Envoyer Data-0 Armer Timer Même état
État 1 nouvel état	Envoyer Data-1 Armer Timer Attente Ack 1			
Attente Ack 1 nouvel état			Arrêt timer Etat 0	Envoyer Data-1 Armer Timer Même état

- le mécanisme du *Send & Wait* conduit à une mauvaise utilisation de la bande passante, quand on attend on n'utilise pas la bande passante alors disponible
- Il est judicieux d'anticiper l'émission sans attendre les acquittements
  - Dans une certaine mesure...
  - Une limite, appelée fenêtre d'anticipation, permet de ne pas trop anticiper et bloque l'émission si les acquittements n'arrivent pas
  - Les acquittements existent toujours
  - La fenêtre «tourne» ou «glisse» lorsqu'un acquittement arrive

La fenêtre d'anticipation



Le contrôle de flux

- Mécanisme permettant à un récepteur d'asservir la capacité à émettre de son correspondant en fonction de ses capacités de traitement
  - Permet d'informer l'émetteur qu'il doit réduire son débit
  - Permet de ne pas inonder les tampons mémoire du récepteur

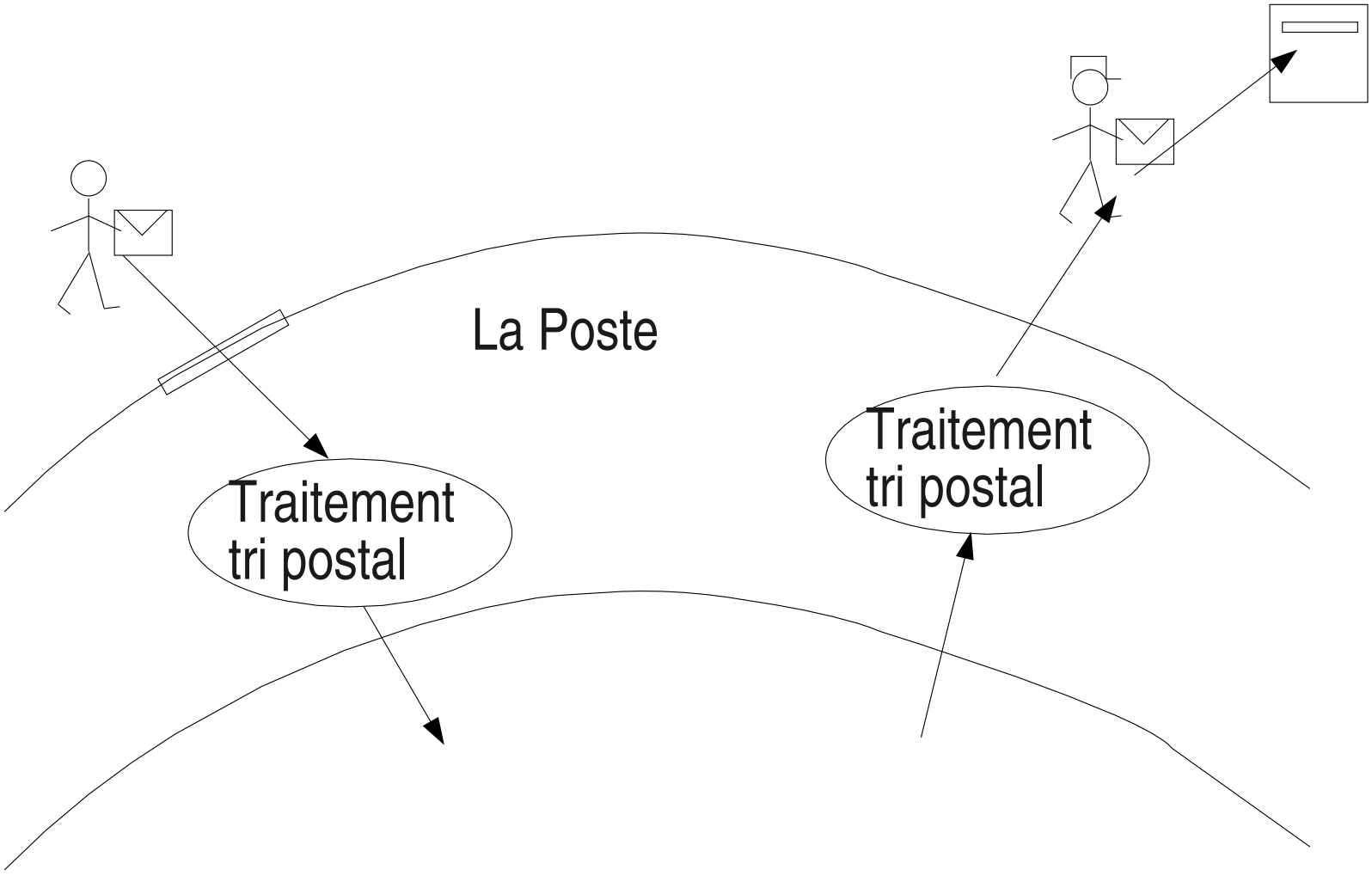
Les organismes de standardisation

- Les organismes officiels nationaux et internationaux
  - OSI : Organisation de Standardisation Internationale (ISO en anglais), et ses branches nationales : AFNOR en France, DIN en Allemagne, ANSI aux USA
  - UIT-T : Union Internationale des Télécommunications, secteur des Télécommunications (il y a aussi le secteur Radio)
  - ETSI : European Telecommunications Standards Institute
- Les organismes de l'industrie et de la recherche
  - IEEE : Institut of Electrical and Electronics Engineers
- L'Internet
  - IETF : Internet Engineering Task Force, étudie et développe les protocoles et services au dessus du protocole IP

- Comment «voir» le réseau ?
  - Comme un utilisateur : «*Je me connecte, je ne sais pas comment ça marche, je ne veux pas savoir comment ça marche, mais ça marche ! Et j'utilise.*»
  - Comme un développeur d'application communicantes : «*Par quel moyen programmatique mes applications peuvent-t'elles communiquer ? Dois-je savoir comment fonctionne le réseau ? Tout le réseau ? Une partie du réseau ?*»
  - Comme le gestionnaire du réseau : «*Dois envisager le réseau dans son ensemble, du câble aux applications ?*»

Exemple d'un réseau existant : La poste

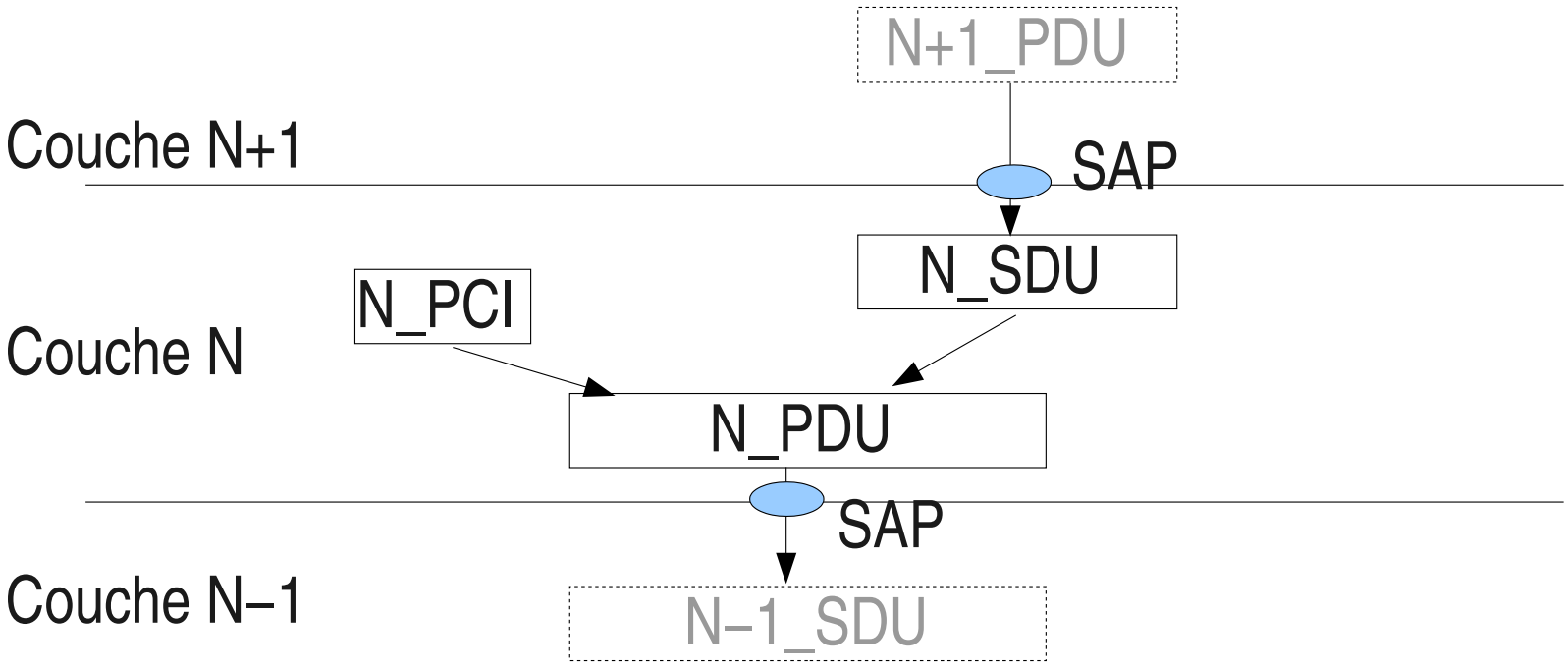
- En tant qu'utilisateur :
  - *Je dois savoir où est situé le bureau de poste, l'adresse de mon correspondant et avoir une boite aux lettres. Je dois pouvoir interagir avec le service offert par La Poste. La poste doit me fournir un service et des moyens d'accès à ce service.*
  - *Je n'ai pas à savoir comment fonctionne le service de La Poste de l'intérieur*
- En tant que postier :
  - *Je dois savoir traiter les lettres, les orienter vers les bons sacs postaux, placer les sacs postaux dans les bons camions, voitures trains, avions...*
  - *Je n'ai pas à savoir comment fonctionne le service de transport qui achemine physiquement les sacs postaux.*


La Poste
Traitement
tri postal
Traitement
tri postal

Le modèle ISO (OSI)

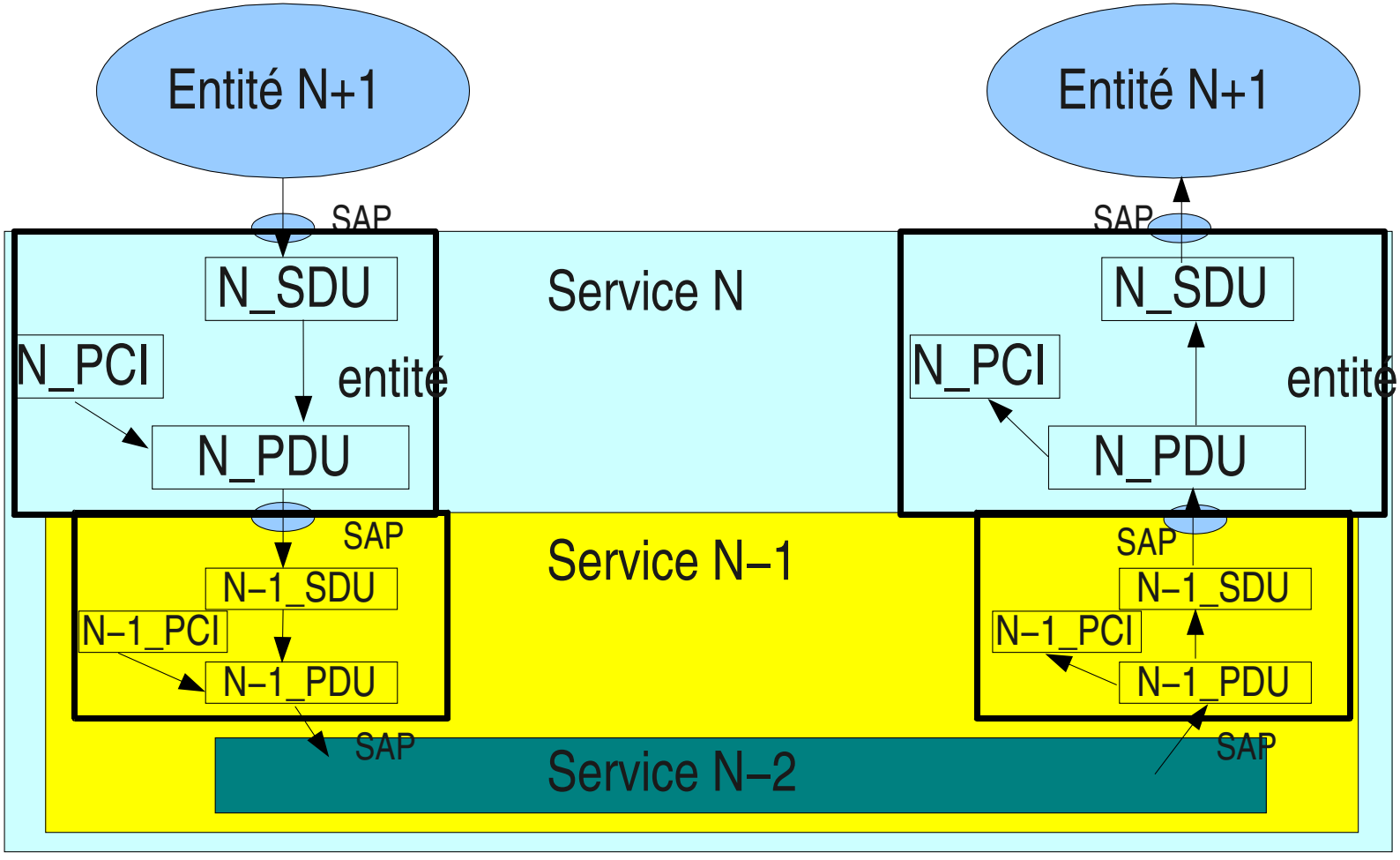
- Interconnexion de Services Ouverts (OSI en anglais)
- Définit la notion de service et de couche de service ainsi que les relations entre les entités distantes d'une même couche (les protocoles de communication)
- Définit aussi les relations entre couches (les primitives de service et les SAP)
- Définit les différentes couches, leur rôle ainsi que leurs protocoles

Le modèle OSI : Couches et unités de données

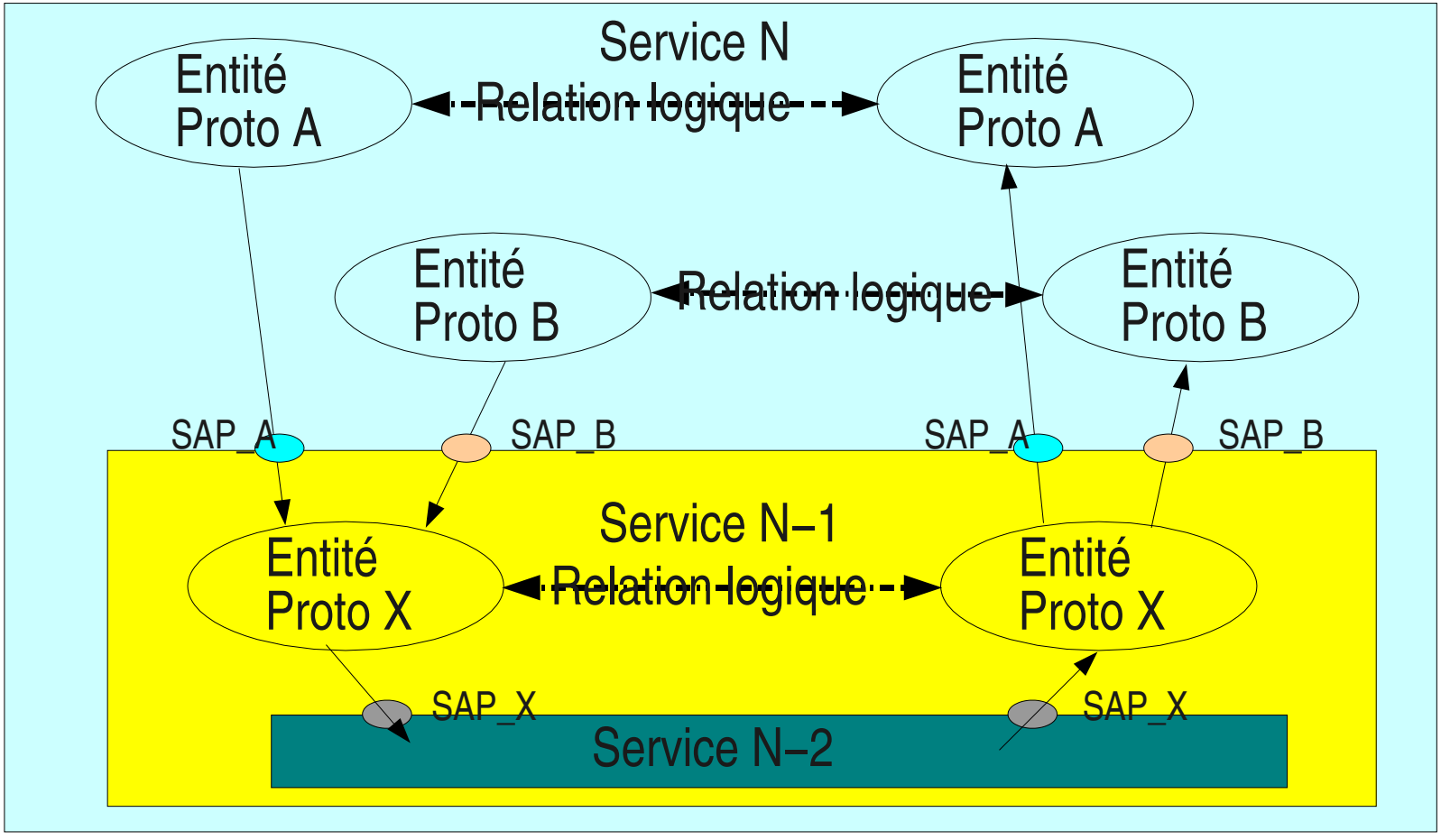


Notion d'entité protocolaire

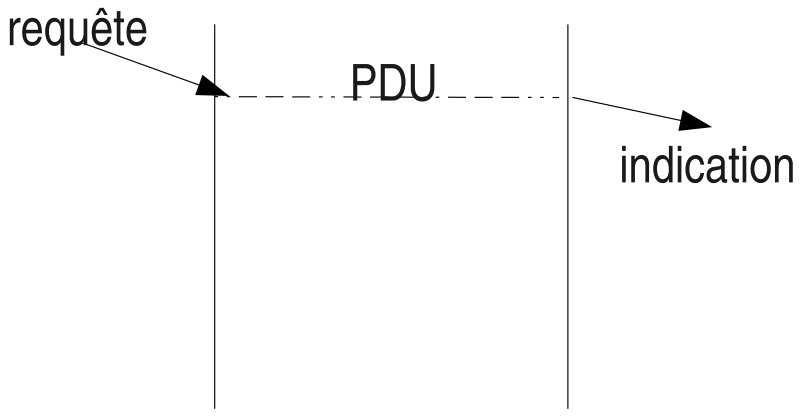
- Une entité protocolaire est un «programme informatique», une application ou un module opérationnel du système d'exploitation mettant en œuvre un protocole
- Une même couche peut être composée de plusieurs entités mettant en œuvre des protocoles différents pour assurer un même service
  - Exemples :
    - Le téléchargement de fichier peut être réalisé via le protocole ftp ou le protocole du web http
    - Des machines peuvent partager des fichiers via les protocoles NFS (monde Unix), SMB (monde Windows) ou AppleTalk (monde MacIntosh)


Entité N+1
Entité N+1
SAP
SAP
N_SDU
Service N
N_SDU
N_PCI
entité
N_PCI
entité
N_PDU
N_PDU
SAP
SAP
Service N–1
N–1_SDU
N–1_SDU
N–1_PCI
N–1_PCI
N–1_PDU
N–1_PDU
SAP
SAP
Service N–2

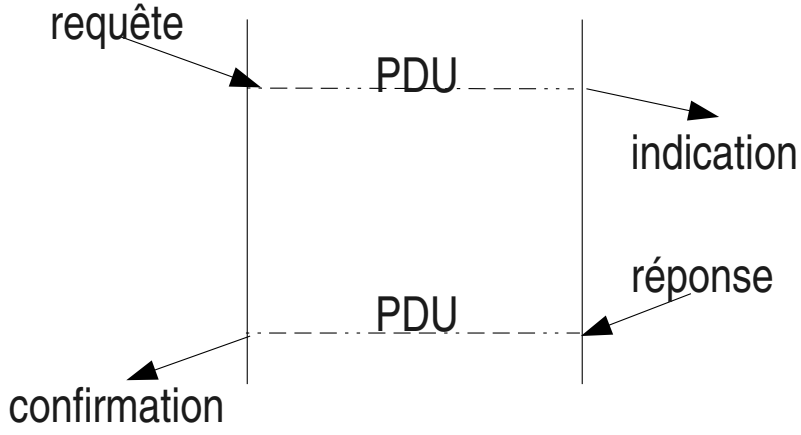
Relations entre entités de même niveau et niveaux différents



Primitives de service : concept et représentation



Service sans confirmation



Service avec confirmation

1 - La couche physique

- Définit les moyens pour transformer les bits constituant les données en information analogique transportable sur les liens physiques entre la machine et son nœud de raccordement au réseau
- Définit les caractéristiques matérielles et électriques (ou optiques) des supports physique

2 - La couche liaison

- Définit les moyens d'acheminer des données structurées au dessus du niveau physique, entre la machine et son nœud de raccordement. La structure de base est la trame
- Définit les moyens de contrôler la fiabilité de la trame en réception
- Peut définir des mécanismes de contrôle de flux et de récupération d'erreurs

3 - La couche Réseau

- Définit les moyens pour acheminer l'information entre machines d'extrémités en traversant les nœuds du réseau
- Implique une nécessité d'identification des machines terminales : un adressage
- Implique de mettre en œuvre des mécanismes de routage dans les nœuds
- Les unités de données de ce niveau sont appelées des paquets

4 - La couche transport

- Dernière des couches basses
- Interface entre les applications et la couche réseau
- Fournit une abstraction du réseau
  - Corrige les imperfections de la couche réseau
    Dernière chance pour que les applications soient assurées du bon transfert de leurs données
  - Le modèle OSI fournit 5 classes de fonctionnalités différentes pour corriger les imperfections du réseau. De la classe 0 pour le très bon réseau à la classe 4 pour le mauvais réseau.

5 - La couche session

- Une communication entre deux applications est considérée comme une session
- La session est organisée, contrôlée
  - Il est prévu des points de synchronisation du dialogue

6 - La couche Présentation

- Fournit les moyens de décrire les types d'information échangés entre les application : langage de description de types : ASN.1
- Fournit les moyens de coder ces types de données dans un format indépendant de celui des machines (problème de la représentation *big* ou *little endian*)

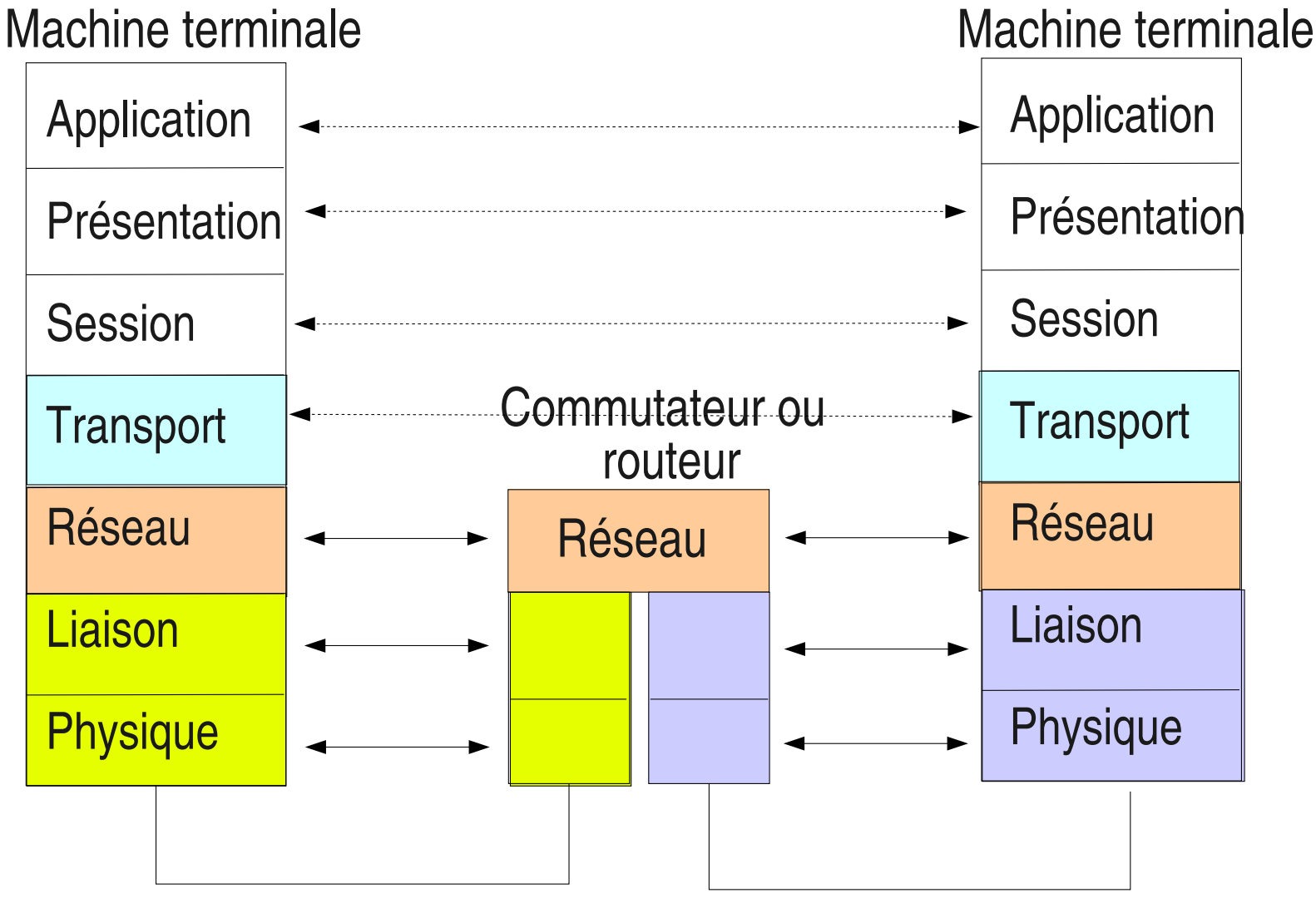
7 - La couche application

- Fournit des sous ensembles applicatifs pour établir et contrôler les communications.

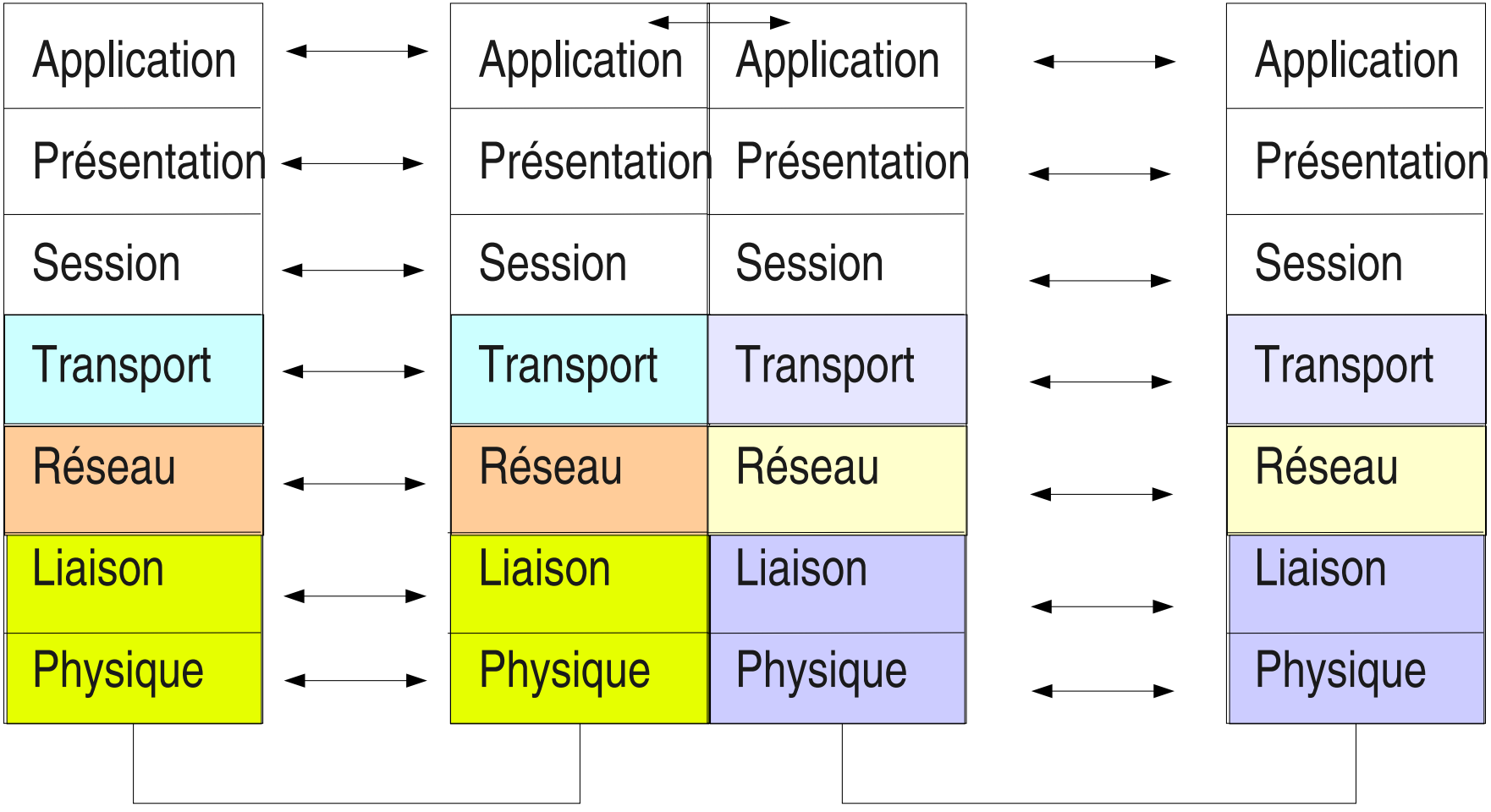
Les 7 couches du modèle OSI : pour résumer...

- Couches physique (1) et liaison (2) : de la machine au réseau
- Couche réseau (3) : de la machine à la machine en traversant le réseau
- Couche transport (4) : une abstraction du réseau (couches basses) pour l'applicatif
- Couches hautes (5 6 7) : des services, géré au niveau applicatif ou middleware

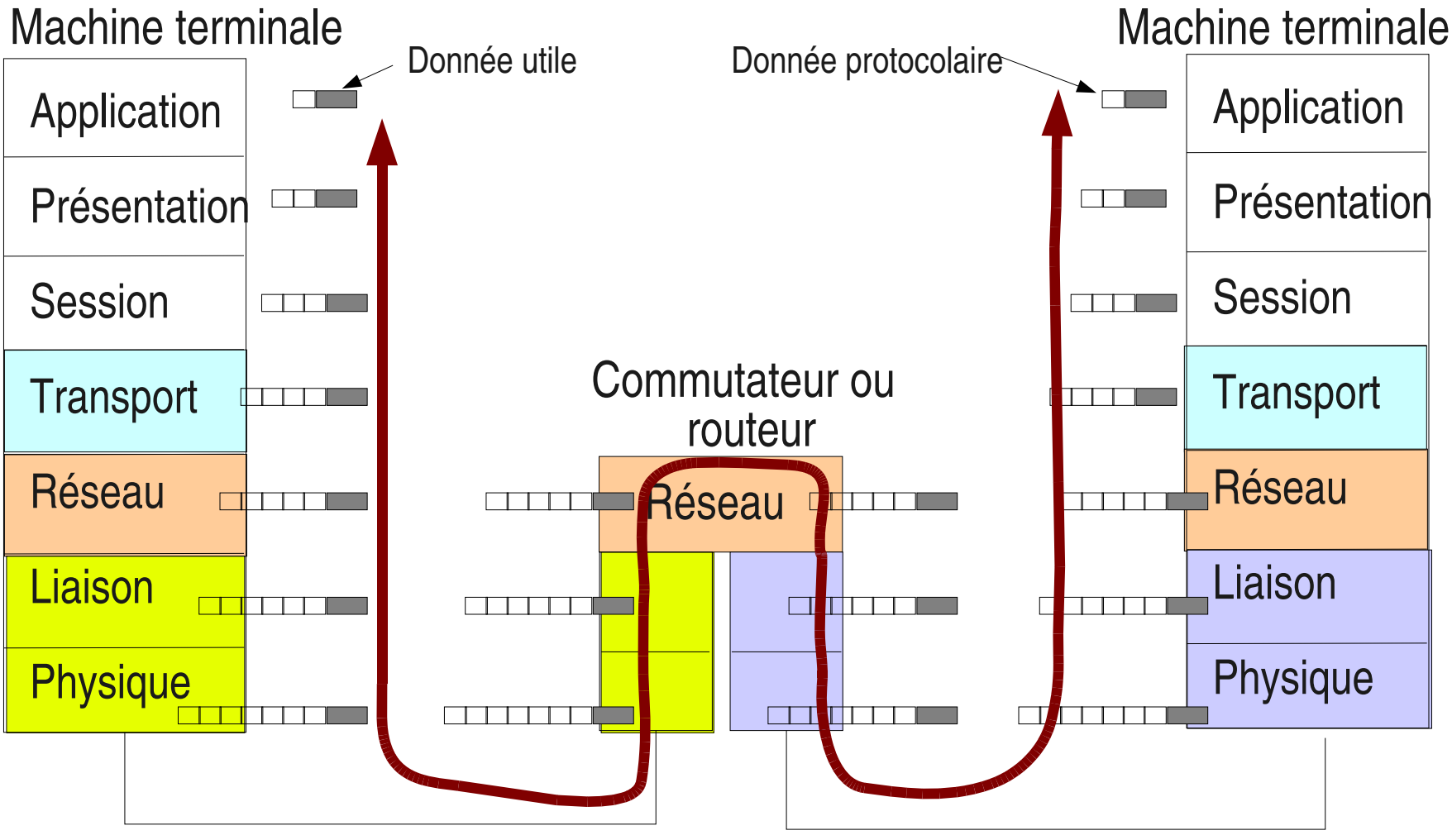
Où sont implémentées les couches ?



Notion de passerelle ou proxy



Chaque couche rajoute ses données protocolaires



Problématiques

- Fonction première : transport des bits sur un support analogique
  - Modulation de signal sinusoïdal (fréquence porteuse) en amplitude et en phase par modems
  - Codage en tension : le plus simple mais limité en distance et en débit. Fragile face aux perturbations électromagnétiques
  - Codage en courant : 2 fils par signal, plus résistant aux perturbations électromagnétiques. Permet de monter très haut en débit (100Mb/s et plus)
  - Codage photonique : sur fibre optique, des impulsions de lumière
- Problème connexe : synchronisation parfaite des horloges des récepteurs sur les horloges des émetteurs

- L'idéal :
  - l'horloge du récepteur doit fonctionner à la même fréquence que celle de l'émeteur
  - l'horloge du récepteur doit fonctionner avec un décalage de phase constant avec celle de l'émetteur (décalage de π pour que le front montant coté recepteur ait lieu en milieu de bit reçu)

- La réalité physique
  - les conditions énoncées ci-dessus sont impossibles à réaliser
  - le récepteur fonctionne avec une horloge qui lui est propre. Il est impossible physiquement de fournir une horloge unique pour tous les émetteurs et récepteurs du réseau. La fréquence de l'horloge récepteur sera naturellement légèrement différente de celle de l'émetteur et en plus la différence variera en fonction du temps et des conditions physiques (température du milieu, perturbations électromagnétiques du milieu).
  - même si les horloges étaient parfaitement à la même fréquence, le problème de la phase demeurerait

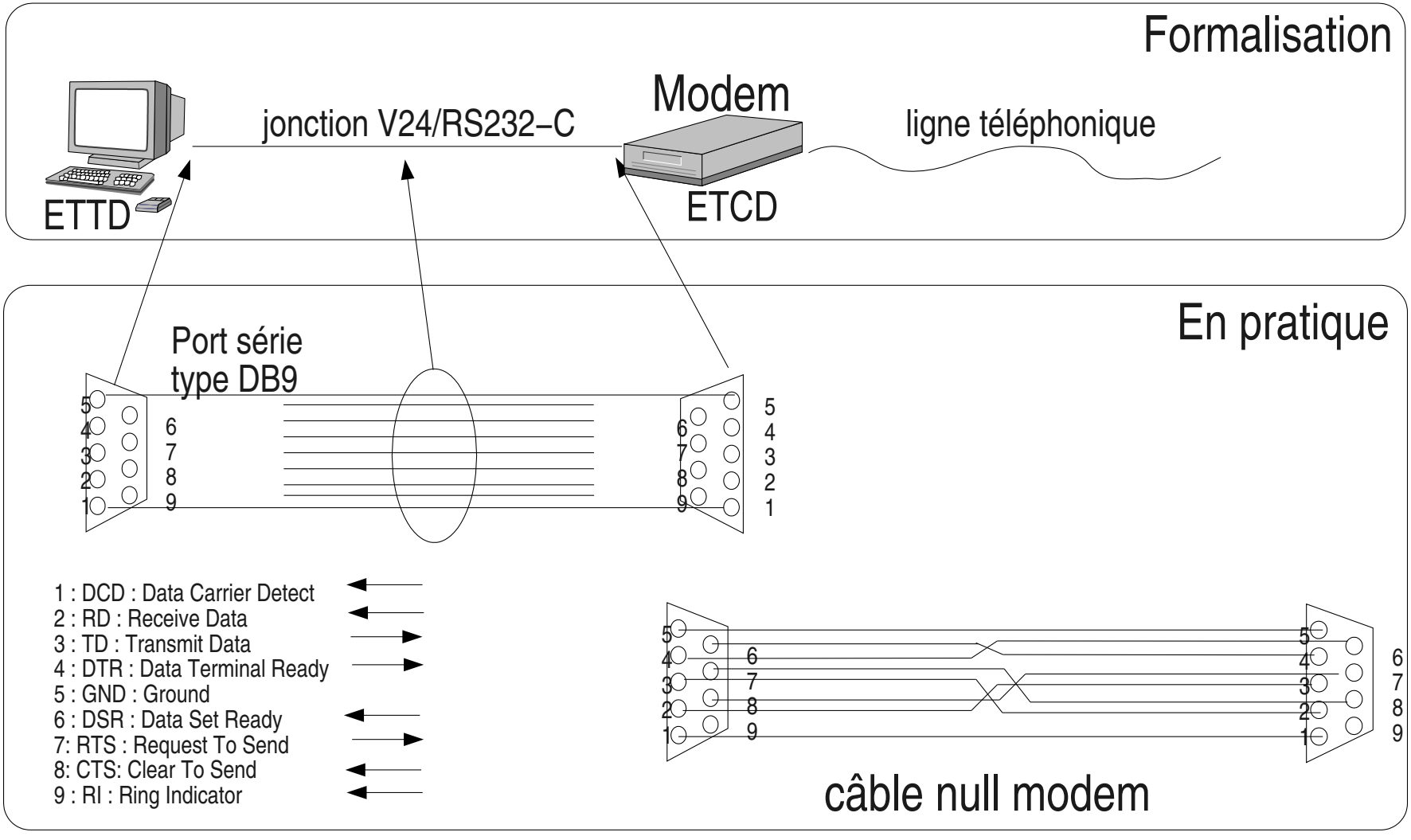
- Le principe fondamental pour remédier au problème :
  - La suite des bits de données est transcodée pour permettre de véhiculer simultanément un signal en phase avec l'horloge d'émission.
  - L'électronique de réception est capable de déduire l'horloge émission du flux binaire reçu. Elle génère alors un signal qui vient synchroniser parfaitement l'horloge du récepteur.

- Forts divers. Ils doivent permettre de respecter les règles suivantes :
  - pas de longues suites de 1 ou de 0, sinon on perd la synchronisation,
  - pas de composante continue dans le flux émis (la décompositon spectrale du signal fait apparaître un courant ou une tension à la fréquence 0 : courant continu). Ce point est finalement identique au précédent, des suites de 0 ou des suites de 1 sont analogues à des séquences de signal continu.

- HDB3 : (haute densité binaire d'ordre 3), les UNs sont codés alternativement +V et -V. Les ZEROs sont codés 0. Si une suite de 3 ZEROs arrive, un UN est inséré à leur suite. C'est un faux UN, pour le reconnaitre on lui donne la même polarité que le UN normal qui a précédé (viol de bi-polarité). Multiplex téléphoniques à 32 voies (2,048 Mb/s).
- AMI : Alternate Mark Inversion, même principe sauf que le bit de poids fort est toujours à 1 et qu'il n'y a pas de longue suite de 0 à priori car on porte de la voix en PCM
- bipolaire (manchester) : chaque bit est codé par une transition +V-V ou -V+V suivant la valeur du bit. Il ne faut pas croiser les fils car les bits seraient décodés inversés (les UN seraient vus commes des ZEROs et réciproquement). Réseau Ethernet.

- bipolaire différentiel (manchester diférentiel) : même principe que ci-dessus mais le codage d'un bit (une transition) dépend de la valeur du bit précédent. Le problème précédent n'existe alors plus. Réseau Token Ring.
- 4B/5B : on transmet 5 bits pour 4 bits utiles. À toute configuration de 4 bits de données on fait correspondre un mot de 5 bits (16 combinaisons, 32 mots possibles). Certains mots de 5 bits sont interdits de par leur configuration. Certains mots sont réservés pour un usage particulier de signalisation sur la liaison
- 8B/10B : principe analogue au précédent

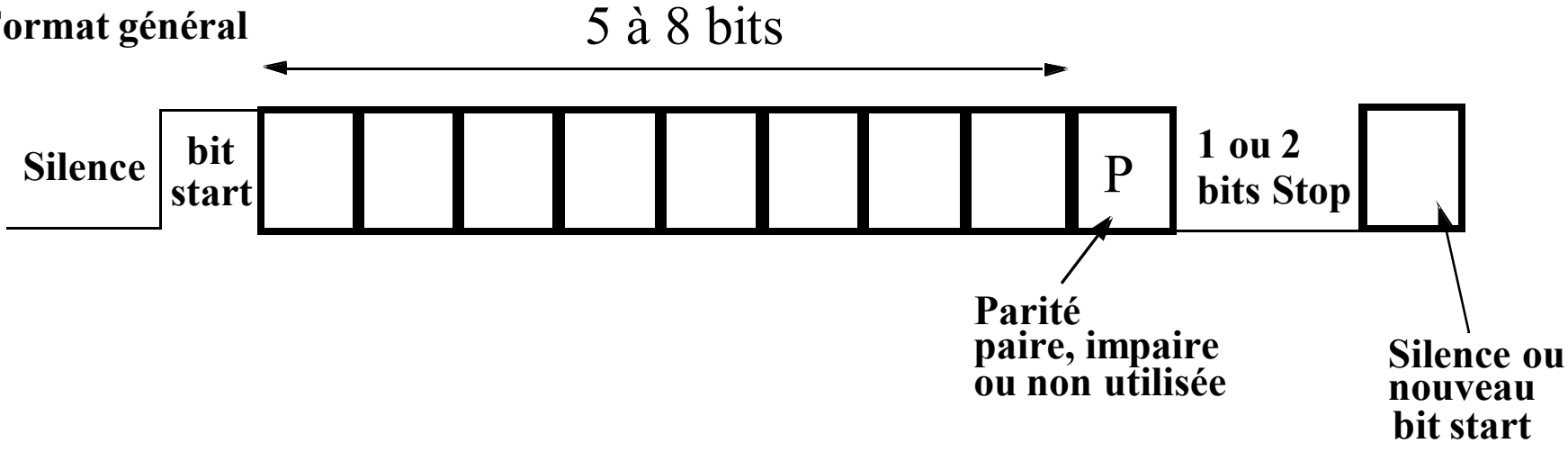
Jonction ITU-T V24 – ANSI RS232-C
Exemple de couche physique



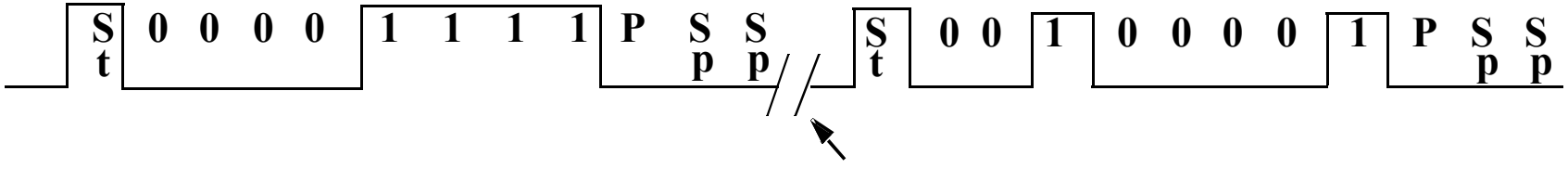
- Selon la recommandation ITU-T V28
  - Niveaux de tension
    - V $>$ +3v $\Rightarrow$ 0
    - V $<$ -3v $\Rightarrow$ 1

Structure orientée octet

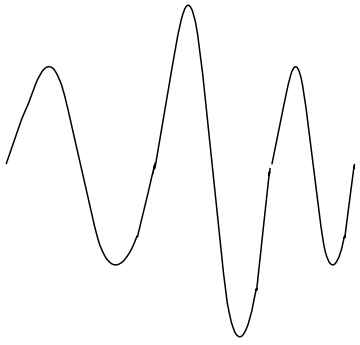
Format général



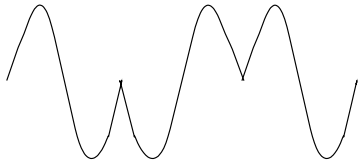
Exemple : transmission de 2 caractères en mode parité paire avec 2 bis Stop



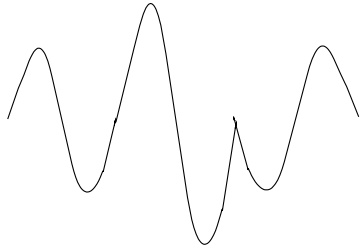
- Modulation en amplitude et phase



Modulation en amplitude



Modulation de phase

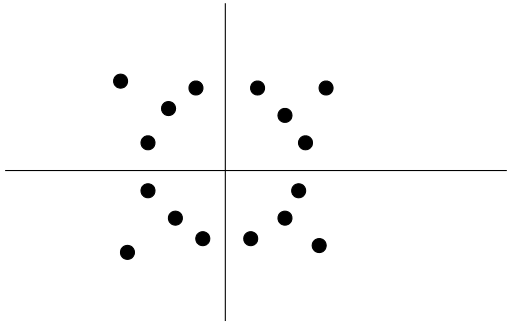


Modulation en amplitude et en phase

- Codage des bits
  - une fréquence,
  - une amplitude et une phase est un état
  - x amplitudes, y sauts de phase, N états
  - un état code $\log_2 N$ bits (exemple : pour 16 états on peut coder 4 bits par état)

Modems RTC : bits et bauds
Codage de l'information

- Le nombre de changement d'états par seconde s'appelle la rapidité de modulation, elle s'exprime en bauds
- Le nombre de bits par seconde s'appelle le débit
- La relation entre le débit et la rapidité de modulation s'exprime par la relation suivante :
  - $D = R \log_2 N$ (N est le nombre d'états)
- Ce type de modulation se nomme «modulation d'amplitude à quadrature de phase» ou QAM (Quadrature Amplitude Modulation) et peut se représenter ainsi :



Phases espacées de 30° (12 phases)
Une amplitudes pour toutes les phases sauf 45°, 135°, 225° et 315° qui ont deux amplitudes.
16 états, 4 bits par états.
Si R = 2400 bauds, D = 9600 b/s

Modems RTC : modulation et débit
Codage de l'information

- Le débit est exprimé en bits par secondes
- Sa valeur maximum dépend des caractéristiques physiques du canal et de la capacité de l'électronique de réception de reconnaître le signal utile dans le signal composite formé par le signal lui même et les perturbations électromagnétiques incontournables (ce qu'on appelle le bruit)
- Les caractéristiques physiques du canal permettent de définir une bande passante maximum : W (coupure à 3db)
- La puissance du bruit est estimée par rapport à la puissance du signal lui même dans le rapport Signal sur Bruit (S/B)
- Résultat fondamental (Shannon 1948) $D_{max} = W \log_2(1 + \frac{S}{B})$

HDLC – High-Level Data Link Control
Exemple de couche liaison

- La trame HDLC

Délimiteur	Adresse	Contrôle	Données	CRC	Délimiteur
1 octet	1 octet	1 ou 2 octets	n octets	2 octets	1 octet

- Fournir une structure en trame dans un «flot de bits» (ou d'octets)
- Le champ délimiteur est aussi appelé le fanion
- Le champ «Contrôle» est aussi appelé «Commande»
- Pour que la structure du fanion ne se retrouve pas dans les données, chaque série de cinq «1» successifs est suivie par un «0» rajouté. Technique dite du «bit stuffing»

Protocole HDLC-LAPB
Exemple de couche liaison

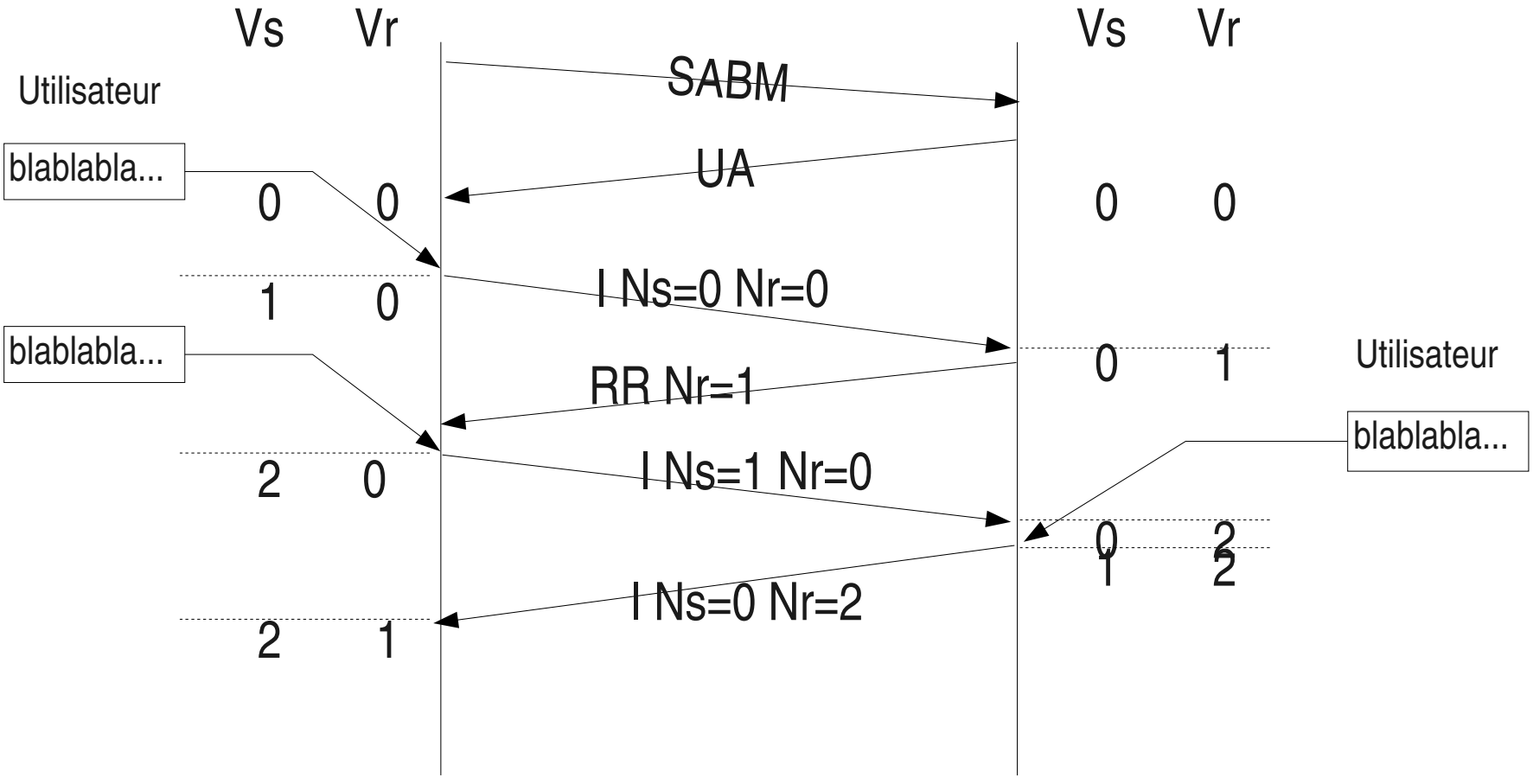
- Une spécialisation des trames HDLC (porter X.25)
- Adresse : 0xC0 ou 0x80 suivant le type Commande ou Réponse et le sens de la trame (ETTD→ETCD ou l'inverse)
- Différents types de trames :
  - I (*Information*) : trames portant des numéros et des acquittements
    - le champ contrôle est sur 1 octet (numérotation modulo 8) ou sur 2 octets (numérotation modulo 128)
  - S (*Supervision*) : trames ne portant que des acquittements
    - le champ contrôle est sur 1 ou 2 octets suivant le modulo de numérotation
  - U (*Unumbered*) : trames de commande
    - le champ contrôle est sur 1 octet
    - trames servant à établir et rompre la relation (communication) entre les deux extrémités de la liaison

TABLEAU 2-7/X.25

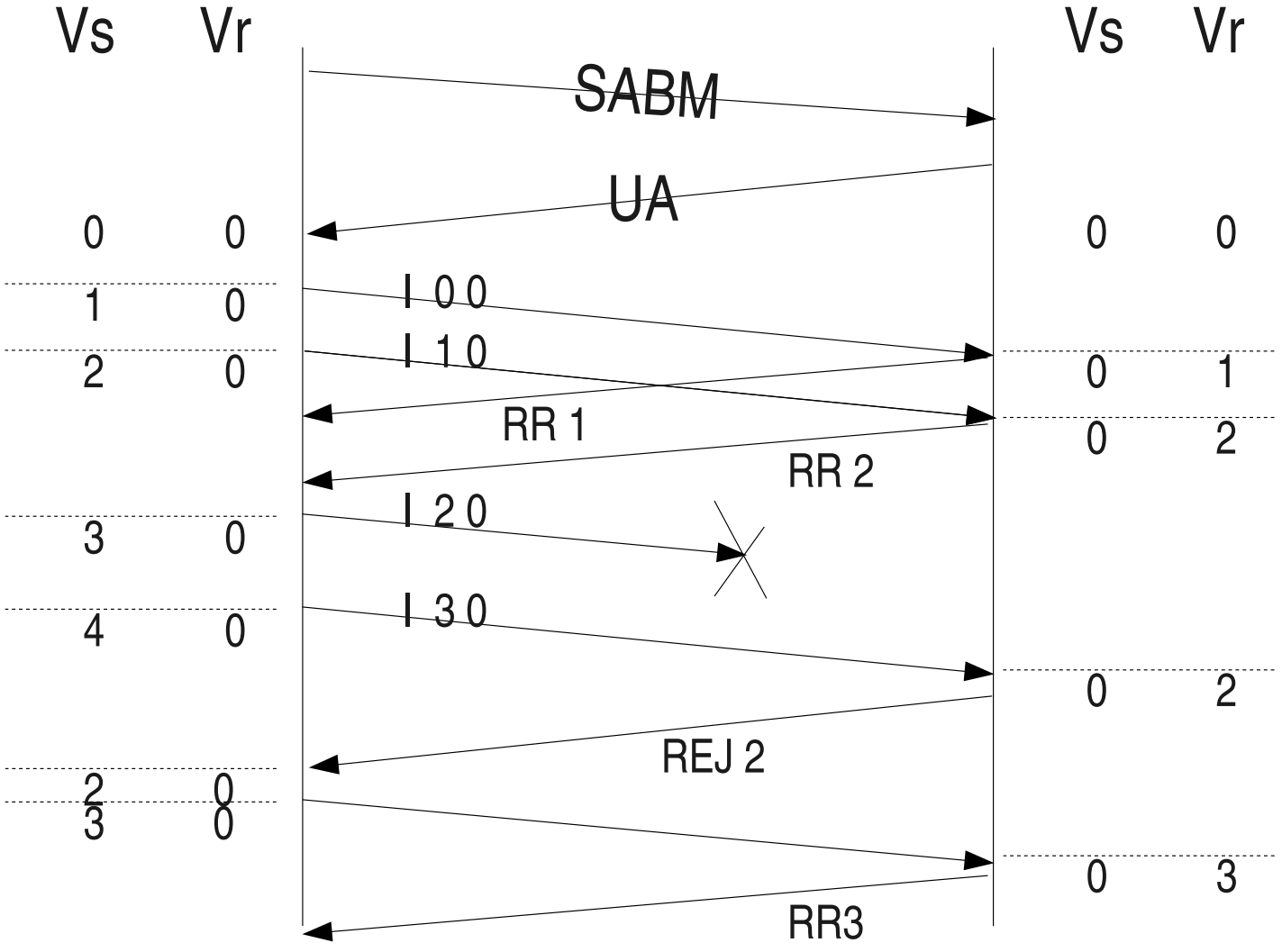
Commandes et réponses LAPB – Fonctionnement de base (modulo 8)

			1	2	3	4	5	6	7	8
Format	Commandes	Réponses	Codage							
Transfert d'information	I (Information)		0	N(S)			P	N(R)		
Supervision	RR (Prêt à recevoir)	RR (Prêt à recevoir)	1	0	0	0	P/F	N(R)		
	RNR (Non prêt à recevoir)	RNR (Non prêt à recevoir)	1	0	1	0	P/F	N(R)		
	REJ (Rejet)	REJ (Rejet)	1	0	0	1	P/F	N(R)		
Non numéroté	SABM (Mise en mode asynchrones symétrique)		1	1	1	1	P	1	0	0
	DISC (Déconnexion)		1	1	0	0	P	0	1	0
		DM (Mode déconnecté)	1	1	1	1	F	0	0	0
		UA (Accusé de réception non numéroté)	1	1	0	0	F	1	1	0
		FRMR (Rejet de trame)	1	1	1	0	F	0	0	1

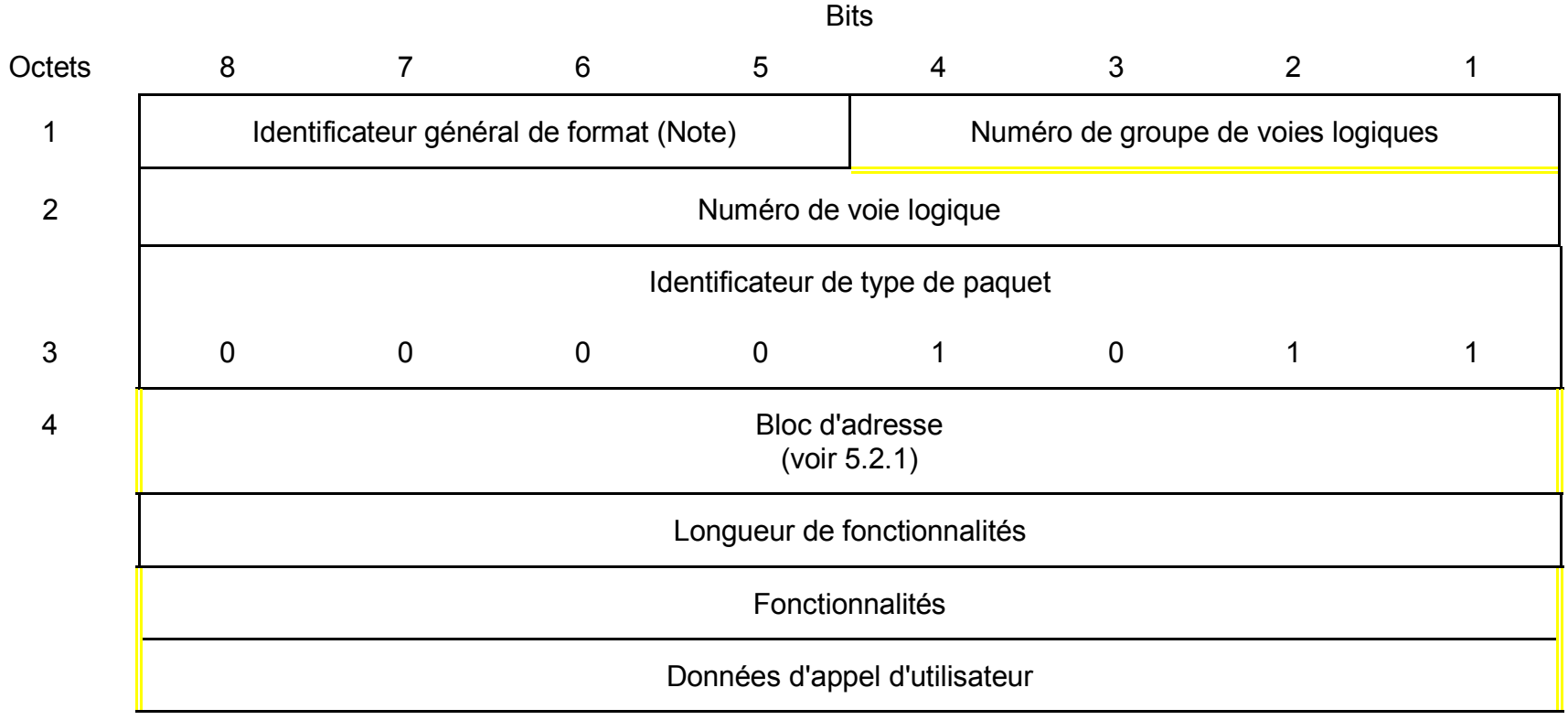
Protocole HDLC-LAPB
Exemple de scénario simple



Protocole HDLC-LAPB
Exemple de scénario plus complexe



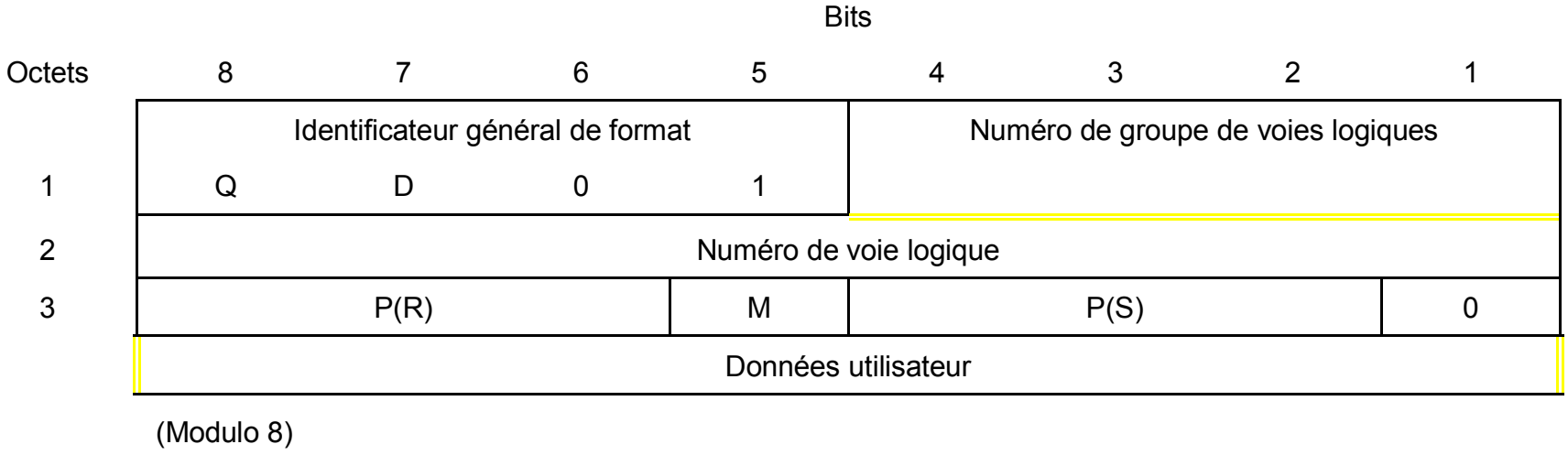
X.25 commutation de paquets en p-à-p
Les paquets d'appels



(Modulo 8 et modulo 128)

NOTE – Codé XX01 (modulo 8) ou XX10 (modulo 128).

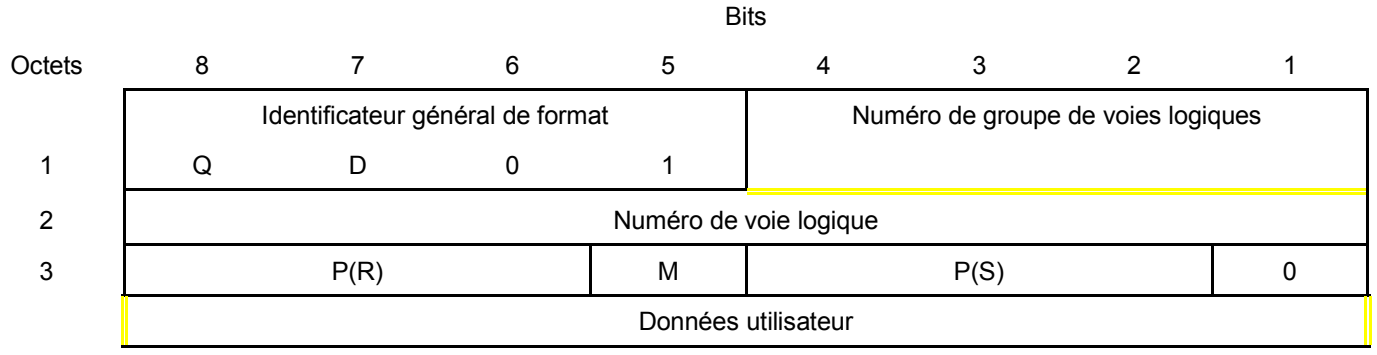
X.25 commutation de paquets en p-à-p
Les paquets de données



(Modulo 8)

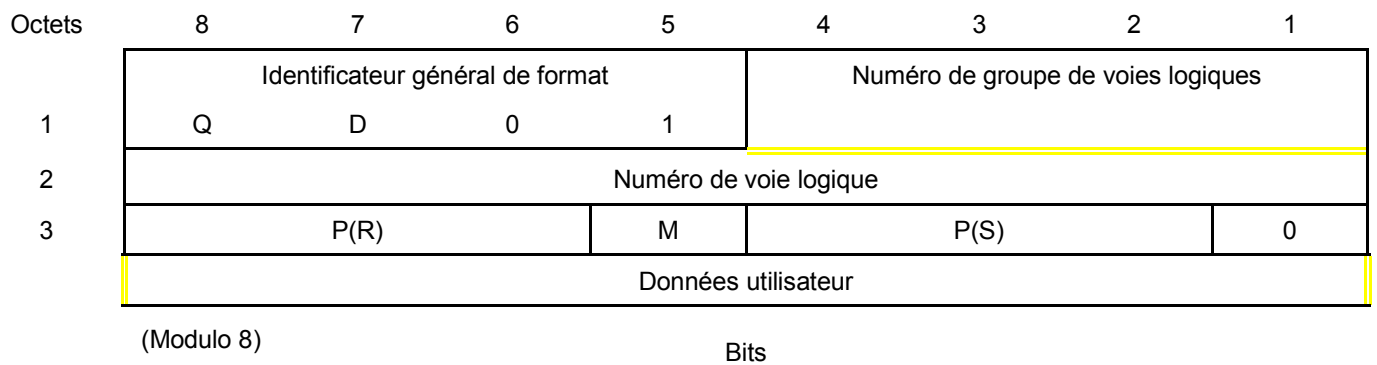
X.25 commutation de paquets en p-à-p
Les paquets de supervision

RR



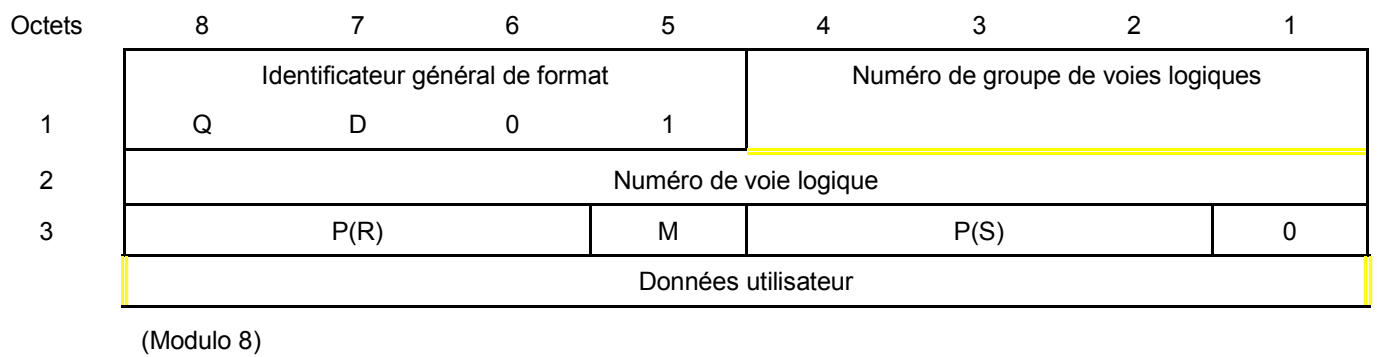
(Modulo 8)

RNR



(Modulo 8)

REJ



(Modulo 8)

X.25 commutation de paquets en p-à-p
Tous les types de paquets

Type de paquet		Service	
de l'ETCD vers l'ETTD	de l'ETTD vers l'ETCD	VC	PVC
Etablissement et libération des communications (Note 1)			
Appel entrant	Demande d'appel	X	
Communication établie	Communication acceptée	X	
Indication de libération	Demande de libération	X	
Confirmation de libération par l'ETCD	Confirmation de libération par l'ETTD	X	
Données et interruption (Note 2)			
Données de l'ETCD	Données de l'ETTD	X	X
Interruption par l'ETCD	Interruption par l'ETTD	X	X
Confirmation d'interruption par l'ETCD	Confirmation d'interruption par l'ETTD	X	X
Contrôle de flux et réinitialisation (Note 3)			
RR par l'ETCD	RR par l'ETTD	X	X
RNR par l'ETCD	RNR par l'ETTD	X	X
	REJ par l'ETTD [a)]	X	X
Indication de réinitialisation	Demande de réinitialisation	X	X
Confirmation de réinitialisation par l'ETCD	Confirmation de réinitialisation par l'ETTD	X	X
Reprise (Note 4)			
Indication de reprise	Demande de reprise	X	X
Confirmation de reprise par l'ETCD	Confirmation de reprise par l'ETTD	X	X
Diagnostic (Note 5)			
Diagnostic [a)]		X	X

VC Communication virtuelle *(virtual call)*

PVC Circuit virtuel permanent *(permanent virtual circuit)*

a) N'est pas nécessairement disponible dans tous les réseaux.

NOTES

1 Voir 4.1 et 6.16 pour les procédures, et 5.2 pour les formats.

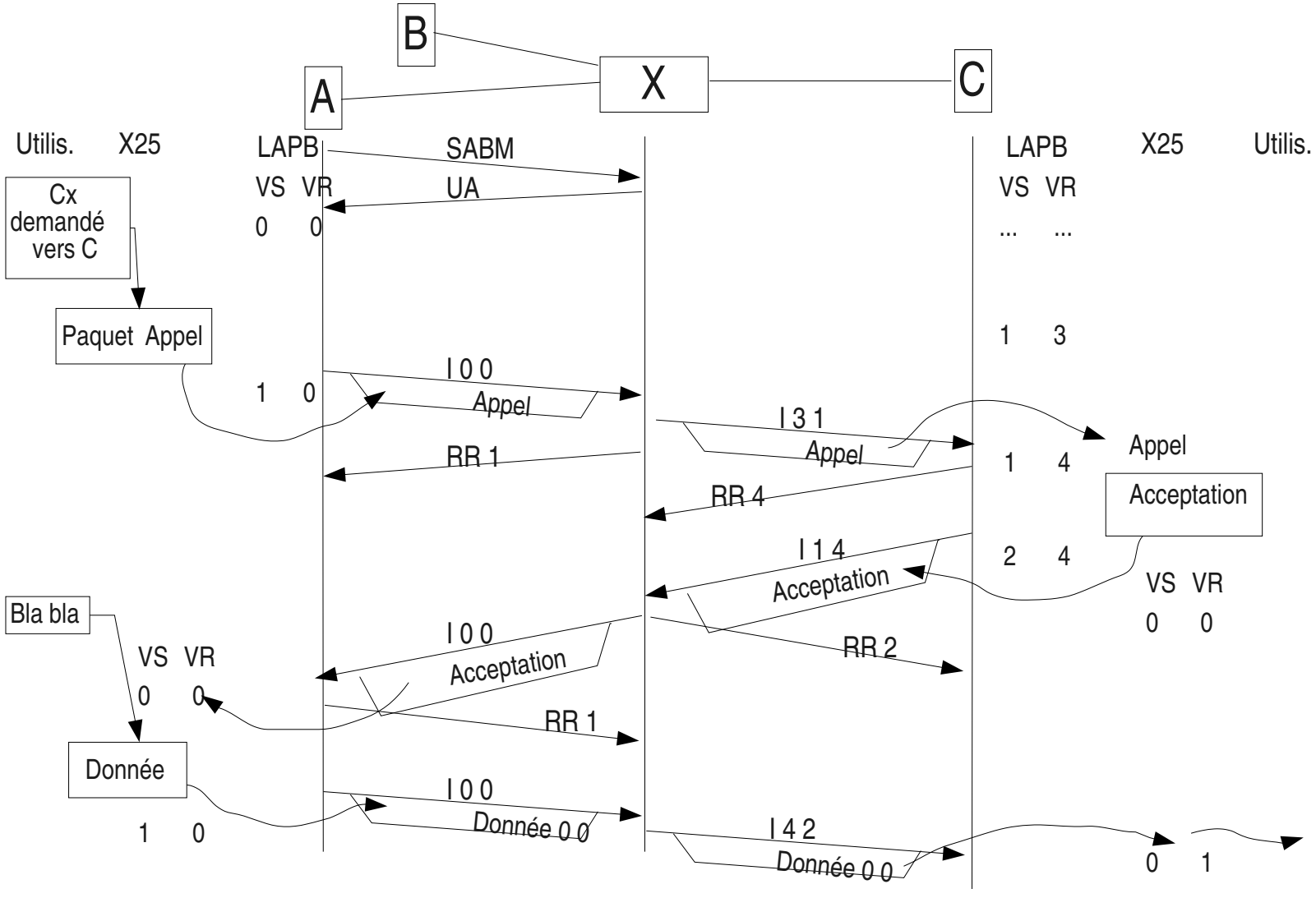
2 Voir 4.3 pour les procédures, et 5.3 pour les formats.

3 Voir 4.4 et 6.4 pour les procédures, et 5.4 et 5.7.1 pour les formats.

4 Voir 3.3 pour les procédures, et 5.5 pour les formats.

5 Voir 3.4 pour les procédures, et 5.6 pour les formats.

X.25 commutation de paquets en p-à-p
Exemple de scénario



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Deuxième partie

Réseaux locaux Ethernet et 802.3

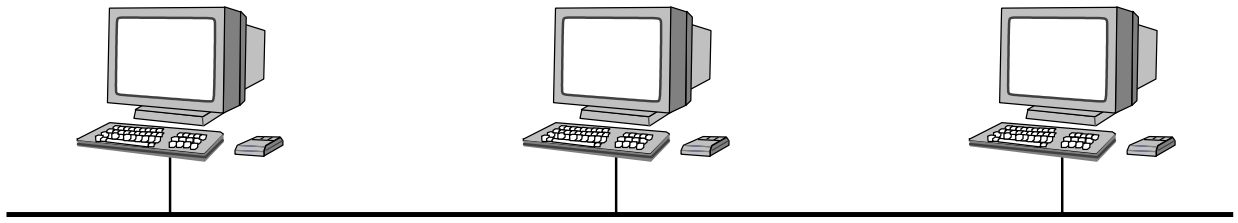
- Les réseaux de l'entreprise
- Caractéristiques :
  - Topologie
    - Bus, avec ou sans fil
    - Anneau
    - Étoile
    - Bande de base ou large bande
  - Caractéristiques physiques des supports (les média)
    - Cuivre
    - Fibre optique
    - radio
  - Méthode d'accès au médium
    - Comment une station peut-elle émettre ses données sur le réseau

Les différents types de réseaux

- Les réseaux locaux : LAN : Local Area Network
  - Réseaux d'entreprise
    - Courtes distances : de quelques centaines de mètres à quelques kilomètres
- Les réseaux métropolitains : MAN : Metropolitan Area Network
  - Interconnexion de réseaux locaux
  - Quelques dizaines à quelques centaines de kilomètres
- Les réseaux grande distance : WAN : Wide Area Network
  - Les réseaux nationaux et internationaux
  - Les réseaux d'opérateurs

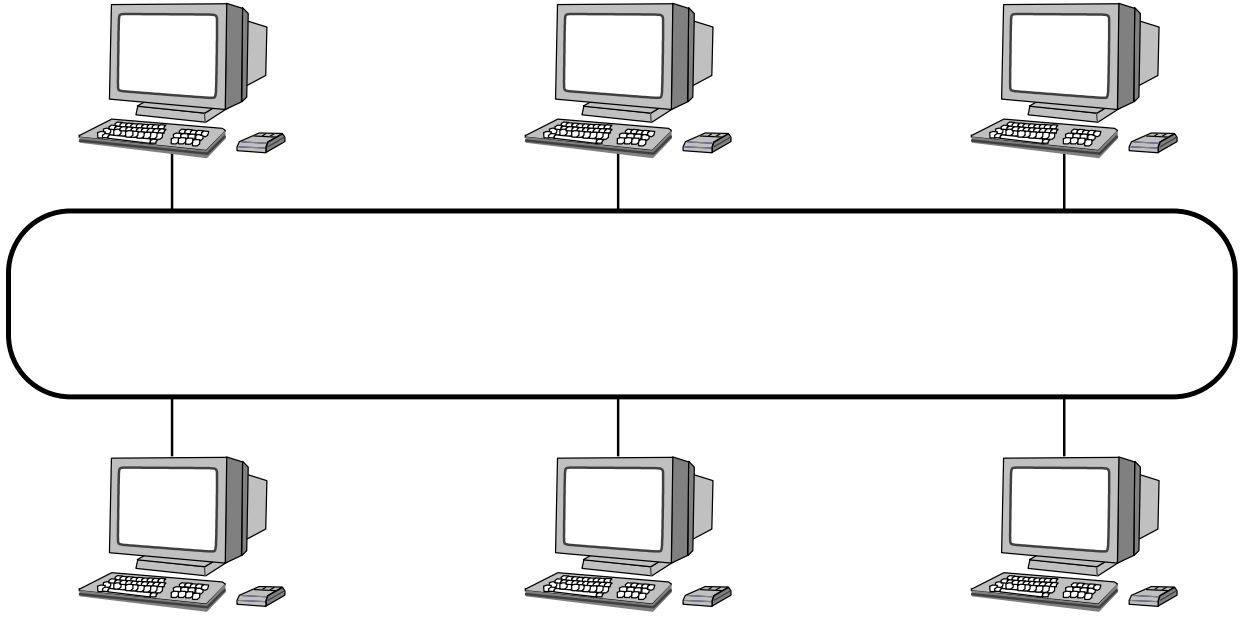
Les topologies des réseaux locaux

Bus



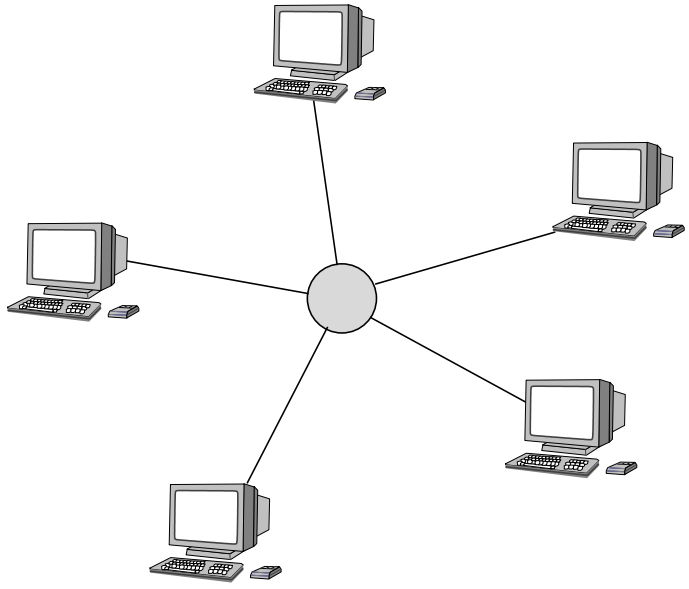
Remarque : un canal radio partagé peut être considéré comme un bus

Anneau

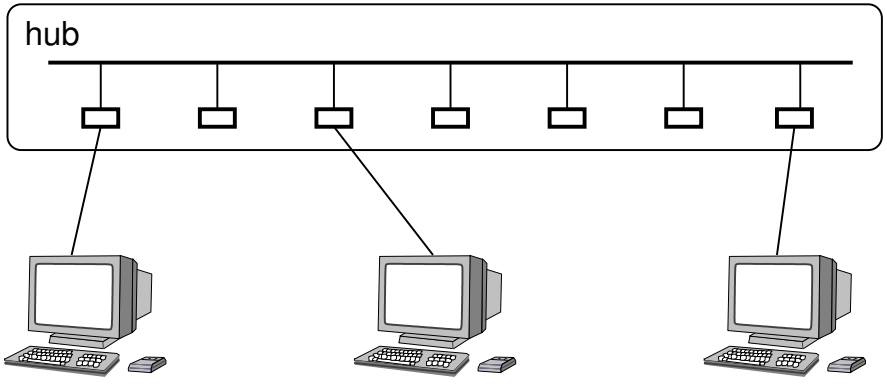


Les topologies des réseaux locaux

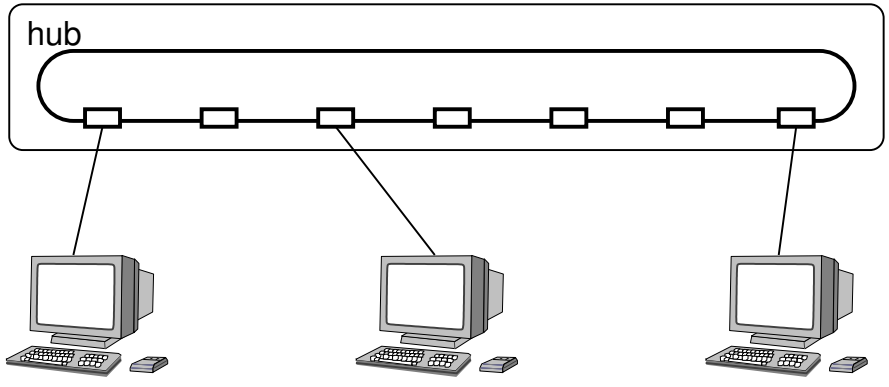
Étoile



Le bus devient étoile



L'anneau devient étoile



Caractéristiques des canaux

- Transmission en bande de base
  - Un seul canal physique pour toutes les stations
  - Problème d'accès concurrent
- Transmission en large bande
  - Plusieurs canaux sur le médium
  - Un canal est caractérisé par une bande de fréquence
  - Les signaux émis par les stations sont transposés dans la bande de fréquence qui est assignée aux stations
  - Un canal donné peut être vu comme un canal en bande de base (émulation d'un bus Ethernet par exemple)

Les méthodes d'accès au médium

Très liées à la topologie (topologie logique)

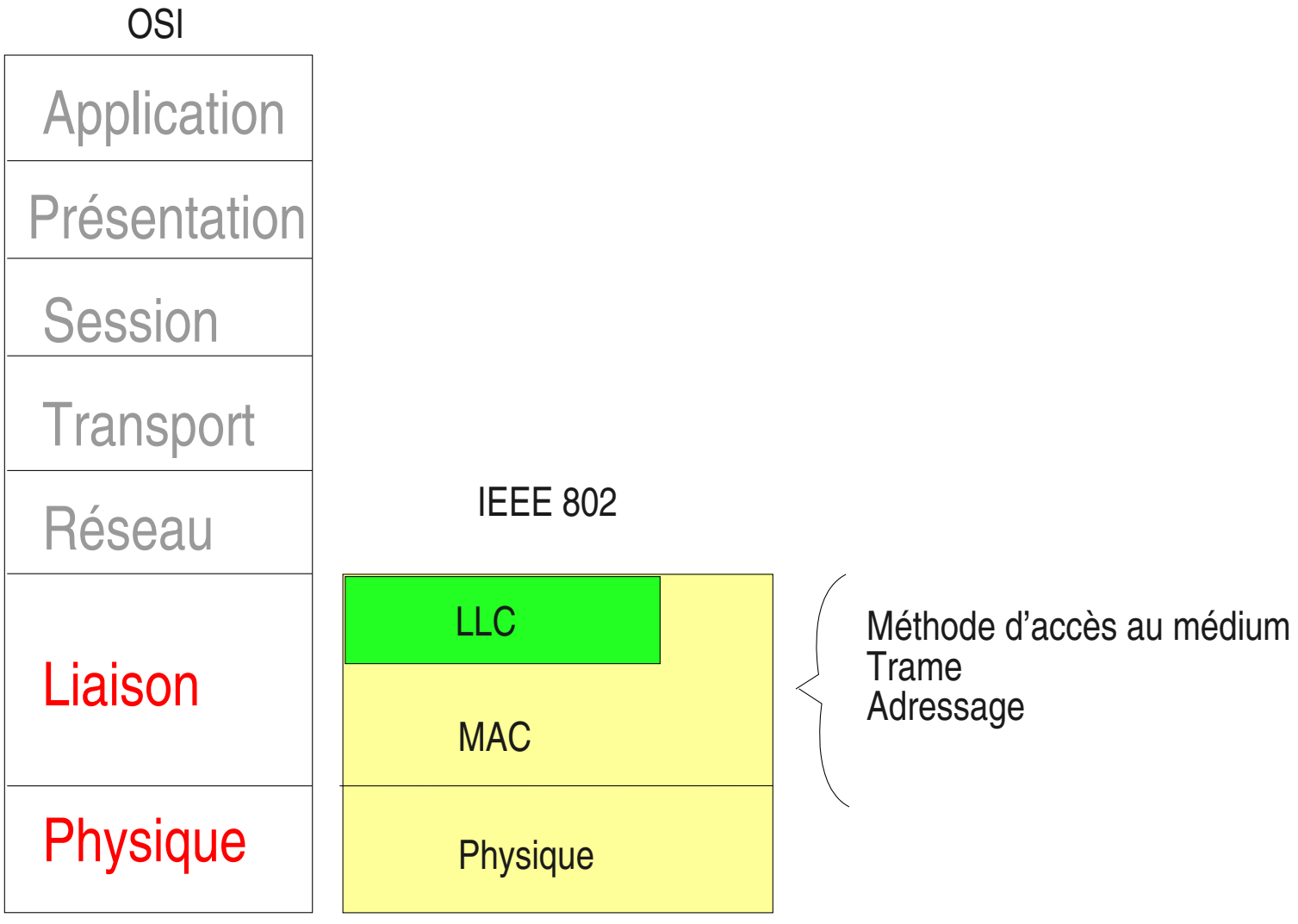
- Bus
  - Tout le monde partage le canal sans priorité
  - Chaque station peut, à priori, émettre lorsqu'elle le désire
    - Il en résulte des collisions
    - Il faut gérer le problème des collisions
      1. Algorithme spécifique de contention (CSMA CD ou CA...)
      2. Émulation d'un anneau logique
- Anneau
  - Le droit d'émettre est transmis sur l'anneau dans une trame spécifique contenant un bit spécial appelé «jeton»
  - Les trames circulent dans un sens donné

- Menée à bien par l'organisme IEEE
  - Plus particulièrement le comité 802 de l'IEEE
    - Chaque topologie et les divers protocoles et caractéristiques sont étudiés et standardisés par un sous-comité 802
    - Exemples : 802.3 pour Ethernet, 802.5 pour Token Ring
- Modèle en couches spécifique
  - Comparable au modèle OSI pour ses deux premières couches
  - Trois couches
    - La couche physique (comme pour OSI)
    - La couche «Medium Access Control» (MAC)
    - La couche «Logical Link Control» (LLC)

Quelques sous-comités IEEE-802

- 802.1 : architecture des réseaux locaux
  - Architecture générale, interconnexion (niveau 2), QoS, etc...
- 802.2 : la couche LLC
- 802.3 : Ethernet
  - 802.3u : Ethernet 100Mb/s
  - 802.3ab, z : Ethernet 1Gb/s
  - 802.3ae : Ethernet 10Gb/s
- 802.4 : le bus à jeton
- 802.5 : l'anneau à jeton (Token Ring)
- 802.11 : les réseaux sans fils (WiFi)
- 802.15 : les WPAN : Wireless Personal Area Network

Les couches du «modèle IEEE-802»



Les méthodes d'accès sur Bus

- L'ancêtre : ALOHA de l'université de Hawaï
  - Tout le monde a le droit d'émettre quand il veut
  - Les collisions sont nombreuses
- Améliorations CSMA Carrier Sense Multiple Access
  - On écoute le canal, s'il est silencieux on peut émettre
  - Les collisions ne sont pas absentes, elles sont moins nombreuses
- Méthodes de contention des collisions
  - CA : Collision Avoidance
    - On envoie une trame test (TRS : Request To Send), si elle ne collisionne pas, on peut émettre
  - CD : Collision Detection
    - On émet et on écoute, il y a collision si le signal écouté est différent de celui qu'on émet, on arrête l'émission qu'on retente après un temps aléatoire

La technologie Ethernet

- La technologie impérialiste, elle a écrasé toutes les autres (ou les autres s'adaptent à elle, exemple le WiFi)
- Topologie logique : le bus
  - Aujourd'hui, topologie physique en étoile avec hubs et commutateurs (switches)
- Méthode d'accès
  - CSMA-CD : Carrier Sense Multiple Access with Collision Detection
- Origine : Intel, Xeros, Digital, première idée en 1976 (Bob Metcalfe)
- Standardisation générale : IEEE-802.3
- Des débits divers : 10, 100, 1Gb/s, voir 10Gb/s

La méthode d'accès CSMA-CD I

- La station A écoute le réseau, il n'y a pas de signal, elle peut émettre
- Le signal se propage
- La station B écoute le réseau, le signal de A ne lui est pas encore parvenu, elle décide d'émettre
- Les deux signaux vont collisionner, le signal résultant va se propager de part et d'autre et va parvenir aux deux stations qui continuent d'écouter
- Chaque station continue à émettre quelques instants pour renforcer la collision et s'arrêtent
- Chacune tire un temps aléatoire au bout duquel elles tentent une ré-émission

- Si une station émet pendant au moins 1 RTT alors il n'y aura plus de collision non détectée
- Si une collision est détectée, chaque station en cause arrête son émission
  - Chaque station considère alors 2 intervalles de temps de valeur RTT et tire aléatoirement 1 ou 2 et ré-émet tout de suite (si 1 est tiré) ou un RTT plus tard (si 2 est tiré)
  - Si une nouvelle collision intervient, on considère alors 4 RTT, et on tire aléatoirement entre 1 et 4. On tente une émission au début de l'intervalle de temps tiré
  - En Ethernet, on peut tenter jusqu'à 16 réémissions mais on ne multiplie par 2 que jusqu'à 10 fois le nombre de RTT

Bus Ethernet : quel est le diamètre du réseau ?

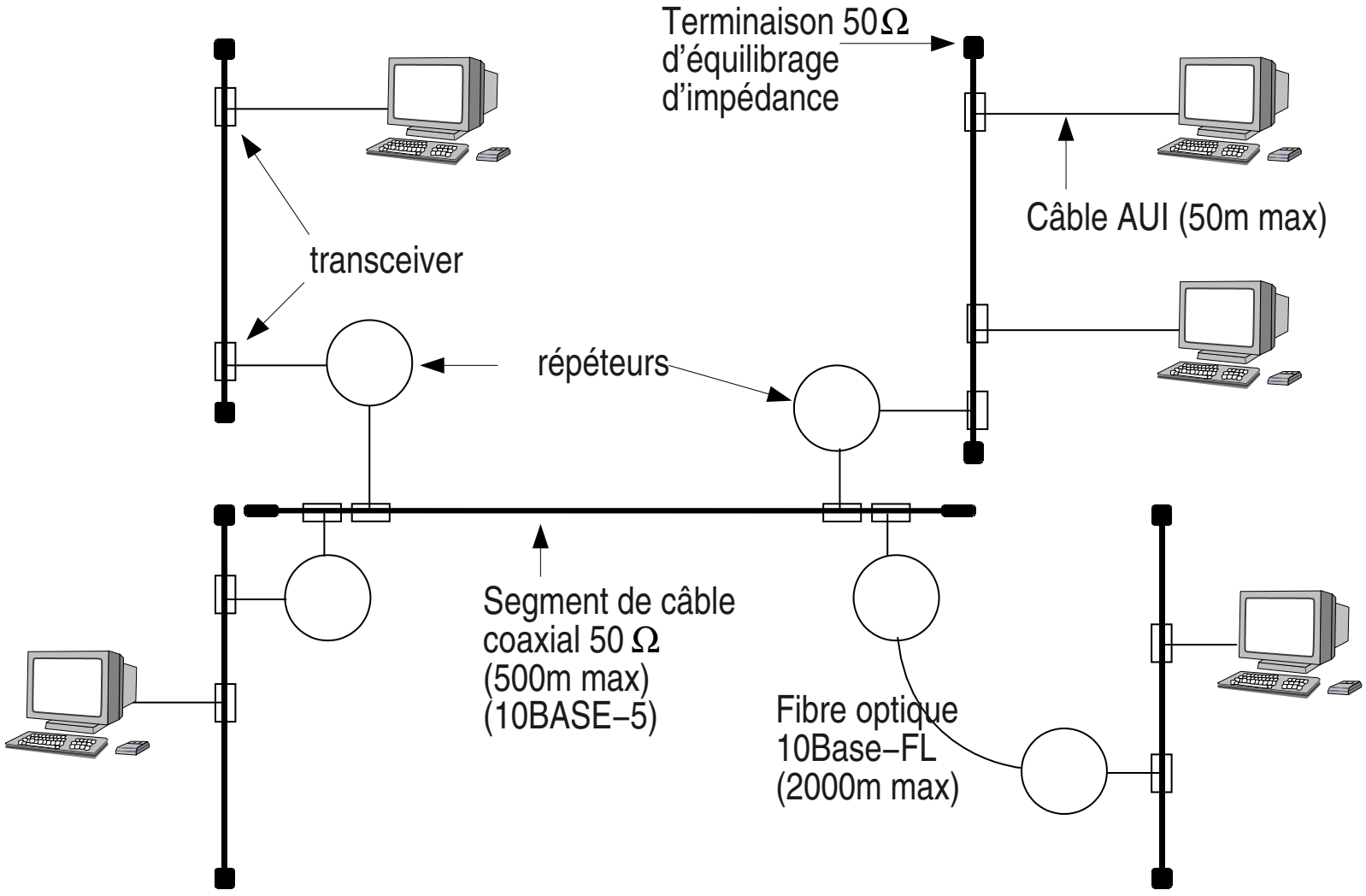
Ou encore : quelle est la distance maximale entre deux stations ?

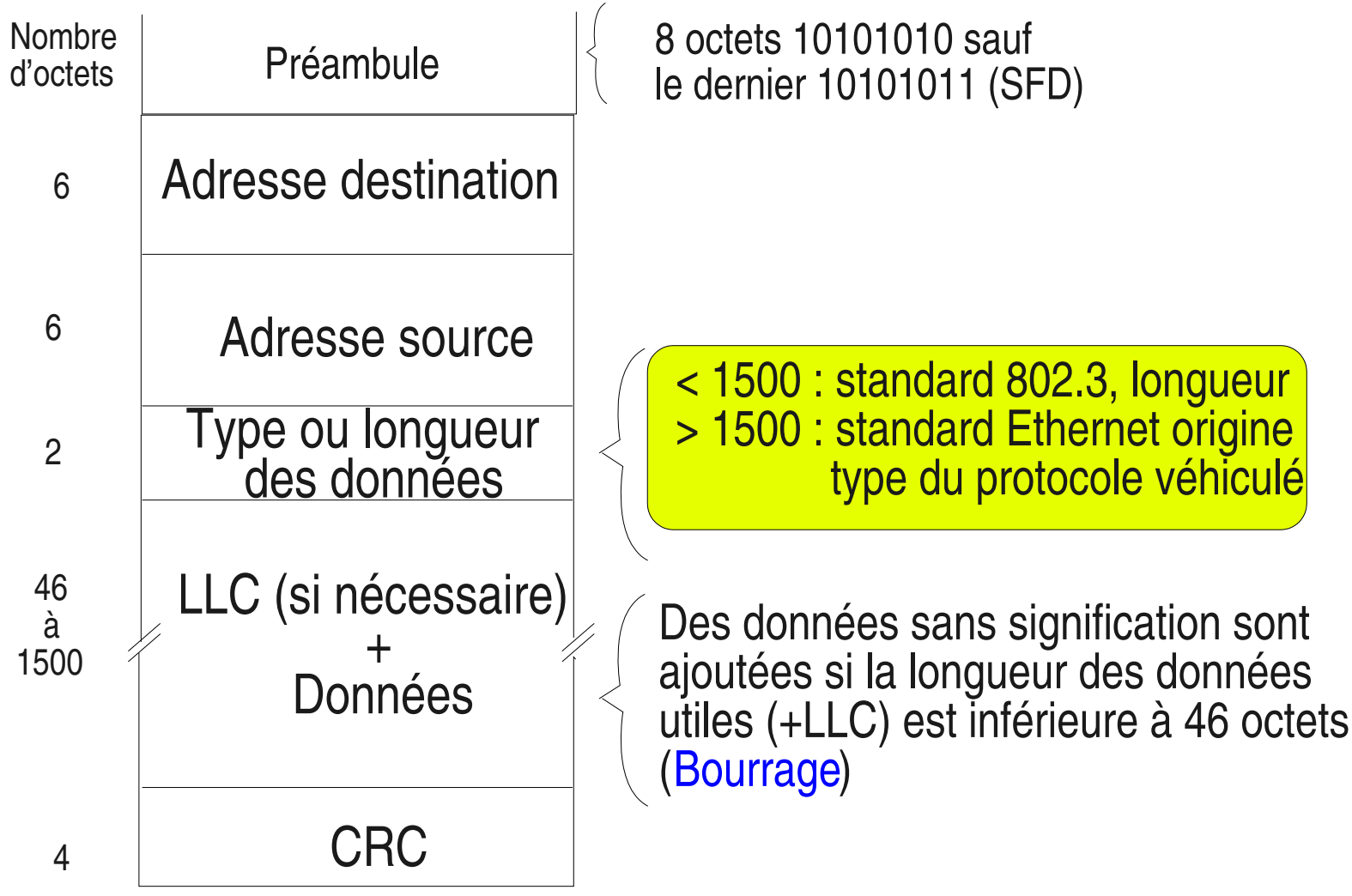
- Pour détecter une collision il faut que les deux stations soient encore en émission lorsque le signal de collision leur revient
- Cas le plus défavorable : elles sont situées aux deux extrémités du réseau et la station B émet un court instant avant que le signal de A ne lui parvienne. La collision a lieu près de B
  - Pour que A détecte la collision il lui faut attendre un temps égal à la durée de propagation de A jusqu'à B et retour
  - On appelle ce temps le Round Trip Delay, le Round Trip Time (RTT), la «tranche canal» ou encore la fenêtre de collision
  - Ce temps dépend : de la taille du segment de données (trame), de la durée d'émission de cette trame et de la distance entre les deux stations les plus éloignées

Ethernet : les caractéristiques générales au débit de 10Mb/s

- Taille minimale de la trame : 64 octets (512 bits)
- Taille maximale : 1518 octets
  - 1500 octets de charge utile (SDU) pour Ethernet pur
  - 1497 ou 1496 ou même 1492 en format 802.3 où la couche LLC est nécessaire
- Taille minimale : 64 octets
  - Bourrage dans le champ «données» si la longueur de celles-ci est inférieur à 46 octets
- Silence inter-trame de 9,6μs
- Tentatives de réémission en cas de collision : 16

Ethernet 10Mb/s en bus : l'ancêtre




Nombre d'octets
Préambule
8 octets 10101010 sauf le dernier 10101011 (SFD)
6
Adresse destination
6
Adresse source
2
Type ou longueur des données
< 1500 : standard 802.3, longueur
> 1500 : standard Ethernet origine type du protocole véhiculé
46 à 1500
LLC (si nécessaire) + Données
Des données sans signification sont ajoutées si la longueur des données utiles (+LLC) est inférieure à 46 octets (Bourrage)
4
CRC

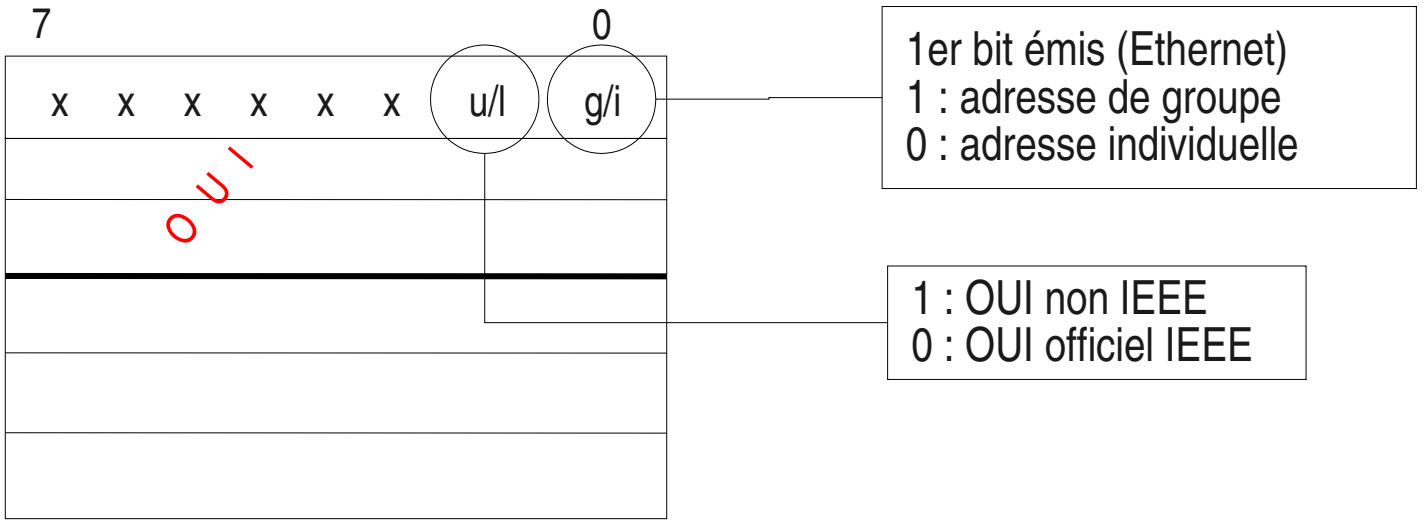
- C'est un datagramme
  - Elle contient l'adresse de la station destinatrice ainsi que l'adresse de la station qui émet
  - Elle contient un CRC, on peut donc vérifier son intégrité en réception
    - Si cette vérification montre une altération, la trame est jetée
- Il n'est pas prévu à ce niveau (MAC) d'échange préalable pour établir une relation entre l'émetteur et le récepteur
  - Il n'y a pas de connexion
  - Il 'y a pas de contrôle de flux
  - Il n'y a pas de récupération sur erreur

Les adresses Ethernet

- Définies dans la couche MAC (adresses MAC)
  - 6 octets
- 3 types d'adresses
  - Les adresses de stations, dites aussi «unicast» :
    - une adresse unique par interface matérielle (une machine peut avoir plusieurs interfaces matérielles)
    - L'adresse d'une interface est affectée par son constructeur
  - L'adresses globale, dite aussi «broadcast»
    - Permet d'envoyer une trame à toutes les stations du réseau, en une seule opération
  - Les adresses de groupes, dites aussi «multicast»
    - Permettent d'adresser un groupe de stations

Format des adresses Ethernet

- Format IEEE-48
  - 6 octets
    - Les 3 premiers affectés au constructeur par l'IEEE (OUI : Organisation Unit Identifier)
    - Les 3 derniers affectés par le constructeur



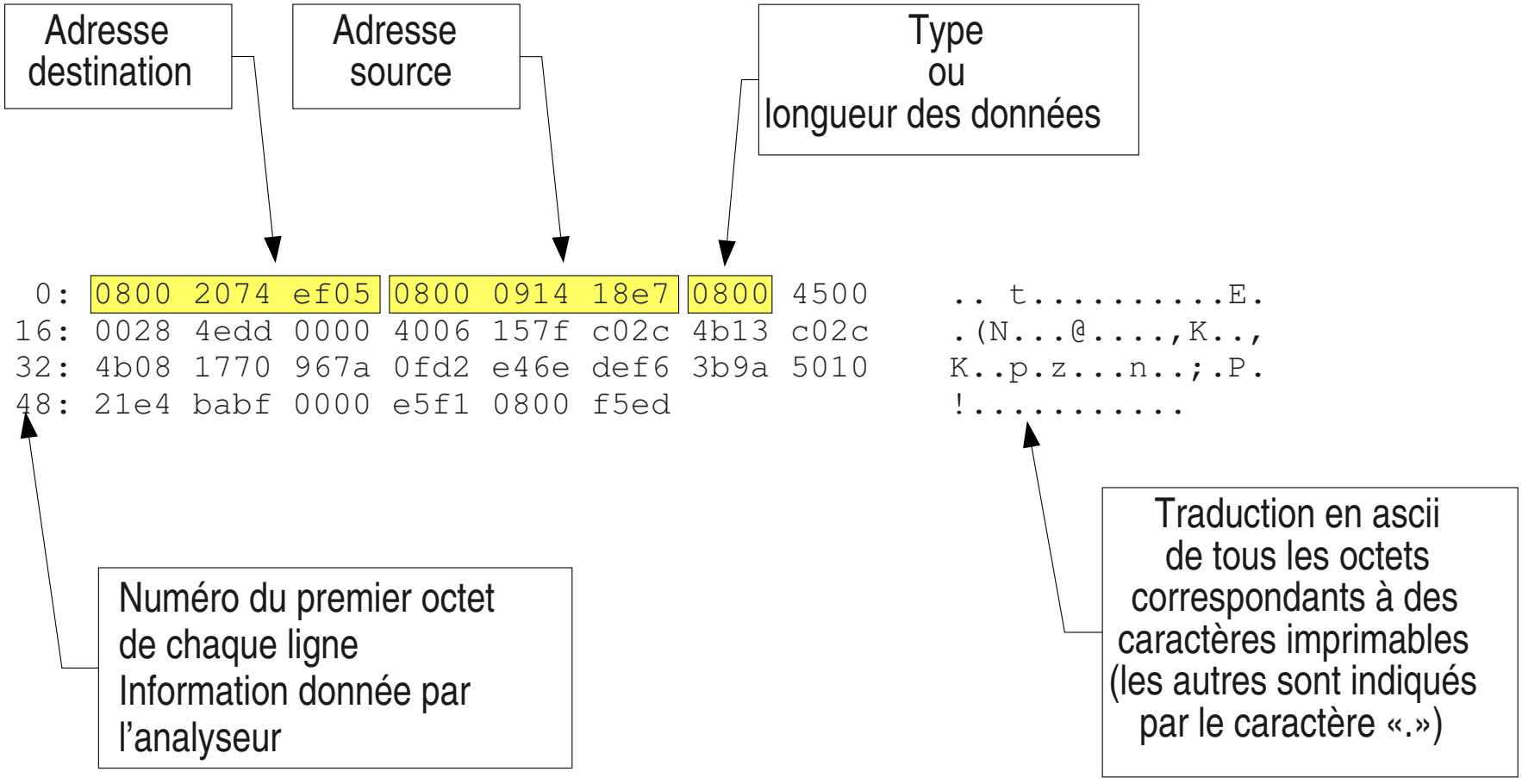
Les adresses de braodcast et de multicast

- Le groupe total : la diffusion ou broadcast
  - `ff:ff:ff:ff:ff:ff`: les 48 bits à 1
- Les groupes restreints : le multicast
  - Le bit de poids faible du premier octet est à 1
  - Quelques exemples :
    - `01:00:5E:xx:xx:xx`: multicast IP
    - `01:80:C2:00:00:00`: spanning tree (protocole de gestion automatique des ponts et commutateurs)
    - `09:00:4E:00:00:02`: multicast IPX

Exemples d'OUI, d'adresses multicats, de types

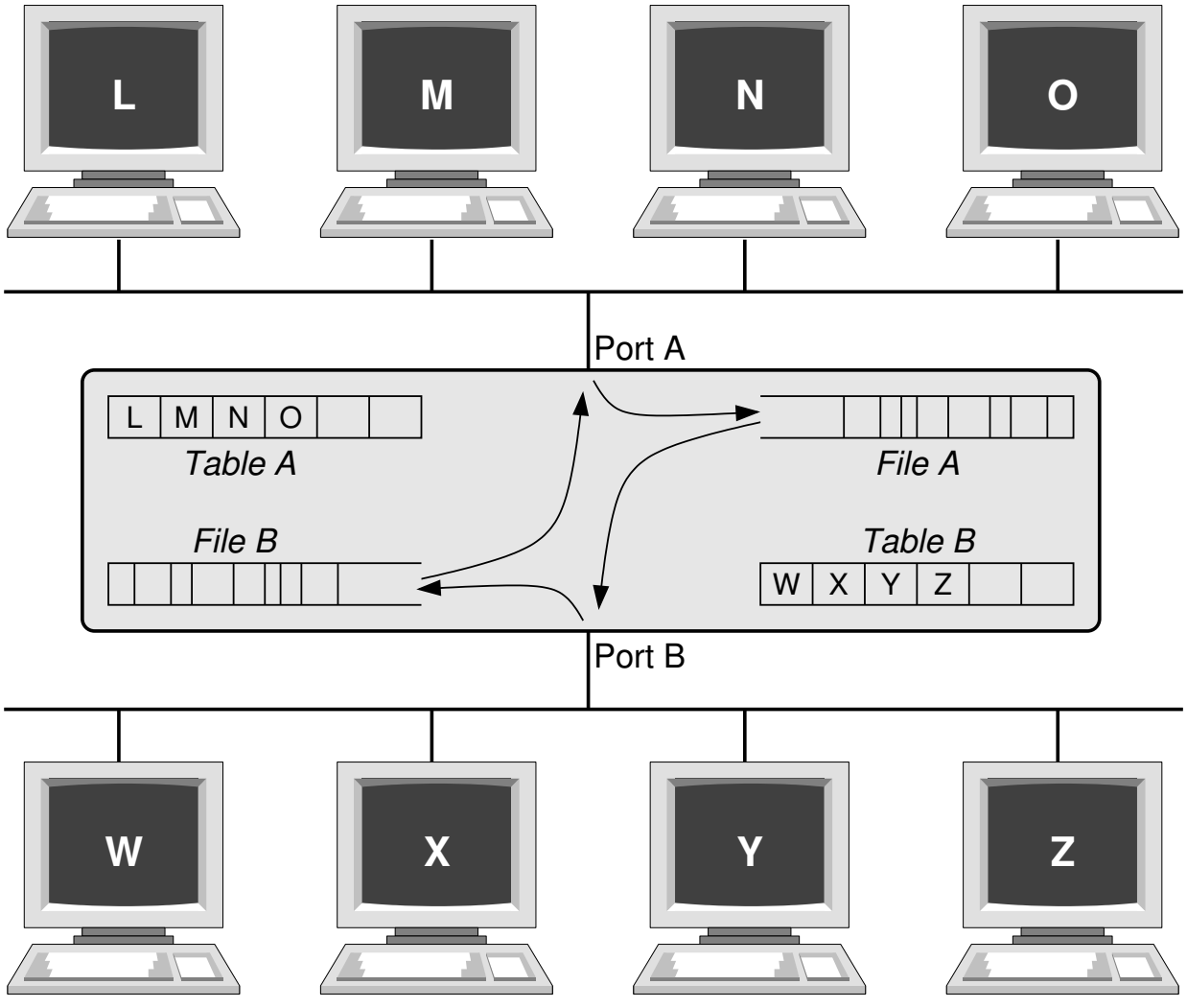
- Quelques OUI
  - 08-00-07 Apple
  - 00-00-0C Cisco
  - 08-00-08 HP
  - 08-00-20 Sun
- Quelques adresses multicast
  - 01-00-5E-xx-xx-xx Multicast internet
  - 01-80-C2-00-00-00 spanning tree

Une trame Ethernet capturée par un analyseur

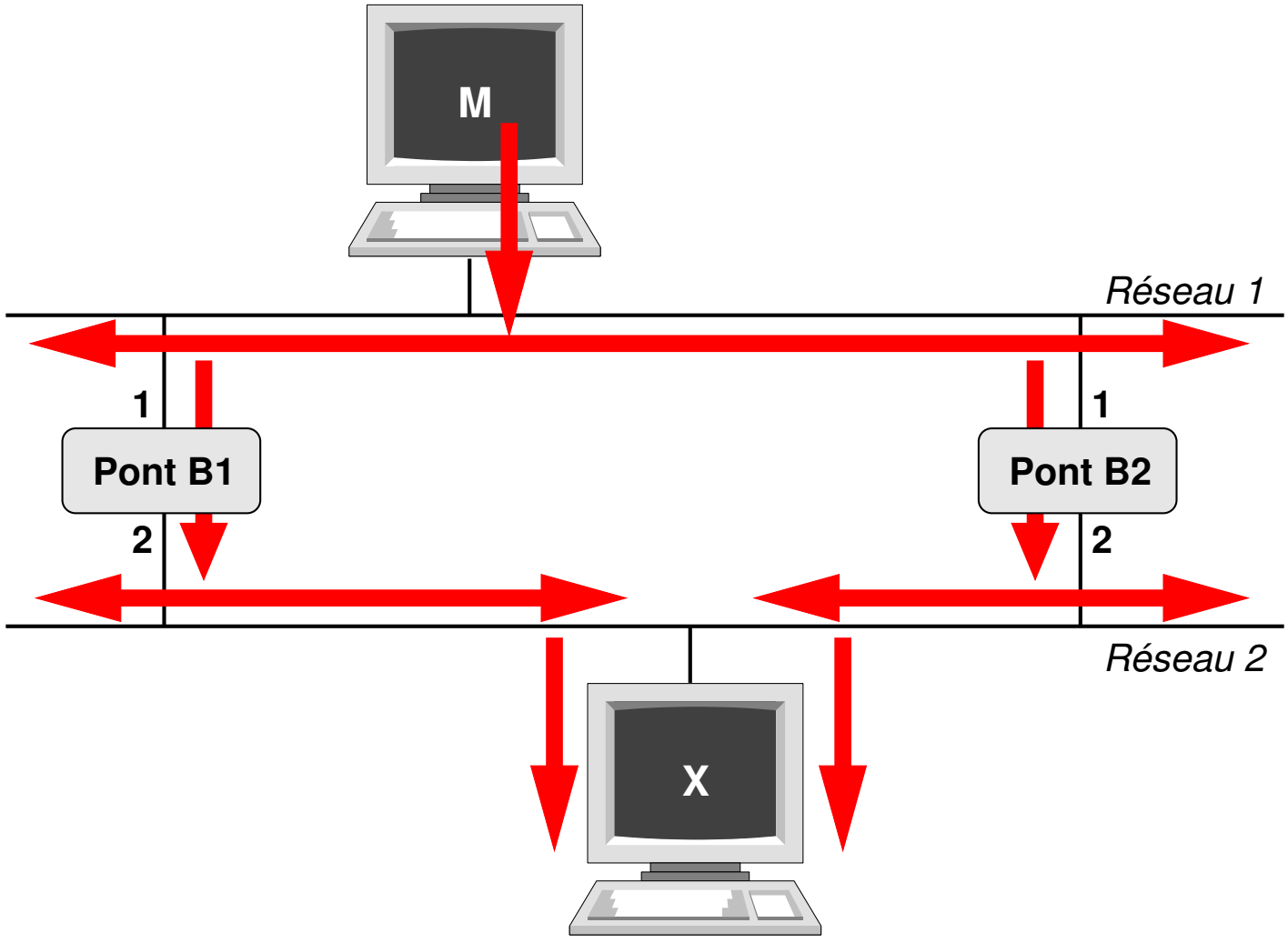


L'analyseur n'affiche pas le CRC

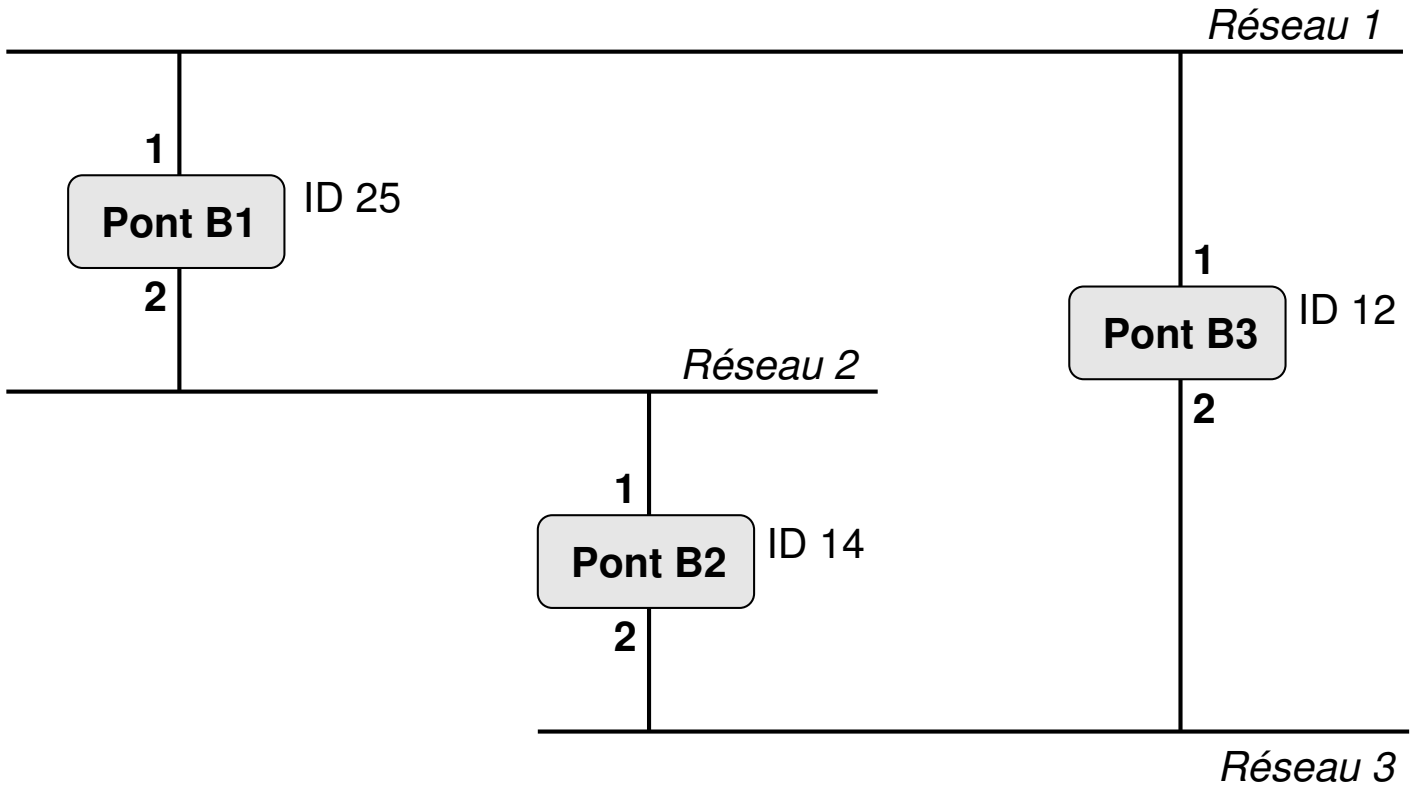
Interconnexion de réseaux locaux Ethernet : les ponts



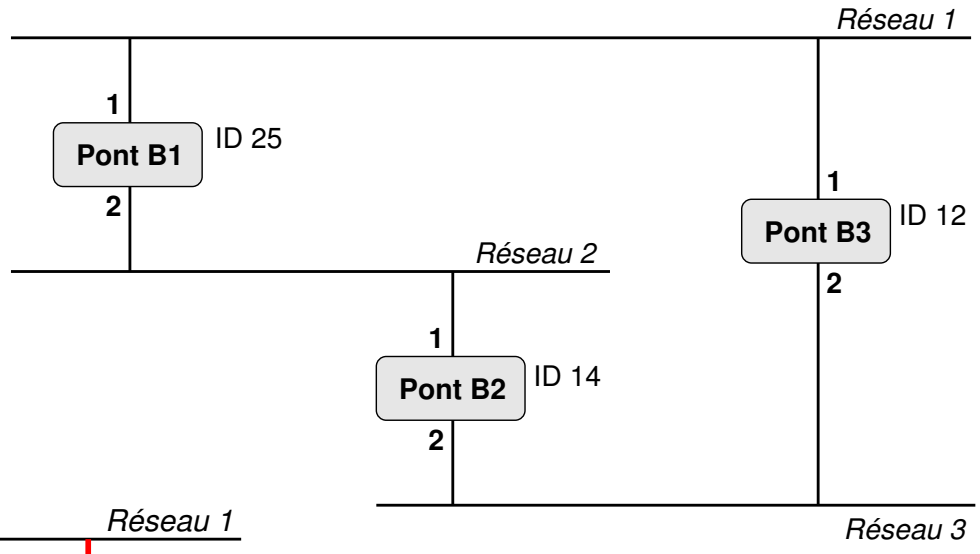
Ponts : attention aux boucles catastrophe assurée !



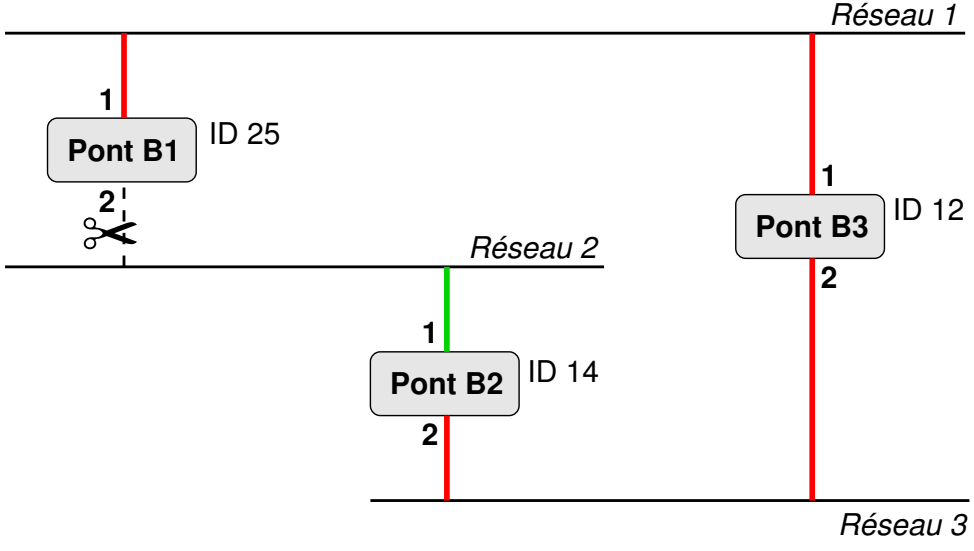
Éviter les boucles : Le mécanisme du «spanning tree»



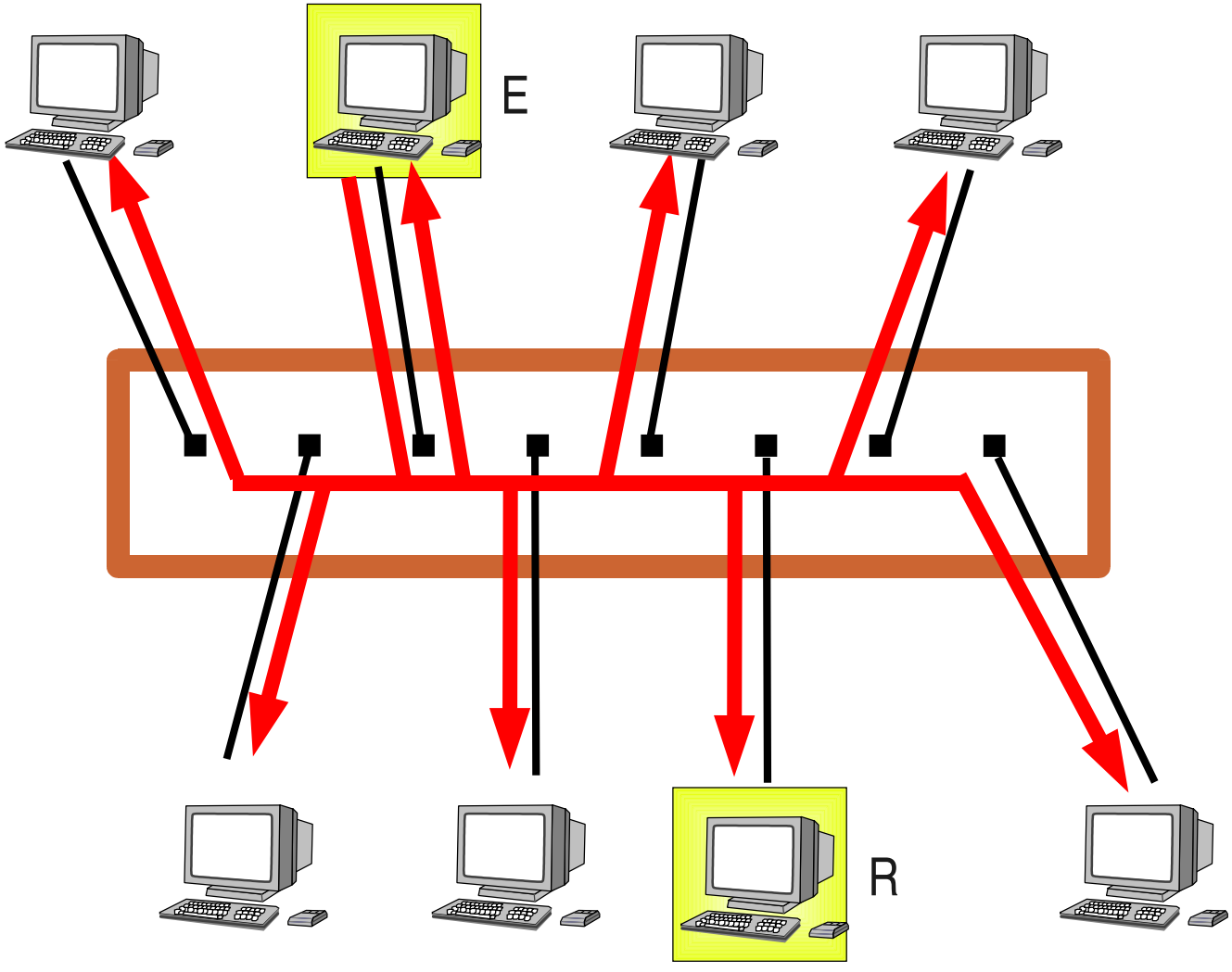
Avant


Réseau 1
1
Pont B1
ID 25
2
Réseau 2
1
Pont B3
ID 12
2
1
Pont B2
ID 14
2
Réseau 3

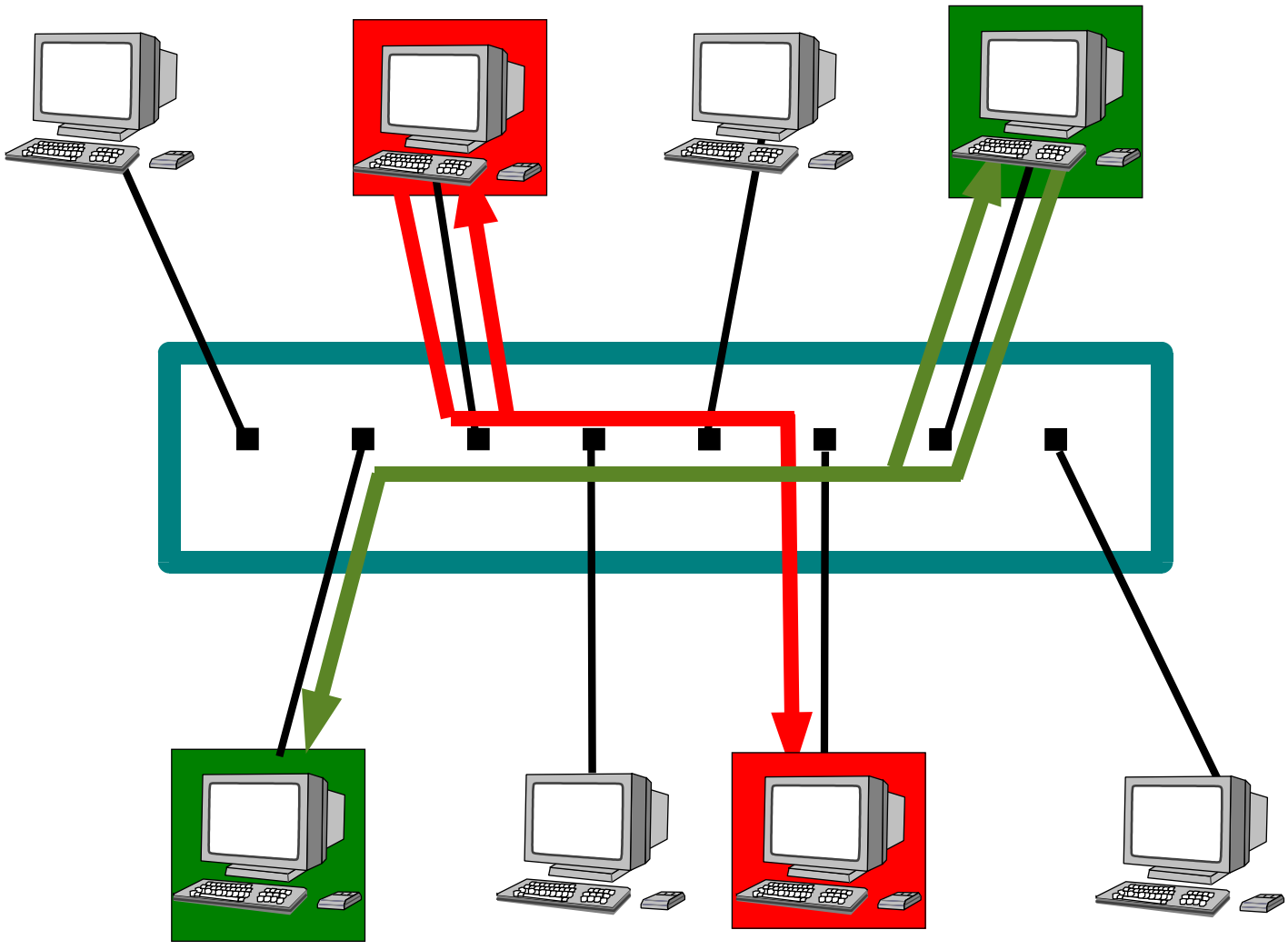
Après


Réseau 1
1
Pont B1
ID 25
2
Réseau 2
1
Pont B3
ID 12
2
1
Pont B2
ID 14
2
Réseau 3

Le bus devient étoile : raccordement par hubs

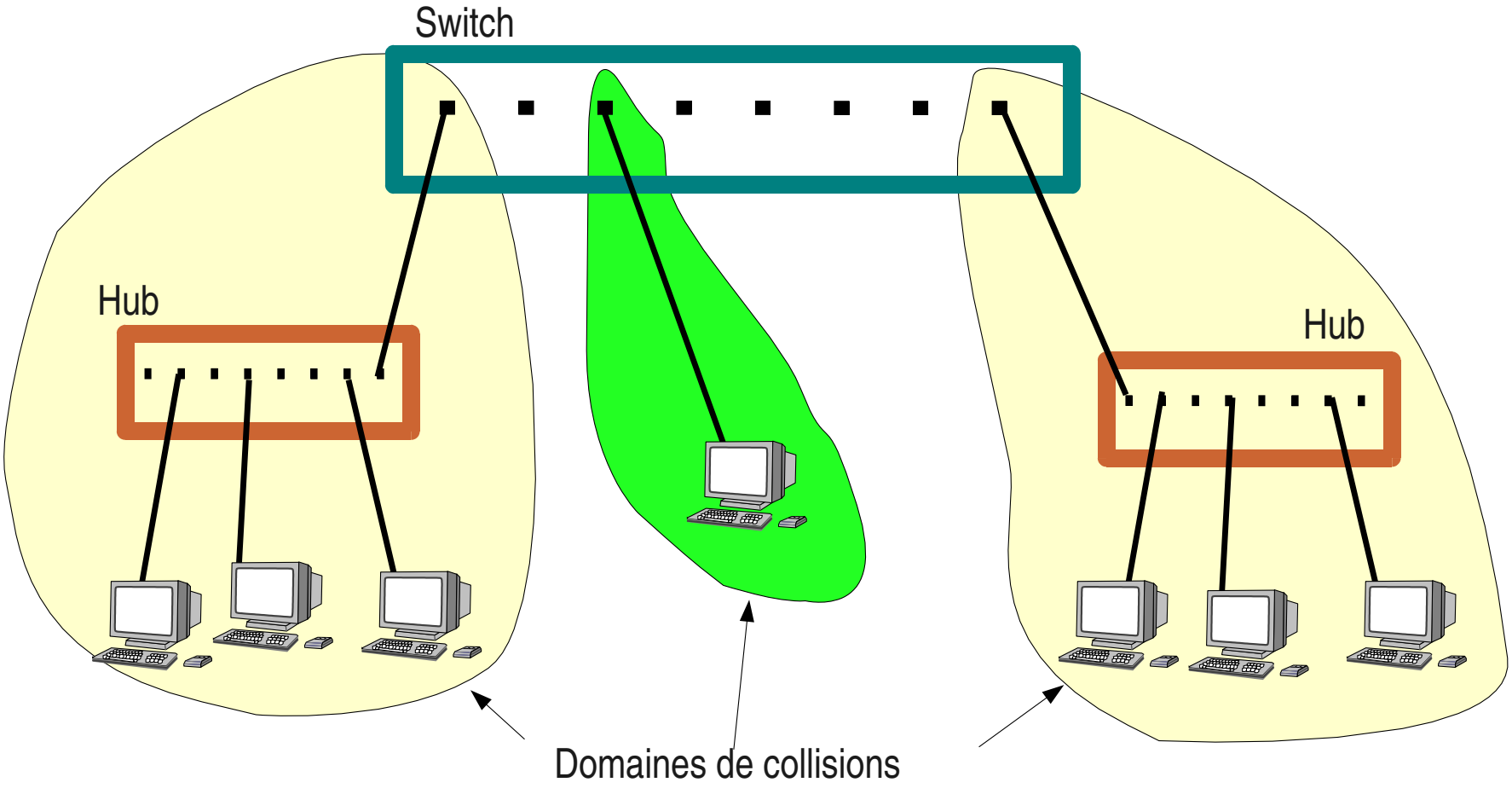


Le bus devient étoile : raccordement par commutateurs (switch)

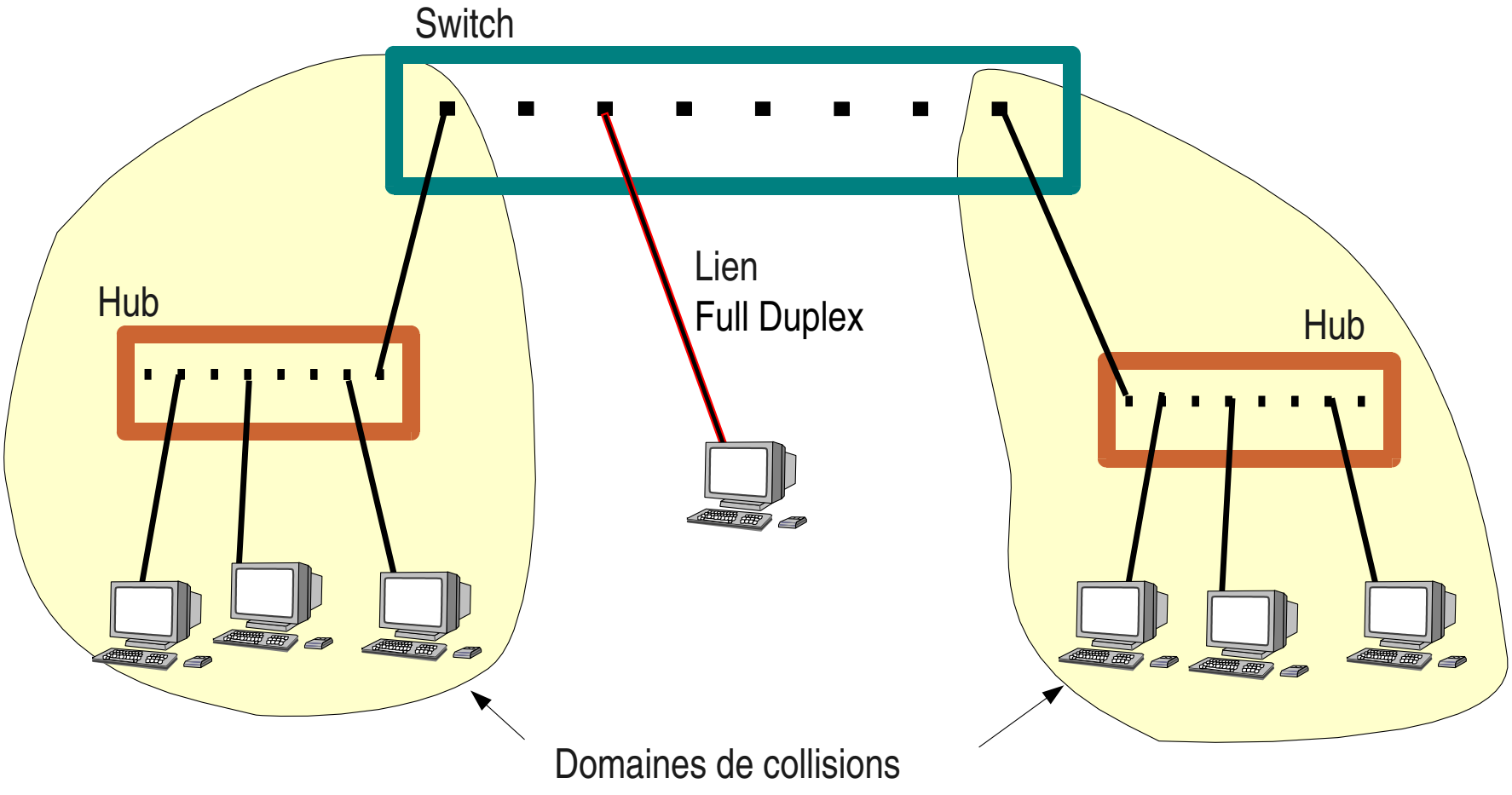


Domaines de collisions

Un switch est un pont multiports



Liens full duplex



- Un VLAN est construit à l'aide des commutateurs dont on restreint les possibilités de commutation à des groupes de ports, la commutation peut être totale entre les membres d'un groupe mais devient impossible entre les membres de groupes différents

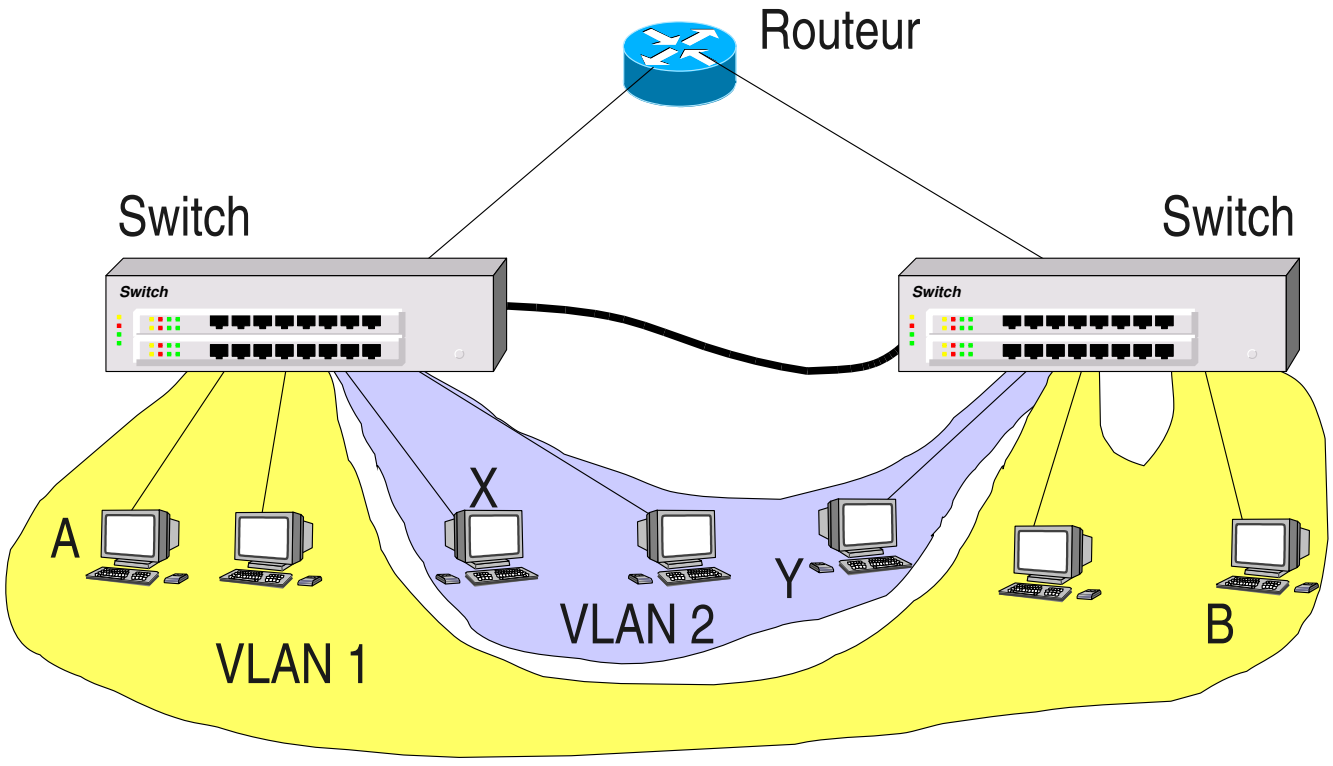
Un groupe définit un domaine de broadcast
domaine de broadcast $\Leftrightarrow$ VLAN (plus petite commune définition)

- suivant les possibilités matérielles des switches, un VLAN peut être défini par port, par adresse MAC, par adresse IP (dans ce dernier cas, ce n'est toutefois pas assimilé à de la commutation niveau 3)

Le concept de LAN virtuel ou VLAN II

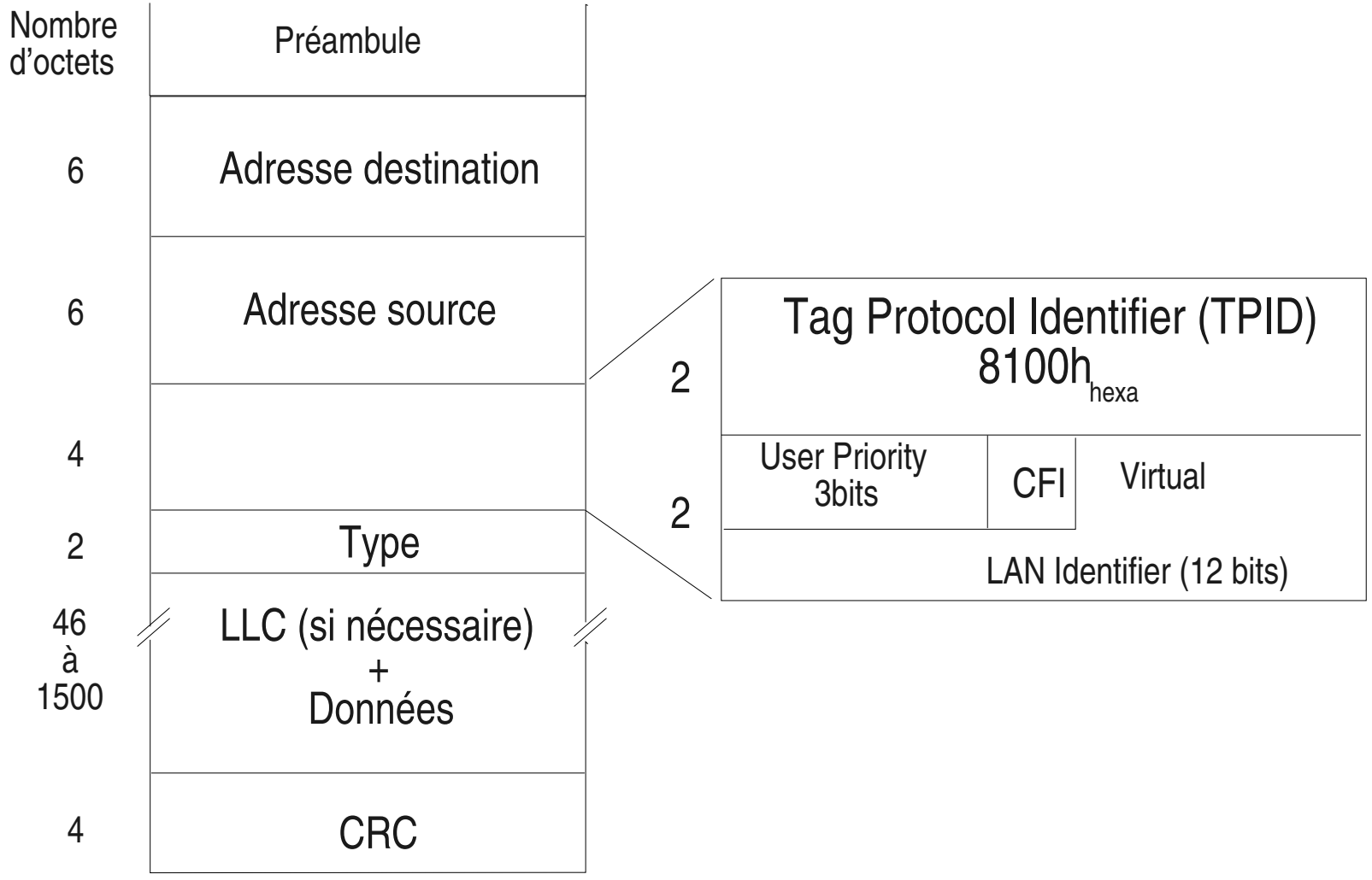
- un VLAN peut être réparti sur plusieurs commutateurs reliés entre eux
- routeur obligatoire entre VLANs, même s'il n'y a qu'un commutateur
- l'attribution d'un élément (port, adresse MAC, adresse IP) à un VLAN est réalisée par une opération de gestion
- l'interconnexion de switchs impose de marquer les trames sur les liens d'interconnexion afin de les associer à un VLAN. Exemple figure suivante : les trames échangées entre A et B ou entre X et Y doivent être différentiées sur le lien entre les deux switches. Solution : étiquetage des trames ; normes IEEE et solutions propriétaires

Les LANs virtuels : VLANs



- Sur un commutateur ou plusieurs interconnectés
  - On subdivise l'espace d'interconnexion en sous espaces étanches
  - On construit ainsi une architecture logique sur une topologie physique
  - Standardisation : IEEE-802.1p et 1q

La trame Ethernet revisitée par IEEE-802.1p/q



Compléments du standard 802.1p/q

- GARP : Generic Attribute Registration Protocol
  - mécanisme de signalisation permettant aux stations de fournir des indications (par des valeurs d'attributs) affectant les paramètres de filtrage des switches.
  - 3 attributs déjà définis :
    - groupe d'adresses MAC, facilite les mécanismes du multicast,
    - mode de filtrage des ports,
    - VLAN
- GMRP : Garp Multicast Registration Protocol
  - mécanisme permettant aux stations terminales et aux switches de s'enregistrer (et se retirer) comme participants à un groupe multicast et de diffuser cette information à l'ensemble de switches du réseau,
  - les switches utilisent cette information comme paramètre de filtrage,
  - utilise GARP comme protocole support.

Les évolutions d'Ethernet

- Années 80 : le 10Mb/s, 802.3, en bus sur câbles coaxiaux
- Années 90 :
  - 10Mb/s sur paires torsadées, raccordement sur hub
  - 100Mb/s
- Fin des années 90 : le 1000Mb/s
- Années 2000 : la 10Gb/s arrive
  - Support physique revisité, utilisation du support Sonet (OC192)
  - Distances visées : 40Km et plus
- Interopérabilité totale ascendante

Le câblage cuivre, les différents types de câbles

- Le standard : ANSI/TIA/EIA 568 B
- Câbles à 4 paires torsadées, connecteurs RJ45
- Catégories
  - 5 : 100MHz, 100Mb/s sur 100m
  - 5e : 100 Mhz
  - 6 : 250MHz : 1Gb/s sur 100m
  - 7 : 600MHz : 10Gb/s sur 100m (15Gb/s sur 15m)
- Blindés ou écrantés
  - UTP : Unshieldes Twisted Pair : non blindé
  - FTP : Foilded TP : écranté
  - STP : Shielded TP : blindé
  - SFTP : écranté et bindé

Les dénominations des différentes versions d'Ethernet

- Format : x BASE/BROAD y
  - x : le débit
  - BASE : bande de base, BROAD : large bande
  - y : indication sur la topologie (et/ou la longueur du câble)
- Bande de base :
  - Le 10 Mb/s
    - 10 BASE-5 : 10 Mb/s, topologie en bus constitué de segments de 500m
    - 10 BASE-2 : 10Mb/s, topologie en bus constitué de segments de 185m
    - 10 BASE-T : 10Mb/s, topologie en étoile, câbles en paires torsadées (T pour *Twisted pair*), longueur 100m

Les dénominations des différentes versions d'Ethernet

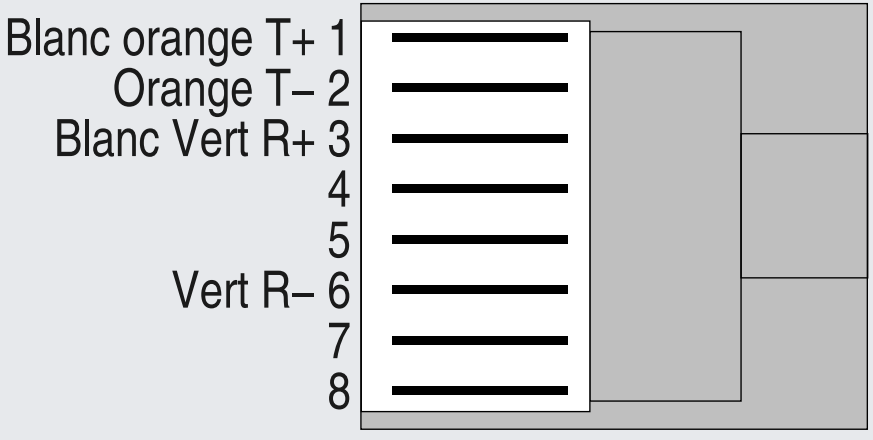
- Le 100Mb/s
  - 100BASE-T4 : 4 paires utilisées, catégorie 3 à 5
  - 100BASE-TX : 2 paires torsadée, catégorie 5
  - 100BASE-FX : 2 fibre optique
- Le 1000Mb/s
  - 1000BASE-LX : fibre optique, grande (Long) longueur d'onde
  - 1000BASE-SX : fibre optique, courte (Short) longueur d'onde
  - 1000BASE-CX : paire torsadée, 25m max
  - 1000BASE-TX : 4 paires torsadées de catégorie 5

Les câblages : les câbles catégorie 5 et leurs connecteurs

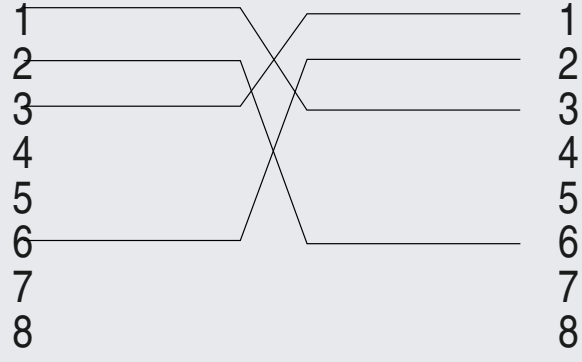
- Les câbles sont classés par catégories en fonction de leurs caractéristiques (en particulier les débits maximum
- Aujourd'hui on recommande la catégorie 5 voire 5e ou 6 pour des câblages cuivre permettant d'atteindre le 100Mb/s

Les câbles pour le 10BASE-T et 100BASE-T (EIA/TIA 568 A)

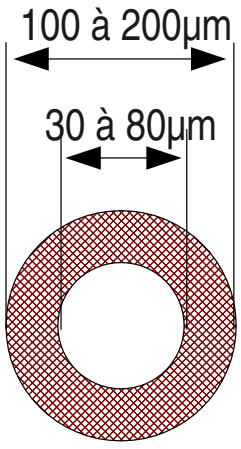
La prise RJ45

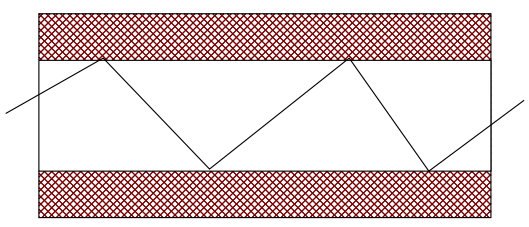


Câble croisé (interconnexion de machines hubs switches)

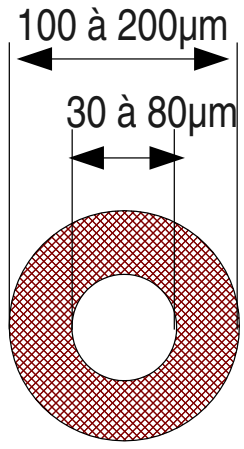


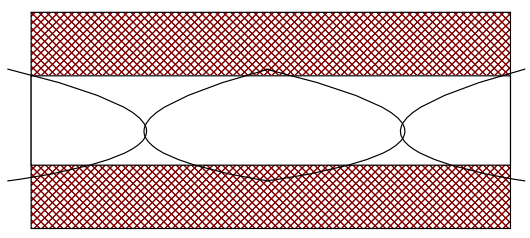
Multimode



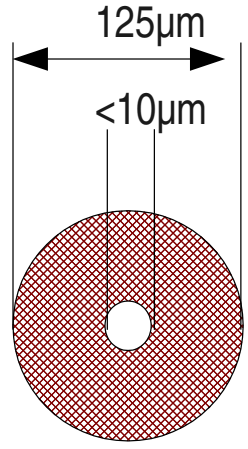


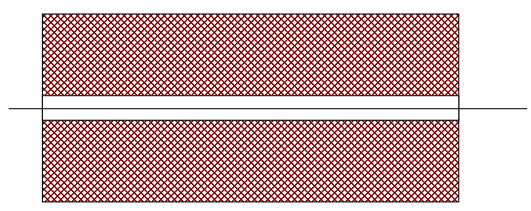
Gradient d'indice





Monomode





Transceiver	Fiber Diameter (microns)	Bandwidth (MHz*km)	Minimum Range (meters)
1000BASE-SX	MM 62.5 MM 62.5 MM 50 MM 50	160 200 400 500	2-220 2-275 2-500 2-550
1000BASE-LX	MM 62.5 MM 50 MM 50 SM 9	500 400 500 NA	2-550 2-550 2-550 2-5000

Transceiver	Fiber Diameter (microns)	Wavelength (nm)	Minimum Range (meters)
1000BASE-LH (Extended distance)	SM 9	1310	1 km - 49 km
1000BASE-LH (Extended distance)	SM 9	1550	50 km - 100 km

- 1000 BASE-SX : (S pour Short), $\lambda = 850nm$, codage en ligne 8B/10B
- 1000 BASE-LX : (L pour long), $1270nm < \lambda < 1355nm$, codage en ligne 8B/10B
- 1000 BASE-LH (Long Haul), fibre mono mode (SM : Single Mode), 9μ

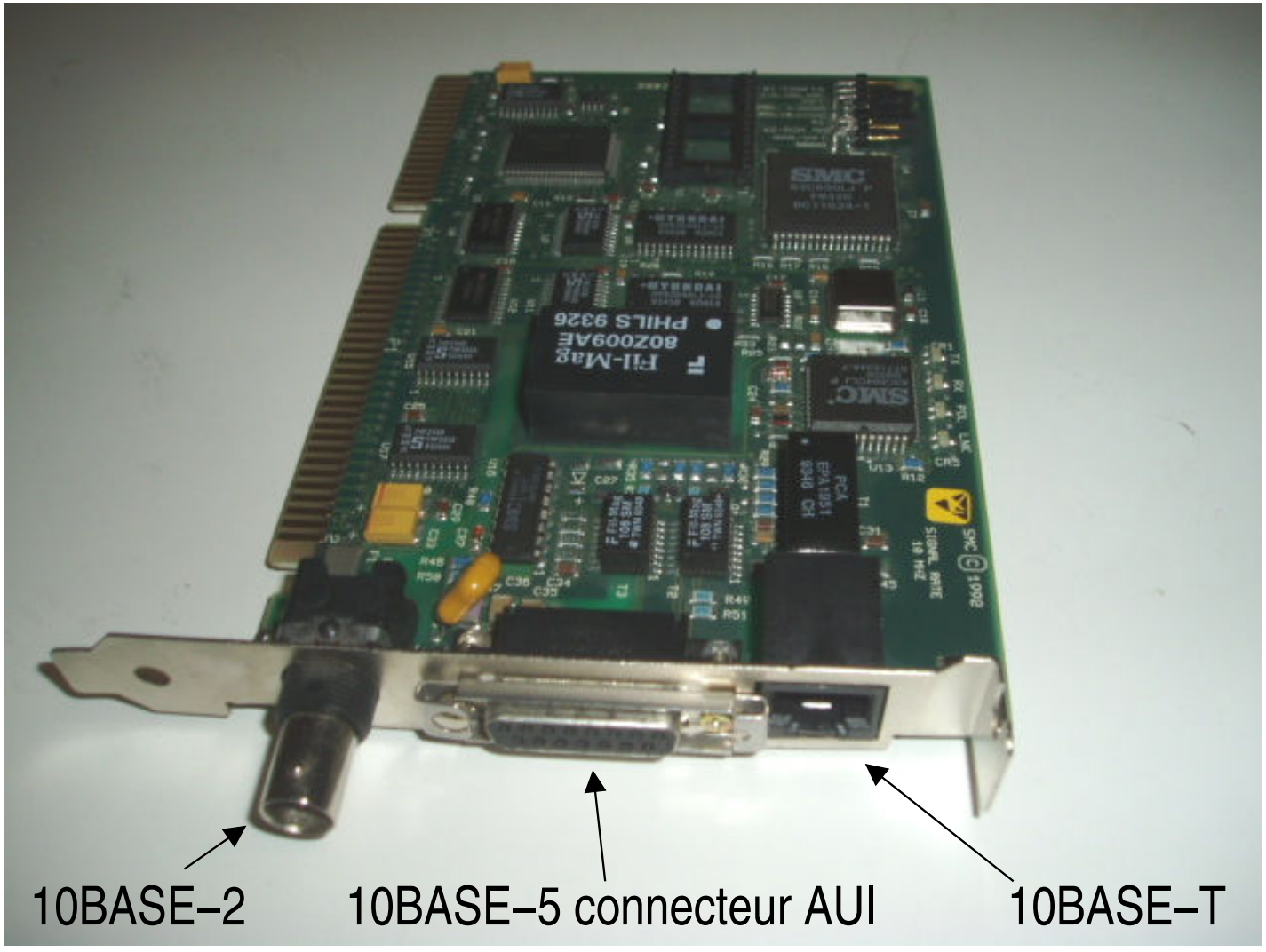
Interopérabilité des débits

- Les hubs et les switches acceptent aujourd'hui les deux débits de base 10Mb/s et 100Mb/s
- Les switches comportant des ports à 1Gb/s se généralisent
  - Un mécanisme d'autonégociation entre les extrémités dce la liaison permettent de détecter le débit maximal accepté (10 ou 100, 1000)
  - L'autonégociation permet aussi de détecter la possibilité du full-duplex et de contrôle de flux

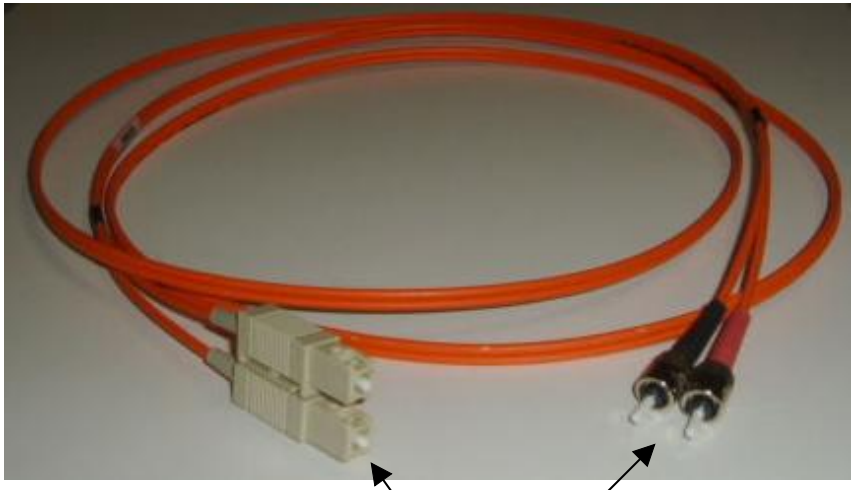
La fonction contrôle de flux

- Disponible sur certains commutateurs
- Trames «pause»
  - Adresse multicast 01-80-c2-00-00-01
  - Type 0x8808
  - Information : sur deux octets ($<$65536), contient le nombre de «pause_times» (périodes de 512 bits)
  - Indiquent la durée pendant laquelle l'extrémité ne désire pas recevoir de nouvelle trame

Une carte interface «multimedium»

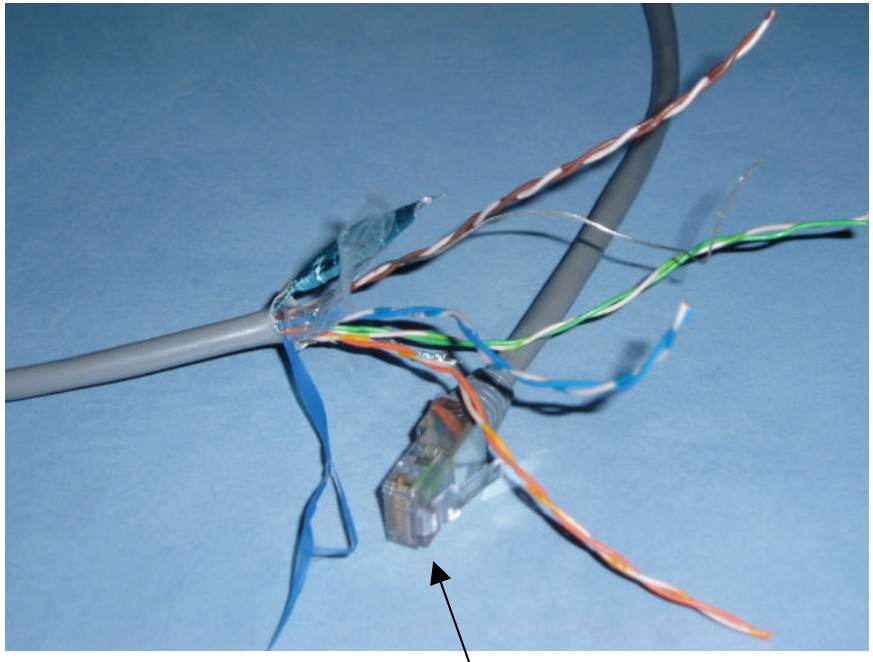


Fibre optique



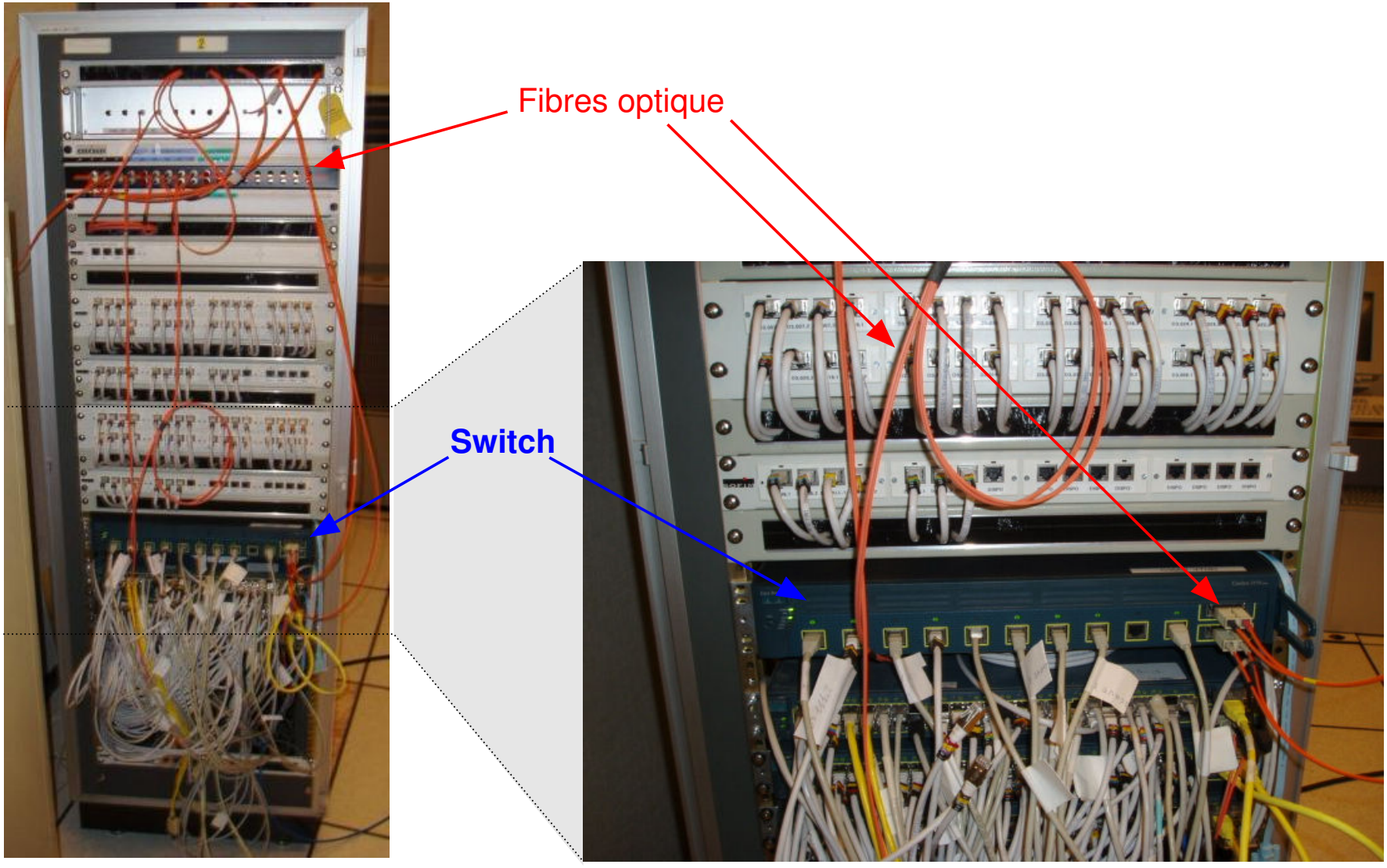
Connecteurs SC et TS

Paires torsadées



Connecteur RJ45

Une baie de brassage de câbles



Administration des commutateurs

- Les commutateurs de bas de gamme ne s'administrent pas
- On peut se connecter sur les autres
  - Par port série dédié (et hyper-terminal sous Windows ou minicom sous linux)
  - Par telnet sur une adresse IP configurée dans l'appareil
  - Par page Web sur une adresse IP configurée dans l'appareil
  - Par SNMP
- Les actions d'administration peuvent être
  - Mettre en place ou inhiber la fonction spanning-tree et gérer cette fonction (priorités du switch, QoS, ...)
  - Mettre en place et gérer des VLANs
  - Monitorer des ports
  - Surveiller les adresses MACs enregistrées
  - Surveiller le trafic

IEEE-802.11 - WiFi

- Canal radio à 2,4GHz (5GHz pour 802.11a)
  - 11b : jusqu'à 11Mb/s – 11g : jusqu'à 54 Mb/s
  - Distance relativement courte : une centaine de mètres dans de bonnes conditions de propagation, ce qui est rarement le cas en intérieur
- Méthode d'accès au médium : CSMA/CA
  - Les collisions sont évitées par un mécanisme de délai avant d'émettre
- Contrôle d'accès
  - Sécurité possible avec chiffrement : Wired Equivalent Privacy (Wep)
- Type d'application :
  - raccordement de mobiles à des réseaux de type LAN dans les entreprises et lieux publics

Technique Courant Porteur

- CPL : courant porteur en ligne
  - Le réseau d'alimentation électrique est le medium
  - Interfaçage simple avec Ethernet ou USB : le câble catégorie 5/RJ45 est relié à une prise spéciale enfich une prise de courant
  - Débits annoncés : 5-20 Mb/s
    - Des produits arrivent annonçant 200Mb/s (théorique) !
  - Modulation en ligne OFDM (certains utilisent CDMA)
  - Solutions propriétaires, pas de compatibilité entre produits
    - Pourtant un standard : Home Plug
    - Des travaux à l'ETSI
  - Applications : «émulation Ethernet», raccordement de quartiers résidentiels, d'entreprises, etc...

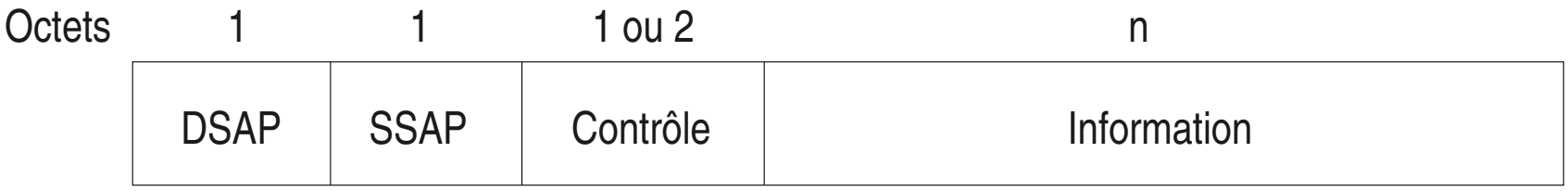
Autres techniques Réseau

- Les PANs : Personnal Area Network
  - Très courtes distances
  - Typiquement : Bluetooth® ZigBee®
    - Interconnexion de petits portables (téléphones, PDAs, etc.), domotique (capteurs, actionneurs)
  - Standardisation en cours : IEEE-802.15
- Les bus spécialisés
  - USB
  - IEEE-1394 (FireWire ou iLink selon les constructeurs)
    - Raccordement d'appareils multimédia (caméra vidéo numériques, audio)
    - Canaux synchrones et asynchrones
    - 800Mb/s
    - Distance : 100m annoncée entre nœuds et hubs (4,5m en standard)
    - En standard sur les PCs aujourd'hui

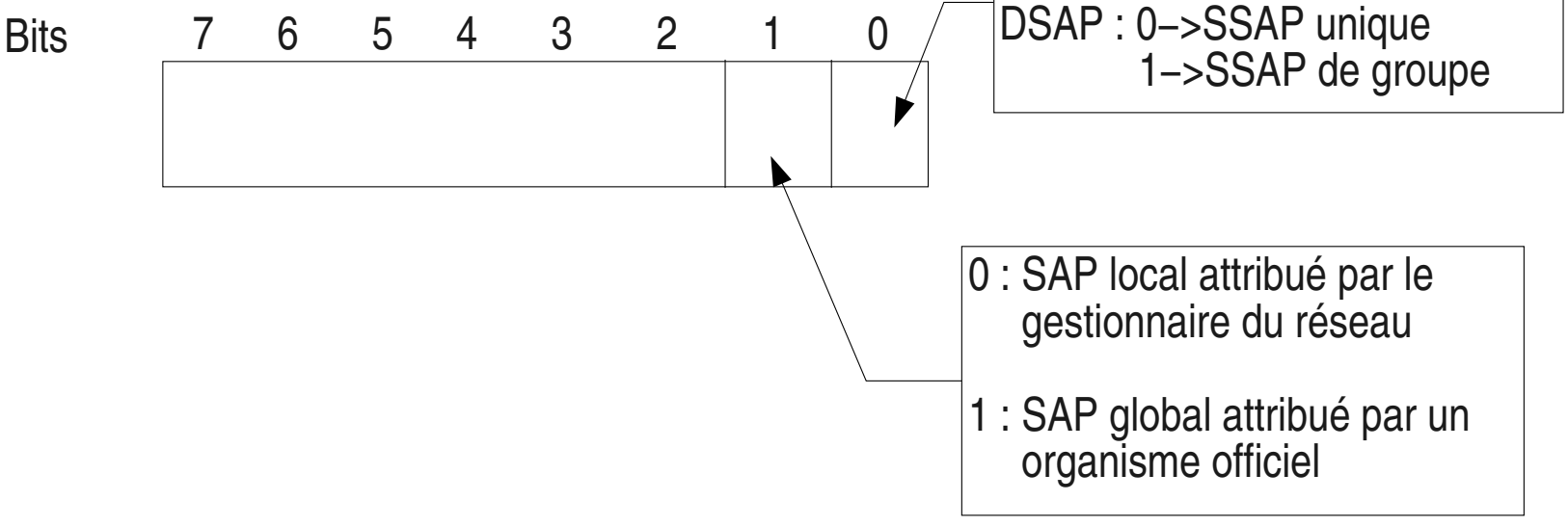
- Permet de combler les manques de la couche MAC par rapport au niveau 2 OSI standard (si nécessaire)
- Permet d'indiquer le SAP du protocole véhiculé dans les données utiles de la trame MAC

- Trois types LLC
  - LLC type 1 : mode datagramme, sert uniquement à véhiculer le SAP des données utiles
  - LLC type 2 : mode connecté, en plus de la fonctionnalité du type 1 on assure des contrôles identiques au HDLC LAPB (avec numérotation des trames modulo 128 : champ contrôle sur 2 octets), sert à véhiculer des paquets X25 par exemple
  - LLC type 3 : mode datagramme avec acquittement. Prévu pour les réseaux industriels

Format du PDU LLC



Codage des DSAP et SSAP



LLC : quelques SAP

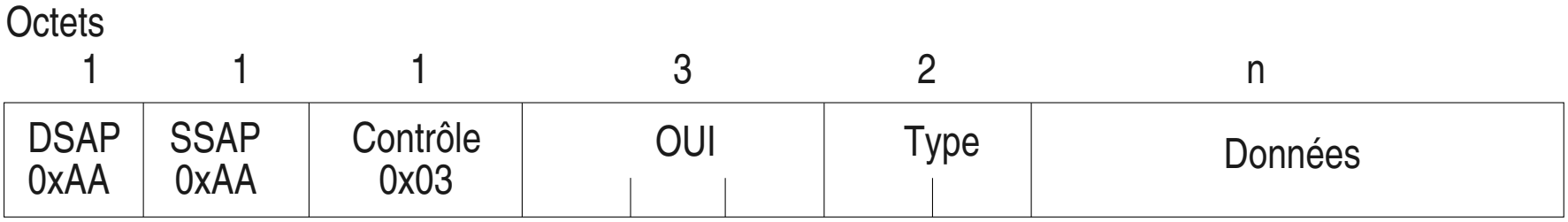
- 0x06 : IP
- 0x42 : Spanning Tree
- 0x7E : X25
- 0xE0 : Novell IPX
- 0xAA : SNAP (voir plus loin)

LLC : le champ contrôle

- LLC type 1
  - Le champ contrôle est codé sur un octet et vaut 0x03
- LLC type 2
  - Format identique au champ contrôle (ou commande) de HDLC LAPB
  - Trames numérotées modulo 128
    - Longueur du champ : 2 octets pour les rames I, RR, RNR et REJ
    - La trame d'établissement de la connexion s'appelle SABME (Set Asynchronous Balanced Mode *Extended*)

- SNAP : Sub Network Access Protocol
- Les champs DSAP et SSAP standards sont trop courts
  - Le champ type de la trame Ethernet pur est très bien...
  - Si on le réutilisait... En le plaçant après le champ contrôle ?
    - Oui mais sa longueur n'est que de deux octets !
    - Le tout ferait donc 5 octets (Contrôle sur 1 octet), ce n'est pas optimum pour une architecture 32 bits (l'architecture de la plupart de nos machines encore pour l'instant)
    - Et si on complétait à 8 octets en rajoutant l'OUI du concepteur du protocole ?

LLC SNAP : le format



- DSAP = SSAP = 0xAA
- Contrôle = 0x03 (trame UI : Unumbered Information)
- OUI : code attribué par l'IEEE à la société créatrice du protocole. Identique aux 3 octets de début des adresses Ethernet. Souvent à 0
- Type : identifie le protocole : identique au champ Type de la trame Ethernet

Exemples de trames LLC


```
 0: 0900 07ff ffff 0040 9c00 0294 0022 aaaa .......@.....".."
16: 0308 0007 809b 001a 0000 0000 03e8 ffe1 ................
32: 0101 0103 e808 e103 e880 03e8 8201 9400 ................
```

```
 0: 0180 c200 0000 0000 1d07 2c63 0026 4242 ..........,c.&BB
16: 0300 0000 8000 0000 0000 0000 0000 0000 ................
32: 0000 0000 0000 00c0 0f00 0050 0000 0000 ...........P....
48: 0000 0000 ....
```

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Troisième partie

Le protocole IP et les protocoles associés

L'organisation de l'Internet

- Organisme de tutelle : l'ISOC (Internet SOCiety), qui regroupe
  - Internet Architecture Board (IAB)
    - Responsable final des travaux de l'IETF
  - Internet Engineering Steering Group (IESG)
    - Contrôle les travaux de l'IETF
  - Internet Engineering Task Force (IETF)
    - L'organisme de standardisation
  - Internet Research Task Force (IRTF)
    - Responsable de la recherche à long terme
  - Internet Assigned Numbers Authority (IANA)
    - La gestion des adresses, des numéros de protocoles, du DNS

- Les membres des groupes de l'IETF travaillent par échange de courriers électroniques, sur des documents appelés *drafts*, dont la validité est de 6 mois
- Périodiquement ils se rencontrent dans des «meetings» pour valider des choix techniques
- Un *draft* peut évoluer vers une nouvelle version, valable 6 mois de plus
- Lorsque que le groupe de travail arrive à un consensus, le *draft* est promu au rang de RFC (Request For Comment)
  - Les RFCs sont les «normes» Internet, l'équivalent des recommandations ISO et ITU-T
  - Mais pas seulement...
- RFC et drafts sont en publications gratuite sur différents sites ftp et web (`www.ietf.org`)

Les RFC

- Format standard : ASCII pur et dur (voir rfc 2223)
- Plusieurs types forts
  - Proposed standard : draft présentant un fort consensus
  - Draft standard : pour un protocole dans cet état de standardisation, il existe au moins deux versions interopérantes
  - Standard : LE document final (mais pas obligatoirement figé pour l'éternité)
- Autres types
  - Expérimental : protocole encore en développement
  - Informatif : comme son nom l'indique
  - Historique
  - Best Current Practice (BCP)
  - Les rfc du premier avril... rfc1084, 1149, etc...

Les classifications des RFCs

- Certains RFCs standards sont classées sous l'appellation std, le std 5 par exemple correspond au rfc791 définissant IP
- Certains sont classées sous l'appellation FYI : For Your Information
- Quelques RFCs :
  - 791 : IP (std5)
  - 793 : TCP (std7)
  - 763 : UDP (std6)

IP et le modèle OSI

Application

Présentation

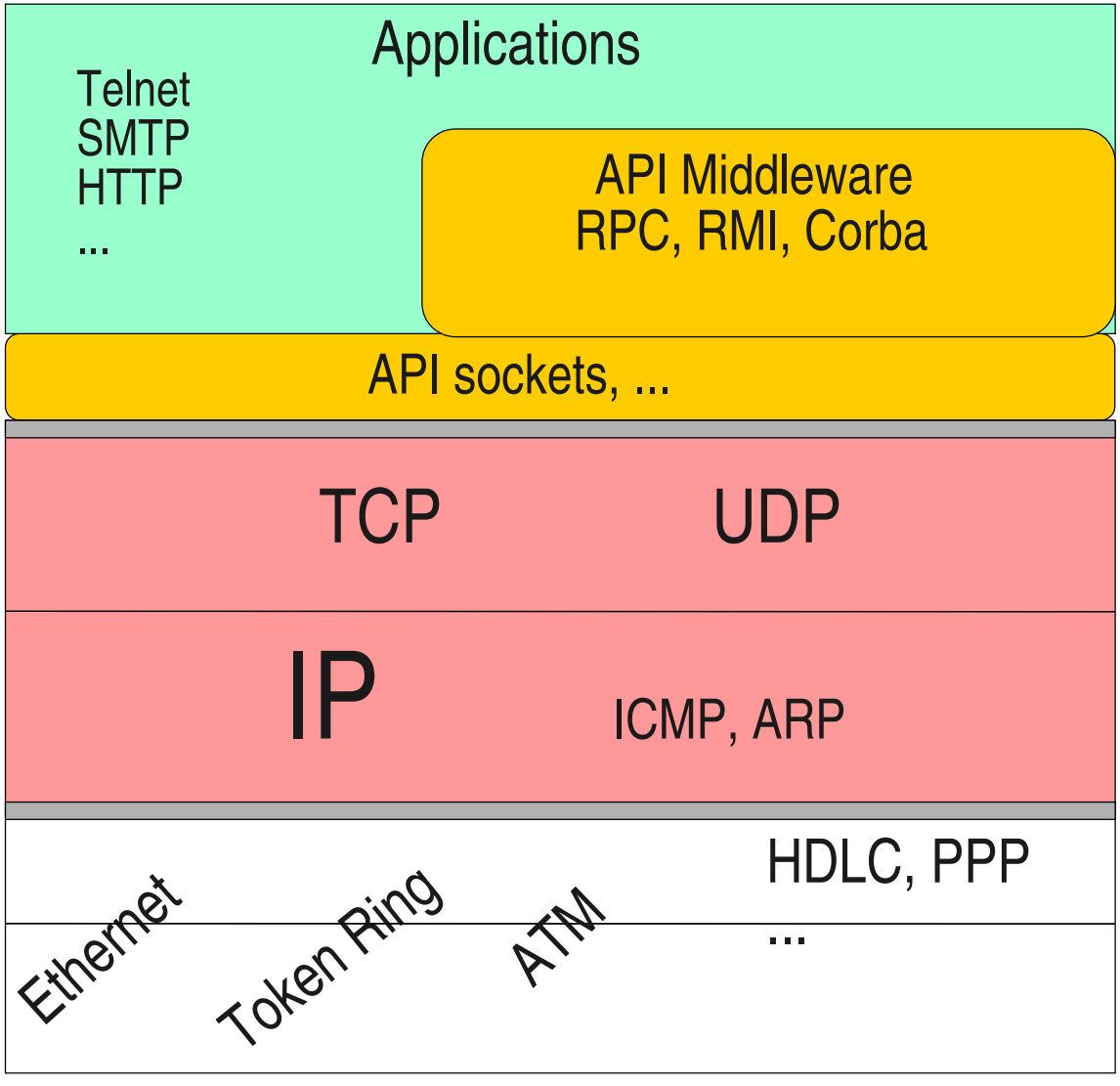
Session

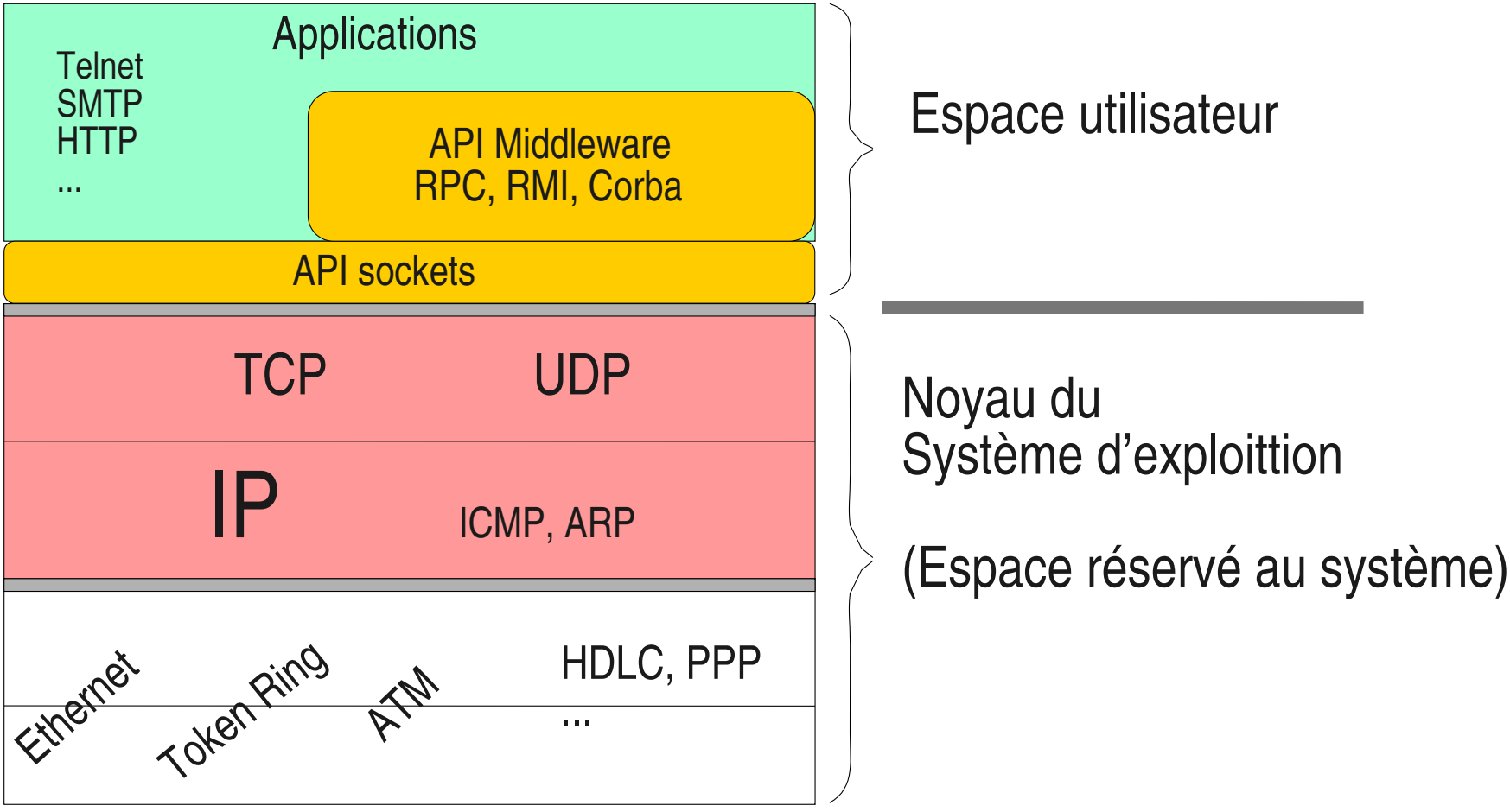
Transport

Réseau

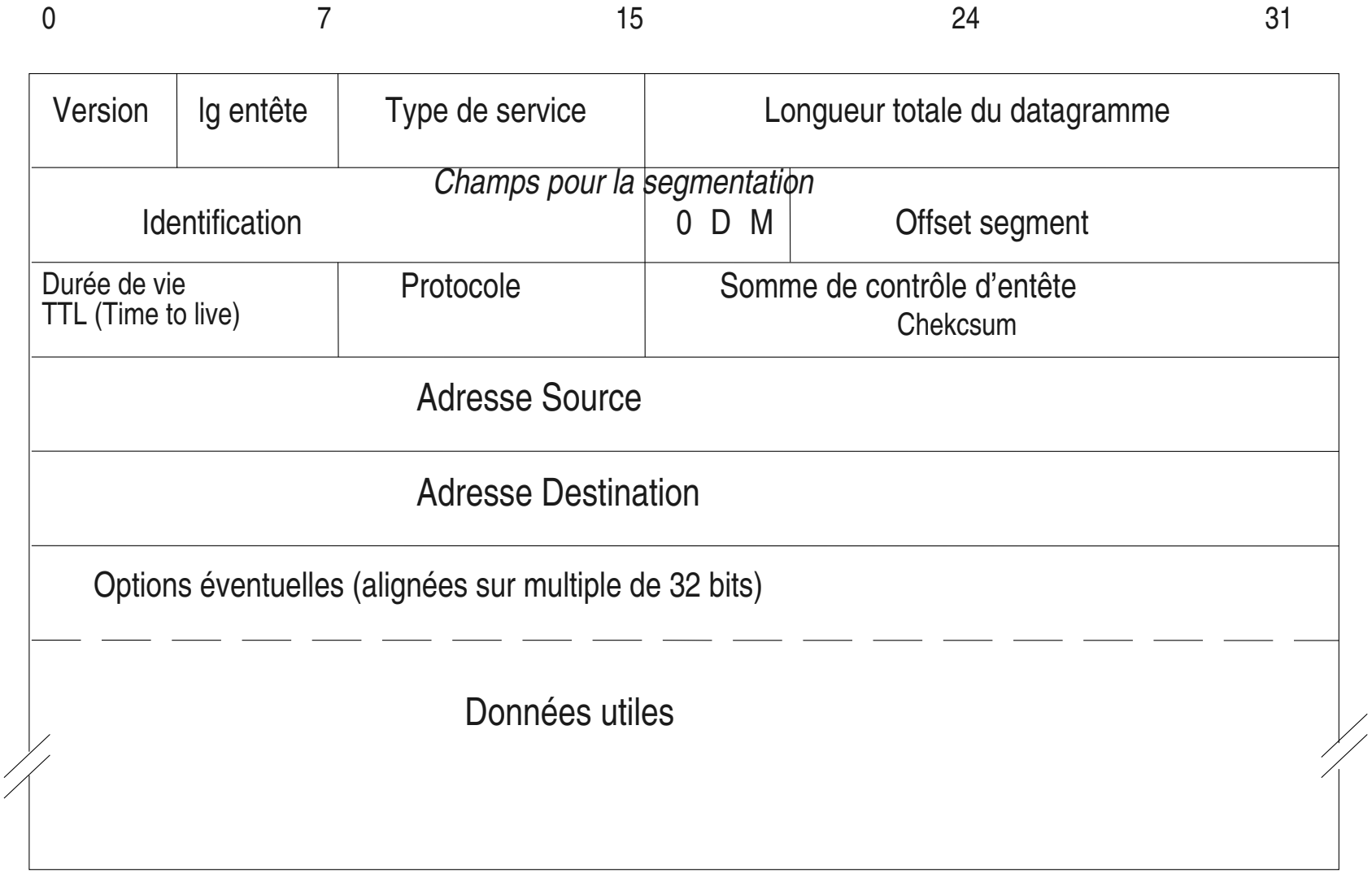
Liaison

Physique



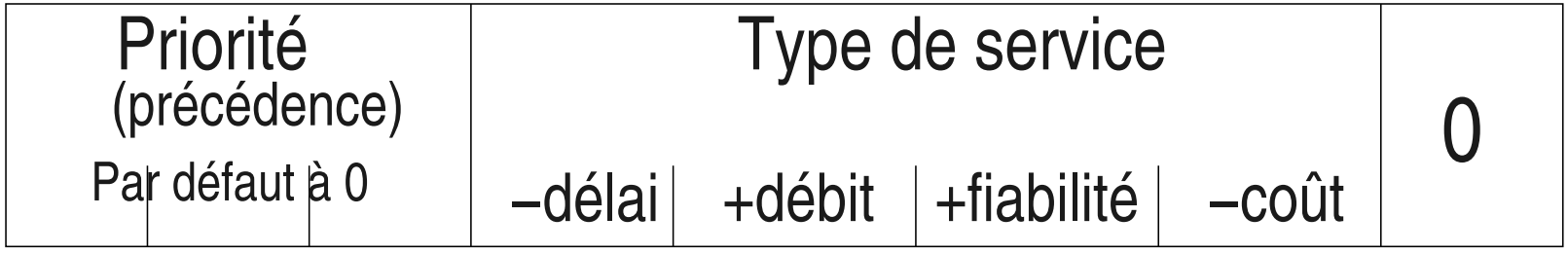

Applications
Telnet
SMTP
HTTP
...
API Middleware
RPC, RMI, Corba
API sockets
TCP
UDP
IP
ICMP, ARP
Ethernet
Token Ring
ATM
HDLC, PPP
...
Espace utilisateur
Noyau du
Système d'exploittion
(Espace réservé au système)

Le datagramme IPv4



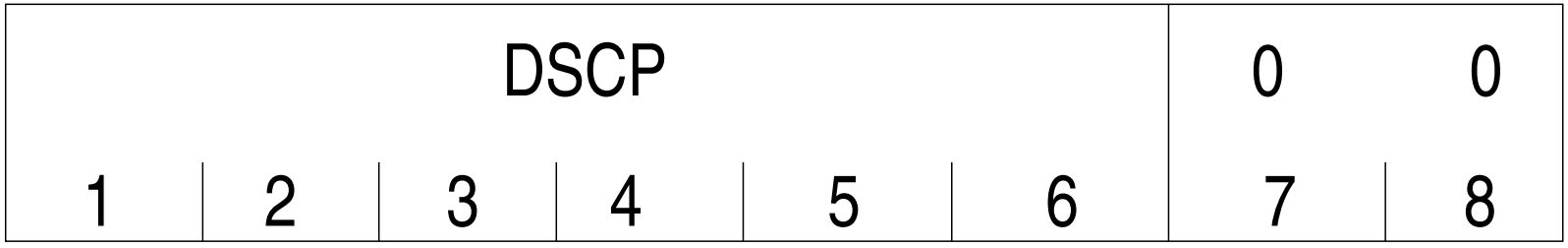
Le champ Type de Service

- Encore appelé ToS (*Type of Service*) *rfc1349*



- Utilisation non généralisée
  - Précédence utilisable pour marquer des flux dans les nœuds du réseau (routeurs)
  - Les bits Type de Service peuvent être positionnés par les applications terminales (API socket sous Windows et linux)
  - Sous Linux ils peuvent être positionnés via la commande iptables selon divers critères

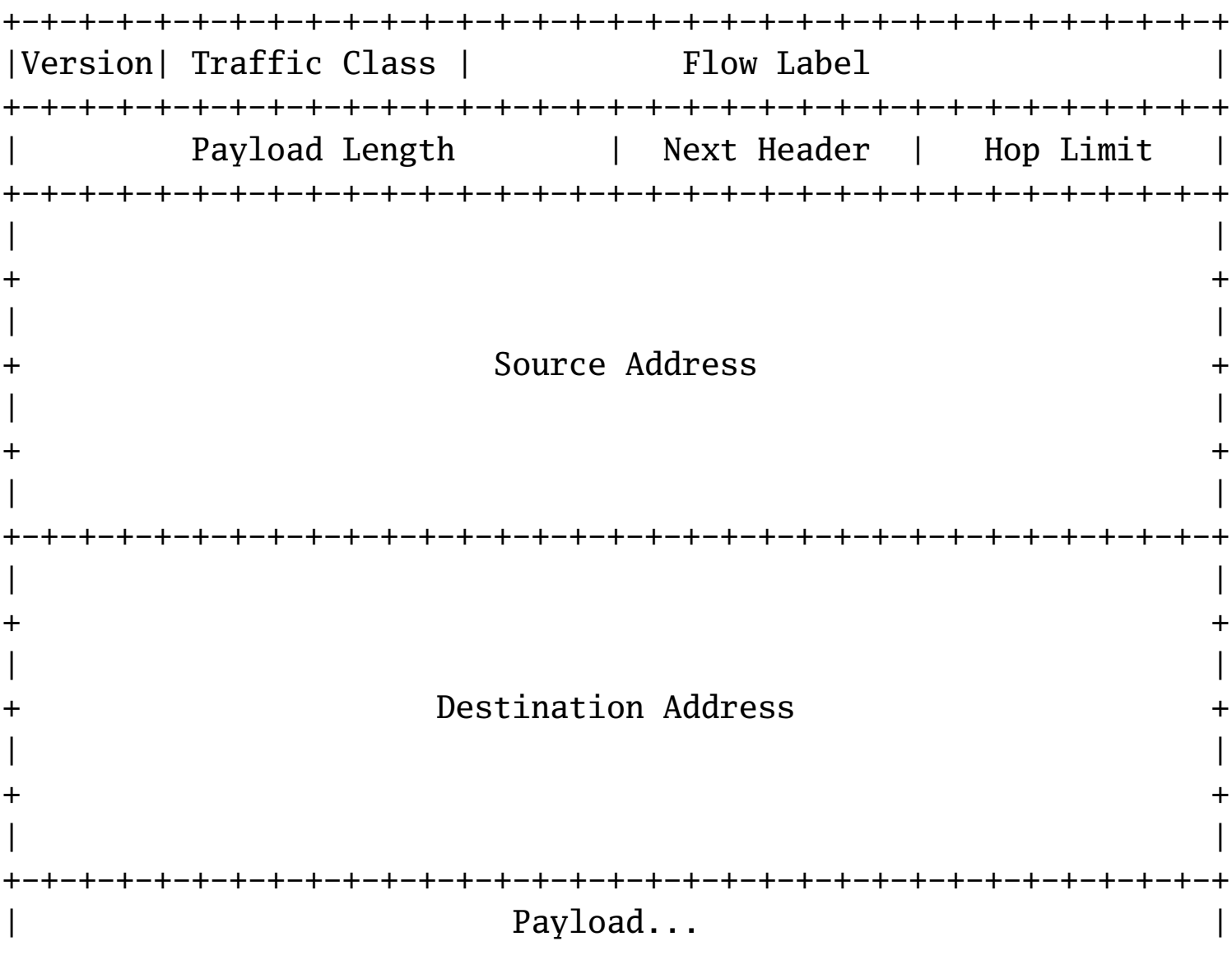
Le champ Type de Service en version DiffServ



- Le champ DSCP sert à coder le *PHB (Per Hop Behavior), paramètre fondamental de DiffServ (rfc 2474)*
- Exemples de PHBs :
  - Expedited forwarding (pertes faibles, latence faible, gigue faible, bande passante assurée)
  - Assured forwarding (groupe de PHBs)
  - Best effort
  - Network control

Les options IPv4

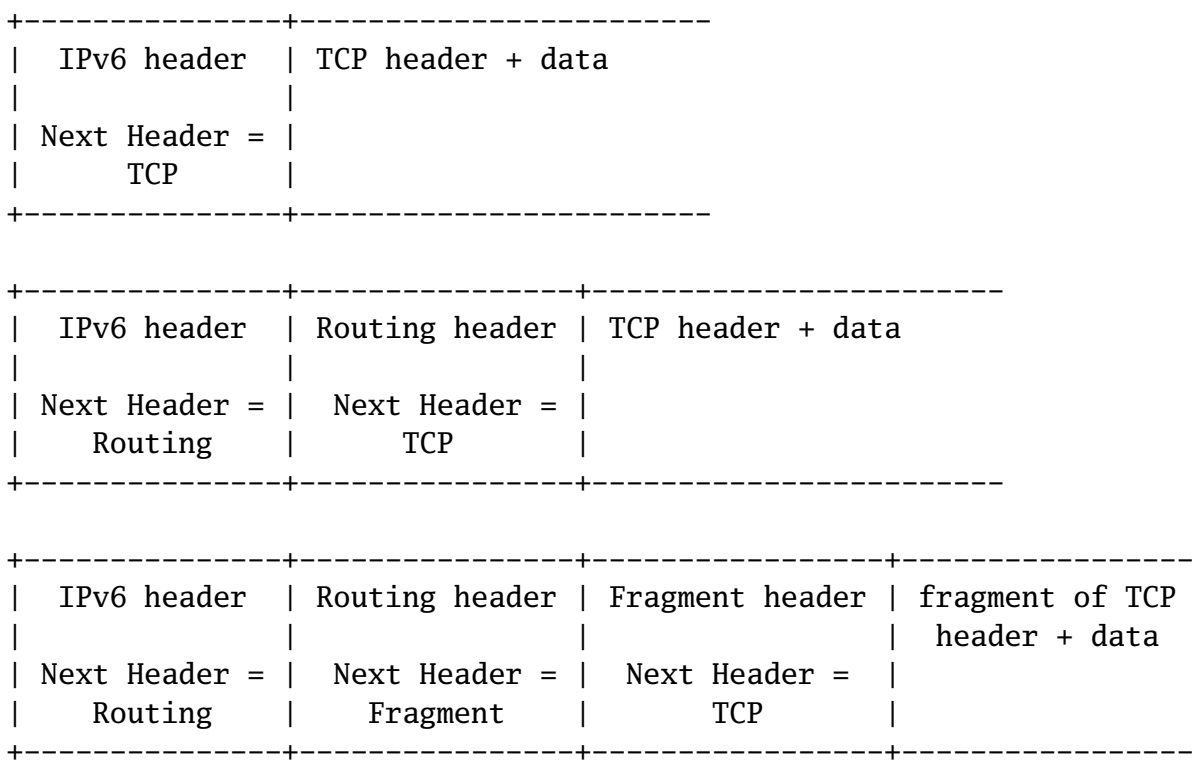
- Format :
- Les options :
  - LSR : *Loose Source Route* : permet d'indiquer la route
  - SSR : *Strict Source Route*, comme précédemment en plus rigoureux
  - RR : *Record Route* : les routeurs traversés rajoutent leur adresse
  - RTALT : *Router Alert*, permet de passer le paquets aux couches hautes des routeurs traversés
  - etc.

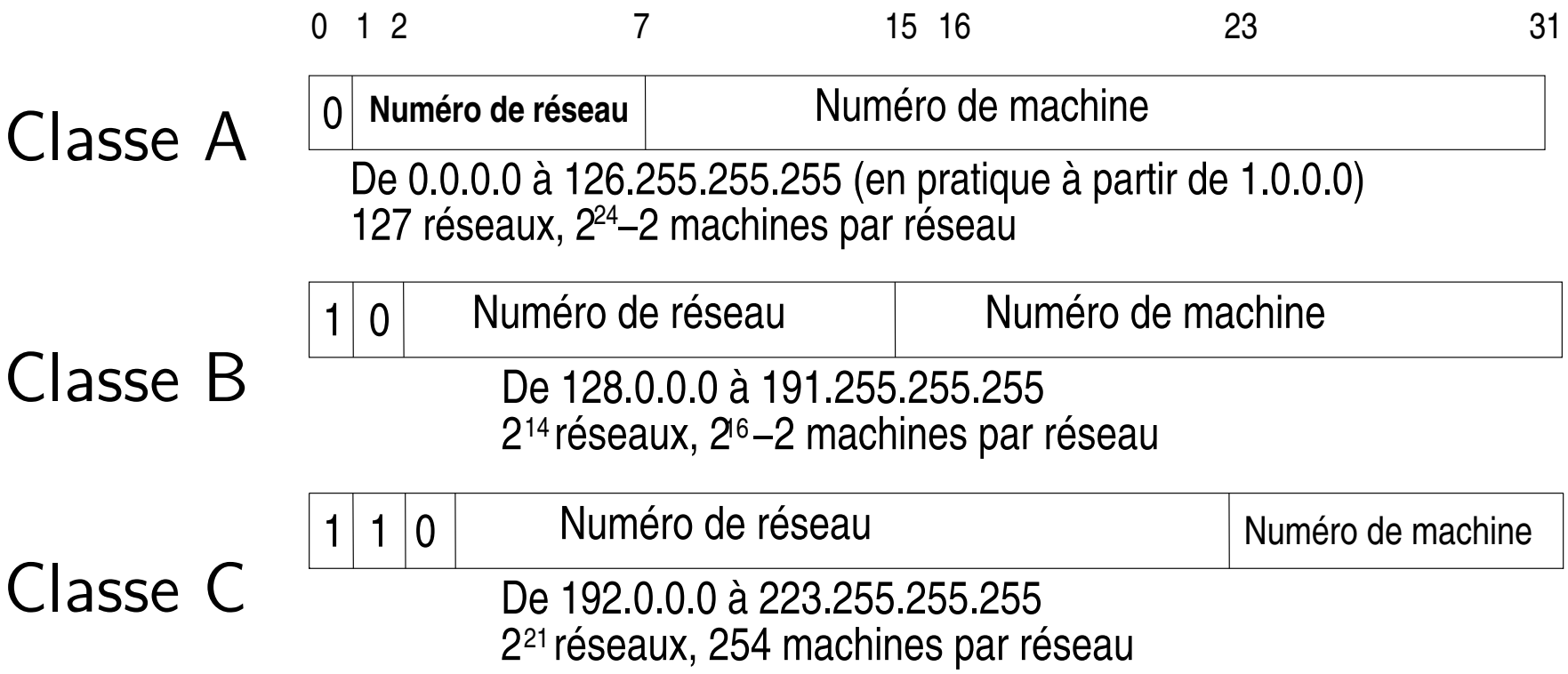

Version
Traffic Class
Flow Label
Payload Length
Next Header
Hop Limit
Source Address
Destination Address
Payload...

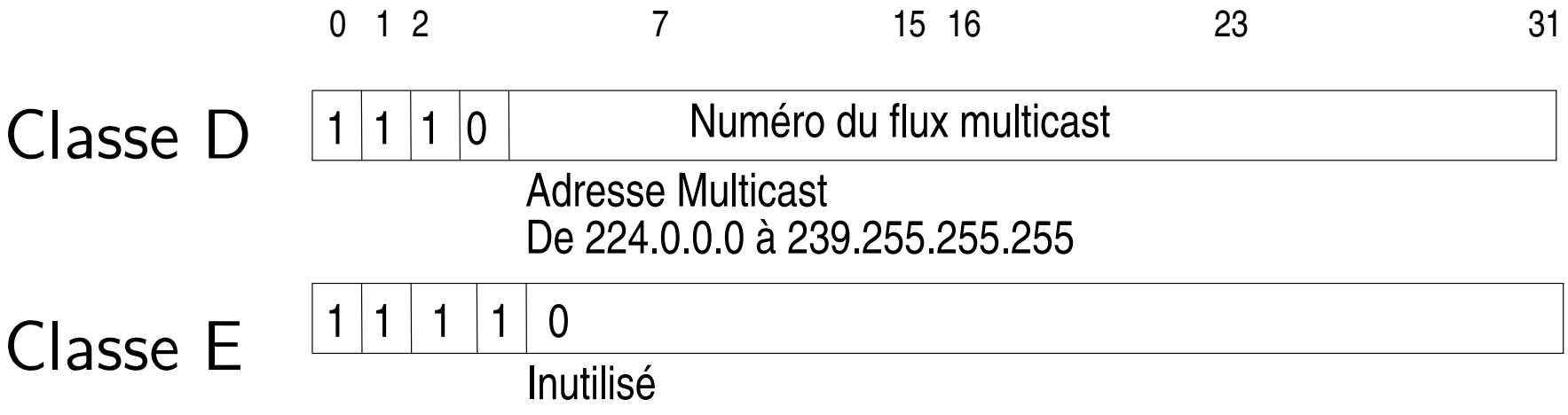
- Version : 6 (!)
- Traffic Class : typiquement Differentiated Services et Explicit Congestion Notification, comme le TOS IPv4
- Flow Label : pour tagger les paquets d'un même flux ; visiblement peu utilisé...
- Payload Length : comme son nom l'indique... (Note : dans IPv4 on a la taille *totale* du datagramme)
- Next Header : remplace le champ Protocole IPv4 et ajoute un mécanisme un peu novateur pour gérer les options (cf. slides suivants)
- Hop Limit : l'équivalent du champ TTL IPv4 (Note : contrairement à IPv4 qui compte en secondes, ici c'est bien en nombre de sauts)
- Source & Destination Addresses : sur 128 bits (Ouf !)

Next Header

- Les headers s'enchaînent dans le payload
- ...les options IPv6 diverses, puis le protocole transporté effectivement (TCP, UDP, ...)




0 1 2 7 15 16 23 31
Classe A
0
Numéro de réseau
Numéro de machine
De 0.0.0.0 à 126.255.255.255 (en pratique à partir de 1.0.0.0)
127 réseaux, 2^{24}–2 machines par réseau
Classe B
1 0
Numéro de réseau
Numéro de machine
De 128.0.0.0 à 191.255.255.255
2^{14} réseaux, 2^{16}–2 machines par réseau
Classe C
1 1 0
Numéro de réseau
Numéro de machine
De 192.0.0.0 à 223.255.255.255
2^{21} réseaux, 254 machines par réseau


0 1 2 7 15 16 23 31
Classe D
1 1 1 0
Numéro du flux multicast
Adresse Multicast
De 224.0.0.0 à 239.255.255.255
Classe E
1 1 1 1 0
Inutilisé

Les adresses particulières

- L'adresse de boucle locale : 127.0.0.1
  - Interface lo sous Linux
- Les adresses privées : rfc-1918
  - 10.0.0.0 à 10.255.255.255
  - 172.16.0.0 à 172.16.255.255
  - 192.168.0.0. à 192.168.255.255
  - Non routables dans l'Internet
    - Les machines munies de ces adresses peuvent cependant accéder l'Internet via des passerelles réalisant une fonction de translation d'adresse appelée **NAT** pour ***Network Address Translation***
  - Routables dans les réseaux privés

Les adresses particulières pour la diffusion

Différentes possibilités :

- *Broadcast* sur le réseau local : 255.255.255.255
  - Peu utilisée
- Partie réseau normale, partie machine à 255
  - Généralement l'adresse de broadcast mise en œuvre
    - Exemples :
    - 192.168.100.255
    - 172.16.255.255
  - Partie réseau normale, partie machine à 0
  - Ancienne adresse de broadcast pouvant encore être utilisées sur des machines SUN dont l'OS est SUNOS-4

Les adresses de réseaux

- Les réseaux avec un netmask standard
  - On indique les 4 octets (entre 0 et 255), comme pour une adresse normale, les derniers octets étant à 0 (le dernier pour une classe C, les deux derniers pour ne classe B, les trois derniers pour ne classe A)
    - Exemple : 192.168.100.0
- Les réseaux «subnettés»
  - On fait figurer tous les octets (entre 0 et 255), y compris les bits de l'extension
    - Exemple : 192.168.100.32 (netmask 255.255.255.224 par exemple)
    - Voir page suivante

Le *subnetting*

- Extension de la partie «Réseau» de l'adresse en empruntant quelques bits de poids forts de la partie machine
- Par exemple 255.255.255.224 pour une classe C
  - Tout à 1 sauf la partie machine (224d = 1110 0000b)
- Le masque indique quels bits
- Permet de créer des «sous» réseaux
  - Les sous réseaux sont raccordés entre eux via des routeurs, comme des réseaux «normaux»
- Il y a toujours un netmask
  - Il est standard s'il ne comporte pas d'extension de bits par rapport à la classe d'adresses : 255.255.255.0 pour une lasse C par exemple

Le netmask général et ses notations

- Les adresses sans classe
  - Concept CIDR (RFC-1519) : *Classless InterDomain Routing*
  - La frontière de l'adresse réseau n'est plus figée selon la loi des classes
  - Utile pour agréger des routes dans les tables de routage des routeurs
  - Utile pour les fournisseurs de service pour affecter un sous ensemble d'adresses à un client...
  - Le netmask doit être précisé avec les adresses
- Notation
  - Classique : 255.255.255.128 (25 bits de masque)
  - Notation CIDR : /25 : exemple 192.168.100.128/25

Affectation d'adresse à une interface IP

- «À la main»
  - Selon les outils offerts par le système d'exploitation
    - À l'aide d'interfaces graphiques d'administration
    - Via des commandes spécifiques
- Dynamiquement
  - Via le protocole DHCP (*Dynamic Host Control Protocol*)
    - Un serveur est configuré pour donner l'information
  - Sur connexion via liaison point à point et le protocole PPP (Point to Point Protocol). La machine à configurer contacte un serveur situé à l'autre extrémité de la liaison. Le serveur peut lui fournir son adresse

Adresses IPv6 – RFC 4291

- Adresses sur 128 bits
- Notation hexa par bloc de 16 bits
  `2001:db8:cafe:deca:0:0:0:1`
- Compression de zéros
  `2001:db8:cafe:deca::1`

Sous-réseaux

- Deux parties :
  - **préfixe** ou **identifiant de sous-réseau** (subnet ID) : partie gauche de l'adresse
  - **identifiant de machine** (host ID) : partie droite
- notation CIDR :
  - `2001:db8:cafe:deca:a9e:1ff:fe6b:25c9/64`
  - subnet correspondant: `2001:db8:cafe:deca::`

Exemples de préfixes

- Préfixe de documentation `2001:db8::/32`
- Préfixe link-local `fe80::/10`
- Préfixe multicast `ff02::/10`
- Exemple de préfixe «end-user» `2001:db8:fada:ba00:/56`

Exemples d'adresses

- Unicast
  `2001:db8:cafe:deca:a9e:1ff:fe6b:25c9/64`
- Link-local
  `fe80::a9e:1ff:fe6b:25c9/64`
- Notation IPv4
  `2001:db8::cafe:192.168.0.1` égal à
  `2001:db8::cafe:c0a8:1`
- Localhost `::1`
- Tout les bits à 0 `::`

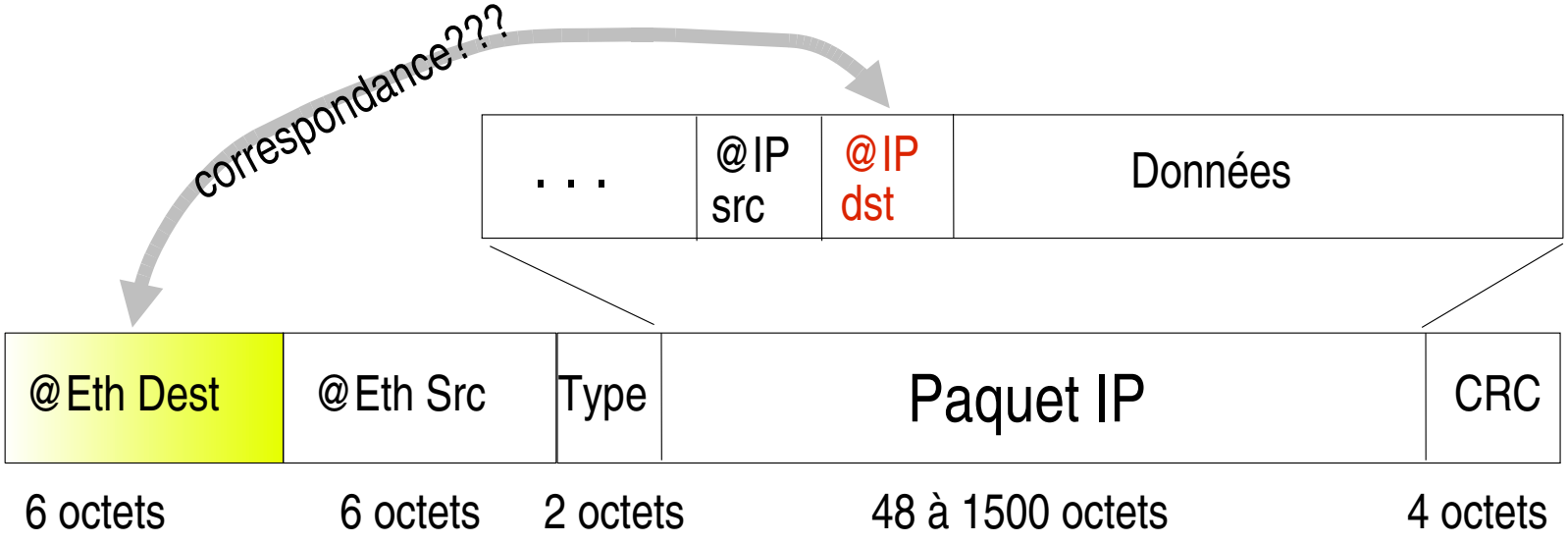
Affectation d'adresses à une interface IP

- Il est *normal* d'avoir plusieurs adresses IPv6 à une interface (adresse de portée *lien*, de portée *globale*)
- Configuration «à la main» : plutôt rare
- Dynamiquement
  - Auto-configuration, grâce au préfixe annoncé par le router
  - Via DHCPv6

Relation entre adresse IP et adresse MAC dans les réseaux locaux

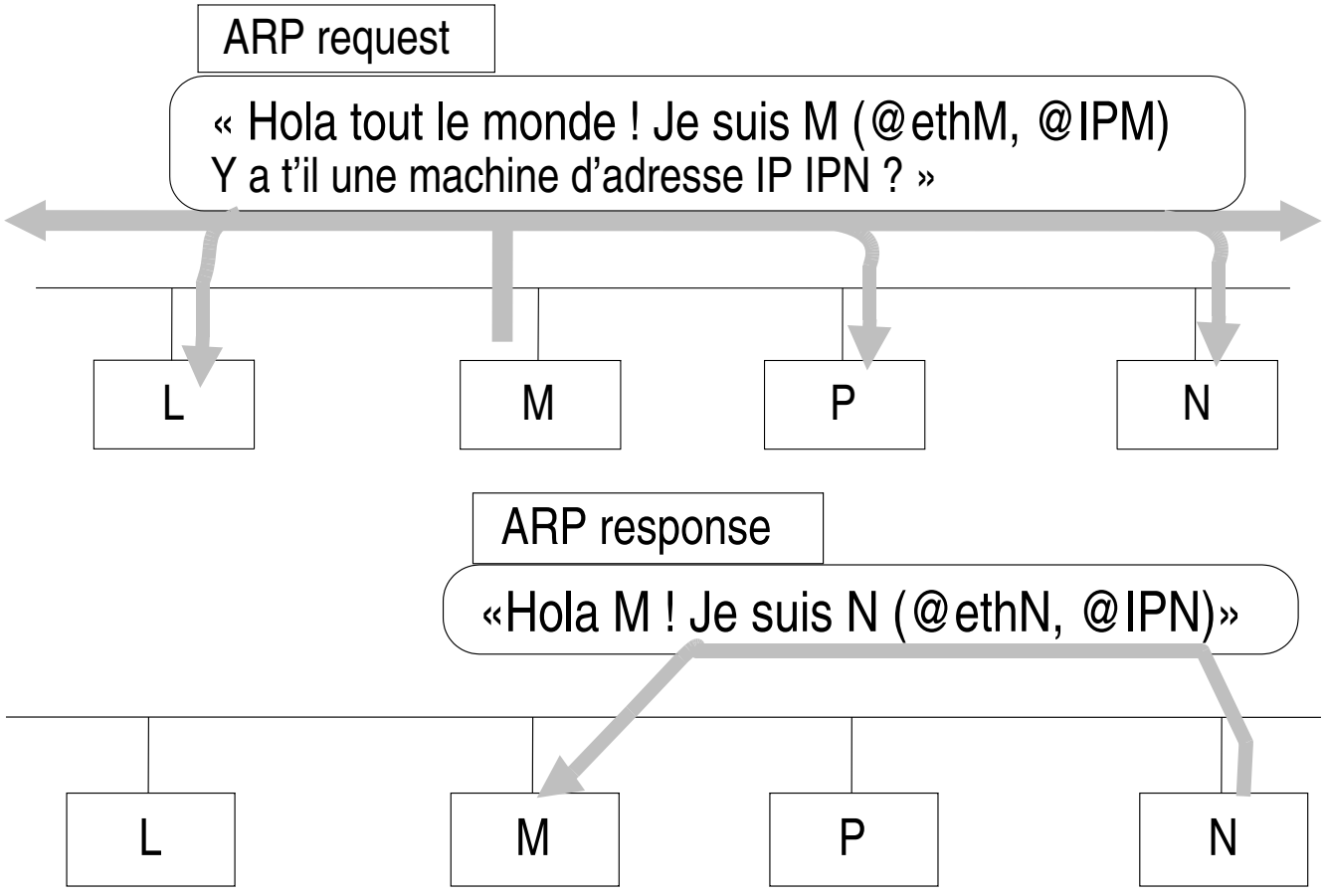
- Concernent les machines reliées à un réseau local de type Ethernet ou 802.11
- Les adresse IP se gèrent
  - Elles sont affectées «à la main» via des outils spécifiques du système d'exploitation des machines (interfaces graphiques ou commandes telles que ifconfig sous Unix/Linux)
  - Elles peuvent être affectées automatiquement via le protocole DHCP, mais cette possibilité est configurée elle aussi à la main
- Les adresses MAC ne se gèrent pas
  - Elles sont préaffectées par le constructeur de la carte interface que l'on achète ou qui est fournie avec la machine
  - Parfois le pilote (drivers) de la carte permet que cette adresse puisse être modifiée (via ifconfig)

- Problème : une machine M doit émettre un paquet IP vers une machine N voisine dont on ne connaît que le numéro IP (sur le même LAN)
- Si on est sur Ethernet le paquet IP sera véhiculé par une trame telle que celle-ci



Adresses MAC et adresses IP : la solution ARP (IPv4)

- Adress Resolution Protocol - rfc826

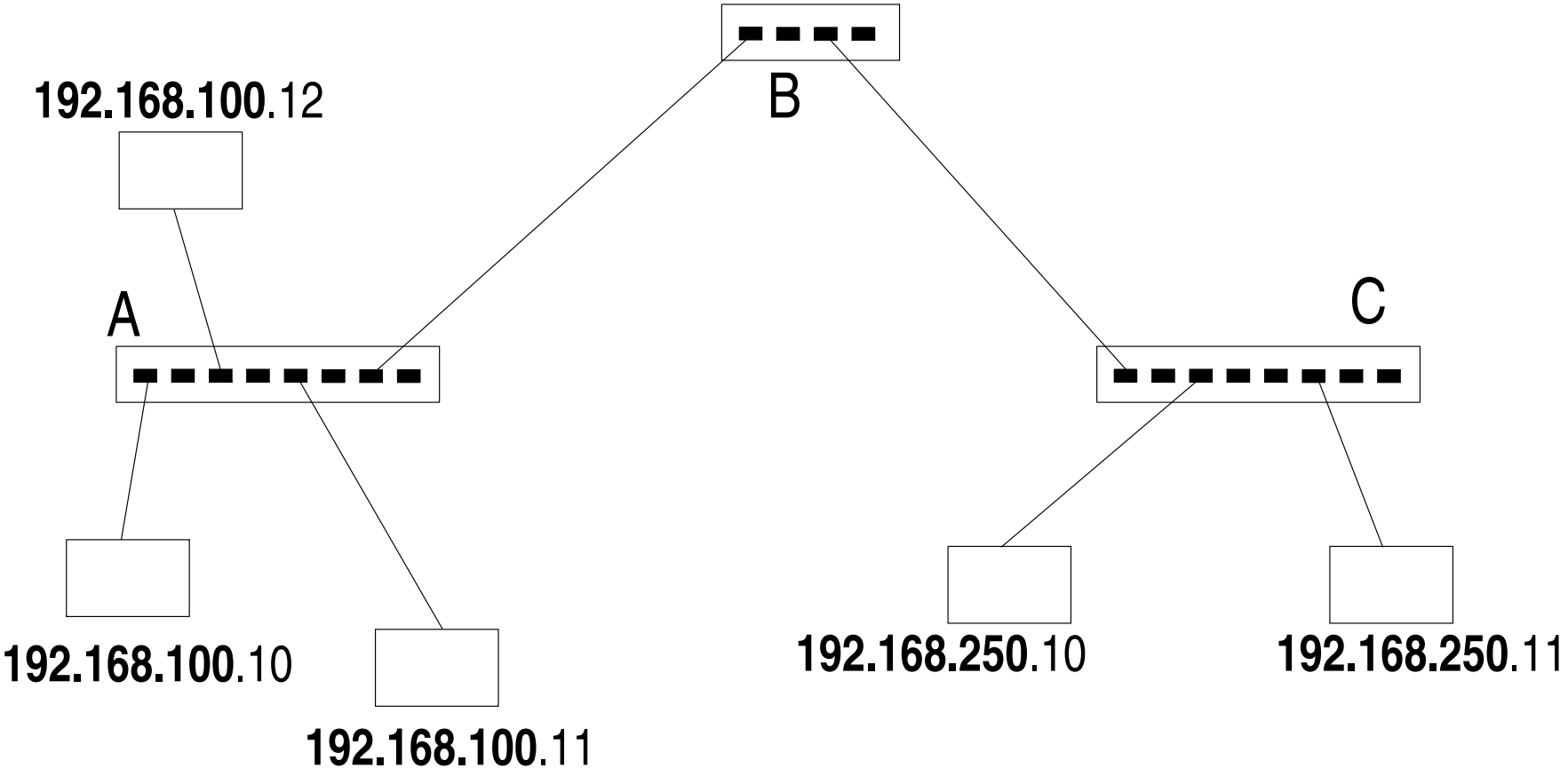


Adresses MAC et adresses IP : Neighbor Discovery Protocol (IPv6)

- Neighbor Discovery Protocol (NDP)
- Intégré dans Internet Control Message Protocol version 6 (ICMPv6) – rfc4443
- Fonctionne de manière similaire à ARP (IPv4)

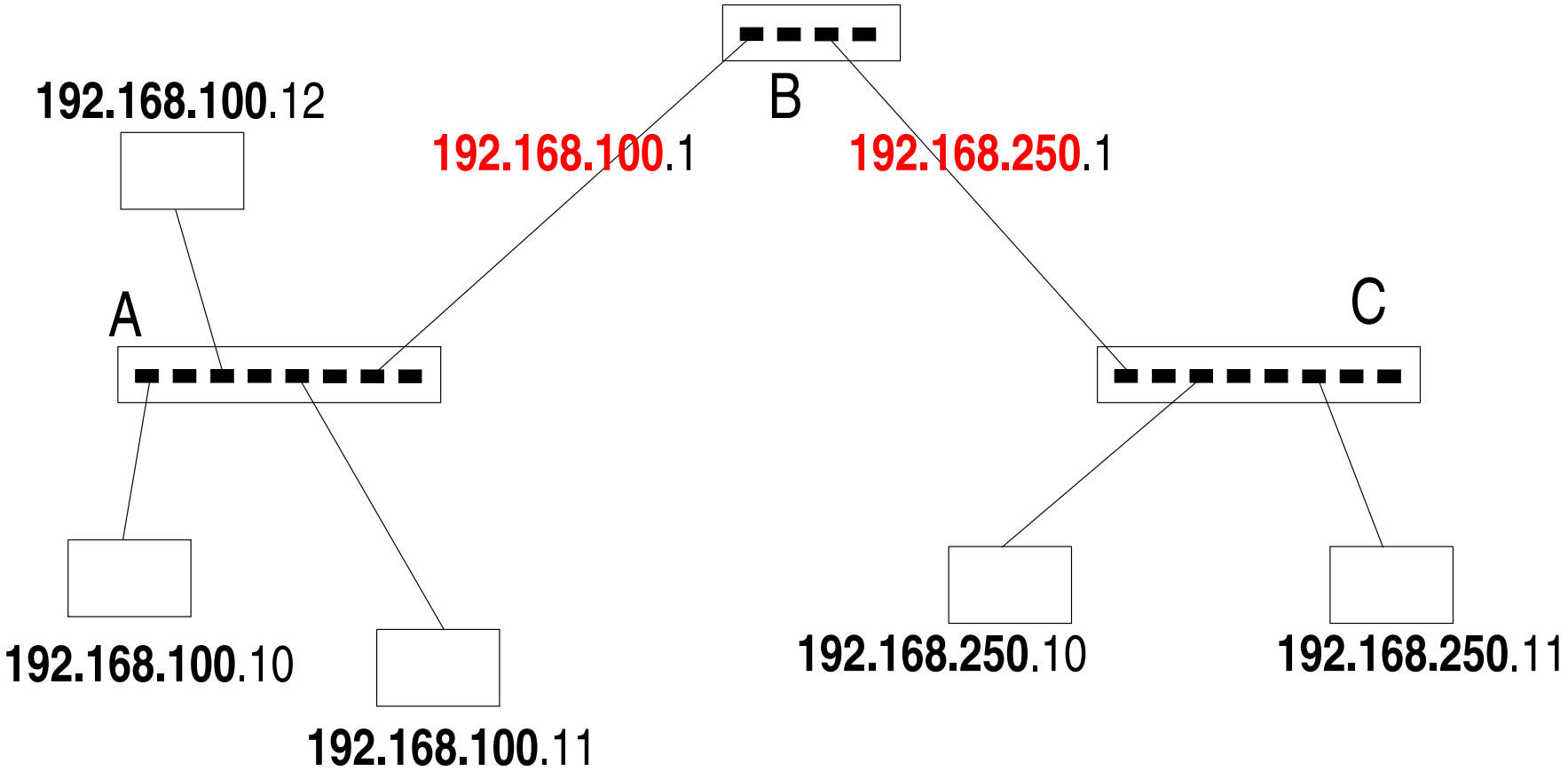
Question...

Quelle est la nature des organes A, B et C ?

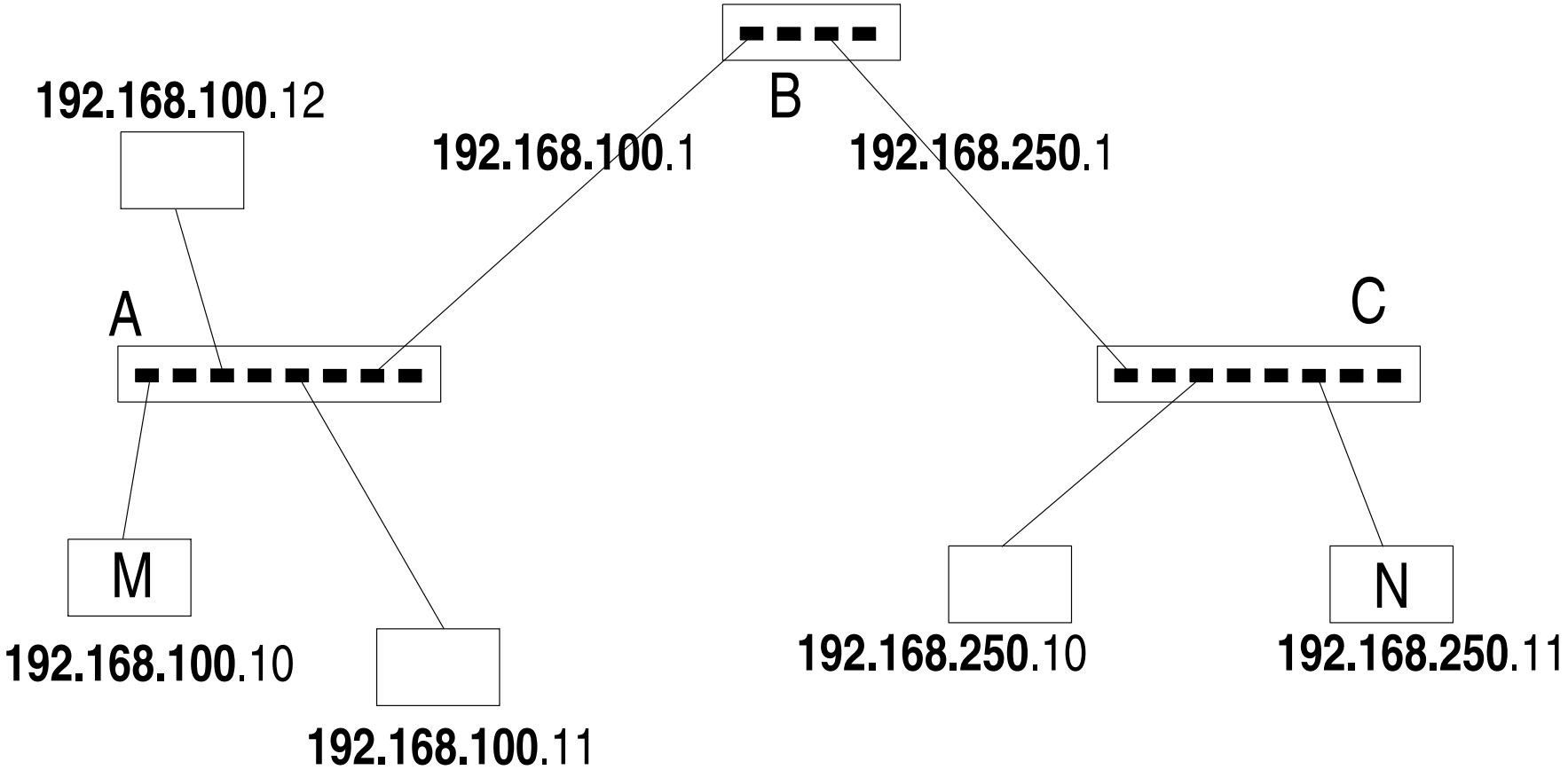


B est un routeur

Alors...

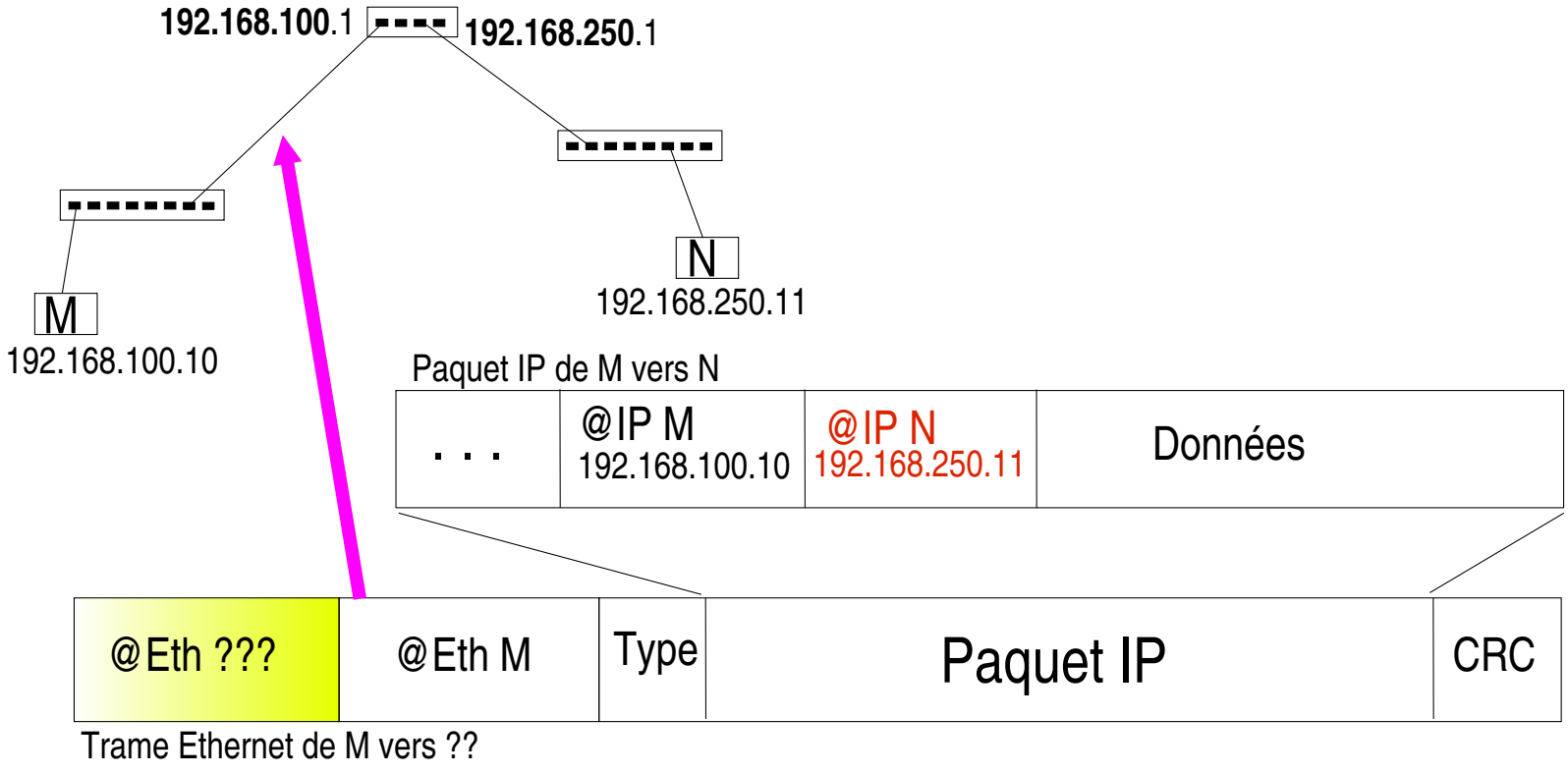


Comment envoyer un paquet de M vers N ?

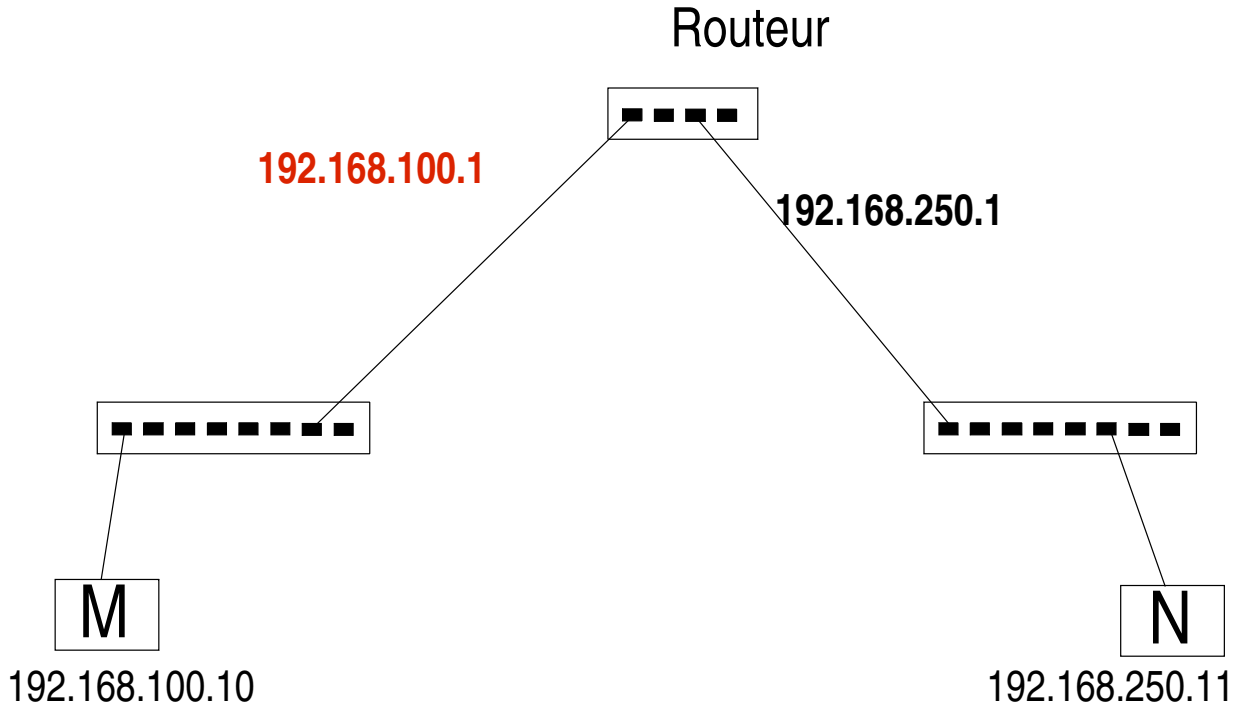


Paquet IP de M vers N, et son porteur... I

En réseau local



Comment, en M, déterminer l'adresse Mac de la trame Ethernet qui va emporter le paquet vers sa destination ?



En M il faut une table de routage qui dit : « pour aller en 192.168.250.11 passer par 192.168.100.1 »
À l'aide de cette table il sera facile de faire une résolution ARP pour trouver l'adresse MAC du routeur. L'adresse destination de la trame Ethernet sera celle du routeur

Paquet IP de M vers N, et son porteur... II

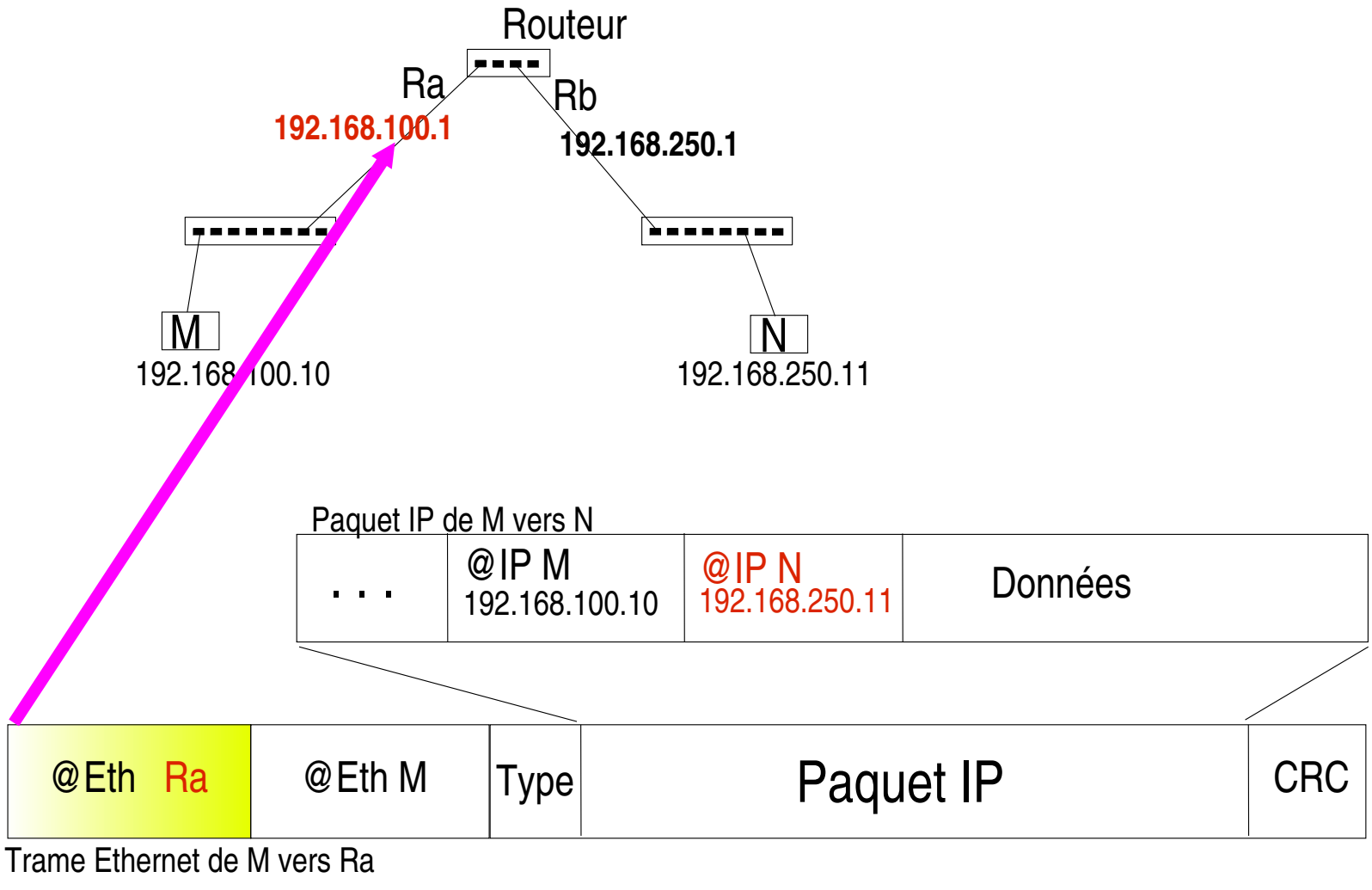
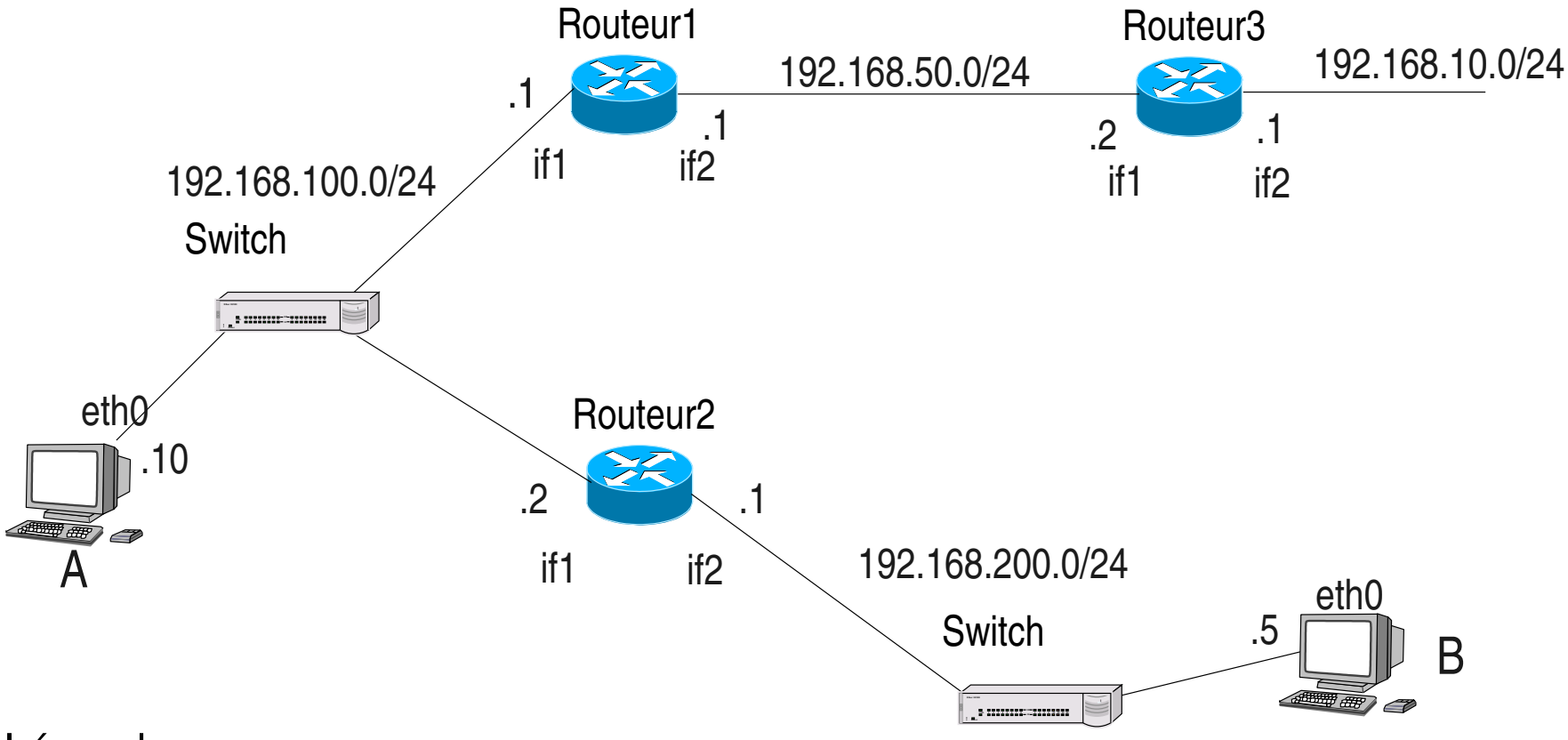


Table de routage

Chaque «entrée» dans la table contient au moins

- Une direction (réseau ou machine)
- Une indication de route
  - machine par laquelle les paquets doivent être acheminés
    - Cette machine doit être accessible
  - interface locale
- Un coût
  - Notion de «distance» ou de «coût»
    - Nombre de sauts nécessaires
    - Débit
    - ...

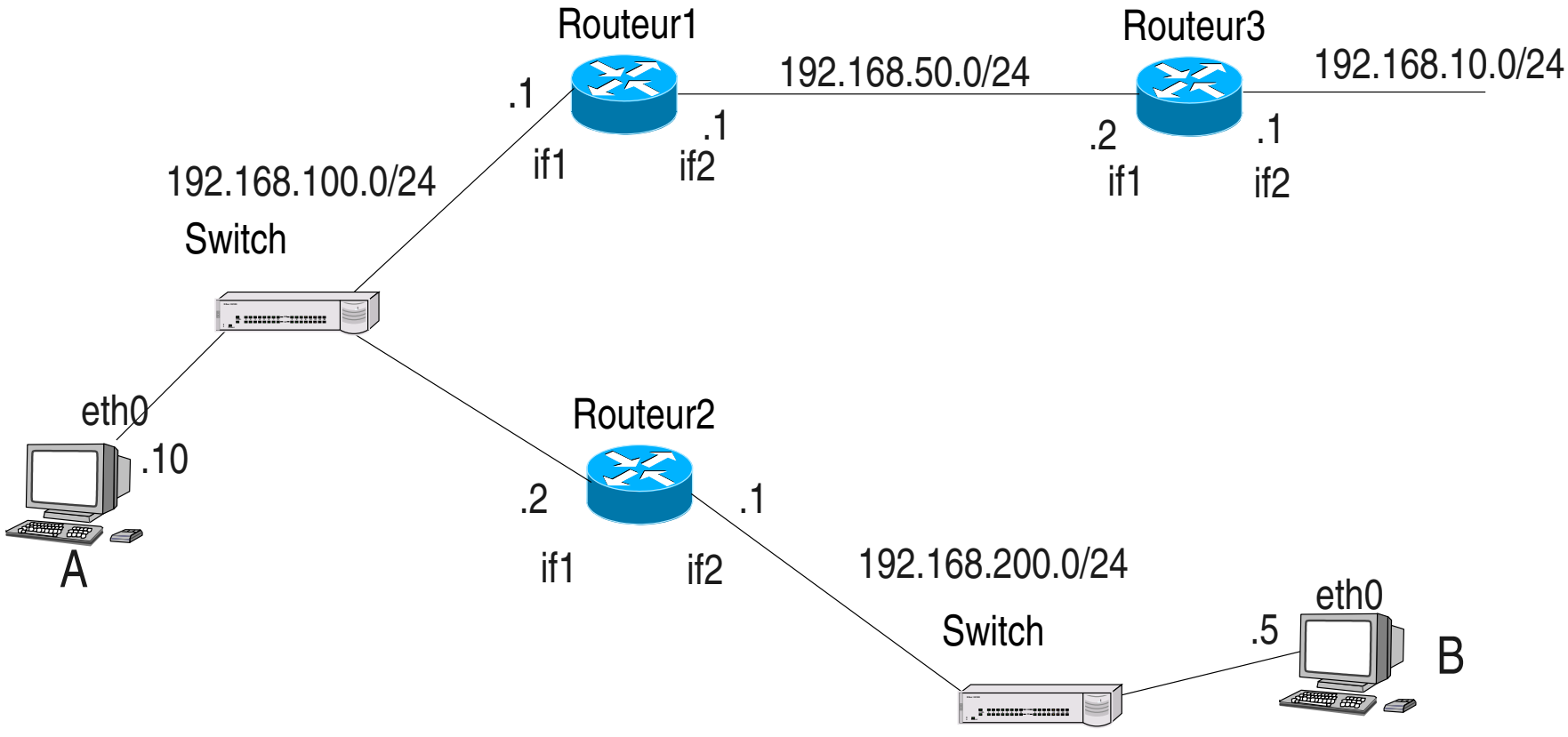
Table de routage



Légende :

`192.168.X.0/24` : adresse de réseau et netmask
`.x` (`.1` ou `.2`, etc.) : partie machine de l'adresse de l'interface
`if1`, `if2`, `eth0` : nom d'interface réseau sur la machine

Table de routage



Considérons les tables de routage correctement servies partout sauf en A
A peut atteindre if1 de Routeur3 (192.168.50.2), mais pas if2 (192.168.10.1)
Pourquoi ne peut-on pas dire en A : pour aller en 192.168.10.0/24 passer par 192.168.50.2 ? *Que faut-il dire ?*

Exemple de table de routage sous Windows

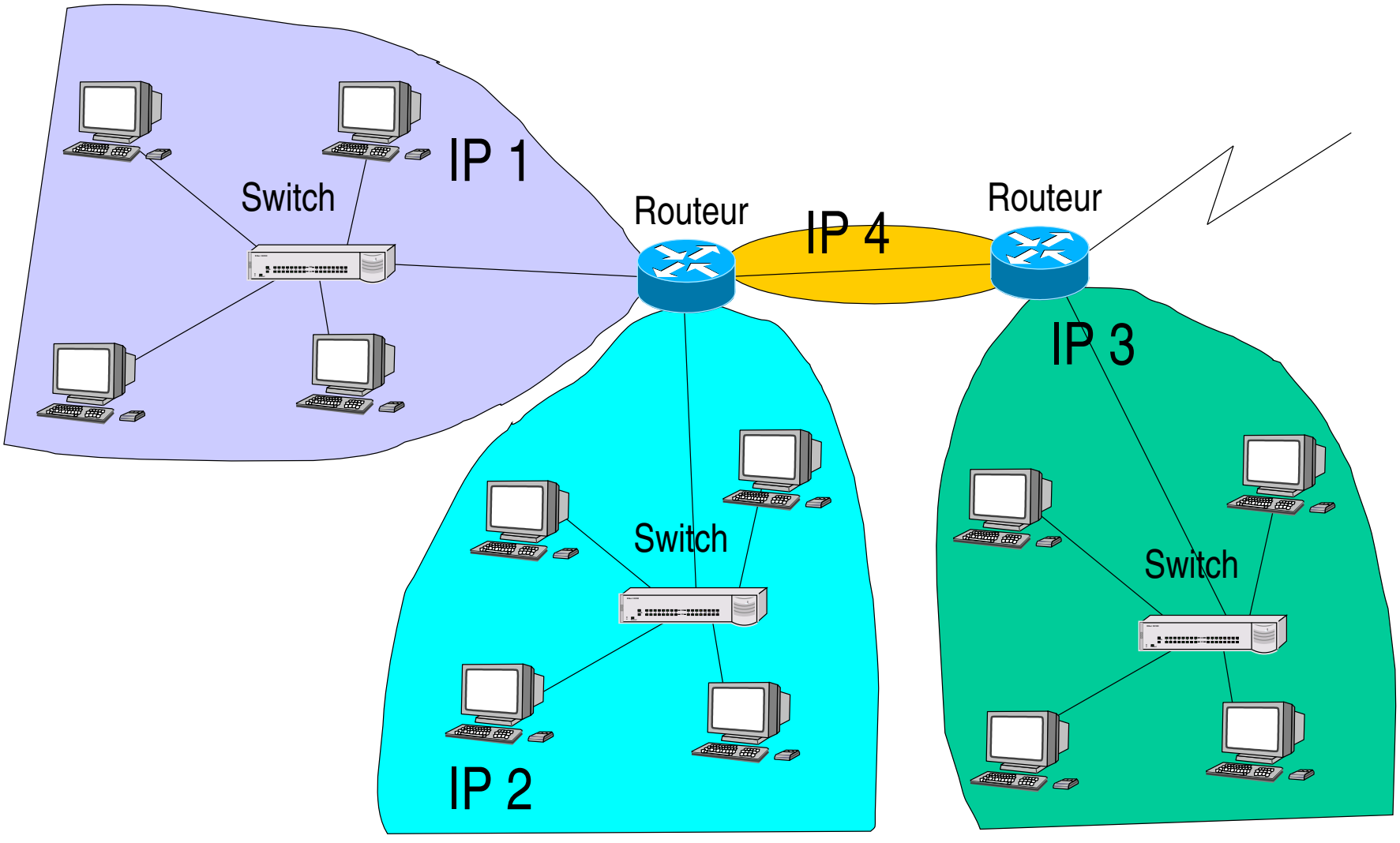

```
C:\>route print
===========================================================
Liste d'Interfaces
0x1 ........................... MS TCP Loopback interface
0x1000003 ...00 01 02 6e 7c 46 ...... 3Com EtherLink PCI
===========================================================
Itinéraires actifs:
  Destination réseau     Masque réseau   Adr. passerelle   Adr. interface   Métrique
            0.0.0.0           0.0.0.0      192.44.75.1    192.44.75.184      1
          127.0.0.0         255.0.0.0        127.0.0.1        127.0.0.1      1
        192.44.75.0     255.255.255.0     192.44.75.184    192.44.75.184      1
      192.44.75.184   255.255.255.255        127.0.0.1        127.0.0.1      1
      192.44.75.255   255.255.255.255     192.44.75.184    192.44.75.184      1
          224.0.0.0         224.0.0.0     192.44.75.184    192.44.75.184      1
    255.255.255.255   255.255.255.255     192.44.75.184    192.44.75.184      1
Passerelle par défaut: 192.44.75.1
===========================================================
```

Et votre système dans tout ça...

Quel est votre paramétrage ?

- Sous Windows (Ouvrez une fenêtre de commande et testez les commandes suivantes :)
  - `ipconfig` (`winipcfg` sous Windows 9x/ME) avec le sélecteur `/all`
  - `route print` pour afficher la table de routage
  - `nslookup` pour traduire des noms de machine en adresse IP et inversement
  - `arp -a` si vous êtes sur un LAN pour lire la table de traduction arp
  - `tracert`
- Sous Linux (Ouvrez un terminal et testez :)
  - `ifconfig` avec ou sans `-a`
  - `route` avec ou sans `-n`
  - `nslookup` ou `host`
  - `arp -a`
  - `traceroute` (avec ou sans `-n`)

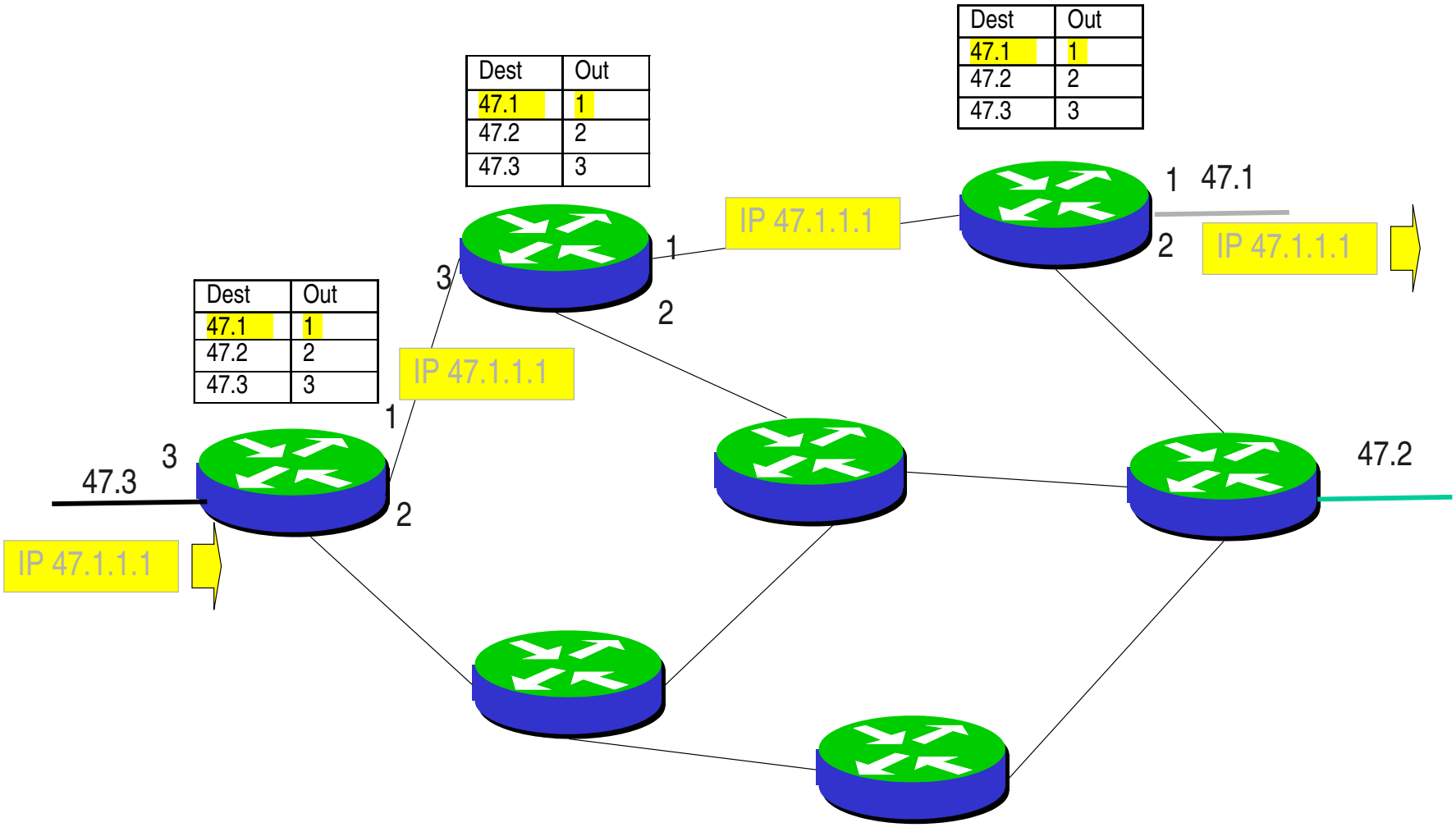
Le réseau IP de l'entreprise



Fonctions des routeurs

- Chaque paquet IP entrant dans le routeur voit son adresse destination examinée et comparée sur un certain nombre de bits avec le contenu d'une table (table de routage) associant des directions avec une interface de sortie
  - Cette fonction est aussi mise en œuvre dans les machines terminales
  - Le champ TTL est décrémenté par chaque routeur, le champ *cheksum* doit être recalculé à chaque fois

Acheminement bond par bond



L'établissement de la table de routage

- Routage statique
  - La table est remplie «à la main» via une commande spécifique ou une interface graphique
- Routage dynamique
  - Via des protocoles appropriés
  - Permettent de détecter des ruptures de liens et de reconfigurer les tables de routage de manière dynamique

Le routage dynamique

- Deux grands types de mise en oeuvre
  - Les protocoles internes
    - Pour les réseaux d'entreprises
    - Notion de système autonome (*autonomous system*)
  - Les protocoles externes
    - Pour les réseau d'opérateurs
    - A prendre en compte aussi par les entreprises pour leurs routeurs de raccordement aux opérateurs

Le principe de fonctionnement du routage dynamique

- La configuration des interfaces d'un routeur «apprend» à celui-ci les numéros des réseaux sur lesquels il est raccordé
  - Le routeur connaît donc déjà quelques réseaux
  - Il peut communiquer cette connaissance aux autres autres routeurs situés sur les mêmes segments de réseau que lui, via un protocole approprié
  - Les autres routeurs peuvent faire de même

../..

Le principe de fonctionnement du routage dynamique II

../..

- Au bout d'un certain temps tous les routeurs connaissent tout sur tout les autres
- Les routeurs continuent périodiquement à s'envoyer des informations. Si l'un d'eux tombe en panne, les autres, ne recevant plus ses informations, le considèrent comme hors service et inhibent les routes qui passaient par lui. Ils peuvent aussi déterminer des routes alternatives

- Deux grands types d'algorithmes
  - Le vecteur de distance
    - Les routeurs s'échangent leur table de routage (destination et distance ou coût) entre voisins
    - Protocole de type interne RIP (Routing Information Protocol), le plus répandu mais peu performant
  - L'état de liens
    - Tous les routeurs d'un domaine s'échangent leurs informations de routage
    - Protocole de type interne OSPF (Open Shortest Path First), plus performant que RIP mais complexe

Routage interne et externe

- Tous les routeurs du monde ne doivent pas avoir la copie des tables de routage de tous les autres routeurs
- Concept de domaines
  - Un domaine est une entreprise, un campus, une entité restreinte
  - Des protocoles de routage sont adaptés au routage intra domaine
    - RIP, OSPF
- Routage inter domaines
  - Les informations sont agrégées
  - Notion d'autonomous système
  - Protocole fortement utilisé : BGP (Border Gateway Protocol)

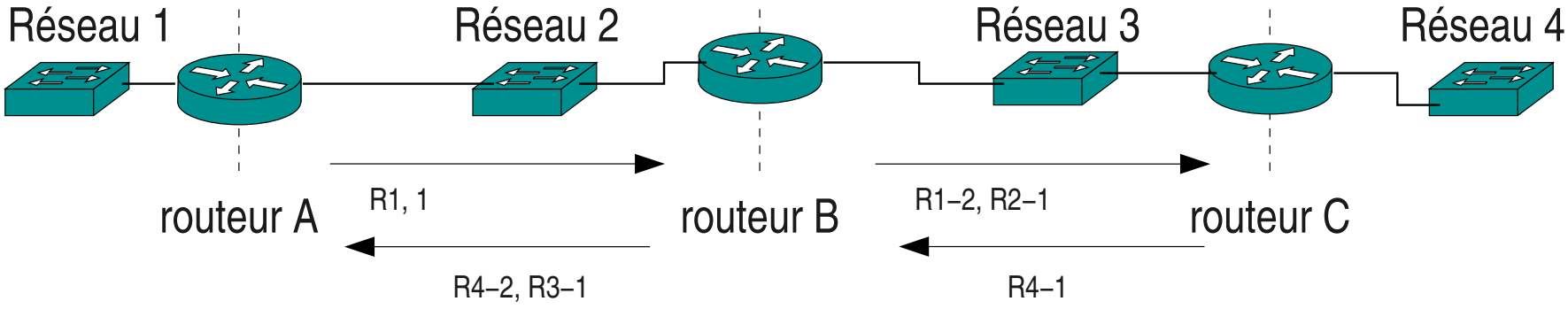
Protocoles de routage les plus répandus I

- RIP : Routing Information Protocol
  - Routage intra-domaine
  - Type vecteur de distance
  - Peu performant
  - Pauvre notion de coût (nombre de saut)
  - Facile à mettre en œuvre
- OSPF : Open Shortest Path First
  - Routage intra-domaine
  - Famille état des liens (link state)
  - Notion de coût plus riche (par défaut : 108/bande_passante)
  - Routage hiérarchique avec délimitation d'aires

Protocoles de routage les plus répandus II

- BGP : Border Gateway Protocol (rfc1771)
  - Routage inter-domaines
  - Prend en compte le routage sans classe (CIDR rfc1466)
  - Véhiculé par TCP

- Vecteur de distance : *direction,coût*



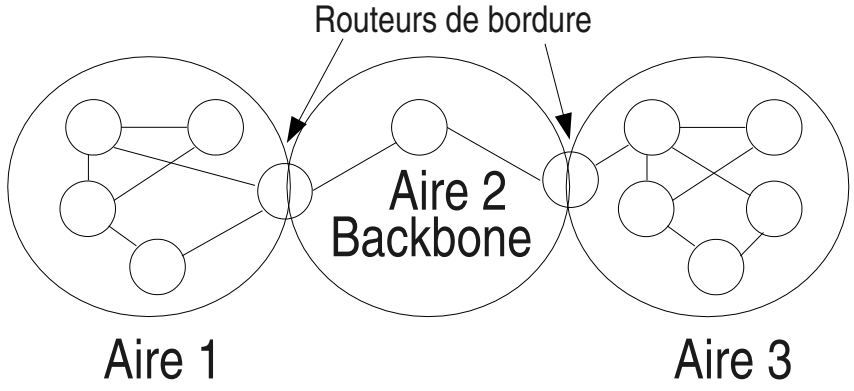
Vers	Via	Coût
R1	loc.	1
R2	loc	1
R3	B	2
R4	B	3

Vers	Via	Coût
R2	loc.	1
R3	loc	1
R1	A	2
R4	C	2

Vers	Via	Coût
R3	loc.	1
R4	loc	1
R2	B	2
R1	B	3

- Pourquoi les routeurs n'annoncent-ils pas tout le contenu de leur table ?

- rfc1583, 1587, 2328
- Routage hiérarchique : notion d'aires



  - Chaque aire doit être reliée à l'aire backbone qui assure le transit entre deux aires
  - Dans chaque aire les routeurs ont une connaissance complète du routage dans leur aire
  - Les routeurs de bordure d'aires ont une connaissance des aires qu'ils raccordent, ils propagent dans chaque aire une connaissance agrégée des autres aires

Un exemple de découverte de route

Sous Windows :

```
C:\>tracert www.redhat.com
Détermination de l'itinéraire vers www.redhat.com [209.132.177.50]
avec un maximum de 30 sauts :
 1 <10 ms  16 ms <10 ms galaxie-75.enst-bretagne.fr [192.44.75.1]
 2 <10 ms <10 ms <10 ms 193.50.69.89
 3 <10 ms <10 ms <10 ms 193.48.78.17
 4 <10 ms  15 ms <10 ms PAO-Rennes2-TR.rrb.ft.net [195.101.145.25]
 5  16 ms <10 ms  16 ms peering-GIP.rrb.ft.net [195.101.145.6]
 6 <10 ms  16 ms <10 ms rennes-a0-0-2.cssi.renater.fr [193.51.181.126]
 7  16 ms  16 ms  15 ms caen-pos1-0.cssi.renater.fr [193.51.180.17]
 8  15 ms  16 ms  15 ms rouen-pos1-0.cssi.renater.fr [193.51.180.22]
 9  31 ms  16 ms  16 ms nri-a-pos6-0.cssi.renater.fr [193.51.179.21]
10  32 ms  15 ms  16 ms 193.51.185.1
11  15 ms  16 ms  15 ms P11-0.PASCR1.Pastourelle.opentransit.net [193.251.241.97]
12  93 ms  94 ms  94 ms P12-0.NYKCR3.New-york.opentransit.net [193.251.241.134]
13  94 ms  94 ms  94 ms ft-gw.n54ny.ip.att.net [192.205.32.137]
14  93 ms  94 ms  94 ms tbr1-p010401.n54ny.ip.att.net [12.123.3.57]
15 110 ms 125 ms 109 ms tbr1-cl1.cgcil.ip.att.net [12.122.10.2]
16 110 ms 109 ms 125 ms tbr2-cl2.cgcil.ip.att.net [12.122.9.134]
17 172 ms 172 ms 172 ms tbr2-cl7.sl9mo.ip.att.net [12.122.10.46]
18 125 ms 140 ms 125 ms tbr2-cl6.dlstx.ip.att.net [12.122.10.90]
19 172 ms 172 ms 172 ms gbr2-p40.phmaz.ip.att.net [12.122.10.86]
20 172 ms 172 ms 172 ms gar2-p370.phmaz.ip.att.net [12.123.142.49]
21 172 ms 172 ms 172 ms mdf1-gsr12-1-pos-7-0.phx1.attens.net [12.122.255230]
22 141 ms 156 ms 141 ms mdf1-bi8k-2-eth-2-1.phx1.attens.net [63.241.128.158]
23  *   *   *  Délai d'attente de la demande dépassé.
```

Le protocole ICMP (IPv4)

Internet Control Message Protocol (rfc 792)

- Sert à véhiculer des messages d'erreur ou de demande d'information
  - Demande d'écho et réponse (commande ping)
  - Destination non accessible
    - Le réseau ne peut être atteint
    - La fragmentation est nécessaire et le bit D est à 1 (PMTU discovery)
    - etc
  - Redirection, il existe une meilleure route
  - Durée de vie dépassé
  - etc

Internet Control Message Protocol v6 (rfc 4443)

- Mêmes fonctions que ICMP (IPv4)
  - Erreurs diverses, ping, redirections, etc.
- En plus :
  - Neighbor Discovery Protocol (équivalent du ARP pour IPv4)
  - Router Solicitation, Router Advertisement : Auto-configuration de l'adresse. Au démarrage, un host recherche s'il y a un routeur présent sur LAN ; le routeur diffuse le préfixe IPv6 à utiliser ; le host se choisit une adresse dans ce préfixe.

- Le protocole DHCP : Dynamic Host Control Protocol (DHCPv4 rfc2131, DHCPv6 rfc3315)
  - Utilisé pour obtenir une adresse IP de manière dynamique ainsi que des informations diverses : serveur DNS, domaine, netmask, routeur par défaut
- Le service de nommage : le DNS
  - Domain Name Service (rfc 1034 et 1035)
    - Permet d'associer un nom à une adresse IP
    - Offre un protocole pour retrouver ce nom en partant de l'adresse ou l'adresse en partant du nom

Le Domain Name Service - DNS

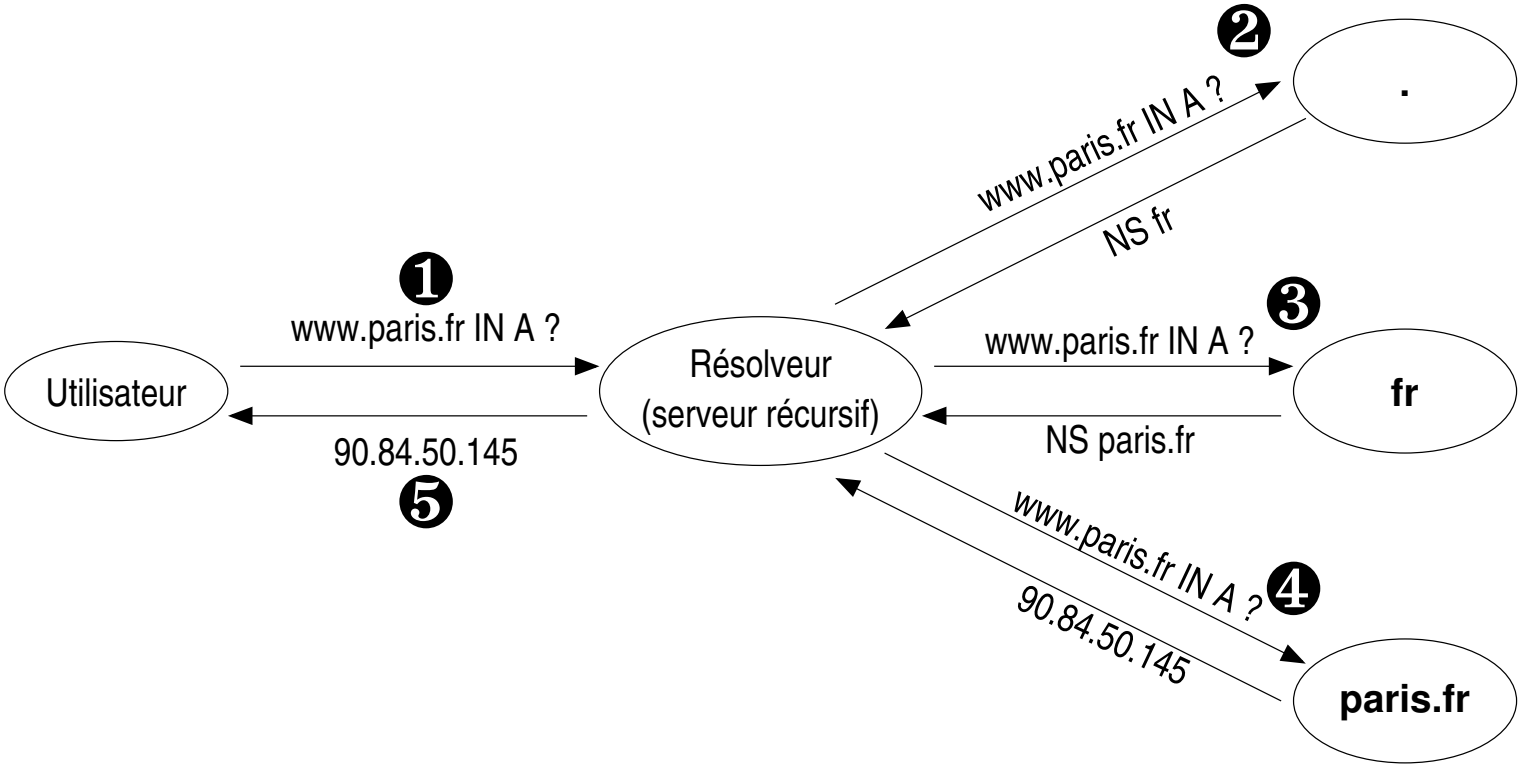
- Service réparti de nommage
- Répartition mondiale
- Gestion indépendante de chaque domaine, pas d'autorité absolue de référence, chaque administrateur est libre de ses choix
- Permet de connaître :
  - les numéros de machine à partir de leur nom et inversement
  - les machines gestionnaires de courrier électronique pour les domaines
  - et bien d'autres informations

DNS : notion de domaine

- Un domaine DNS peut être composé de plusieurs réseaux IP (identifiés par des numéros IP différents)
- Un domaine peut être subdivisé en sous domaines
- Les domaines forment une arborescence
- La racine de l'arborescence est subdivisée en «top level domains» : .edu, .org, .gov, .mil, .net, .com, .us pour les USA, .fr, .ca, .uk, .de, .jp, etc pour le reste du monde
- De nouveaux «top level» apparaissent ou vont apparaître : .int, ...

DNS : exemple de requête

- Des *résolveurs* (forward les requêtes et cachent les réponses)
- Des *serveurs faisant autorité* (qui possèdent l'information originelle)



Serveurs DNS faisant autorité
pour leurs zones respectives

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

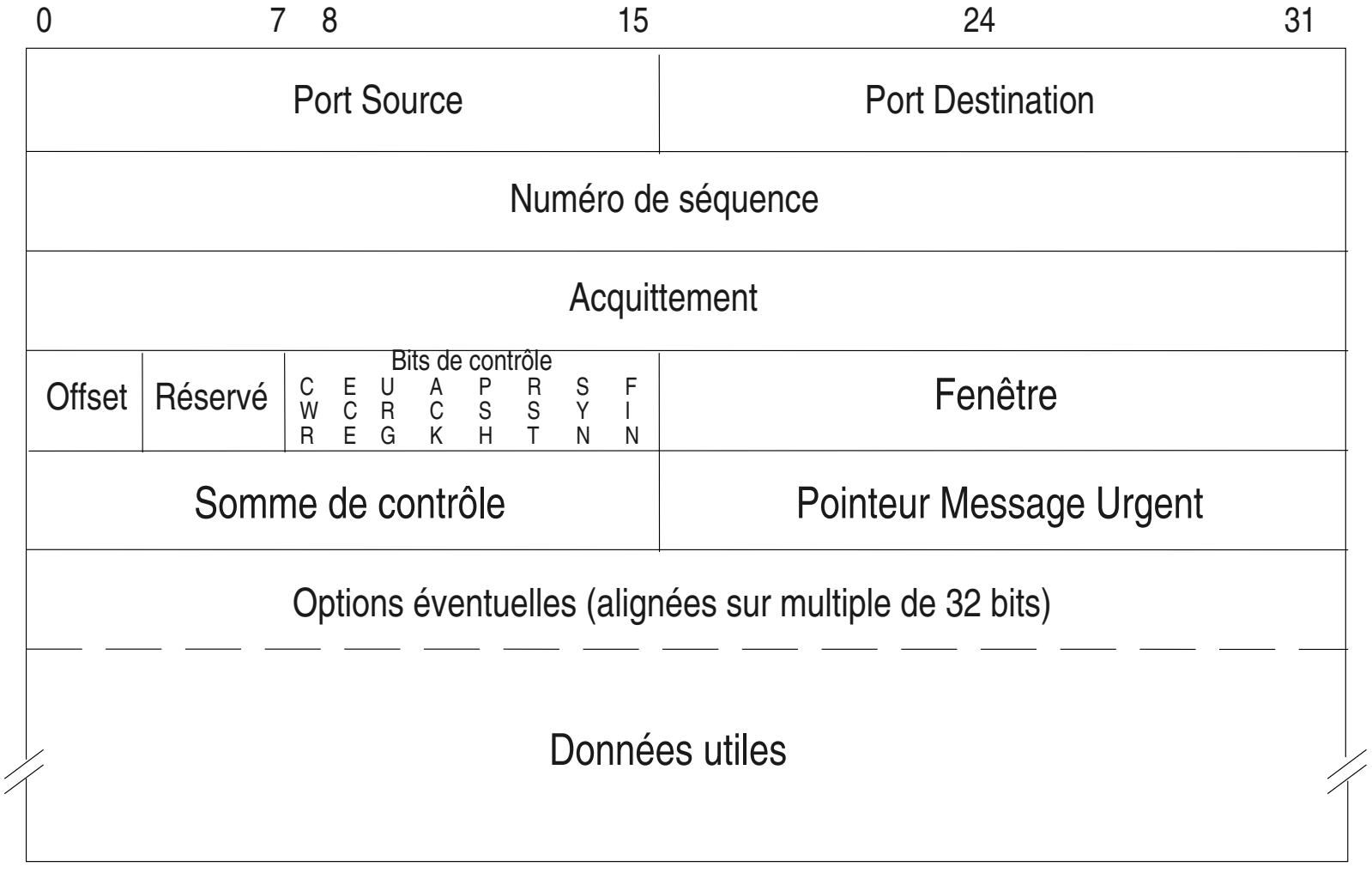
Quatrième partie

Les protocoles TCP et UDP
Quelques protocoles applicatifs

Les protocoles de transport au dessus de IP

- TCP : Transmission Control Protocol rfc793
  - Transport en mode «connecté»
  - Contrôle de flux et détection d'erreur avec retransmission
- UDP : User Datagram Protocol rfc763
  - Mode datagramme
  - Pas de contrôles, pas d'assurance de délivrance
- Nouveaux, expérimentaux
  - DCCP : Datagram Congestion Control Protocol, rfc4340
  - SCTP : Stream Control Transmission Protocol, rfc4960
  - MPTCP : Multipath TCP, rfc6826

Le segment TCP



CWR	ECE	URG	ACK	PSH	RST	SYN	FIN

- ECE Explicit Congestion Notification-Echo (rfc 3168), l'entité TCP recevant un segment contenant ce bit à 1 doit réduire son débit
- CWR Congestion Window Reduced (rfc 3168) réponse à la réception du bit ECE pour indiquer que la requête a bien été prise en compte
- URG indique que le champ *Urgent Pointer* est significatif.
- ACK indique que le champ *Acknowledgement Number* est significatif.
- PSH fonction *PUSH*. Les données doivent être immédiatement remises à la couche supérieure.
- RST *reset*, la connexion est rompue.
- SYN synchronisation des indicateurs numéro de séquence. Segments d'initialisation de connexion.
- FIN terminaison de connexion.

Exemple : segment IP et TCP dont le bit ACK est à 1

```
 0: 0800 2074 ef05 0800 0914 18e7 0800 4500
16: 0028 8954 0000 4006 db07 c02c 4b13 c02c
32: 4b08 1770 fdbe 0162 6e86 8e21 a873 5010 /* 0001 0000 */
48: 2210 bba3 0000 023a b3a1 1829
```

Les données «urgentes»

- Données appelées parfois "hors bande" ou *Out of Band* (OOB)
- Données à traiter en priorité par la couche réceptrice
- Elles sont véhiculées dans le flux normal en suivant le chemin normal. IP n'est pas sensible à ces données, leur caractère "urgent" est significatif seulement aux extrémités
- L'arrivée de ces données a un caractère aléatoire pour les applications destinatrices
- Les applications ne lisent pas ces données dans le flux normal
- Une application devant pouvoir accepter de telles données doit avertir le système pour que celui-ci lui envoie une interruption (un signal logiciel) afin qu'elle puisse traiter en priorité la donnée. L'application doit prévoir une routine spéciale de traitement pour la lecture de ces données
- Sémantique mal définie : le RFC6093 recommande aux nouvelles application ne ne plus l'utiliser...

Le bit de contrôle RST

- Utilisé par une entité TCP connectée avec une autre entité TCP distante pour avertir d'un problème
- Une application se terminant normalement fait une fermeture sur le port TCP utilisé (souvent une *socket*), ceci se concrétise par un échange de segments avec le bit FIN positionné et la connexion est rompue
- Si l'application se termine brutalement sans fermer la connexion, l'entité TCP associé envoie vers l'autre extrémité un segment avec le bit RST positionné. L'autre extrémité est ainsi prévenue de la terminaison de la connexion

TCP est orienté connexion

- La connexion TCP :
  - Relation établie point à point entre les deux extrémités
  - Normalement transparente aux routeurs (sauf si ceux-ci mettent en œuvre du filtrage où des mécanismes de QoS)
  - Caractérisée par un contexte dans les machines d'extrémité (numéros IP local et distant, ports local et distant)

- La connexion est établie par un échange de paquets initiaux : le *three way handshake*
  - Ce n'est pas une connexion au sens strict du terme, il n'y a pas de chemin virtuel établi dans le réseau, les segments TCP sont portés par IP, protocole orienté datagramme
  - La connexion TCP se traduit par un contexte mémorisé dans les machines d'extrémité (le client et le serveur), ce contexte est le quintuplet :

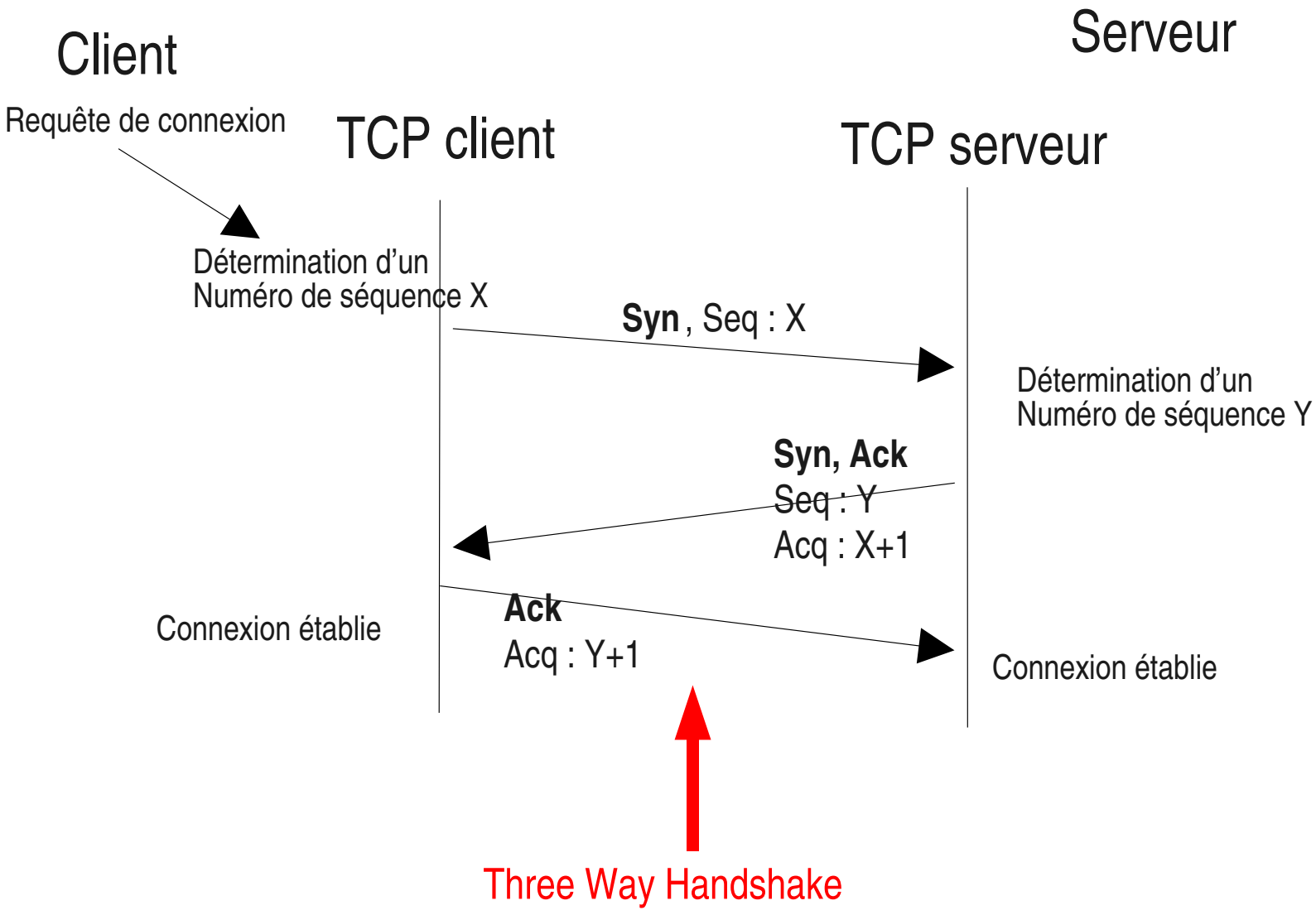
```
[protocole, port local, port distant, @IP locale, @IP distante]
```

- Si les applications connectées n'ont rien à se dire, les entités TCP correspondantes ne s'envoient aucun segment, il n'y a plus alors de trafic
- Il est possible de faire en sorte que les entités s'échangent des segments d'alerte (sans données utiles puisqu'il n'y en a pas à envoyer) toutes les deux heures en positionnant une option dans TCP
- Si l'entité TCP distante ne répond pas au segment d'alerte ceci signifie que l'application associée s'est terminée sans prévenir ou que la machine s'est arrêtée. L'application locale sera prévenue lors de la prochaine lecture du port TCP

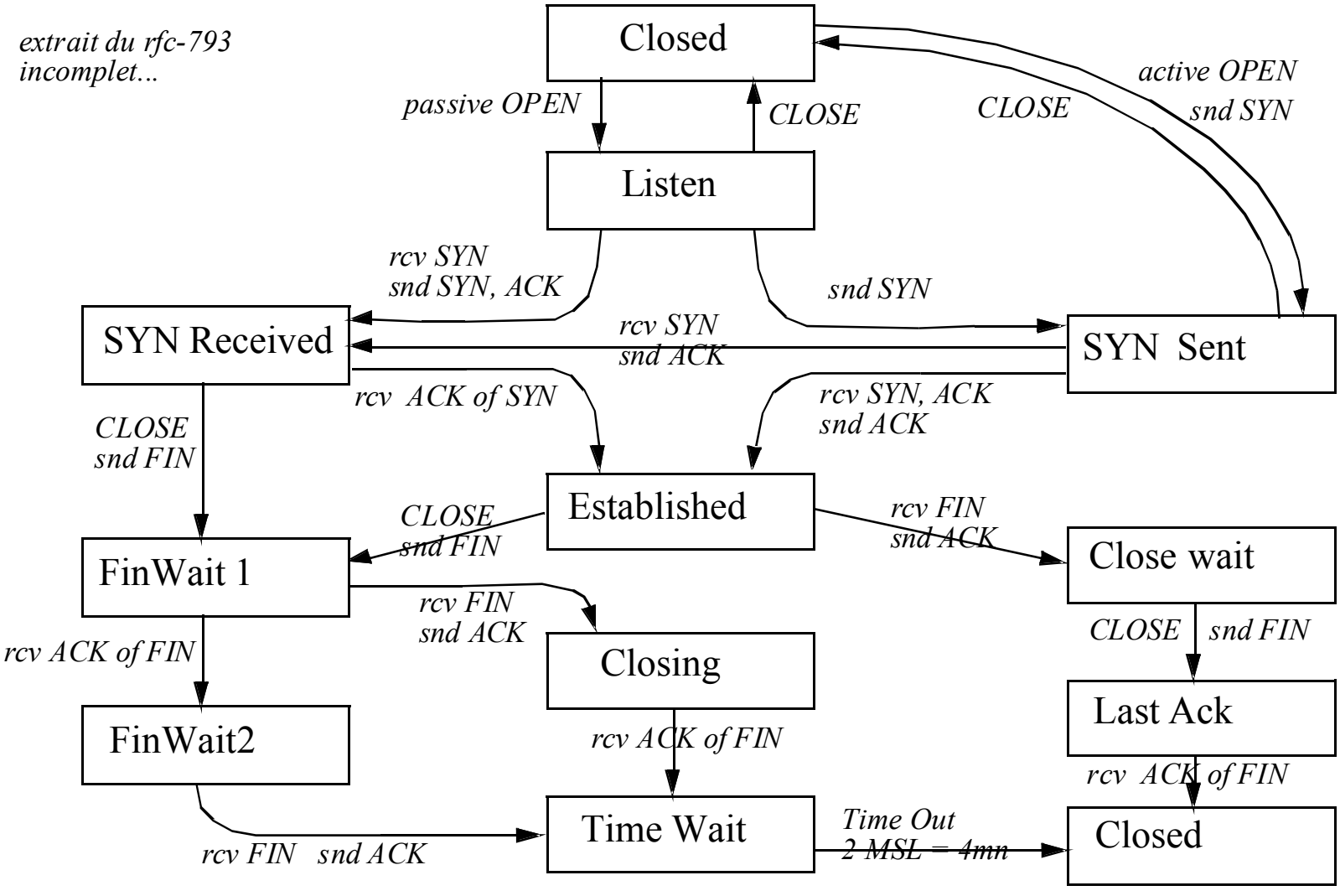
Connexions actives et passives

- Connexion passive
  - Ce n'est pas une connexion mais un point d'accès TCP ouvert par une application en fonction de serveur
  - Le contexte TCP créé attend une requête de connexion émanant d'un client
- Connexion active
  - La connexion réelle, initialisée par une application en mode client

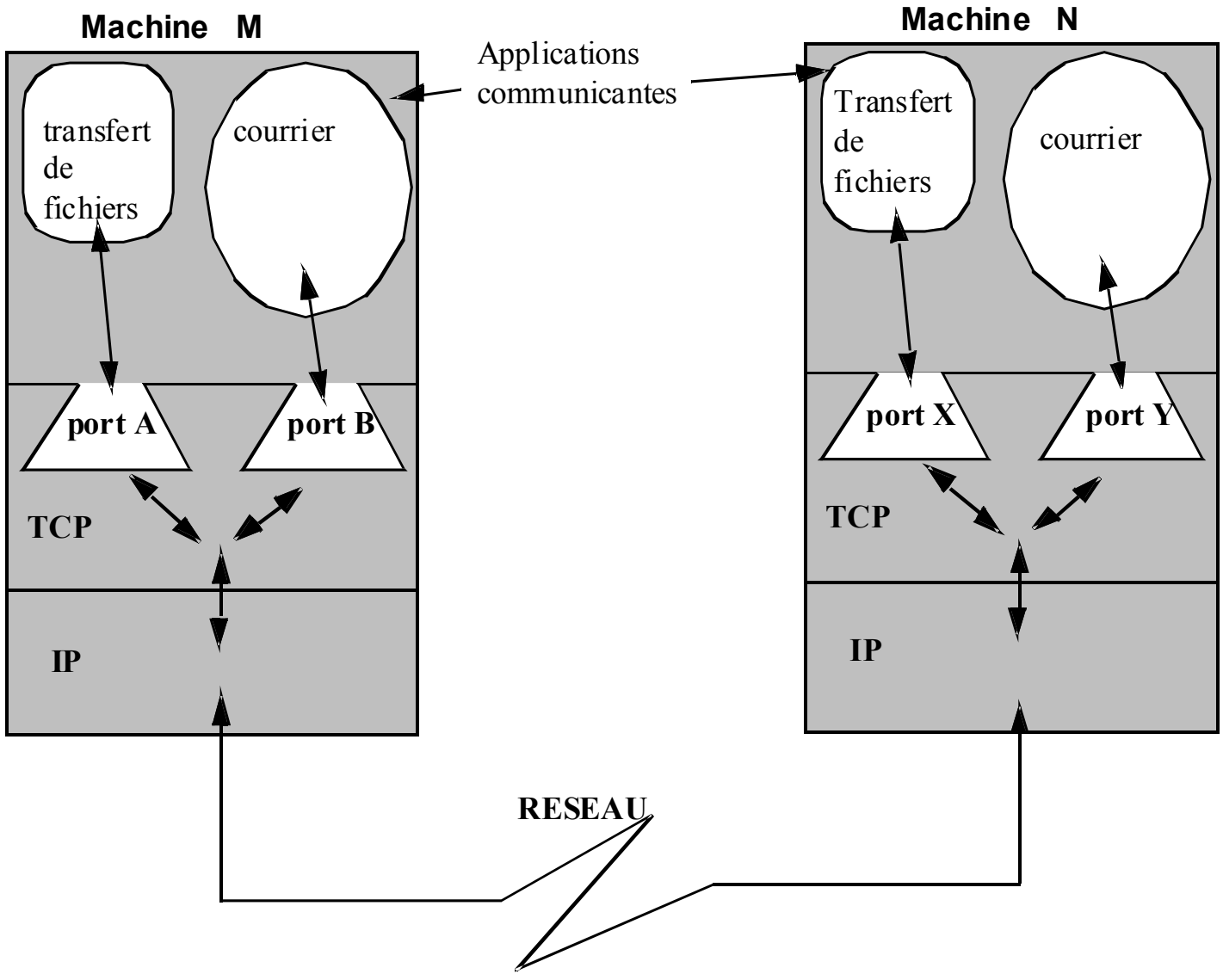
Établissement de la connexion



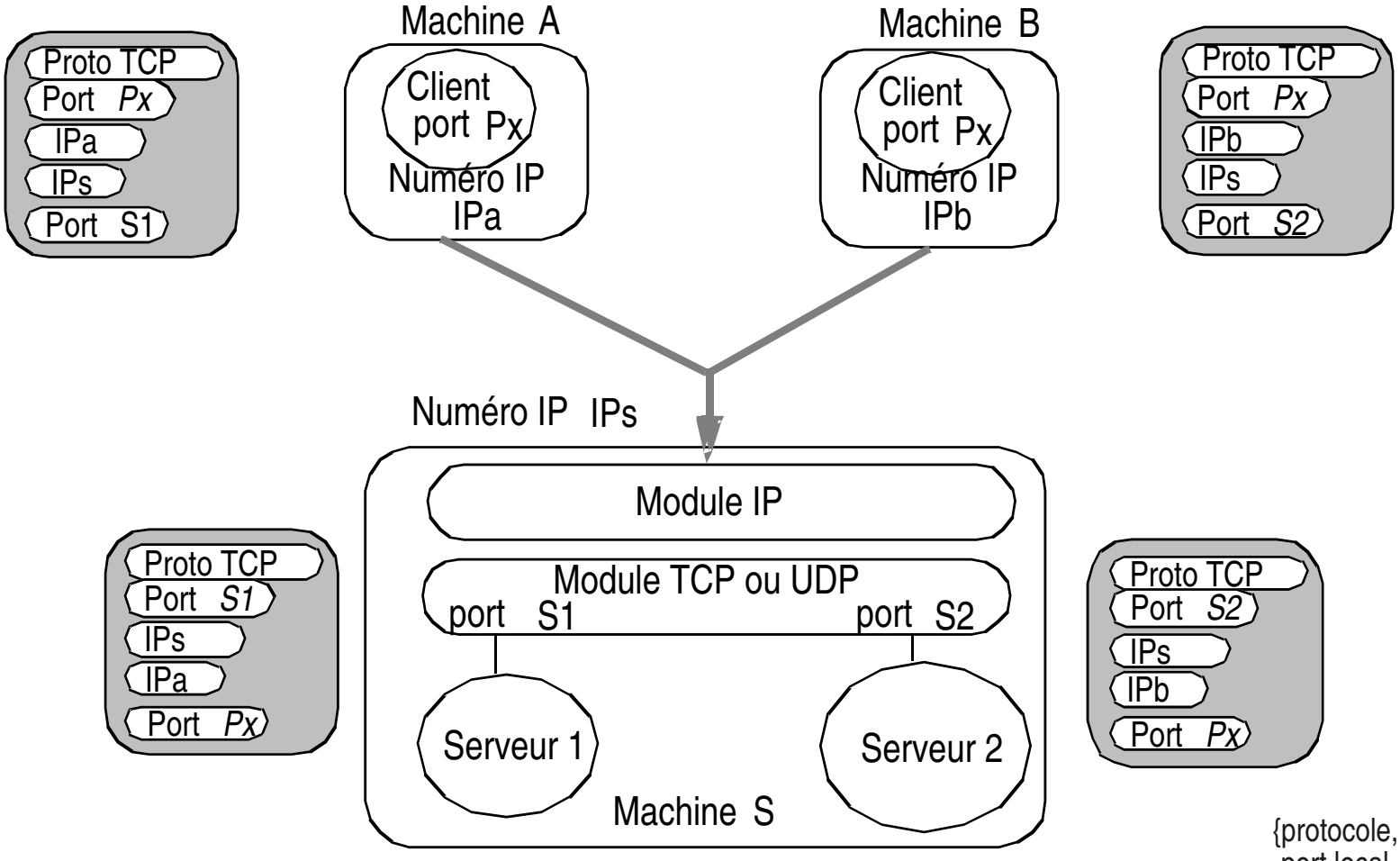
La machine d'états TCP



La notion de port



Identification des «connections»



Identification des connexions sans ambiguïté par le quintuplet {protocole, port local, n° IP local, port distant, n° IP distant}

- Caractéristiques
  - Le champ offset de l'entête étant codé sur 4 bits (il indique le nombre de mots de 32 bits de l'entête), la longueur max de l'entête est de 15 mots de 32 bits (15dec = 1111bin). L'entête minimum étant de 5 mots de 32 bits (20 octets), la longueur du champ option est de 10 mots de 32 bits max.

- Options standards
  - mss : *maximum segment size*, permet de négocier la taille maximale des segments TCP afin d'éviter des segmentations coûteuses. cette option est utilisées au moment de la connexion. Longueur 4 octets.
  - pas d'opération (NOP) : option nulle, permet d'aligner la prochaine option sur un début de mot de 32 bits. Plusieurs NOP peuvent se suivre si besoin. (lg : 1 octet)
  - fin de liste : permet d'aligner la fin des options sur un mot de 32 bits (lg 1 octet)

- Multiplicateur du champ fenêtre (window scale)
  - valeur permettant de multiplier la fenêtre par un multiple d'une puissance de 2 (jusqu'à 2^{14}) (rfc 1323). Ceci permet d'utiliser plus efficacement la bande passante pour des sources à haut débit et/ou des communications à grande distance en permettant un flux le plus continu possible
- Horodatage des données (rfc 1323)
  - L'option contient deux valeurs : un indicateur d'heure d'émission et un acquittement. Une source peut ainsi calculer le temps d'aller et retour sur un chemin en plaçant sa valeur d'horloge dans le premier champ. Lorsque l'acquittement du segment correspondant arrivera, cette valeur sera contenue dans le second champ de cette option. Il sera alors facile de retrancher cette valeur de la valeur courante de l'horloge.

- Négociation de l'acquittement sélectif (rfc 2018)
- Acquittement sélectif : la première option permet d'indiquer qu'une source TCP est capable d'utiliser cette fonctionnalité, la seconde met en œuvre ce type d'acquittement permettant de demander uniquement la retransmission de données perdues en évitant de retransmettre des données suivantes bien reçues

Capture d'une séquence TCP

rfc-1323 et rfc 2018 (relevé avec tcpdump lors d'un début de requête web)

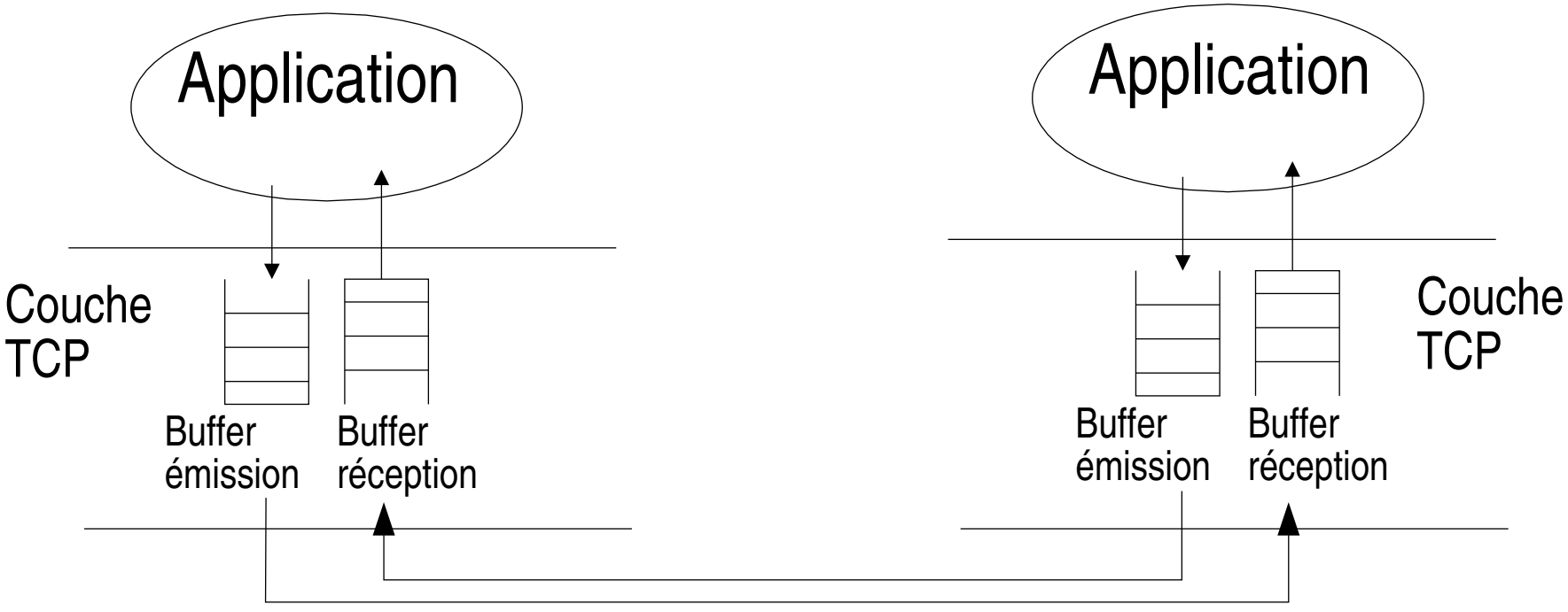
```
linux1.1082 > 206.132.41.203.www: S 2047350264:2047350264(0) win 16060 <mss
1460,sackOK,timestamp 43244276>
206.132.41.203.www > linux1.1082: S 2319802134:2319802134(0) ack 2047350265 win
32120 <mss 1460, sackOK, timestamp 190973015>
linux1.1082 > 206.132.41.203.www: . ack 2319802135 win 16060 <nop,nop,timestamp
43244311 190973015>
linux1.1082 > 206.132.41.203.www: P 2047350265:2047350896(631) ack 2319802135 win
16060 <nop, nop, timestamp 43244311 190973015>
206.132.41.203.www > linux1.1082: . ack 2047350896 win 31856 <nop, nop, timestamp
190973051 43244311>)
206.132.41.203.www > linux1.1082: P 2319802135:2319803583(1448) ack 2047350896 win
31856 <nop, nop, timestamp 190973063 43244311>
linux1.1082 > 206.132.41.203.www: . ack 2319803583 win 14612 <nop,nop,timestamp
43244367 190973063>
206.132.41.203.www > linux1.1082: P 2319803583:2319805031(1448) ack 2047350896 win
31856 <nop, nop, timestamp 190973063 43244311>
```

Le contrôle des connexions : la commande `netstat`

- La commande `netstat` avec l'option `-a` (sous Unix/linux, ou `-p tcp` sous Windows)

```
[linux30]$ netstat -atn
Connexions Internet actives (serveurs et établies)
  Proto   Recv-Q   Send-Q   Adresse locale      Adresse distante          Etat
  tcp          0        0   0.0.0.0:32768       0.0.0.0:*               LISTEN
  tcp          0        0   127.0.0.1:32769     0.0.0.0:*               LISTEN
  tcp          0        0   0.0.0.0:111         0.0.0.0:*               LISTEN
  tcp          0        0   0.0.0.0:6000        0.0.0.0:*               LISTEN
  tcp          0        0   0.0.0.0:22          0.0.0.0:*               LISTEN
  tcp          0        0   0.0.0.0:631         0.0.0.0:*               LISTEN
  tcp          0      272   192.168.100.30:22   192.168.100.18:1994   ESTABLISHED
```

TCP et les applications



Les données sont bufferisées et TCP les émet selon ses propres règles de contrôle de flux. Les applications n'ont pas le contrôle.

- Contrôle de flux d'extrémité à extrémité : le champ window
  - Les routeurs ne mémorisent pas "la connexion" TCP, ils ne peuvent donc pas participer au contrôle de flux. Celui-ci ne peut être réalisé que par les extrémités via le champ fenêtre.
  - Les entités TCP d'extrémités peuvent enregistrer les données reçues dans un tampon mémoire de taille limitée (8, 16, 32 Ko, ...)
  - Si l'application réceptrice ne lit pas suffisamment vite les octets s'accumulent dans le tampon mémoire de réception.

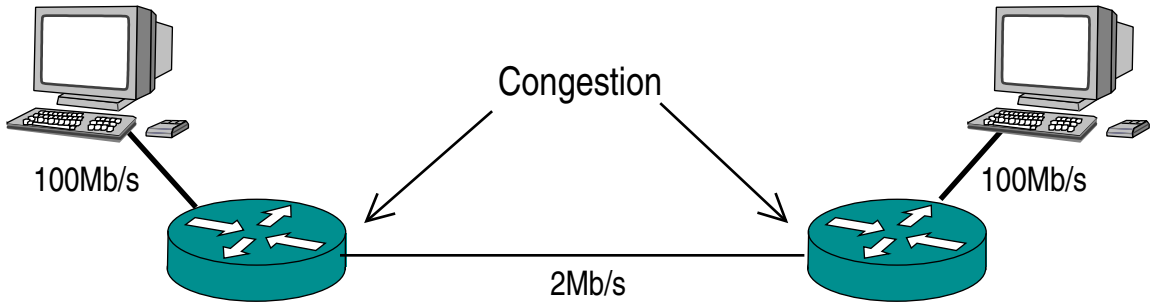
../..

../..

- L'entité TCP réceptrice envoie des acquittements avec une valeur de fenêtre qui diminue pour venir à 0 s'il le faut.
- 5s après avoir reçu un tel acquittement (`win`=0), l'entité émettrice teste le récepteur en lui envoyant un octet et continue ainsi en doublant l'intervalle (jusqu'à une borne de 1 min). Ce moyen permet de forcer le récepteur à renvoyer un acquittement et donner ainsi sa fenêtre. La valeur de 5s est en réalité variable selon les implémentations.
- Dès que les octets reçus par l'entité TCP réceptrice seront consommés par l'application réceptrice, la fenêtre pourra revenir à une valeur différente de 0.

Le problème de la congestion

- Il n'y a pas que les entités d'extrémités en jeu, il y a aussi le réseau. Le mécanisme de fenêtre ne permet pas d'adapter le contrôle de flux aux congestions du réseau.
- S'il y a congestion dans le réseau, des paquets peuvent être perdus, les retransmissions qui en découleront participeront au renforcement de la surcharge totale.
- Le congestion peut être due à des liens saturés mais aussi à des différences de débits entre segments de réseau.



Comment adapter TCP ?

Adaptation de TCP pour le contrôle de congestion

- Détermination du RTT (*Round Trip Time)*
  - Avec les options spécifiques
  - Permet de «régler» un temporisateur de retransmission

Adaptation de TCP pour le contrôle de congestion II

- Le mécanisme du «*slow start*»
  - Une fenêtre de congestion est définie : `cwnd` (*congestion window*). On ne peut pas émettre plus que `cwnd` octets.
  - Juste après la connexion `cwnd` vaut 1mss (*max segment size*).
  - Un RTT plus tard, ce nombre double, et double à chaque RTT jusqu'à un seuil fixé à l'origine à 65535 (on ne peut cependant émettre que `min(cwnd, w)`, `w` étant la valeur de la fenêtre de l'entête TCP).
  - Lorsqu'une congestion a lieu, une perte se produit, le temporisateur expire et provoque une retransmission, le seuil est divisé par 2 et `cwnd` est remis à 1mss.
  - Si le seuil est dépassé, la progression de `cwnd` devient linéaire.

Adaptation de TCP pour le contrôle de congestion III

- Le mécanisme du «*fast recovery*» (entre autres) vient compléter le low start
  - Si on reçoit un même acquittement plusieurs fois, cela signifie deux choses : d'une part c'est qu'on a bien envoyé des segments et que ceux-ci ont été bien reçus (sinon on n'aurait pas reçu d'acquittement du tout) et d'autre part il doit y avoir un «trou» dans la séquence de segments transmis. Il suffit alors de ne retransmettre que le segment qui semble perdu.
  - Lorsque que plusieurs pertes successives ont lieu, ce mécanisme ne suffit plus car il ne permet que de récupérer le premier segment perdu. Il faut alors utiliser le mécanisme des acquittements sélectifs.

Comment éviter la segmentation avec TCP

- Le mécanisme du *Path MTU Discovery (rfc 1191)*
- le problème
  - la couche TCP émission connaît le MTU de l'interface IP de sortie (1500 par défaut sur Ethernet). Sur le chemin vers le récepteur un lien peut avoir un MTU inférieur (700 par exemple), le routeur amont sur ce lien devra segmenter.
- la solution
  - les routeurs sont interdits de segmentation TCP (bit D à 1). Si un paquet se présente de taille trop grande il est rejeté et le routeur qui le rejette émet un paquet ICMP vers l'émetteur. Ce paquet contient une information significative. Voir exemple :

Exemple de PATH MTU Discovery

Exemple : extrait d'un relevé avec tcpdump

- émission du premier paquet (seq = 1, taille 1448=MTU(1500)-TCPhead(20)-TCPoptions(12)-IPhead(20))

```
Em > Rec P 1:1449(1448) ack 1 win 16060 <nop,nop,timestamp 830140
827243>
```

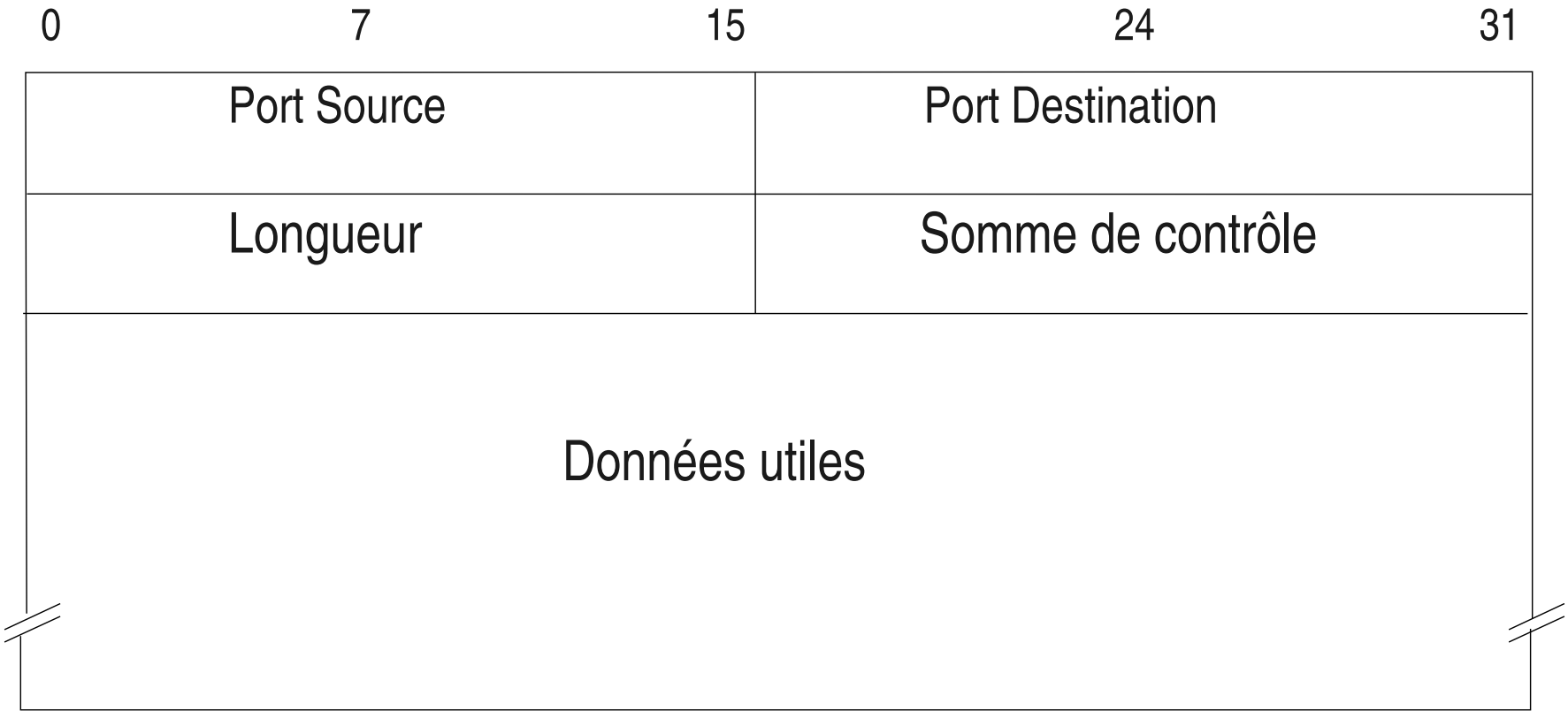
- réception du mesage ICMP émis par un routeur intermédaire (celui qui devrait segmenter)

```
192.168.200.17 > Em: icmp: Rec unreachable - need to frag (mtu 700)
```

- On a bien compris... On émet des paquets de 648 octets (700-12-20-20=648)

```
Em > Rec . 1:649(648) ack 1 win 16060 <nop,nop,timestamp 830140
827243>
```

Le datagramme UDP



UDP

- Pas de connexion
- Pas de contrôle de flux
- Pas d'assurance de la remise
- Mode message : alignement des données reçues sur les données émises

- TCP est orienté «flot d'octets», il ne permet pas d'assurer le «cadrage» des données reçues sur celui des données émises

  Exemple de possibilité :

- UDP est orienté «Message»

- Mis en œuvre par les applications
- Des API existent
  - Sockets : le réseau est vu comme un fichier (on écrit et on lit le réseau comme un fichier), le protocole de communication est à implémenter «à la main»
  - RPC : l'API masque les aspects réseau, on appelle des procédure distantes de la même manière que des procédures locales
  - CORBA : *Common Request Broker Architecture*, extension aux applications réparties du concept «programmation objet»

- Codage des données
  - Protocoles «texte» : smtp, http (1.x), sip, sdp
  - Protocoles codés :
    - ASN.1/BER/PER : H323, snmp
    - XDR : NFS, NIS
- Notions de «session»
  - p.ex. protocole sip, sdp, beep, et même les *cookies* sur le web, etc.

Exemple d'échange HTTP : la requête


```
GET /index.fr.php HTTP/1.1
Host: www.enst-bretagne.fr
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1
Gecko/20030624 Netscape/7.1 (ax)
Accept: text/xml,application/xml,.....
Accept-Language: fr,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 300
Connection: keep-alive
Referer: http://www.enst-bretagne.fr/
```

Exemple d'échange HTTP : la réponse


```
HTTP/1.1 200 OK
Date: Fri, 09 Jul 2004 09:20:06 GMT
Server: Apache/1.3.22 (Unix) PHP/4.0.4pl1 mod_fastcgi/2.2.10 PHP/3.0.....
X-Powered-By: PHP/4.0.4pl1
Keep-Alive: timeout=15, max=100
Connection: Keep-Alive
Transfer-Encoding: chunked
Content-Type: text/html
Content-Language: fr

<HTML>
<HEAD>
<TITLE>[ENST Bretagne] école formation ingénieur</TITLE>
<!-L'école nationale supérieure des télécommunications de Bretagne offre
un large choix de formation: ingénieur, mastères, télécom, thèse, DEA...->
<meta name="robots" content="noindex,nofollow">
<meta name="description" content="L'école nationale supérieure des télécommunications de Bretagne offre un
large choix de formation: ingénieur, mastères, télécom, thèse, DEA...">
<meta name="keywords" content="école ingénieur, formation ingénieur, école nationale supérieure
télécommunications, Bretagne, mastères, télécom, thèse, DEA, ENST, GET, étude télécommunication,
enseignement supérieur, technologie information, communication, TIC, concours, mines-ponts, alternance,
projets européens, recherche, laboratoire, grandes écoles, formation continue, diplôme, Brest, Rennes,
ECOLE INGENIEUR, FORMATION INGENIEUR">
<LINK rel="stylesheet" href="css/style.css">
...
```

Exemple d'échange smtp

(Exemple extrait du rfc-821)

```
S: MAIL FROM:<Smith@Alpha.ARPA>
R: 250 OK

S: RCPT TO:<Jones@Beta.ARPA>
R: 250 OK

S: RCPT TO:<Green@Beta.ARPA>
R: 550 No such user here

S: RCPT TO:<Brown@Beta.ARPA>
R: 250 OK

S: DATA
R: 354 Start mail input; end with <CRLF>.<CRLF>
S: Blah blah blah...
S: ...etc. etc. etc.
S: <CRLF>.<CRLF>
R: 250 OK
```

Fonctionnement du courrier électronique I

- Les utilisateurs ont une adresse de format standard
  - Par ex. :
    - `nom_utilisateur@nom-du-serveur.domaine`
  - ou plus concis :
    - `nom_utilisateur@domaine`
- Les outils mis en œuvre sont
  - Le client (MUA : *Mail User Agent* en langage OSI) : netscape, outlook, lotusNotes, etc...
    - Parle SMTP lors de l'envoi
    - Parle POP ou IMAP (sécurisé ou non) pour recevoir
  - Les serveurs (MTA : *Message Transfert Agent*) : ISS, sendmail, Postfix, etc...
    - Intermédiaires et final
  - Le DNS : requêtes MX (*Mail Exchanger*)

Fonctionnement du courrier électronique II