Local Area Network: Ethernet and Advanced Bridging

http://cours.touta.in/?page_id=516 for discussion and exercises

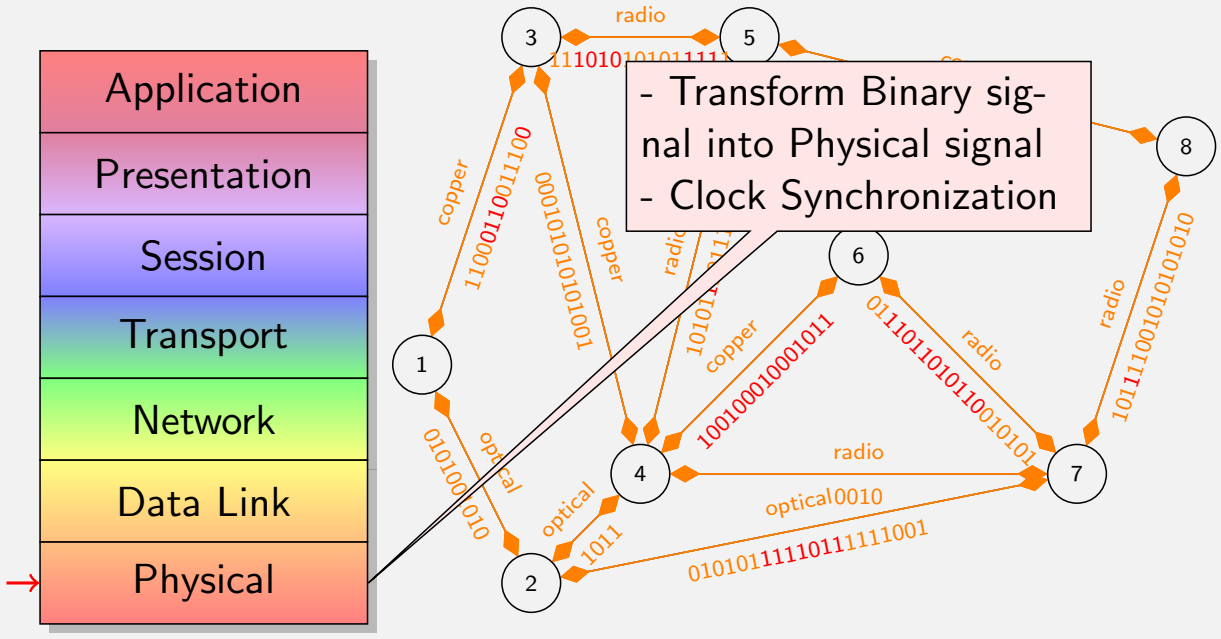
Laurent Toutain

August 26, 2011

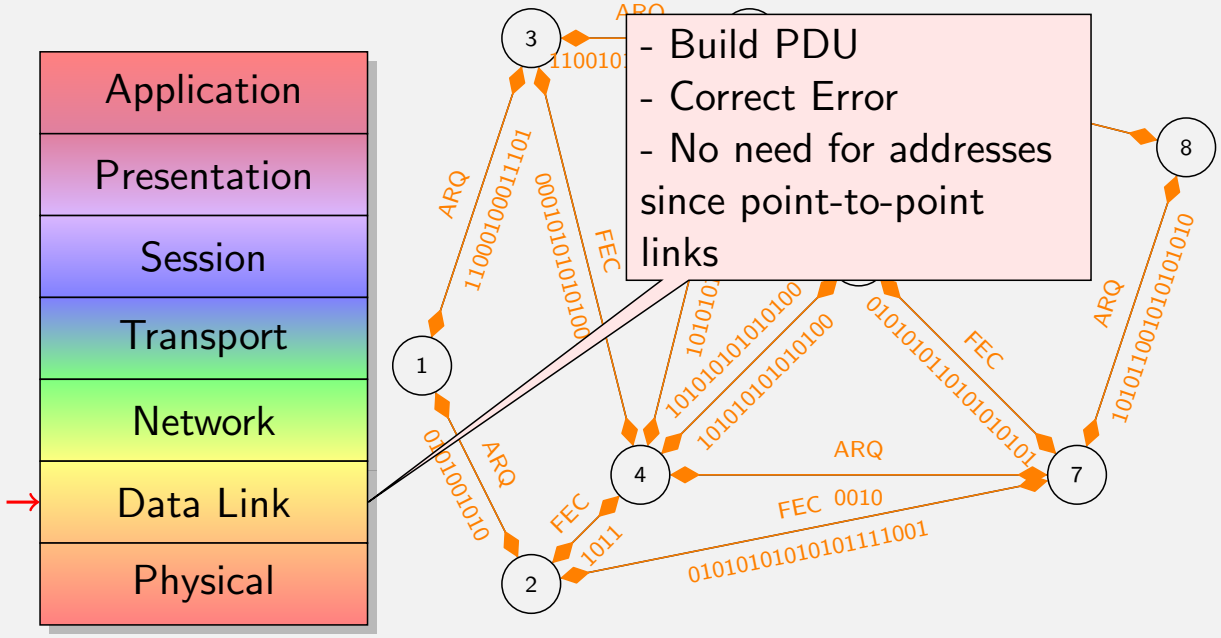
Table of Contents

Introduction

ISO Reference Model

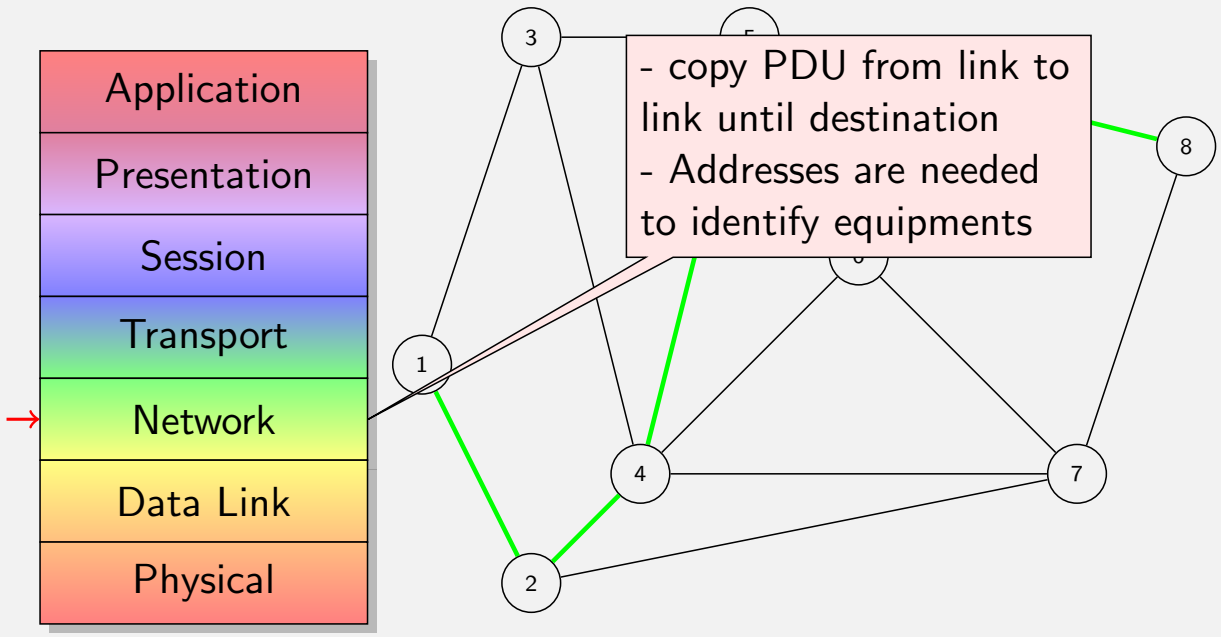


ISO Reference Model



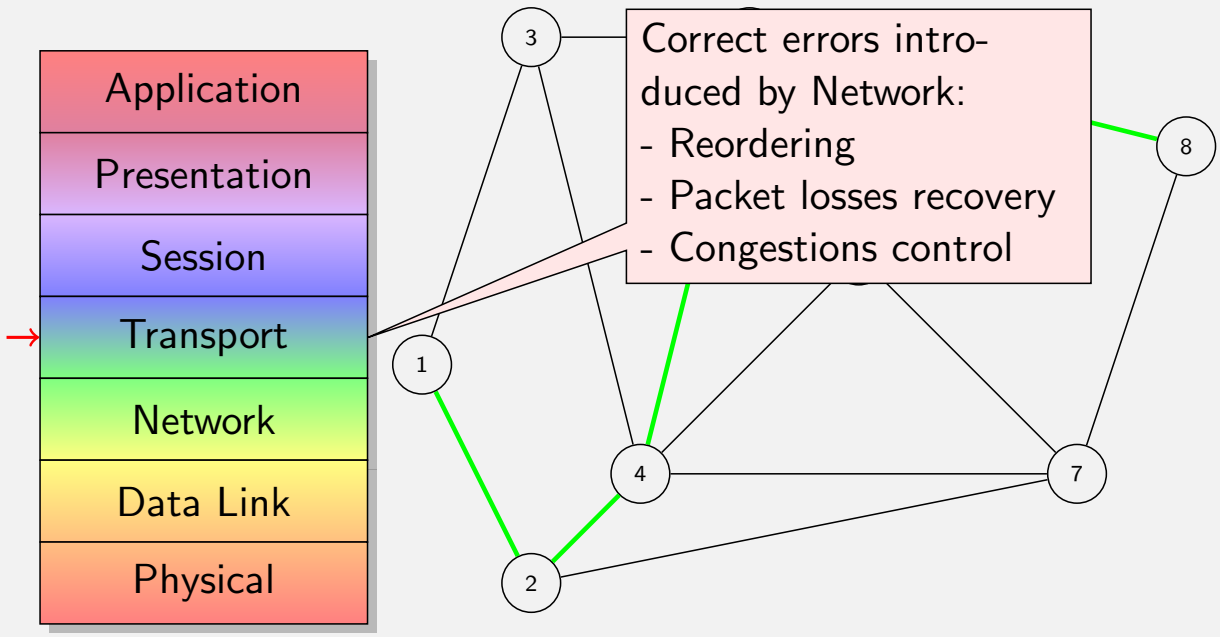
TELECOM Bretagne

ISO Reference Model

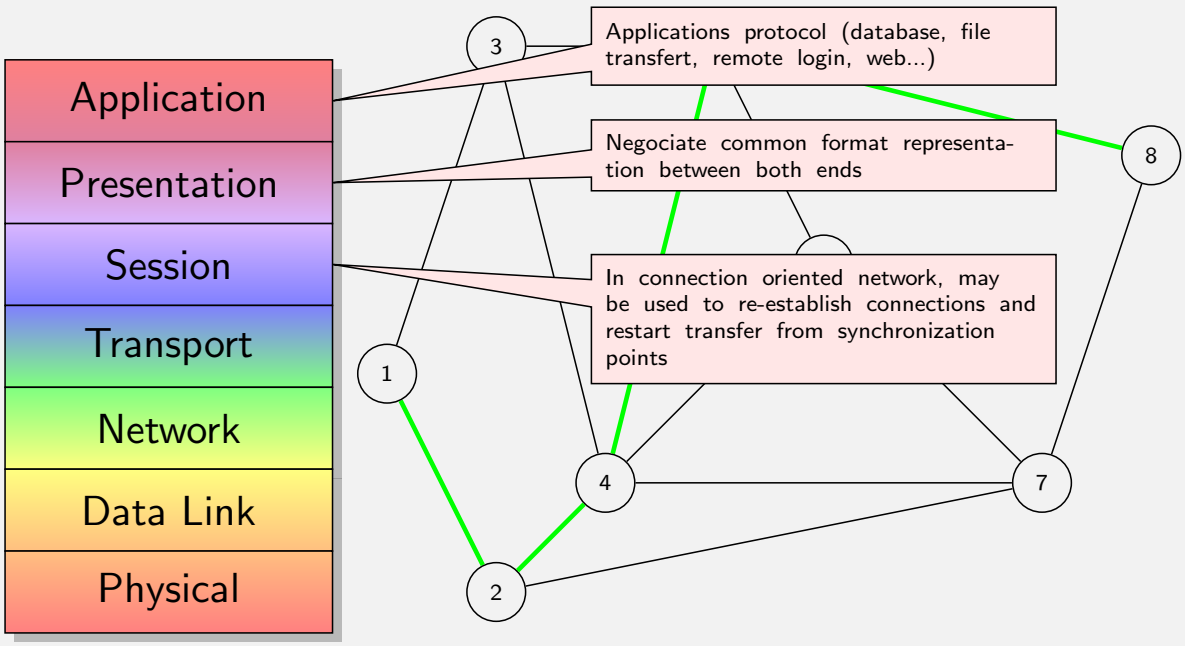


TELECOM Bretagne

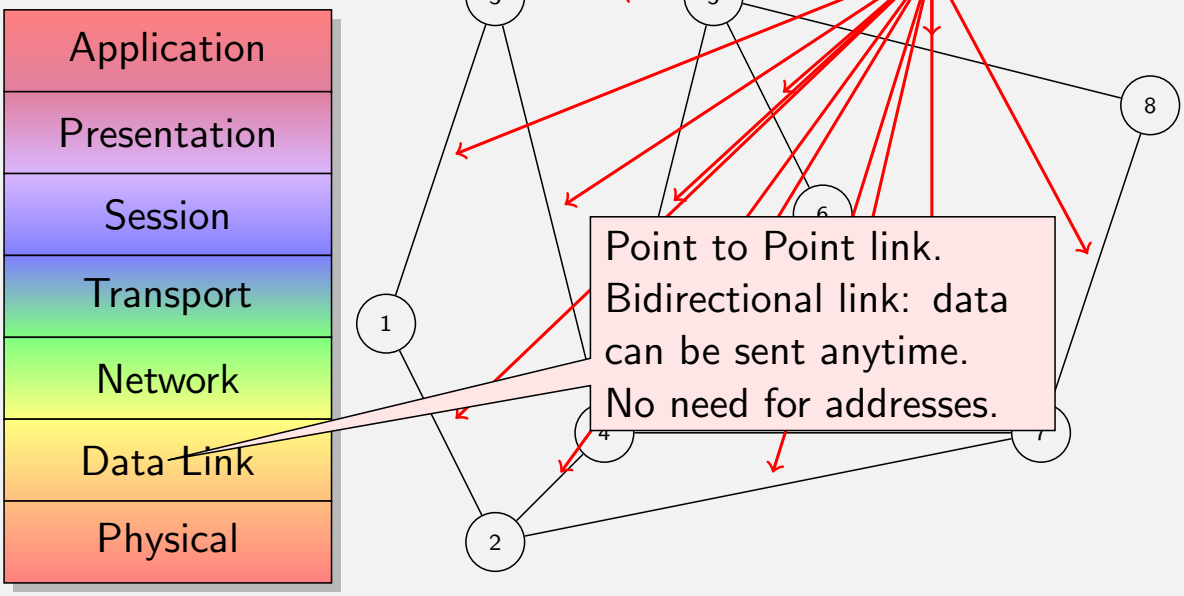
ISO Reference Model



ISO Reference Model

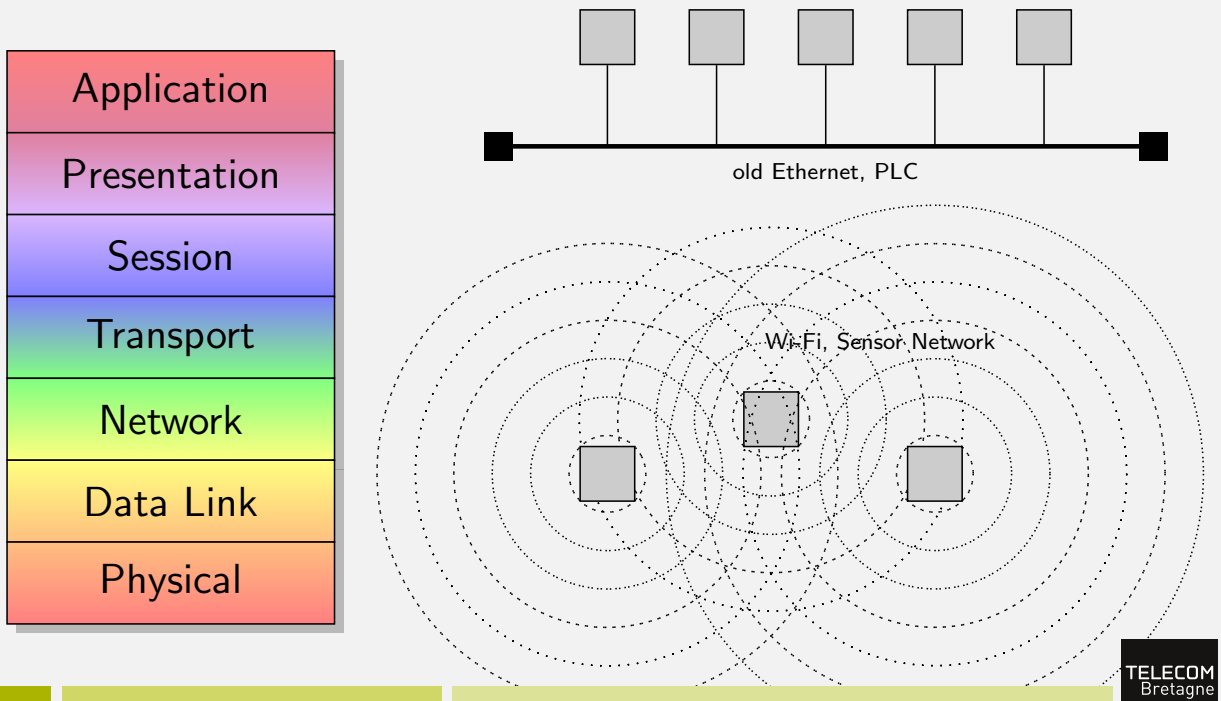


ISO Reference Model: Limits

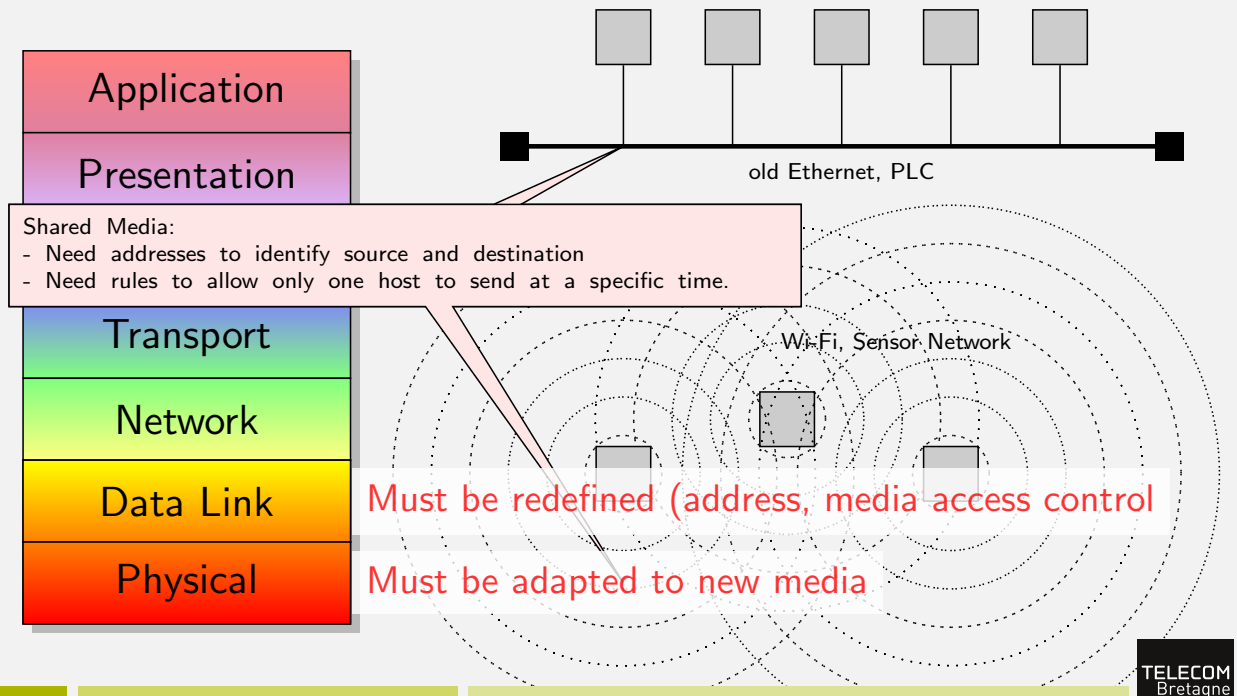


ISO Reference Model: Limits

Introduction ▶



ISO Reference Model: Limits



Comments I

Le comité IEEE 802

Les efforts pour normaliser les réseaux locaux ont commencé en 1979 sous le patronage de l'IEEE (*Institute of Electrical and Electronics Engineers*). Le but de la normalisation était de reprendre les couches 1 et 2 du modèle OSI (*Open System Interconnection*) de l'ISO (*International Standard Organization*) pour les adapter aux particularités des réseaux locaux et métropolitains. En février 1980, le groupe de travail a pris le nom de groupe 802 (80 pour l'année et 2 pour le mois).

Le but du comité 802 de l'IEEE est de développer un standard permettant la transmission de trames d'information entre deux systèmes informatiques de conception courante, à travers un support partagé entre ces systèmes, ceci quelle que soit leur architecture. En effet, le modèle de référence de l'ISO est construit à partir d'une architecture maillée, les équipements sont reliés par des liaisons point-à-point. Dans les réseaux locaux, la manière de connecter les équipements est différente. Ces réseaux sont construits sur un support de transmission partagé par tous les équipements. Les principaux concepts devant être ajoutés au modèle de référence de l'ISO sont :

- des adresses pour pouvoir différencier chaque équipement au niveau 2 ;
- une méthode d'accès garantissant qu'un seul équipement à la fois émettra des données à un instant donné sur le support partagé.

Les protocoles développés par l'IEEE couvrent les couches 1 et 2 du modèle OSI. Au niveau 1, les standards de l'IEEE vont définir les moyens de coder l'information sur le support physique. Au niveau 2, les protocoles de l'IEEE vont définir les méthodes d'accès pour garantir qu'un seul équipement puisse émettre simultanément, ainsi que d'autres fonctionnalités l'authentification des équipements, l'interconnexion des différentes technologies,. . . .

Main characteristics

- Transmission support (media) shared by several equipments:
  - Operating mode: broadcast + Media Access Control
  - When an equipment send data, all others may receive it
  - A equipment should not send data when another one is sending.
  - ⇒ Possible conflict
  - Avoid centralized transmission right management
- A shared media imposes:
  - A way to identity unambiguously any equipments,
  - Decentralized transmission right management
  - No configuration from the user
  - Scalability limitations due to broadcast media
- IEEE defines a model for those kind of networks (layer 1 and 2 of OSI Reference Model)

Comments I

Le comité IEEE 802

Les efforts pour normaliser les réseaux locaux ont commencé en 1979 sous le patronage de l'IEEE (*Institute of Electrical and Electronics Engineers*). Le but de la normalisation était de reprendre les couches 1 et 2 du modèle OSI (*Open System Interconnection*) de l'ISO (*International Standard Organization*) pour les adapter aux particularités des réseaux locaux et métropolitains. En février 1980, le groupe de travail a pris le nom de groupe 802 (80 pour l'année et 2 pour le mois).

Le but du comité 802 de l'IEEE est de développer un standard permettant la transmission de trames d?information entre deux systèmes informatiques de conception courante, à travers un support partagé entre ces systèmes, ceci quelle que soit leur architecture. En effet, le modèle de référence de l'ISO est construit à partir d?une architecture maillée, les équipements sont reliés par des liaisons point-à-point. Dans les réseaux locaux, la manière de connecter les équipements est différente. Ces réseaux sont construits sur un support de transmission partagé par tous les équipements. Les principaux concepts devant être ajoutés au modèle de référence de l'ISO sont :

- des adresses pour pouvoir différencier chaque équipement au niveau 2 ;
- une méthode d'accès garantissant qu'un seul équipement à la fois émettra des données à un instant donné sur le support partagé.

Les protocoles développés par l'IEEE couvrent les couches 1 et 2 du modèle OSI. Au niveau 1, les standards de l'IEEE vont définir les moyens de coder l'information sur le support physique. Au niveau 2, les protocoles de l'IEEE vont définir les méthodes d'accès pour garantir qu'un seul équipement puisse émettre simultanément, ainsi que d'autres fonctionnalités l'authentification des équipements, l'interconnexion des différentes technologies,. . . .

Network Classification

By scope

By scope

	LAN	Access	Metropolitan	WAN
Scope	*hist: 1m to 2 km*		*hist: 100 Mbit/s*	No limit
	1 m to 100 m	up to 10 km		
# equipments	*hist: 2 to 200*		200 km	No limit
	2 to 50	2 to 100	2 to 100	
Owner	User	provider : direct facturation	provider (indirect facturation)	Provider (indirect facturation)
Cost	No charge	f(throughput)	Subscription (SLA: Service Level Agreement)	Distance, Duration, Subscription (SLA)
Throughput	*hist: 4 to 10 Mbit/s*			50 bit/s to 2 Mbit/s
	100 Mbit/s to 1 Gbit/s	2 to 20 Mbit/s	10Gbit/s	Multiple of 2.5 Gbit/s
Error rate	$< 10^{-9}$	$< 10^{-6}$	$< 10^{-9}$	10^{-9}
Delay	1 to 100μ	10 to 15 ms		Mostly propagation delays
Technologies	Ethernet - Wi-Fi	ADSL (ATM Ethernet) WiMAX (Ethernet) 3G	Ethernet	SDH/WDM, Ethernet

Comments I

Ce premier critère peut être la superficie couverte. Les technologies peuvent être divisées en plusieurs catégories, dont les frontières sont plus ou moins floues et qui peuvent évoluer avec le temps avec les progrès technologiques. La terminologie anglaise les désigne sont le nom de WAN (*Wide Area Network*), MAN (*Metropolitan Area Network*) et LAN (*Local Area Network*). En français, on les désigne respectivement sous les noms de réseaux publics, réseaux fédérateurs et réseaux locaux. Le tableau précédent indique les caractéristiques de ces différents types de réseaux.

Réseaux locaux

Un réseau local (en anglais LAN : *Local Area Network*) se caractérise surtout par des performances réduites : la distance relativement courte et la résistance au facteur d'échelle, c'est-à-dire qu'avec l'augmentation du nombre d?équipements connecté les performances se dégradent, est beaucoup plus petite.
Un réseau local dessert donc généralement un bureau, un étage ou un bâtiment dans une entreprise.
L'administration du réseau et des machines est généralement assurée par le même service. Le coût d'utilisation d'un réseau local se concentre surtout dans l'achat du matériel et de la mise en place des câbles.
Les réseaux Ethernet et Wi-Fi sont les technologies de réseaux locaux les plus répandues. Historiquement la portée d?un réseau Ethernet était théoriquement de 2,5 km, mais avec les progrès fait en électronique et en particulier la baisse des coûts des équipements d'interconnexion et leur fiabilité accrue, la taille des réseaux a fortement diminuée. Les règles de câblage actuelle impose qu'un réseau filaire ne dépasse pas quelques centaines de mètres. Pour les réseaux sans fils la portée est d'une dizaine de mètres.
Parallèlement à la diminution de la taille du réseau, le nombre d'utilisateurs directement connectés a également chuté. Historiquement un réseau pouvait connecter entre deux (utilisation du réseau pour interconnecter deux équipements) à quelques centaines. Actuellement, une cinquantaine d'utilisateurs est un nombre acceptable. Sur les réseaux filaires (comme Ethernet) avec les techniques de commutation qui s'opposent au partage du média par plusieurs équipements, on revient à des communication point-à-point entre deux équipements sur le réseau : la station et le commutateur.

Comments II

Le débit est généralement de 100 Mbit/s pour les réseaux filaires et varie entre quelques dizaines de Kbit/s à une centaine pour les réseaux sans fils. A part dans des cas particuliers, l'augmentation du débit pour cette catégorie de réseau n'est plus nécessaire, car le débit de 100 Mbit/s dans le cas des réseaux filaire, est de moins en moins partagée entre les utilisateurs, est dédié à chaque équipement. Néanmoins les technologie à 1 Gbit/s se répandent de plus en plus. Elle peuvent être bénéfiques dans le cas de transfert de très gros fichiers, comme par exemple pendant la sauvegarde d'un disque dur.
Si plusieurs technologies de réseaux locaux ont existé, le marché en a retenue principalement deux: Ethernet pour les réseaux utilisant un support filaire et Wi-Fi pour les réseaux sans fil.

Réseau d'accès

Les technologies comme l'ADSL, le WiMAX où les accès données offert par la téléphonie de troisième génération (3G) peuvent entrer dans cette catégorie car ils interconnectent les réseaux locaux aux réseaux publics.
Pour les réseaux d'accès, les débits sont généralement moins élevés que pour les réseaux locaux et les réseaux public. Ils forment généralement un goulot d'étranglement (parfois volontaire quand il s'agit de facturer en fonction du débit ou technologique).
Les réseaux d'accès filaires comme l'ADSL sont construits autour d'une topologie point-à-point, mais par exemple, les réseaux d'accès en fibre optique pourraient utiliser un mode d?accès partagé. Pour les réseaux hertzien, la question ne se pose pas, vue la nature diffusive du support.
Au dessus de la technologie support de transmission, une couche de convergence peut être mise en place pour offrir un tramage rendant cette technologie compatible avec Ethernet.

Les réseaux métropolitains

La frontière entre un réseau local et un réseau métropolitain (MAN : *Metropolitan Area Network*) peut être très floue. Les principes de fonctionnement sont parfois à peu près les mêmes. Les réseaux métropolitains ou MAN

Comments III

permettent d'interconnecter un certain nombre de sites entre-eux ou de les raccorder à un réseau public. Ils sont souvent appelés dans la littérature anglaise *backbone*, c'est-à-dire l'épine dorsale.
S'il s'agit d'interconnecter des réseaux locaux entre eux, l'administration d'un réseau métropolitain peut être confiée à une structure commune qui regroupe tous les utilisateurs, voire à l'entreprise elle-même si elle est la seule utilisatrice du réseau. La facturation ne se fait pas à l'octet transmis mais est forfaitaire. Elle recouvre les frais de fonctionnement, de maintenance et d?administration du réseau. S'il s'agit de raccorder à des réseaux public, le coût peut être fonction du débit de raccordement.
Comme il s'agit d'un réseau en fibre optique, et construit dans un site protégé, le taux d'erreur est très faible, les délais de transmission réduits et le routage (principalement du pontage) est assez simple. Ces réseaux couvrent une distance de 200 km et offrant un débit allant jusqu'à une dizaine de Gigabit/s.

Les réseaux publics

Ces réseaux (WAN : Wide Area Network) sont en général des réseaux maillés constitués de liaisons point-à-point à haut débit entre des nœuds d'interconnexion. Historiquement, le débit était relativement faible, il pouvait descendre à 50 bit/s pour le réseau Télex et atteindre 2 Mbit/s pour les utilisateurs de technologies comme X.25. Une des plus grandes révolutions dans les réseaux a concerné la forte montée en débit de ce type de réseau. Pendant des années, il a représenté un goulot d'étranglement pour les communications. Or, actuellement, avec les progrès faits dans les technologies de transmission, il est possible d?atteindre des débits de plusieurs Gigabit/s voire quelques Terabit/s.
Si ces réseaux ne se présentent plus dans beaucoup de cas un goulot d'étranglement, dans la partie transmission, il est toujours difficile d'effectuer la commutation, c'est-à-dire le traitement dans le réseau pour aiguiller les données sur un lien ou sur un autre. La nature des lignes et leur longueur faisaient que le taux d'erreur était relativement élevé. Des codes correcteurs ou détecteurs d'erreurs ont dû être utilisés, ce qui peut réduire encore le débit. Les erreurs provoquées par des parasites sur la ligne de transmission sont de plus en plus rares, le plus souvent elles sont dues à des saturations des équipements intermédiaires qui perdent des informations.

Comments IV

Le délai de transmission est assez important. En plus du délai de propagation (par exemple dû à l'utilisation d'un satellite dans certains réseaux), le message est recopié de nœud en nœud jusqu'à atteindre sa destination.
Enfin, le routage, c'est-à-dire le chemin que doit suivre l'information pour atteindre son destinataire, est très complexe, il est la conte-partie de la contrainte de résistance au facteur d'échelle. Dans les technologies LAN ou Metro, l'adressage est généralement à plat et les équipement sont découvert grâce aux propriétés de diffusion du réseau. Dans le cas des réseaux WAN, l'adresse (ou une partie de l'adresse) est utilisée pour localiser le destinataire. L'adressage est le plus souvent hiérarchique (comme l'adressage postal qui indique, le pays, la ville, la rue, le numéro dans la rue et finalement l'identité du destinataire). Il faut également que les équipements d'interconnexion possèdent des tables de routage pour aiguiller les données en fonction de leur adresse de destination. L'utilisation de l'adressage hiérarchique contribue à une taille relativement limité.
Le routage permet de trouver le meilleur chemin qui, du point de vue de l'utilisateur, optimise le débit et minimise le délai de transmission et du point de vue de l'opérateur optimise la charge sur l'ensemble des liaisons du réseau. Pour ce faire, chaque nœud devrait avoir, à chaque instant, une vision complète du réseau. Ceci conduirait à la situation paradoxale où toute la capacité du réseau serait utilisée pour transmettre l'état du réseau aux différents autres nœuds, sans laisser de place pour le trafic utile. Des heuristiques plus ou moins compliquées doivent être employées pour essayer de se rapprocher du routage optimal.

Longtemps constitué de liaison SDH, on retrouve de plus en plus Ethernet.

Questions

Where was historically located the bottleneck ? and currently ?

- ◯ LAN
- ◯ Access
- ◯ Metro
- ◯ WAN

Will Future WAN technologies will reduce delays ?

- ◯ Yes
- ◯ No

Categories of network

This listing gives the name of interconnection equipments and the time (in ms) to reach a web site in Korea. From this information, associate a possible categories of network to each equipment:

```
1  mgs-c5.ipv6.rennes.enst-bretagne.fr  0  ms
2  << unknown >>
3  te4-1-caen-rtr-021.noc.renater.fr  7  ms
4  te4-1-rouen-rtr-021.noc.renater.fr  7  ms
5  te0-0-0-1-paris1-rtr-001.noc.renater.fr  7  ms
6  renater.rt1.par.fr.geant2.net  7  ms
7  so-3-0-0.rt1.lon.uk.geant2.net  14  ms
8  so-2-0-0.rt1.ams.nl.geant2.net  22  ms
9  xe-2-3-0.102.rtr.newy32aoa.net.internet2.edu  106  ms
10  ge-0-0-0.0.rtr.chic.net.internet2.edu  133  ms
11  kreonet2-abilene.kreonet.net  188  ms
12  134.75.108.209  302  ms
13  supersiren.kreonet.net  302  ms
14  kaist-gw.kaist.ac.kr  295  ms
15  143.248.119.85  295  ms
16  143.248.116.253  296  ms
17  143.248.116.218  296  ms
18  iccweb.kaist.ac.kr  296  ms
```

Network Classification

By Access method

By Access method

- Broadcast: Ethernet, Wi-Fi, PLC, Sensor Networks
- Non Broadcast Multiple Access:
  - Any equipment can be joined, but only once at a specific time
  - Telephony Network, ATM.
- Point-to-Point: Leased Line, Virtual Circuit once established
- Master/Slave: communication only possible between the master and a slave (no direct communication between slaves)
  - ISDN, GSM (communication between the mobile phone and the relay)

Questions

Which technologies require addresses (and how many) ?

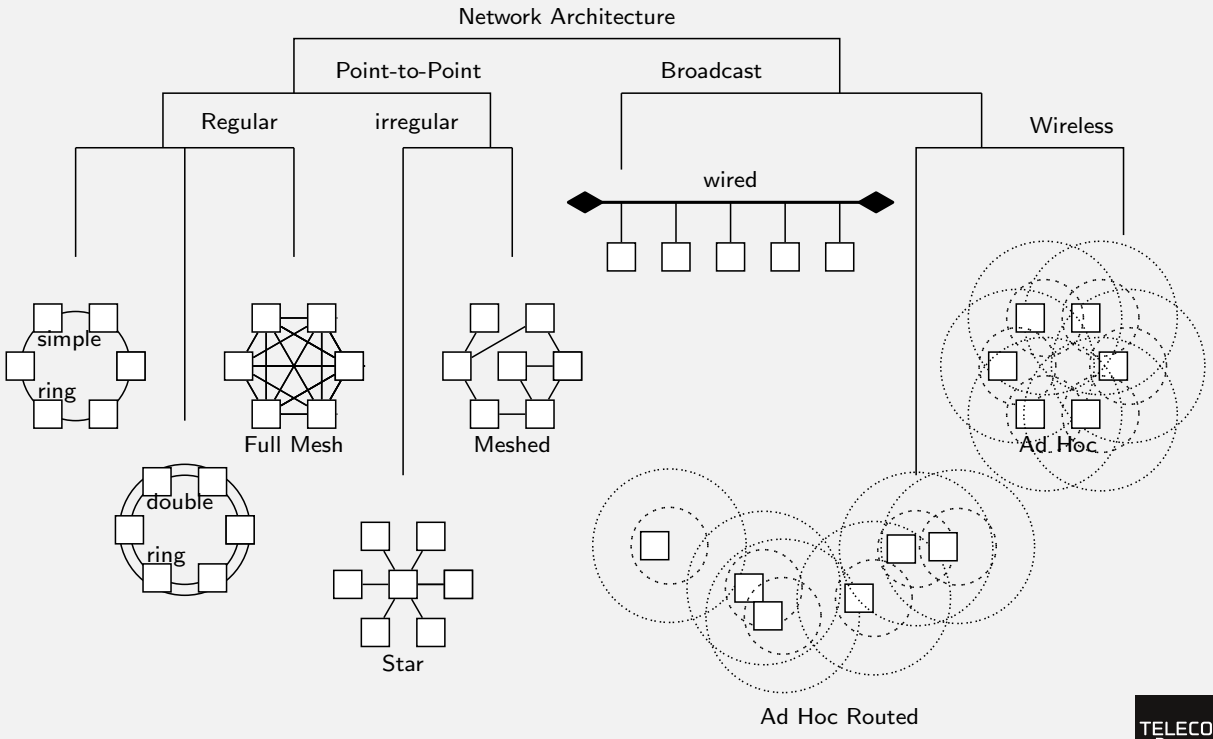
- ◯ Broadcast
- ◯ NBMA
- ◯ point-to-Point
- ◯ Master/Slave

Which of these technologies are scalable ?

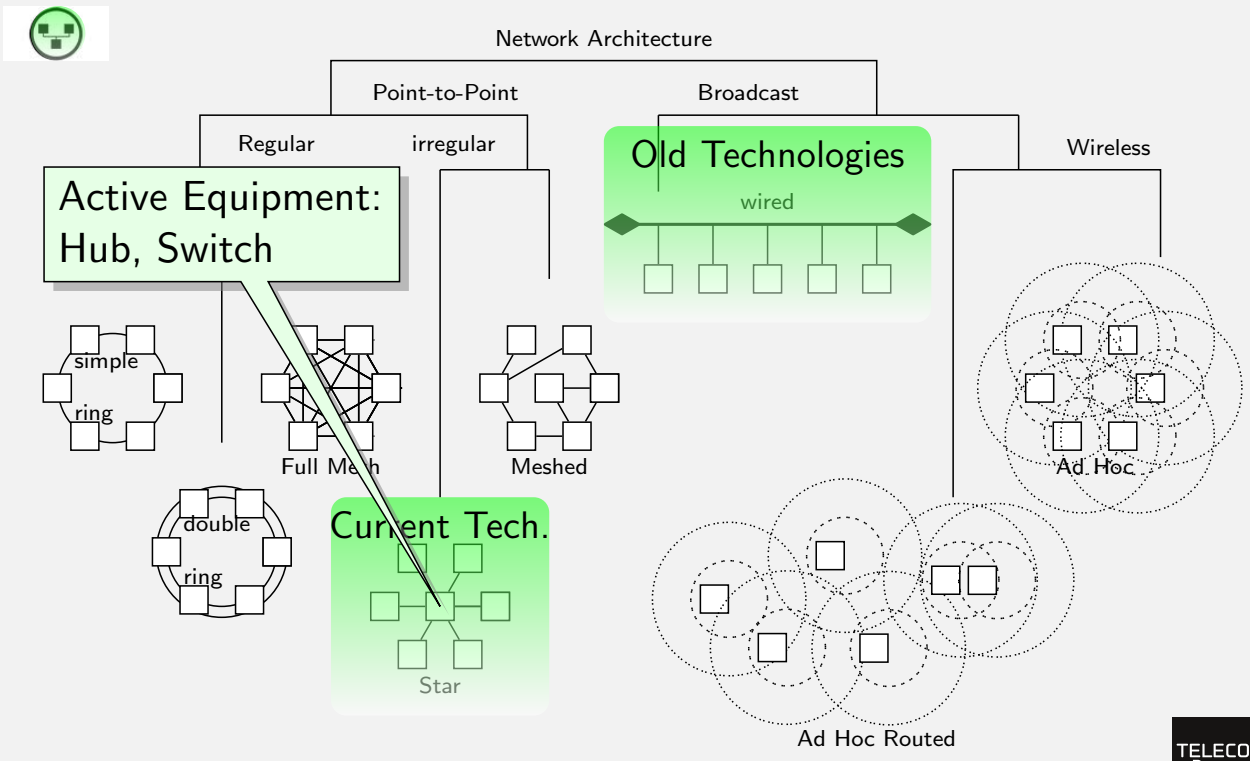
- ◯ Broadcast
- ◯ NBMA
- ◯ point-to-Point
- ◯ Master/Slave

Network Classification

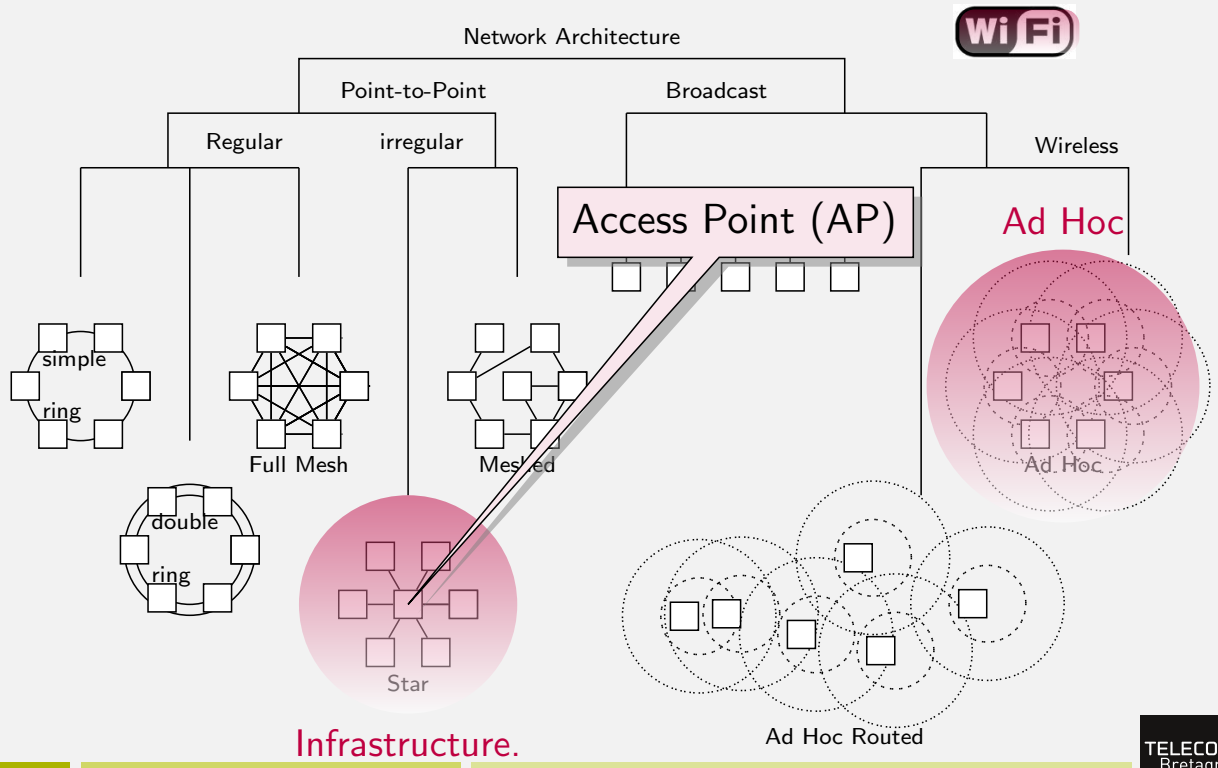
By topology



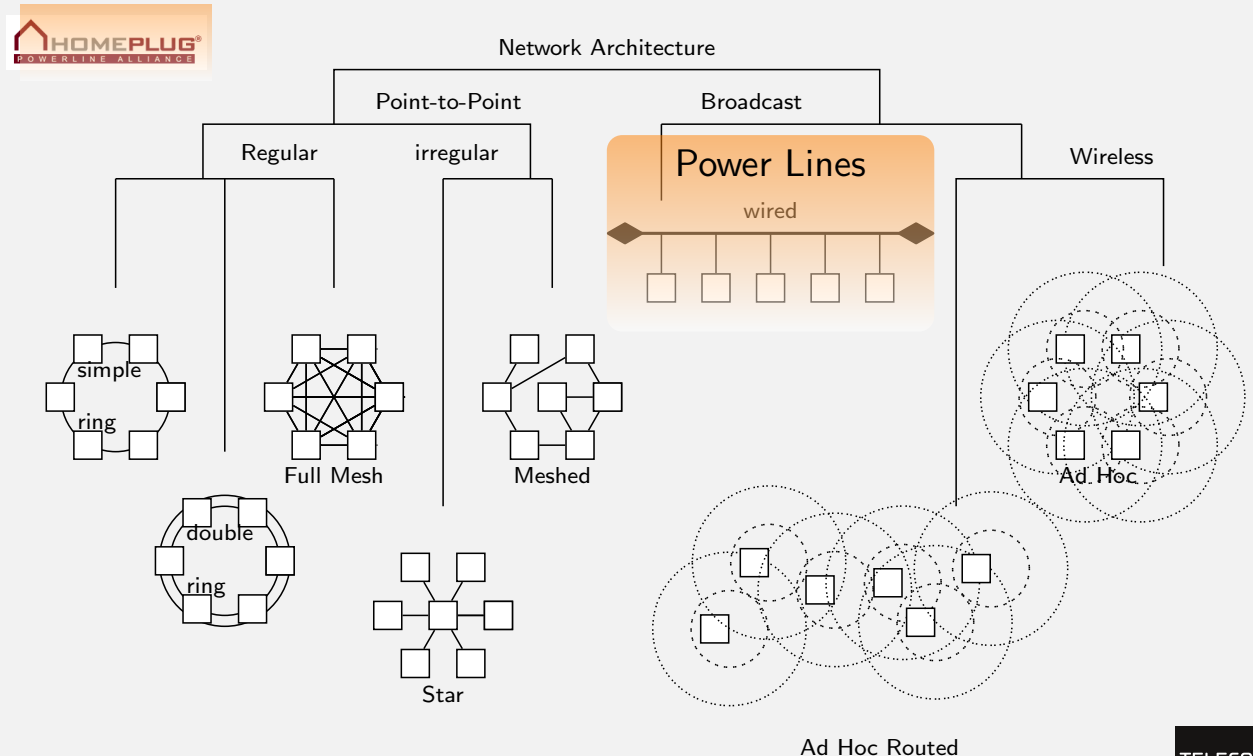
Network Architecture
Network Classification ▶ By topology

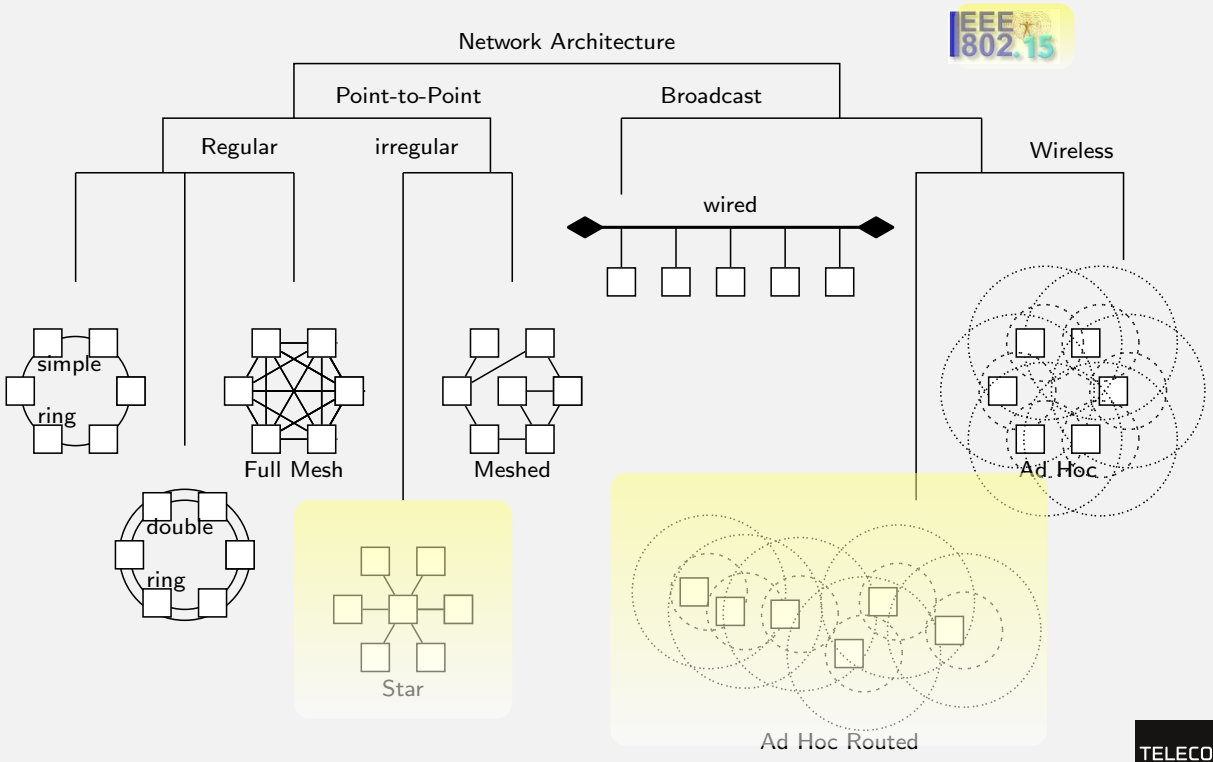


Network Architecture



Network Architecture





Comments I

Différentes topologies, c'est-à-dire formes de réseau, peuvent aussi servir à classer les types de réseau. Chaque topologie possède ses forces et ses faiblesses. A chaque topologie correspondent des méthodes d'accès au support physique différentes. Le transparent précédent essaie de représenter les possibilités réellement exploitées dans les architecture de réseau.

- Les liaison point-à-point sont les liaisons les plus faciles à exploiter car elles ne nécessitent pas d'adressage ni pour identifier l'émetteur (le message provient toujours de l'autre extrémité) ni le récepteur, il se propage jusqu'à l'autre extrémité. En général, ces liaisons sont bi-directionnelles, elles ne nécessitent pas de mécanismes de gestion d'accès; dès qu'un équipement souhaite transmettre un message, il peut le transmettre sur le support dédié. Malheureusement, une liaison point-à-point ne permet de joindre qu'un seul équipement. Pour permettre de plusieurs équipements et donc de former un réseau, plusieurs architectures sont construites autour de liaisons point-à-point :
  - Maillage complet (*full mesh*): Il s'agit de mettre en place des liaison point-à-point entre tous les équipements qui souhaitent communiquer. Cette solution n'est généralement pas économiquement viable, car on se retrouve avec des liaisons sous exploitées, même dans le cas de réseaux locaux.
  - Etoile (*Star*): Il s'agit de faire converger les liaisons point-à-point vers un équipement central qui se charge de retransmettre les informations vers le (ou les) destinataires. Cette architecture est très prisée car elle se rapproche des câblages utilisés dans les bâtiments pré-câblés. Les liaisons filaires partant des bureaux convergent vers une armoire de brassage. Ce type d'architecture requiert généralement un adressage pour permettre d'identifier l'émetteur et le destinataire du paquet. Les mises en ?uvre actuelle d'Ethernet sont basées sur ce type d'architecture. L'équipement central est appelé un hub ou un commutateur (switch en anglais) suivant la technologie utilisée. Cette architecture se retrouve également au niveau 2 des réseaux Wi-Fi en mode infrastructure. Tous les équipements dialoguent avec un point d'accès (*access point*) qui redistribue l'information vers le réel destinataire.

Comments II

- Anneau (*Ring*) : Un anneau est composé de liaisons point-à-point entre toutes les stations qui composent le réseau de manière à former une boucle. Cette topologie a été un temps populaire avec l'anneau à jeton, mais elle demande une gestion complexe des droits à la parole. Elle est maintenant abandonnée dans les architectures des réseaux locaux. Par contre du fait qu'il existe deux chemins possible pour aller d'un point à un autre, cette architecture est plus robuste à la coupure d'une liaison. Les réseaux SDH l'utilisent pour cette raison.

- multipoint : Une liaison multipoint permet de joindre plusieurs équipements à la fois. C'est par exemple le cas avec les anciennes versions d'Ethernet où un câble co-axial reliait tous les équipements. Un support Hertzien, comme pour le réseau Wi-Fi est naturellement à diffusion puisque tous les équipements partagent la même fréquence pour communiquer. On peut également le retrouver quand des données sont transmises sur les fils électriques (CPL : Courant Porteur en Ligne/ PLC : Power Line Communication). Pour le Wi-Fi, il existe plusieurs modes de fonctionnement :
- Ad-Hoc : Dans ce mode les équipements dialoguent directement entre eux. Contrairement à un réseau filaire où la topologie du réseau ne change pas, dans un réseau Ad-Hoc, le réseau doit se reconfigurer dynamiquement en fonction des arrivées et des départs des équipements souvent lié à leur mobilité ou aux conditions de propagation raidio. Dans les versions de base de la norme, il faut que les équipements soient en portée radio pour que le dialogue soit possible, il n'existe pas de possibilité de relayage par des stations intermédiaire.
- Ad-Hoc Routé : Dans ce mode les équipements peuvent servir de relais pour permettre d'acheminer la trame vers le destinataire. Cela demande la mise en place de protocole de routage pour permettre de trouver ce chemin. Les réseaux de capteurs (par exemple ceux définit dans la norme IEEE 802.15.4 utilisé par ZigBee) du fait de la faible portée des transmission doivent avoir recours au routage ad-hoc.

Comments III

- Infrastructure : Toutes les communications se font avec un point central appelé *Access Point* (AP). L'AP facilite la configuration des équipements en envoyant des paramètres et assure l'interconnexion avec les réseaux externes. Si une communication doit se faire entre deux équipements situés dans le même réseau Wi-Fi, les données seront transmises deux fois (source vers AP puis AP vers destinataire).
  L'adressage est important dans ce type de réseau pour identifier d'où vient le message et à quel équipement il est destiné. L'adresse doit être forcément unique pour identifier l'équipement, par contre elle n'a pas à indiquer son emplacement (a l'instar d'une adresse postale qui hiérarchiquement permet de trouver le destinataire), pour que le réseau à diffusion permet de l'envoyer vers tous les équipements. Le réseau multipoint peut être construit à partir de liaisons point-à-point. Par exemple, dans le cas d'une architecture en étoile, si l'équipement central retransmet les données binaires reçues vers sur un câble vers les autres supports, cela permet de construire un réseau à diffusion. Les hubs utilisés par Ethernet réalisent cette fonction.

IEEE Model

Introduction

IEEE Model

- Effort started in February 1980 (802) to specify LAN protocols
  - Currently covers also all others technologies (Access, Metro and some part of WAN)
- Adapt ISO Reference Model to broadcast media.
- No centralized equipment (i.e. Master/Slave):
  - Any host can leave or enter the network at any time.
  - USB[1] is not a LAN technology, since a master is needed.
- Not a single solution:
  - No universal solution (depending on the media)
    - Wired and wireless media react differently
    - Industrial may push different technologies
    - Competition exists between IEEE 802 solutions
    - Market will decide
- But a current architecture to ease interconnection.

[1] Universal Serial Bus

IEEE groups

- Standardization is a dynamic process:
  - Some technologies disappear due to la lack of interest from the market
  - Some technologies appear:
    - New areas (wireless, Wireless Sensor Network, Personal Area Network,...)
  - All technologies must evolve
    - Increment speed
    - Use new media
    - New usages (energy aware,...)
- IEEE 802 is divided into subgroups.
- Each Working Group is designated by a number[2].
  - For instance IEEE802.3 does standardization for Ethernet

[2] See; http://ieee802.org/dots.shtml

IEEE groups

IEEE Model ▶ Introduction

- Each Working Group issues regularly a new version of the standard:
  - IEEE 802.3-2008, IEEE 802.3-2005, IEEE Std 802.3-2002,
- Subworking group are working on make evolve some specific part of the standard:
  - They are designated by a letter.
  - Capital letter: a new document
  - Small letter: a modification to an existing document.
  - These results are merged in the next issue of the full standard
  - Sometime the letter continues to be used to refer to the technology:
    - IEEE 802.1Q: Virtual LANs
    - IEEE 802.11g: Wi-Fi at 54 Mbit/s
    - ...
- Standards and approved extensions to standards are available on IEEE Explorer:
  - http://ieeexplore.ieee.org/xpl/standards.jsp#

Comments I

Les efforts pour normaliser les réseaux locaux ont commencé en 1979 sous le patronage de l'IEEE (*Institute of Electrical and Electronics Engineers*). Le but de la normalisation était de reprendre les couches 1 et 2 du modèle OSI (*Open System Interconnection*) de l'ISO (*International Standard Organization*) pour les adapter aux particularités des réseaux locaux et métropolitains. En février 1980, le groupe de travail a pris le nom de groupe 802 (80 pour l?année et 2 pour le mois).

Le but du comité 802 de l'IEEE est de développer un standard permettant la transmission de trames d'information entre deux systèmes informatiques de conception courante, à travers un support partagé entre ces systèmes, ceci quelle que soit leur architecture.

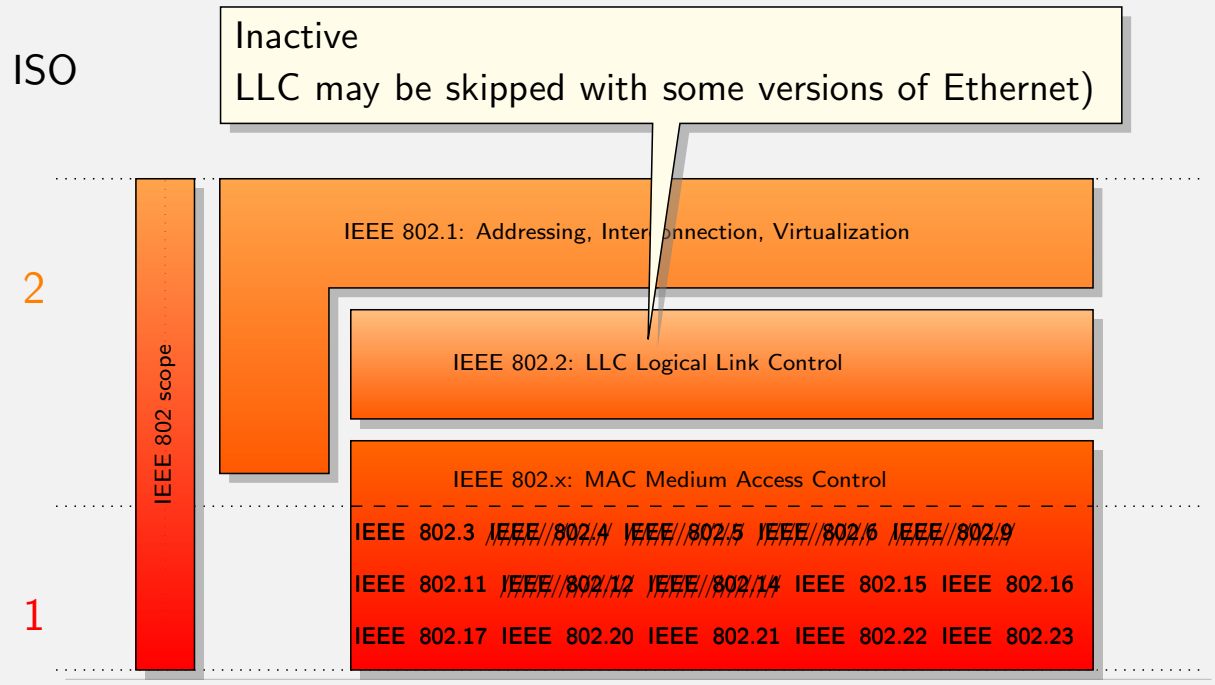
En décembre 1981 trois méthodes pour accéder au support de transmission étaient pressenties. Cette offre multiple a fait dire que le groupe "ne savait pas prendre de décision". Mais en réalité, de même qu'il existe plusieurs moyens de transport pour les personnes et les produits, il existe plusieurs moyens d?accéder au support physique suivant le type d'application. Les trois méthodes étaient le CSMA/CD (*Carrier Sense Multiple Access/Collision Detect*), le bus à jeton et l'anneau à jeton.

Pour qu'un nouveau sujet soit étudié, il faut qu'un PAR (*Project Authorisation Request*) soit voté. La désignation de celui-ci dépend de la nature de l'étude. Si le sujet peut être traité par un sous-comité, il est référencé par une lettre après le nom du sous-comité. Les documents produits seront intégrés dans les prochaines révisions des normes. Une lettre majuscule désigne un document autonome tandis qu'une lettre minuscule désigne un document complémentaire. Il n'y a généralement pas de correspondance entre les lettres majuscules et minuscules. Par exemple, le document IEEE 802.1p est un complément au document IEEE 802.1D décrivant le pontage dans les réseaux locaux. Si les travaux s'éloignent trop des sous-comités existants, un nouveau sous-comité est créé.

IEEE Model

Architecture

IEEE Architecture



Comments I

Le modèle de référence de l'ISO est construit à partir d'une architecture maillée, les équipements sont reliés par des liaisons point-à-point. Dans les réseaux locaux, la manière de connecter les équipements est différente. Ces réseaux sont construits sur un support de transmission partagé par tous les équipements. Les principaux concepts devant être ajoutés au modèle de référence de l'ISO sont :

- des adresses pour pouvoir différencier chaque équipement au niveau 2 ;
- une méthode d'accès garantissant qu'un seul équipement à la fois émettra des données à un instant donné sur le support partagé.

En plus, au niveau 1, les méthodes de codage de l'information sont adaptés à la nature des supports choisis et aux conditions dans lesquels ils sont utilisés.

En 1982, le comité 802 fut réorganisé et divisé en plusieurs groupes. La figure précédente en donne l'architecture générale. Certains groupes sont maintenant en inactifs (comme le groupe IEEE 802.2) ou abandonnés (groupes hachurés sur la figure). les groupes définissant l'architecture sont les suivants :

- le groupe IEEE 802.1 pour l'architecture générale du réseau :
  - le modèle architectural en couche présenté dans cette section ;
  - format des adresses, le comite IEEE 802.1 définit des adresses sur 16, 48 et 64 bits. Dans les réseaux locaux, la taille des adresses est de 48 bits, ce qui facilite l'interconnexion de différentes technologies comme entre Ethernet et Wi-Fi ;
  - techniques d'interconnexion des réseaux par pontage entre technologies identiques ou entre différentes technologies;
  - techniques de virtualisation qui permettent de construire un réseau logique différent de l'architecture physique donnée par le cablage ;

Comments II

- ...

- le groupe IEEE 802.2 pour la sous-couche LLC (*Logical Link Control*) : un protocole avec trois classes appelées LLC type 1, LLC type 2 et LLC type 3 pour gérer le transfert de données. Ces trois classes sont respectivement :
  - un service simple en mode non connecté. La reprise sur erreur, le contrôle de séquencement et de duplication sont laissés à la couche supérieure,
  - un service en mode connecté comparable aux services offerts par le protocole HDLC (High level Data Link Control) : acquittement, contrôle de séquencement,
  - un service non connecté mais avec acquittement permettant des temps de transmission faibles et sécurisés ;

  Le groupe est actuellement inactif car plus aucune standardisation n'est en cours. En fait, seul le premier mode, très simple, est utilisé dans les réseaux locaux. Il est même possible dans certaines configurations très populaires d'Ethernet de ne pas utiliser cette sous-couche.

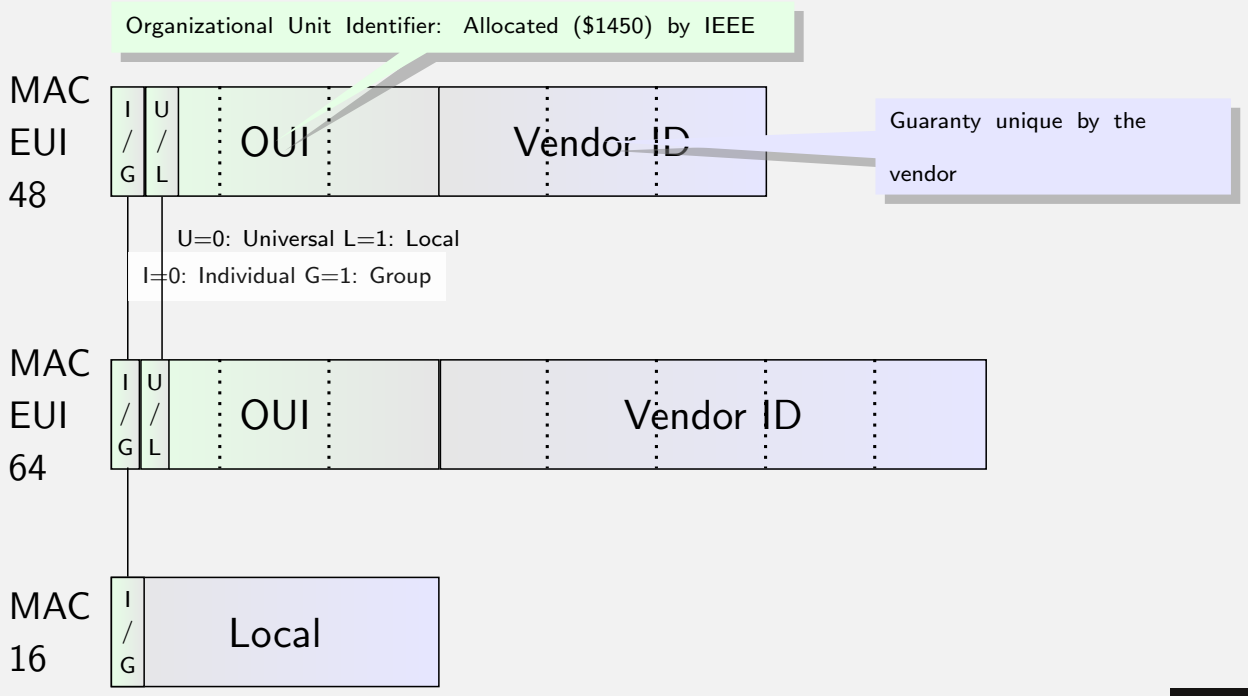
IEEE Model

Addresses

IEEE 802 Addresses

- Addresses need to be unique on the link,
- Addresses must be assigned without any centralized server,
- Addresses are assigned by the manufacturer:
  - Network card or Equipment manufacturer,
  - They are globaly (worldwide) unique
  - Addresses can be changed by the network manager (do not base security on them).
- IEEE defines three length:
  - MAC-16: 16 bit long, used on network where frame size is important (e.g. Wireless Sensor Networks)
  - MAC-48: 48 bit long, mainly used in LAN technologies (Ethernet, Wi-Fi)
    - no exhaustion before 2100.
  - MAC-64: 64 bit long, not actually used.
- IEEE makes the difference between MAC and EUI (Extended Unique Identifier):
  - used to identify a equipment (not used to send frames)
  - EUI-48 or EUI-64.

How to make addresses globally unique ?

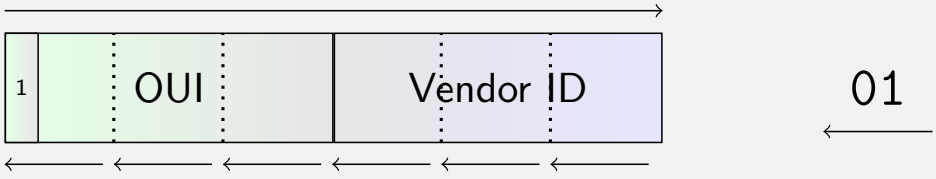


Different Addresses I

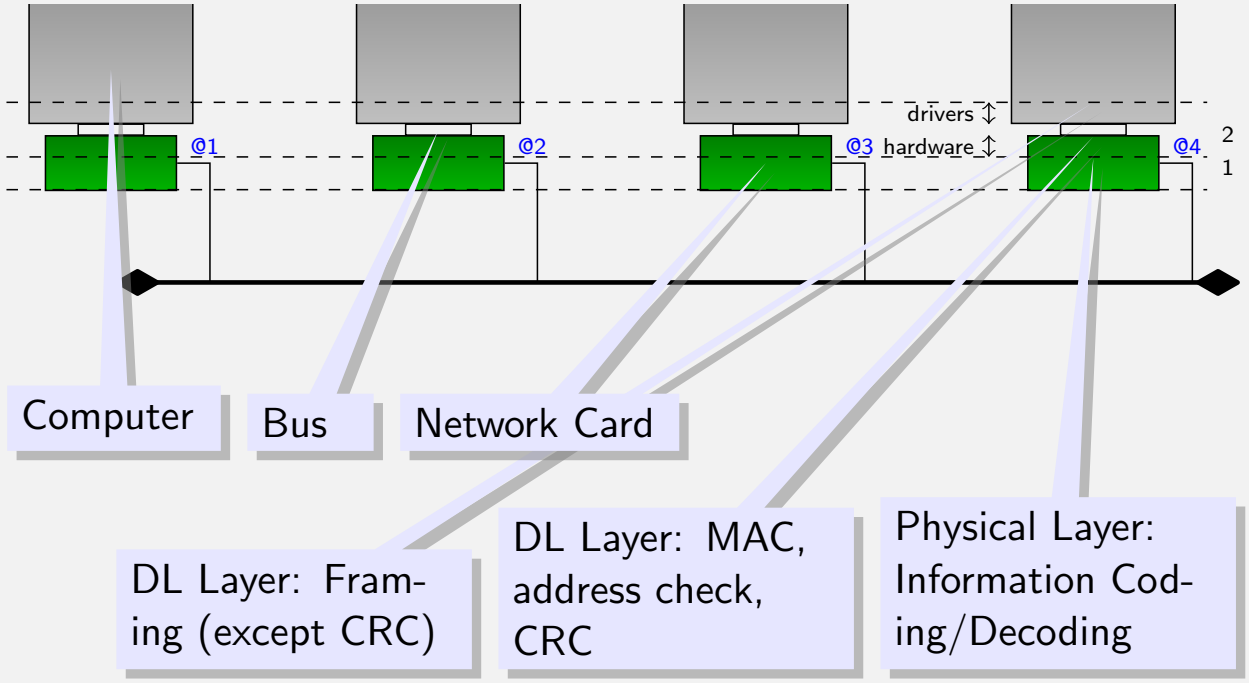
- Representation:
  - Bytes separated by dashes (sometime by column)
  - Ex: `00-23-df-a9-f7-ac`
- Unicast Addresses:
  - I/G bit = 0
  - The frame is sent to only one host on the link
  - OUI is allocated by IEEE; See http://standards.ieee.org/regauth/oui/ to find some values.
  - Who is the vendor of `00-23-df-a9-f7-ac`
- Broadcast Address
  - Sent to all the hosts connected to the link.
  - All bits equal to 1
  - `ff-ff-ff-ff-ff-ff`

Different Addresses II

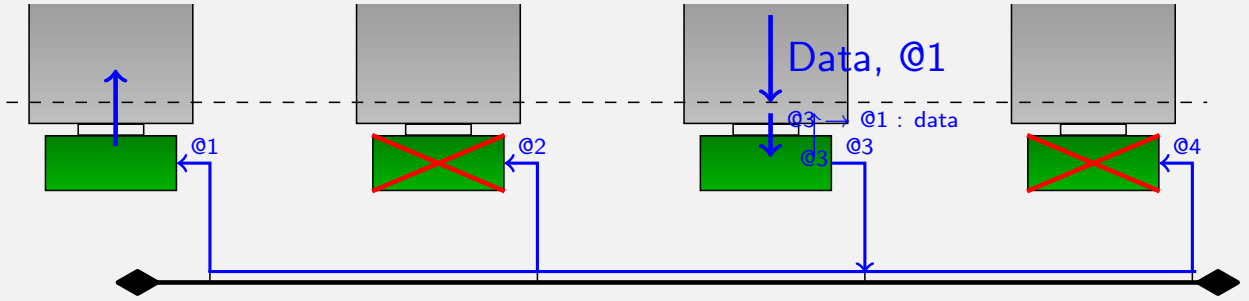
- Multicast Address:
  - bit I/G = 1 and address is different from Broadcast address.
  - Nodes must subscribe to this address to receive data.
  - No need to subscribe to send.
- Warning:
  - First bit send is the I/G bit
  - Inform interconnection elements (bridges) to copy the frame on other media.
  - In numerical representation, this bit is on the right.



Addresses Management: Computer Architecture

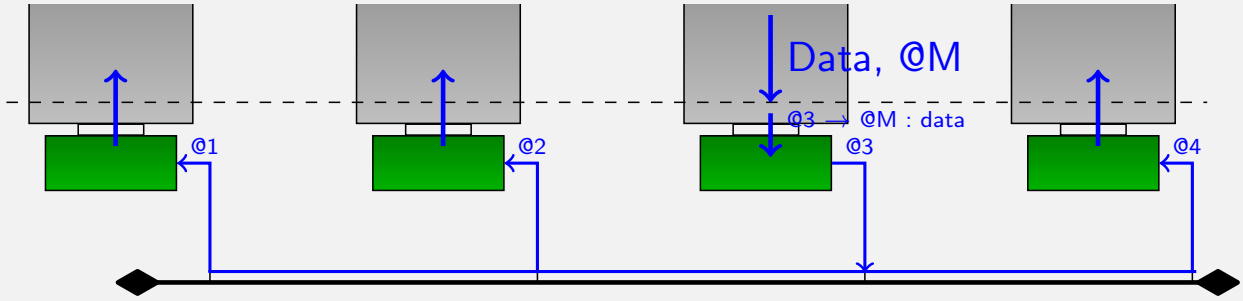


Addresses Management: Unicast



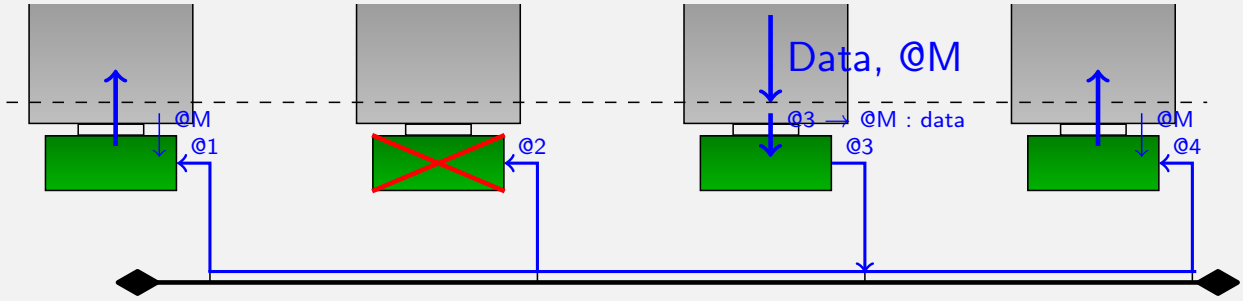
1) MAC Layer receives data.request for @1
2) MAC Layer reads source MAC address (can be overloaded)
3) MAC Layer creates the frame and sends it to the Network card
4) MAC Layer computes the CRC and perform the MAC protocol and sends the frame on the wire
5) Only @1 recognizes its address and sends it to upper layer. Others discard the card and the frame is not seen by equipments.

Addresses Management: Broadcast



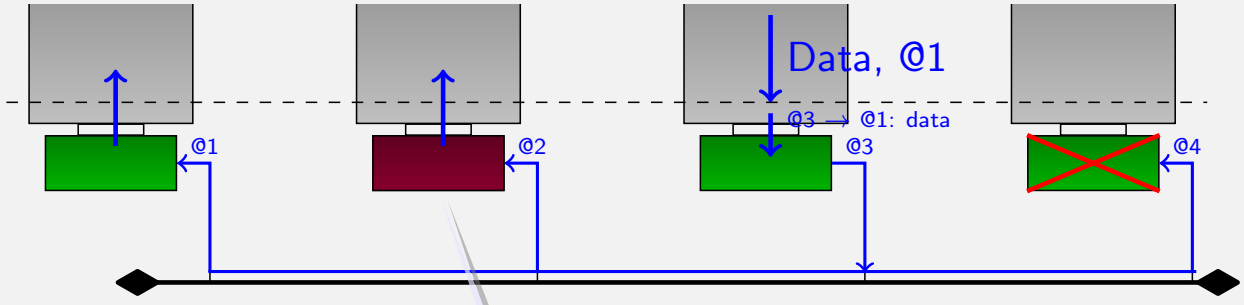
1) MAC Layer receives data.request for FF
2) MAC Layer reads source MAC address (can be overloaded)
3) MAC Layer creates the frame and sends it to the Network card
4) MAC Layer computes the CRC and perform the MAC protocol and sends the frame on the wire
5) All equipments recognize the broadcast address and transmit the information to the upper layer.

Addresses Management: Multicast



0) Equipments register the multicast address
1) MAC Layer receives data.request for @M
2) MAC Layer reads source MAC address (can be overloaded)
3) MAC Layer creates the frame and sends it to the Network card
4) MAC Layer computes the CRC and perform the MAC protocol and sends the frame on the wire
5) Equipment which have registered the multicast address receive the information..

Addresses Management: Promiscuous



1) @2 sets promiscuous mode (some privileges are needed)
2) @3 sends a frame to @1

3) @1 accepts the frame (its address), @2 accepts the frame (promiscuous), @4 discards the frame

Promiscuous is used either to analyze traffic (see wireshark, tcpdump) or by L2 interconnection elements such as bridges

Comments I

Bien que les méthodes d?accès soient différentes, l'adressage des stations est le même. La norme IEEE 802.1 propose trois formes d'adresses :

- une adresse courte sur 16 bits, pour les réseaux locaux ayant des contraintes énergétiques comme les réseaux de capteurs (IEEE 802.15.4),
- une adresse sur 48 bits pour les réseaux comme Ethernet (IEEE 802.3) ou Wi-Fi (IEEE 802.11). Il s'agit de la forme d'adressage la plus utilisée. Elle peut également se retrouver dans des standards proposé par des organismes différents de l'IEEE.
- plus récemment une adresse sur 64 bits permettant un déploiement plus massif. En 1995, l'IEEE a défini la norme IEEE 1394 pour un réseau à haut débit prévu pour les applications informatiques et domestiques grand public comme les téléviseurs, les magnétoscopes, les chaînes hi-fi, etc. Les quantités de matériels vendues seront nettement supérieures au marché du matériel informatique. L'IEEE a défini une nouvelle structure d'adresse où le numéro de série est étendu à 5 octets. Une adresse EUI-64 a une longueur de 8 octets. L'IEEE prévoit de n'attribuer une autre adresse à un constructeur que si ce dernier a déjà utilisé plus de 90 % des valeurs possibles.Elle se retrouve dans les équipements de grande consommation. Elle se retrouve également dans les réseaux de capteurs (IEEE 802.15.4), bien que la longueur ne soit pas favorable aux contraintes énergétiques liée à la transmission des données, elle permet d'assigner de facto une adresse unique à un très grand nombre d'équipements.

Comments II

L'IEEE fait la subtile différence en une adresse appelée MAC et un identifiant appelé EUI (*Extended Unique Identifier*). Les adresses MAC vont se retrouver dans les en-têtes des trames, tandis que les EUI seront mises dans les données. Ainsi dans la norme IEEE 1394, connue aussi sous le nom commercial FireWire, l'adresse correspond à la position de la machine dans l'arbre que constitue le réseau (elle peut donc varier en fonction des modifications de la topologie), l'identifiant EUI-64 est par contre stable.
Une adresse universelle (c'est-à-dire avec le bit U/L à 0) est divisée en deux parties. L'IEEE attribue aux vendeurs (ou aux constructeurs) de cartes, les trois octets de gauche appelés aussi OUI (*Organizational Unit Identifier*). Les OUI sont attribués aux compagnies qui ont font la demande contre 1 450 dollars. Les trois octets (ou cinq) de droite servent à désigner le numéro de série dans la production du vendeur. Par construction chaque adresse est unique. Par contre, il ne faut s'attendre à aucune logique dans la numérotation quand on considère un réseau particulier. Une liste non exhaustive des adresses de vendeurs se retrouve dans le site web
Whttp://standards.ieee.org/regauth/oui/
L'adresse universelle permet de simplifier la gestion des réseaux puisque l'administrateur n'a pas à attribuer de valeurs. A tout moment, un administrateur peut en modifier la valeur, il devrait mettre le bit U/L à 1 pour éviter tout conflit avec les valeurs préalablement attribuées.
Les adresses MAC universelles ou locales servent à désigner de manière unique une station dans le monde. Pour les adressages de groupe, il existe deux méthodes le broadcast ou diffusion généralisée :

- l?adresse de broadcast est unique et reconnue par toutes les stations. Cette adresse est égale à `FF-FF-FF-FF-FF-FF` (tous les bits sont à 1). Toutes les stations connectées sur le réseau local lisent les trames portant cette adresse. Le filtrage, pour savoir si la trame était effectivement destinée à la station, est effectué par les couches d?un niveau supérieur ;

Comments III

- le multicast ou diffusion restreinte : l'inconvénient majeur du broadcast provient du filtrage des messages par les couches hautes. Pour chaque message en diffusion généralisée, la couche MAC réveille les couches hautes. Le filtrage est effectué par le système d'exploitation et consomme des ressources sur la machine (CPU, mémoire,...). Cela se traduit par une perte de performance de l'ensemble des stations du réseau. Pour le multicast, les stations qui veulent accéder à un service (ou groupe) doivent explicitement s'abonner. Elles donnent au composant MAC l'adresse du groupe. Quand le composant reconnaît un paquet portant une adresse de groupe préalablement enregistrée, il transmet ce paquet aux couches supérieures. Les stations qui n'ont pas enregistrée une adresse de multicast particulière filtrent ces trames comme des trames ne leur étant pas destinées. Le filtrage se fait par le contrôleur de communication de niveau MAC et ne pénalise pas les performances des stations.

Une trame de diffusion commence par un bit à I/G à 1. Il s'agit du premier bit transmis sur le réseau, il permet en particulier aux équipements d'interconnexion de détecter que la trame devra être recopiée sur d'autres supports.
La représentation des données circulant sur le réseaux va parfois poser des problèmes et créer des confusions dans la lecture des tables. L'IEEE considère que le premier bit émis est le bit de poids faible. Cette représentation n'est pas intuitive car ne correspond pas au sens naturel de lecture de la gauche vers la droite. La valeur hexadécimale `0x7A` (soit `0111 1010`) indique que les bits `0`, puis `1` puis `0`,... sont transmis sur le support physique. Si l'on écrit ces bits dans l'ordre où ils sont transmis, on obtient la valeur binaire `0101 1110` soit la valeur hexadécimale `0x5E`. Cette dernière représentation est utilisée par l'Ethernet ou l'Internet. Ceci se remarque sur les adresses de multicast, la valeur du premier octet de l?adresse servant à indiquer une trame en diffusion est 0x01 et non 0x80.

Questions

	F	T
- Multicast is possible with MAC-16?	○	○
- Local addresses can be set with bit U/L=0	○	○
- `FF-01-02-03-04-06` is a broadcast address	○	○
- It is possible to forge the MAC address of another equipment	○	○

- IPv6 uses MAC addresses such as `33-33-FF-31-67-A1`, What is the nature of this address ?

IEEE Model

Interconnection

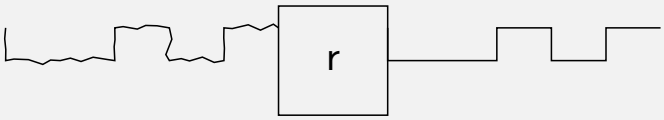
Interconnection

- IEEE 802.1 group defines interconnection method.
- Interconnection can occur at different level in the OSI Reference Model
  - Layer 1: Repeater
  - Layer 2: Bridge
  - Layer 3: Router
  - Layer 7: Gateway, Proxy, ALG (*Application Level Gateway*)
    - Some old documentation uses Gateway for Router
- IEEE works on Repeaters . . .
  - at bit level
- . . . and Bridges
  - In store and forward mode; store frames, analyze them and forward on appropriate interface,
  - Bridges learns addresses location reading the frame source field
  - No configuration is needed

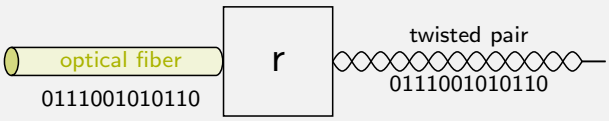
Repeater

Different functions:

- Regenerate signals



- Change media

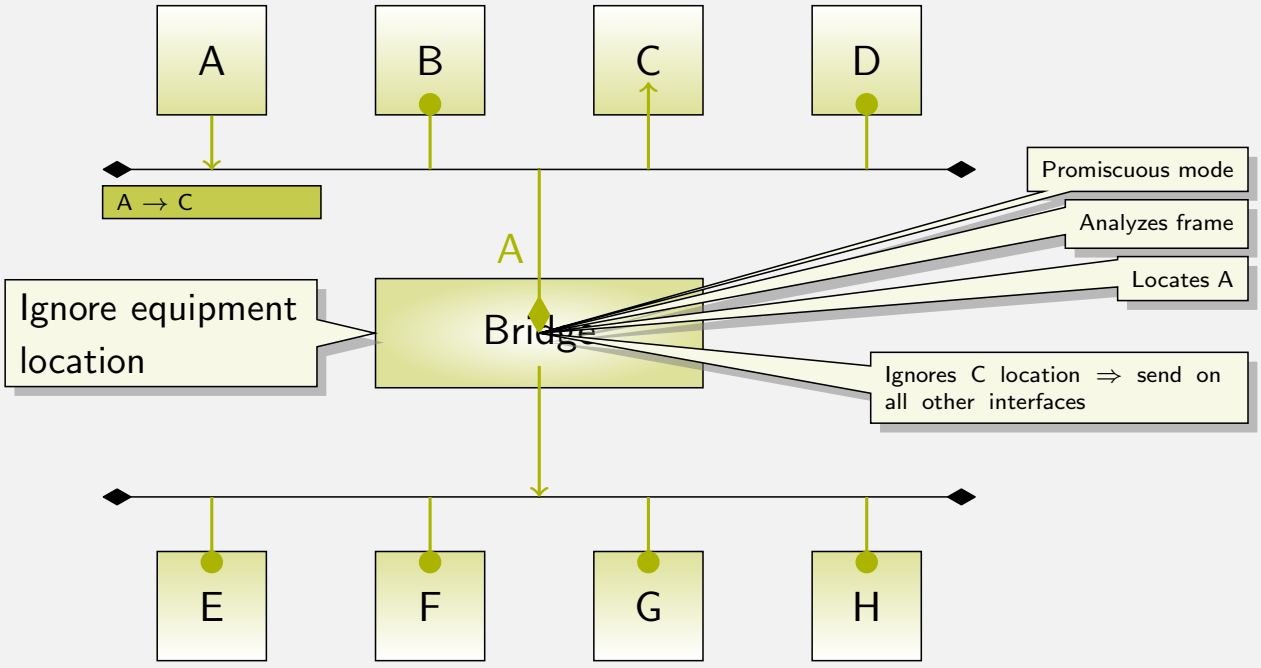


- Propagation delay inside the repeater can be longer than equivalent cable length:
  - time to detect binary values and generate a new signal.
- Almost invisible to upper layers

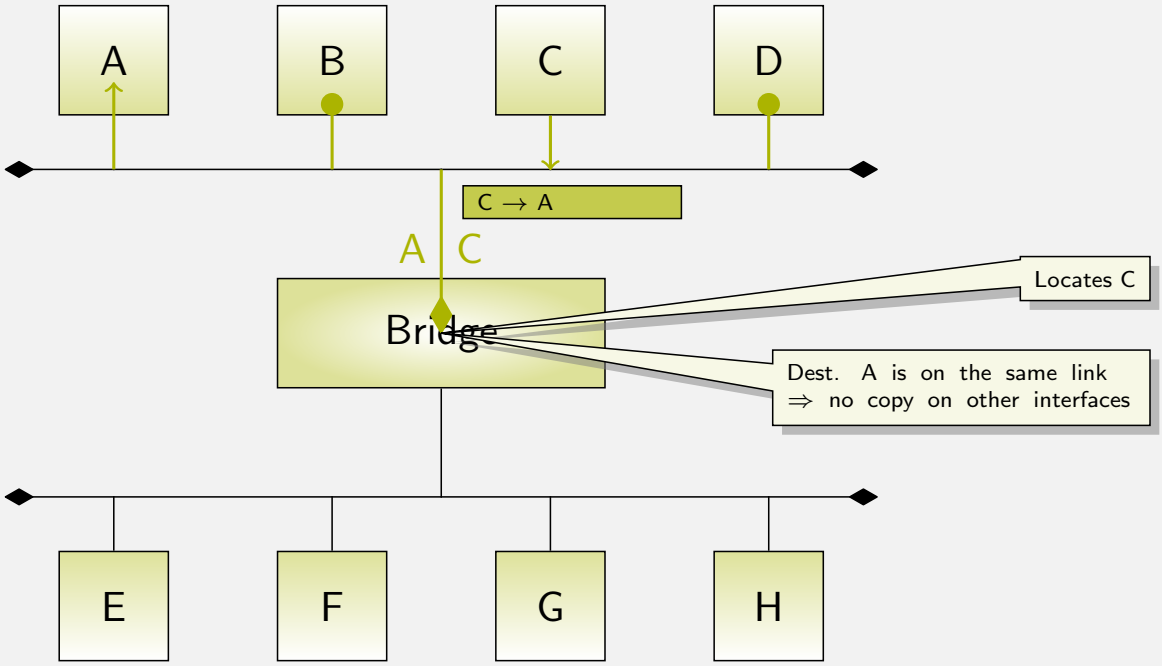
Comments I

Le répéteur permet de régénérer un signal. Avant que le codage des bits soit trop atténué par la traversée du support, le répéteur lit les bits qui circulent et les recopie vers un autre support. Il permet d'augmenter la portée du réseau. Le répéteur peut aussi permettre de passer d'un type de support physique à un autre, par exemple, d'un support en paire torsadée vers une fibre optique. Le temps de traversé d'un répéteur peut être relativement long (comparé au temps de propagation d'un signal sur un support physique) car il faut pouvoir décoder le signal binaire avant de le retransmettre.
Un répéteur ne permet pas de changer de technologie de réseau, par contre il est invisible aux autres couches et ne demande aucune administration particulière.

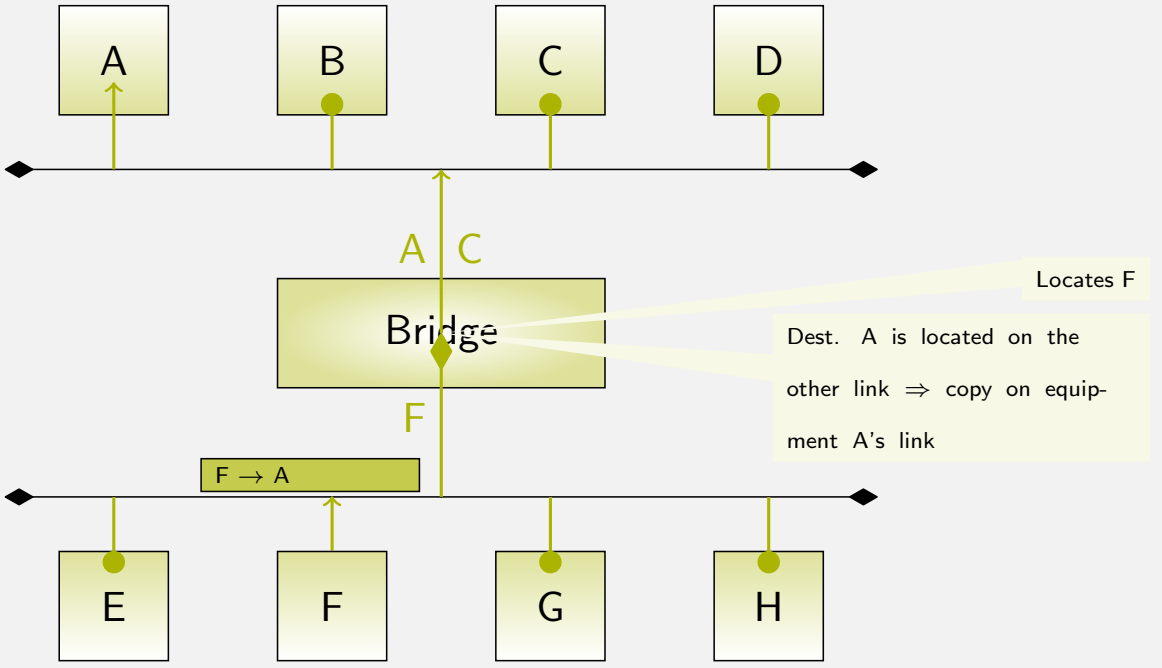
Bridge



Bridge



Bridge



Advantage

- Bridges learns equipments' location
- Copy frame on other interface only when needed
  - improve security: cannot listen traffic between two equipments located on another link
  - decrease traffic on a link.
- Can work with different L2 technologies:
  - Addressing rules must be the same: e.g. MAC-48
  - All fields must be easily filled:
    - No configuration,
    - No queries on the link.
  - For instance: Wi-Fi and Ethernet
    - Bridge is called an Access Point
- No scalability
  - All equipments' addresses must be memorized
  - Broadcast/Multicast is sent everywhere

Comments I

Par convention, le terme pont désigne un équipement permettant une interconnexion au niveau 2 du modèle de référence de l'ISO. Un pont travaille donc au niveau trame. Pour qu'il ait un intérêt, il doit être connecté à au moins deux sous-réseaux. Le principe de fonctionnement d?un pont est, dans les grandes lignes, relativement simple :

- le pont écoute toute l'activité (c'est-à-dire toutes les trames qui circulent) sur chaque sous-réseau, on appelle ce mode *Promiscuous* ;
- il stocke dans sa mémoire les trames qu'il sélectionne suivant des critères que nous détaillerons dans la suite ;
- il retransmet vers le (ou les) autre(s) sous-réseau(x) les messages stockés en mémoire.

Comments II

Un pont n'a pas besoin d'une adresse MAC pour fonctionner. On le dit transparent. Le pont recopie dans sa mémoire certaines des trames qui l'atteignent. Pour l'émission, le pont se comporte comme une station ordinaire sur le réseau, sauf qu'il ne modifie pas l'adresse MAC de la source dans la trame. Sa présence ne peut pas être détectée. La seule différence, mais difficilement observable, est l'augmentation des délais de propagation dus au stockage de la trame complète dans la mémoire du pont et au temps d'accès aux autres sous-réseaux..
Si les ponts ne font que retransmettre les données d'un réseau vers l'autre, cela peut conduire à des engorgements du réseau. Par exemple, dans la topologie représentée dans les transparents précédents, supposons qu'à un instant donné la station A dialogue uniquement avec la station C et la station F avec la station G. Si la somme de tous les trafics est supérieure à la capacité de chaque sous-réseau, le pont ne va pas pouvoir retransmettre le trafic d?un réseau vers l'autre. Sa mémoire va se remplir et aléatoirement des trames de données seront perdues. Ce n'est pas grave pour le trafic de A vers C et de F vers G, par contre, cela peut affecter fortement le trafic entre les stations C et F.

Comme les trames émises par la station A n'intéressent pas les stations situées sur le réseau 2, et les trames émises par la station F n'intéressent pas les stations situées sur le réseau 1, le pont peut filtrer les trames pour ne laisser passer que celles qui doivent nécessairement aller sur un autre sous-réseau. Pour cela, le pont doit connaître le lien sur lequel est connectée la station. Un administrateur peut se charger de cette opération, mais elle est lourde à mettre en place, car si une station est déplacée, le pont doit être reconfiguré. Une procédure automatique rendant le pont complètement transparent permettrait de le brancher pour qu'il filtre correctement les messages. Cette connaissance de la "position" des stations peut être facilement acquise par les ponts. Par construction un pont est en mode *Promiscuous*, il prend une copie de toutes les trames qui circulent sur un sous-réseau et il en connaît le format. Il peut donc en extraire le champ adresse source. A partir de l'adresse de la source, le pont connaît l'existence des stations se trouvant sur le sous-réseau. Le pont construit une table qui contient la localisation de chaque station. L'algorithme de sélection des trames, quand le pont voit passer une trame, sera le suivant :

- si le destinataire est sur le même sous-réseau, la trame est ignorée,

Comments III

- si le destinataire est sur un autre sous-réseau, le message est recopié sur cet autre sous-réseau,
- si le destinataire est inconnu, le pont recopie le message sur tous les autres sous-réseaux,
- si le message est un message en diffusion (premier bit de l'adresse MAC à 1) le pont le recopie sur tous les autres sous-réseaux.

Les ponts qui utilisent cette technique sont appelés ponts filtrants auto-apprenants ou ponts transparents car ils ne nécessitent aucune configuration. La plupart des ponts actuellement utilisés sont de ce type. Ces ponts présentent de nombreux avantages :

- ils n'ont besoin d'aucune configuration. Le pont est connecté aux sous-réseaux et acquiert automatiquement la configuration du réseau,
- ils n'obligent pas à modifier la configuration des stations. Le pont étant invisible aux stations, celles-ci ont une vision totale du réseau,
- ils acceptent tous les types de protocole de niveau supérieur. Le pont recopie les trames d'un réseau à un autre sans modification,
- ils diminuent la charge totale du réseau en limitant la propagation d'un message à un sous-réseau,
- ils augmentent la sécurité du réseau en ne faisant pas circuler sur tout le réseau les messages émis par une station.

Questions

What does a bridge when it receives a Broadcast frame ?

- ◯ Discard frame
- ◯ Copy frame on all other interfaces

	F	T
A bridge may change the frame format	◯	◯
A bridge does not work if interfaces' speed is different	◯	◯
A bridge is limited to two interfaces	◯	◯
It is possible to put more than one bridge on a network	◯	◯

Routers

- Work at L3 (packet) level:
  - IP Layer
- Addresses follow a logic (generally hierarchical)
  - ⇒ common part of the address (prefix) is used to locate equipments
- Routers uses this property for scalability
- Routers have to be configured
  - Manually and through routing protocols
- Bridges and routers have almost the same behavior, they can be implemented in the same equipment
  - Bridge some protocols, router the others
  - Bridge some parts of the network, route traffic between these parts

Comments I

D'un point de vue technique, un routeur et un pont fonctionnent a peu près de la même manière; prendre les PDU (trames ou paquets) circulant sur un lien et les recopier sur une autre. Mais les propriétés de l'adressage change radicalement radicalement entre le niveau 2 et le niveau 3. Au niveau liaison, les adresses sont réparties de manière aléatoire, forçant les ponts à mémoriser l'ensemble des adresses. Au niveau réseau, les adresses sont structurées et réparties suivant une logique; toutes les adresses des équipements connectés au même réseau local possèdent une partie commune. Le routeur sera configuré pour envoyer vers ce réseau tous les paquets ayant cette partie commune dans l'adresse de destination. Ainsi, le routeur n'a pas besoin d'avoir une entrée par équipement, mais une seule entrée qui peut correspondre à un nombre important d'équipement. De cette manière, la résistance au facteur d'échelle est augmenté et il est possible de construire des réseaux mondiaux, comme le réseau Internet. En contre partie, la facilité de configuration que l'on avait au niveau 2 disparait et il est nécessaire de configurer le routeur. cela doit se faire de manière manuelle pour les réseaux auxquels le routeur est directement connecté et par des protocoles dit de routage pour apprendre les configurations distantes.

Un même équipement peut intégrer les fonctions de pontage et de routage; certains protocoles pourront être routés et d'autres pontés. Dans les réseaux IP, il est courant d'avoir des équipements qui pontent certaines parties du réseau pour profiter des facilités d'auto-configuration et de router le trafic entre ces parties pontées.

Gateways

- Work at application level
- Depend of the application
  - From a L7 application to another
  - i.e. change Telephony Signalization (SIP to H.323)
- Application can be the same on both side:
  - Security: filtering (proxy)
    - Only known and well formed application protocols are authorized.
  - Performances (cache)
    - Some information are stored locally to reduce traffic (P2P, HTTP, ...)
  - Transition from IPv4 to IPv6 (ALG: Application Level Gateway)
    - Same application, but different L3 protocols
    - Also used from a private Internet to the public Internet
- Not all L7 protocols can use gateways

Comments I

autrefois fois, une passerelle permettait de passer d'un protocole d'une famille à l'autre, comme le courrier X.400 de l'ISO vers le courrier SMTP du monde IP. Outre le fait que ce genre de traduction est relativement complexe et peu performant,Avec la généralisations des protocoles au dessus d'IP, la notion de passerelle (Gateway) a légèrement évoluée. Au lieu de passer d'un protocole applicatif à un autre, la passerelle va garder le même protocole applicatif des deux côtés. Il s'agit de faire le relais entre deux réseaux qui ne sont pas naturellement interconnectés, pour par exemple:

- contrôler les protocoles utilisés. La passerelle va recevoir les requêtes venant d'un équipement, si celles-ci ne sont pas syntaxiquement correcte ou utilisent des valeurs interdites, la passerelle refusera de les relayer ou de les reformuler vers l'extérieur.
- augmenter les performances du réseau. La passerelle peut mémoriser les requêtes et les réponses associées et en cas de demande similaire retourner la valeur précédemment mémorisées.
- permettre de passer d'un réseau à un autre en cas d'impossibilité de connexion aux couches inférieures, comme par exemple quand une partie du réseau est en IPv4 et l'autre en IPv6 ou si un site utilise un adressage privé non routable à l'extérieur.

IEEE Model

IEEE groups overview

IEEE groups overview

- IEEE 802.1: Architecture, Addresses, Interconnection, Virtualization
- IEEE 802.2: Equivalent of Link Layer (see later)
- IEEE 802.3: Ethernet
  - Physical Layer (see later)
    - from 10 Mbit/s to 100Gbit/s
    - support: Coaxial (obsolete), twisted pair, optical fiber
  - MAC Algorithm:
    - From shared media with CSMA/CD (Carrier Sense Multiple Access/Collision Detect) to switched (brigded) full duplex links
  - Compatible Frame Format since first proposals

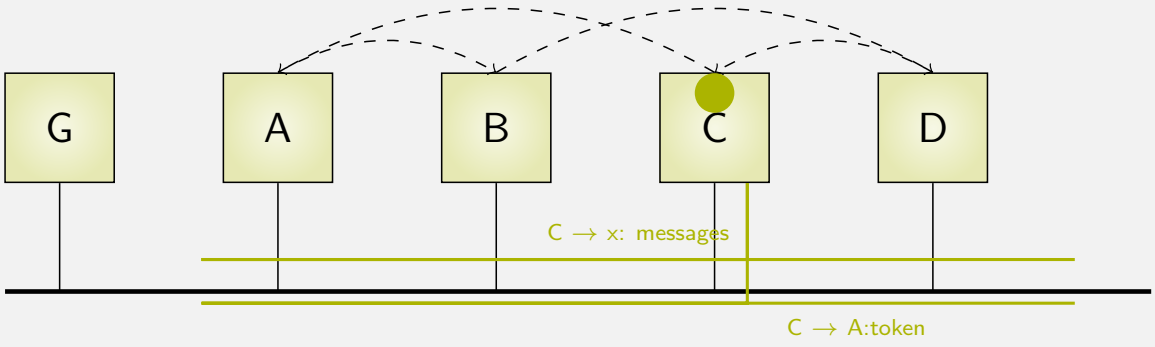
Comments I

Le groupe IEEE 802.3 pour le CSMA/CD (*Carrier Sense Multiple Access/Collision Detect*). La topologie est en bus, le principe d'accès est simple. Les équipements écoutent le canal avant d'émettre. Si le canal est silencieux, la station peut émettre sa trame, sinon l'émission est différée. Au lieu d'éviter à tout prix l'émission simultanée par plusieurs sources (appelée collision), le protocole essaie de résoudre ces conflits. Les stations impliquées attendent un délai aléatoire avant de tenter une nouvelle transmission. Le protocole est très simple à mettre en œuvre, il ne nécessite pas l'échange d'information entre les équipements pour gérer le droit de parole. Cette simplicité se traduit par un très faible coût des équipements. Le protocole et sa variante qu'est Ethernet, sont présentés au en détail dans la suite.

IEEE 802.4; Token Bus

- Bus: Coaxial cable where all equipments are connected
  - Broadband: only a range of frequencies where allocated to token bus
- Token: Right to talk, must circulate among equipments
  - more deterministic than CSMA/CD
  - each equipment can keep the token for a maximum time
  - can easily compute maximum waiting time
  - better performances than CSMA/CD when the network is loaded,
  - worse performances when the network is not loaded.

Token Management

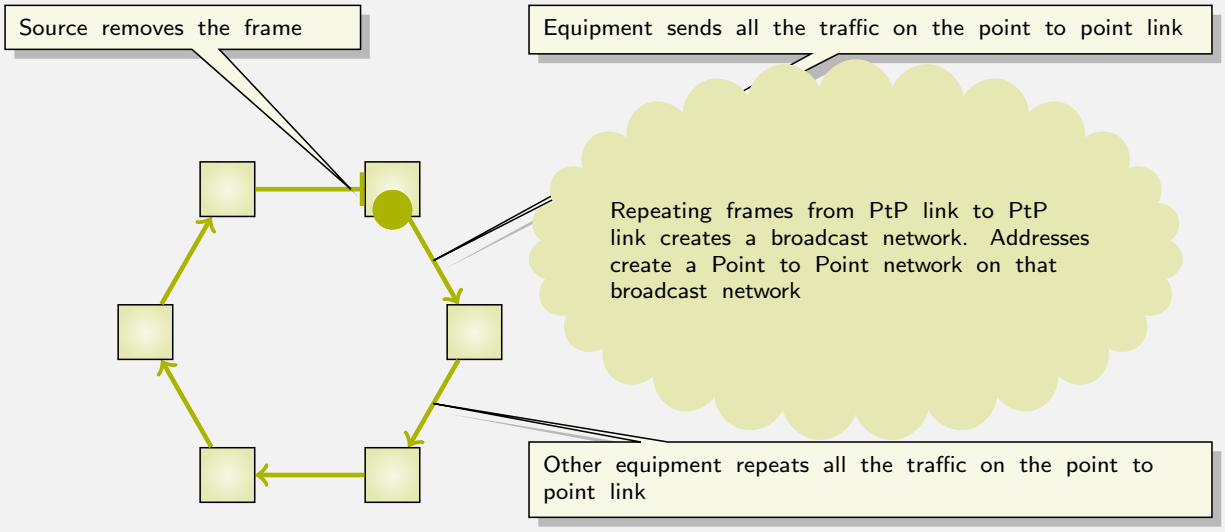


- How to manage token circulation on a bus ?
  - What his the next equipment?
  - How new equipments can enter into the network?
  - What append if the token disappear ?
- Expensive technology, reserved to factories:
  - broadband, token management, less production

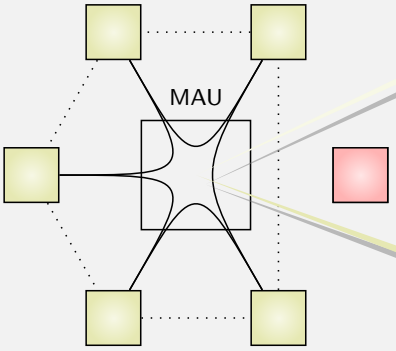
Comments I

le comité IEEE 802.4 pour le bus à jeton (Token Bus) défini par General Motor pour des applications industrielles car, contrairement à Ethernet ou au IEEE 802.3, ce protocole permet de garantir une borne maximale pour le temps d'émission d'un message sur le réseau. Le droit à la parole est symbolisé par la possession d'un message spécial appelé jeton. La transmission d'un message se base sur les propriétés de diffusion naturelle des réseaux locaux. Par contre, la transmission du jeton doit absolument être en point-à-point puisqu'une seule station à la fois doit en être en possession. Il faut construire artificiellement au-dessus du bus un anneau virtuel permettant de faire circuler le jeton. Une grande partie de la complexité du protocole va venir de cette gestion. La figure précédente permet d'illustrer quelques-uns de ces problèmes. Supposons que la station C tombe en panne, la station D va lui envoyer le jeton qui sera perdu. La station D va devoir entrer dans une phase de reconfiguration de l'anneau pour trouver un autre successeur. Un autre problème est posé pour l'insertion d'une station dans l'anneau virtuel, la station G ne peut pas s'insérer dans l'anneau puisque qu'aucune station ne lui envoie le jeton. Le protocole prévoit que périodiquement les stations actives doivent tester la présence de nouveaux équipements. Le protocole de gestion de l'anneau sur un bus est relativement complexe, impliquant que les cartes le mettant en œuvre déroulent un algorithme. Cela requiert sur la carte, un CPU, de la mémoire, etc. De plus, les garanties déterministes offertes par le bus à jeton ne sont pas nécessaires pour les applications bureautiques. Ce protocole n'est presque plus utilisé ;

IEEE 802.5; Token Ring



IEEE 802.5; Token Ring



Medium Attachment Unit (MAU) emulates ring topology.
Real topology is a start topology
MAU detects if an equipment is down. In that case it is excluded from the ring

Almost all the LAN are based on that topology, cabling is the same, only the central equipment gives the nature of the network.

Token Ring management

- Token management is simpler than Token Bus;
  - Token follow the physical link
  - No need to address the token destination, just send it on the wire
- Token Ring was very popular:
  - Supported by IBM,
  - Offer some delays guaranties
- Two speeds 4 Mbit/s the 16 Mbits
  - 100 MBit/s standardization arrived too late
  - Some proprietaries solutions exists, but without standards it was risky to move
  - Ethernet evolutions: 100Mbit/s, switching made Token Ring not so competitive

Comments I

le groupe IEEE 802.5 pour l'anneau à jeton. Le mécanisme de gestion du droit de parole est aussi basé sur un jeton, mais la circulation de celui-ci est simplifiée du fait qu'un anneau existe physiquement. IBM a annoncé les premiers prototypes d'anneau à jeton en 1981. Et c'est en 1985 que le premier réseau ayant un débit de 4 Mbit/s était commercialement disponible en même temps que la norme ISO 8802.5. Le principe de fonctionnement est relativement simple dans les grandes lignes. Un médium de communication composé de N liaisons point-à-point relie circulairement l'ensemble des N stations voulant être membres du réseau (cf. figure page 81). Pour que le réseau puisse fonctionner, il faut qu'une et une seule station puisse émettre des données à un instant précis. Le droit à la parole est symbolisé par la possession d'un jeton. Il s'agit d'une trame particulière qui circule de station en station en suivant la topologie en anneau du réseau. Si une station veut émettre un message, elle attend de recevoir le jeton, le retire du réseau, émet son message, puis réinsère le jeton dans l'anneau. Si une station n'a rien à émettre, elle laisse passer le jeton. Pour qu'une trame arrive à destination, elle doit être recopiée de station en station. Le destinataire continue la retransmission, tout en gardant une copie pour lui-même. Quand le message a fait un tour complet, l'émetteur le retire du réseau en ne le recopiant pas sur l'autre support. Le message faisant un

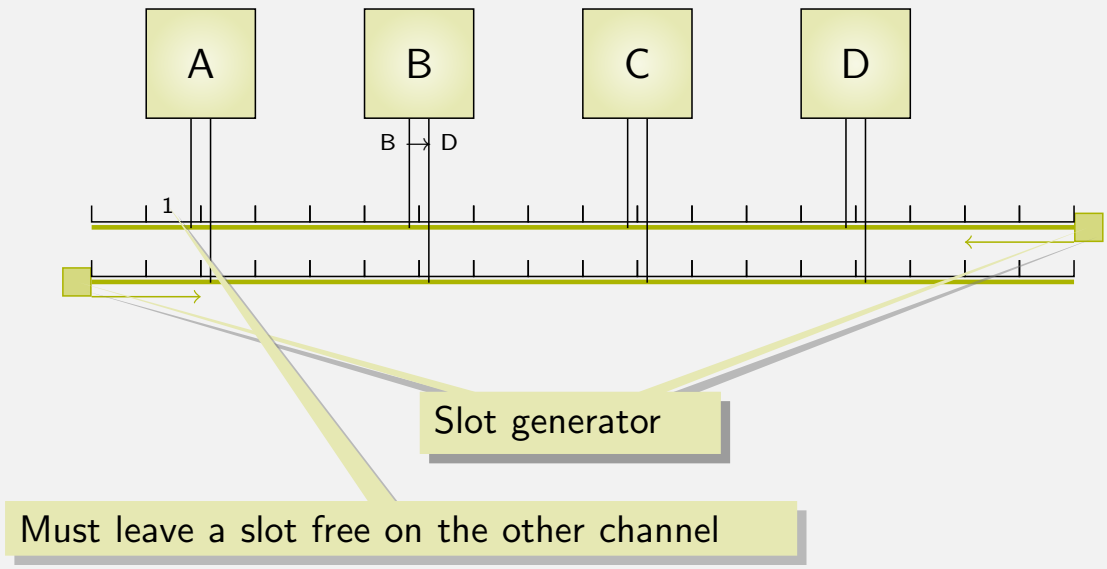
Comments II

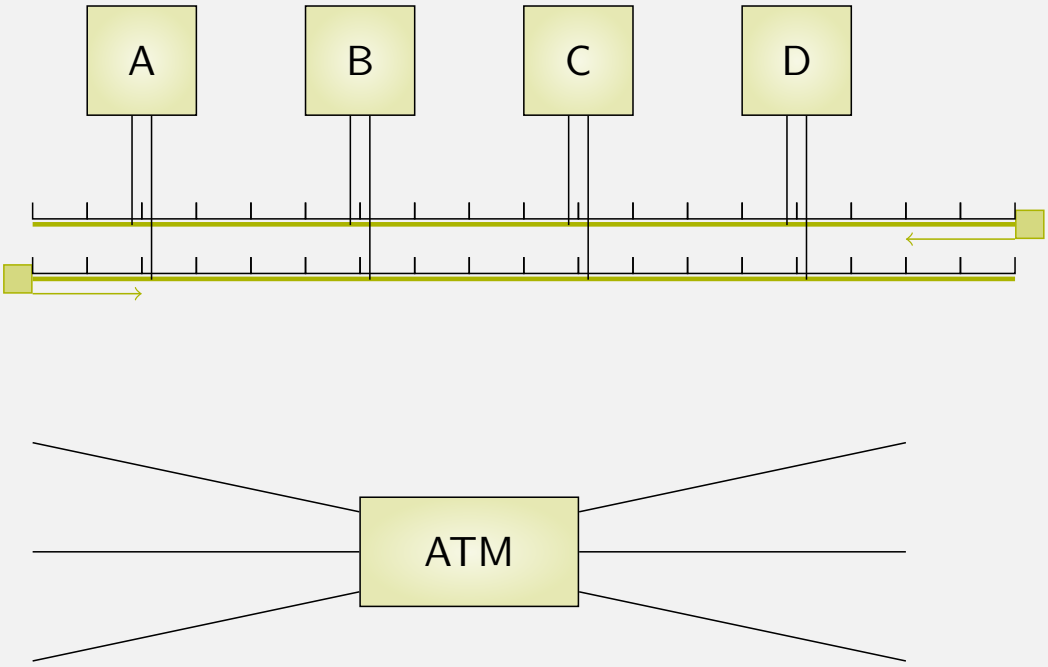
tour complet permet de traiter les cas de multicast, mais fait aussi office d?acquittement pour l'émetteur qui voit revenir intact son propre message. Pour résoudre ces problèmes, un placement en étoile est préféré à un placement en boucle. Un équipement central contiendra en interne la topologie en anneau. Un double câblage permettra au signal d'aller et de revenir de la station. L'équipement MAU (*Medium Attachement Unit*), bien qu'il ne modifie pas le signal, doit être assez "intelligent" pour détecter : une coupure du câble, une panne de la station ou une mise hors tension de celle-ci. Dans ces cas, le MAU doit fermer le circuit comme indiqué sur le transparent page 81.
Bien que le méthode d'accès au canal présente des avantages, comme un délais d'émission borné des trames et la possibilité de définir des priorités, cette technologie est très fortement en perte de vitesse au profit d'Ethernet. En effet, elle n'a pas été aussi réactive qu'Ethernet pour s'adapter aux nouveaux modes de transmission.

IEEE 802.6; Distribute Queue Dual Bus

- Protocol for Metropolitan Area Networks:
  - Long propagation delay
  - None of the original solutions (shared Ethernet, Token Bus or Token Ring have good performances if propagation delays are high)
- Based on two buses, each with a direction:
  - A slot generator gives the place where equipment can send information
- Equipments knows their and others positions.
- Send information on the right bus to reach the destination
- To maintain fairness between equipments, DQDB develops an algorithm:
  - set a bit to one on slot going the opposite way.
  - when receiving a slot with a bit set to one, don't use the slot

DQDB





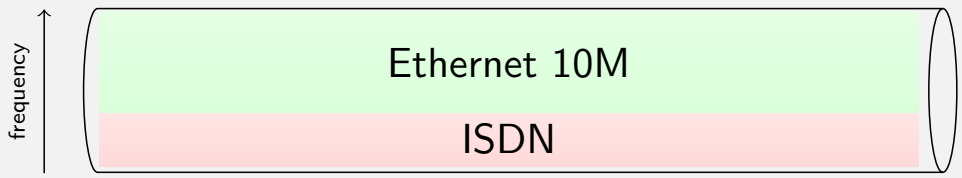
- No real fairness:
  - End equipments got more bandwidth than central ones.
- Last attempt to use a distributed shared media:
- ATM was used instead
  - Star topology around a switch;
  - Also based on fixed slots (cells)
- then Ethernet
  - switched mode has no distance limitation
  - optical fiber allows long distance
- Metro Internet Forum focuses on those architectures:
  - add scalability
  - add flow separation between groups of users.
  - W http://metroethernetforum.org

Comments I

En 1990, le groupe IEEE 802.6 fut ajouté pour traiter le cas des réseaux métropolitains (MAN). Ce protocole, aussi appelé DQDB (Distributed Queue Dual Bus), est basé sur deux bus véhiculant l'information dans un sens différent. A chaque extrémité, un générateur produit des slots dans lesquels les équipements pourront émettre leur message. Quand une station veut émettre des messages, elle détermine quel bus permettra de joindre le destinataire. Elle positionne dans un slot de l'autre bus, un drapeau indiquant aux stations en amont son intention d'émettre. Dans la station, un mécanisme d'accès basé sur des compteurs permet de déterminer le slot libre dans lequel le message sera émis. Le protocole DQDB devait être utilisé dans les réseaux métropolitains pour transmettre à la fois des communications téléphoniques et des données. Il est arrivé trop tard, car avec les progrès technologiques en électronique, il était plus préférable d'utiliser une topologie en étoile, comme ATM, à une topologie sur support partagé. ATM garde la notion de PDU de taille limitée et fixe, appelés dans ce cas cellules. Puis les évolutions d'Ethernet permettant aussi bien l'utilisation de la fibre optique que la commutation ont permis de créer ces réseaux métropolitains. Le forum MetroEthernet développe ces architectures.

IEEE 802.7-9

- Technical Advisery Group (TAG):
  - don't produce standards, help other groups and make liaison with other organizations
  - IEEE 802.7: Broadband
  - IEEE 802.8: Optical Fiber
- IEEE 802.9: Integrated Services LAN (isoEthernet)
  - Share physically telephony and data networks



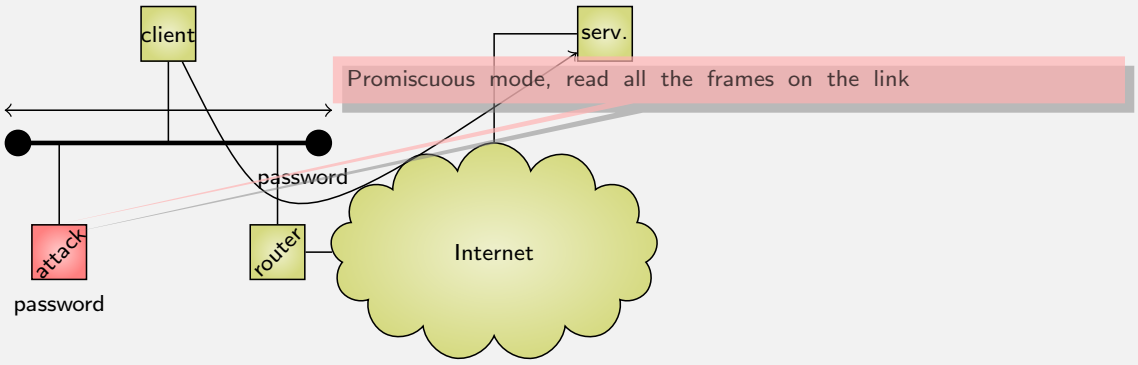
- Too expensive and too limited
  - replace with Voice over IP (VoIP)
  - voice is digitalized and put into frames
  - possible with the throughput raise and frame switching

Comments I

Deux groupes de conseils techniques (TAG : *Technical Advisery Group*) furent créés pour servir de liaison avec les autres groupes et pour les aider techniquement dans les bons choix technologiques (802.7 et 802.8). Ces deux TAG ne produisent pas de norme. Le document produit par le groupe 802.7 spécifie la conception, l'installation et les paramètres de test pour les réseaux utilisant un codage par fréquences des informations binaires (10BROAD36, IEEE 802.4,...). Le codage en fréquence permet un multiplexage des données et la coexistence de réseaux de nature diverses (données, vidéo,...) sur le même support. Le groupe de travail IEEE 802.8 traite de la même manière du câblage en fibre optique pour les réseaux locaux et métroplitains ; le groupe IEEE 802.9, ou isoEthernet, normalise des techniques d'accès pour les réseaux intégrant la voix et les données. Le coût du câblage des bâtiments intervient pour une part importante dans le prix d'installation d'un réseau. Or, pour câbler un bureau, il faut apporter au moins deux réseaux : le réseau téléphonique et le réseau informatique. La norme permet de faire cohabiter sur un même support des réseaux RNIS et des réseaux de données (IEEE 802.x ou FDDI). En fait, la tendance actuelle est plutôt de véhiculer des données multimédias comme la voix sur IP sur les réseaux locaux plutôt que de partager physiquement le support entre plusieurs réseaux spécialisés ;

IEEE 802.10; Security

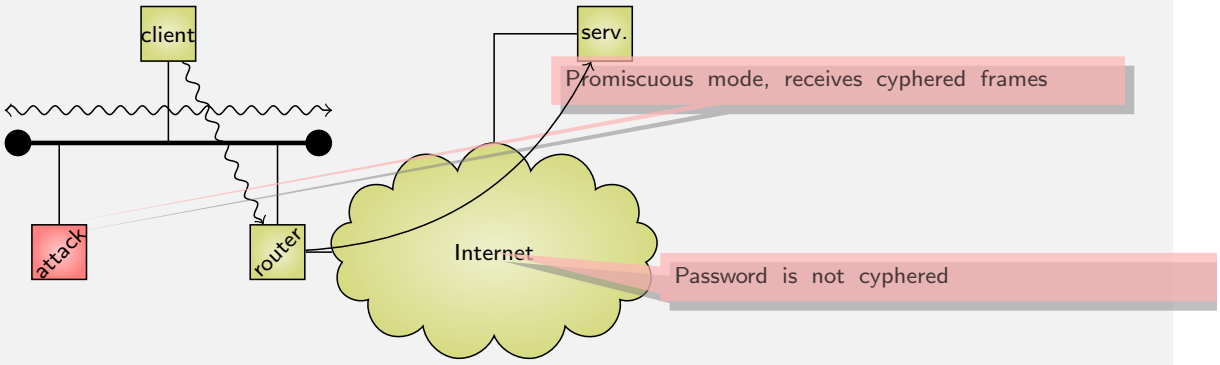
- Cypher frames on the LAN



- only efficient on LAN
  - At layer 3 (IPsec) or 7 (TLS) cyphering is end-to-end
- Ethernet with switching makes this attack impossible
- Wi-Fi uses WEP or WPA (IEEE 802.11i)

IEEE 802.10; Security

- Cypher frames on the LAN



- only efficient on LAN
  - At layer 3 (IPsec) or 7 (TLS) cyphering is end-to-end
- Ethernet with switching makes this attack impossible
- Wi-Fi uses WEP or WPA (IEEE 802.11i)

Comments I

le groupe IEEE 802.10 traite de la sécurité des transmissions. La sécurité des transmissions n'est pas assurée dans les réseaux locaux. Ils sont basés sur des supports à diffusion. Il suffit d'un PC connecté au réseau pour capturer tout le trafic et en particulier les mots de passe qui circulent en clair. Le protocole IEEE 802.10 propose en particulier de chiffrer les données émises entre les équipements. Mais des problèmes liés à la législation ont freiné son déploiement. De plus, d'autres techniques de chiffrement ont été développées pour les protocoles de la couche réseau, ce qui réduit encore l'intérêt de ce protocole.

IEEE 802.11-14

- IEEE 802.11: Wi-Fi since IEEE 802.11b (shared 11 Mbit/s)
  - Evolution up to 100 MBit/s
  - See later in this class
- IEEE 802.12: Attempt to develop a 100 Mbit/s network: failure
- IEEE 802.13: does not exist
- IEEE 802.14: Cable Television Network: failure
  - ITU succeed with DOSCIS (J.112, J.122, J.222)

Comments I

le groupe IEEE 802.11 traite les réseaux sans fils ou WLAN (Wireless LAN). En plus du coût élevé, la nécessité de connecter un équipement au réseau n'est pas adaptée aux nouvelles contraintes liées à l'utilisation d'ordinateurs portables. La norme IEEE 802.11 permet la transmission d'information à des débits de 1 ou 2 Mbit/s en utilisant des ondes hertziennes dans la bande des 2,4 GHz ou des liaisons infrarouges. La portée des émissions peut être de 100 mètres, mais dans des bureaux, où les obstacles sont nombreux, elle est réduite à une trentaine de mètres ;

le groupe IEEE 802.12 propose une alternative pour les réseaux à 100 Mbit/s, appelée aussi 100VG-AnyLAN car elle utilise un câblage adapté à la voix (VG : Voice Grade) pour transmettre des données à 100 Mbit/s. Ce protocole est aussi compatible avec les format de trames IEEE 802.3 et IEEE 802.5 d?où le nom commercial d'AnyLAN. L'élément central des réseaux 100VG-AnyLAN est un hub. Il possède des ports qui permettent de connecter les équipements informatiques. Il possède aussi un port particulier qui lui permet de se connecter à un autre hub 100VG-AnyLAN. Les hubs peuvent être mis en cascade sur trois niveaux. Cette technologie n'a pas réussie à s'imposer face aux évolutions de la norme IEEE 802.3 vers les plus hauts débits ; le groupe IEEE 802.13 par superstition n'existe pas. le groupe IEEE 802.14, créé en 1996, s'occupait de la transmission numérique sur les réseaux de télévision câblés. Il a échoué et le standard que le marché a retenu a été défini par l'UIT sous le nom de DOCSIS (J.112, J.122 et J.222).

IEEE 802.15

- 802.15.1: Bluetooth
- 802.15.2:
- 802.15.3: UWB Ultra Wide Band network:
  - Dismantled
- 802.15.4: WSN Wireless Sensor Network
  - Low Power Network
  - Used by ZigBee, 6LoWPAN (IETF)

Comments I

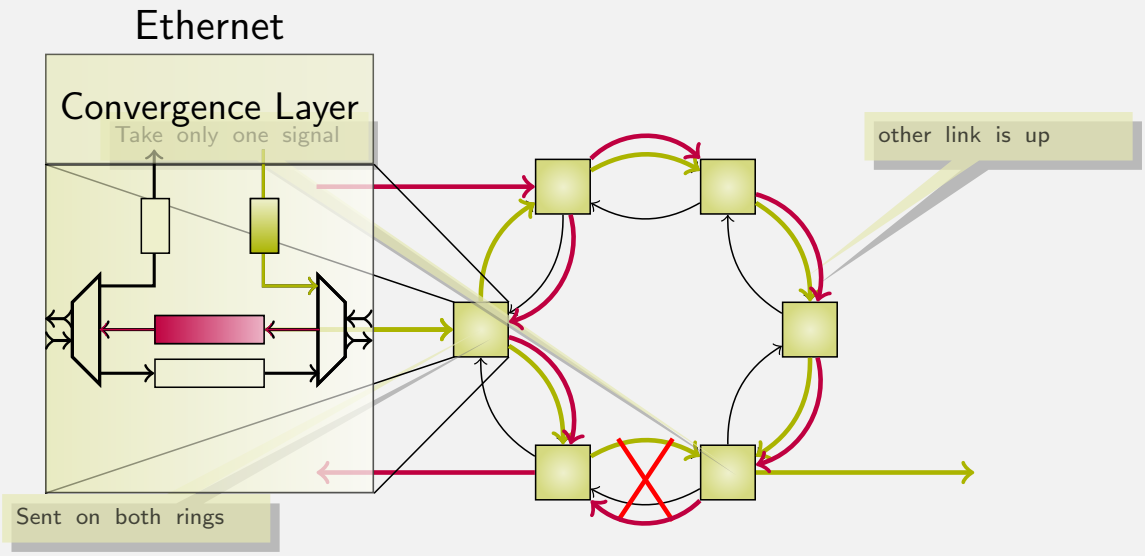
le groupe IEEE 802.15, créé en mars 1999, s'intéresse aux réseaux sans fil personnels (WPAN : *Wireless Personal Area Network*). Plusieurs sous-groupes traitent de technologies différentes. Le sous-groupe :

- IEEE 802.15.1 reprend le standard bluetooth défini par plusieurs constructeurs. Il s'agit de pouvoir communiquer à des débits d?environ 1 Mbit/s dans un rayon de 10 mètres autour d?un individu.
- IEEE 802.15.2 étudie l'intégration des réseaux personnels sans fils et des réseaux locaux sans fils qui peuvent partager les mêmes fréquences de manière différente.
- IEEE 802.15.3 définit des réseaux personnels sans fil a haut-débit (plus de 20 Mbit/s)
- IEEE 802.15.4 définit à l'opposé des réseaux sans fils à très faible débit mais également à très faible consommation. Il pourra être utilisé dans les réseaux de capteur. La partie MAC et les protocoles applicatifs sont connus sous le nom commercial de ZigBee ;

IEEE 802.16: WiMAX

- Oct 2004 : Publication of IEEE 802.16-2004 standard, first WiMAX complete standard, known as Fixed WiMAX. Based on OFDM PHYsical Layer
  - Advertisement Range: 50 kms
  - Practical value: 5-10 kms
- Dec 2005 : 802.16e amendment of 802.16, base of Mobile WIMAX technology
- 2010 : 802.16m Next generation of WiMAX. Data rates of the order of 100 Mb/s, designed as a 4G cellular network (competitor to LTE)

IEEE 802.17: Resilient Packet Ring



More details see:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.8026&rep=rep1&type=pdf

IEEE 802.18-23

- IEEE 802.18 and IEEE 802.19 : TAG
  - Radio Regulatory: Liaison with other standardization bodies (such as ITU-T)
  - Wireless Coexistence: How standards can work together in the unlicensed frequencies.
- IEEE 802.20: Mobility
- IEEE 802.21: hand over
- IEEE 802.22: Regional
- IEEE 802.23: Emergency Services

Comments I

- le groupe IEEE 802.16 traite des accès sans fil pour les réseaux à large bande (BWA : Broadband Wireless Access) ;
- le groupe IEEE 802.17 appelé RPR (Resilient Packet Ring) traite des problèmes de reconfiguration des anneaux SDH. deux groupes techniques (TAG) IEEE 802.18 et IEEE 802.19 traitent respectivement des aspects gestion des fréquences et de la coabitation des différents standards IEEE entre eux.
- le groupe IEEE 802.20 est une alternative à WiMax (IEEE 802.16) en intégrant les aspects mobilités.
- le groupe IEEE 802.21 s?intéresse au passage d'une technologie IEEE à une autre par un utilisateur mobile.
- le groupe IEEE 802.22 utilise le spectre UHF/VHF non utilisé par la télévision pour construire des réseaux régionaux sans fil.

Aloha

Aloha

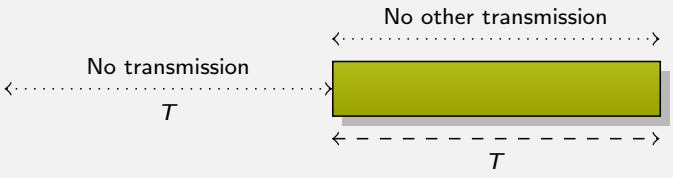
- Simplest protocol on a shared media
- Send when a station want to send
  - Does not listen to neighbor's traffic
  - Collisions (two stations sending at the same time) can occur
  - Unreceived messages are sent again latter.
- Was primary used by Hawaiian universities
  - works well when traffic is very limited
  - unstable when the network is loaded
- Aloha is not CSMA since the media is not sensed before sending

Hypothesis

- All messages have the same size and the then the same duration
- The number of stations is ∞
- The global traffic can be modelized by a Poisson process
  - Inter-arrival is given by an exponential law
- T: Message duration
- λ: number of messages generated in the system per second
- g: number of messages sent per second
- s: number of successful messages per second
  - message without any collision
- Reminder: for a Poisson process with parameter g the probability of having k messages during a period T is $P_k(T) = \frac{(gT)^k e^{-gT}}{k!}$

First step: no retransmission

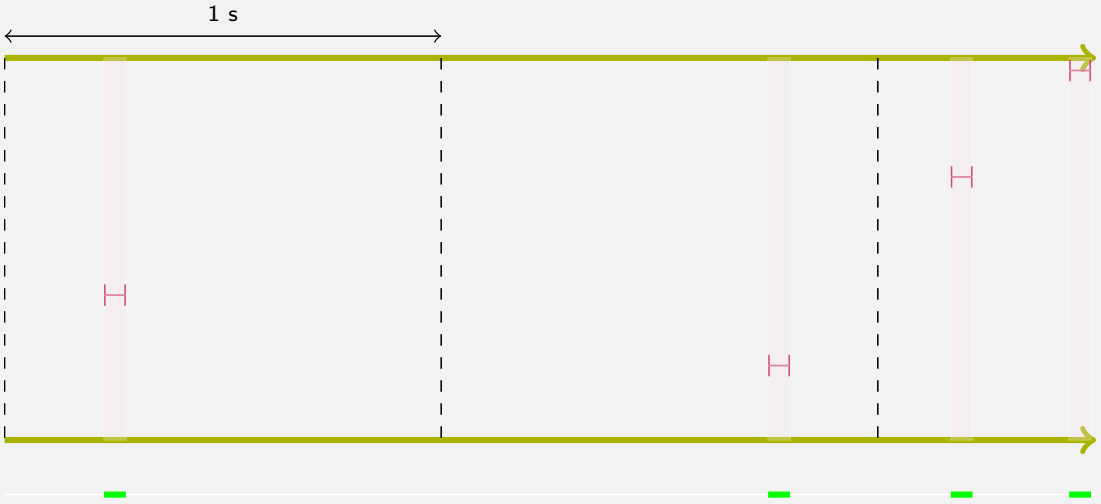
- λ: number of messages generated in the system per second
- g: number of messages sent per second, $g = \lambda$
- s: number of successful messages per second



- $P_{Succ} = P_0(2T) = \frac{(2gT)^0 e^{-2gT}}{0!} = e^{-2gT}$
- $P_{Succ} = \frac{s}{g}$
- $TransmissionNumber = \frac{g}{s}$

S and G

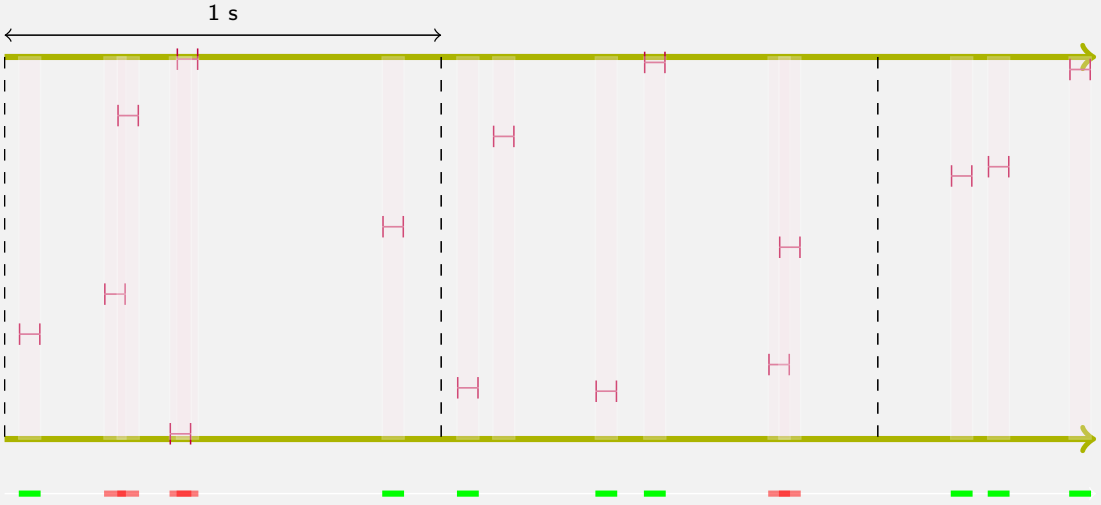
- Normalized throughput $G = g.T$
  - percentage of successful transmission
- Offered load $S = s.T$
  - percentage of channel usage
  - can be > 1



S and G

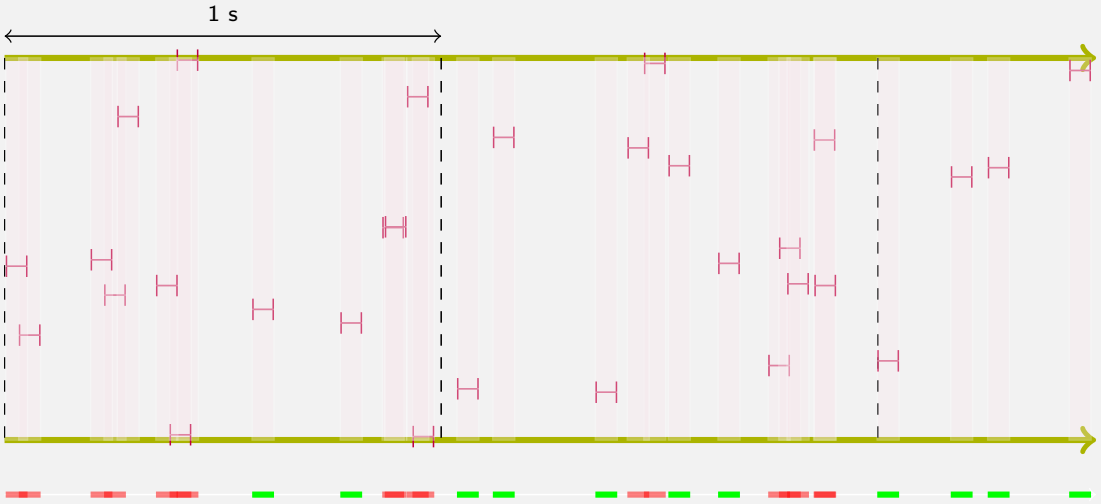
Aloha ▶

- Normalized throughput $G = g.T$
  - percentage of successful transmission
- Offered load $S = s.T$
  - percentage of channel usage
  - can be > 1



S and G

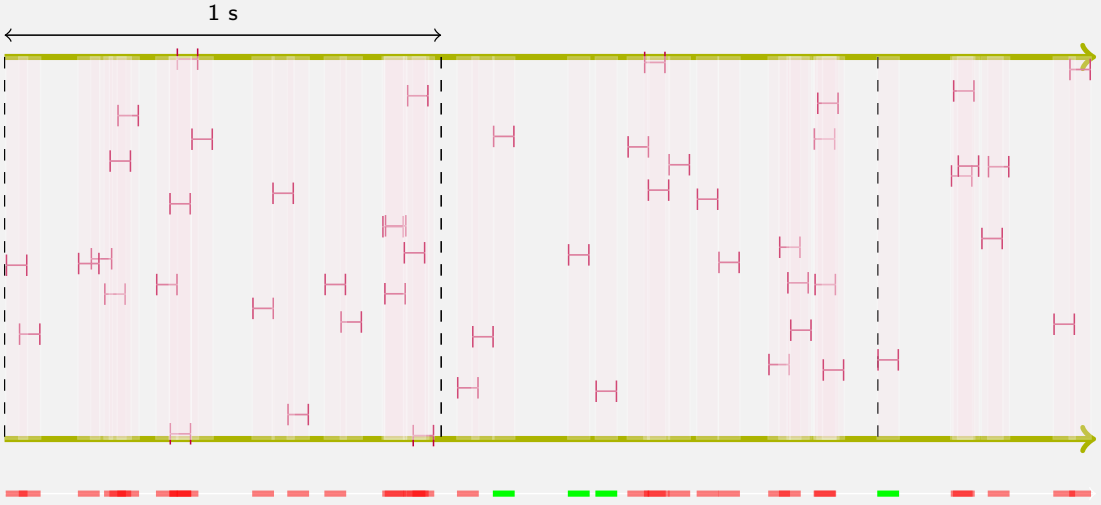
- Normalized throughput $G = g.T$
  - percentage of successful transmission
- Offered load $S = s.T$
  - percentage of channel usage
  - can be > 1



S and G

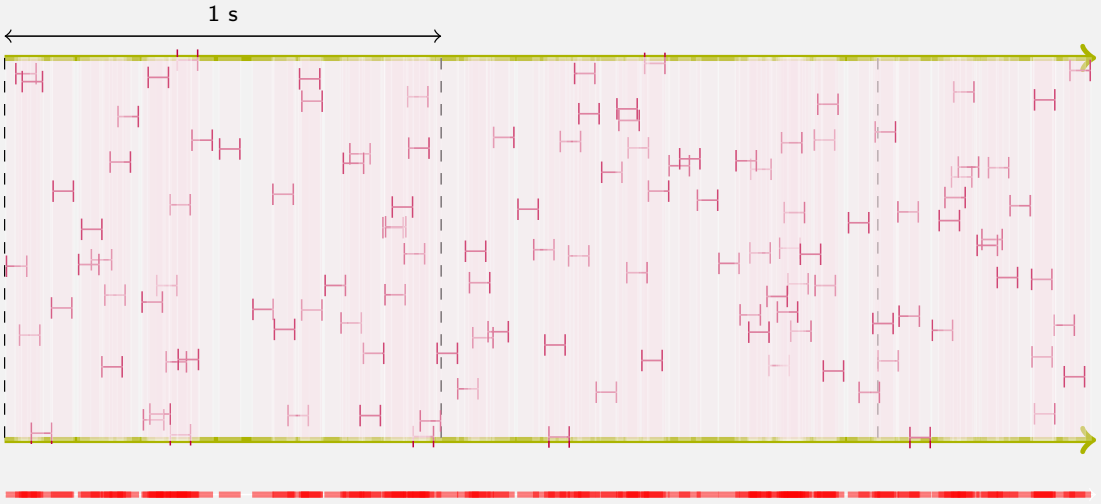
Aloha ▶

- Normalized throughput $G = g.T$
  - percentage of successful transmission
- Offered load $S = s.T$
  - percentage of channel usage
  - can be > 1

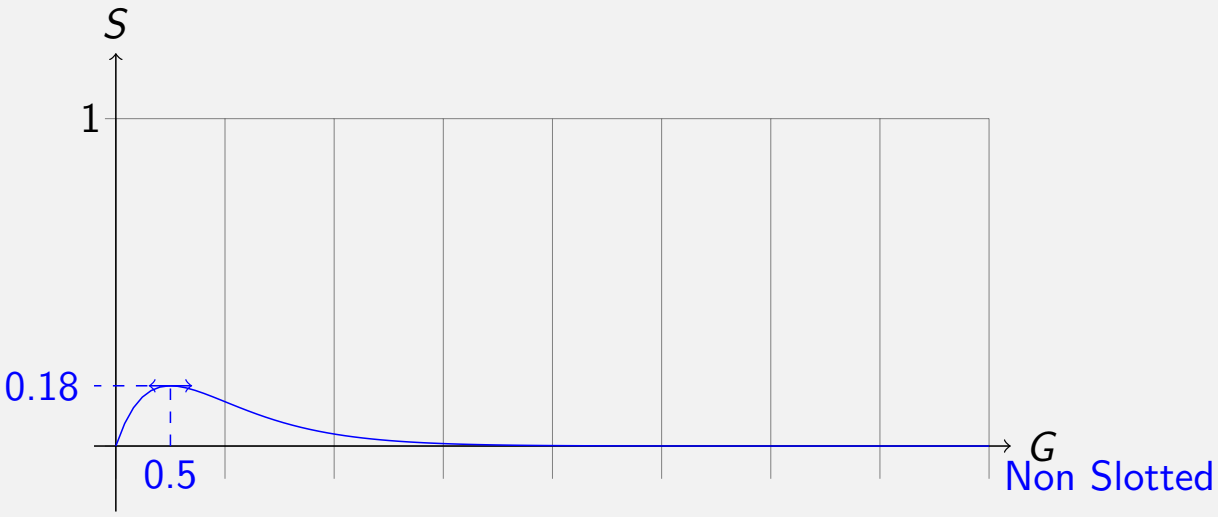


S and G

- Normalized throughput $G = g.T$
  - percentage of successful transmission
- Offered load $S = s.T$
  - percentage of channel usage
  - can be > 1



Performance of non-slotted Aloha

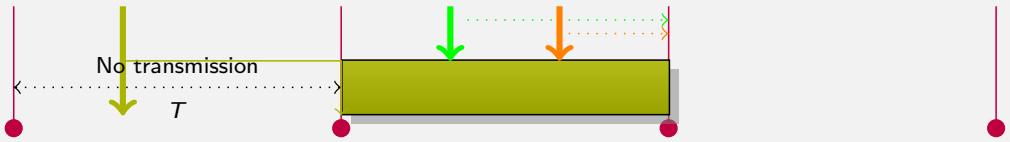


- $P_{Succ} = e^{-2gT} = \frac{s}{g}$ and $G = gT$, $S = sT$
- $S = G.e^{-2G}$
- The maximum of S is $1/2e$ for $G = 1/2$

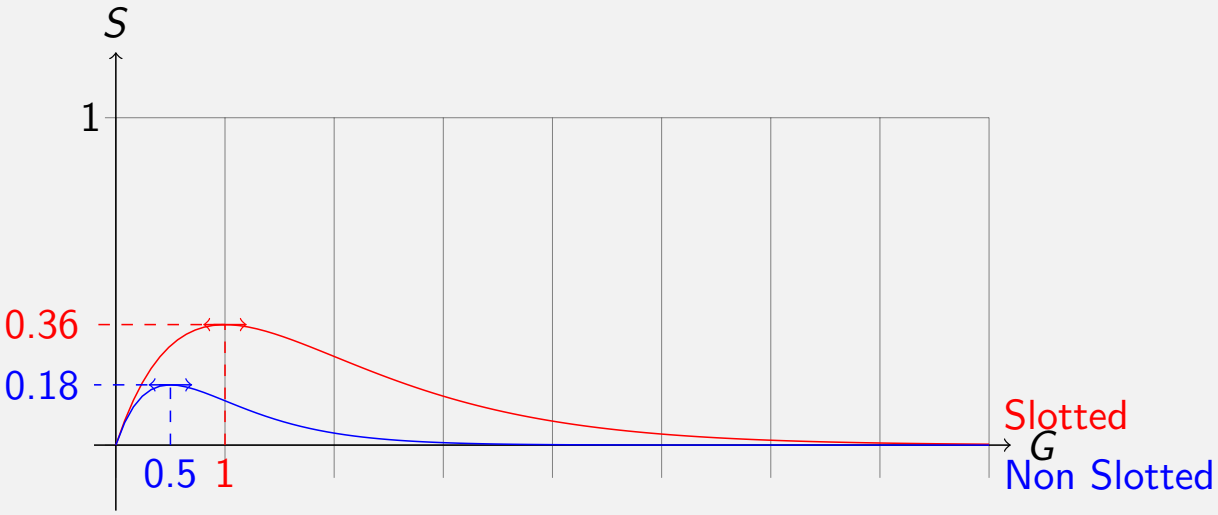
Aloha with a retransmission procedure

- λ: number of messages generated in the system per second
- g: number of messages sent per second
  - Arriving messages (fresh load) plus retransmissions
- s: number of successful messages per second
  - message without any collision
  - when the system works correctly $s = \lambda$
- The global transmission process is assumed to be a Poisson process
- Same result: $S = G.e^{-2G}$ but
  - S is fixed
  - G is to determined
- G/S gives the average number of transmissions per message

Slotted Aloha: Success Probability



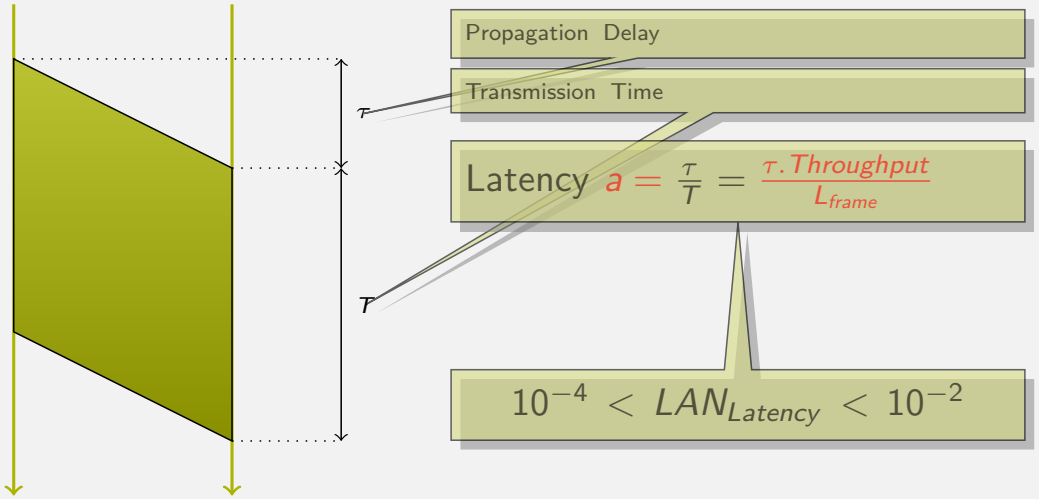
- $P_{Succ} = P_0(T) = \frac{(gT)^0 e^{-gT}}{0!} = e^{-gT} = e^{-G}$
- $P_{Succ} = \frac{S}{G}$
- $S = G.e^{-G}$



- Non Slotted Aloha: max 18% of the channel bandwidth
- Slotted Aloha: max 36% of the channel bandwidth
- When this limit is reached, system become instable
  - Before the limit, increasing the load increase success
  - After the limit, increasing the load, decrease success
- Getting closer to limit, increases instability risk

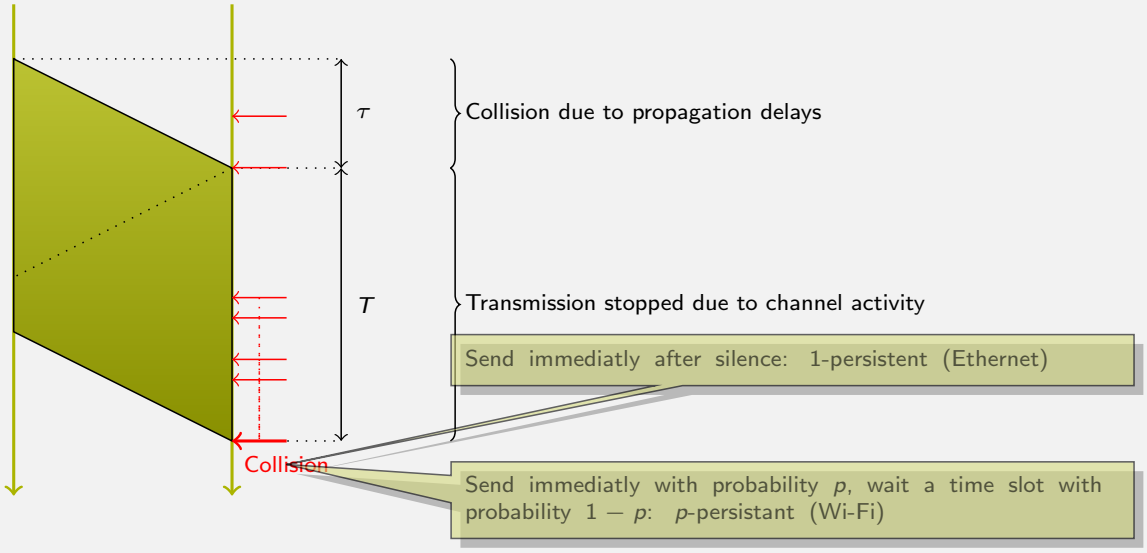
CSMA: Carrier Sense Multiple Access

- Listen to channel before sending
  - Easy with wired technologies, difficult with wireless



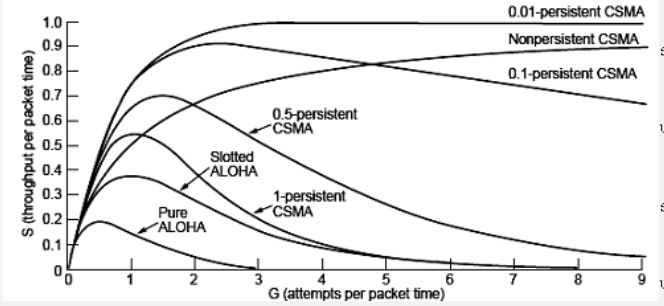
CSMA: Carrier Sense Multiple Access

- Listen to channel before sending
  - Easy with wired technologies, difficult with wireless



Access Method comparison

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.74.8411&rep=rep1&type=pdf



Pure ALOHA $s = ge^{-2g}$

slotted ALOHA $s = ge^{-g}$

unslotted 1 - persistant CSMA $s = \frac{g\left[1 + g + ag\left(1 + g + \frac{ag}{2}\right)\right]e^{-g(1+2a)}}{g(1+2a) - \left(1 - e^{-ag}\right) + (1+ag)e^{-g(1+a)}}$

slotted 1 - persistant CSMA $s = \frac{g\left[1 + a - e^{-ag}\right]e^{-g(1+a)}}{(1+a)\left(1 - e^{-ag}\right) + ae^{-g(1+a)}}$

unslotted nonpersistant CSMA $s = \frac{ge^{-ag}}{g(1+2a) + e^{-ag}}$

slotted nonpersistant CSMA $s = \frac{age^{-ag}}{1 - e^{-ag} + a}$

Comments I

Auteur Xavier Lagrange
Le mécanisme Aloha a été mis en oeuvre dans les années 70 au sein de l'université d'Hawaï par Norman Abramson. L'université était en effet répartie sur plusieurs iles et il fallait développer des moyens d'échange d'information entre les sites à faible coût. C'est le système le plus simple qu'on puisse imaginer : chaque station est équipée d'un émetteur-récepteur et fonctionne sur la même fréquence. Si une station veut transmettre une trame d'information, elle la transmet directement (sans faire aucun test préalable). Il est donc possible que 2 stations transmettent en même temps et il y a dans ce cas une collision entre les trames : aucune trame n'est correctement reçue.
A partir de ce processus très simple, des variantes ont été imaginées.
Présentation du modèle
On considère un ensemble de stations qui transmettent entre elles des trames d'information. Ces trames sont générées, par exemple, par une application. On suppose que le processus de génération est un processus aléatoire de Poisson. En d'autres termes, la durée entre 2 trames sucessives est une loi exponentielle. On suppose que le nombre de stations est très grand et peut donc être considéré comme infini. De ce fait, le processus de génération des trames n'est pas modifié par l'état du système : la probabilité que dans les δt secondes suivantes, une trame soit générée est la même quel que soit le nombre de transmissions en cours. On peut donc considérer que le processus de génération de trames pour l'ensemble du système est un processus de Poisson.
Toute trame envoyée alors que le support n'est pas occupé par une autre station (et ce pendant toute la durée de la trame) est considérée comme bien transmise. Dans tout autre cas la trame est considérée comme brouillée dans sa totalité et non reçue par les stations. Par exemple, si une trame commence à être transmise alors qu'une autre trame est en cours de transmission, aucune des 2 trames n'est reçue même si le chevauchement des 2 trames a lieu pendant un temps très court. On suppose enfin que toutes les trames ont une même durée constante.
On définit les paramètres suivants : T la durée des trames et λ le nombre moyen de trames générées par seconde pour l'ensemble de la population.
Aloha non synchronisé sans retransmission
On considère tout d'abord le cas où les stations émettent immédiatement les trames générées et ne les répètent pas en cas d'échec. Nous cherchons à calculer le nombre moyen s de trames correctement transmises (i.e. sans

Comments II

collision) par unité de temps. Pour cela, nous allons calculer de 2 manières différentes la probabilité de succès P_{succ} de transmission d'une trame.
Soit g le nombre de trames émises sur le support de transmission. On a dans le cas présent $g = \lambda$.
Du fait des hypothèses considérées, l'émission des trames respecte un processus de Poisson. Soit $P_k(\delta t)$ la probabilité d'avoir exactement k trames émises pendant une durée δt. Pour un processus de Poisson de paramètre g, on a (formule supposée connue) :

$$P_k(\delta t) = \frac{(g\delta t)^k}{k!} e^{-g\delta t} \quad (1)$$

Une trame est brouillée lorsqu'il y a émission simultanée de plusieurs stations pendant un temps, aussi petit soit-il. En considérant le schéma 1, on observe que si une trame commence à être émise à t, il faut qu'il n'y ait aucune autre émission pendant $[t - T, t + T]$, soit pendant une durée $2T$. Cette durée est appelée période de vulnérabilité.

Comments III

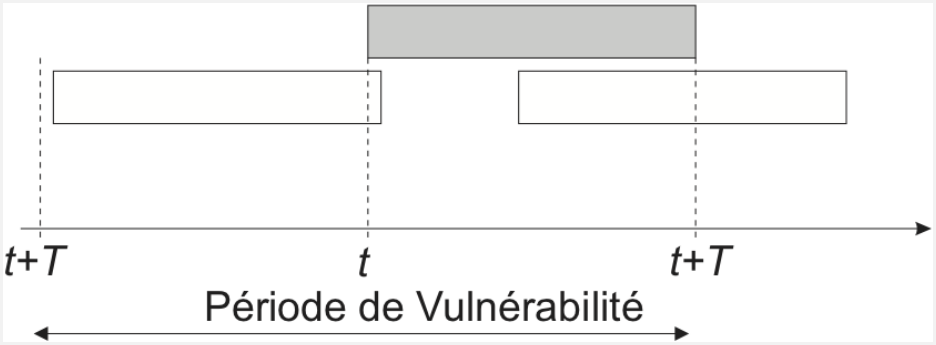


Figure: fenêtre de collision dans le cas de l'Aloha pur et période de vulnérabilité

Comments IV

La probabilité P_{succ} est égale la probabilité qu'il n'y ait aucune autre émission pendant la durée de $t - T$ à $t + T$ sachant qu'il y a une transmission à t. Avec les hypothèses de population infinie, on sait que la probabilité qu'il ait d'autres émissions ne dépend pas du fait qu'il y ait une émission ou non en cours. La probabilité P_{succ} est donc égale à la probabilité qu'il n'y ait aucune émission pendant $2T$. On peut donc appliquer la formule 1 pour $k = 0$ et $\delta t = 2T$:

$$P_{succ} = e^{-2gT} \tag{2}$$

Prenons maintenant le point de vue d'un observateur qui cherche à mesurer la probabilité de succès. Il compte le nombre de trames bien transmises pendant une durée T_{obs} et le divise par le nombre total de trames transmises, soit sT_{obs}/gT_{obs}. On a donc

$$P_{succ} = s/g \tag{3}$$

En combinant les équations 2 et 3, on en déduit

$$s = ge^{-2gT} \tag{4}$$

Etant donné que la durée des trames est constante, il est très utile de normaliser les paramètres s et g par cette durée T. On pose donc $G = gT$ et $S = sT$. Comme s représente le nombre de trames bien transmises par seconde, $1/s$ représente la durée moyenne entre deux débuts (ou deux fins) de trames bien transmises. Pendant cette durée moyenne $1/s$, on transmet correctement une trame donc la durée moyenne d'occupation efficace du canal est T. Comme on peut écrire S sous la forme $S = T/(1/s)$, on en déduit d'une part que S est nécessairement inférieur à 1 et d'autre part que S représente le pourcentage d'occupation efficace du canal.
On peut réécrire l'équation 4 sous la forme

$$S = Ge^{-2G} \tag{5}$$

Comments V

Prise en compte des retransmissions
Les stations qui émettent des trames désirent qu'elles soient bien reçues. Supposons qu'elles aient un moyen de savoir après l'émission d'une trame si elle est effectivement bien reçue (par exemple par un mécanisme d'acquittement envoyé par le destinataire).
Si une station a émis une trame qui a été brouillée, elle garde la trame en mémoire et passe dans un état appelé backlog. Elle lance une temporisation aléatoire (suivant une loi exponentielle) puis répète l'émission de la trame après une temporisation aléatoire si celle-ci a été mal reçue. Dans ce cas le paramètre représente le nombre de trames nouvelles générées et donc le nombre d'entrées dans le système. Il est appelé quelquefois trafic frais. Le paramètre g représente le nombre de trames émises sur le canal qui est égal au nombre de nouvelle trames générées plus le nombre de trames réémises après un échec. Il est appelé charge offerte au système ou offered load. Le paramètre s représente le nombre de trames émises avec succès et donc le nombre de sorties du système. On a donc $g > \lambda$.
Le schéma 2 illustre le système. A l'équilibre du système on a $s = \lambda$. En considérant que la charge offerte est un processus de Poisson, le raisonnement précédant reste applicable et on peut écrire : $S = G.e^{-2G}$

Comments VI

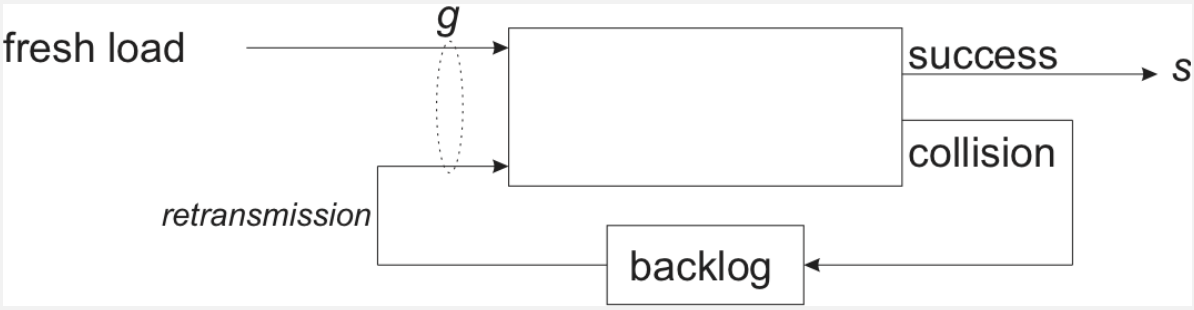


Figure: Mécanisme de retransmission en Aloha

L'Aloha discrétisé
L'Aloha discrétisé ou slotted Aloha repose sur les mêmes hypothèses que l'Aloha pur mais en discétisant le temps. On définit une base de temps commune à toutes les stations, qui peut être transmis par exemple par une station de base. Sa période est égale à la durée de trame.
Dans ce cas la période de vulnérabilité est réduite de $2T$ à T. On peut donc écrire : $P_{succ} = e^{-gT}$. Donc $S = Ge^{-G}$.
Avec l'Aloha pur, le débit utile normalisé maximal est de 0,18 ($1/2e$) pour une valeur de G de 0,5. Avec l'Aloha synchronisé, il est de 0,36 pour une valeur de G e 1.

Ethernet

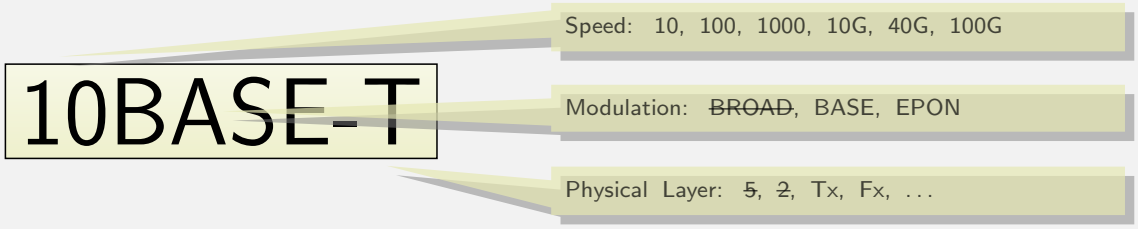
Introduction

History and Evolutions

- Developed initially by Digital, Intel and Xerox to share first laser printers
- First version in 1 976 at 3 Mbit/s
- Second version standardized by IEEE at 10 Mbit/s:
  - Subtle changes in frame field leads to two way to manage protocol stack.
  - Frame format slightly changed but physical medium evolved:
    - from 10 Mbit/s to 100 MBit/s them 1GBit/s and 10 or 40 GBit/s.
    - standardization is currently working on 100 GBit/s links.
    - From coaxial link to twisted pair and optical fiber (No support for radio link)
  - MAC get simplified:
    - Shared media (coaxial or twisted pair with hub) uses CSMA/CD algorithm
    - Switched media or point to point links do not require MAC algorithm (there is a dedicated resource between the equipment and the switch).

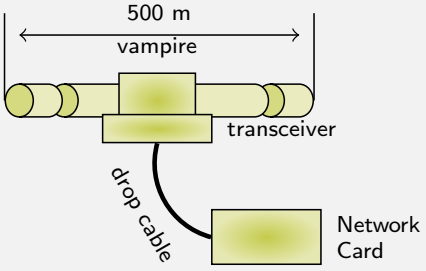
IEEE at 10 Mbit/s

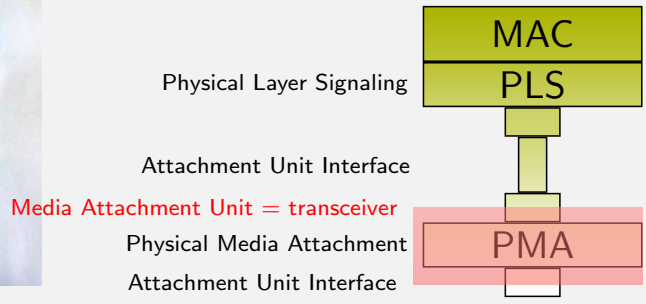
Technology names: 10BASE-5 & 10BASE-2



Obsoleted technologies:

- 10BASE-5 (*Thick Ethernet*) and 10BASE-2 (*Thin Ethernet*)
  - Based on co-axial





http://en.wikipedia.org/wiki/10BASE5

Comments I

Le protocole IEEE 802.3 est mis en œuvre sur plusieurs types de supports physiques. Chaque support est désigné par un code qui indique :

- sa vitesse de transmission en Mbit/s (1, 10, 100 ou 1 000),
- son procédé de codage:
  - BASE: bande de base, l'ensemble du spectre est utilisé pour coder des signaux Ethernet, c'est le cas avec un codage où des tensions électriques servent à coder le signal binaire.
  - BROAD : Broadband ou modulation cette technologie est maintenant abandonné, elle permettait de faire cohabiter le signal Ethernet avec d'autres réseaux en n'utilisant qu'une gamme de fréquences.
  - EPON (Ethernet Passive Optical Network): utilisée par les opérateurs pour transporter des données vers les habitations en replacement de l'ADSL.
- un chiffre indiquant la longueur maximale d'un segment pour les câbles coaxiaux ou une lettre donnant le type du support (T pour la paire torsadée et F pour la fibre optique).

Comments II

A l'origine les réseaux étaient conçus au-dessus d'un câble coaxial (cf. transparent page 129). Dans ce mode tous les équipements partagent le même support physique. Il faut mettre en œuvre un mécanisme d'exclusion mutuel pour qu'à un instant donné, un seul équipement émette des données. En conséquence, quand un équipement émet des données, les autres équipements ne peuvent rien émettre. Ce mode de fonctionnement est dit half-duplex.

La norme IEEE 802.3 définit plusieurs interfaces représentées page 129. Cette figure peut sembler complexe et le nombre d'interfaces relativement important. En fait, toutes ces interfaces ne se retrouvent pas dans tous les équipements. Leur rôle est de rendre indépendantes les fonctions liées à la transmission et celles liées à la commutation.

La plus connue et la plus anciennement définie est l'interface AUI (*Attachment Unit Interface*) dont la partie visible est le connecteur 15 broches femelle que l'on retrouve sur certaines cartes réseaux. Cette interface rend la couche physique indépendante du médium utilisé (câbles coaxiaux, paire téléphonique, fibre optique...).

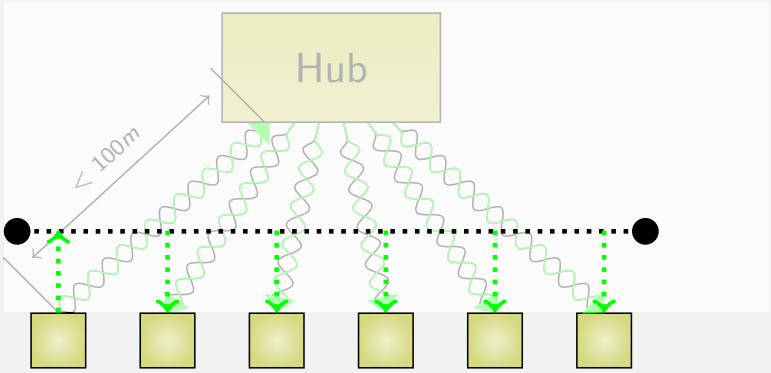
Un transceiver aussi appelé, pour cette technologie MAU (*Medium Attachment Unit*) ou génériquement pour toutes les solutions PMA (*Physical Media Attachment*), effectue la conversion des signaux pour les adapter au support physique. En plus de l'adaptation au support de transmission, un transceiver contrôle la durée d'émission de la station à laquelle il est connecté. En effet, si celle-ci, à la suite d'une erreur logicielle ou matérielle, se met à émettre en permanence toutes les autres stations connectées à un support partagé seront bloquées. Le mécanisme du jabber autorise des émissions comprises entre 20 et 150 ms. L'interface MDI (*Media Dependent Device*) est le nom générique des prises qui permettent de connecter le transceiver au support de transmission. Il peut s'agit d'une prise BNC dans le cas d'un réseau 10BASE2, d'une RJ-45 dans le cas d'un réseau sur paires torsadées ou des connecteurs pour la fibre optique.

Comments III

La définition des sous-couches a évolué avec les différentes versions de la norme comme le montrent les autres piles protocolaires de la figure 4.3. Avec l'apparition des hubs et des commutateurs, l'interface AUI a montré ses limites. La première est d'ordre mécanique. Il était relativement difficile de connecter les équipements avec la prise AUI. De plus la communication entre l'équipement et le transceiver se fait en utilisant qu'une paire dans chaque sens et un codage Manchester, ce qui limite l'évolution vers des débits plus élevés.

Le manque de souplesse du câblage et sa spécificité font que les câblages co-axiaux sont maintenant abandonnée, car en plus de leur manque de souplesse, ce câblage est dédié à un type de réseau et ne peux évoluer ou être utilisé par une autre technologie.

10BASE-T



- Hub: Active equipment repeating signal on other ports
- Generate collision signal when two equipments send at the same time
- Hub is viewed as a shared network by equipments.
  - More reliable than co-axial cable. Cable cut just isolates one equipment.
  - Use standard wires (telephony wires for 10BASE-T)

Comments I

Les réseaux 10BASE-T (T pour Twisted pair : paire torsadée) ne sont pas bâtis sur une topologie en bus. Il s?agit d?une topologie en étoile nécessitant un équipement actif appelé hub (moyeu en anglais) qui émule un bus. Initialement ce réseau fonctionne sur paire téléphonique standard à 10 Mbit/s. Cela permet d?utiliser le câblage téléphonique préexistant (donc d?un coût faible). Le câblage est réalisé par une hiérarchie de grappes étoilées.

Il est intéressant de noter que ce type de câblage s'éloigne des règles initiales des réseaux locaux stipulant qu'il ne faut pas dépendre d'un équipement actif dont la panne pourrait paralyser le réseau.. En effet, le bon fonctionnement du réseau va dépendre d'un élément central actif. Malgré tout, la fiabilité des composants électroniques est moins à remettre en cause que la coupure accidentelle du bus. Cette topologie est de plus en plus répandue car elle permet d'utiliser des précâblages existants (téléphonique, anneau à jeton...). De plus les nouveaux protocoles de réseaux à haut-débit seront basés sur l'étoile (Ethernet à 100 Mbit/s, ATM...). Les équipements actifs présentent aussi l'intérêt de pouvoir être gérés à distance (activation/désactivation d'un port, configuration, collecte de statistiques, etc.).

Généralement pour les précâblages, des câbles comprenant quatre paires torsadées sont posés. Seulement deux paires sont nécessaires pour relier une station au hub : une paire dédiée à l'émission et une à la réception. Les deux autres paires sont inutilisées. Elles pourront servir pour des transmissions à un débit supérieur. Cette topologie initialement prévue pour émuler un bus à 10 Mbit/s a permis l?évolution de la norme vers trois techniques complémentaires :

- l'augmentation des débits pour atteindre 100 Mbit/s. Cette augmentation passe principalement par un câblage de catégorie 5 adapté à ces vitesses de transmission (100BASE-TX) et plus marginalement par des techniques de codage adaptées(100BASE-T2) ou par l'utilisation en half-duplex des quatre paires tordasées pour le (100BASE-T4) ;

Comments II

- le remplacement des mécanismes d'émulation d'un bus par les hubs par un mécanisme de commutation, permettant la transmission simultanée de plusieurs trames ;
- la création de réseaux virtuels permettant une réelle séparation des trafics transitant sur la même infrastructure

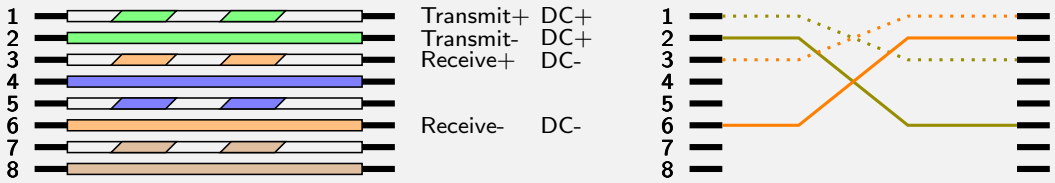
Les hubs effectuent la concentration et la retransmission des messages qu'ils reçoivent d'une paire vers toutes les autres paires. Quand deux stations émettent en même temps, le hub génère des signaux de collision vers tous les équipements. Ainsi, pour les stations, la méthode d'accès au support est identique à celle sur un bus. Le hub peut être considéré comme un répéteur ayant plusieurs ports d'entrée/sortie.

Il n'est pas possible sur une liaison point-à-point d'avoir des conflits d'accès. Chaque sens se voit attribuer une paire torsadée. Les premiers équipements d'interconnexion (appelés hubs) ne faisaient que répéter les signaux reçus sur un port sur les autres ports et produire un signal ressemblant à une collision si deux ports sont activés simultanément. Les conflits d'accès doivent toujours être arbitrés par les équipements terminaux en utilisant l'algorithme du CSMA/CD.
Cette topologie en étoile à été le point de départ d'une autre évolution. Le mode d'accès a pu être modifié et permet aux deux extrémités d'émettre et de recevoir des données simultanément. Ce mode est appelé full-duplex et les équipements d'interconnexion sont appelés des commutateurs ou dans la terminologie la anglaise switches.

Il faut bien noter que le fonctionnement en mode half-duplex ou full-duplex ne dépendent pas uniquement de la topologie du réseau, mais aussi des fonctionnalités mises en œuvre dans les équipements terminaux et ceux d'interconnexion. Dans le standard, tous les supports permettent un fonctionnement en half-duplex, mais avec l'évolution des débits (100 ou 1 000 Mbit/s), les inconvénients liés à la méthode d'accès font que le full-duplex est majoritairement utilisé.

RJ-45

- RJ-45 plug and cable contains 4 twisted pairs
  - 10M and 100M generally use only 2 pairs, 1G uses 4 pairs
- Modulation signal is the difference between both wires
  - PoE (Power on Ethernet) (IEEE 802.3at): continuous voltage is added to data or use the blue and brown pairs
- Straight cable to connect equipment to switch/hub, cross cable to connect equipment/switch/hub together
  - equipments may select to the appropriate cabling.



Straight Cable | Cross Cable

http://www.ertyu.org/steven_nikkel/ethernetcables.html

Comments I

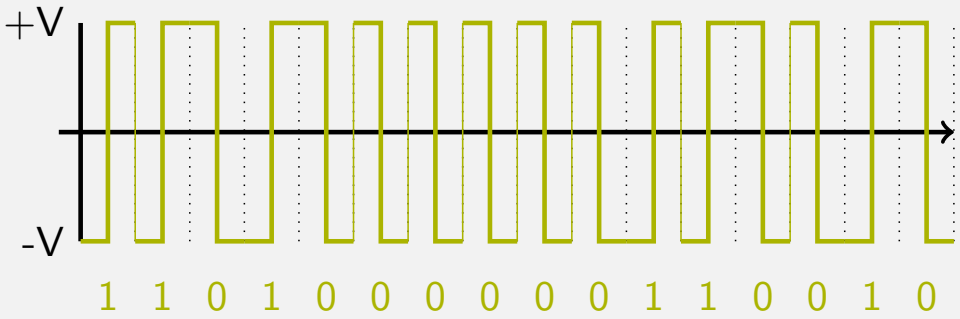
La prise RJ-45 est aussi une interface communément rencontrée dans les réseaux Ethernet. Le câble se termine aux deux extrémités par une prise RJ-45 dont le brochage est donné figure 4.4. A noter que certaines solutions comme le 100BASE-T4 ou le 1GBASE-TX utilisent les quatre paires. Des fonctions d'auto-négociation sont prévues dans les équipements pour éviter toute confusion et trouver le mode de fonctionnement optimal.

Il est également possible d'alimenter des équipements distants via ce câblage. La norme IEEE 802.3af définit deux modes de fonctionnements. le plus simple consiste à utiliser les paires non utilisées pour transporter les données Ethernet pour envoyer une tension continue. La seconde qui est compatible avec les technologies qui requièrent l'utilisation des quatre paires (comme le 1GBASE-TX) ajoute une composante continue aux signaux. Cette composant, appelée courant fantôme, est neutre pour le signal binaire car le codage se fait de manière différentiel.

Le câble reliant un hub à une station est droit. Le câble reliant deux hubs doit avoir ses paires de fils croisées, sinon les paires en émission et en réception se retrouvent face à face.

Coding technique at 10 Mbit/s

- Manchester coding:
  - transition $0 \rightarrow 1 = 1$
  - transition $1 \rightarrow 0 = 0$
- Each bit sent allows clock synchronization by the receiver
  - Cannot be used a 100 Mbit/s and higher due to the high number of transitions



http://en.wikipedia.org/wiki/Manchester_code

Comments I

Dès que le débit devient plus élevé, il est souhaitable d'introduire des transitions fréquentes dans le signal transmis. Les codages biphasés (ou Manchester) utilisent ce principe en forçant une transition (alternance) au milieu de chaque période binaire. Les transitions en milieu de bit servent à la synchronisation des horloges. Dans le codage Manchester normal, une transition de l'état haut vers l'état bas code un bit à 0. Une transition de l'état bas vers l'état haut code un bit à 1. Le réseau Ethernet utilise ce type de code.

Les avantages du codage Manchester sont les suivants :

- introduction de transitions qui permettent à l'horloge du récepteur de se synchroniser sur celle de l?émetteur ; les câbles de transmission peuvent aussi servir à transporter un courant continu d'alimentation des équipements du réseau (voir les courants fantômes utilisés en Power over Ethernet (PoE)) ;
- pas de composante électrique continue qui peut provoquer des claquages des composants électroniques si les 1 ou les 0 prédominent ;
- présence de deux symboles J et K autorisant la signalisation des trames. Les symboles J et K sont obtenus en violant les règles précédentes et en maintenant le même état pendant deux demi-périodes. Elle ne sont pas utilisées par Ethernet, mais on retrouvait leur usage dans d'autres technologies, comme l'anneau à jeton pour signaler le début et la fin d'une trame.
- détection facile des erreurs de transmission quand aucun changement de polarité n'est détecté pendant plus d'une période.

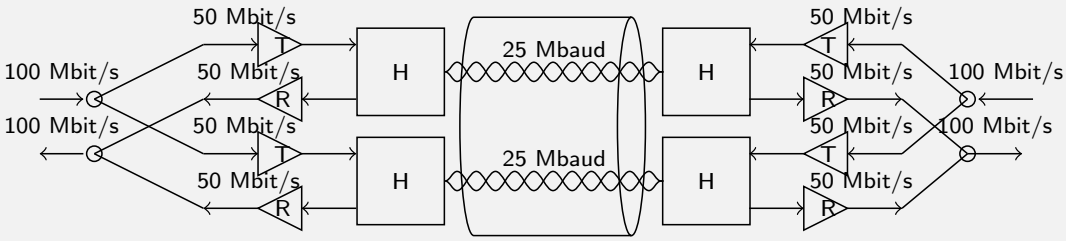
L'inconvénient majeur du codage Manchester est d'utiliser une plus grande bande passante puisque la fréquence de modulation du codage (en bauds) est le double de la vitesse de transmission (en bit/s). Les réseaux plus récents, comme Ethernet à 100 Mbit/s, utilisent des codages de type 4B/5B où 4 bits d'information utile sont codés sur le support à l'aide de 5 bits. Ce qui permet d'introduire au moins une transition tous les 4 à 5 bits transmis. Suivant cette désignation le codage Manchester serait un codage 1B/2B.

Ethernet

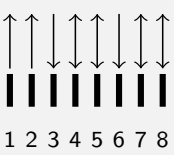
100 Mbit/s

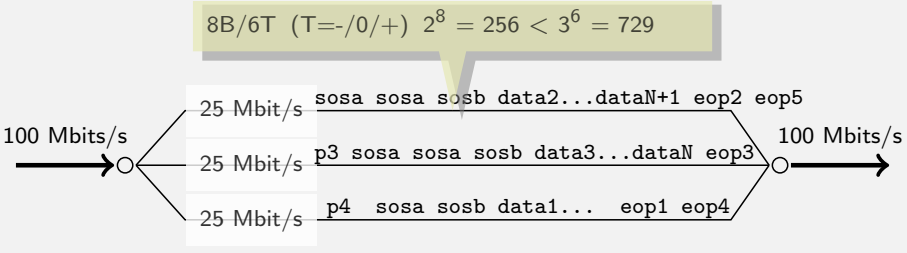
Ethernet 100 Mbit/s on Twisted Pairs

- Only based on star topologies (hub or switch);
- 100BASE-T2 (cat 3)



- 100BASE-T4 (cat 3) half-duplex





Comments I

Le 100BASE-T4 est une solution économique pour permettre une montée en débit sur un câblage de catégorie 3 ou supérieure. Cette solution ne peut être mise en œuvre que si les quatre paires généralement attribuées à une prise RJ-45 sont disponibles. Le mode de transmission est forcément half-duplex. La figure page 141 donne le câblage des prises RJ-45. Les paires 1-2 et 3-6 sont connectées de la même manière, elles restent spécialisées dans l'émission ou la réception de données. Les deux autres paires (4-5 et 7-8) servent soit pour l'émission, soit pour la réception. Les paires Tx et Rx permettent de détecter l'émission simultanée des deux extrémités. L'algorithme du CSMA/CD est utilisé pour arbitrer les conflits.

A un instant donné, trois paires sont utilisées soit en réception soit en émission. Sur chaque paire, le débit nécessaire est limité à 33,333 Mbit/s. Un codage 8B/6T permettant de coder 8 bits sur 6 symboles et 3 niveaux de signaux (+, 0 et -) permettent de réduire la vitesse de modulation à 25 Mbaud. Ainsi par exemple :

- 0 est codé par la séquence +-00+-,
- 1 est codé par la séquence 0+-+-0,
- ...

Le codage 8B/6T autorise aussi des séquences permettant la signalisation :

- SOSA (+-+-+-) et SOSB (+-+−+) codent le préambule de la trame,
- P3 (+-) décale l?émission sur la deuxième paire de deux temps bits,
- P4 (+-+-) permet de décaler l?émission sur la troisième paire de quatre temps bits,
- EOP_1 (++++++), EOP_2 (++++−), EOP_3 (++−00), EOP_4 (——) et EOP_5 (−00000) codent la fin de transmission de la trame.

Chaque octet est transmis alternativement sur chaque paire. Pour éviter un brouillage trop important l'émission des octets est, pour chaque paire, décalée dans le temps. La figure précédente donne l'ordre d'émission des octets d'une trame.

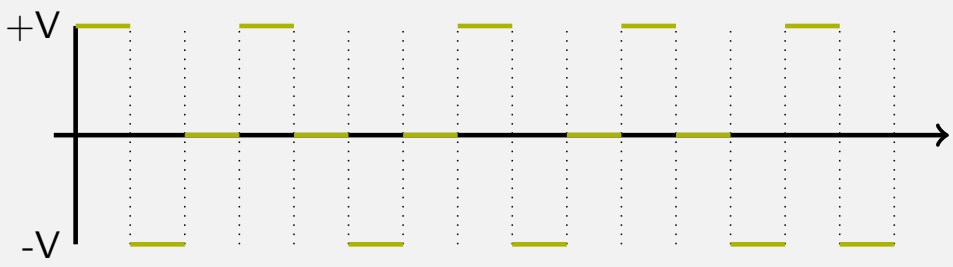
100Base-TX coding

Use 4B/5B coding (4 bits are coded into 5 bits).

Binary values		5B	Binary values		5B	Signal	5B	Meaning
0	0000	11110	8	1000	10010	Q	00000	Quiet (signal lost)
1	0001	01001	9	1001	10011	I	11111	Idle
2	0010	10100	A	1010	10110	J	11000	Start #1
3	0011	10101	B	1011	10111	K	10001	Start #2
4	0100	01010	C	1100	11010	T	01101	End
5	0101	01011	D	1101	11011	R	00111	Reset
6	0110	01110	E	1110	11100	S	11001	Set
7	0111	01111	F	1111	11101	H	00100	Halt

Coding uses MLT-3

- 1 are alternatively coded by +V and -V, 0 are coded with 0V



Which value is sent ?

Comments I

Le 100BASE-TX utilise deux paires câblées de la même manière que le 10BASE-T. Le codage au niveau physique est le même que celui-spécifié par la norme FDDI sur paires torsadées. Le codage des informations binaires s?effectue en deux phases. Dans un premier temps, le codage 4B/5B est utilisé, il transforme un groupe de 4 bits en une séquence de 5 bits comme indiqué tableau précédent. Le débit effectif sur le support de transmission est donc de 125 Mbit/s.
En plus des 16 séquences codant les données binaires, le codage 4B/5B autorise des séquences de signalisation :

- les symboles /J/ et /K/ servent pour coder le premier octet du préambule (SSD : Start of Stream Delimitor) ;
- les symboles /T/ et /R/ servent à indiquer la fin de la trame (ESD : End of Stream Delimitor) ;
- le symbole /I/ est utilisé sur la liaison point-à-point quand aucune trame n?est transmise ;
- le symbole /H/ est utilisé pour indiquer une erreur de transmission.

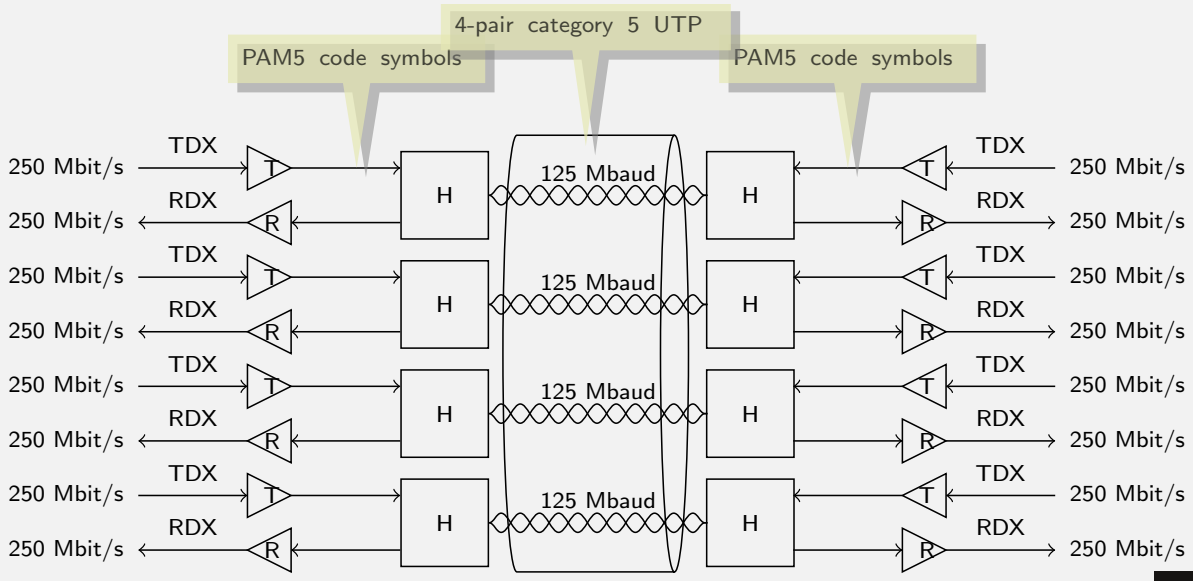
Les séquences non représentées tableau 4.5 sont invalides. La réception d'une séquence invalide ou du symbole /H/ est interprétée comme une collision.
La figure 4.25 représente le codage utilisé pour transmettre une trame sur le support physique. Les données arrivent par bloc de 4 bits par l'interface MII. Le premier octet du champ préambule est remplacé par la séquence /J/K/. Le reste de la trame est interprété comme une séquence binaire et est codé comme indiqué figure 4.5. En fin de trame, une séquence /T/R/ est insérée pour indiquer la fin de la transmission. En 100BASE-TX, les données sont transmises en utilisant le codage MLT-3 (Multi-Level Transition-3). Ce codage utilise 3 états -1, 0 et 1. Chaque valeur binaires à 1 est transmise par un changement d'état. Les états successif sont -1, 0, +1, 0, -1... Une valeur binaire à 0 en restant dans le même état que précédemment. La figure 4.26 donne un exemple de codage. Ce codage peut sembler inefficace pour maintenir la synchronisation puisqu'une longue séquence de bits à 0 maintiendrait le signal dans le même état. Ceci n'est pas possible car les valeurs binaires sont précédemment codées en utilisant le codage 4B/5B.

Ethernet 100 Mbit/s on Optical Fiber

- 100BASE-FX
  - Multimode Optical Fiber
  - up to 2 km in switched full duplex mode
  - limited to 400 m in half duplex CSMA/CD mode
- 100BASE-SX
  - Cheaper lasers, less distance, about 550 m.
- 100BASE-LX
  - Up to 10 km
  - monomode fiber
- ...

Ethernet 1 Gbit/s

- 1000BASE-T (IEEE 802.3ab)
  - category 5 cable and over all 4 pairs
  - up to 100 m



1000BASE-X

- 1000BASE-TX
  - Use two twisted pairs of cat6 cable, using 8B/10B coding
  - less popular than 1000BASE-T widely deployed on computers
- 1000BASE-CX (IEEE 802.3z)
  - initial standard for twisted pair on a max distance of 25 m, replaced by 1000BASE-T
- 1000BASE-LX
  - Multi-mode optical fiber, up to 550
- 1000BASE-ZX
  - Single mode optical fiber, up to 70 km
- ...

Comments I

Le support visé pour l'Ethernet à 1 Gigabit/s est principalement la fibre optique. Le comité IEEE 802.3ab a défini la possibilité d'utiliser les quatre paires de catégorie 5 pour transmettre à 1 Gbit/s sur 100 mètres. Comme pour le 100BASE-T2, l'utilisation d'un tel câblage est problématique. Les phénomènes d'écho au niveau des connecteurs sont aussi plus importants, ce qui implique de placer dans les cartes des équipements d'annulation d'écho augmentant de ce fait le coût de l'installation.
Par contre si un câblage de catégorie 6 est disponible, la norme 1000BASE-TX peut être utilisée, et il est possible de connecter deux équipements distants de 100 m en n'utilisant que 2 paires torsadées. Mais cette solution a du mal a s'imposer face au 1000BASE-T qui utilise des câbles de catégorie 5 relativement répandus dans les bâtiments et facile a manier. De plus le faible coût des adaptateurs 1000BASE-T n'incite pas à utiliser 1000BASE-TX.

Ethernet

CSMA/CD

Carrier Sense Multiple Access / Collision Detect

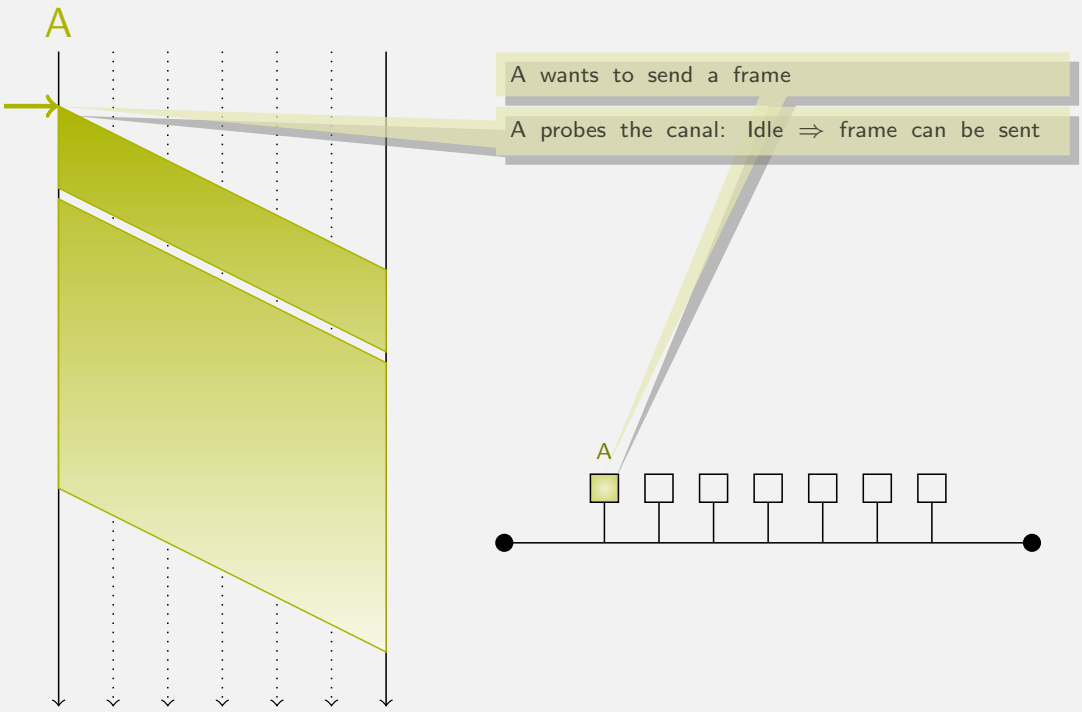
- Carrier Sense: Listen to the channel to detect the status
  - Idle: Not transmission
  - Busy: Regular transmission is occuring
  - Collision: Several equipments are sending at the same time.
    - Standard impose that signal level cannot be reduce by more than half in the whole link
    - If reception level is higher than the standardized transmission level, then at least two equipments are sending.
    - This not work on radio network, reception level is much smaller than transmission lever and some station may be hidden
  - if the channel is Idle, then send the frame
  - if the channel is busy, then delay the transmission
- Collision Detect: If a collision occurs during the frame transmission: stop and restart.
- Impossible to detect on radio links

Comments I

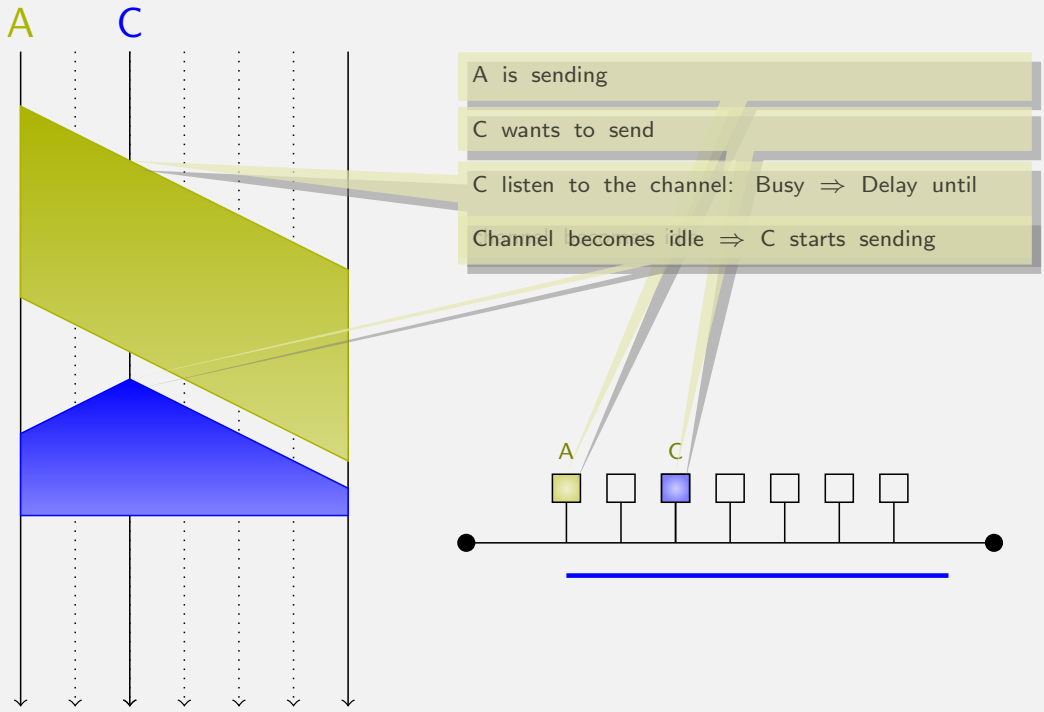
Avant de voir plus en détail les équipements qui permettent de concevoir un réseau Ethernet, il est intéressant de comprendre le fonctionnement de l'algorithme du CSMA/CD, car cet algorithme basé sur un délai de propagation maximum des signaux, aura pour conséquence de limiter le nombre d'équipements d?interconnexion et d'imposer des règles de câblages relativement strictes. Le protocole d'accès au médium repose sur deux principes :

- CSMA (Carrier Sense Multiple Access) (accès multiple après écoute de la porteuse). Cette méthode permet de réduire le nombre de collisions. Avant d'émettre la station écoute le canal. Si celui-ci est libre, elle émet son message, sinon elle diffère son émission. Mais, à cause des délais de propagation, tout risque de collisions n'est pas supprimé, comme le montre le schéma slide page 155 ;
- CD (Collision Detect) (détection de collisions). Si une station émettrice se rend compte que son message participe à une collision, elle arrête l'émission du message.

CSMA/CD



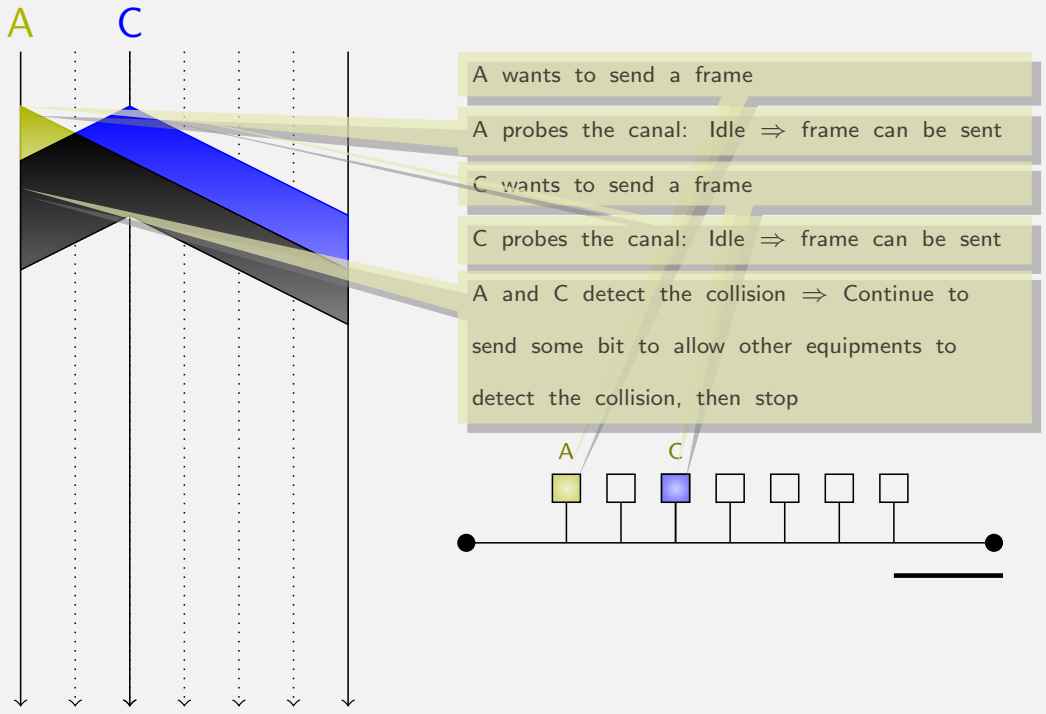
CSMA/CD: Delayed transmission



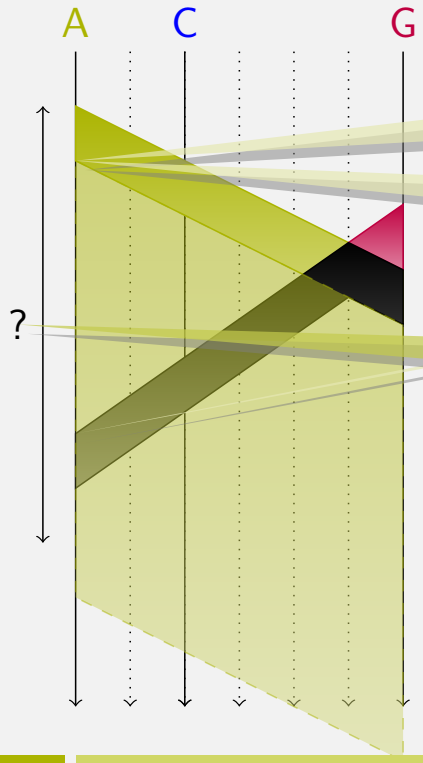
Comments I

Globalement, les stations émettent des messages quand elles le désirent. De ce fait, une station dispose entièrement du canal si les autres stations n'ont rien à émettre. Cette propriété s'appelle transparence de canal. Mais, à cause de cet algorithme d'accès aléatoire au canal, des problèmes peuvent survenir. On appelle collision le fait que deux ou plusieurs stations émettent un message en même temps. Ceci peut être dû à l'émission simultanée de deux ou plusieurs équipements.
Une autre cause est due à la synchronisation des émissions en attendant que le canal se libère. La figure 4.6 illustre ce phénomène. Les stations A et B ont leur émission retardée à cause de la transmission d'une trame par une autre station. Quand celle-ci va se terminer, au délai de propagation près, les deux stations vont commencer à émettre simultanément leurs données provoquant une collision. L'algorithme du BEB (voir page 163) permet de départager les équipements en conflit.

CSMA/CD: Collision



CSMA/CD: Minimal frame size

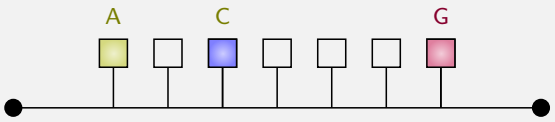


A send succefully its frame, even if some of the equipment on the link will not receive it due to collision

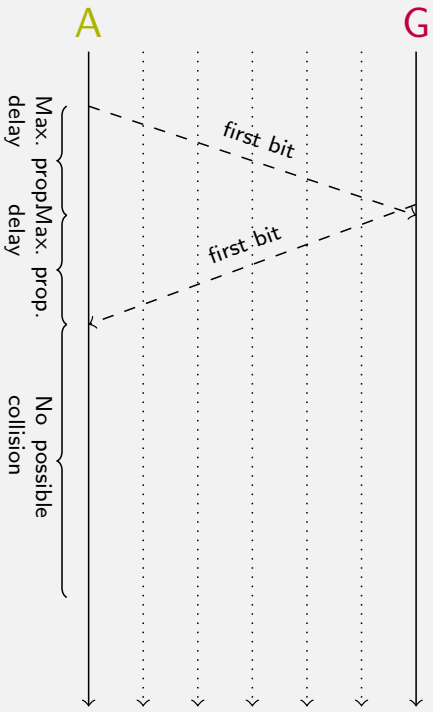
Solution: Artificially raise the frame size until this problem could append. This way A will be aware that its frame got a collision.

Collision detected during the transmission..

What is the length of these period ?



CSMA/CD: Minimal frame size computation

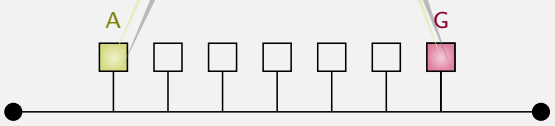


Take the worst case: Larger distance between equipments

$Min_{frame} = 2.MaxPropagationDelay$

At 10 Mbit/s standard gives:
$MaxPropagationDelay = 51.2\mu s$

$Min_{frame} = 64Bytes$

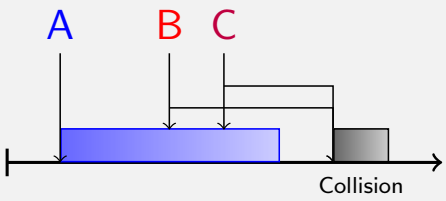


Comments I

On appelle Tranche Canal (TC) ou Time Slot la durée nécessaire à une application pour que celle-ci soit certaine que son message soit transmis sans problème. Cette période est au minimum égale à deux fois la durée maximale de propagation d'un message sur le câble. Ceci justifie les contraintes de câblage vues précédemment. Si l'on additionne tous les délais dans l'aller et le retour d'un signal introduit pour la transmission de celui-ci et si l'on considère les deux stations les plus éloignées, le calcul donne une durée maximale de propagation de 44,99 ms. Ainsi, pour un réseau Ethernet classique, la norme indique une valeur légèrement supérieure. La durée d?une tranche canal est équivalente à la durée d'émission de 512 bits soit 51,2 ms à 10 Mbit/s.

La durée d'émission des trames doit toujours être supérieure ou égale à la tranche canal. Pour un réseau à 10 Mbit/s, 51,2 ms correspondent à la durée d'émission d'une trame de 64 octets. Si le paquet est plus petit, des bits de bourrage (ou padding) sont introduits en fin de trame pour atteindre cette taille. Cette durée minimale a été introduite pour que toutes les stations soient dans le même état à la fin d'une transmission.
Le transparent page 157 montre un protocole dans lequel la durée d'émission est plus petite que le délai de propagation. La station `A` émet un message M1 et la station `G` émet un message M2 avant de pouvoir détecter l'émission de `A`. Comme la durée d'émission est inférieure à la durée de propagation, `A` n'a pas le temps, pendant l'émission, de découvrir que ce message connaîtra une collision. Elle va donc transmettre au niveau supérieur un compte-rendu positif, indiquant que la transmission s'est bien déroulée. De plus on s'aperçoit que la station `C` a reçu correctement la trame M1 mais pas la trame M2, la station S6 a reçu correctement la trame M2 mais pas la trame M1.
Quand une collision est détectée par une station, celle-ci n'interrompt pas immédiatement la transmission mais continue à émettre des données de brouillage (ou jamming) pour permettre la détection de la collision par les autres stations. La taille du brouillage est de 32 bits. L'émission d'une trame en collision peut donc durer moins d'une tranche canal. Si les règles de câblage ne sont pas respectées, une collision peut être détectée après cette durée de 51,2 ms. Ce comportement fautif du réseau est appelé collision tardive (late collision).

Maximum Frame size



- A senses the channel: Idle ⇒ send frame
- B senses the channel: Busy ⇒ wait until channel becomes Idle
- C senses the channel: Busy ⇒ wait until channel becomes Idle
- A and C send frame at the same time ⇒ collision

Maximum Frame Size

Standard fixes the $Max_{frame} = 1\ 500\ Bytes$.

Frame Length

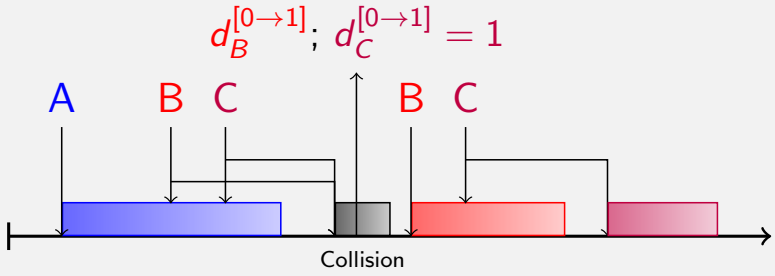
- Minimum 64 Bytes is given by CSMA/CD propagation delays
  - All versions of the standard impose this limit.
- Maximum 1 500 Bytes is given by the collision propability
  - This value is very small
  - If CSMA/CD is disabled, this value can be up to 9 Kilobyte
    - Jumboframe
  - Can be use either:
    - to increase transmission speed (reduce copy from card to memory)
    - to allow VLAN by adding a value to give a virtual network number
    - to encapsulate Ethernet frame into another Ethernet frame (used by Metro Ethernet)

Comments I

La taille maximale d'une trame est de 1 518 octets (1 500 octets de données, 14 octets d'en-tête et 4 octets de CRC) pour éviter qu'une station monopolise le canal. Cette taille est fixée arbitrairement, mais une taille plus importante aurait pour conséquence d'augmenter le risque de collision.

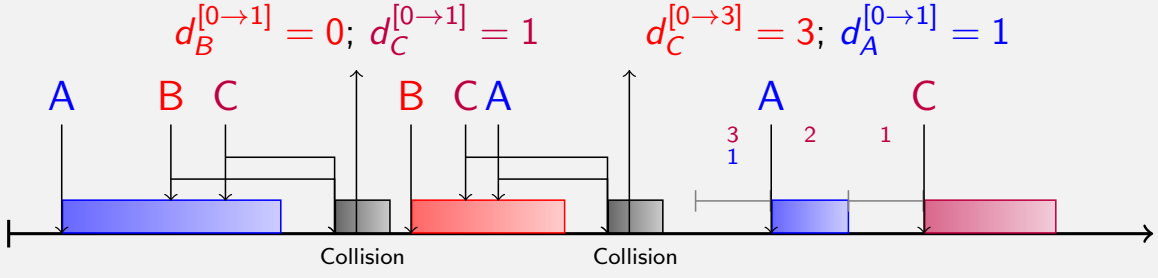
Binary Exponential Backoff

- Previous slides shows that collisions can occur after the busy period.
- Each equipment must have a different behavior:
  - Otherwise a new collision is generated
- Based on random number
  - First attempt: draw a random number d between [0, 1];
  - Wait $51.2 \cdot d$ μs.



Binary Exponential Backoff (continued)

- In the previous example, if both equipment draws a different random number, the collision is solved
  - if both select the same number, the collision occurs again.
- The space double every attempt until the10^{th} then remains the same until the 16^{th}
  - If after 16 attempts the frame is not sent, the transmission is aborded.
  - No guaranty (and not guaranteed delay) to send a frame.



Comments I

On appelle BEB (Binary Exponential Backoff), l'algorithme qui permet de limiter la charge du réseau quand une collision se produit. Lors d'une collision, les stations impliquées arrêtent leur émission après que celle-ci ait duré une tranche canal. Il reste à définir ce que font les stations après la collision. Si celles-ci recommencent à émettre aussitôt après, une autre collision se produira et ainsi de suite ; plus aucun message ne sera émis sur le support. Il faut trouver un mécanisme pour départager les stations (bien entendu sans échanger de message !) Il faut aussi s'arranger pour limiter l'accès au support physique en cas de congestion. Les résolutions des collisions, plus l'arrivée des nouveaux messages, risquent d'induire de nouvelles collisions qui vont encore plus limiter la bande passante utile, ce qui va entraîner de nouvelles collisions, et ainsi de suite en s'amplifiant. L'algorithme du BEB permet aux stations de tirer au sort la durée d'attente avant la prochaine tentative de réémission. Notons 0 et 1 les deux choix possibles. En supposant que deux stations seulement participent à la résolution de la collision, on trouve les quatre possibilités suivantes :

- la première station et la seconde station tirent 0 : les deux stations recommencent à émettre juste après la collision et reproduisent une collision ;
- la première station tire 0 et la seconde station tire 1 : la première station commence à émettre, au bout d'une tranche canal la seconde veut émettre, ainsi détectant une activité sur le médium, elle va attendre la fin de la transmission du message de la première station pour émettre son message. La collision est résolue ;
- la première station tire 1 et la seconde station tire 0 : le cas est identique au cas précédent. La collision est résolue ;
- la première station et la deuxième station tirent 1 : les deux stations vont attendre une tranche canal, puis vont simultanément tenter d?émettre leur message en produisant une nouvelle collision.

Comments II

Sur cet exemple on remarque qu'il existe une chance sur deux de résoudre la collision. Dans le cas où celle-ci n'est pas résolue (les deux stations ont tiré le même nombre où plus de deux stations sont en cause dans la collision), on double l'espace de tirage. Les stations pourront attendre 0, 1, 2 ou 3 tranches canal. Ce qui réduit à 1/4 la probabilité que deux stations émettent simultanément. De plus les stations vont attendre plus longtemps en moyenne pour émettre leur message, ce qui va réduire la charge sur le réseau. Par défaut, l'espace de tirage est doublé jusqu'à la dixième tentative. Si au bout de seize tentatives la trame n'est toujours pas émise, le protocole abandonne et informe la couche supérieure de l'échec.

CSMA/CD and 100 Mbit/s

At 10 Mbit/s, standard mandate at minimum duration of 51.2 μs

- or 64 Bytes

At 100 Mbit/s if the same duration is taken, the minimal frame size must be increased to 640 Bytes

- no gain: small frame sending time remains the same (51.2 μs)
- when bridging from 100BASE-TX to 10BASE-T, duration will be increased (512 μs)

So at 100 MBit/s, minimum duration is 5.12 μs

- Minimum size remain compatible with 10 Mbit/s technologies
- Network size must be decreased by a factor 10
  - 10BASE-5 allowed 2.5 km
  - 100BASE-TX allows 250 m
  - Compatible with cabling constraints

Comments I

La norme pour les réseaux à 10 Mbit/s précise que la durée d'émission minimale d'une trame est de 51,2 ms. A 100 Mbit/s cela correspondrait à une taille minimale de 640 octets. Les petites trames, qui constituent une part essentielle du trafic interactif, ne contiendraient pratiquement plus d'information utile mais presque uniquement des bits de bourrage. L'émission d'une petite trame à 100 Mbit/s mettra autant de temps qu'à 10 Mbit/s, soit 51,2 ms ! Autre inconvénient, lors de la recopie d'une trame d'un réseau à 100 Mbit/s sur un réseau à 10 Mbit/s, le bourrage ne peut pas être éliminé par Ethernet et la taille des trames serait comprise entre 640 et 1 518 octets. La conséquence paradoxale de l'augmentation de la vitesse de transmission est que, dans cette situation, les performances sont divisées par 10. La seule solution pour que les principes de fonctionnement du CSMA/CD restent valables consiste à réduire la durée de propagation maximale du signal dans le réseau. Il est physiquement impossible d'augmenter la célérité du signal sur le support. Le temps de traversée des couches électroniques est difficilement réductible pour un coût abordable. La seule manière de réduire la durée de parcours du signal est de limiter la taille du réseau. Ceci est aujourd'hui possible. Ethernet 10 Mbit/s a été conçu à une époque où les équipements d'interconnexion étaient peu nombreux et très chers. L'Ethernet 10 Mbit/s devait pouvoir couvrir de grandes étendues : la distance entre deux stations étant au maximum de 2,5 km. A l'heure actuelle, les réseaux locaux sont surtout employés pour interconnecter des machines qui sont à un même étage d'un immeuble ou dans une même pièce. Les distances de câblage peuvent être réduites sans pour autant pénaliser les performances du réseau. De plus, la présence de ponts sur le réseau permet de limiter la superficie des domaines de collision. Le paramétrage d'Ethernet 100 Mbit/s a été fait pour que la tranche canal soit de 5,12 ms. Ainsi la taille minimale de la trame est toujours de 64 octets. La taille maximale est toujours fixée à 1 518 octets. Les réseaux à 10 Mbit/s et 100 Mbit/s sont entièrement compatibles. Les contraintes de câblage, d'électronique pour les équipements découlent de ce choix. L'espacement entre les trames (IFS : *Inter Frame Spacing*) vaut 0,96 ms.

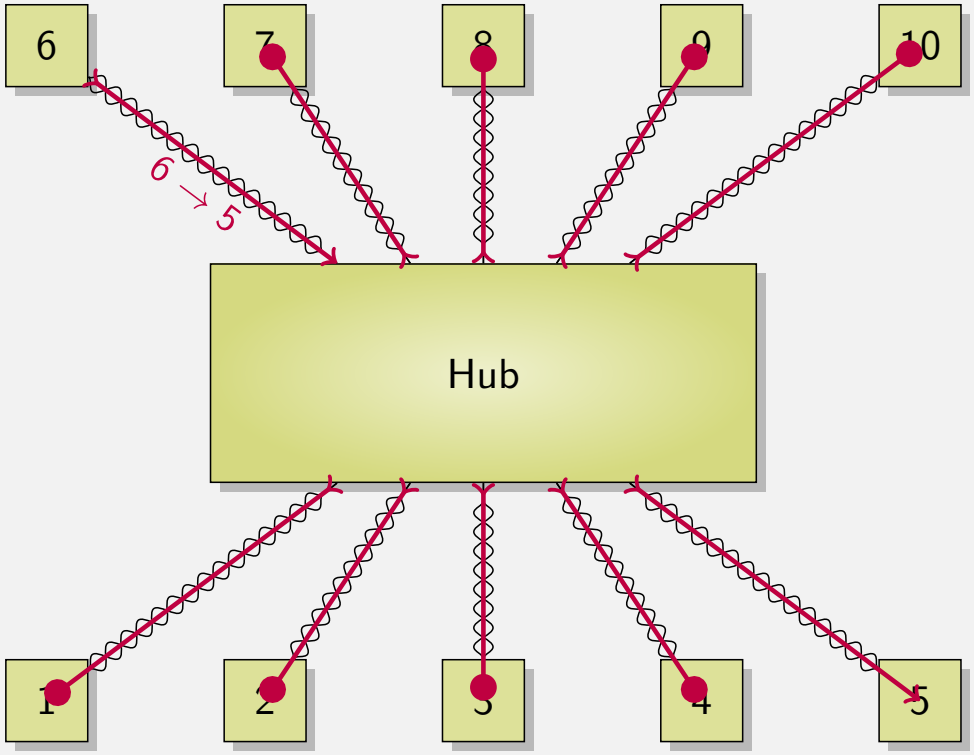
Questions

	F	T
An shared Ethernet network can be 10 km long ?	○	○
An shared Ethernet network can be 10 cm long ?	○	○
An switched Ethernet network can be 10 km long ?	○	○
An switched Ethernet network can be 10 cm long ?	○	○
64 Bytes are not necessary in switched mode	○	○
We have the guaranty that a frame is sent in less than 16 attempts	○	○
CSMA/CD guaranty that frames are sent is the same order they are submitted to the MAC layer	○	○
CSMA/CD can be used with 100BASE-TX and 1GBASE-TX	○	○

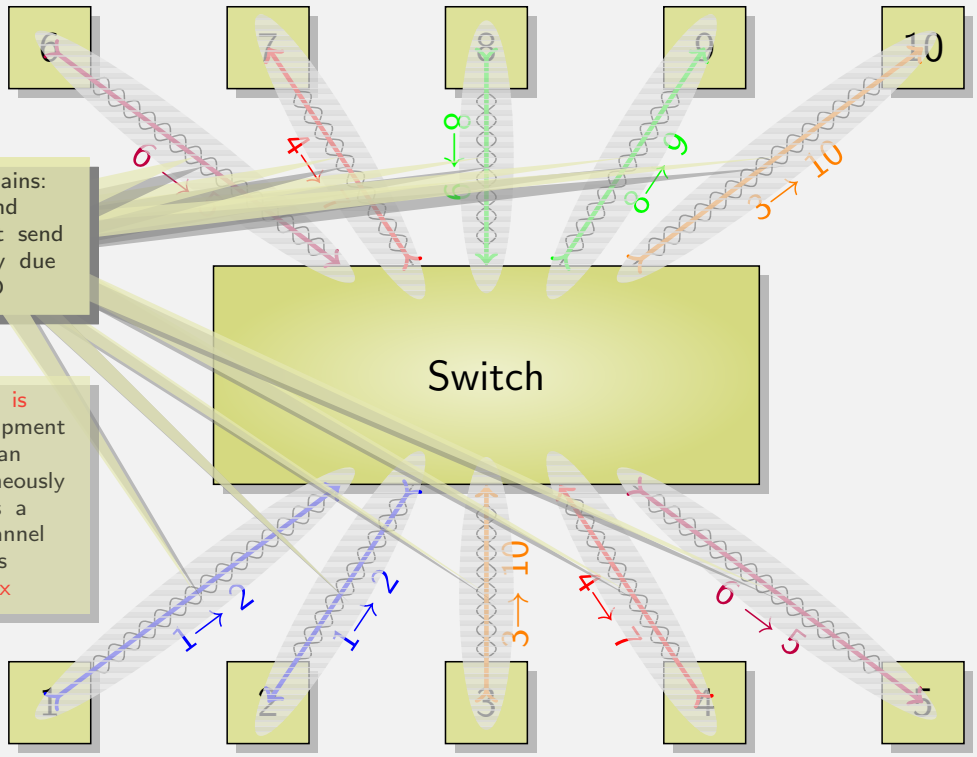
Ethernet

Switching

Switching

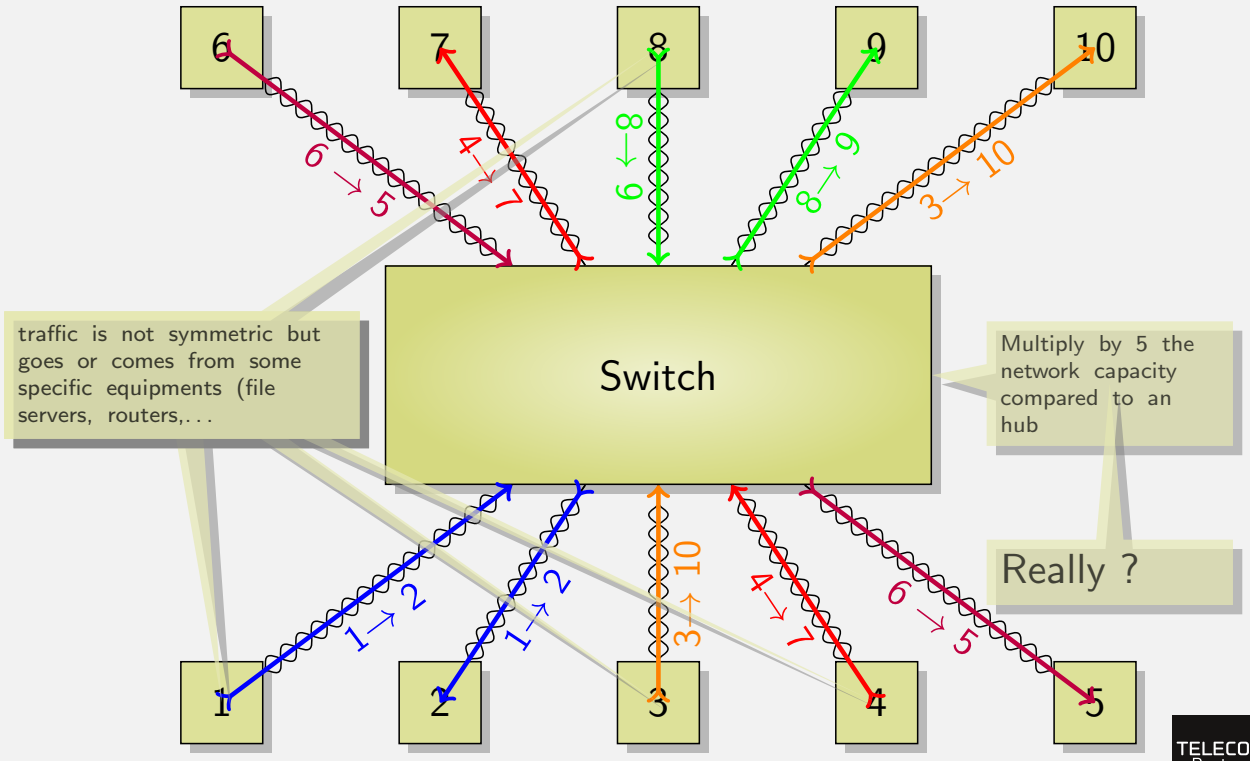


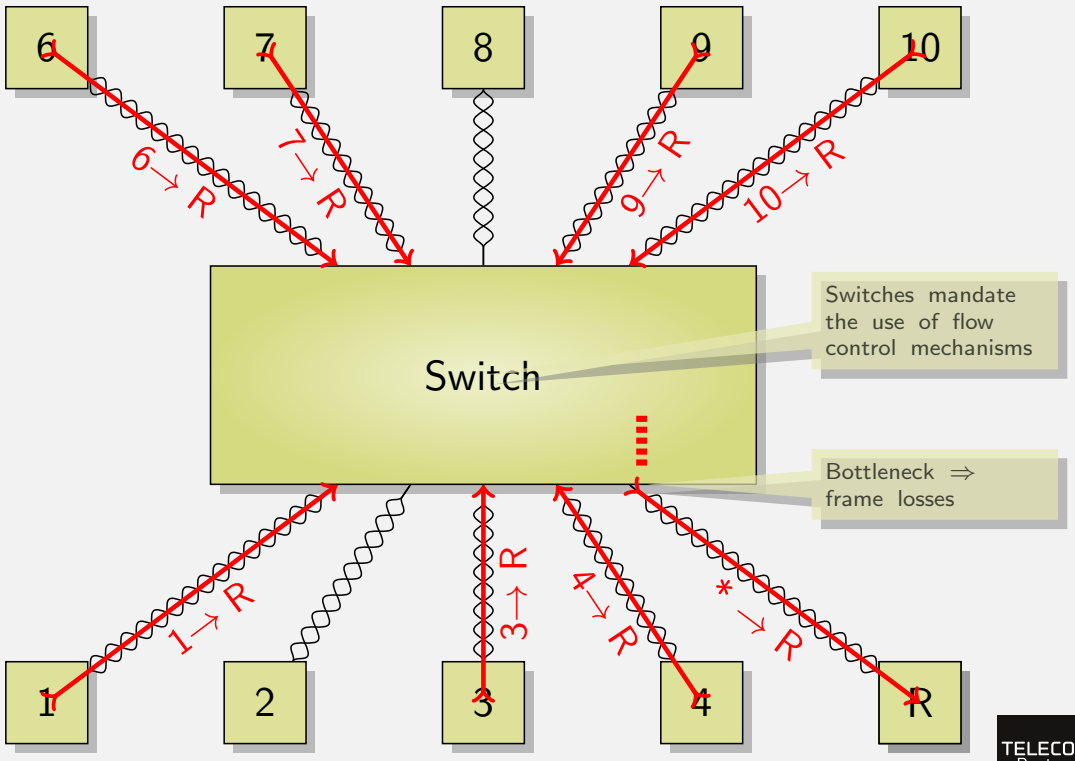
Switching



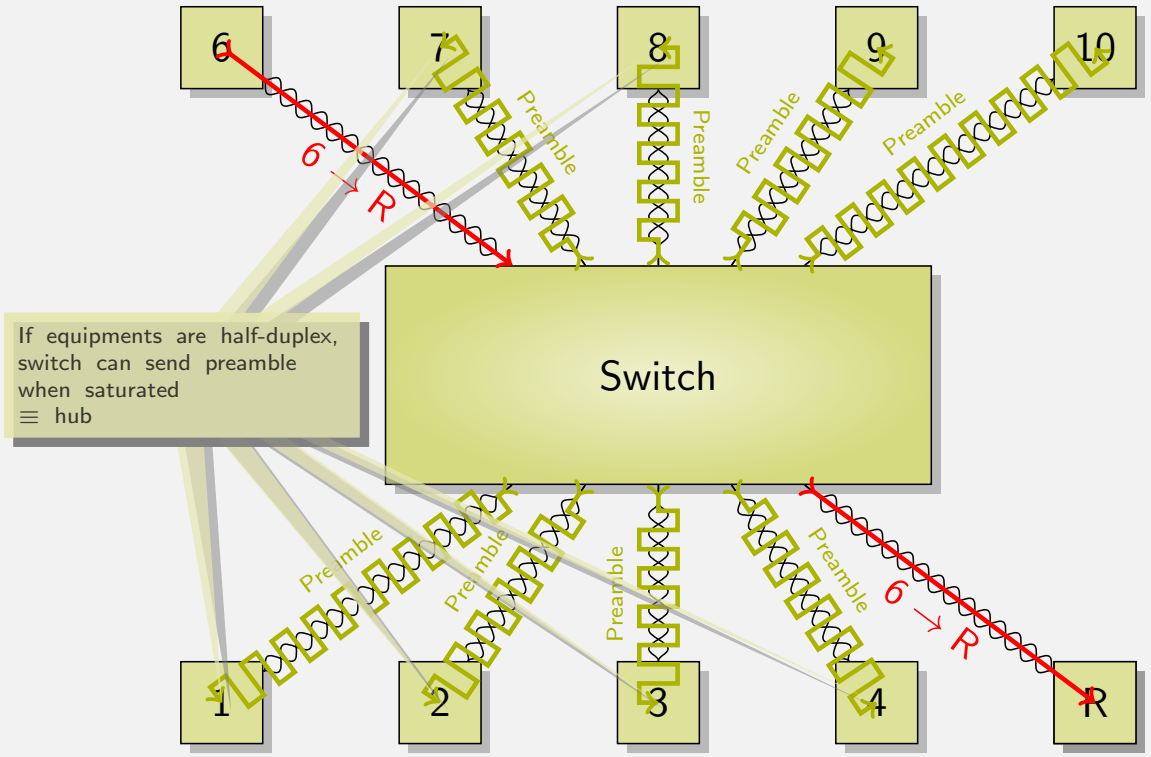
Collision domains: Equipment and switch cannot send simultaneously due to CSMA/CD

If CSMA/CD is disabled, equipment and switch can send simultaneously since there is a dedicated channel for both ways
⇒ full duplex


6
7
8
9
10
6→5
4→7
6→8
8→9
3→10
Switch
traffic is not symmetric but goes or comes from some specific equipments (file servers, routers,...
Multiply by 5 the network capacity compared to an hub
Really ?
1→2
1→2
3→10
4→7
6→5
1
2
3
4
5

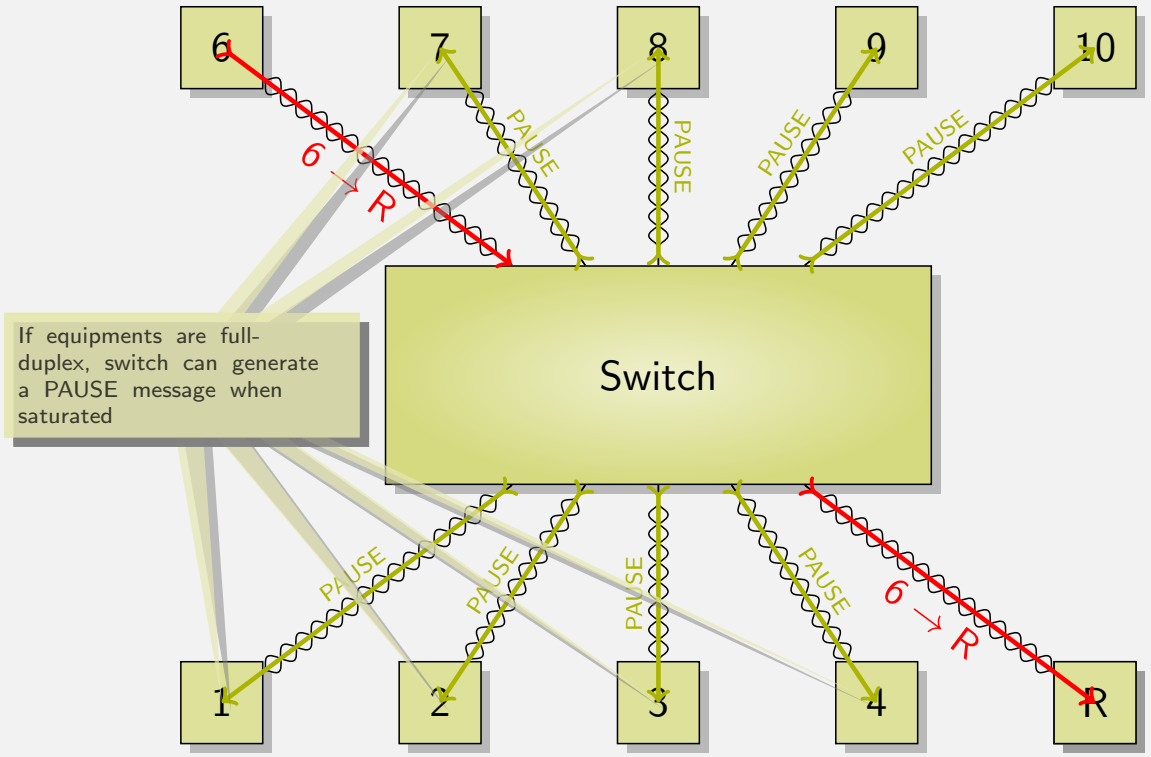

6
7
8
9
10
6→R
7→R
9→R
10→R
Switch
Switches mandate the use of flow control mechanisms
Bottleneck ⇒ frame losses
1→R
3→R
4→R
*→R
1
2
3
4
R

Switching



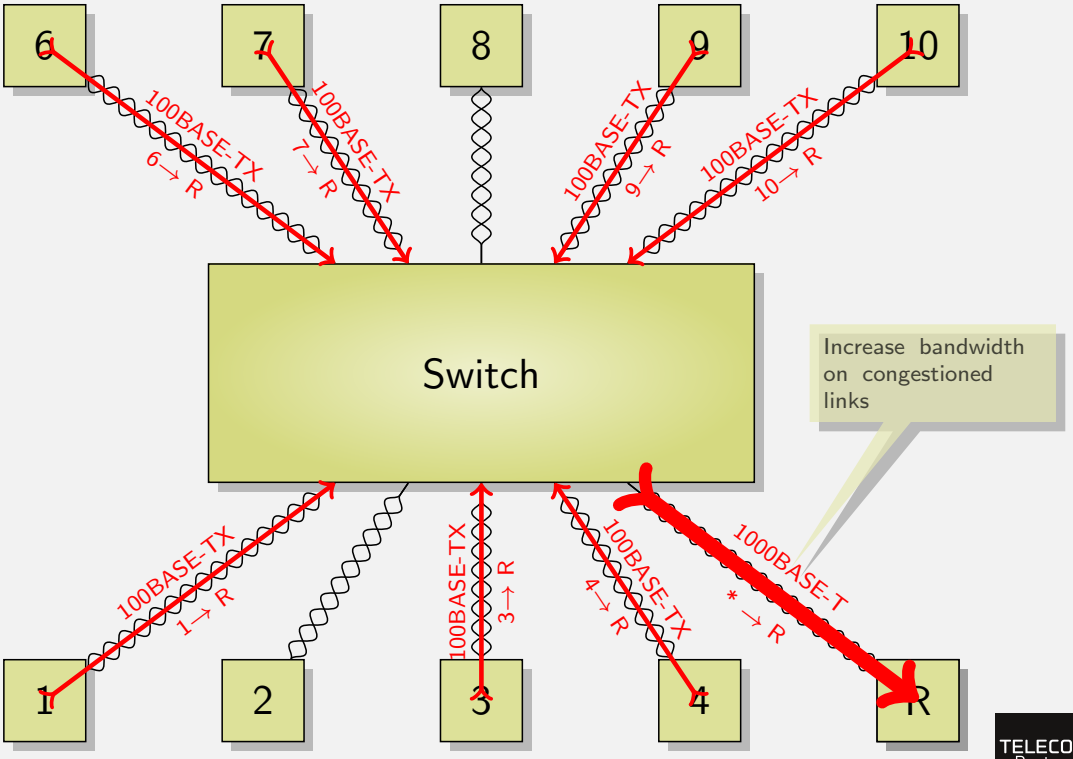
TELECOM Bretagne

Switching



TELECOM Bretagne

Switching



Comments I

Un commutateur (*switch* en anglais) peut être assimilé à un pont comportant autant de ports d'entrée/sortie que de connecteurs. Le commutateur lit le message sur un port en entrée et le recopie sur un port en sortie en fonction de l'adresse de destination contenue dans la trame. Il localise la position des destinataires en lisant le champ adresse de la source des trames qu'il reçoit. A l'intérieur du commutateur, des composants spécialisés et des processeurs rapides réalisent une matrice de commutation permettant de traiter simultanément plusieurs trames. Cette technique permet d'éviter les collisions entre les équipements en ne diffusant plus les messages sur tous les brins. Les stations n'ont pas été modifiée et fonctionnent toujours en half-duplex en mettant en œuvre l'algorithme du CSMA/CD qui écoute le canal avant de transmettre. Par contre, il est possible d'améliorer encore les performances en permettant aux équipements terminaux d'émettre et de recevoir simultanément en désactivant cet algorithme. Dans ce cas, lorsque la station 5 envoie une trame vers la station 6, elle peut recevoir une trame venant de la station 4.

En théorie, la moitié des stations peut communiquer simultanément avec l'autre moitié, multipliant ainsi la bande passante totale du réseau. Chaque station dispose alors d?un débit de 10 Mbit/s garanti. La figure page 173 représente le cas idéal dans lequel chaque station du réseau est en relation avec une autre. Dans cet exemple, la bande passante totale du réseau est de 50 Mbit/s.
Mais dans la pratique le trafic n'est jamais aussi bien réparti entre les stations et celui-ci converge vers les routeurs ou les serveurs de fichiers. Sur un réseau Ethernet traditionnel lorsqu'une station envoie des données, elle empêche toutes les autres d'émettre. Une seule machine à la fois peut parler avec un routeur ou un serveur. Dans le cas d'un réseau Ethernet commuté, cette limitation n'existe plus. Les stations peuvent émettre quand elles le désirent. Si le port de sortie est libre, le message est tout de suite recopié vers le destinataire. Si le port de sortie est occupé, le message est mis dans une file d'attente du commutateur en attendant que le port de sortie se libère. Les commutateurs vont poser des problèmes de congestion du réseau. Avec Ethernet en mode partagé, l'algorithme du CSMA/CD limitait le droit de parole de chaque station créant naturellement un contrôle de flux sur le segment. Avec le commutateur, ce contrôle de flux n'existe plus. Si, par exemple, plusieurs stations émettent à 10 Mbit/s vers un serveur de fichiers, les files d'attente du commutateur vont se saturer et des trames vont être perdues. Les

protocoles de niveau supérieur des stations devront récupérer ces erreurs ce qui peut se traduire par des performances réduites.

Une première solution consiste a utiliser des messages entre le commutateur et les stations pour indiquer une congestion liée soit à la congestion de la matrice de commutation qui ne peut pas traiter simultanément autant de messages ou d'un de ses ports qui reçoit trop de trafic. Si les équipements sont toujours en half-duplex, l'envoi de préambules (appelée en anglais *back pressure*) va empêcher ceux-ci d'émettre puisse qu'une activité est détectée sur le lien.
Si les équipements sont en full-duplex, l'utilisation des préambule n'aura aucun effet et il est nécessaire d'envoyer explicitement un message. Le comité IEEE 802.3x a proposé une méthode basée sur l'émission d'une trame particulière (message PAUSE) indiquant un délai pendant lequel l'équipement ne doit plus émettre de trame :

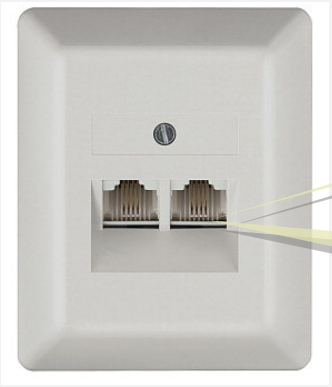
- le champ destination contient soit l'adresse de l'équipement soit l'adresse de multicast 01-80-C2-00-00-01. La source peut être un commutateur ou un équipement terminal distant,
- le champ source contient l'adresse de l'équipement qui a émis le message,
- e champ protocole contient la valeur 0x8808. A noter que l?IEEE a utilisé pour ce message l'encapsulation Ethernet et non l'encapsulation IEEE 802.3,
- le champ code contient la valeur 0x0001 qui indique le message pause, les autres valeurs étant réservées,
- le champ durée contient la durée (en 512 temps bits) pendant laquelle la station ne doit plus émettre de trames ce qui donne pour l?Ethernet à 100 Mbit/s une durée comprise entre 0 et 335,54 millisecondes,
- les 42 octets suivant servent de bourrage pour atteindre la taille minimale d'une trame Ethernet.

Une seconde solution, de type ingénierie réseau consiste, pour limiter ces phénomènes d?engorgement, à attribuer plus de bande passante aux liens qui risquent d'être saturés.

Ethernet

Auto-Configuration

Auto-Configuration



What are the switch/hub capabilities: 10BASE-T, 100BASE-TX, 1000BASE-T? Full-duplex, Half-duplex?

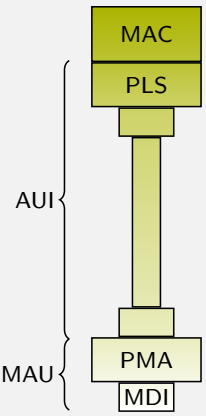
Same question for the equipment connected

Same Media Dependant Interface (RJ-45)
Find the best Ethernet technologies compatible to both ends

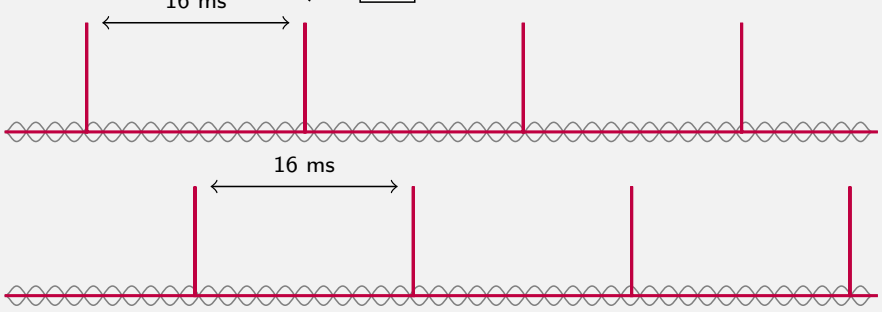
⇒ Auto-negociation

TELECOM Bretagne

Auto-Configuration

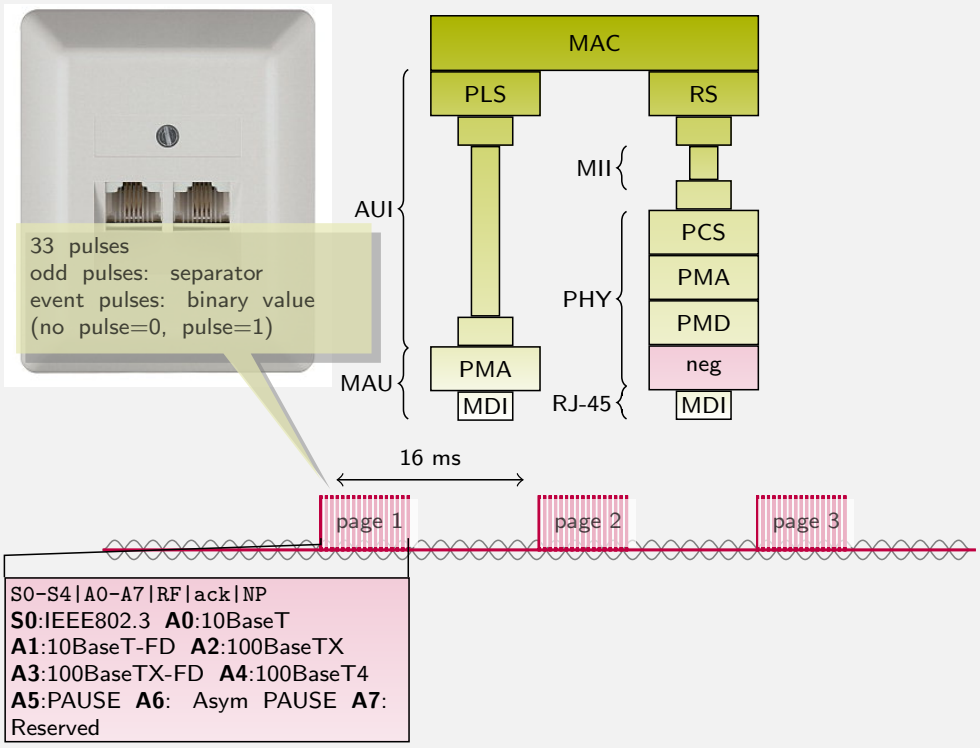


Link test: hen no traffic a pulse every 16ms
switches on interface's leds

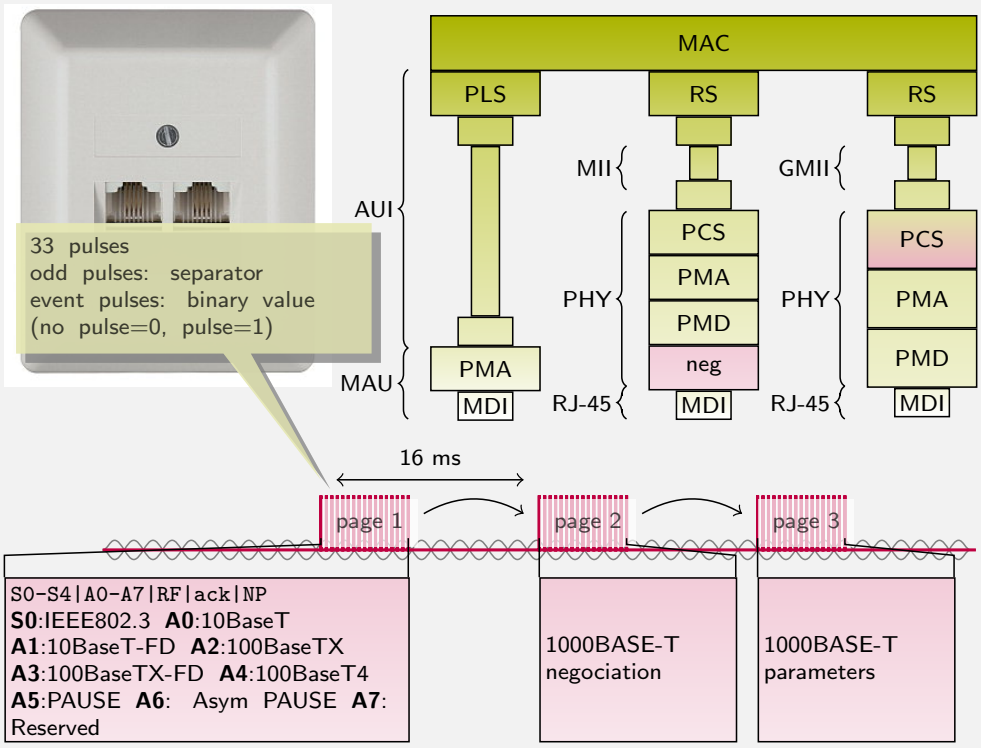


TELECOM Bretagne

Auto-Configuration



Auto-Configuration



Comments I

L'auto-négociation permet à l'équipement d'interconnexion (hub, répéteur, commutateur) de détecter la présence d'un équipement actif connecté à l'autre extrémité et ses caractéristiques pour adapter la vitesse de transmission. La mise en œuvre de ces mécanismes pour les débits de 10 et 100 Mbit/s n'est pas obligatoire, elle l'est par contre pour le 1000BASE-T.

Initialement pour le 10BASE-T, seule la fonction de détection d'un équipement était présente puisqu'il n'y avait pas d'ambiguïté quant à la vitesse de transmission. Quand il n'y a pas de transmission de trames sur le support physique, les transceivers émettent une impulsion toutes les 16,8 millisecondes composée d'une unique oscillation appelée Normal Link Pulse (NLP) ou Link Test Pulse (LTP) . Un équipement qui détecte de 2 jusqu'à 10 impulsions déclare que l'équipement distant est actif. Une absence d'impulsion pendant une période de 50 à 150 millisecondes fait considérer l'équipement distant comme inactif.

Pour les versions suivantes de la norme, ce mécanisme a été modifié pour permettre de transmettre les données nécessaires à la négociation. Un train de 33 impulsions appelé FLP (*Fast Link Pulse*) est émis par les transceivers. Les 17 positions impaires contiennent toujours une impulsion. Les 16 positions paires permettent de transmettre un mot de 16 bits (une impulsion correspond à un bit à 1 tandis qu'une absence d'impulsion correspond à un bit à 0). Les cinq premiers bits échangés permettent de sélectionner la norme. Pour l'instant l'auto-négociation est définie pour les normes IEEE 802.3, IEEE 802.9, IEEE 802.5 et IEEE 1394 (FireWire) qui peuvent utiliser un câblage RJ-45. Pour la norme IEEE 802.3, les 8 bits suivants permettent de désigner la technologie utilisée suivant le bit positionné à 1 :

- A0 : 10BASE-T,
- A1 : 10BASE-T Full Duplex,
- A2 : 100BASE-TX,

Comments II

- A3 : 100BASE-TX Full Duplex,
- A4 : 100BASE-T4.

Le bit A5 s'il est à 1 indique que l'équipement peut recevoir des trames PAUSE permettant d?effectuer un contrôle de flux. Les bits A6 et A7 sont réservés. Le bit RF (*Remote Fault*) indique que l'équipement a détecté une erreur lors de la négociation avec l'équipement distant. Le bit Ack indique que les données ont été correctement reçues et le bit NP (Next Page) indique que d'autres informations vont suivre.
Les pages suivantes peuvent être de type :

- message (bit MP à 1) contenant un code définit par l?IEEE désignant une information ;
- non formaté (bit MP à 0) contenant la valeur correspondant au code transporté par le message.

Le bit T (Toggle) est utilisé pour numéroter les pages, il prend alternativement les valeurs 0 et 1. Le bit Ack2 sert à indiquer que les paramètres transportés dans la trame sont acceptés par l'équipement distant (le bit Ack qui ne sert qu'à acquitter la trame).

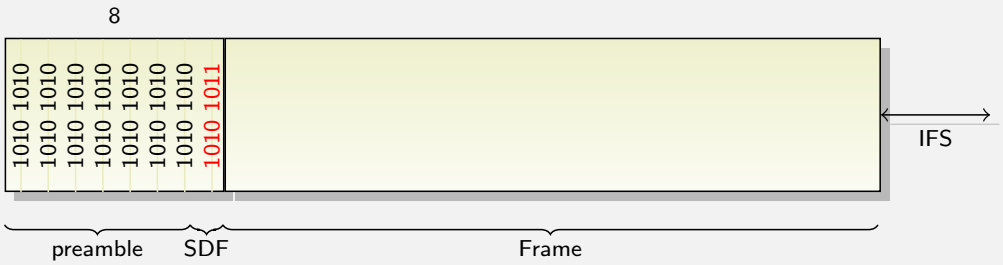
Si un transceiver à une extrémité implémente uniquement la norme 10BASE-T, il répondra à un train d?impulsions FLP par une simple impulsion NLP. Dans ce cas la liaison est configuré en 10BASE-T. Si les deux transceivers sont capables d'échanger des informations par le biais des trains d'impulsions FLP, la liaison se fera avec la meilleure technologie commune aux deux transceivers. La norme définit les priorités de la plus performante à la moins performante dans l'ordre suivant : 100BASE-TX Full Duplex, 100BASE-T4, 100BASE-TX, 10BASE-T Full Duplex, 10BASE-T.

Cette hiérarchie ne prend en compte que les équipements aux extrémités et ignore la qualité du câble, ce qui peut conduire à des problèmes de connexion. Ainsi si le câble est de catégorie 3 et que les deux transceivers ont la possibilité de se configurer en 100BASE-TX, le train d'impulsions passera correctement sur le câble, mais par contre la transmission de données sera impossible.

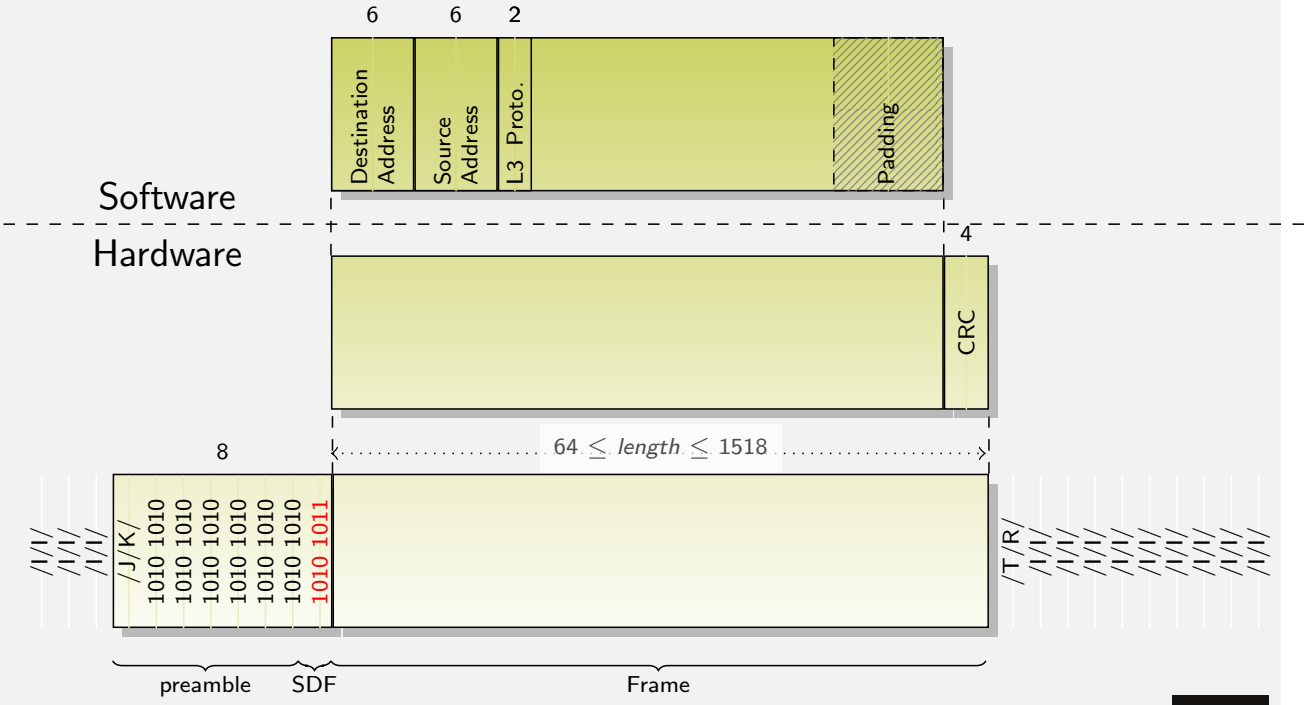
Ethernet

Frame Format

Frame Format

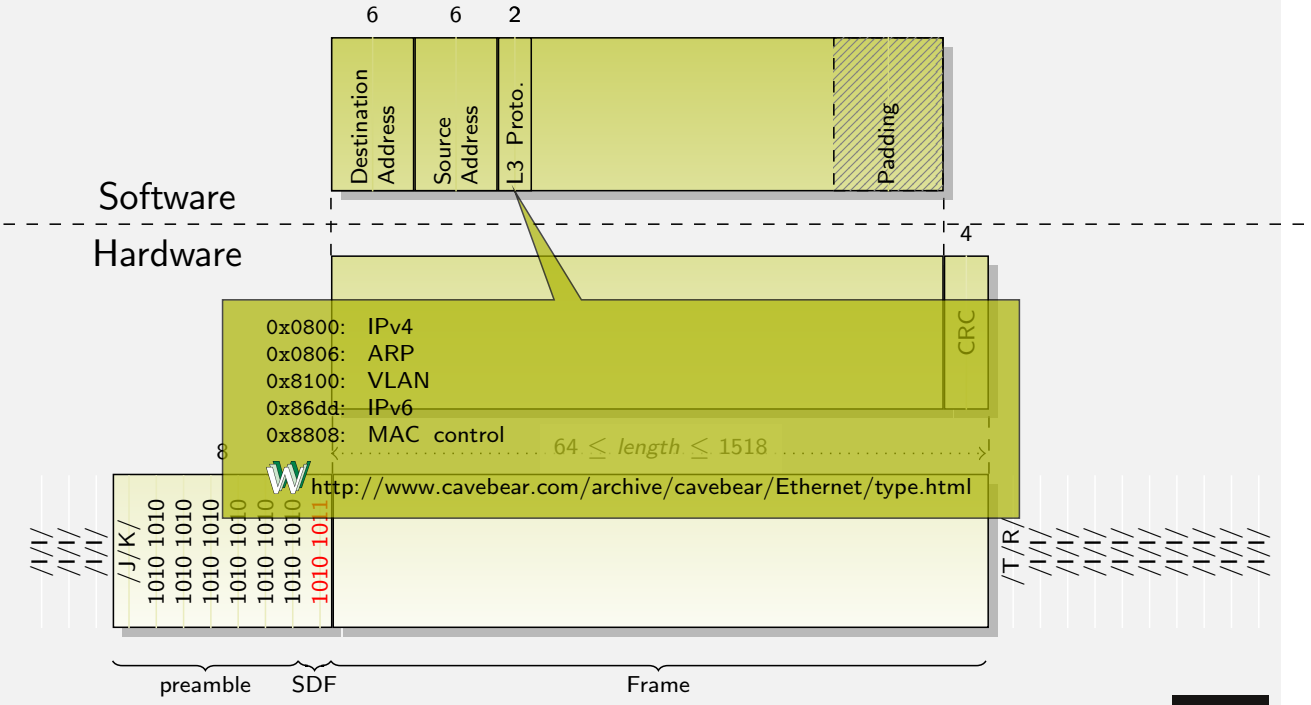


Frame Format



TELECOM Bretagne

Frame Format



TELECOM Bretagne

Comments I

Le niveau physique de l'IEEE 802.3 commence l'émission des trames par un préambule, c'est-à-dire une séquence de 7 octets ayant pour valeur binaire 10101010. Ces octets ont pour but de synchroniser l'horloge du récepteur avec l'horloge de l'émetteur. Comme aucun trafic n'existe quand les stations du réseau n'ont rien à émettre, il n'y a aucune raison pour que les horloges des différentes stations restent synchronisées. Les premiers bits du préambule peuvent être détruits par la traversée de répéteurs pendant la transmission de la trame. Les répéteurs peuvent ne pas récupérer les premiers bits du préambule, par contre, ils doivent reconstruire un préambule entier lors de la retransmission des données.
Cette valeur a été choisie pour que le signal émis avec un codage Manchester (10BASE5, 10BASE2) soit carré . Comme les premiers bits peuvent être détruits, la transmission doit utiliser un codage Manchester classique, car le premier bit reçu ne peut pas servir de référence.
Le début d'émission des données du niveau supérieur est repéré par le champ SFD (*Starting Delimitor Frame*) ayant pour valeur 1010 1011.

Pour permettre aux récepteurs de séparer les différentes trames qui se succèdent sur le bus, celles-ci sont séparées par un silence de 9,6 ms appelé silence intertrame (IFS : Inter Frame Spacing). Le tapis roulant des caisses des supermarchés est une bonne illustration de cette technique. Pour que la caissière puisse faire la différence entre les achats des différents clients, ceux-ci laissent un espace significatif entre leurs achats et ceux de la personne précédente. L'IFS permet aussi aux phénomènes d'écho qui se produisent aux extrémités des câbles de se résorber avant une nouvelle transmission, sinon ces phénomènes pourraient créer des brouillages qui pourraient être interprétés par les stations comme des collisions intempestives.

Dans les technologies point-à-point, les équipements n'ont plus à garder le silence entre les périodes d'émission de trames. Grâce au codage de l'information, il existe des symboles qui ne codant pas des données binaires permettent une signalisation. Ainsi entre les trames, il est possible d'envoyer des symboles /I/ qui permettent de garder une synchronisation permanente entre les horloges de l'émetteur et le récepteur. Les bits de synchronisation du préambule perdent donc de leur signification, mais ils sont conservé pour permettre un espacement correct entre les

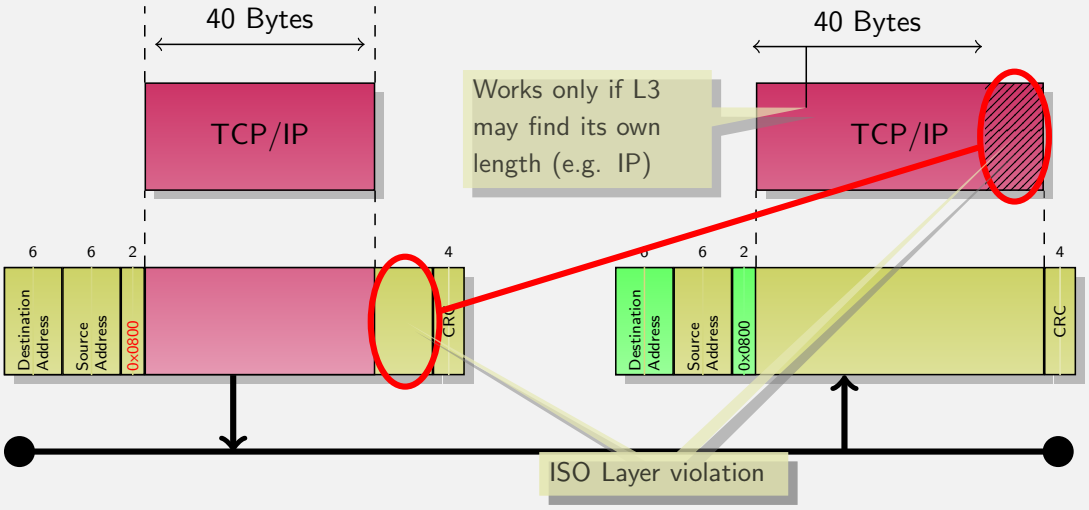
Comments II

trames, si celles-ci doivent être recopiées sur des réseaux CSMA/CD. Le premier octet de synchronisation est remplacé par la séquence /J/K/ et après la transmission du CRC dans ce qui étant réservé au silence inter-trame, le symbole /T/R/ est transmis pour indiquer la fin de la trame.

Le format de la trame Ethernet le plus couramment utilisé est le suivant:

- l'adresse de destination respecte le format précédemment défini . Le choix d?une adresse sur 6 octets dépend de la mise en œuvre du réseau, mais en pratique ce choix est toujours sur 6 octets.
  Ce champ est le premier de la trame pour permettre aux stations réceptrices de déterminer, dès le début de la transmission, si la trame leur est destinée ;
- l'adresse de la source indique la station qui a émis la trame. La taille est la même que celle de l'adresse de destination ; Ce champ va permettre aux ponts et aux commutateurs d'apprendre la localisation des équipements sur le réseau.
- Le troisième champ contient l?identificateur du protocole de niveau supérieur au lieu de contenir la longueur des données.
- Les données provenant du niveau supérieur sont mise dans la partie suivante.
- Si la taille minimale de la trame de 64 octets, conséquence de l'algorithme du CSMA/CD n'est pas respectée, des octets de bourrage sont introduits après les données.
- Finalement la trame se termine par un CRC ($x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$). Généralement le calcul et la vérification du CRC est effectué par le matériel (i.e. la carte réseau). Les équipements informatiques ne le voient donc pas et des outils comme *wireshark* ou *tcpdump* vont visualiser des trames de longueur minimale de 60 octets.

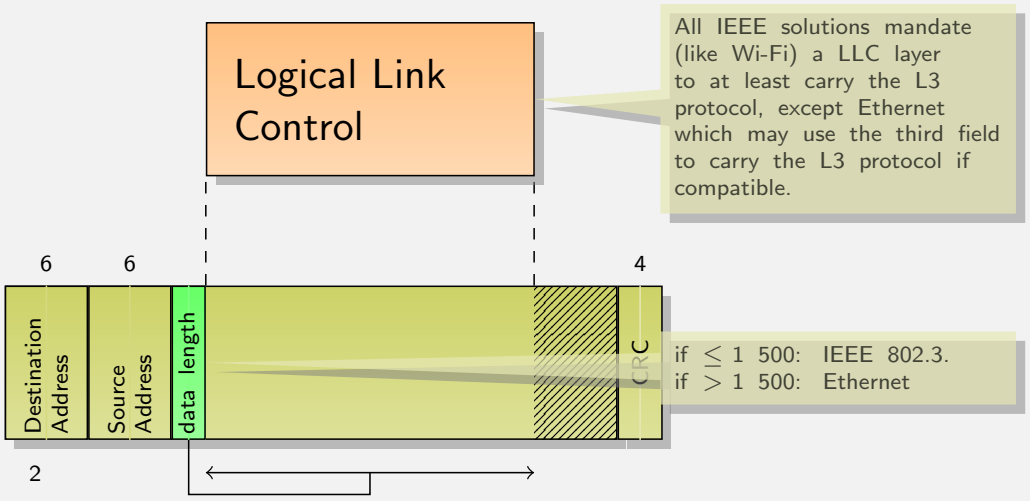
Frame Format



Comments I

Les bits de bourrage introduits au niveau MAC par l'émetteur ne peuvent pas être supprimés par le niveau MAC. Ces octets sont donc retransmis au niveau 3. On assiste ici à une violation du fonctionnement de l'architecture en couches puisque des données produites au niveau 2 vont remonter les couches de l'équipement distant ; le protocole de niveau 3 doit contenir un moyen quelconque (par exemple un champ longueur du paquet) pour éliminer les octets de bourrage s'ils existent. Un protocole populaire comme IP (version 4 ou 6) contient un champ qui indique sa propre longueur. Il est donc possible d'utiliser IP directement sur Ethernet. Si un protocole de niveau 3 n'est pas en mesure d'éliminer les octets de bourrage, il n'est pas possible d'utiliser le format de trame définit par Ethernet, il faut utiliser celui-ci définit par l'IEEE est qui est compatible avec le modèle ISO.

Ethernet vs IEEE 802.3



Comments I

La mise en trame des premiers champs est faite par le logiciel, il est donc indépendant du matériel. La norme IEEE 802.3 recommande d'utiliser le troisième champ de manière différente d'Ethernet. Au lieu de coder le protocole de niveau supérieur, ce champ indique la taille des données en octets (sans le bourrage). Cela permet de rendre compatible le protocole avec le modèle de référence de l'ISO (des données produites au niveau 2 comme les octets de bourrage sont consommée au même niveau chez le récepteur). On perd donc la possibilité au niveau MAC d'aiguiller vers le niveau supérieur. Il faut donc ajouter une sous-couche protocolaire LLC qui va permettre (en autre) de coder le protocole de niveau supérieur.

La différence entre une trame IEEE 802.3 et une trame Ethernet se fera au niveau de la valeur du troisième champ. La taille des données étant limitée à 1 500 octets, les valeurs supérieures serviront à coder des protocoles pour les trames Ethernet.

Questions


```
0x0000:  0180 c200 0000 0019 e79e 1dae 0026 4242
0x0010:  0300 0000 0000 8000 0002 b972 6601 0000
0x0020:  0008 8005 0019 e79e 1d80 802e 0200 1400
0x0030:  0200 0f00 0000 0000 0000 0000
```

This dump gives a IEEE 802.3 frame:

- What is the nature of the destination address ?
- What is the nature of the frame (Ethernet or IEEE 802.3) ?
- Can you identify padding bytes ?
- Can you give the frame CRC ?

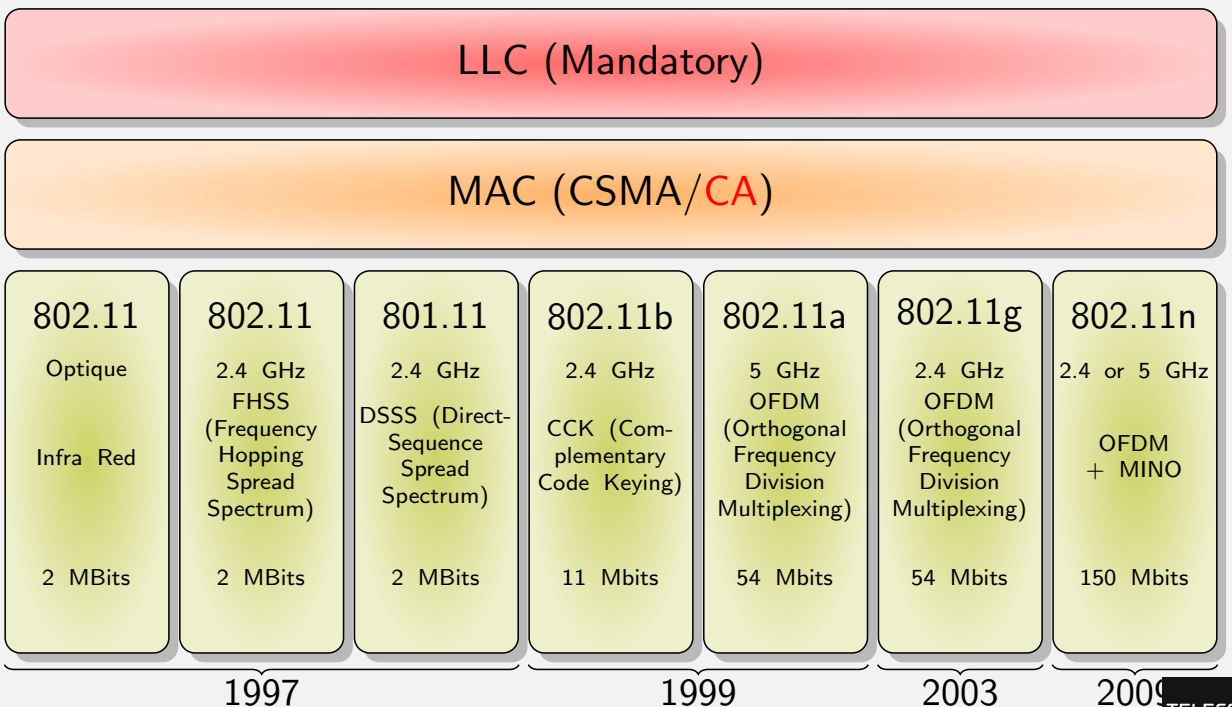
Ethernet: References

- Cisco tutorial:
  http://www.cisco.com/en/US/docs/internetworking/technology/handbook/Ethernet.html#wp1020985
- Auto-configuration: www.ethermanage.com/ethernet/pdf/dell-auto-neg.pdf

IEEE 802.11

Physical Layer

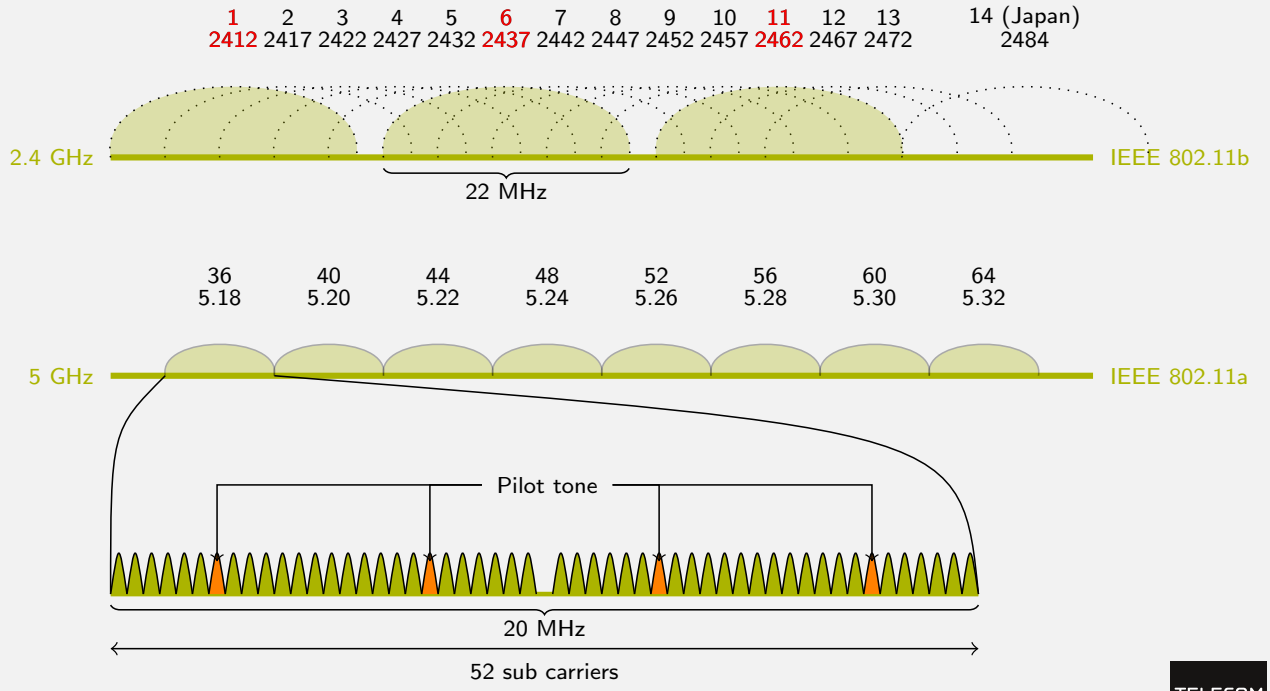
IEEE 802.11 Standards



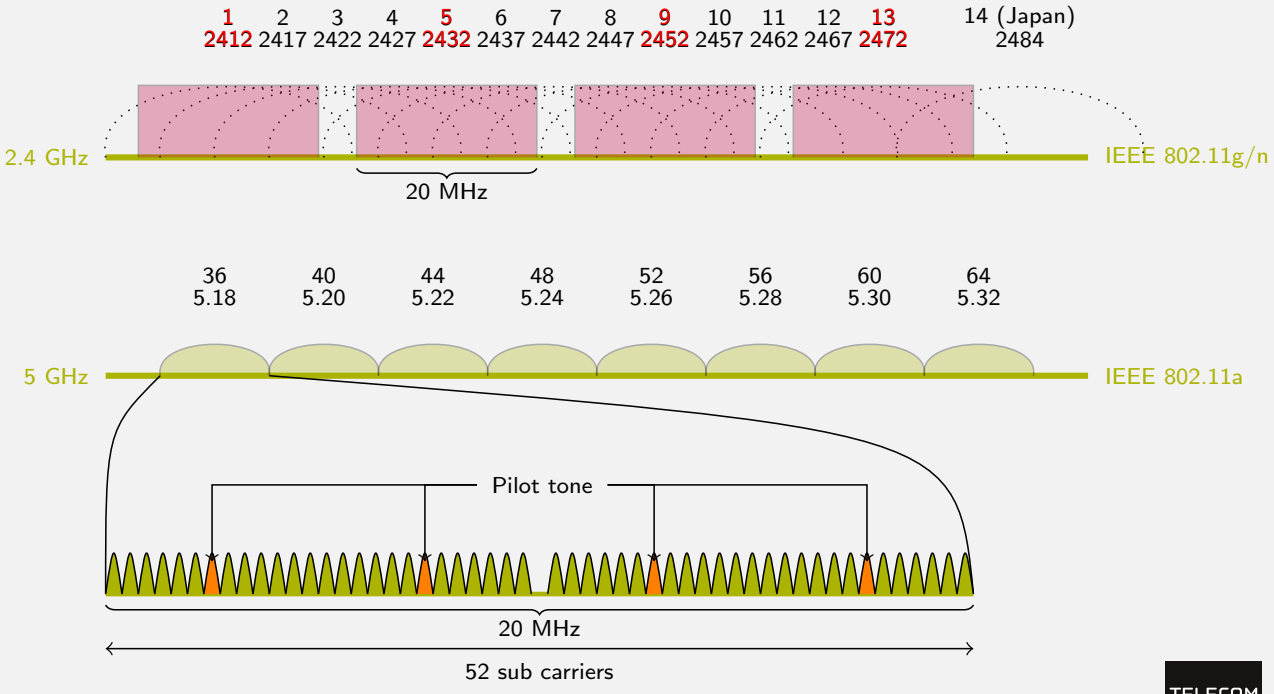
Comments I

Le standard IEEE 802.11 publié en 1997 utilisait 3 types de modulations. Une en infra-rouge et deux en radio en utilisant soit l'étalement de spectre (DSSS : *Direct-Sequence Spread Spectrum*), soit le saut de fréquence (FHSS : *Frequency Hopping Spread Spectrum*) offrant un débit partagé de 2 Mbit/s. Néanmoins, la norme ne connu un réel succès que lorsque la version IEEE 802.11b fut publié permettant un débit de 11 Mbit/s la rapprochant des performances qu'offrait Ethernet à cette époque.
La norme IEEE 802.11b utilise CCK (*Complementary Code Keying*) pour coder l'information. Les autres évolutions du standard utilisent OFDM (*Orthogonal Frequency Division Multiplexing*) permettant des débits plus élevés, une utilisation plus large du spectre et plus de souplesse pour adapter la modulation aux conditions de propagation. Néanmoins, pour des raisons de compatibilités entre les technologies, un replis vers CCK est possible généralement pour les messages de signalisation (trame MAC).

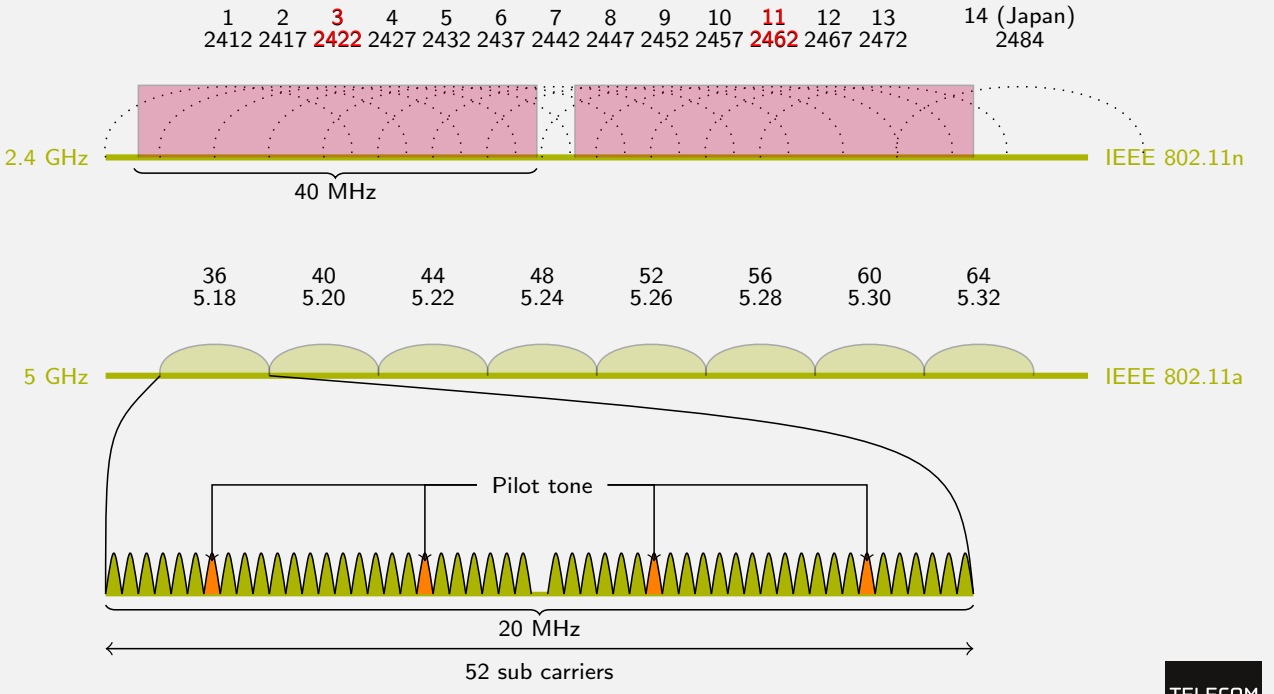
Channels



Channels



Channels



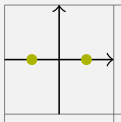
Comments I

La bande des 2.4 GHz est décomposée en 13 canaux d'une largeur de 22 MHz se recouvrant. Au maximum, seuls trois canaux le 1, le 6 et le 11 ne se recouvrent pas. Ces canaux peuvent être exploité en utilisant les modulation CCK, soit une modulation OFDM. Dans ce cas, chaque canal est découpé en 52 sous-fréquences, dont 4 pour la correction d'erreur. Dans le cas d'OFDM, il est possible de regrouper des canaux pour permettre de plus grands débits.

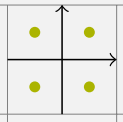
WiFi Throughputs

Data Rate (Mbit/s)	Transmission Type	Modulation Scheme	Standard
1	DSSS	Baker code & BPSK	802.11-1999
2	DSSS	Baker code & QPSK	802.11-1999
5.5	DSSS	CCK & QPSK	802.11b/g
6	OFDM	BPSK	802.11a/g
9	OFDM	BPSK	802.11a/g
12	OFDM	QPSK	802.11a/g
11	DSSS	CCK & QPSK	802.11b/g
18	OFDM	QPSK	802.11a/g
24	OFDM	16 QAM	802.11a/g
36	OFDM	16 QAM	802.11a/g
48	OFDM	64 QAM	802.11a/g
54	OFDM	64 QAM	802.11a/g

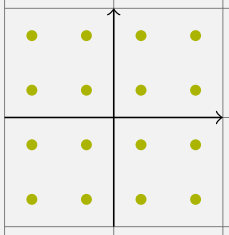
Modulations



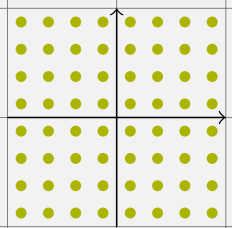
BPSK: Binary Phase Shift Keying (20 MHz)
Coding rate = 1/2 ⇒ 6 Mbit/s
Coding rate = 3/4 ⇒ 9 Mbit/s



QPSK: Quadrature Phase Shift Keying (20 MHz)
Coding rate = 1/2 ⇒ 12 Mbit/s
Coding rate = 3/4 ⇒ 18 Mbit/s



16-QAM: Quadrature Amplitude Modulation (20 MHz)
Coding rate = 1/2 ⇒ 24 Mbit/s
Coding rate = 3/4 ⇒ 36 Mbit/s



64-QAM: Quadrature Amplitude Modulation
Coding rate = 2/3 ⇒ 48 Mbit/s
Coding rate = 3/4 ⇒ 54 Mbit/s

Comments I

Les spécifications initiales de la norme en 1997 et 1999 utilisent un codage par étalement de spectre (nous ne nous interessons pas ici au saut de fréquence. Un codage BPSK (Binary phase-shift keying) utilise une fréquence et la module en changeant sa phase. Ainsi un bit à 0 peut être codé sur une phase et un 1 par un déphasage de 180° (π). Le code de Baker consiste dans ce cas à émettre 11 éléments pour coder un bit. Ainsi le bit 0 est codé par la séquence +1 +1 +1 -1 -1-1 +1 -1 -1 +1 -1 (valeur choisie car elle a un faible taux d'auto-correlation). Le bit a 1 est codé en inversant les valeurs soit -1 -1 -1 +1 +1 +1 -1 +1 +1 -1 +1. Ainsi si le signal est modulé a 11 méga symboles/s, le débit binaire sera d'1 Mbit/s. En utilisant un codage QPSK (Quadrature phase-shift keying), on utilise 4 déphasages 0, $\pi/2$, π, $3\pi/4$ pour coder respectivement les valeurs binaires 00, 01, 11, 10. Ainsi a chaque modulation on double le nombre de bit transportés. Un signal est modulé a 11 méga symboles/s offrira un débit binaire sera d'2 Mbit/s.
Le codage CCK (Complementary Code Keying), utilisé par la norme IEEE 802.11b, est plus complexe. Ce codage permet de n'utiliser que 8 symboles (à la place de 11) pour coder une valeur binaire. Donc la vitesse de modulation se trouve multipliée d'un facteur 11/8 soit 1.375. En utilisant une modulation QPSK on transmet 2 bit à chaque modulation, et en utilisant 4 décalage du codage initial, 2 bits sont également transmis . En jouant donc sur les codage et la phase on peut donc transmettre 4 bits par modulation d'où un débit de 4*1.375= 5.5 Mbit/s. En utilisation 8 décalage du code, 4 bits sont transmis, le débit de 11 Mbit/s est atteint.
Le standard IEEE 802.11a est lui basé sur OFDM. Il s'agit de diviser une bande de fréquences en sous-porteuses pouvant se chevaucher, mais sans interférences. Le signal ODFM est composé de 52 sous-porteuses (dont 4 pilotes) s'étalant sur 20 Mhz. Chaque sous-porteuse peut être modulée en BPSK, QPSK, 16-QAM (Quadrature Amplitude Modulation) ou 64-QAM. Si, deux modulations doivent être simultanément utilisées, une sera forcement BPSK. Ce codage permet suivant le code correcteur utilisé pour coder l'information et le codage des débits de: 6, 9, 12, 18, 24, 36, 48 et 54 Mbit/s.
Le standard IEEE 802.11g transpose dans la gamme de fréquences du 2.5 GHz la modulation OFDM. Dans ce cas le codage CCK reste disponible et l'on peut donc avoir les débits suivants: 1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48 et 54 Mbit/s.

Versions compatibility

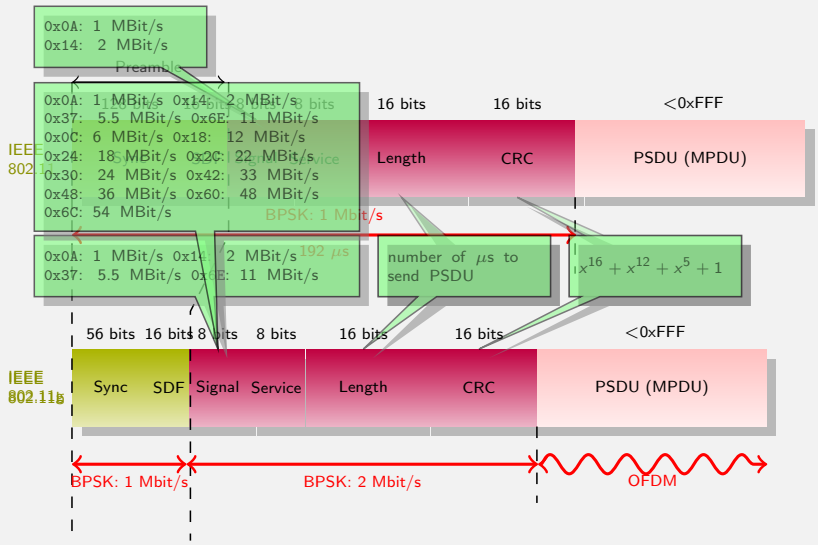
IEEE 802.11b introduces short preamble

- ERP: Extended Rate Physical
- Long preamble: 192 bits or 192 μs
- Short preamble: 120 bits or 96 μs

IEEE 802.11g introduces OFDM

- 3 main physical layers
  - ERP-DSSS/CCK: 1 or 2 Mbit/s legacy physical layer
  - ERP-OFDM: 6 to 54 MBit/s, pure OFDM
    - Invisible from legacy nodes (IEEE 802.11 or IEEE 802.11b)
  - DSSS-OFDM: Start in DSSS and continue with OFDM
    - Ensure compatibility with IEEE 802.11 and IEEE 802.11b

Physical format



http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.8843&rep=rep1&type=pdf

Comments I

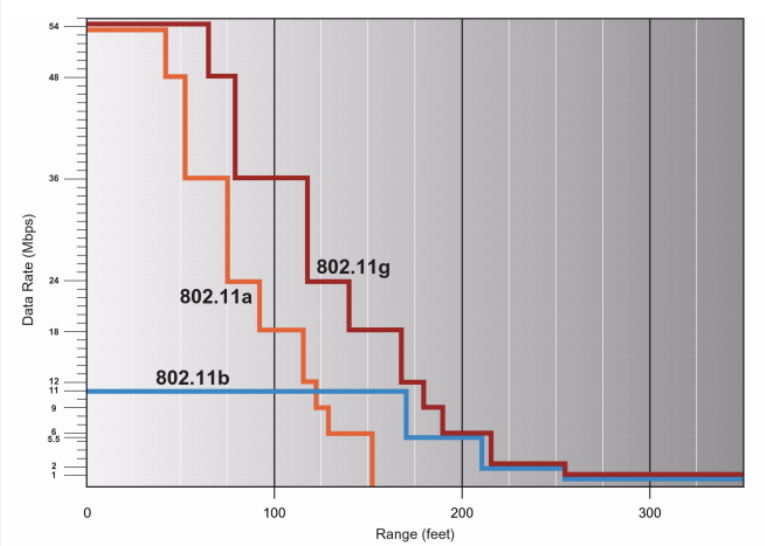
La norme IEEE 802.11 ayant évoluée au cours du temps, il faut s'assurer de la compatibilité des nouvelles versions avec les anciennes et les communications doivent être possibles dans les deux sens. La norme IEEE 802.11 initiale ne prévoyait qu'un mode de fonctionnement à 1 et 2 Mbit/s. Au niveau physique, la norme définit un en-tête PLCP (*Physical Layer Convergence Procedure*). d'une longueur de 192 bits. Il commence par un champ sur 128 bits pour synchroniser l'émetteur avec le récepteur, comme pour Ethernet, ce champ se termine par une séquence bien connue pour indiquer la fin de la période de synchronisation (SDF: *Start Frame Delimiter*). Le champ *Signal* contient un code donnant le débit de transmission des données binaires: `0x0A` pour 1 MBit/s soit un codage BPSK 0x14 pour 2 MBit/s soit un codage QPSK. L'en-tête est toujours émis a 1 Mbit/s, cette indication de vitesse ne concerne que la partie données du PDU. Le champ *Service* contient quelques informations sur la porteuse utilisée et un bit permettant d'éviter des erreurs d'arrondis lors du calcul de la longueur des données. Le champ textitLength contient la durée d'émission en micro-seconde (arrondi au bit supérieur) des données de la trame PLCP, soit les information venant de la couche supérieure (MAC-PDU).
La norme IEEE 802.11a a permis une montée en débit à 54 Mbit/s en utilisant OFDM. Mais comme les bandes de fréquences sont différentes, il n'y a pas de problème de compatibilité. La norme IEEE 802.11b travaillant dans la même bande de fréquence que la norme initiale, il faut indiquer au début de la trame la modulation possible pour atteindre un correspondant. Pour se faire le nombre de valeur possible pour le champ Signal a été revu pour introduire les vitesses de 5.5 Mbit/s (0x37) et 11 Mbit/s (0x6E). Il est également possible de réduire la longueur du préambule de synchronisation pour permettre d'augmenter les performances de transmission, cela pose un problème de compatibilité avec les premières versions de la normes. Cette possibilité ne peut être activée que si le correspondant est capable de la traiter.
La norme IEEE 802.11g reprend le codage OFDM défini par IEEE 802.11a mais dans la gamme de fréquence des 2.4 GHz. En plus de définir d'autres vitesses de codage, l'OFDM peut poser des problèmes de compatibilité car les anciennes stations considèrent ces émissions comme du bruit et peuvent décider d'émettre des données qui vont brouiller les communications OFDM. Pour éviter des problèmes une approche consiste à émettre un en-tête à 1 Mbit/s compréhensible par l'ensemble des équipements et n'utiliser le codage OFDM que pour les données (trame MAC).

Questions

Is there compatibility problems between IEEE 802.11a and IEEE 802.11g ?

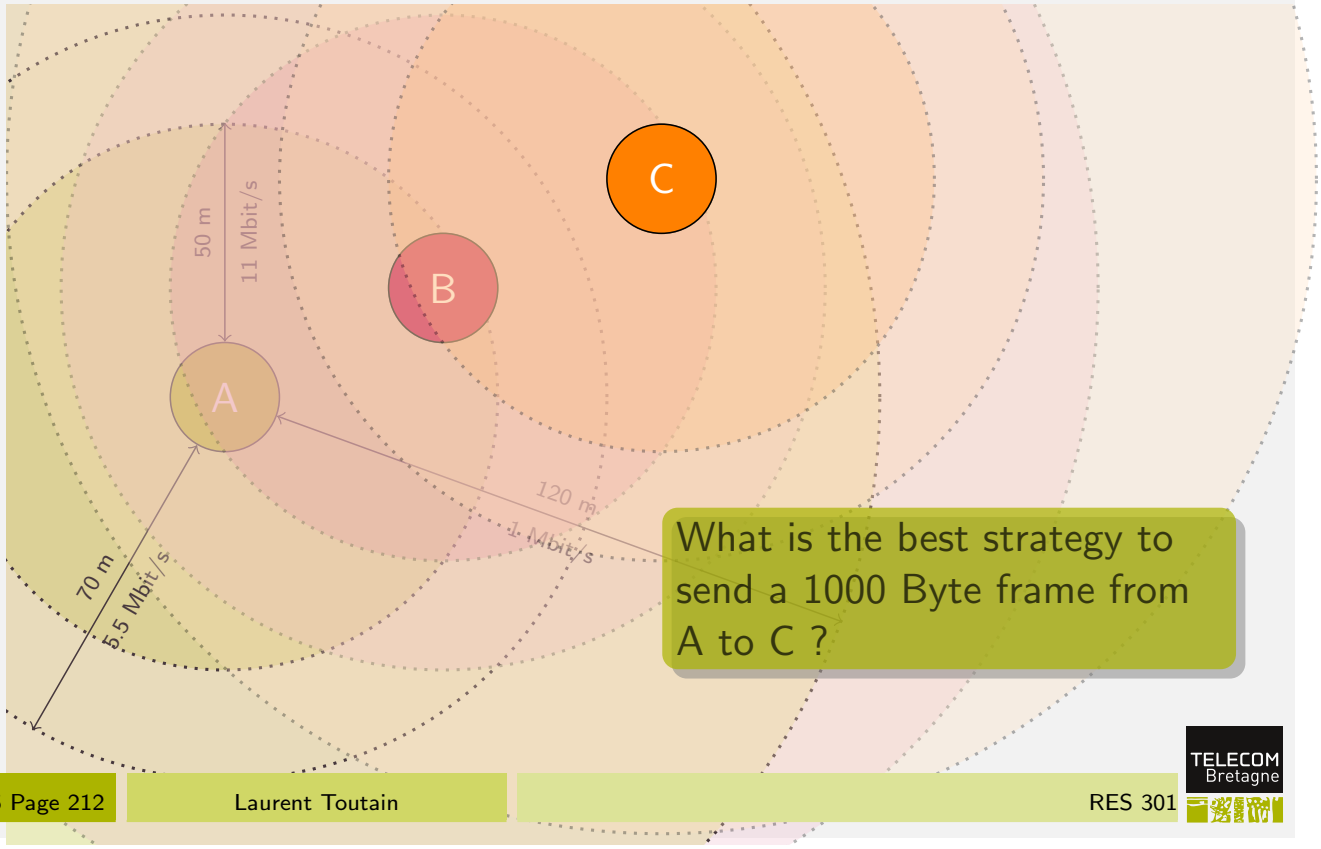
A= B	IEEE 802.11	IEEE 802.11b	IEEE 802.11g
IEEE 802.11			
IEEE 802.11b			
IEEE 802.11g			

Speed versus Distance



http://www.dell.com/downloads/global/shared/broadcom_802_11_g.pdf

Questions



IEEE 802.11n

- OFDM: 48 → 52 subcarriers
  - ⇒ 54 Mbit/s → 58.5 Mbit/s
- FEC: 3/4 → 5/6
  - ⇒ 58.5 Mbit/s → 65 Mbit/s
- Guard Interval: 800 ns → 400 ns
  - ⇒ 65 Mbit/s → 72.2 Mbit/s
- Channel Bounding: 20 Mhz → 40 MHz
  - ⇒ 72.2 Mbit/s → 150 Mbit/s
- 4 spacial streams
  - ⇒ 150 Mbit/s → 600 Mbit/s

http://www.wireshark.ch/download/Cisco_PSE_Day_2009.pdf

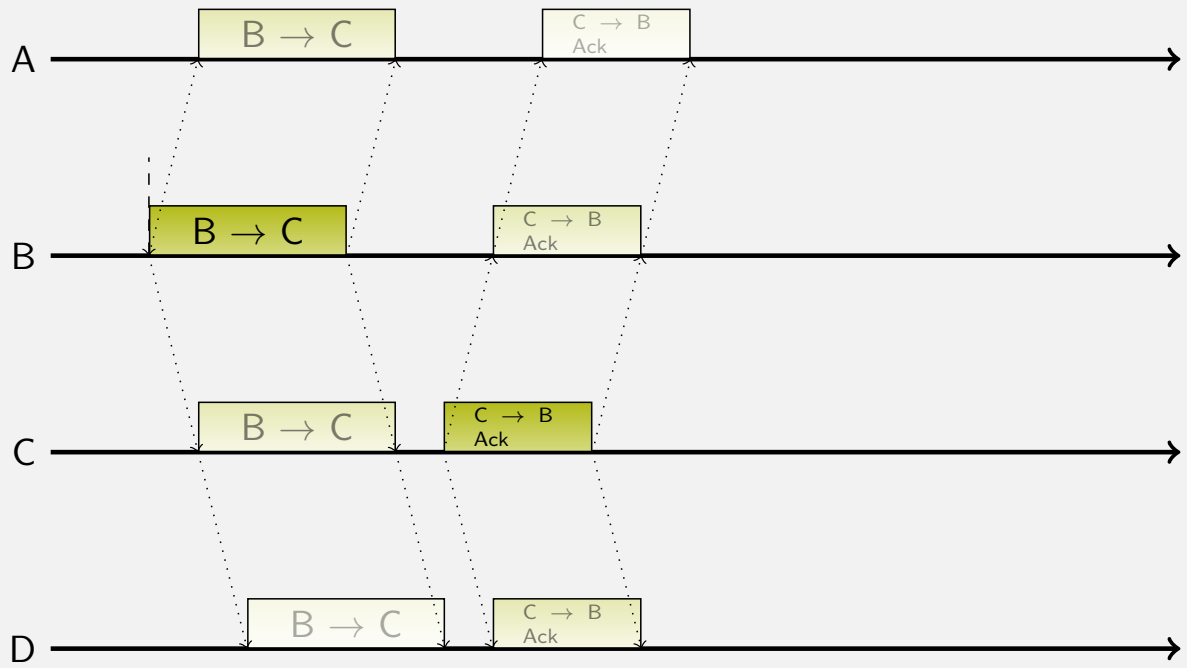
IEEE 802.11

CSMA/CA

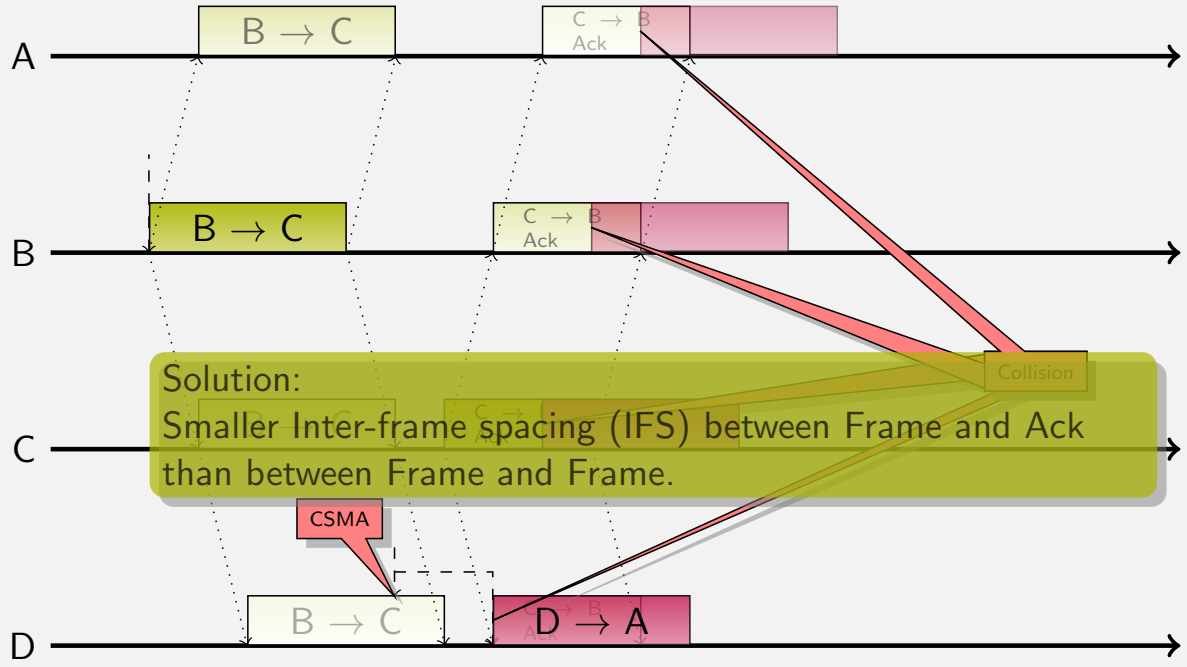
CSMA/CA

- Ethernet uses Carrier Sense Multiple Access/Collision Detect
  - Wait for free channel, if collision stop sending and try later
  - Collision is detected either because received intensity is higher than maximum sending intensity
  - or generated by the Hub
  - work only on wired network
- On radio channel, received signal intensity is small compared to sending intensity.
  - Instead of detecting collision: try to avoid collision and
  - Solution: Acknowledge frames

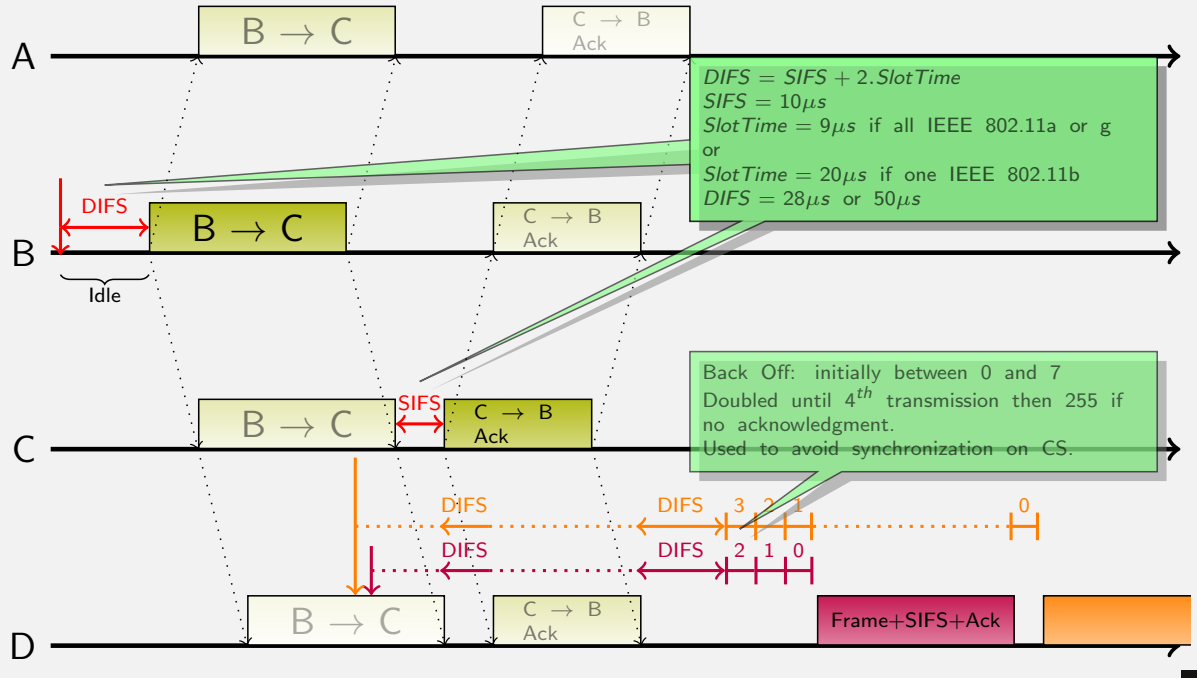
Acknowledgment: what does it change?



Acknowledgment: what does it change?



Acknowledgment: what does it change?



Comments I

Le protocole CSMA/CD (*Carrier Sense Multiple Access/Collision Detect*) défini pour Ethernet ne peut pas être utilisé sur les liens radio. Si sur un réseau filaire, il est relativement aisé de détecter une collision (la puissance du signal reçu est supérieure à la puissance d'émission, montrant ainsi l'émission simultanée de plusieurs trames), cette détection s'avère très difficile car la puissance reçue est très faible comparée à celle utilisée pour l'émission. Comme la détection et la correction par retransmission des trames collisionnées n'est pas possible dans a norme IEEE 802.11, une autre stratégie est mise en œuvre. Il s'agit d'éviter au maximum les collisions et d'utiliser un mécanisme d'acquittement pour être certain que le récepteur a reçu l'information.
Chaque trame en point-à-point est acquittée par le récepteur. Pour éviter qu'une autre station n'émette entre le moment à la trame a été émise et le moment où elle est acquitté, un certain nombre de temporisateurs sont mis en œuvre. Le plus court des temporisateurs SIFS (*Short Inter Frame Space*), il sépare l'émission de la trame de son acquittement. La valeur est de 10 μs avec les technologies IEEE 802.11b/g et de 16 μs pour IEEE 802.11a.

Quand une station cherche a émettre, elle teste le canal pendant une durée DIFS (*Short Inter Frame Space*) supérieure au SIFS; Si aucune activité n'est détectée, la station peut émettre sa trame. La durée de DIFS est données par la formule suivante $SIFS + 2 * SlotTime$, où Slot Time définit une durée permettant de s'assurer que si une autre station a commencé a émettre durant le précédent slot, elle sera détectée par les autre équipement durant ce Slot. Initialement dans les technologies IEEE 802.11 et IEEE 802.11b, la durée du Slot Time était de 20 μs. Avec l'augmentation des débits, cela conduisait à des pertes de performances, avec la norme IEEE 802.11g, ce délai a été réduit à 9 μs si toutes les équipements sont compatibles, sinon le la durée du Slot Time revient à 20 μs. DIFS varie donc entre 28 pour la norme IEEE 802.11g, 34 μs pour la norme IEEE 802.11a àt 50 μs dès qu'un équipement IEEE 802.11b ou antérieur est présent.
La valeur de 2 Slot Time est utilisée car il existe un autre temporisateur utilisé par un coordinateur pour émettre du trafic prioritaire, La durée de PIFS (*Point Coordination Inter Frame Space*) est de $SIFS + SlotTime$. Plus généralement, la norme IEEE 802.11e permet de faire varier la durée du DIFS pour introduire des priorités entre les stations.

Comments II

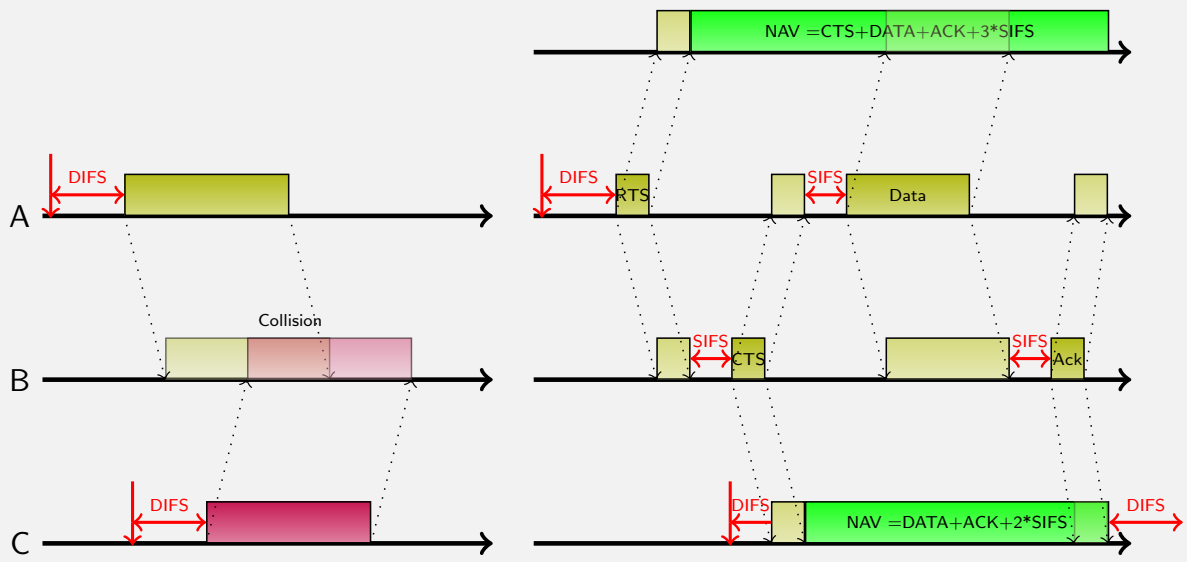
Si par contre une activité est détectée pendant la période DIFS, elle va attendre la fin de celle-ci et attendre a nouveau une durée DIFS. Si aucune activité n'est détectée, elle ne va pas directement transmettre la trame, mais attendre une durée aléatoire (initialement comprise entre 0 et 7 Slot Time) avant d'émettre. Si une activité est détectée durant cette période, le décompte s'arrête. Cette technique permet d'éviter les synchronisations entre stations durant une attente sur une occupation du canal. En effet quand l'activité s'arrête les stations ne vont pas émettre (comme c'est le cas pour Ethernet) mais attendre une durée aléatoire. Si la station ne reçoit pas d'acquittement positif, réitère la procédure en doublant l'intervalle de tirage jusqu'à atteindre au bout de la quatrième tentative 255 Slot Time. Au bout de la sixième tentative, la transmission de la trame est annulée.

Questions

What is the real speed of an empty channel with IEEE 802.11g?

Suppose largest values of DIFS and frame size of 1000 Bytes (Ack frame is 14 Byte long)

Hidden hosts



Disabled by default. If collision rate is too high, activate for frame larger than 500 Bytes. Not implemented on every IEEE 802.11 cards.

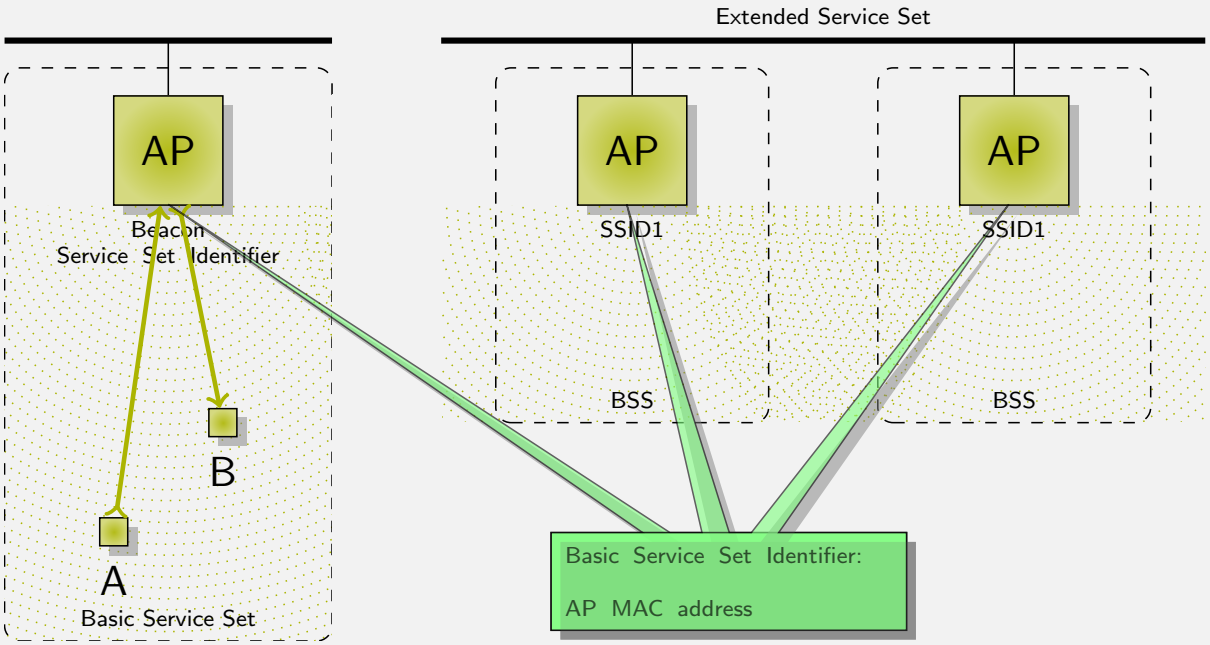
Comments I

Le mécanisme de CSMA/CA peut être également perturbé par des stations cachées (du à un obstacle ou à la limite de porté du signal). Pour éviter que deux stations n'émettent simultanément considérant que le canal est vide et que le signal arrivant à destination ne soit perturbé par l'autre, un mécanisme optionnel de demande et d'autorisation à émettre (RTS: *Request To send*, CTS: *Clear To send*, a été mis en œuvre dans la norme. Ces messages courts ont moins de chance de rentrer en collision que le message de données. Ils transportent la durée d'émission du message. Toute station recevant ce message doit rester silencieuse pendant cette durée.
Ainsi quand, comme dans l'exemple précédent, la station A veut émettre, elle envoi le message RTS, les stations voisine vont rester silencieuse pendant la durée qui correspond à l'émission des messages CTS et de données ainsi qu'à trois temporisations SIFS. Seul le récepteur répond en envoyant le message CTS, ses voisins à leur tour doivent rester silencieux pendant une durée qui correspond à l'émission de données et 2 temporisation SIFS. Cela résout dans ce cas, le problème de la station cachée car C ne pourra plus émettre de données pendant cette durée.

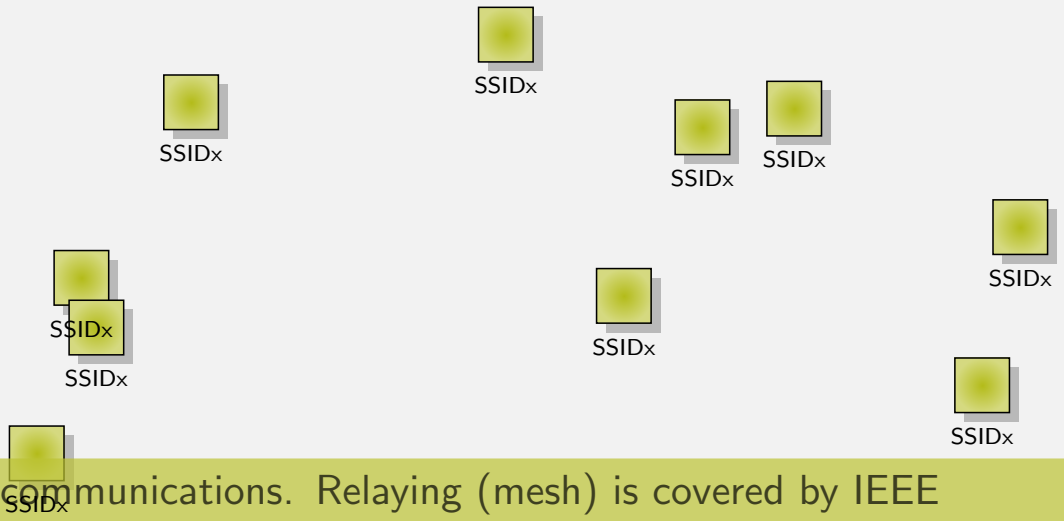
IEEE 802.11

Architecture and Frames

Architectures (infrastructure)



Architectures (Ad Hoc)



Direct communications. Relaying (mesh) is covered by IEEE 802.11s or with OLSR

All node mot be manually configured, or use iBSS

Comments I

ïl existe plusieurs modes de fonctionnement pour les réseaux IEEE 802.11. Dans le mode infrastructure, un point d'accès (AP: *Access Point*) est nécessaire, il agit comme un hub (ou un commutateur) dans le réseaux, les communications passeront systématiquement par cet élément. Le mode infrastructure permet un meilleur contrôle du réseau car il est possible d'y adjoindre des fonctions d'authentification ou de chiffrement. Pour l'utilisateur, il simplifie la configuration car celui-ci reçoit des trames de balisage (textitBeacon) contenant le nom du réseau (SSID (*Service Set Identifier*), en le sélectionnant la station initie un processus d'attachement au réseau. Cet ensemble, AP et station rattachées, forme un BSS (*Basic Service Set*). Du point de vue des stations, l'AP est identifié par son adresse MAC appelée BBSid. Si la couverture d'un AP n'est pas suffisante, la portée du réseau peut être étendue en ajoutant plusieurs BSS annonçant le même SSID, cela forme un ESS (*Extended Service Set*). Les AP peuvent être utilisé également pour émettre du trafic de manière prioritaire, mais cette fonctionnalité est rarement utilisée.

Dans le mode Ad-Hoc, il n'existe pas de point central et les stations s'échangent directement des trames. Il faut tout de fois que les transmission se fassent directement car il n'existe pas dans les normes initiales de fonction de relayage. celles-ci se retrouvent soit au niveau 2 dans l'extension de la norme IEEE 802.11s, soit au niveau 3 dans d'autres travaux comme ceux de l'IETF pour le routage entre les équipements. Normalement chaque équipement doit être configuré manuellement pour entrer dans un réseau ad-hoc. Par contre, la première station entrant dans le réseau peut annoncer un SSID qui permettra aux autres de choisir ce nom pour obtenir la configuration. Ensuite, ces stations propageront le SSID, mais toujours sans routage entre le stations ni point central. ce mode de fonctionnement est appelé iBSS (*independant Basic Service Set*)

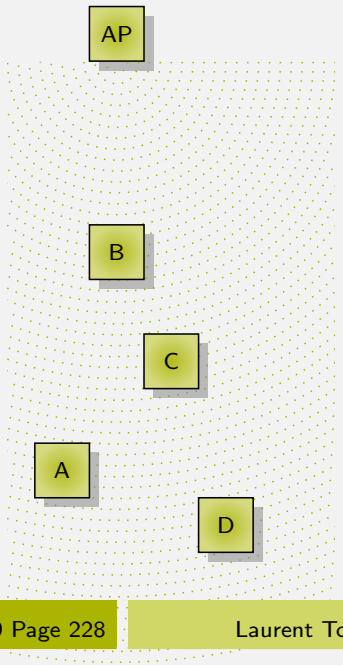
Questions

What happen if B send a broadcast Frame ?

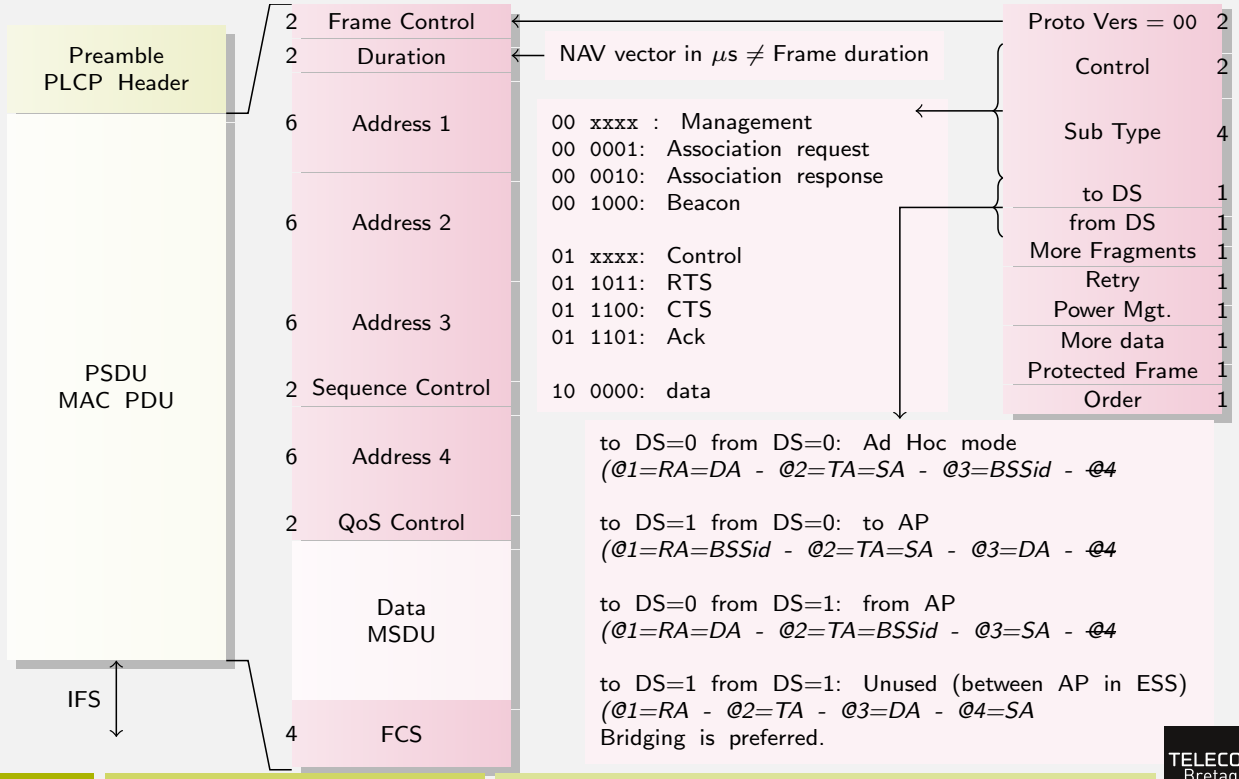
Do we have a guaranty that all receivers got the frame ?

Can RTS/CTS be used ?

At which rate should be sent the multicast frame ?

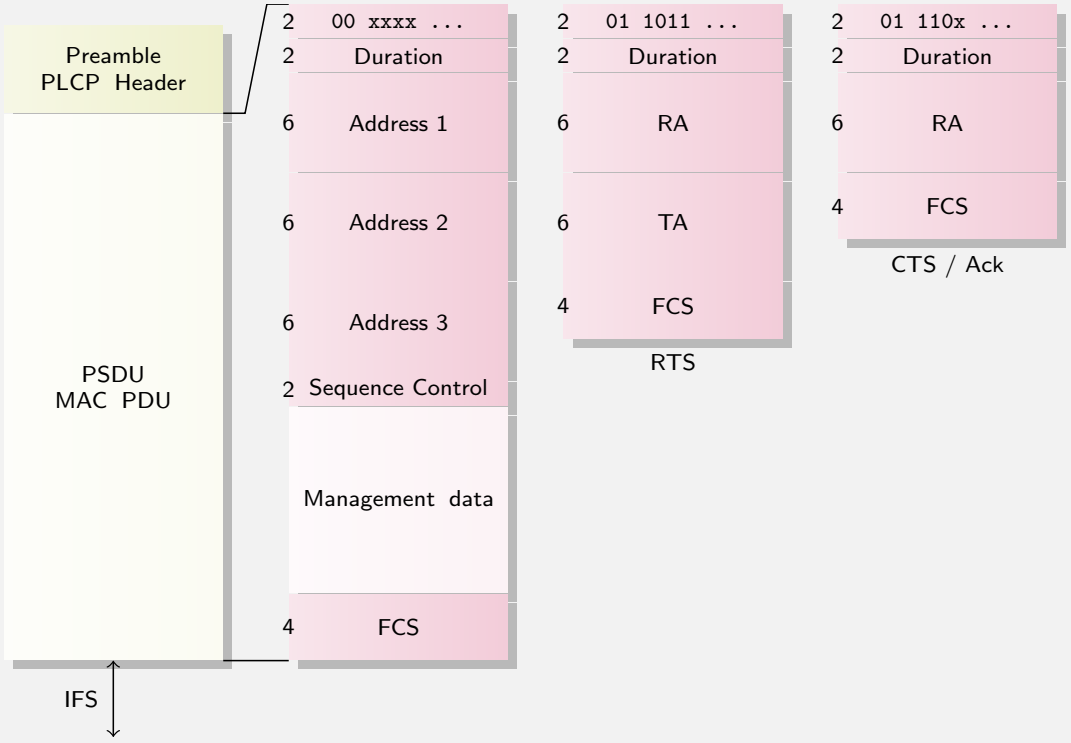


MAC Frame Format

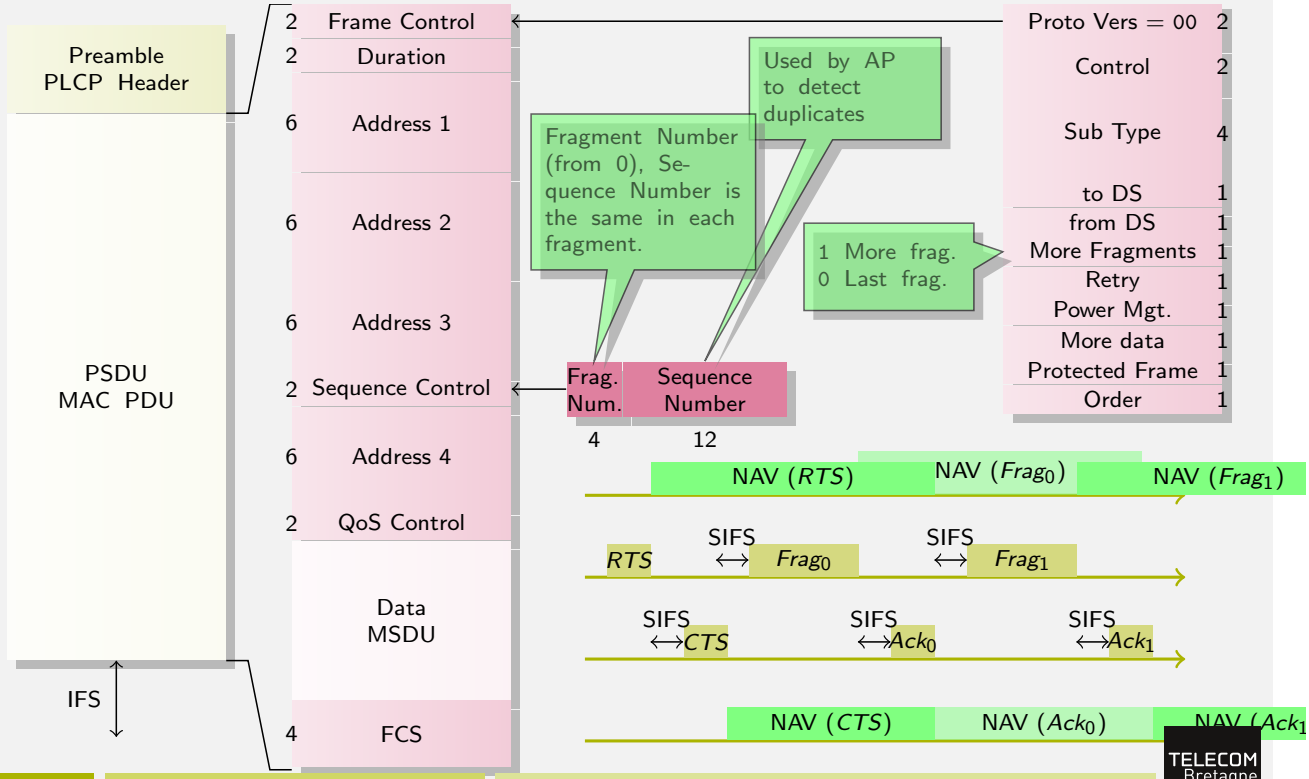


MAC Frame Format

IEEE 802.11 ▶ Architecture and Frames



MAC Fragmentation



Comments I

Une trame MAC commence par un en-tête dont le format va varier (en tout cas le nombre de champs) en fonction de sa nature. (*Frame Control*) contient un certains nombre champs. Après un champ version sur 2 bits qui doit être à 0 dans les versions actuelles de la norme, le champ *Control* permet définir la nature de la trame. La norme en prévoit de trois types, contenus :

- 00: gestion (*Management*). Ces trames vont servir principalement pour gérer la relation entre une station et un point d'accès. Il existe une trame de balise (*Beacon*) qui permet aux stations de trouver les points d'accès, leurs caractéristiques et les SSID qui y sont associés. Les trames de gestion permettent également à une station de s'associer à un réseau, d'authentifier le terminal,...
- 01: control, ces trames vont servir à acquitter une donnée ou par exemple à mettre en œuvre l'algorithme du RTS/CTS.
- 10: Données. Il existe plusieurs types de trames de données suivant le type d'acquittement.

Comments II

Le champ suivant donne le sous type en fonction de la nature de la trame.
Les autres champs seront vu au fur et à mesure dans ce cours.

Le champ *Duration* est différent de celui que l'on trouve au niveau physique dans le PDU PLCP. Ici, il indique la durée du vecteur NAV quand la gestion des stations cachées par l'échange RTS/CTS est utilisé. Ainsi dans le cas d'une trame de contrôle CTS, il inclut la durée d'émission de la trame de données qui a provoquée l'émission de la trame, l'émission de l'acquittement et la durée correspond à 3 SIFS.

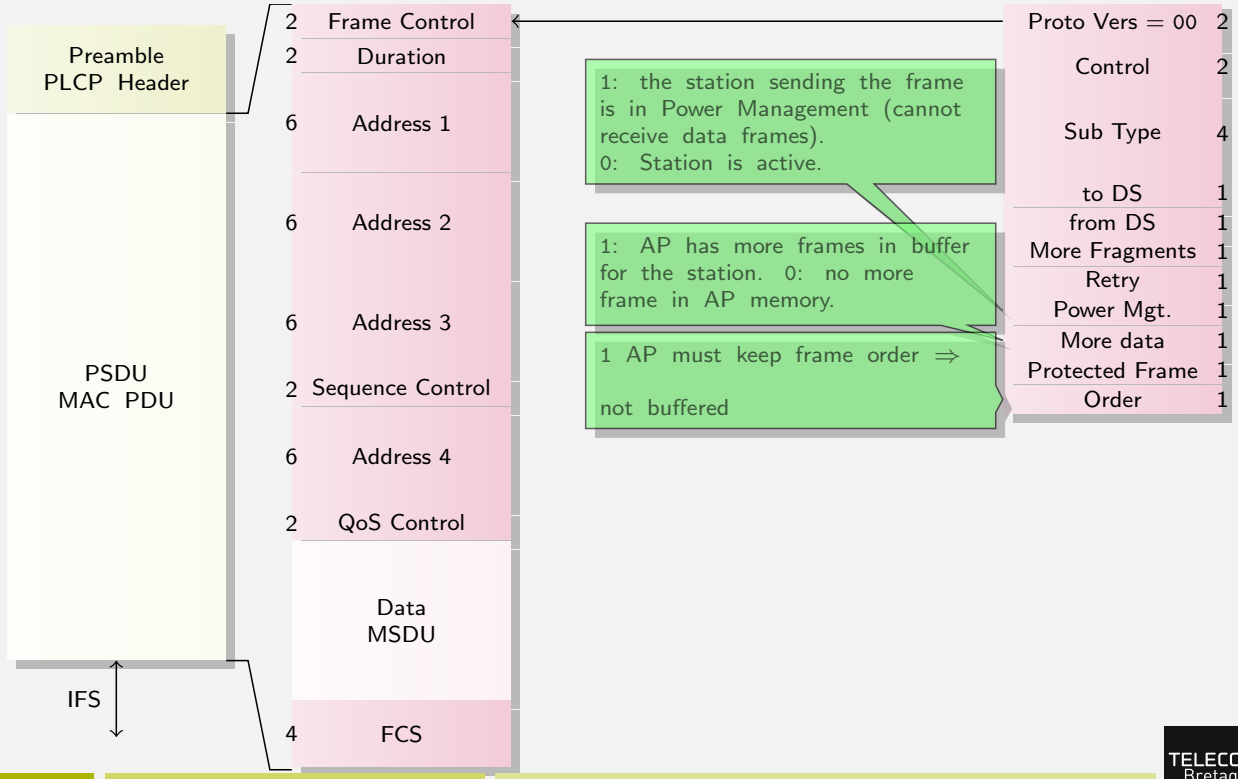
Comme le médium radio est a diffusion, il faut désigner un récepteur du message (RA: *Receiver Address*) qui va recevoir la trame, du réel destinataire (DA: *Destination Address*) qui est le destinataire final du message. De même, TA (*Transmitter Address*) désigne l'équipement qui émet le message et SA (*Source Address*) celui qui en est à l'origine. Dans la cas d'un réseau ad-hoc deux adresses suffisent (RA=DA et TA=SA), mais dans le cas d'un réseau avec un seul point d'accès (BSS), il faut un champ supplémentaire. Quand une station émet une trame, elle l'envoie explicitement au point d'accès (l'adresse est le BSSID)mais le destinataire est généralement différent, par contre TA=SA. Quand le point d'accès relaye la trame, RA=DA, mais la source est différent, l'origine est SA mais l'émetteur est le BSSID est le réel destinataire est DA.
Un quatrième champ est prévu dans le cas d'un réseau composé de plusieurs points d'accès (ESS), dans ce cas en particulier sur le réseau d'interconnexion entre les point d'accès, les quatre adresses sont différentes. En fait, ce dernier mode est peu utilisé car un pontage classique est suffisant et sur le réseau d'interconnexion, il peut transiter que des trames Ethernet grâce aux propriétés d'auto-apprentissage des ponts (cela sera vu en exercice plus tard dans ce cours).
Les bits *toDS* et *from DS* du champ *Frame Control* permettent de spécifier ces différents modes de fonctionnement.

La fragmentation est une autre fonction mise en œuvre par défaut. La longueur maximale des trames est de 0xFFF octets soit 4 095 octets, cette longueur est suffisant puisque, en pratique, Ethernet limite toujours la longueur des

Comments III

trame à 1 500 octets. Si l'on veut transmettre des trames plus grande, ou si l'on doit diminuer la taille des trames, car en cas de fort trafic, les autres stations vont différer leur émission jusqu'à ce que le médium redevienne silencieux, puis suivant les principes de l'algorithme de CSMA/CA, vont attendre un silence d'une durée de DIFS et tirer un nombre aléatoire entre 0 et 7 pour déterminer une attende supplémentaire. Or, si le nombre de station en attente est important, la probabilité que deux stations tirent la même valeur aléatoire et entre en collision n'est plus négligeable. Donc sur des réseaux ayant un fort taux de collision, il peut être intéressant de configurer les stations pour qu'elles diminuent la taille des trames envoyées sur le médium radio.
L'algorithme mis en œuvre par la norme est très simple, quand une station veut émettre une trame de longueur supérieur à la limite, elle la découpe en plusieurs fragments. Le premier fragment est émis (peut être précédé d'un échange RTS/CTS comme sur l'exemple. Le champ *duration* protège la trame en empêchant les voisins d'émettre, l'acquittement permet également de se protéger des stations cachées. Quand la trame de données est émise, le champ *duration* va protéger la trame suivante. Comme les durées inter-trames sont de SIFS, il est impossible à un autre équipement de s'insérer dans cet échange. Un bit *More Fragment* dans le champ *Frame Control* de l'en-tête indique s'il y a de nouveau fragment à émettre (1) ou s'il s'agit du dernier fragment (0).
Le champ *Sequence Number* est utilisé par les stations pour éviter que le point d'accès ne duplique des trames liés à des retransmission. Une station émet chaque nouvelle trame avec un champ *Sequence Number* différent. Si le point d'accès reçoit la même valeur d'un même source, il détruit les duplicats. Pour la fragmentation, les 12 bits de poids faible sont recopiés dans toutes les en-têtes des fragments, par contre les 4 bits de poids fort servent à numéroter les fragments. A noter que le bit *Retry* du champ *Frame Control* quand il est positionné à 1 permet également d'indiquer qu'une trame de données ou de gestion est retransmise.

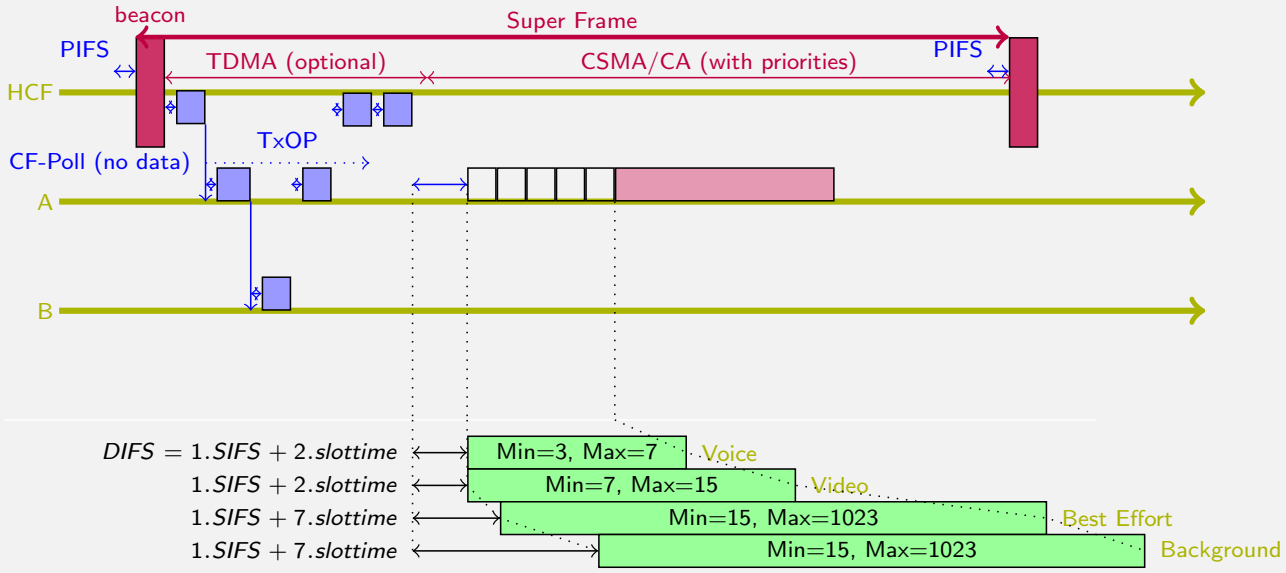
Energy Management



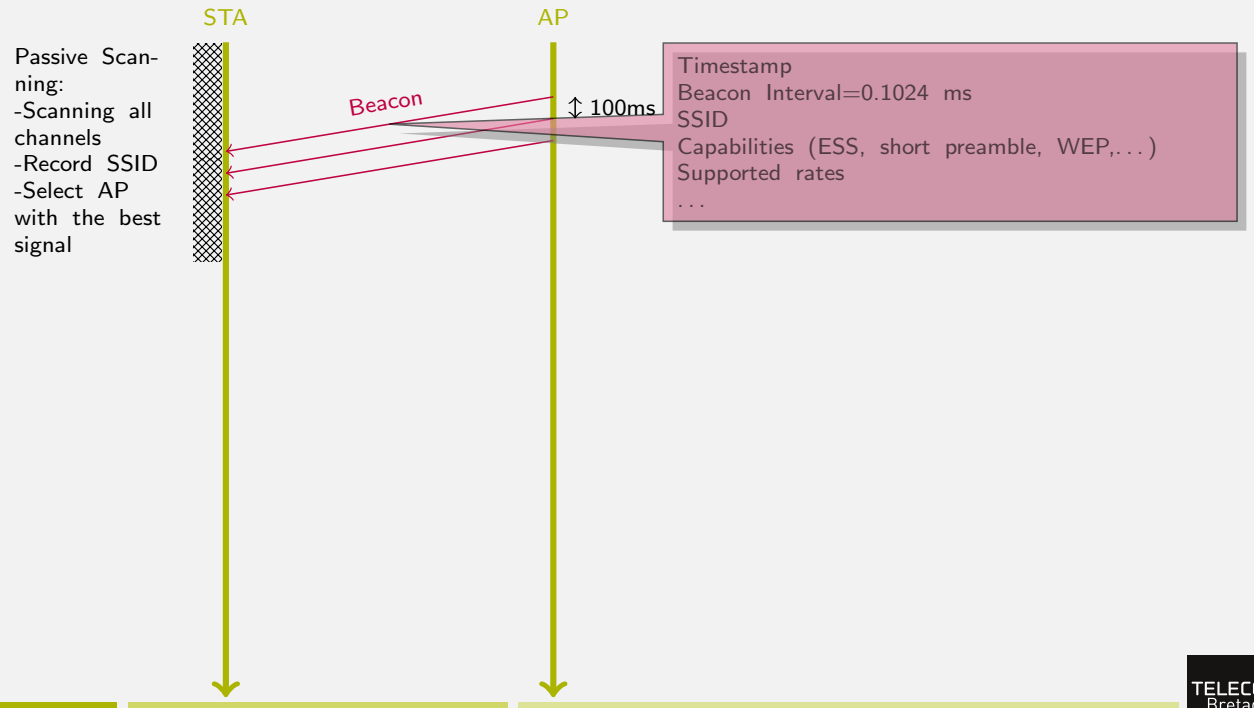
Comments I

La norme IEEE 802.11 prévoit qu'une station station puisse mettre sa carte réseau "en sommeil" soit pour économiser de l'énergie, soit par permettre à la carte de scanner les autres canaux pour découvrir d'autres points d'accès. Quand une carte se met en sommeil, elle prévient l'AP en mettant le bit *Power Management* du champ *Frame Control* en émettant par exemple un trame de données vide. Le point d'accès va mémoriser le trafic vers cette station et émettre périodiquement un message indiquant les stations pour lesquels il a mémorisé des trames. A son réveil la carte demande la transmission de ces trames en attente. Le bit *More Data* permet d'indiquer que l'AP a encore des trames en mémoire pour cette destination.

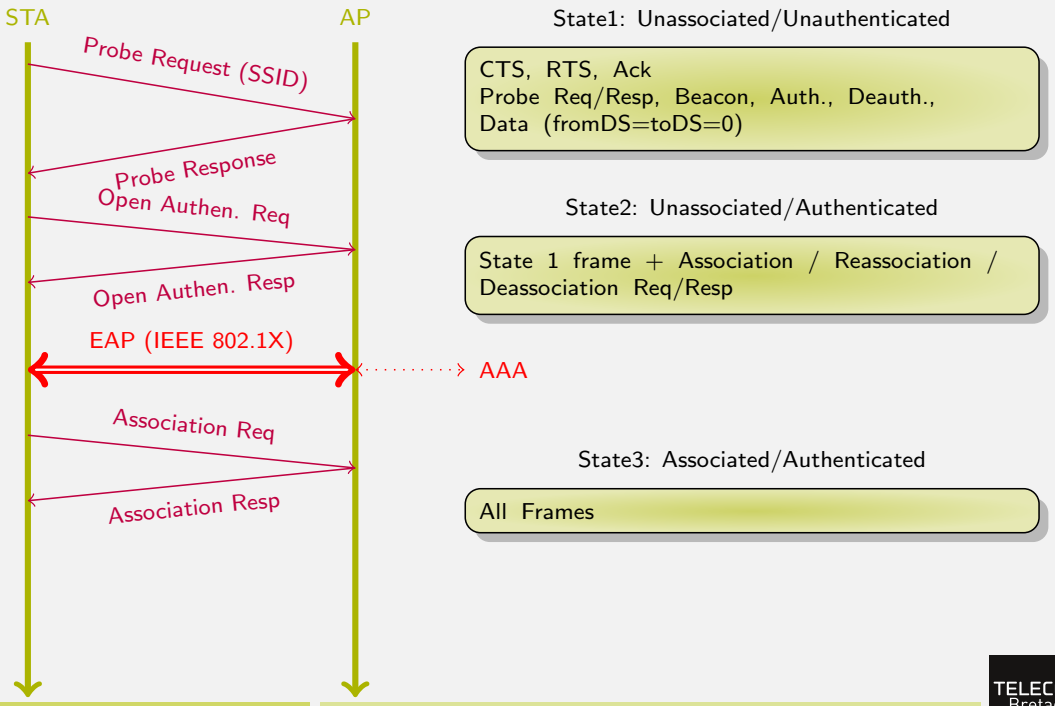
QoS: IEEE 802.11e



Connecting to an AP



Connecting to an AP



References

- Channels a/b/g/b : ˜jian/Course_mc_files/802.11n.pdf -o qrcode6.png"

http://users.cecs.anu.edu.au/˜jian/Course_mc_files/802.11n.pdf

LLC Layer

Introduction

LLC: Logical Link Control

- MAC Layer emulate a point-to-point network
  - MAC algorithm: one sender
  - Destination address: one receiver
- LLC Layer plays the role of the Link Layer in OSI model
  - Detect and correct transmission errors
  - Identify Layer 3 protocol
- LLC has three types:
  - Type 1: Datagram
  - Type 2: Connection oriented
  - Type 3: Acknowledged datagram
- In practice, only type 1 is used
  - Error rate in LAN/MAN/WAN is very low, connection oriented mode is too complex to manage compared to the benefits

Comments I

La sous-couche LLC (Logical Link Control) repose sur la sous-couche MAC. La sous-couche MAC contient des mécanismes pour obtenir une exclusion mutuelle entre les stations qui partagent le même support (bus, anneau). Quand la station a gagné son droit à la parole, la sous-couche LLC contrôle la transmission des données. Trois types de services de transmission sont offerts :

- LLC type 1 ou mode datagramme. Aucune fonction de contrôle d'erreur sur les trames n'est effectuée. La couche LLC aiguille les données vers les différents protocoles de la couche de niveau 3. La grande majorité des protocoles utilisés dans les réseaux locaux utilisent cette encapsulation ;
- LLC type 2 ou mode connecté. En plus des fonctions d'aiguillage du type 1, un contrôle d'erreur du séquencement des données et du flux est effectué. Le protocole est identique à HDLC. Ce type d'encapsulation sera utilisé par exemple pour véhiculer des paquets X.25 sur un réseau local ;
- LLC type 3 ou mode datagramme acquitté. Ce mode de fonctionnement a été ajouté à la norme initiale pour les besoins des réseaux industriels. Il permet entre autre l'acquittement des datagrammes et la réponse automatique.

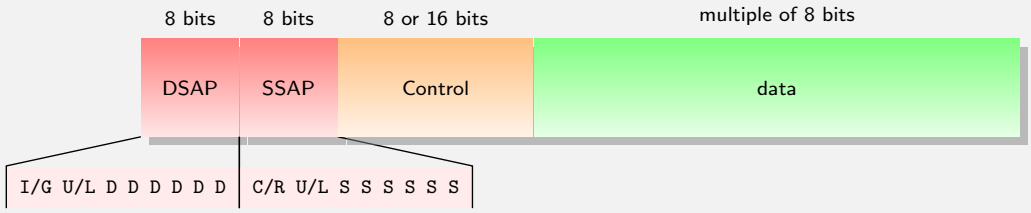
Le principe de fonctionnement de la sous-couche LLC est décrit dans les normes IEEE 802.2 ou ISO 8802-2.

LLC Layer

Frame Format

LLC framing

Based on HDLC framing:



- DSAP: Destination Service Access Point
  - The address of the L3 service that must process the PDU
    - First bit indicates Individual (=0) or group (=1)
    - U=1: standardized by ISO; L=0 local to the LAN
    - in practice only individual are currently used
- SSAP: Destination Service Access Point
  - The address of the L3 service that sends the PDU
    - First bit indicates command (=0) or respond (=1)
    - in practice only individual are currently used
- In practice DSAP=SSAP

SAP Values

hexa	dec.	binary	protocol
0x00	0	0000 0000	Null SAP
0x02	2	0000 0010	LLC Management
0x06	6	0000 0110	~~reserved to IPv4~~
0x0A	10	0000 1010	Secure Data Exchange PDU (IEEE 802.10)
0x42	66	0100 0010	Spanning Tree Protocol
0x7E	126	0111 1110	X.25
0xAA	170	1010 1010	SNAP
0xE0	224	1110 0000	IPX
0xFE	254	1111 1110	ISO, CLNP
0xFF	255	1111 1111	All entities

- Value 0x06 for IP is forbidden by IETF
- Most common values are STP (0x42) and SNAP (0xAA)

Comments I

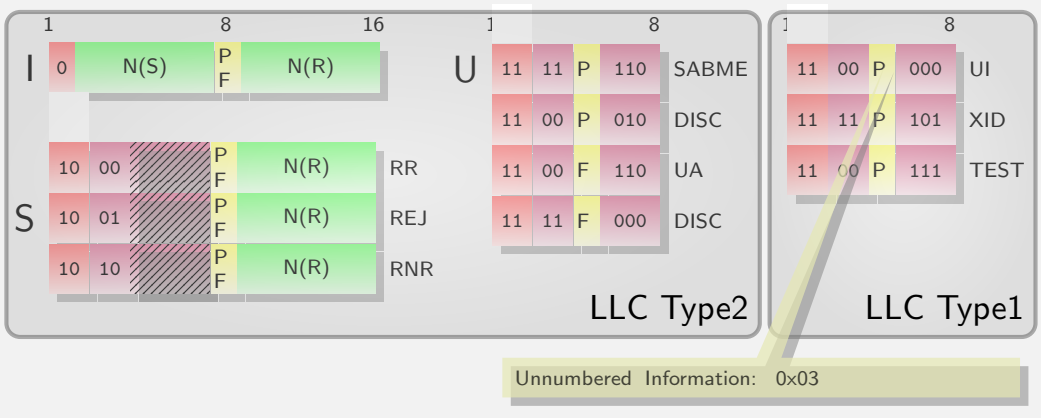
Le champ DSAP (*Destination Service Access Point*) permet de désigner le ou les protocoles de niveau supérieur auxquels seront fournies les données de la trame LLC. Le champ SSAP (*Source Service Access Point*) permet de désigner le protocole qui a émis la trame LLC. Les champs DSAP et SSAP sont codés sur 1 octet. Les 7 bits de poids forts servent à coder les adresses des différents SAP. Le premier bit sert à coder :

- pour une adresse de SSAP si la trame LLC est une trame de commande (bit C/R à 0) ou de réponse (bit à 1),
- pour une adresse de DSAP si la trame est destinée à un SAP unique (bit à 0) ou à un groupe de SAP (bit à 1).

Le deuxième bit sert à coder, si le bit est à 0, qu'il s'agit d'un SAP local attribué par le gestionnaire du réseau, ou, s'il est à 1, qu'il s'agit d'un SAP global, attribué par les organismes de normalisation (IEEE, ISO,...). Le tableau précédent donne les différentes valeurs pour les SAP.

Control field

- As for HDLC, LLC control field defines 3 types of frames:
  - Information Frames: used to carry LLC type 2 data
  - Supervision Frames: used to acknowledge I frames
  - Unnumbered Frames: used to open, close connection and for LLC type 1 and 3.
- Counters are on 7 bits: I and S frames are 2 byte long, U frame 1 byte long



Comments I

Le champ contrôle est codé sur 1 octet pour les trames en datagrammes (trames Non Numérotées). Pour les trames de type I (Information) et S (Supervision) ce champ est codé sur 2 octets. Trois types de trames peuvent être définies (cf. figure 5.4) :

- les trames de type I transportent de l'information utile en mode connecté (c'est-à-dire LLC type 2). Elles contiennent deux compteurs. Le compteur N(S) numérote (modulo 128) les trames émises. Le compteur N(R) permet d'acquitter les trames reçues. Le compteur N(R) contient toujours le numéro de la prochaine trame attendue ;
- les trames de type S permettent la gestion des trames d'information en mode connecté. Il existe 4 bits, non utilisés dans le champ, qui sont réservés à un usage futur. Dans les implémentations actuelles, ces bits sont à 0. Trois trames sont définies parmi les quatre valeurs possibles :
  - RR (Receiver Ready). Cette trame permet d'acquitter les trames déjà reçues. Cette trame est utilisée quand le récepteur n'a pas de données à émettre. La trame RR permet aussi d?indiquer que le récepteur a de la mémoire disponible pour recevoir des trames (contrôle de flux),
  - RNR (Receiver Not Ready). Cette trame est émise par le récepteur pour indiquer qu'il ne peut plus recevoir de trames (problème d'allocation des tampons de stockage). Le compteur N(R) indique toujours le numéro de la prochaine trame attendue,
  - REJ (Reject). Cette trame est utilisée par le récepteur quand le numéro de la trame attendue ne correspond pas à la trame reçue. Le compteur N(R) indique le numéro de la trame à partir de laquelle l'émetteur doit retransmettre ;

Comments II

- les trames de type U ou non numérotées (Unnumbered) permettent de gérer une connexion (ouverture, fermeture) ou d'envoyer des données en mode datagramme. Les sept premières trames de la figure précédente sont utilisées par LLC type 2 et les trois dernières par LLC type 1.
  - SABME (Set Asynchronous Balanced Mode Extented). Cette trame est utilisée pour établir une connexion. Elle ne peut pas contenir de données,
  - UA (Unnumbered Acknowledgment). Cette trame est émise en réponse à une demande d'établissement de connexion. Elle ne peut pas non plus contenir d?information,
  - DISC (DISConnect). Cette trame permet de terminer une connexion précédemment établie avec une trame SABME,
  - DM (Disconnect Mode). Cette trame est émise quand l'équipement distant n'est pas connecté,
  - FRMR (FRaMe Reject). Cette trame est émise en réponse à un dysfonctionnement du protocole (erreur dans l'ouverture d'une connexion, taille du champ donnée trop importante, compteur N(S) ou N(R) invalide,...),
  - XID (eXchange IDentification). Cette trame est émise en commande ou en réponse. Elle permet de tester la présence d'une station (en utilisant le SAP nul), les membres d'un groupe en diffusion, la présence de deux stations avec la même adresse,...
  - TEST. Cette trame est émise en commande ou en réponse. Elle permet de tester un chemin entre deux sous-couches LLC,
  - UI (Unnumbered Information). Cette trame est utilisée en LLC de type 1 pour échanger les datagrammes d'information.
  - AC0 et AC1 (Acknowledged Connectionless). Ces trames en commande ou en réponse sont utilisées par le type 3.

Comments III

Le bit P/F se retrouve dans chacune des trames LLC. Une convention de notation fait que ce bit n'est pas appelé de la même manière suivant que la trame est une requête ou une réponse à une requête (par exemple une trame SABME est une requête et une trame UA est une réponse à cette requête). Pour une requête le bit s'appelle P (pour Poll). L'émetteur de la requête demande une réponse immédiate du récepteur. Dans le cas d'une réponse, ce bit s'appelle F (pour Final), il indique que le récepteur répond bien à la précédente requête. A l'origine du protocole, le droit d'émission était géré par une station maître à l'aide du polling (invitation à émettre). Les stations esclaves pouvaient émettre une suite de trames dont la dernière était marquée Final. Ces noms ont été conservés dans les évolutions du protocole. Il ne faut pas confondre ce bit P/F avec le bit C/R présent dans le champ SSAP. Le bit C/R sert à coder la nature de la trame commande ou réponse, tandis que le bit P/F sert à demander une réponse immédiate dans le cas d'une trame de commande et à donner une réponse immédiate dans le cas d'une trame réponse. Autrement dit, dans le cas d'une trame de commande (bit C/R à 0), le bit du champ contrôle s'appelle P et dans le cas d'une trame de réponse (bit C/R à 1), le bit du champ contrôle s'appelle F.

L'utilisation actuelle de la couche LLC est très simplifiée. Comme les réseaux locaux introduisent peu d'erreur de transmission et que le protocole dominant IP est en mode datagramme, il n'est pas nécessaire de recourir à LLC type 2. Seul le type 1 est utilisé et le champ contrôle indique une trame de type UI (*Unnumbered Information*) dont la valeur est 0x03.

Questions

How many L3 protocol ISO can standardize ?

How many L3 protocol Ethernet can identify ?

What is the length of a LLC type1 header ?

Does Ethernet require LLC encapsulation ?

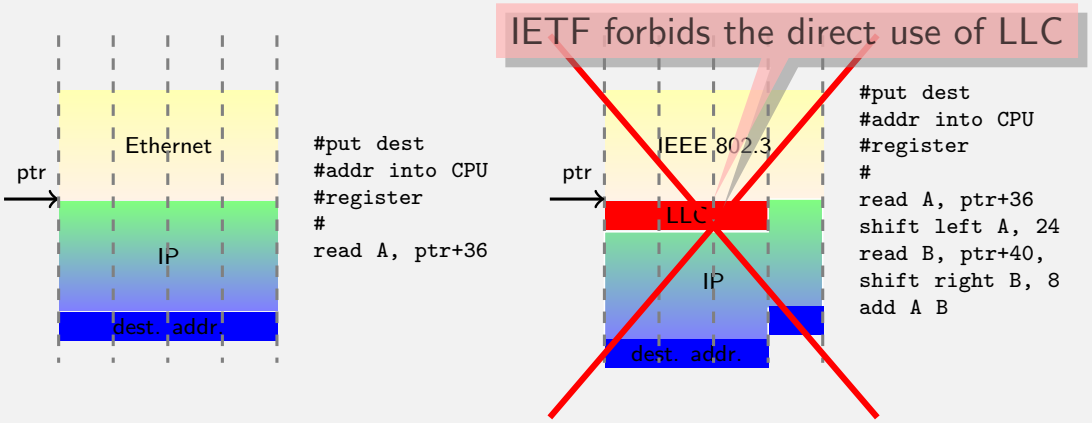
```
0x0000: 0180 c200 0000 0019 e79e 1dae 0026 4242
0x0010: 0300 0000 0000 8000 0002 b972 6601 0000
0x0020: 0008 8005 0019 e79e 1d80 802e 0200 1400
0x0030: 0200 0f00 0000/0000/0000/0000
```

What is the L3 protocol ?

Which type of LLC ?

LLC Drawbacks

- IEEE mandate LLC in its archtecture, but :
  - protocols are not named the same way by Ethernet and LLC
    - and Ethernet is the dominant protocol.
  - LLC type 1 breaks memory alignment
    - 3 byte long
    - IP relies on memory alignment to speed up address processing

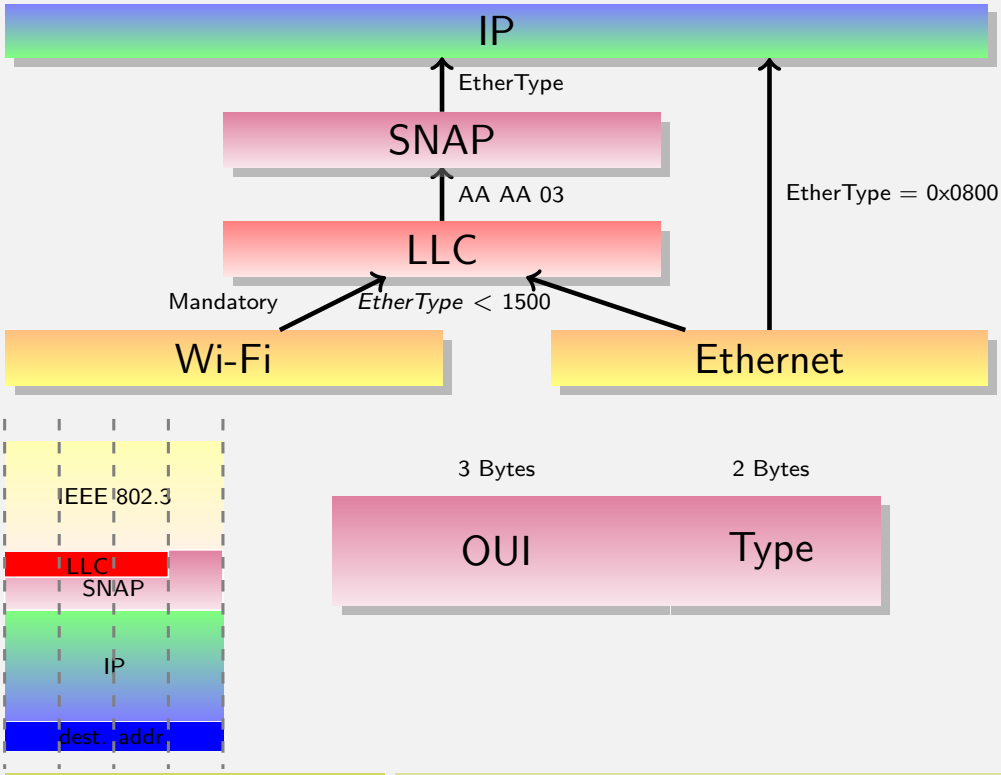


Comments I

L'intégration de piles de protocoles de niveaux 3 et supérieurs, dans des réseaux conformes aux recommandations de l'IEEE, c'est-à-dire avec une sous-couche MAC et une sous-couche LLC, ne se fait pas sans peine. En effet, les protocoles de niveau 3 ont été développés à l'origine pour les réseaux Ethernet qui n'utilisent pas l'encapsulation LLC. Ainsi, avec l'encapsulation LLC, les SAP sont codés sur 7 bits. Il ne peut y avoir que 128 protocoles au-dessus de LLC, ou seulement 64 si l'on ne considère que les adresses globales et plus généralement les SAP de LLC sont différents des EtherType d'Ethernet qui sont codés sur deux octets.

De plus en LLC type 1, la taille de l'en-tête est de 3 octets et pose des problèmes d'alignement qui réduisent les performances de la machine. Ainsi, dans le protocole IP utilisé dans l'Internet, un soin particulier a été pris pour aligner les informations sur des mots de 32 bits (64 pour IPv6). Le décalage introduit par l'encapsulation LLC va rendre plus compliquée l'extraction, par exemple des adresses source et destination. En conséquence, bien que le SAP numéro 6 soit attribué par l'IEEE au protocole IP, son emploi est rigoureusement interdit par l'IETF, l'instance de standardisation de l'Internet. On remarquera que sur le site de l'IANA, cette valeur est indiquée comme réservée à IP et non utilisée par IP!

LLC: dead end ?



LLC Layer

SNAP

SNAP RFC 1042

- Sub-Network Access Protocol
- Mainly used to maintain alignment in equipment memory
  - 5 Byte long + LLC (3 Bytes) = 8 Bytes
- If OUI = 00-00-00 then the last two bytes contains the EtherType
  - Facilitate bridging since protocols are identified the same in Ethernet and LLC networks.
  - Make Ethernet EtherType universal
- OUI can be used by its owner to develop proprietary protocols
  - Cisco Discovery Protocol: OUI 0x00000C protocol ID 0x2000.

Comments I

Le protocole SNAP (RFC 1042) peut être utilisé pour résoudre les problèmes introduit par LLC. SNAP (*Sub-Network Access Protocol*) n'effectue aucun traitement sur les données. Il offre simplement une encapsulation supplémentaire qui se place entre la couche de niveau 3 et la sous-couche LLC.

Un en-tête de trame SNAP fait 5 octets. Avec les 3 octets d'en-tête de la sous-couche LLC, l'encapsulation totale fait 8 octets, ce qui résout les problèmes d'alignement. Les 5 octets codent :

- sur 3 octets l'*Organizational Unit Identifier* (OUI). Il s'agit en principe du code du vendeur fourni par l'IEEE que l'on retrouve au début des adresses MAC. Il permet à un fabriquant de matériel de spécifier ses propres protocoles sans avoir à demander à l'IEEE de EtherType particuliers.
  Plus généralement, ces octets sont souvent à 0 ;
- sur 2 octets le code du protocole. Si les octets du champ précédent sont à 0, ce champ utilise le même codage que la trame Ethernet pour coder les protocoles de niveau supérieur. Sinon le codage est propre au possesseur de l'OUI.

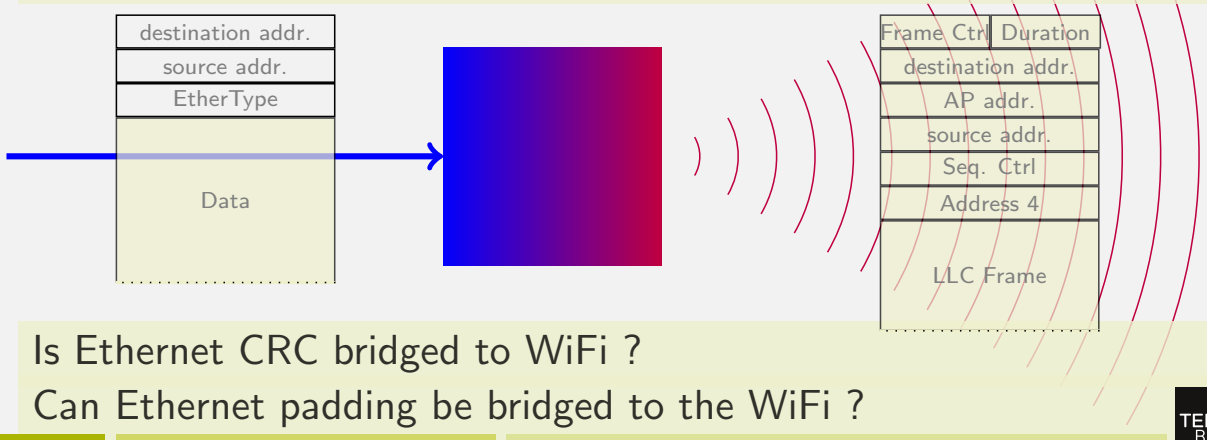
Un DSAP à 0xAA (170) correspond à une trame SNAP. Le champ contrôle code une trame d'information en "datagramme" UI (Unnumbered Information).

L'emploi de SNAP ne se limite pas à résoudre les problèmes conflictuels entre deux organismes de standardisation, il permet de rendre universelle l'encapsulation Ethernet. Ainsi si l'on veut transmettre un protocole de niveau 3 sur un circuit virtuel (X.25, ATM, Frame Relay...), une solution consiste à utiliser l'encapsulation SNAP. Autre avantage, il permet d?avoir un codage uniforme entre un réseau imposant l'encapsulation LLC et un réseau Ethernet, rendant le pontage possible.

Questions


```
0x0000: 000c 295d 3a94 0011 43aa f48b 86dd 600c
0x0010: cef8 0036 1140 2001 0660 7301 0001 0211
0x0020: 43ff feaa f48b 2001 0660 7301 0001 0000
0x0030: ...
```

Describe bridging from Ethernet to a IEEE 802.11. Which fields are added by the bridge/Access Point.



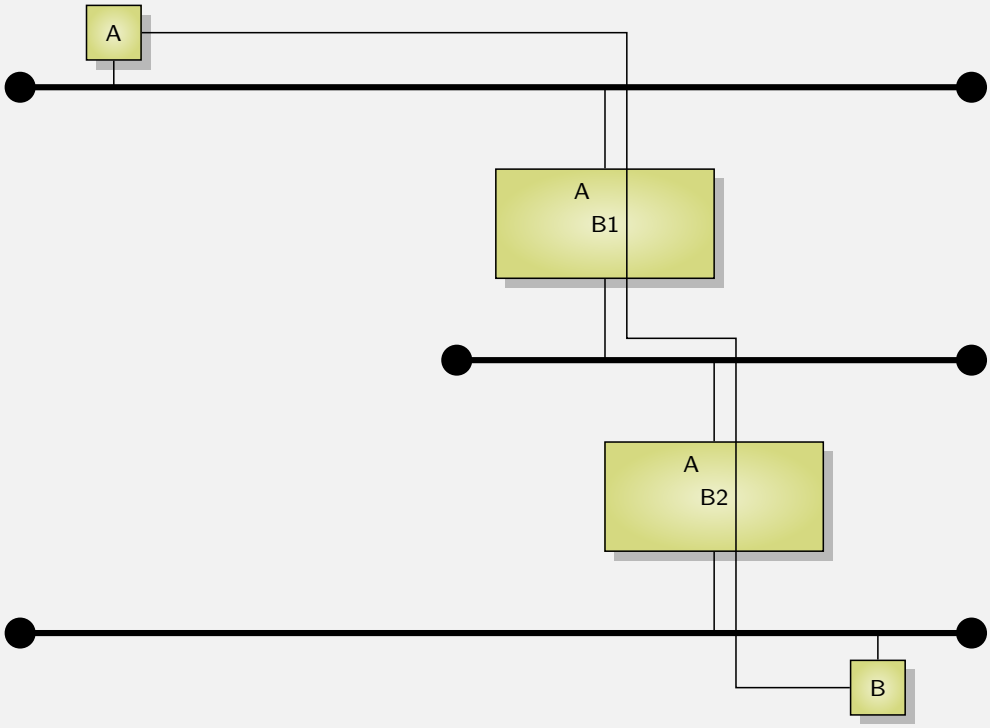
Is Ethernet CRC bridged to WiFi ?

Can Ethernet padding be bridged to the WiFi ?

Spanning Tree

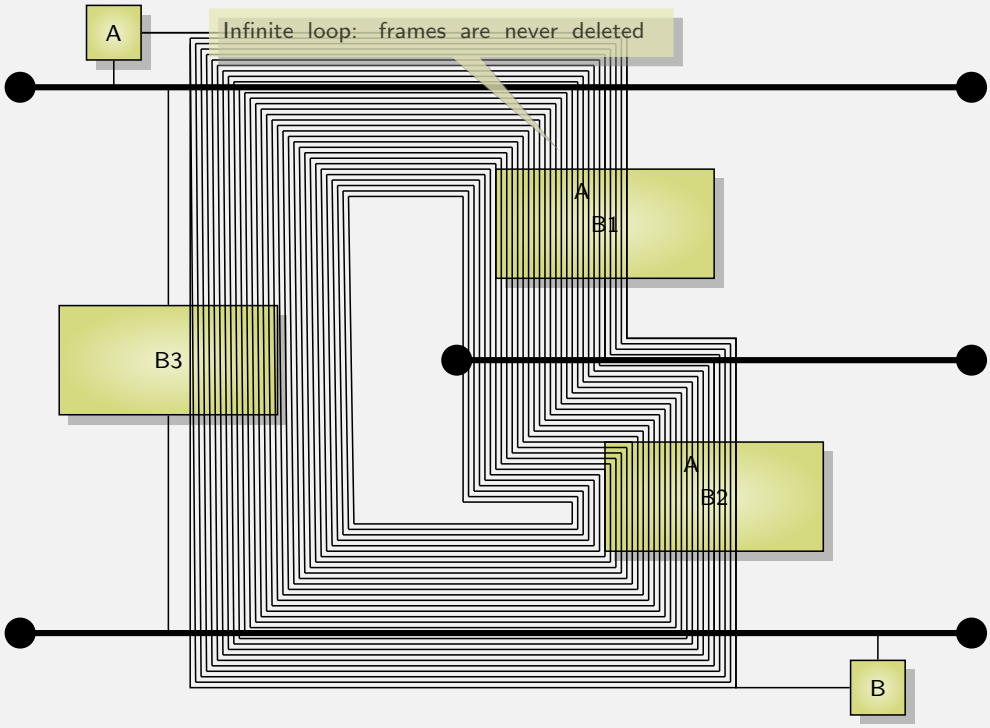
Bridge loops

Several bridges in a network



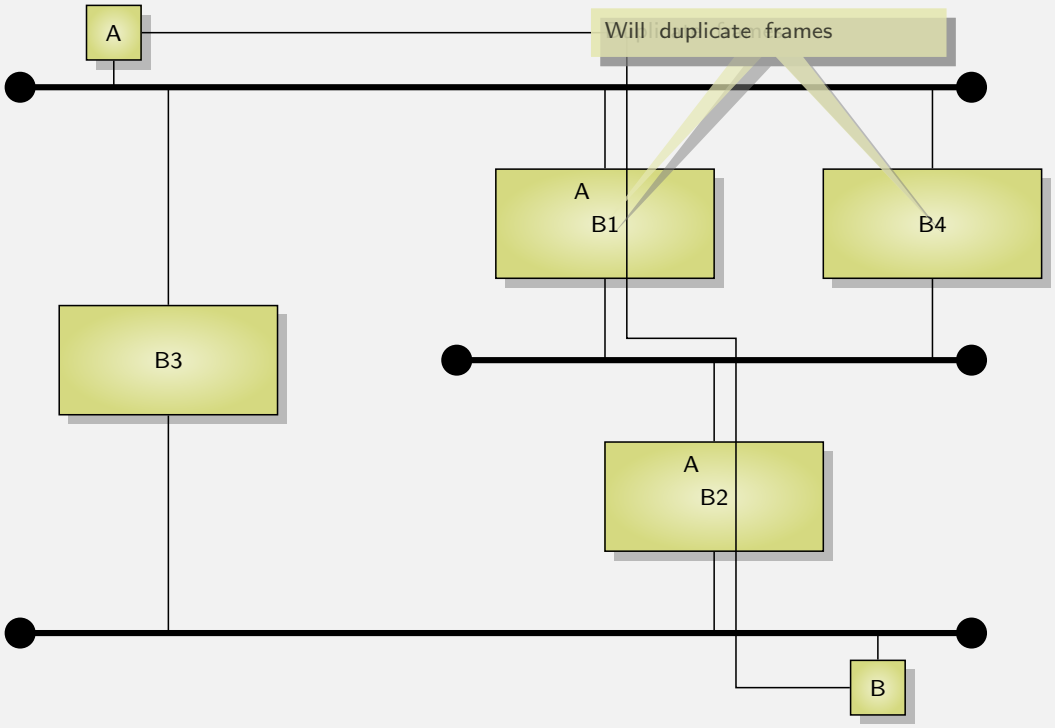
TELECOM Bretagne

Several bridges in a network



TELECOM Bretagne

Several bridges in a network



Spanning Tree

- Bridging works well, if there is a single path to go from one point to another
- When there is several way to go to the same destination, there is a loop.
- Frame will turn forever in the network
  - Since frame are not modified by bridging it is not possible to stop them
  - If several paths, frame can also be duplicated
  - Network will collapse by overloading
- Several possible path = a graph
- Single path = a tree
- The goal is to disable some bridge interfaces to get back to a tree structure
- The tree must cover all the graph: Spanning Tree

Comments I

L'interconnexion redondante de réseaux en utilisant plusieurs ponts pose des problèmes. Les messages arrivent en copies multiples au destinataire. Les ponts ne modifiant pas les trames qu'ils recopient d'un réseau à un autre, une marque ne peut pas être ajoutée dans la trame pour indiquer que celle-ci est déjà passée par ce pont. La recopie ne s'arrête jamais et finalement la bande passante du réseau est complètement saturée. Or la possibilité de multiples interconnexions doit pouvoir être admise dans les réseaux :

- soit pour augmenter la fiabilité : si un pont tombe en panne ou si un sous-réseau est coupé, les stations peuvent toujours communiquer par un autre chemin,
- soit parce que le réseau est complexe et l'ajout par erreur d'un pont ne doit pas pénaliser le trafic.

Il faut revenir au cas qui fonctionnait convenablement. Cela revient à éviter la formation de boucles sur le réseau.

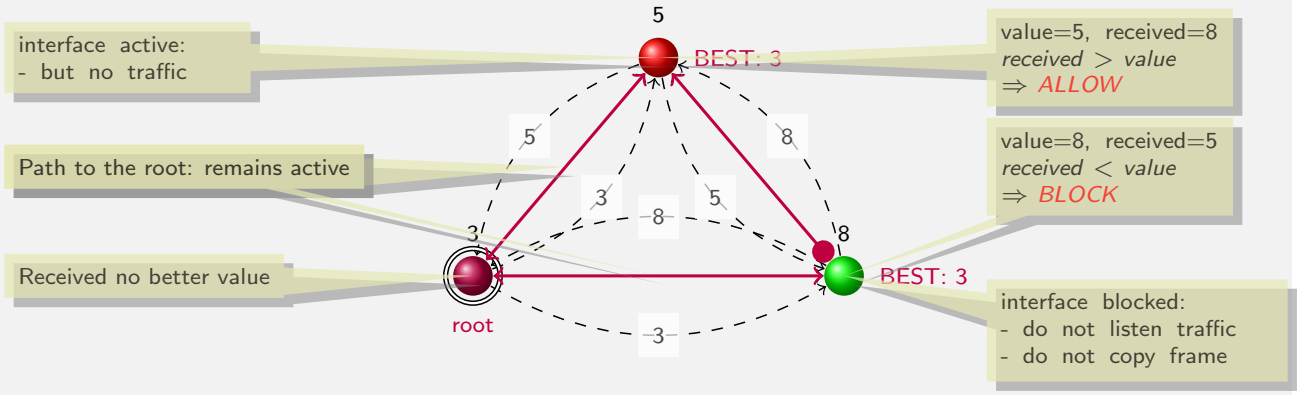
Un ensemble de sous-réseaux interconnectés par un pont peut être assimilé à un graphe. Si on supprime les boucles dans un graphe on obtient un arbre. Si cet arbre passe par tous les arcs, on obtient un arbre de recouvrement total, appelé en anglais *Spanning Tree*.

Spanning Tree

Algorithm

Main idea

- Compute a root (shortest value)
- Take the shortest path to the root
  - Guaranty that all costs will be different
- Disable interfaces leading to a larger cost



Comments I

Le transparent précédent donne une idée intuitive du fonctionnement de l'algorithme du Spanning Tree sur un exemple très simple. Le but est de désactiver des interfaces des ponts pour ne plus écouter ou émettre de trafic. Cet exemple montre clairement une boucle, mais la désactivation d'une interface parmi les 6 permettra de revenir a une structure en arbre.

Dans un premier temps, les nœuds vont échanger leur poids (qui par construction sera unique). Le nœud ayant le poids 3 ne reçoit aucune offre avec une valeur inférieure. Il se déclare racine de l'arbre. Le nœud 5 reçoit sa meilleure offre du nœud 3. Le chemin ayant reçu cette offre devient le chemin vers la racine et sera actif. Même chose pour le nœud 8.

Pour supprimer la boucle la seule possibilité consiste à désactiver une interface sur le chemin qui ne mène pas à la racine. Les nœuds 5 et 8 continue d'envoyer leur poids. Le nœud 5 reçoit une annonce qui est supérieure à la sienne, il bloque l'interface. Par contre le nœud 8 reçoit une annonce inférieure à la sienne, il laisse l'interface active.

On revient à une structure d'arbre, donc sans boucle. Les mécanismes de pontage des trames peuvent donc fonctionner sans problème.

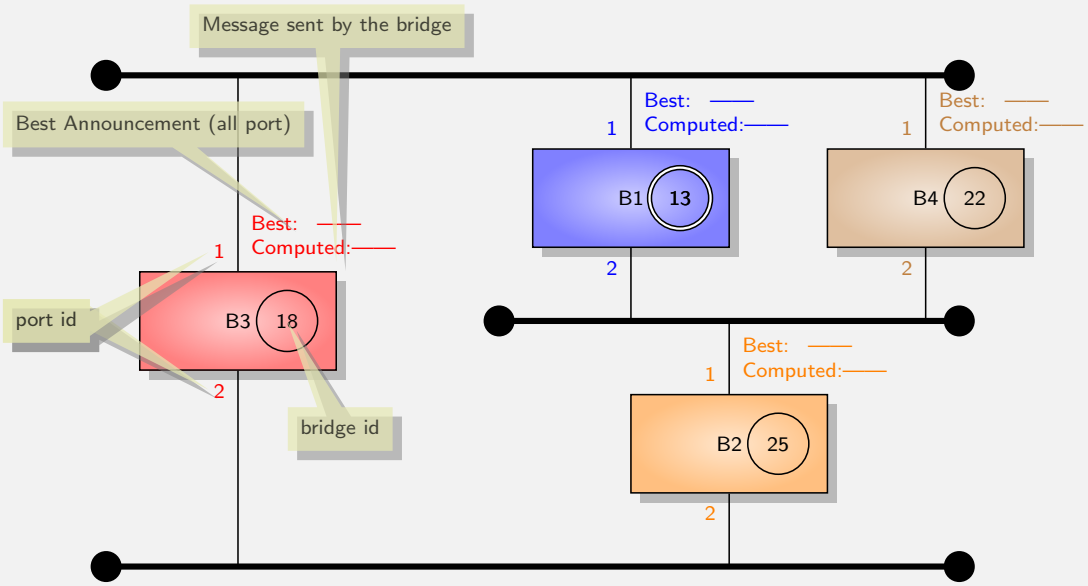
Cost vector

- Each bridge must have an unique id:
  - by construction id are not equal
  - MAC address of one of the interface can be used.
- Cost vector announced by bridge is composed of 4 values
  - Root id: the lowest value the bridge received
    - if no smaller value are received, root id = bridge id
  - Cost to the root: used to select the smallest path to the root
  - Bridge id
  - Port number or interface number
- $V = \langle V_1, V_2, V_3, V_4 \rangle < U = \langle U_1, U_2, U_3, U_4 \rangle$
  - if $V_1 < U_1$ or
  - if $V_1 = U_1$; $V_2 < U_2$ or
  - if $V_1 = U_1$ and $V_2 = U_2$; $V_3 < U_3$ or
  - if $V_1 = U_1$ and $V_2 = U_2$ and $V_3 = U_3$; $V_4 < U_4$

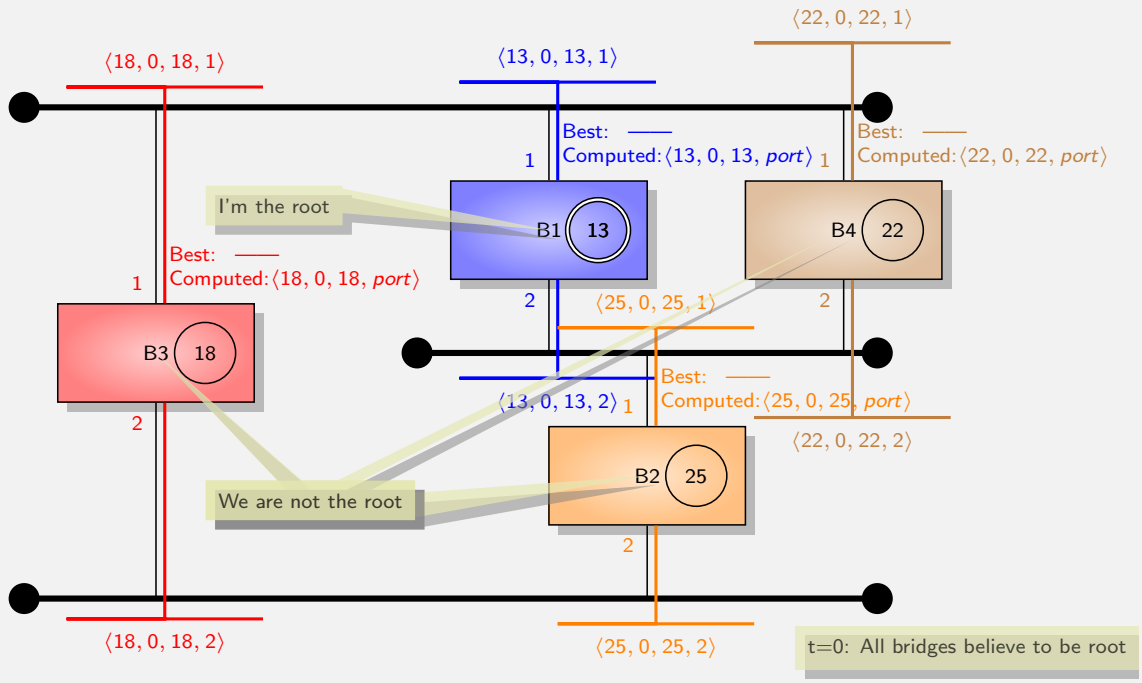
Algorithm overview

1. Each bridge/switch believes to be the root
2. When receiving a message smaller all the bridges except one will not be the root
   - 2.1 The port on which the smallest (best) message is received becomes the root port
   - 2.2 Other port must be active (Designated) or passive (no forwarding: Alternate) ports
     - if a smaller message than the one generated by be bridge is receive the port is alternate
     - if the bridge generate the best message for the link, the port is designated.
3. if the topology changes, the bridge informs the root:
   - 3.1 the root informs all the bridges of a topology change
   - 3.2 all bridges discard their forwarding table (i.e. MAC address location)
   - 3.3 all bridges start again learning process (and continue to forward).

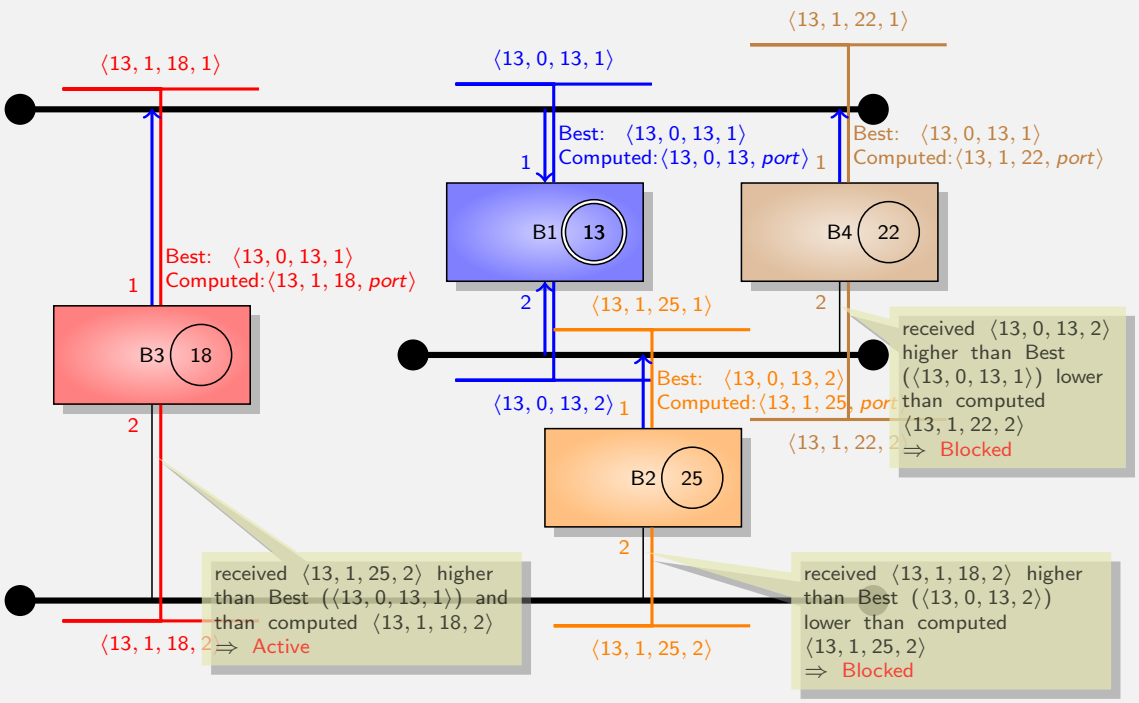
Example



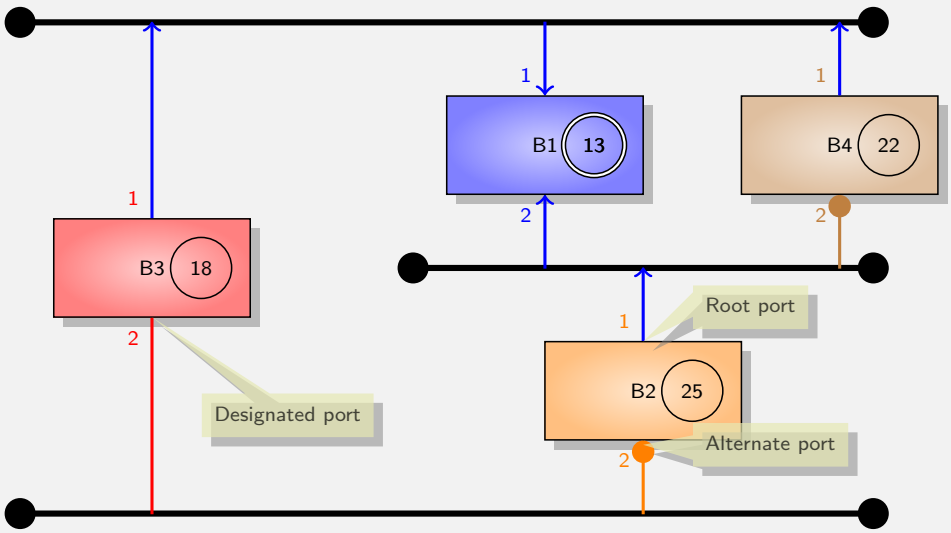
Example



Example



Example



Comments I

L'algorithme de Spanning Tree est relativement simple à mettre en œuvre. Il consiste à élire un pont particulier appelé racine et à choisir un chemin unique entre les ponts et la racine. Contrairement aux ponts transparents qui ne nécessitent pas d'adresse MAC, l'algorithme du Spanning Tree requiert que chaque pont ait une adresse sur le réseau pour échanger des messages conduisant à l'élection de la racine.

Chaque pont, en plus d'une adresse MAC, possède un identificateur. L'attribution de bonnes valeurs aux identificateurs par l'ingénieur réseau permet d'influer sur l?arbre couvrant pour qu'il utilise au mieux les ressources du réseau. La réalisation de l'arbre couvrant va se faire en désactivant (inhibant) certains des ports des ponts. Ainsi les boucles seront éliminées sur le réseau. L'envoi des messages appelés BPDU (pour le terme anglais : Bridge Protocol Data Unit) se fait en utilisant une adresse de multicast de niveau MAC. Cette adresse réservée à l?algorithme du Spanning Tree dispense l'administrateur du réseau d'indiquer les autres ponts voisins. Le pont les découvrira en écoutant le réseau.

Les ponts vont échanger des messages contenant :

- l'identificateur supposé de la racine par le pont émetteur du message,
- le coût de la liaison entre le pont et la racine supposée, c'est-à-dire le nombre de ponts qu'un message devra traverser pour atteindre la racine. Pour un pont se supposant être la racine, le coût est nul,
- l'identificateur du pont émetteur,
- le numéro du port sur lequel le message est émis.

Une configuration est meilleure qu'une autre si :

- l'identité de la racine est la plus petite,

Comments II

- en cas d'égalité sur l'identité de la racine, le coût du chemin vers la racine est plus petit,
- en cas d'égalité des deux premiers champs, l'identificateur de l'émetteur du message est le plus petit,
- finalement, si les trois premiers champs sont identiques, le message a été émis sur le port ayant le numéro le plus petit.

Intuitivement, cet ordre de priorité se comprend aisément. Le consensus va se faire pour désigner le pont racine comme le pont ayant le plus petit identificateur. Si plusieurs chemins sont possibles pour aller vers la racine, le chemin le plus court sera choisi. S'il existe plusieurs possibilités pour aller vers la racine avec le même coût, arbitrairement, le chemin proposé par le pont ayant le plus petit identificateur sera choisi. Si ce pont propose plusieurs fois le même chemin (cas où plusieurs ports sont connectés sur le même sous-réseau), le port le plus petit sera choisi.

A l'initialisation d'un pont, celui-ci se considère comme racine, il émet périodiquement sur chacun de ses ports un message $\langle id, 0, id, n^{o}\ du\ port \rangle$. Si, sur un des sous-réseaux, auquel il est connecté, circule un message contenant une meilleure configuration :

- la voie par laquelle cette meilleure configuration a été reçue devient le chemin pour la racine,
- une nouvelle configuration est calculée. Le premier champ reprend le champ du meilleur message, le coût est augmenté de 1. Les deux derniers champs ne sont pas modifiés. Cette nouvelle configuration sera émise périodiquement.

Comments III

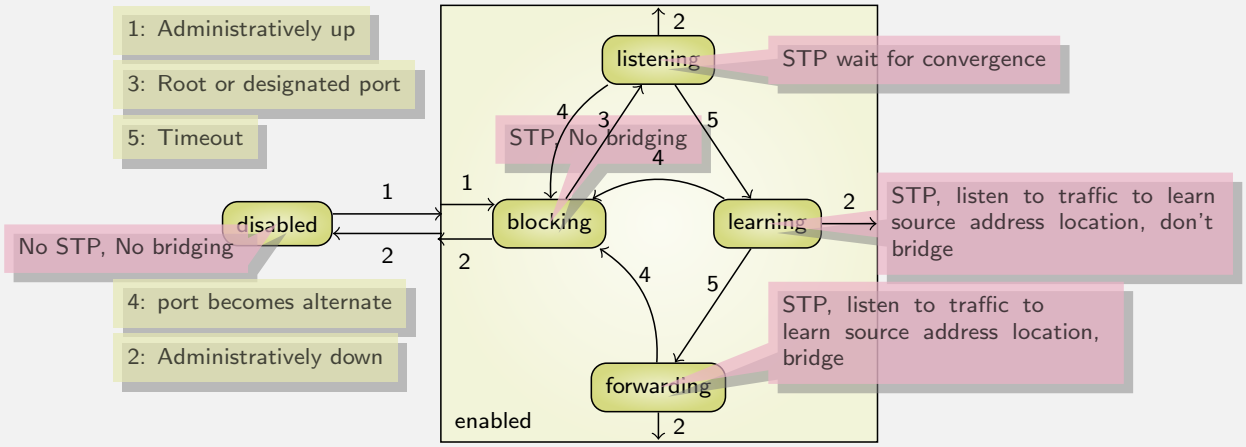
Le pont détermine ensuite quels ports ne menant pas à la racine doivent être activés ou inhibés en transmission. Il regarde le meilleur message de configuration reçu sur chacun de ses ports. Si le meilleur message de configuration pour un port donné est compris entre la meilleure configuration reçue (vrai par construction de l'algorithme) et la configuration calculée, alors le port est inhibé. Par contre, si aucun des messages reçus sur un port n'est inférieur à la configuration calculée, alors le port reste actif.

Spanning Tree

Rapid STP

Interface states

In STP, an interface may have different states:



Timer 5 is about 15 sec + up to 40 sec to detect failure ⇒ more than 1 minute to forward frames.

Comments I

Un port est dans l'état désactivé (*Disabled*) s'il n'est pas connecté à un sous-réseau ou si l'administrateur du réseau l'a explicitement configuré ainsi. Un port dans cet état ne participe ni à la retransmission des trames, ni à l'algorithme du Spanning Tree.

Dans l'état activé, déclenché par configuration par l'administrateur du réseau (action 1 sur le schéma précédent), le port va participer à l'algorithme du Spanning Tree qui permettra de déterminer s'il activera sur ce port à le relayage des trames. Un port qui ne participe pas au processus de relayage, ne peut ni recopier en mémoire les trames qui circulent sur le sous-réseau auquel il est attaché, ni émettre les trames que d'autres ports ont reçues. On distingue quatre sous-états qui peuvent être quittés à n'importe quel instant pour revenir dans l'état désactivé, sous l'action de l'administrateur réseau ou la déconnexion du réseau (action 2):

- dans l'état bloqué (*blocked*), le port ne participe pas à l'interconnexion, il participe au déroulement de l'algorithme du Spanning Tree ;
- dans l'état écoute (*listening*), le port a reçu des messages de configuration qui l'ont fait passer de l'état bloqué à cet état (action 3). Il sait que son port sera potentiellement un port vers la racine ou un port désigné. Le pont ne participe toujours pas à l'interconnexion, ceci pour éviter la formation de boucles qui pourraient survenir dans l'état transitoire. Le fait que le port participe de nouveau à l'interconnexion des réseaux est probablement dû à la défaillance d'un autre pont sur le réseau ou à la coupure d'un lien. L'algorithme privilégie une rupture de la connectivité pendant une courte période. Le pont n'écoute pas non plus le trafic pour former une base de données d'adresses car pendant la période transitoire, la localisation des émetteurs peut changer. Dans cet état (mais également dans les autres) le port peut imédiattement retourner dans l'état bloqué (action 4) si l'algorithme du Spanning Tree reçoit des messages de configuration contraires indiquant que l'état du port est alternatif ;

Comments II

- dans l'état apprentissage (*learning*), le port ne participe toujours pas à l'interconnexion, mais écoute le trafic pour repérer les stations qui lui sont rattachées. Cette période d'apprentissage permet au port, lorsqu'il deviendra actif, de ne pas inonder le réseau de recopies inutiles dues à une absence de l'adresse de l'émetteur dans la base de données de filtrage.
- dans l'état relayage (*forwarding*), le pont participe à l'interconnexion des réseaux tout en continuant a dérouler l'algorithme du Spanning Tree.

Network recovery

- Each bridge send periodically a ST message
  - interval: between 1s and 10s (recommended: 2s)
- Each message contains an age field giving the time the root has been seen.
- If age reach a limit (between 6s and 40s (recommended 20s)), this implies a connectivity problem between the bridge and the root.
  - Spanning Tree construction is restarted (start in the blocking state)
  - new root may be selected.
- Network may take about 1mn to recover:
  - Up to 20s to detect a path to root failure
  - 15s to move from learning
  - 15s to move to forwarding state.

Comments I

En cas de panne d'un équipement ou d'une liaison, le temps de passage de l'état bloqué à l'état relayage est relativement long puisque les temporisateurs (action 5) sont configurés avec des durées de 15 secondes et avec le paramétrage par défaut du Spanning Tree, il faut jusqu'à 20 secondes pour qu'un pont détecte une rupture de connectivité vers la racine.

A l'origine du protocole, cette longue rupture de connectivité était considérée comme salutaire comparée aux dommages des boucles. Mais avec le développement de nouvelles applications multimédias ou de voix sur IP, cette perte de connectivité peut avoir des conséquences graves pour les applications. Pour réduire ces durées, l'algorithme du Spanning Tree a été adapté pour réagir plus vite. Il a été défini par le groupe IEEE 802.1w sous le nom de Rapid Spanning Tree Protocol (RSTP).

Slow Convergence

- When SPT has been developed, preventing loops was more important than time to recover when a bridge or a link fails.
  - File transfer, mail,. . . are not real time
  - It may take more than a minute to recover
- Nowadays with real time protocol such as VoIP or streaming these delays are not acceptable
- Rapid Spanning Tree Protocol developed by IEEE 802.1w is backward compatible with STP.
  - if 3 consecutive ST messages are lost, ST topology is restarted
    - up to 6s to restart Spanning Tree construction
- Adapt to switched technologies
  - Link to other switches
  - Link to hosts (No Bridge PDU).

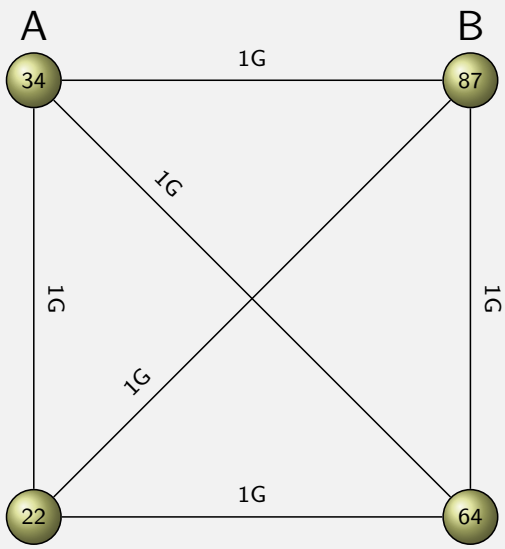
Comments I

La principale modification concerne la détection de la perte de connectivité avec la racine. RSTP considère à la place la perte de connectivité avec le pont qui conduit vers la racine. Chacun des ponts émet une message de Spaning Tree (Bridge PDU ou BPDU) toutes les 2 secondes. Si un pont perd 3 BPDU consécutifs, il lance une phase de reconfiguration avec recherche d'une nouvelle racine. L'algorithme convergeant rapidement que quelques secondes pour reconstruire l'arbre couvrant. De plus, il ne repasse pas par les phases d'écoutés et d'apprentissage. La perte de connectivité est donc réduite à quelques dizaines de secondes.

RSTP utilise aussi les propriétés des réseaux construit sur des architectures commutées où il existe que deux types

de liaisons: celles entre deux commutateurs, celles entre un commutateur et un équipement terminal.

Questions



Which router becomes root?

Which ports are root ports?

Which ports are alternate ports?

Will a frame from A to B follows the optimal path?

Are all the link used ?

What happen if root bridge links were 10 Mbit/s links? How to solve this?

Alternatives to Spanning Tree

- Spanning Tree can be viewed as a fuse to protect network but
  - Traffic Engineering is quite complex
  - The shortest path is not always used
  - Some links are not used
- Alternatives based on routing protocol are currently standardized:
  - IETF TRILL (Transparent Interconnection of Lots of Links)
    W http://datatracker.ietf.org/wg/trill/
  - Shortest Path Bridging (SPB) (IEEE 802.1aq)
    W http://www.ieee802.org/1/pages/802.1aq.html

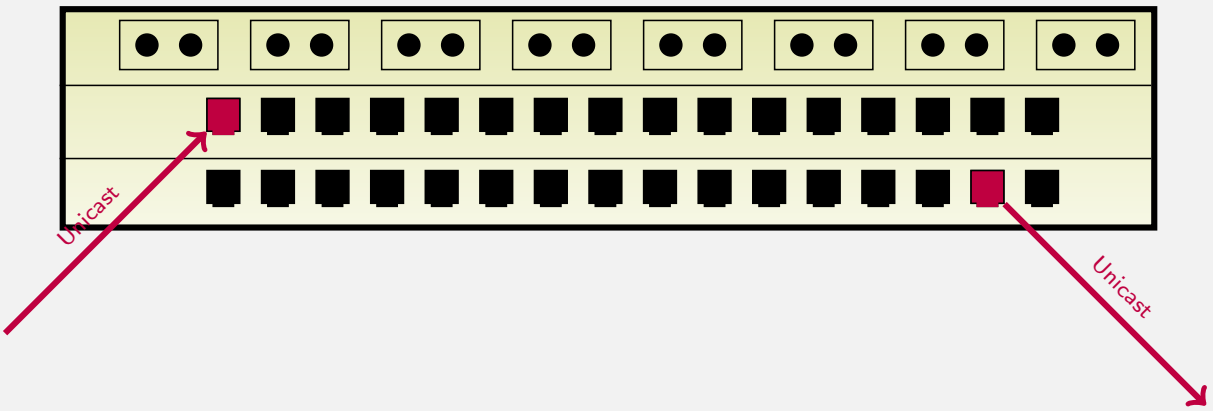
VLAN

Introduction

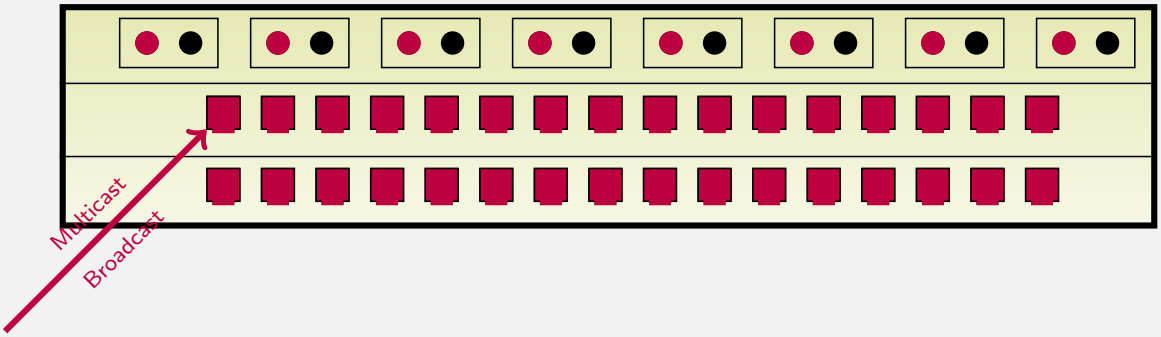
Virtual Local Area Network

- LAN allows a simple management:
  - No address allocation procedures
  - Simple interconnection with bridges
- But suffers from some scalability and security problems
  - Address list must be construct for each port
  - Traffic cannot be filtered
  - Broadcast or Multicast frames flood the network
  - Cabling constraints: Moving one equipment from one network to another implies modifying the topology.
- VLAN solves this problem separating the physical infrastructure into logical (virtual) ones.
  - Physical: Links, switches
  - Logical: Configure switches to assign a virtual network to a group of ports.

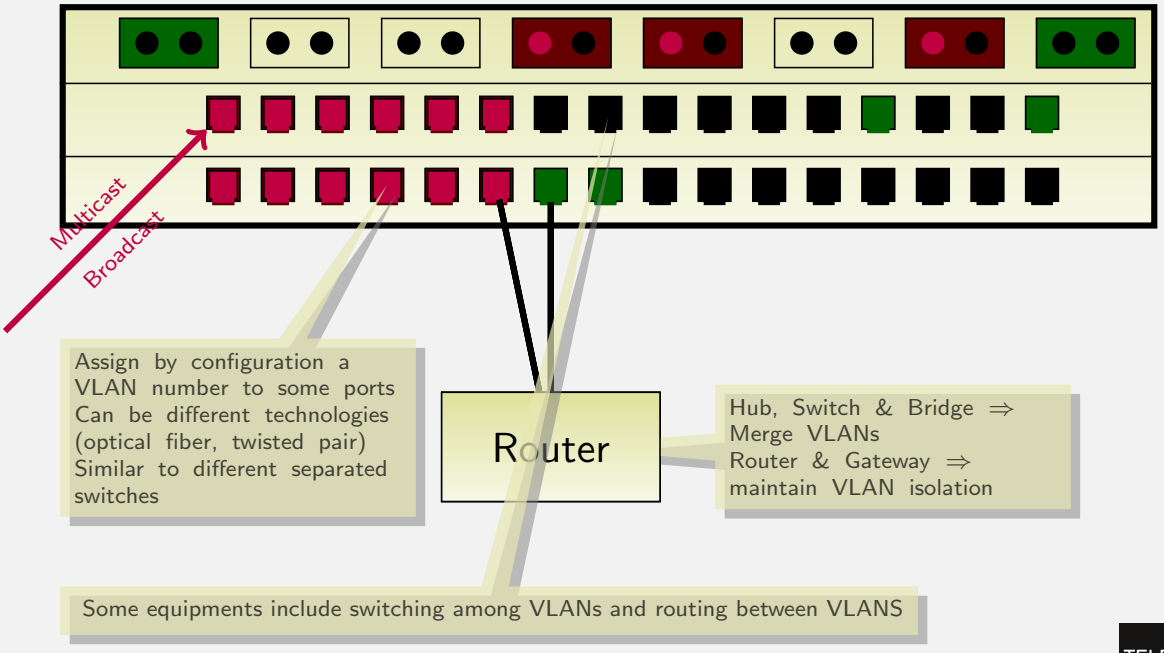
VLAN on a switch



VLAN on a switch



VLAN on a switch



Comments I

Les réseaux virtuels ou VLAN (*Virtual LAN*) constituent une alternative à la construction de réseaux de niveau 3 à l'aide de routeurs. Construire un réseau virtuel revient à avoir sur une même infrastructure physique (câblage, équipements d'interconnexion), plusieurs réseaux de niveau 2 complètement indépendants. Les réseaux virtuels présentent plusieurs avantages :

- facilité de mise en œuvre : contrairement aux réseaux de niveau 3 où une gestion relativement stricte du plan d'adressage est nécessaire, les réseaux virtuels gardent la souplesse des réseaux de niveau 2. Des logiciels d'administration facilitent leur configuration ;
- confidentialité : sur un réseau de niveau 2, il est relativement difficile de filtrer les trafics, un équipement peut dialoguer avec n'importe quel autre, les ponts ne filtrant que les trames d'équipements situés sur un même réseau.
- Au niveau 3, l'attribution d'adresses liées à la topologie du réseau permet la mise en place de règles de filtrage dans les routeurs permettant la création de firewall. Comme le trafic entre les réseaux virtuels est isolé, il est possible de limiter les accès à certains équipements.
  Il est à noter que les stations devront posséder une adresse IP par réseau virtuel ;
- souplesse d'utilisation : il est facile de donner ou de retirer les accès aux différents réseaux virtuels de l'entreprise. Cela évite de modifier le câblage dans les armoires de brassage.

Comments II

Un réseau virtuel peut être vu comme une réduction de la portée des trames en diffusion. Les transparents précédents montrent comment sont gérées les trames dans un commutateur. En règle générale, le trafic point-à-point est aiguillé uniquement vers le destinataire, cela permet des échanges simultanés entre plusieurs équipements. Par contre, quand une trame avec une adresse de diffusion (par exemple un broadcast ARP) est émise par un équipement, celle-ci est envoyée sur tous les ports du commutateur.

Dans l'exemple, les ports marqués en routage (port 3, 4 et 5 optiques et 1 à 6 des ports paires torsadées) peuvent être alloués au réseau virtuel 1et les ports marqués en vert au réseau virtuel 2. Le comportement est identique à l'utilisation de deux commutateurs différents. Une trame en diffusion (ou multicast) émise par un équipement du réseau virtuel 2 ne sera reçue que par les autres équipements de ce réseau virtuel. Pendant ce temps, les équipements des réseaux virtuels 1 peuvent continuer à dialoguer.

En conséquence, une trame émise par un équipement situé sur le port 1 ne pourra pas directement joindre un équipement situé sur le port 5. Un routeur devra être utilisé pour interconnecter les réseaux virtuels entre eux. La définition de réseaux virtuels nécessite plusieurs modifications au modèle de pontage établi par l'IEEE pour les réseaux de niveau 2, décrite dans les paragraphes suivants.

L'appartenance à un VLAN est généralement lié à la configuration des ports physiques des commutateurs. Il s'agit de la méthode la plus simple : elle associe aux ports du commutateur, un numéro de VLAN. Cette méthode est la plus sûre car l'appartenance à un réseau virtuel ne dépend pas de facteurs extérieurs au commutateur. Un utilisateur ne peut pas, en modifiant la configuration de sa machine, changer de réseau virtuel.
Les équipements terminaux ignorent la notion de VLAN. Ce sont les équipements d?interconnexion qui déterminent l'appartenance. Plusieurs méthodes de configuration des équipements d'interconnexion pour décrire cette appartenance à un réseau virtuel sont possibles.

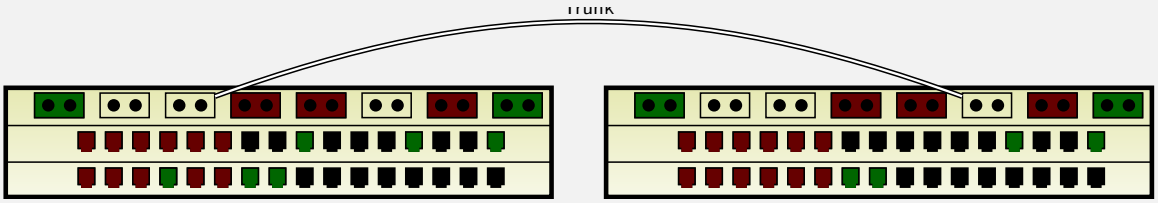
Comments III

- Manuelle: Cette méthode ne convient que lorsque le réseau virtuel n'est constitué que d'un seul commutateur. L'administrateur du réseau se connecte sur l'équipement et entre les fichiers de configuration. Le risque d'erreur est relativement important.
- Semi-automatique: Cette méthode se base sur les plates-formes d'administration SNMP. Généralement, ces outils ne fonctionnement que si tous les équipements d'interconnexion sont du même constructeur. Il peut y avoir la découverte automatique des commutateurs présents sur le réseau. L'administrateur peut ensuite attribuer un numéro de VLAN à chaque port. La configuration est ensuite stockée dans chaque commutateur.

VLAN

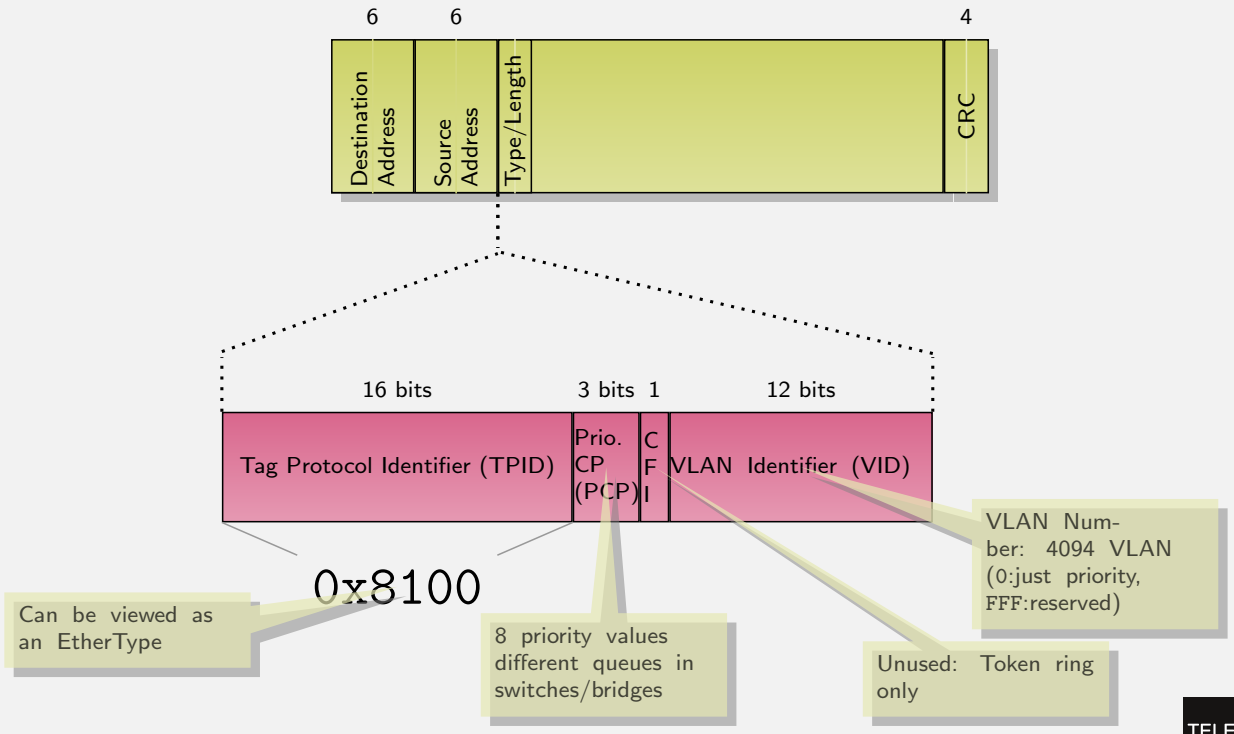
IEEE 802.1p/Q

Tags IEEE 802.1p/Q

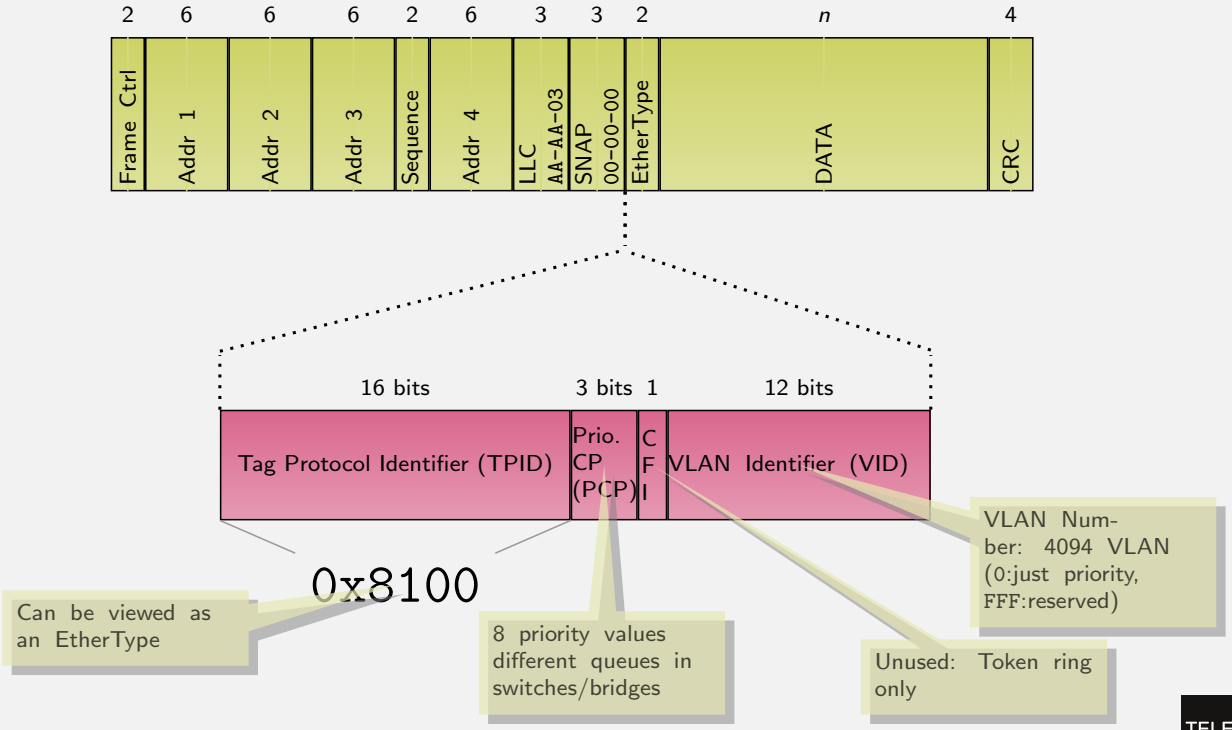


- When several switches are involved, VLAN value must be carried between equipments.
- Switches add a tag in the frame containing the VLAN number.
- IEEE 802.1p/Q defines tag format for VLAN and priority.

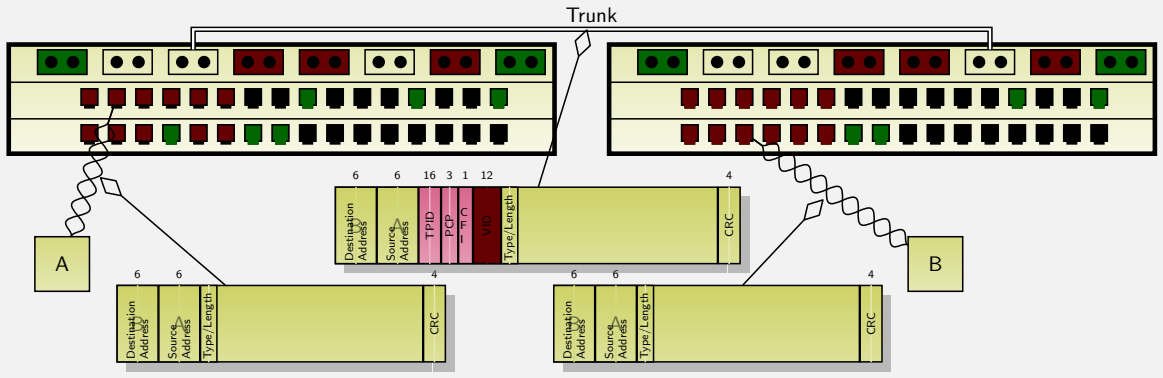
Ethernet vs IEEE 802.3



Ethernet vs IEEE 802.3



Tags IEEE 802.1p/Q



- VLAN tag is added between switches:
  - Host generally do not tag frames
  - Maximum frame size can be exceeded when tag is added
    - not a problem with jumboframe (9 KB)
    - IEEE 802.3ac also allow Ethernet frame with up to 1522 Bytes
- There is a default VLAN without tagging, switch can be configured to associate internally a VLAN value.

Comments I

Quand le réseau comprend plusieurs commutateurs, il est nécessaire de pouvoir transporter l'information d'appartenance d'un équipement d'interconnexion à un autre. Ceci se fait en ajoutant une étiquette aux trames transportées indiquant le numéro du VLAN. Les équipements d'interconnexion ne doivent connaître que les appartenances locales aux réseaux virtuels. Cet étiquetage est souvent fait par les équipements d'interconnexion car les équipements terminaux émettent généralement des trames non étiquetées.

La norme IEEE 802.1p permet de définir un format commun pour les étiquettes indépendamment des constructeurs. Pour ce faire, l'en-tête des trames MAC doit être modifié pour pouvoir y insérer des informations supplémentaires, tout en garantissant une compatibilité avec les anciens équipements. Pour ceux-ci, les informations supplémentaires seront vues comme un protocole de niveau supérieur. Ainsi pour Ethernet, le champ protocole 0x8100 définit l'encapsulation IEEE 802.1p. Cette encapsulation se résume à ajouter une étiquette de deux octets appelée TCI (Tag Control Information). Elle peut être aussi interprétée comme l'ajout de deux champs : TPID (Tag Protocol Identifier) et TCI entre l'adresse de la source et le champ type/protocole de la trame MAC. Pour le Wi-Fi, cette encapsulation se fait en utilisant SNAP pour placer l'identificateur de protocole 0x8100.

Le champ TCI se décompose en trois parties :

- un champ priorité sur 3 bits ;
- un drapeau appelé CFI (Canonical Format Indicator) sur un bit. qui permet en théorie d'insérer des informations de routage par la source (liste des ponts à traverser), mais en pratique ce bit n'est pas utilisé.
- un champ VID (VLAN Identifier) permet de marquer l'appartenance de la trame à un VLAN particulier.

L'insertion de cette information peut être réalisée par la station émettrice de la trame ou par les équipements d?interconnexion (Hub, commutateurs, ponts, etc). Dans ce dernier cas :

- le checksum de la trame doit être recalculé ;

Comments II

- dans le cas des réseau IEEE 802.3, les octets de bourrage devenus inutiles peuvent être retirés.

Surtout dans le cas d'Ethernet et du protocole IEEE 802.3, la longueur des trames peut excéder la taille maximale autorisée (c'est-à-dire 1 500 octets pour les données et 18 octets pour l'en-tête). La norme IEEE 802.3ac propose d'étendre la taille maximale des trames à 1522 octets si le champ type/longueur contient le TPID. Mais cette longueur peut être incompatible avec d'anciens pilotes de carte mais comme ce format ne concerne que des équipements récents, le risque d'incompatibilité est très faible.

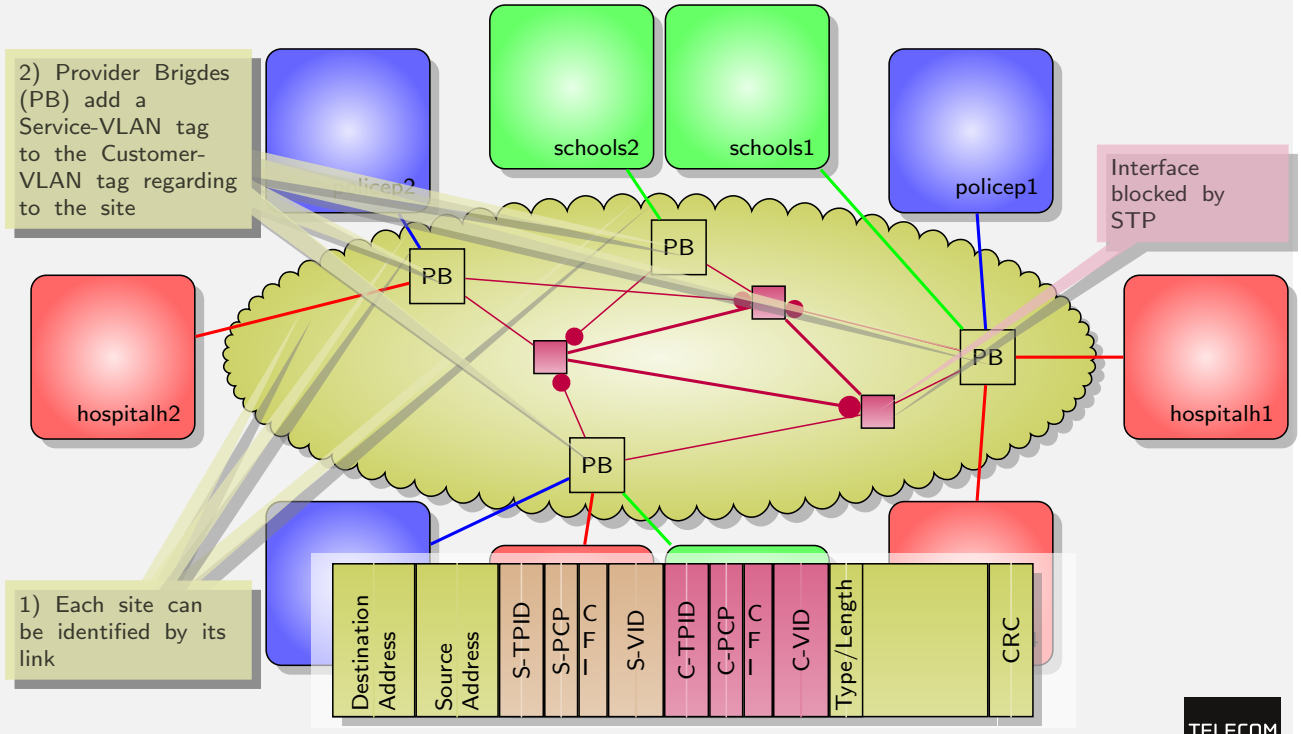
Metropolitan Networks

QinQ

Metropolitan Ethernet

- Switched networks with VLAN are particularly powerful:
  - frame switching is efficient
  - distance is no more a limitation (no CSMA/CD)
  - OPEX is reduced (equipments automatically configure themselves)
- But scalability and security are limited:
  - Switches must learn all the MAC address
  - An attacker may inject fake addresses
  - VLAN numbers must be globally unique
- QinQ allows to stack VLAN tag
- MACinMAC(IEEE 802.1ah) allows to encapsulate Ethernet frame into Ethernet frame.

IEEE 802.1ad (QinQ)



TELECOM Bretagne

IEEE 802.1ad (QinQ)

- IEEE 802.1ad defines two categories of VLAN tags:
  - Customer VLAN (or C-VLAN) generated by the customer switches
  - Service VLAN (or S-VLAN) added by the provider.
    - S-VID are chosen by the provider to link a group of sites.
    - S-VID designates a customer.
- S-VLAN guaranty unicity between two customers
  - if two independent company select the same VLAN number, traffic will not be merged
- S-VLAN guaranty security
  - A company cannot access to another company traffic.
- S-LAN are defined by the value 88a8 in the S-TPID
- More simple to manage than Layer 2 VPN based on IP routing and MPLS.

TELECOM Bretagne

Comments I

Avec le développement des réseaux métropolitains, il est nécessaire de pouvoir protéger des trafic venant d'une société. Les VLAN sont évidemment la solution la plus adaptée mais impliquent certaines contraintes en terme de numérotation. Il est possible pour l'opérateur de définir sur les liaisons trunk vers ses clients les numéros de VLAN autorisés ou non. Mais cela impose au client de numéroter ses VLAN dans une plage restreinte. Pour offrir plus de souplesse, le document IEEE 802.1ad propose d'empiler deux tags, seul le premier étant actif.

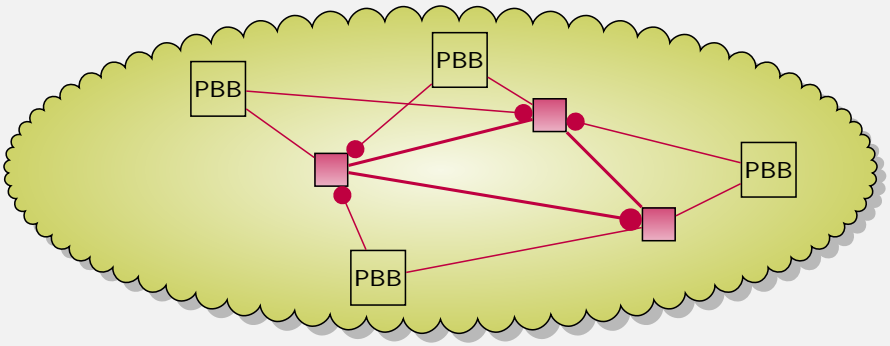
Le premier tag est appelé S-VLAN (pour Service VLAN) est purement interne au réseau de l'opérateur métropolitain. Il sert à désigner le client de l'opérateur. Le second tag est appelé C-VLAN (pour Customer VLAN) et est celui choisi par le client pour ses besoins. Il n'y a dons pas de risque de conflit entre deux clients puisque même si ils choisissent la même valeur de C-VLAN, la valeur du S-VLAN permettra de les différencier). Comme le S-VLAN est ajouté ou retiré par le commutateur en entrée ou en sortie de l'opérateur (PB :*Provider Bridge*), les clients n'ont pas la possibilité de le modifier. Cela renforce la sécurité du réseau car une société ne peut pas avoir accès au trafic d'une autre.

Cette solution de pontage sur de grandes distances est une alternative intéressante à des solutions de VPN (*Virtual Private Network*) basées sur du routage IP et sur l'utilisation de MPLS.

Metropolitan Networks

Mac In Mac

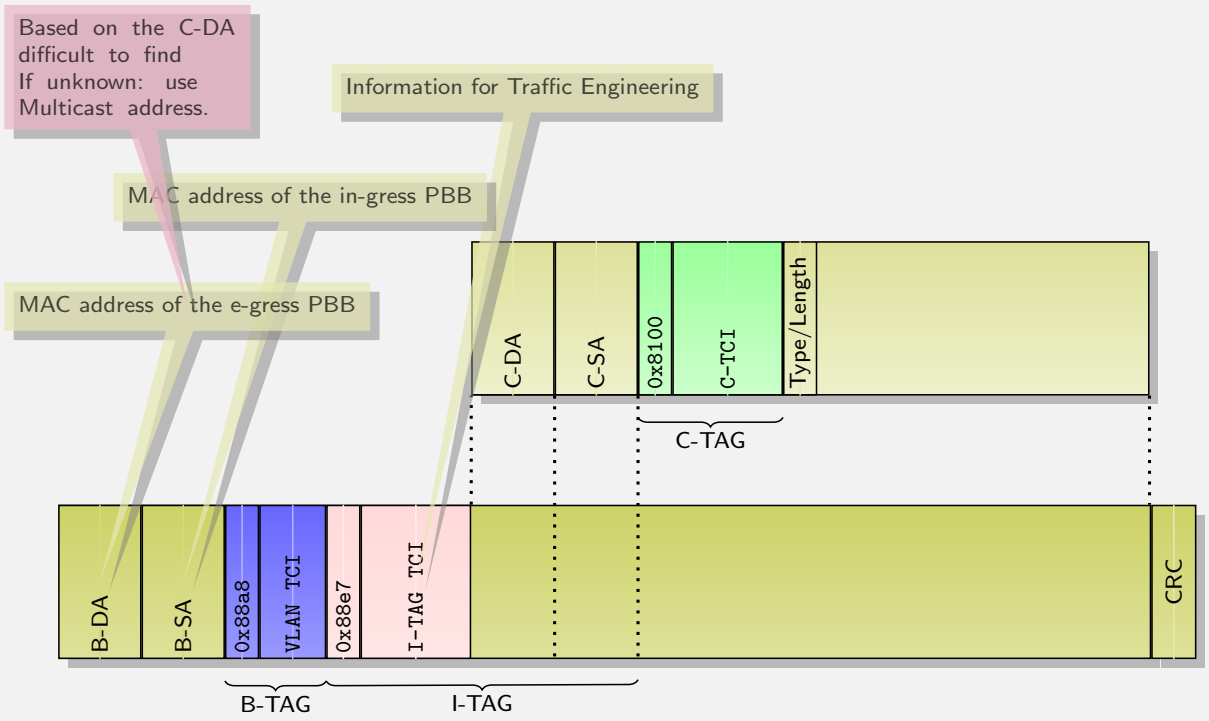
IEEE 802.1ah (MACinMAC)



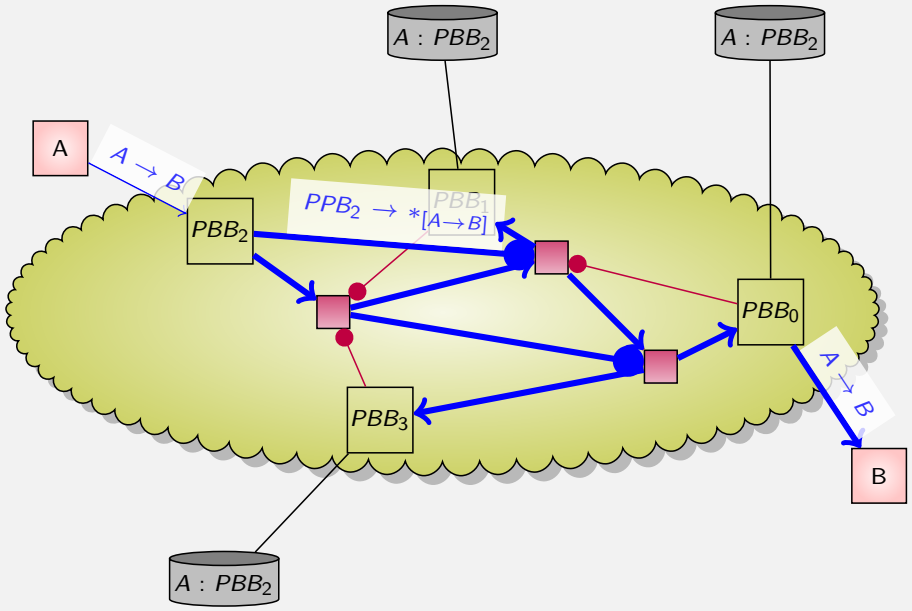
- Switches in the provider network have to memorize all the MAC addresses
  - require a lot of memory.
- an attacker forging the source mac address may:
  - slow down the network saturating core tables,
  - make some equipments unreachable using their MAC address.
- Provider Backbone Bridge adds an Ethernet header.

IEEE 802.1ah (MACinMAC)

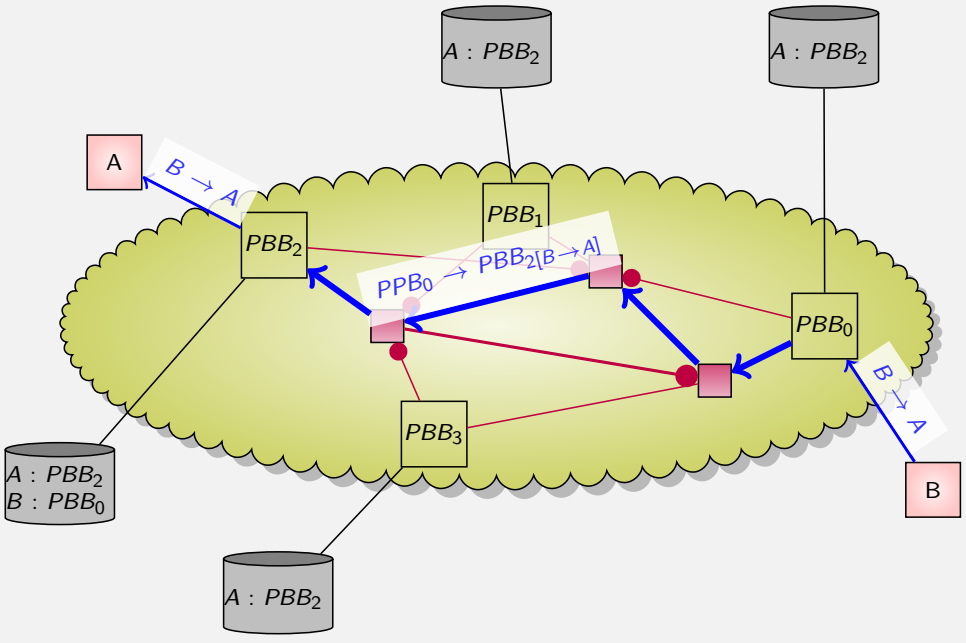
Metropolitan Networks ▶ Mac In Mac



IEEE 802.1ah (MACinMAC)



IEEE 802.1ah (MACinMAC)



Comments I

La norme IEEE 802.1ah permet d'encapsuler une trame Ethernet dans une autre. Cela permet de réduire le nombre d'entrées dans les tables de relayage des équipements du cœur de réseau. Comme pour la norme QinQ, deux niveaux de VLAN sont possible. le premier (appelé B-TAG) contient le VLAN ajouté par l'opérateur pour distinguer les utilisateurs. La norme prévoit d'insérer également d'autres informations dans un I-TAG qui serviront à la gestion de la qualité de service dans le réseaux de l'opérateur.
Comme il existe deux niveaux d'adresses, il faut pouvoir établir la correspondance entre l'adresse de destination contenue dans la trame du client et l'adresse MAC de destination du Provider Backbone Bridge (PBB) qui devra retirer l'encapsulation IEEE 802.1ah. Le moyen le plus simple consiste à construire une table dans les ponts en bordure du réseau de correspondance entre les adresses MAC du client et les adresses mac des PBB de sortie. Cet apprentissage se fait automatiquement en lisant les adresses sources au niveau du backbone et du client. Quand un pont n'a pas dans ses tables la correspondance, il utilise à la place une adresse de multicast. L'information est donc transmise à l'ensemble des ponts de bordure de l'opérateur. Cette trame en multicast a permis également à tous les ponts en bordure d'apprendre la localisation de la source. Quand le destinataire répond, le pont en bordure est capable de trouver l'adresse du pont de sortie du réseau de l'opérateur. Les mécanismes de commutation font que cette trame est directement et uniquement envoyé vers cette destination.

References

- W http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4463774
- W www.itg523.de/oeffentlich/08april/Knoll_Carrier_Ethernet.pdf