

18 Ethernet commuté et les VLAN

18.1 ETHERNET COMMUTÉ

18.1.1 Introduction

Issue de la téléphonie et des réseaux grande distance (WAN), puis mise en œuvre dans le monde Ethernet (*Switched Ethernet*) pour résoudre les problèmes d'effondrement des réseaux (figure 18.1) les techniques de commutation sont aujourd'hui largement utilisées pour réaliser tout type de réseaux.

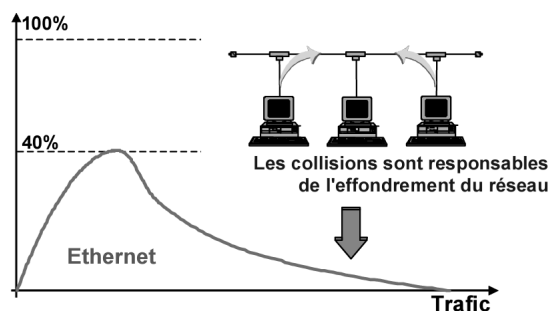


Figure 18.1 Effondrement des réseaux à contention.

La fonction d'acheminement des commutateurs permet de ne diffuser le trafic que sur le seul segment du réseau accueillant la (les) station(s) destinatrice(s) des messages (domaine de diffusion). Cette faculté, associée à celle de non-retransmission des trames erronées (erreurs de FCS, trame incomplète...), permet de découper un réseau physique en plusieurs sous-réseaux logiques, dits domaines de collision. Une collision sur un brin (segment) est invisible sur un autre. L'architecture adoptée est généralement du type *backbone* (réseau fédérateur), ce *backbone* pouvant être un lien à haut débit ou un simple commutateur. Dans ce type d'architecture représenté figure 18.2 seul, le trafic interdépartemental transite sur le réseau *backbone*.

Dans la représentation de la figure 18.2, les commutateurs segmentent les réseaux en trois sous-réseaux constituant chacun un domaine de collision. Ce type d'architecture correspond à l'utilisation traditionnelle des ponts¹ dont les commutateurs ne constituent qu'une évolution.

1. Voir chapitre 20.

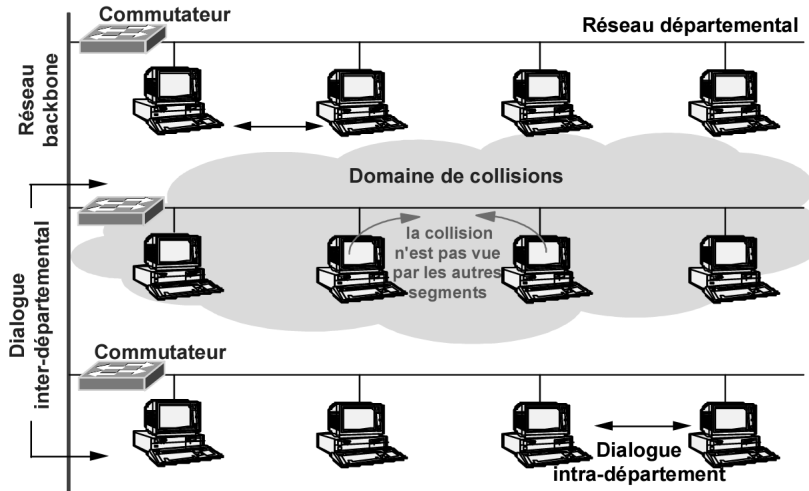


Figure 18.2 Architecture d'un réseau commuté de type backbone.

Aujourd'hui, les commutateurs ne sont pratiquement plus utilisés pour leur fonctionnalité de pontage mais surtout en lieu et place des hubs apportant ainsi la commutation jusqu'au poste de travail. Le tableau de la figure 18.3 compare les fonctionnalités d'un hub traditionnel et celles d'un commutateur LAN.



	 HUB	 Commutateur
Topologie	Étoile physique Bus logique	Étoile physique Étoile logique pour le trafic <i>unicast</i> Bus logique pour le trafic de <i>broadcast</i>
Acheminement	Diffusion sur tous les ports	Acheminement sur : – le port concerné (<i>unicast</i>), – les ports concernés (<i>multicast</i>), – tous les ports (<i>broadcast</i>).
Parallélisme	Ne traite qu'une seule trame	Traite simultanément plusieurs trames
Analyse de l'en-tête	N'interprète pas les en-têtes	Analyse les en-têtes (adresses) Les ponts analysent en plus le FCS
Bande passante	Partagée	Intégralité de la bande passante
Collision	Un seul domaine de collisions	Un domaine de collisions par port
Dialogue	<i>Half duplex</i>	<i>Half duplex</i> <i>Full duplex</i>

Figure 18.3 Comparaison hub/commutateur.

18.1.2 Principe de l'acheminement dans les commutateurs

Les différentes techniques de commutation

En ne mettant en relation que les ports intéressés par l'échange, la commutation améliore la gestion de la bande passante. L'acheminement s'effectue au niveau MAC, ce qui autorise des performances élevées. Deux techniques de base sont mises en œuvre (figure 18.4). Les premiers commutateurs (commutateurs Kalpana) relayaient la trame vers le port destination dès la lecture de l'adresse. Cette technique : lecture de l'adresse au vol et commutation rapide (*Fast forward* ou *Cut through*) est très performante mais elle propage les trames sans en avoir vérifié la validité et en particulier les trames corrompues par une collision.

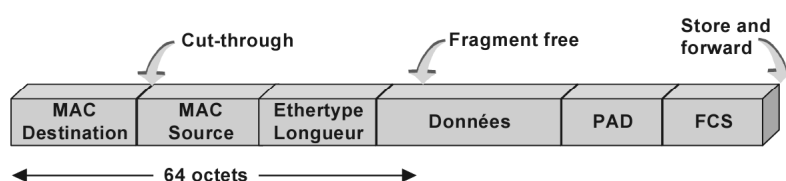


Figure 18.4 Techniques de commutation.

Ce mode de fonctionnement distingue les commutateurs² des ponts, les ponts traditionnels ne propageant les trames qu'après en avoir vérifié l'intégrité, d'où la distinction d'appellation. Cependant, les trames de collision sont identifiables, leur longueur est inférieure à la fenêtre de collision (64 octets). Aussi, pour éviter la propagation des collisions, certains commutateurs ne retransmettent les trames qu'après avoir reçu les premiers 64 octets (*fragment free*). Réémettant les trames avant d'en avoir reçu l'intégralité, ce mode de fonctionnement interdit l'adaptation des débits.

L'accroissement des performances des mécanismes de commutation a autorisé un retour au mode initial de fonctionnement des ponts. Les trames reçues sont intégralement lues avant retransmission (*Store and forward*), ce qui autorise le contrôle d'erreur, ne sont alors retransmises que les trames non erronées. Assurant une mémorisation provisoire des trames, ce mode permet l'adaptation des débits et du codage (port *auto sense*).

La première technique est plus performante en termes de performance (faible temps de latence). Cependant, elle propage les trames erronées. L'*adaptive error free* combine les deux techniques, initialement le commutateur démarre en mode *Cut through*, et si le taux d'erreur atteint un seuil prédéterminé, le commutateur bascule en mode *store and forward*. Rappelons qu'en mode *cut through*, il n'y a pas de mémorisation des données, il ne peut donc y avoir adaptation de débit entre le débit entrée et celui du port de sortie. Ces modes de fonctionnement ne sont guère possibles dans les réseaux à débit multiple 10/100/1000 où seul le mode « *store and forward* » est possible.

Les différents modes de commutation

La configuration du système peut être statique (les tables de commutation sont introduites par l'administrateur) ou dynamique (les tables de commutation sont construites par analyse de trafic et apprentissage des adresses MAC, voir paragraphe suivant). À l'instar des *hubs*, les

2. Notons que l'IEEE dans toutes ses recommandations concernant la commutation continue d'utiliser la dénomination de pont (*bridge*), les deux appellations seront employées indifféremment dans ce chapitre.

commutateurs peuvent mettre directement en relation des stations (commutation par port) ou des segments de réseaux (commutation de segment). La figure 18.5 illustre ces différents types.

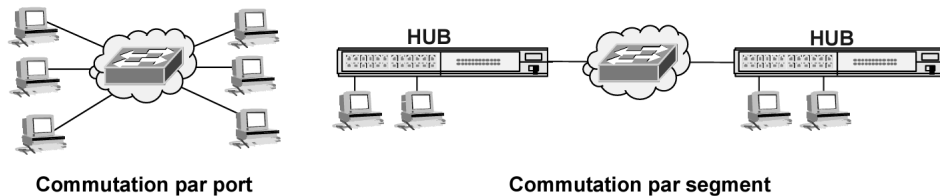


Figure 18.5 Différents modes de commutation.

Établissement dynamique de la table d'acheminement

Traditionnellement, la commutation consiste, en fonction d'un identifiant (label), à mettre en relation directe un port d'entrée avec un port de sortie. La mise en relation est établie préalablement à tout envoi de donnée par un protocole de signalisation qui établit une table de mise en relation dite table de d'acheminement ou table de commutation. Les commutateurs LAN (*Lan switch*) n'ouvrent pas explicitement de circuit virtuel. Il n'y a aucun protocole de signalisation de mise en œuvre. À l'instar des ponts, dont ils ne constituent qu'une évolution, la table d'acheminement (FDB, *Forwarding Data Base*) est construite dynamiquement par écoute du trafic (figure 18.6).

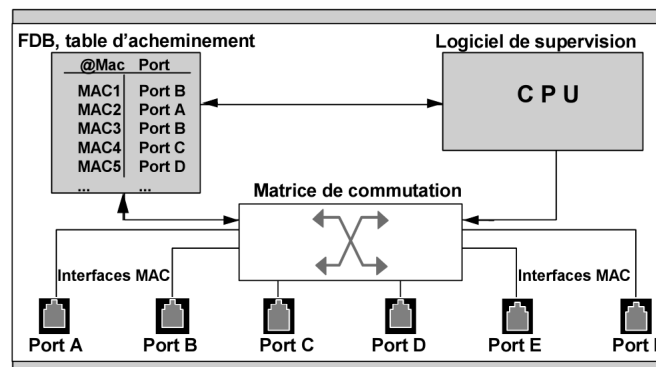


Figure 18.6 Principe d'un commutateur.

À cet effet, le commutateur examine le trafic reçu sur chacun de ses ports et associe à ces ports l'adresse MAC source de la trame reçue. Le commutateur apprend ainsi la localisation géographique des stations. Ainsi, à réception d'une trame (figure 18.7), le commutateur :

- ❑ lit l'adresse source, si celle-ci lui est inconnue il l'associe au port d'arrivée de la trame (localisation de la source) ;
- ❑ examine l'adresse destination, si celle-ci lui est inconnue il diffuse la trame sur tous ses ports sauf sur le port d'arrivée (inondation) et il inscrit dans sa table le couple « @MAC/Port » source (apprentissage) ;
- ❑ si l'adresse destination est inscrite dans sa table, diffuse la trame sur le port associé à cette adresse (*forwarding*), sauf si l'adresse destination est associée au port d'origine (filtrage, la

trame est destinée à une machine présente sur le même segment que la machine source). Dans ce dernier cas, il élimine la trame.

Enfin, les trames de *broadcast* sont diffusées sur tous les ports sauf celui d'arrivée.

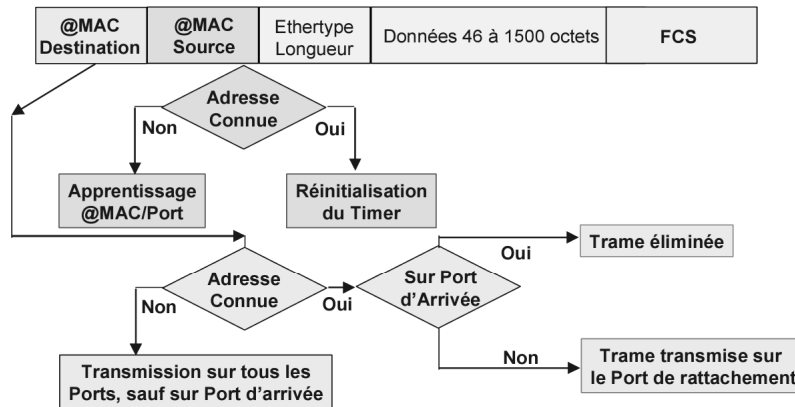


Figure 18.7 Construction de la table d'acheminement.

En acheminant les trames directement sur le ou les seuls ports où sont localisés les destinataires, la commutation autorise l'acheminement simultané de plusieurs trames (figure 18.8).

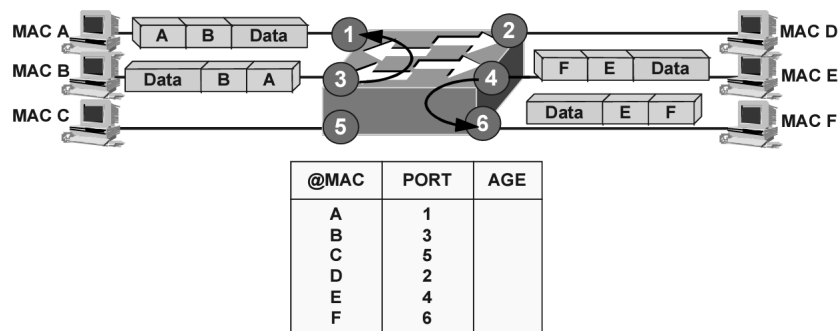


Figure 18.8 La commutation autorise le parallélisme des communications.

Les tables ne pouvant posséder autant d'entrées que de stations raccordées, périodiquement, les adresses les plus anciennes sont effacées (figure 18.8). À cet effet, à chaque entrée de la table, est associé un temporisateur réinitialisé à chaque réception d'une trame de même origine (adresse MAC source). À l'échéance du temporisateur, l'entrée est effacée (vieillesse de l'adresse ou *aging*).

Ce procédé permet la mobilité (changement de port de raccordement d'une machine) et évite l'engorgement des tables. Par défaut, la valeur du temporisateur est de 300 s (5 min). Cette durée est paramétrable : trop courte, le commutateur se comporte en *hub* (absence de l'adresse en table, diffusion) et, trop longue, les tables peuvent être pléthoriques et leur lecture occasionner une baisse de performance.

Ethernet full duplex

Dans la commutation par port, une station est reliée directement à un port (architecture de type *hub*), c'est une liaison en mode point à point et par conséquent, les risques de collision sont inexistant. L'adaptateur peut alors émettre et recevoir en même temps des messages différents, l'échange est *full duplex*. La technologie *full duplex* (**FDSE**, *Full Duplex Switched Ethernet*) permet de doubler la bande passante d'un réseau local. Initialement réservée aux liens inter-commutateurs, la technologie *full duplex* est aujourd'hui supportée par la plupart des adaptateurs. Il suffit pour cela d'invalider la détection de collision.

18.1.3 Notion d'architecture des commutateurs

Historiquement, dans les réseaux de téléphonie les premiers commutateurs spatiaux étaient constitués de relais électromécaniques qui mettaient en relation un circuit d'entrée avec un circuit de sortie (commutateur *crossbar*). Par la suite, les relais furent remplacés par des semi-conducteurs (transistors).

Le principe des commutateurs *crossbar* est représenté figure 18.9, à chaque croisement, un transistor établit ou rompt la liaison entre les deux conducteurs. Les points « noirs » de la figure schématisent une connexion (état passant du transistor), les points « clairs » représentent l'état bloqué du transistor et donc la coupure entre les deux circuits.

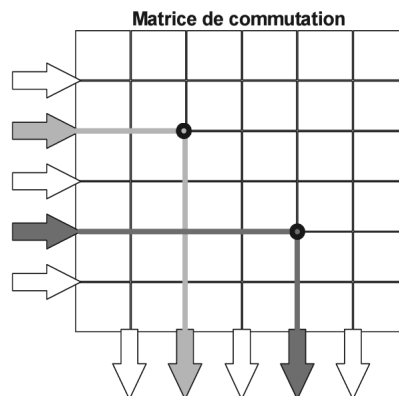


Figure 18.9 Principe d'une matrice de type crossbar.

Autorisant le parallélisme dans le traitement des trames, les commutateurs de type *crossbar* sont très efficaces mais aussi très complexes ; le nombre de points de connexion évolue en N^2 où N représente le nombre de ports. Une solution à cette complexité est fournie par les commutateurs multi-étage dont la réalisation la plus courante est le commutateur Banyan. Les commutateurs du type Banyan ne comportent que $N \cdot \text{Log}N$ composants.

La figure 18.10 illustre le principe de mise en relation d'un port d'entrée avec n'importe quel port de sortie. Chaque port d'entrée de ce système peut être mis en relation avec tout port de sortie.

Ces architectures répondent mal au problème de la diffusion. Aussi, d'autres architectures, certes peut-être moins efficaces que celles du type Banyan mais beaucoup simples, ont la faveur des constructeurs.

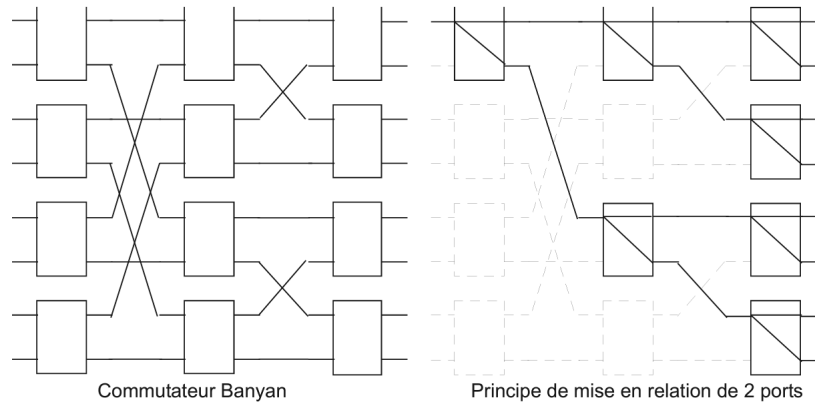


Figure 18.10 Principe d'un commutateur de type Banyan.

Ce sont les architectures à bus ou à mémoire partagée. La première correspond à la réalisation d'un réseau très haut débit (*Collapsed backbone*), le débit du bus devant être au moins égal à la demi-somme des débits incidents. La seconde solution, la plus courante, est celle de la mémoire partagée à accès multiples simultanés. Toutes les trames sont copiées en mémoire centrale avant d'être commutées. Ce dernier type d'architecture permet de résoudre simplement les mécanismes de blocage. La figure 18.11 associe un mode de commutation à une architecture interne.

Type de commutation		
Cut-through	Store and forward	
Crossbar	Hard	Bus + Asic ^a
	Soft	Mémoire partagée

a. ASIC (*Application Specific Integrated Circuit*), circuit intégré dédié à une application spécifique.

Figure 18.11 Association architecture et technique de commutation.

18.1.4 Contrôle de flux dans les commutateurs

Un commutateur peut mettre en relation n'importe quel port d'entrée avec n'importe quel port de sortie. Si simultanément, plusieurs flux d'entrée convergent vers un même port de sortie (par exemple pour l'accès à un serveur), il peut y avoir dépassement des capacités d'émission du port de sortie et pertes de données. Afin d'éviter ces pertes, le système est doté de *buffers* d'attente.

Les commutateurs peuvent disposer exclusivement de *buffers* d'entrée ou de sortie.

En cas de conflit d'accès à un port de sortie, si le système ne possède que des *buffers* en entrée, il peut se produire un blocage en sortie (perte de données). Si le système n'est doté que de *buffers* de sortie, on ne peut exclure un blocage en entrée que si les *buffers* de sortie sont suffisamment dimensionnés pour admettre une convergence de tous les flux d'entrée vers un unique port de sortie. Cette solution conduit à la réalisation de *buffers* de grande taille.

Dans ces conditions, les constructeurs adoptent généralement une solution mixte : *buffers* en entrée et *buffers* en sortie (figure 18.12).

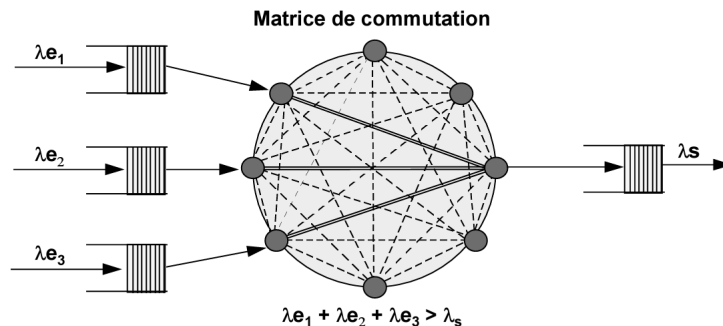


Figure 18.12 Congestion dans les commutateurs.

Enfin, quelle que soit la taille des *buffers*, une perte de trames par débordement n'est jamais à exclure. Aussi, les commutateurs mettent-ils en œuvre un contrôle de flux de type *Xon/Xoff* par l'intermédiaire de la trame « Pause ».

La figure 18.13 illustre, dans les réseaux locaux (IEEE 802.3), le fonctionnement du contrôle de flux.

Lors de l'apparition d'un état de saturation (seuil des files d'attente), le commutateur émet sur son ou ses port(s) saturé(s) une trame « Pause ». Celle-ci fixe la durée de blocage du port amont exprimée en *Time Slot*. Une valeur de durée à zéro correspond à une autorisation de reprise des émissions. Ce mécanisme est évidemment incompatible avec le support de la téléphonie sur IP dans les réseaux locaux.

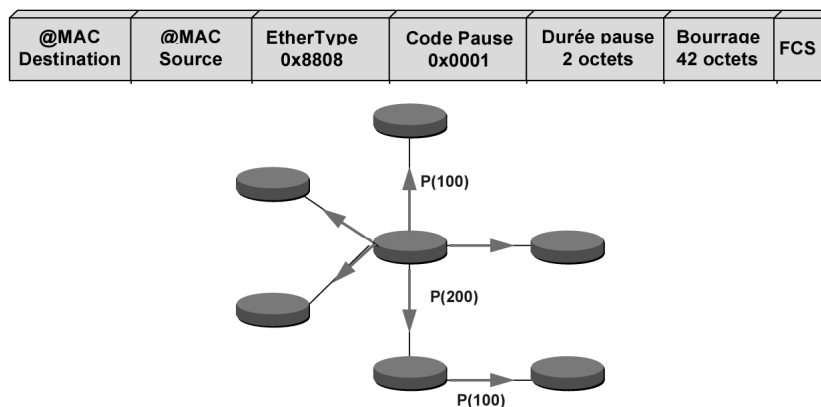


Figure 18.13 Contrôle de flux dans les LAN (IEEE 802.3x).

18.1.5 Commutation et trafic multicast

Le protocole GARP

La norme IEEE 802.1p définit deux modes de fonctionnement des ponts et commutateurs. Dans le mode de type 1, les ponts ne mettent en œuvre aucune règle de filtrage spécifique, ils ne réalisent que l'acheminement du trafic vers le port auquel l'*host destination* est raccordé (fonctionnement

traditionnel des ponts). Dans le mode de type 2, les ponts/commutateurs mettent en œuvre des règles de filtrage étendues :

- ❑ le mode A correspond au mode de fonctionnement traditionnel des ponts/commutateurs (type 1). Les trames *unicast* dont l'adresse destination est connue sont acheminées, les trames *unicast* d'adresse inconnue et les trames de *multicast* et *broadcast* sont diffusées ;
- ❑ le mode B est similaire au type 1. Cependant, les trames *unicast* ne sont acheminées que si leur adresse destination ne figure pas dans une liste de filtrage d'interdiction (liste rouge) ;
- ❑ enfin dans le mode C, la trame *unicast* n'est transmise que si l'adresse destination figure explicitement dans la base de filtrage (acheminement des seules trames autorisées, liste blanche).

Le protocole GARP (*Generic Attribute Registration Protocol*) assure la diffusion des informations de filtrage (inscription ou suppression d'attributs) à travers le réseau ponté ou commuté. En particulier, il sert à la diffusion des informations d'acheminement des trames de *multicast* (**GMRP**, *Generic Multicast Registration Protocol*) et de constitution des VLAN (**GVRP**, *Generic VLAN Registration Protocol*). Le format générique des messages GARP, illustré en figure 18.14, utilise l'encapsulation LLC de l'IEEE.

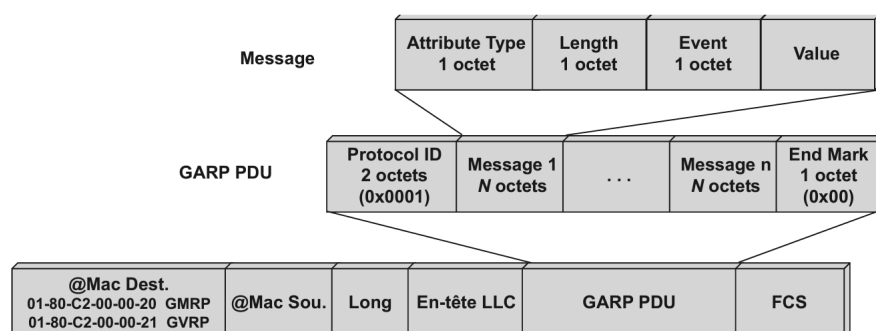


Figure 18.14 Format des messages GARP.

Les protocoles utilisateurs de GARP sont distingués par leur adresse *multicast* MAC destination. La GARP PDU contient une liste de messages. Chaque message est identifié par le champ *Attribute Type* dont les valeurs dépendent du protocole (GMRP ou GVRP). Le champ longueur indique, en octets, la longueur totale de chaque message (type, longueur, événement, valeur). L'action associée à l'attribut (enregistrement ou désenregistrement d'attribut) est définie par le champ Événement (*Event*). Le champ Valeur contient l'adresse de *multicast* (GMRP) ou l'identification du VLAN concerné (GVRP).

Le protocole GMRP

Le protocole GMRP (**GMRP**, *Generic Multicast Registration Protocol*) permet de limiter la diffusion du trafic *multicast* aux seuls ports d'un pont ou commutateur sur lequel l'adresse du groupe *multicast* a été enregistrée.

Le principe est simple, une station mettant en œuvre le protocole GMRP demande explicitement au pont auquel elle est raccordée d'être destinataire des messages *multicast* des groupes dont elle communique l'adresse (enregistrement). Cette information se propagera dans le réseau global limitant ainsi le trafic *multicast* aux seuls éléments ayant fait l'objet d'une annonce d'inscription. En

conjugaison avec le *Spanning Tree*, GMRP construit un arbre de diffusion *multicast*. Ce protocole est peu utilisé.

18.1.6 Le STP (Spanning Tree Protocol) ou arbre recouvrant

Généralités

La disponibilité du système d'information est un facteur de réussite d'une entreprise. Aussi, pour pallier toute défaillance d'un équipement, il est nécessaire de réaliser des connexions maillées entre les équipements d'accès et les équipements de concentration, ce qui implique une gestion de la redondance et éventuellement une « clusterisation » des serveurs de communication. En local, cette redondance se réalise en deux niveaux : la redondance des liens et le maillage du réseau (redondance des équipements). Sur les commutateurs, la redondance des liens est généralement réalisée par des liens *trunk* (accroissement du débit d'interconnexion entre commutateur et sécurisation des liens), elle est prise en compte par la norme IEEE 802.3ad.

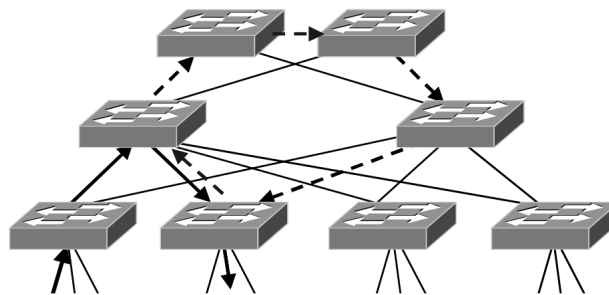


Figure 18.15 STP et la gestion de la redondance.

La mise en parallèle volontaire d'équipements réseaux (commutateurs ou ponts) par mesure de sécurité (figure 18.15), ou par erreur, dans un réseau complexe, engendre un phénomène de bouclage qui conduit à l'effondrement du réseau notamment lorsqu'il s'agit de messages de *broadcast* (tempête de *broadcasts*). Ce phénomène est illustré par la figure 18.16.

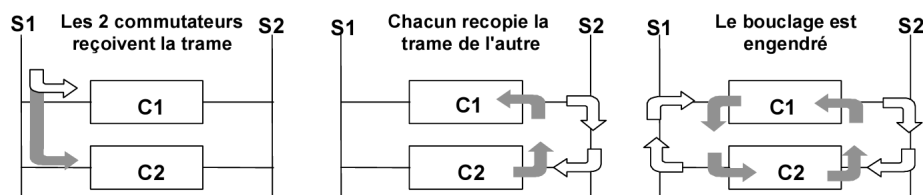


Figure 18.16 Bouclage des trames sur des ponts en parallèle.

La trame émise sur le segment S1 à destination d'une station non encore enregistrée dans les commutateurs est reçue par les deux commutateurs (C1 et C2), elle est retransmise sur le segment S2, la trame émise par le commutateur 1 sur le segment 2 est reçue par le commutateur 2, tandis que celle émise par le commutateur 2 est recopiée par le commutateur 1. Chacun recopie alors la trame sur le segment 1... Une situation de boucle est engendrée.

Développé à l'origine par DEC et normalisé par l'IEEE³ (IEEE 802.1D), l'algorithme du *spanning tree* (STP, *Spanning Tree Protocol*) ou « arbre recouvrant » est un protocole d'apprentissage de la topologie du réseau dont le but est :

- ❑ d'éliminer les boucles en désactivant les ports des éléments qui engendrent ces boucles ;
- ❑ de contrôler en permanence la disponibilité des commutateurs actifs ;
- ❑ et, en cas de défaillance d'un commutateur actif, de basculer le trafic sur son *backup* mis en sommeil.

Le principe en est relativement simple, il s'agit de construire un graphe en arbre. À partir d'un commutateur élu, désigné commutateur racine (*Switch root*), l'algorithme du *spanning tree* détermine le chemin le plus court en éliminant les risques de bouclage. Les commutateurs en boucle sont déclarés commutateurs *backup* et, tant que les commutateurs actifs dont ils sont le *backup* sont en fonction, ils sont en sommeil (figure 18.17).

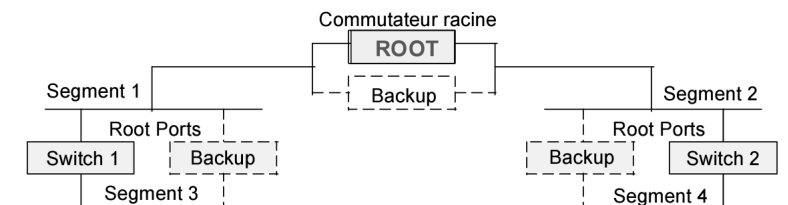


Figure 18.17 Exemple de configuration du Spanning Tree Protocol.

L'algorithme du *spanning tree*

Pour construire l'arbre recouvrant (*Spanning tree*), les ponts (commutateurs) s'échangent des messages de diffusion (BPDU, *Bridge Protocol Data Unit*). Le *spanning tree*, défini à l'origine pour les ponts, utilise deux types de message, les messages de configuration (figure 18.18) et les messages d'indication de changement de topologie.

Le message de configuration (BPDU de configuration) utilise l'encapsulation IEEE 802.2 LLC1 (*Logical Link Control mode datagramme*) avec un SAP de 0x42 et une adresse MAC destination de diffusion 01-80-C2-00-00-00. Les champs identification du protocole, *Type* sur 2 octets, et *Version* (1 octet), non utilisés, doivent toujours être mis à 0. Le champ suivant sur 1 octet distingue un message de configuration (0), d'un message d'information de topologie (128).

Chaque pont possède un identificateur (ID, *bridge IDentifier*) construit à partir d'un indicateur de priorité (2 octets de poids fort du champ) et de l'adresse MAC du pont. L'adresse MAC du pont est celle de son port de plus faible adresse. Le champ priorité est fixé à 0x8000 par défaut. Sa valeur peut être modifiée par l'administrateur (0 à 0xFFFF). Plus ce nombre est faible, plus la priorité est élevée. Le pont racine élu est le pont de plus petit ID, à priorité égale, c'est celui de plus petite adresse MAC. Au démarrage, chaque pont émet une BPDU dans laquelle il se déclare pont racine. Tout pont qui reçoit une BPDU avec un ID inférieur au sien, cesse ses émissions et rediffuse les trames reçues. À la fin du processus, seul le pont racine continue à émettre son identifiant. Il est élu pont racine.

Le port par lequel un pont reçoit en premier la trame d'identification du pont racine est appelé « *root port* ». Pour déterminer l'arbre minimal, chaque pont se voit affecter, par l'administrateur,

3. La recommandation 802.1D concerne aussi la gestion des liens *trunk*.

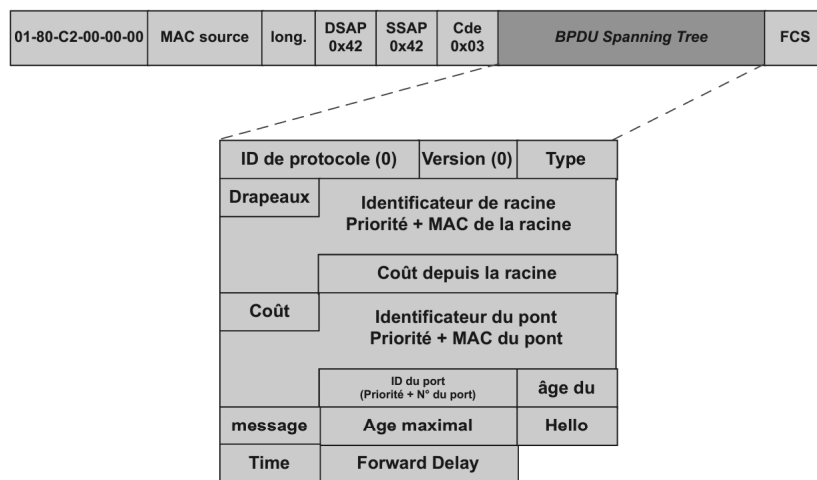


Figure 18.18 BPDU de configuration.

un coût. Le pont racine émet une trame de diffusion avec un coût nul, chaque pont répète cette trame en incrémentant le coût du sien (*Root path cost*). Un pont qui reçoit, sur un port « non-racine », une trame dont le coût, depuis la racine, est inférieur au coût des trames qu'il émet, en déduit qu'il existe, depuis la racine, une route de moindre coût. Il se met alors en sommeil (pont *backup*). En cas d'égalité entre deux ponts, c'est le pont de plus petit ID qui est élu. L'éventuel pont en boucle sur la racine détermine qu'il est pont *backup* simplement parce qu'il reçoit des trames identiques sur ses deux ports et que son ID est supérieur à celui de la trame reçue. En principe, le coût de traversée des ponts est le même pour tous, la valeur recommandée était « 1 000/Débit en Mbit/s ». L'accroissement des débits a rendu cette valeur obsolète, la recommandation IEEE 802.1p fixe de nouvelles valeurs (figure 18.19). Il est, aussi, possible, en jouant sur les valeurs du niveau de priorité et de coût, de privilégier un chemin par rapport à un autre.

Débit du réseau	Valeurs recommandées	Plages admises
Réseaux IEEE 802.3		
10 Mbit/s	100	50-600
100 Mbit/s	19	10-60
1 Gigabit/s	4	3-10
10 Gigabit/s	2	1-5
Réseaux IEEE 802.5		
4 Mbit/s	250	100-1 000
16 Mbit/s	62	40-400
100 Mbit/s	19	10-60

Figure 18.19 Valeurs des coûts recommandées par l'IEEE.

Périodiquement une trame de diffusion est émise (paramètre *Hello Time* valeur par défaut de 2 s, valeur recommandée 20 s). Si un pont *backup* reste plus de cet intervalle de temps sans rien recevoir, il en déduit que le pont, dont il est le *backup*, est défaillant. Il émet alors une trame de configuration. Ce fonctionnement par diffusion de BPDU peut, sur un lien WAN, consommer une

partie importante de la bande passante. Aussi, lorsque les réseaux sont interconnectés *via* des liens WAN à faible bande passante, il est recommandé de ne pas mettre en service le *spanning tree*.

Les ponts transparents (commutateurs) peuvent être dans cinq états :

- ❑ l'état **Disabled**, dans cet état, le pont ne participe à aucune activité, il est inerte ;
- ❑ l'état **Listening** durant la phase de configuration et de construction de l'arbre recouvrant. Dans cet état, le pont n'accepte ni ne retransmet les trames utilisateurs ;
- ❑ l'état **Learning**, correspond à la phase d'apprentissage, le pont met à jour sa base de données d'acheminement. Dans cet état, le pont ne participe toujours pas à la retransmission des trames ;
- ❑ l'état **Forwarding**, c'est l'état de fonctionnement normal d'un pont, il participe alors à l'acheminement des trames sur le réseau ;
- ❑ enfin, l'état **Blocking**, dans cet état, le pont est en sommeil, il n'achemine aucune trame mais participe aux opérations du *spanning tree* et d'administration des ponts.

Conclusion

L'algorithme du *Spanning Tree Protocol* utilise une adresse de diffusion. Certains constructeurs utilisent une adresse privée, aussi, en cas de système multiconstructeur, il est important de vérifier que les adresses de diffusion soient identiques (figure 18.20).

	Ethernet
IEEE	01-80-C2-00-00-00
IBM	03-00-00-00-80-00
DEC	09-00-2B-01-00-00
Retix	09-00-77-00-00-01
Spider Systems	09-00-39-00-70-00
Ungerman-Bass	01-DD-01-00-00-00

Figure 18.20 Exemples d'adresses de diffusion du *spanning tree*.

Les informations de coût et d'ID peuvent être initialisées par le constructeur. Dans ce cas, l'administrateur ne maîtrise ni le pont racine, ni la topologie. Un pont racine mal déterminé peut constituer un véritable goulet d'étranglement. En effet, chaque pont retransmet le trafic vers le port racine ainsi de suite jusqu'à ce qu'il atteigne sa destination. Si le destinataire n'est pas localisé dans la branche montante, le pont racine reçoit tout le trafic.

Le *Spanning tree* définit des routes statiques qui ne prennent pas en compte le trafic réel sur les branches du réseau. Considérons le réseau représenté figure 18.21, compte tenu des coûts indiqués (C), le pont P3 est en sommeil. Si le trafic entre le segment R1 et R2 est important, le délai de retransmission des trames du segment R1 vers le segment R3 peut être prohibitif. Il eut été plus intéressant de configurer le réseau pour que le pont *backup* soit le pont P2, sauf si le trafic P1, P2 est important ! Le routage par la source (**SR**, *Source Routing*), autre mode d'acheminement dans les réseaux pontés (commutés), remédie à cet inconvénient, il détermine la route optimale dans le réseau en fonction de critères prédéfinis (charge, délai...).

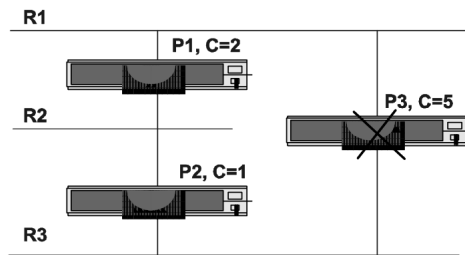


Figure 18.21 Topologie du réseau et Spanning tree.

Le *Spanning Tree* se décline aujourd'hui en trois adaptations :

- ❑ Le *Spanning Tree* standard (STP), IEEE 802.1D.
- ❑ Le *Rapid Spanning Tree* (RSTP), 802.1w porte le temps de convergence à quelques secondes contre quelques dizaines pour le *Spanning Tree* standard. S'il existe dans le même domaine de *broadcasts* des éléments ne supportant pas le RSTP, les implémentations du RSTP se replient en mode dit compatible.
- ❑ Le Multiple STP ou MSTP (802.1s repris dans 802.1q en 2003) est destiné aux environnements comportant plusieurs VLAN (une instance du *Spanning Tree* par VLAN). Ce mode de fonctionnement répartit la charge du *Spanning Tree* sur l'ensemble des liens de redondance, ce qui améliore la bande passante disponible et le temps de convergence.

Le *Spanning Tree* et ses évolutions (IEEE 802.1D STP, IEEE 802.1w RSTP, IEEE 802.1s MSTP) éliminent les boucles dans un réseau en bloquant des ports, voire des équipements, les liens de redondance sont inutilisés alors que le besoin de bande passante augmente. Le protocole *Shortest Path Bridging* (SPB, IEEE 802.1aq) non seulement simplifie la gestion des grands réseaux mais encore partage la charge du réseau en utilisant tous les liens de redondance (routage à trajets multiples).

18.2 LES RÉSEAUX VIRTUELS OU VLAN (*VIRTUAL LOCAL AREA NETWORK*)

18.2.1 Principes généraux des VLAN

Avec l'accroissement des réseaux, les messages de diffusion (ARP, DHCP, annonces de service...) occupent une part de plus en plus importante de la bande passante. En définissant, indépendamment de la situation géographique des systèmes, des domaines de diffusion (domaine de *broadcast*), les VLAN autorisent une répartition et un partage optimal des ressources de l'entreprise. Application directe de la commutation statique, les VLAN associent un port à un identifiant. Ne peuvent communiquer que les machines raccordées à des ports de même identifiant. Ainsi, sur le commutateur de la figure 18.22 deux VLAN sont déclarés. La communication entre stations n'est possible qu'entre les stations A, C et D d'une part et les stations B, E et F d'autre part. Il en est de même pour les *broadcasts* qui ne sont diffusés qu'au sein de leur VLAN respectif (domaine de diffusion).

La communication n'est autorisée qu'entre machines d'un même VLAN. Les communications inter-VLAN doivent transiter par un routeur.

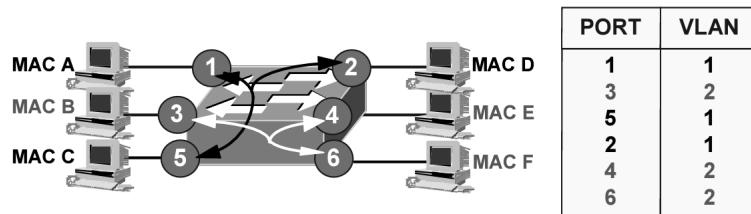


Figure 18.22 Principe des VLAN.

La figure 18.23 compare une architecture traditionnelle de cloisonnement physique de segments de réseaux au cloisonnement logique réalisé par les VLAN.

Dans la segmentation traditionnelle, l'appartenance à un réseau dépend du commutateur de rattachement (brassage), alors que dans la segmentation logique celle-ci est indépendante du brassage, elle n'est déterminée que par le critère d'appartenance à tel ou tel VLAN.

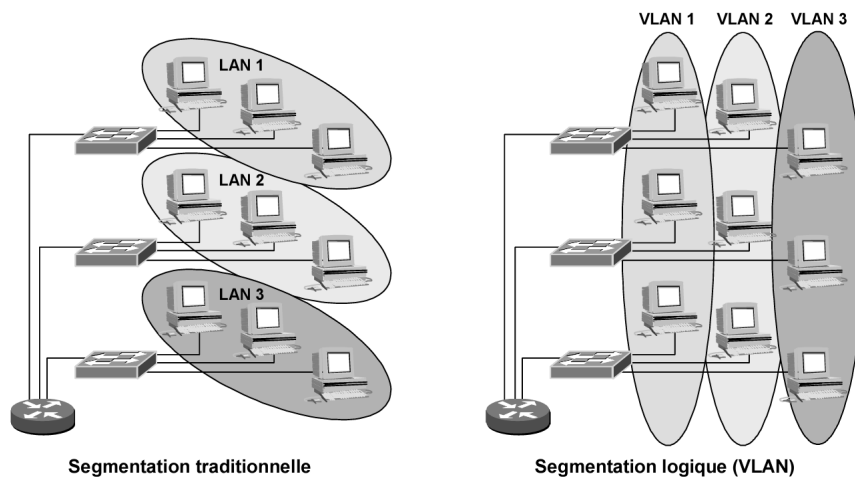


Figure 18.23 Comparaison de la segmentation physique et logique.

Ainsi, les réseaux virtuels permettent de réaliser des réseaux axés sur l'organisation de l'entreprise tout en s'affranchissant de certaines contraintes techniques, notamment celles liées à la localisation géographique des équipements.

La figure 18.24 illustre une réalisation calquée sur l'organisation de l'entreprise.

En fait, les VLAN introduisent la notion de segmentation virtuelle, qui permet de constituer des sous-réseaux logiques selon des critères prédéfinis (ports, adresses MAC, adresses IP...).

Un logiciel d'administration permet d'affecter chaque système raccordé à un commutateur à un réseau logique d'appartenance. L'affectation peut être introduite manuellement par l'administrateur station par station (VLAN statiques) ou réalisée automatiquement par rapport à un identifiant propre à la définition du VLAN comme un ensemble d'adresses MAC, d'adresses IP... (VLAN dynamiques).

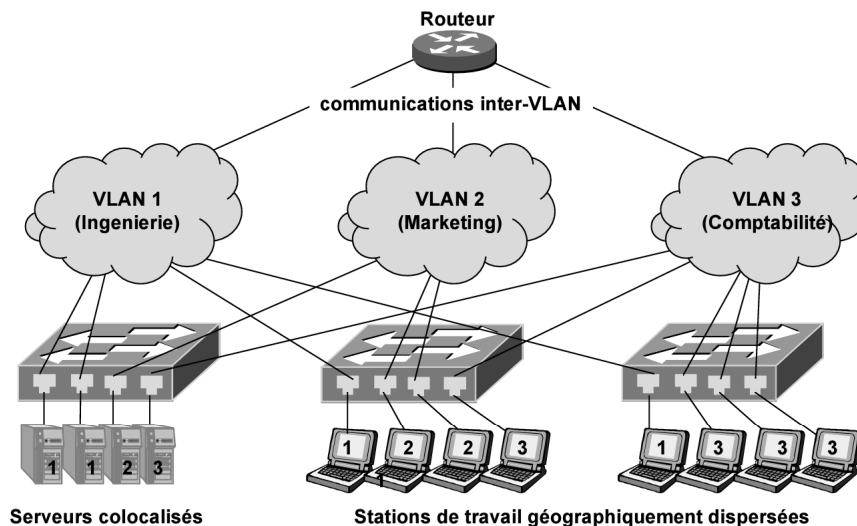


Figure 18.24 VLAN et organisation de l'entreprise.

18.2.2 Les différents niveaux de VLAN

Les échanges à l'intérieur d'un domaine sont sécurisés et les communications inter-domaines sont explicitement autorisées et contrôlées par les filtres configurés dans le routeur d'interconnexion.

L'appartenance à un VLAN étant définie logiquement et non géographiquement (figure 18.23), les VLAN permettent d'assurer la mobilité (déplacement) des postes de travail. Selon le regroupement effectué, on distingue :

- ❑ **Les VLAN de niveau 1 ou VLAN par port (*Port-based VLAN*)** : ces VLAN associent chaque port d'un commutateur à un VLAN. Une station raccordée à 1 port est automatiquement affectée au VLAN du port. Si le port est raccordé à un *hub*, toutes les stations de ce *hub* appartiennent au même VLAN (VLAN par segment). La configuration est statique (VLAN statique), le déplacement d'une station implique son changement de VLAN. C'est le mode le plus sécurisé, un utilisateur ne peut changer sa machine de VLAN. Un port, donc les stations qui lui sont raccordées, ne peut appartenir qu'à un seul VLAN.
- ❑ **Les VLAN de niveau 2 ou VLAN MAC (*MAC Address-based VLAN*)** : ces VLAN associent les stations par leur adresse MAC. De ce fait, deux stations raccordées à un même port (segment) peuvent appartenir à deux VLAN différents. Les relations adresses MAC/VLAN sont introduites par l'administrateur. En fonction du critère d'appartenance à un VLAN, ici l'adresse MAC, les ports déterminent automatiquement leur VLAN d'appartenance (VLAN dynamique). Il existe des mécanismes d'apprentissage automatique d'adresses (lecture des adresses MAC des stations raccordées), l'administrateur n'ayant plus qu'à effectuer les regroupements par simple déplacement et regroupement de stations dans le logiciel d'administration (*Drag&Drop*). Une station peut appartenir à plusieurs VLAN. Les VLAN de niveau 2 sont indépendants des protocoles supérieurs. La commutation, s'effectuant au niveau MAC, autorise un faible temps de latence.
- ❑ **Les VLAN de niveau 3 ou VLAN d'adresses réseaux (*Network Address-based VLAN*)** : ces VLAN sont constitués de stations définies par leur adresse réseau (plage d'adresses) ou par masque de sous-réseau (*Subnet* d'IP). Les utilisateurs d'un VLAN de niveau 3 sont affectés

dynamiquement à un VLAN. Une station peut appartenir à plusieurs VLAN par affectation statique. Ce mode de fonctionnement est le moins performant, le commutateur devant accéder à l'adresse de niveau 3 pour définir le VLAN d'appartenance. L'adresse de niveau 3 est utilisée comme étiquette, il s'agit bien de commutation (commutateur de niveau 3) et non de routage. L'en-tête n'est pas modifié.

Il est aussi envisageable de réaliser des VLAN par :

- ❑ protocole (IP, IPX...), la communication ne pouvant s'établir qu'entre stations utilisant le même protocole ;
- ❑ par application (N° de port TCP), la constitution des VLAN est alors dynamique, un utilisateur pouvant successivement appartenir à des VLAN différents selon l'application qu'il utilise ;
- ❑ par mot de passe (constitution dynamique des VLAN au login de l'utilisateur).

La figure 18.25 illustre ces différentes approches.

Les VLAN peuvent être définis sur un ou plusieurs commutateurs, que ceux-ci soient locaux ou distants. Cependant, il devra y avoir, entre chaque commutateur, autant de liens (physiques ou virtuels) que de VLAN interconnectés.

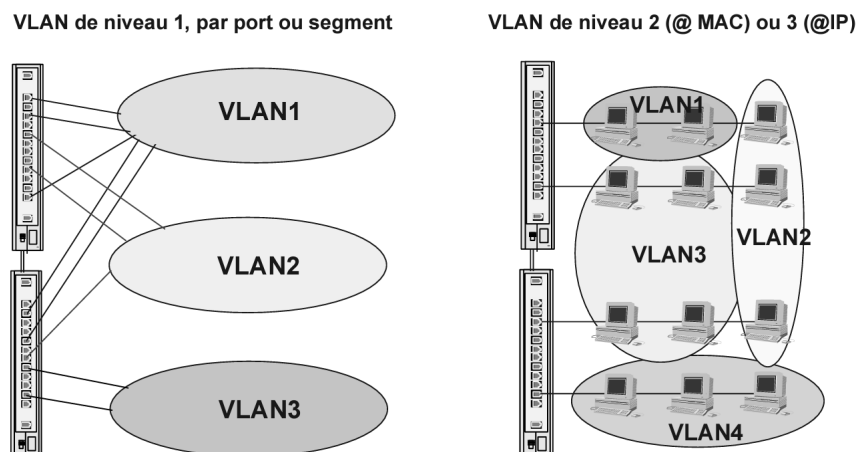


Figure 18.25 Différents niveaux de VLAN.

18.2.3 L'identification des VLAN (802.1Q)

Principe

Lorsqu'un réseau comporte plusieurs commutateurs, chaque commutateur doit pouvoir localiser toutes les machines (table d'acheminement) et connaître le VLAN d'appartenance de la source et du destinataire (filtrage de trafic). Lorsque le réseau est important, les tables peuvent devenir très grandes et pénaliser les performances. Il est plus efficace d'étiqueter les trames. L'étiquette identifie le VLAN de la station source, le commutateur n'a plus alors qu'à connaître les VLAN d'appartenance des stations qui lui sont raccordées. Ainsi, on distingue deux types d'équipement, ceux qui savent gérer l'étiquetage et qui ont connaissance des VLAN (les VLAN *aware*) et ceux qui ignorent cette appartenance (VLAN *unaware*).

Dans le réseau de la figure 18.19 cohabitent des équipements *aware* et *unaware*. Les trames émises par les équipements *aware* sont marquées (*tagged*), celles émises par les équipements *unaware* ne sont pas marquées (*untagged*). La mixité des équipements nécessite que soit défini un VLAN par défaut : le VLAN auquel sont rattachés les équipements *unaware* (VLAN C de la figure 18.26). Lorsqu'un équipement *aware* reçoit une trame marquée à destination d'un équipement *unaware*, il en extrait le *tag*.

La norme IEEE 802.1p/Q

► Identification des VAN, marquage des trames

Un VLAN correspond à un domaine de *broadcast*. Cependant, lorsque plusieurs VLAN sont définis sur un même segment, cette définition est mise en défaut. Il est évidemment possible d'imaginer que le commutateur transforme le *broadcast* en une rafale d'*unicast*. La solution adoptée par l'IEEE est toute différente : un seul VLAN peut être déclaré par port⁴, sauf pour les liaisons inter-commutateur supportant le trafic de VLAN différents (liens dits : *trunk link*, figure 18.26).

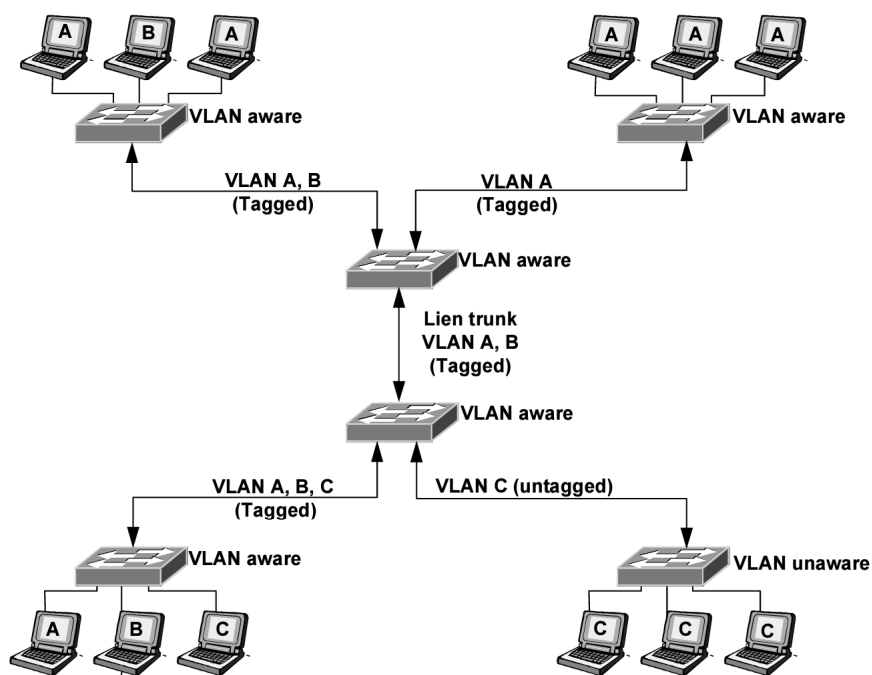


Figure 18.26 Principe de l'étiquetage des trames dans les VLAN.

Les VLAN sont définis dans les normes 802.1Q et 802.1p (802.1p/Q⁵) qui introduisent quatre octets supplémentaires dans la trame MAC. Ces quatre octets permettent d'identifier les VLAN

4. Certaines implémentations autorisent le raccordement de périphériques partagés par plusieurs VLAN (superposition de ports).

5. 802.1Q concerne les VLAN, 802.1p la qualité de service. Cependant, la recommandation 802.1p utilise le « tag » de la recommandation 802.1Q, c'est pourquoi ces deux recommandations, bien que différentes, sont généralement associées. La recommandation 802.1p a été fusionnée avec la recommandation 802.1D en 2004.

(VLAN tagging) et de gérer huit niveaux de priorité (COS, Class of Service, ou encore QoS). La figure 18.27 illustre l'étiquetage d'une trame MAC des réseaux de type 802.3.

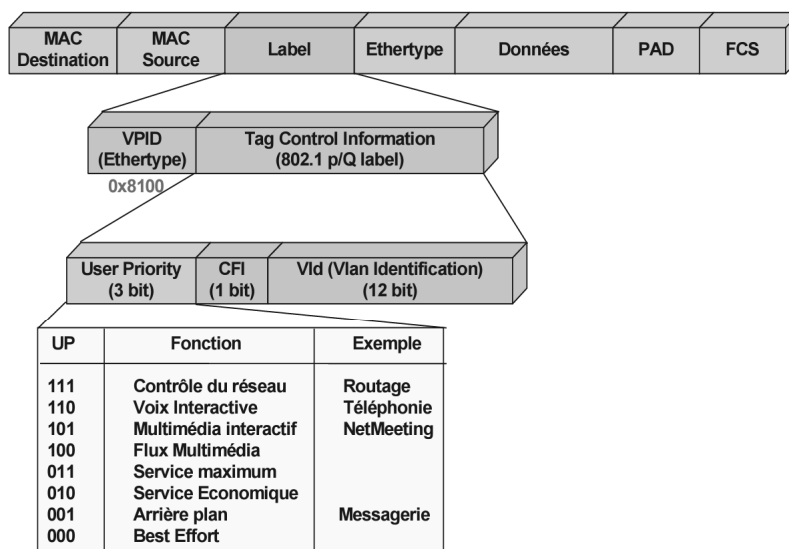


Figure 18.27 Format de la trame 802.1p/Q.

La trame 802.1p/Q augmente la taille de la trame 802.3, la taille maximale passe ainsi de 1 518 à 1 522 octets. Cet accroissement de la taille des trames limite l'usage des trames marquées aux équipements *aware*. En particulier, la plupart des équipements terminaux (station, périphérique) sont encore de type *unaware*, en conséquence, sur les liens d'accès aux commutateurs (*Access link*) ne circulent que des trames non marquées. Les trames sont identifiées par le port d'entrée et le *tag* est extrait par le port de sortie (figure 18.28). Lorsqu'un port d'un équipement *aware* reçoit une trame non marquée, le commutateur affecte le numéro du VLAN du port d'arrivée ou le numéro du VLAN par défaut si ce port n'a pas été configuré.

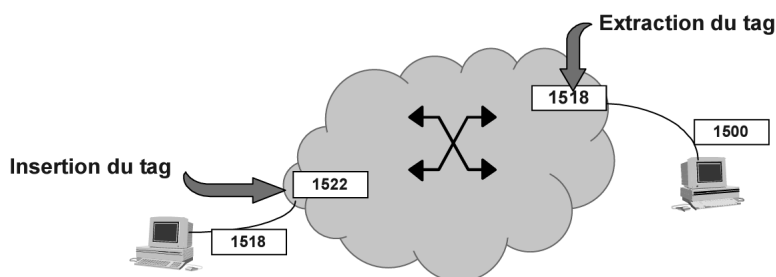


Figure 18.28 Identification des VLAN internes au réseau.

Pour garantir la compatibilité avec l'existant, le marquage des trames est vu comme une encapsulation supplémentaire. Ainsi, le champ **VPID** (*VLAN Protocol ID*) est similaire au champ Ethertype de la trame 802.3, il identifie le format 802.1p/Q, sa valeur est fixée à 0x8100 (figure 18.27). Les

2 octets suivants permettent de définir 8 niveaux de priorité (*User Priority*). Les commutateurs de dernière génération disposent de plusieurs files d'attente. Suivant leur niveau de priorité, les trames sont affectées à telle ou telle file.

Le bit *CFI* (*Canonical Format Identifier*) est, en principe, inutilisé dans les réseaux 802.3, il doit être mis à 0. Dans les réseaux Token Ring, à 1, il indique que les données du champ Routage par la source sont au format non canonique. Le champ *VID* (*VLAN Identifier*) identifie, sur 12 bits, le VLAN destination (4 094 VLAN identifiables). L'introduction de 4 octets supplémentaires implique que les commutateurs d'entrée et de sortie recalculent le FCS.

► La qualité de service dans les LAN

Dans les commutateurs, la QoS est définie par la recommandation 802.1p incluse dans 802.1D (1998). Cette recommandation définit 8 classes de services (champ de 3 bits) qui devraient correspondre à 8 files d'attente (traitement différencié). La plupart des commutateurs, pour ne pas dire tous, pour des raisons d'économie ne disposent que de 4 voire 2 files d'attente (figure 18.29).

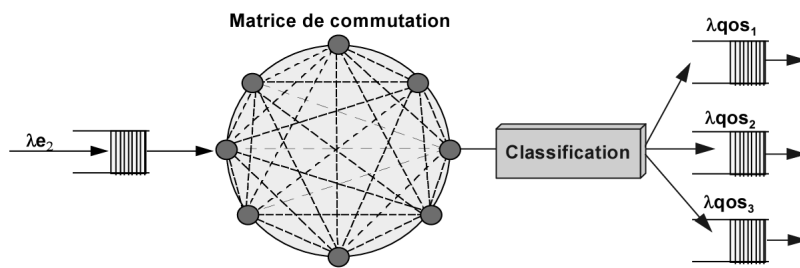


Figure 18.29 Gestion de la QoS dans les commutateurs.

Cependant, les constructeurs ne développent que 4 voire 3 files d'attente dans leur système, ce qui conduit à définir un « mappage » entre la priorité marquée et la priorité gérée. Il faut alors déterminer en fonction du marquage la file d'attente qui doit traiter le flux. Le tableau de la figure 18.30 indique, en fonction du niveau de priorité de l'application (marquage 802.1p) et en fonction du nombre de files d'attente disponibles à quelle file doit être affecté le flux.

		Nb. files d'attente →	Affectation à la file (file 7 plus prioritaire que la file 0)								
			1	2	3	4	5	6	7	8	
		Priorité ↓									
BK	Background	0	0	0	0	1	1	1	1	2	
BE	Best effort	1	0	0	0	0	0	0	0	0	
EE	Excellent effort	2	0	0	0	0	0	0	0	1	
CA	Critical application	3	0	0	0	1	1	2	2	3	
VI	Video < 100 ms (latence, jitter)	4	0	1	1	2	2	3	3	4	
VO	Voice < 10 ms (latence, jitter)	5	0	1	1	2	3	4	4	5	
IC	Interconnect control	6	0	1	1	3	4	5	5	6	
NC	Network control	7	0	1	1	3	4	5	6	7	

Figure 18.30 Affectation des flux aux files d'attente.

Dans cette typologie, un trafic marqué 0, 1, 2, 3, dans un système à 2 files, est affecté à la file 0, alors qu'un trafic marqué 4 et plus sera affecté à la file 1 (plus prioritaire). En principe, les données les plus prioritaires sont toujours émises avant les autres (*Priority Queuing*, ou *strict priority*), ce qui peut provoquer un blocage du trafic si la priorité la plus haute est trop invoquée. Il existe de nombreux autres algorithmes de gestion des files d'attente dont certains seront développés au chapitre 21. Cependant, une bonne gestion de la QoS dans les réseaux consiste toujours à dimensionner le système pour que les flux prioritaires ne correspondent qu'à un certain pourcentage (max. 50 %) de la bande passante globale offerte.

► Communication intra et inter-VLAN

Lorsqu'une trame doit être diffusée sur un port appartenant au même commutateur que le port d'origine, la communication est réduite au commutateur concerné. Mais lorsque le destinataire n'est pas situé sur le même commutateur, la trame doit être diffusée sur le réseau. Pour limiter la diffusion aux seuls équipements appartenant au VLAN concerné, les équipements *aware* annoncent périodiquement les VLAN qu'ils gèrent (figure 18.31). Le protocole **GVRP** (*GRAD VLAN Registration Protocol*) permet de faire connaître aux autres éléments du réseau les VLAN gérés.

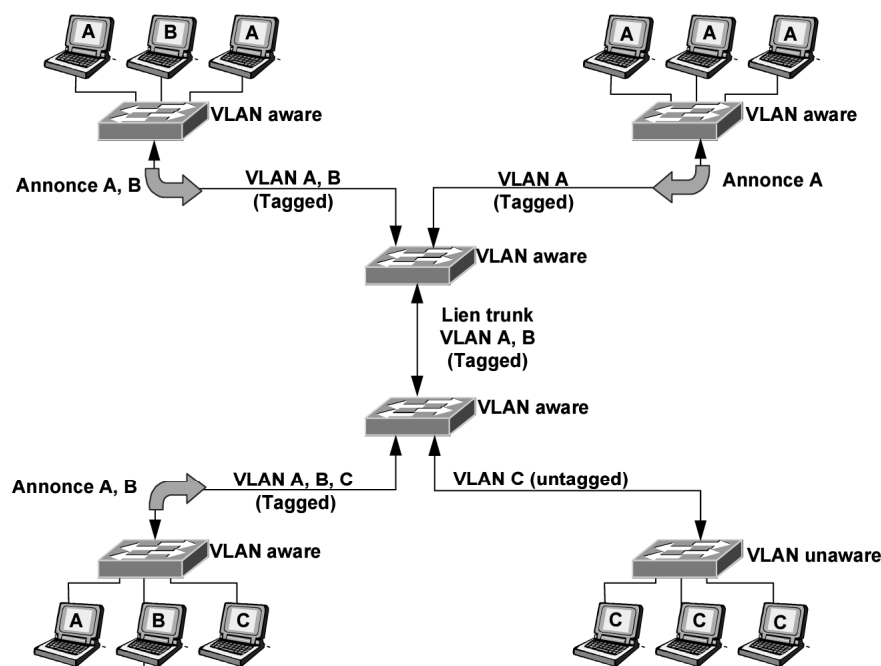


Figure 18.31 Annonces des VLAN (GVRP).

Les communications entre VLAN doivent, en principe, transiter par un routeur. Ce dernier doit posséder un attachement sur chaque VLAN routé. La notion d'interfaces virtuelles évite la multiplication des interfaces physiques sur le routeur. Celui-ci possède une seule interface physique reliée à un seul port du commutateur (lien et port *trunk*), sur le routeur plusieurs interfaces virtuelles peuvent être définies, à chaque interface virtuelle est associée un VLAN (figure 18.32). Lorsque le routeur reçoit une trame marquée, il retire le *tag*, consulte la table de routage et les filtres de trafic associés, il détermine alors l'interface virtuelle de sortie et la marque

en conséquence. L'acheminement peut aussi être réalisé directement par un commutateur enrichi de fonctions de routage (commutation de niveau 3).

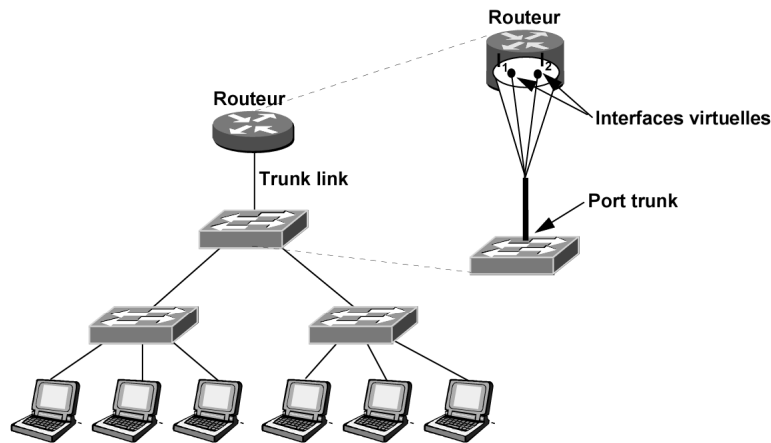


Figure 18.32 Notion d'interfaces virtuelles.

18.2.4 Les VLAN dans les réseaux d'opérateurs

Q in Q, IEEE 802.1ad (Provider Bridge Network)

Les VLAN constituent un moyen de cloisonnement de flux et donc de sécurité. Pour garantir, à l'utilisateur final, un service de bout en bout à travers un réseau public Ethernet métropolitain, il est nécessaire de prolonger les VLAN clients sur l'infrastructure partagée. La prolongation des VLAN « privés » sur une infrastructure publique se heurte à deux problèmes :

- ❑ le nombre de VLAN identifiables ou VID (VLAN ID) est limité à 4094 VLAN (12 bits, les valeurs 0 et 4095 étant réservées), cet espace est insuffisant pour un opérateur ;
- ❑ la collision d'identifiant entre deux clients différents du service de transport.

Aussi, l'IEEE (802.1ad, *Provider Bridge Network* ou encore « Q-in-Q ») a-t-elle défini deux niveaux de VLAN :

- ❑ Les VLAN clients dits « *Customer VLAN* » (**C-VLAN**, Ethertype 0x8100) identifiés par le champ C-VID (*Customer VID*).
- ❑ Les VLAN opérateurs dit *Service VLAN* (**S-VLAN**, Ethertype 0x88A8) identifiés par le champ S-VID (*Service VID*) sur 12 bits aussi.
- ❑ L'opérateur peut donc ainsi transporter dans un « Super » VLAN, les VLAN clients (C-VLAN). La recommandation 802.1ad ajoute à la trame Ethernet 4 octets d'identification des VLAN (figure 18.33) portant sa taille à 1 526 octets et nécessitant l'activation dans les commutateurs de l'option *jumboframe*. La signification des champs est identique à celle de la recommandation 801.1Q, sauf en ce qui concerne le bit CFI (802.1p) remplacé par le bit **DEI** (*Drop Eligible Indicateur*, en cas de congestion trame à écarter en priorité).
- ❑ La notion de S-VLAN correspond sur une infrastructure Ethernet à celle de VPN (*Virtual Private Network*) des réseaux MPLS.