

Les transmissions multicasts

(/home/terre/d01/adp/bcousin/Polys/multicast.fm- 25 Octobre 1999 06:42)

PLAN

- Introduction
- L'adressage multicast
- Le protocole IGMP
- Le protocole DVMRP
- Le protocole MOSPF
- Les protocoles PIM
- Conclusion

Bibliographie

- K. Savetz & al., Mbone multicasting tomorrow's internet, IDG Books, 1995
- C. Huitema, Le routage dans l'Internet, Eyrolles, 1995
- W. Stallings, High speed networks, Prentice Hall, 1998
- C. Comer, TCP/IP: architectures, protocoles, applications, InterEditions, 1998

1. Introduction

1.1. Présentation

De nombreuses applications nécessitent d'adresser les mêmes informations à **plusieurs** destinataires :

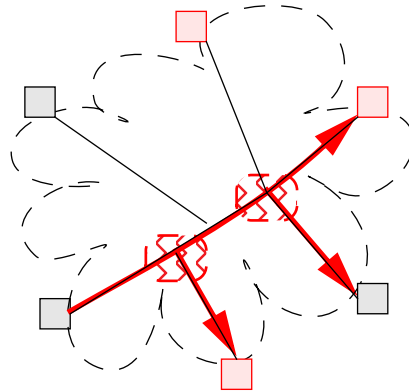
- Téléconférence : audioconf., vidéoconf., etc.
 - . ex : CU-SeeMe, IVS, Netmeeting
- Outils collaboratifs : “whiteboard”, éditeur partagé, etc.
- Radio-diffusion, web phone, etc.
- Distribution information : messagerie électronique, applications réparties

Bâtir le réseau de distribution :

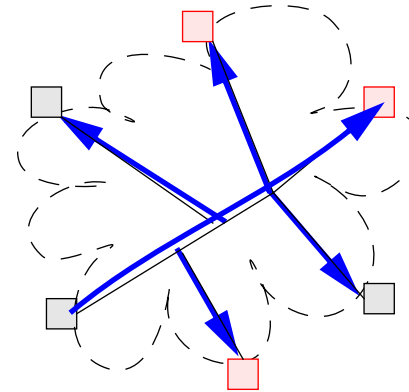
- soit au niveau **applicatif** :
 - . nombreuses recopies, cheminement sous-optimal, à base de liaisons bi-points
 - ex : “reflectors” de CU-SeeMe
- soit de le faire au niveau du **réseau de transport** :
 - . notion de liaison multipoint
 - . le réseau connaît le meilleur chemin !

- . bonne efficacité

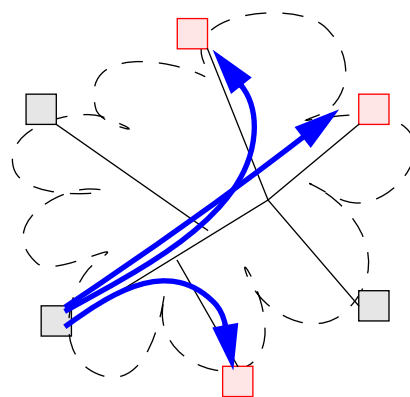
Quelques techniques de diffusion :



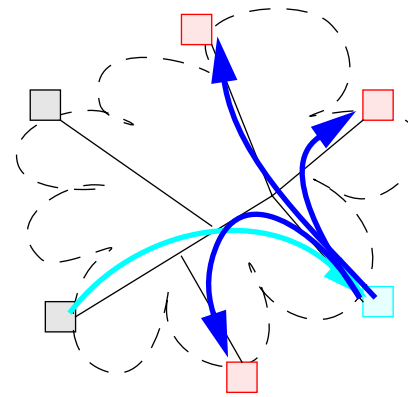
Diffusion sélective



Diffusion totale



Multi- monotransmission



Serveur de diffusion

1.2. On distingue différents types de protocoles multicasts :

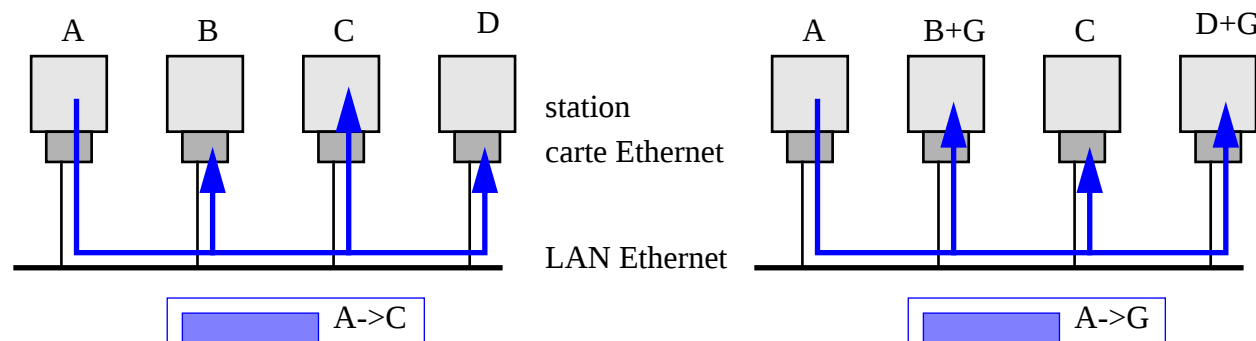
- les protocoles du routage
 - ex : IGMP, PIM, MOSPF, DVMRP, etc
- les protocoles chargés de la transmission des données multicasts
 - ex : RTSP, MTP, TFTP, SPAM et RSVP, RTP, etc
- les protocoles chargés de la gestion des groupes :
 - ex : SAP, SIP, SCCP, SDP, etc

 les protocoles de routage

1.3. La diffusion dans les réseaux locaux

Le partage du même support physique entre toutes les stations fait que **le service de diffusion est naturel** dans les réseaux locaux.

- Que les trames soient adressées à une, plusieurs ou toutes les stations, toutes les cartes d'accès au réseau local **reçoivent systématiquement une copie de toutes les trames** circulant sur le RL.
- Même coût entre une simple transmission ou une diffusion partielle ou totale
- Les cartes d'accès recevant une trame qui ne leur est pas adressée, l'ignorent simplement!
 - note : les cartes d'accès n'ayant que cela à faire, les traitements intempestifs et superflus ne sont pas pénalisants

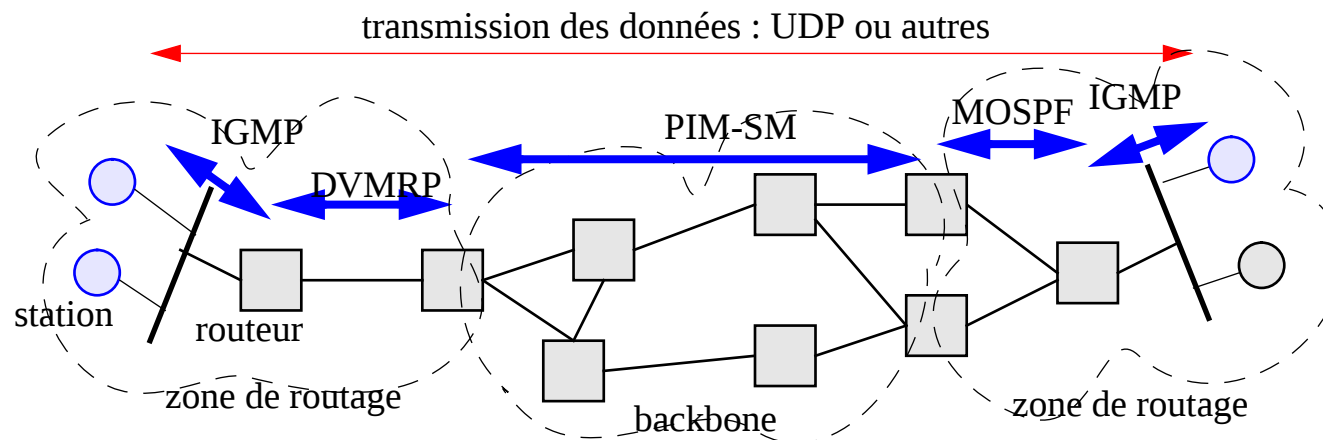


On rappelle que les LAN disposent d'un adressage avec la notion d'adresse de groupe (bit G).

1.4. Les protocoles de routage multicasts

On discerne 3 niveaux de routage multicast :

- le protocole IGMP régit les informations de localisation des stations appartenant à un groupe, informations échangées entre les stations et les routeurs multicasts.
- les protocoles de routage multicast (par exemple : DVMRP ou MOSPF) régissent les informations de routage échangées entre routeurs d'une zone de routage réduite.
- les protocoles de routage multicast (par exemple : PIM) régissent les informations de routage échangées entre routeurs multicasts d'une zone de routage externe ("backbone").



2. L'adressage multicast

2.1. Introduction

Une adresse multicast **identifie un groupe** (de stations)

- Adresse IP de classe D, de préfixe “1110₂”,
- Adresse comprise entre 224.0.0.0 et 239.255.255.255

Toutes les stations possédant cette adresse multicast reçoivent une copie de chaque paquet ayant pour adresse de destination celle-ci.

Pour émettre un datagramme multicast une station n'a pas besoin d'appartenir au groupe.

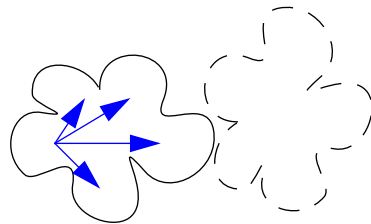
Attention :

- Utilisation interdite des adresses multicasts comme adresse d'émission.
- Contrôle des règles de gestion du groupe : hors du domaine d'IP
- Ne pas faire l'hypothèse que les datagrammes multicasts émis ne sont uniquement reçus que par les récepteurs envisagés. Le partage de la même adresse multicast par plusieurs applications est possible et doit être résolu au niveau applicatif.

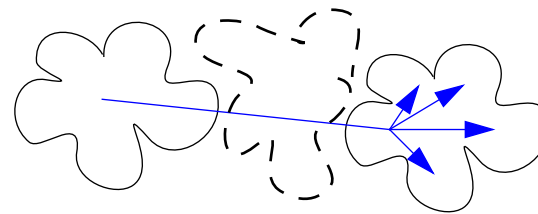
2.2. Adresses multicasts réservées

2.2.1 Diffusion

- Plein “1” !
 - **Diffusion locale** : 255.255.255.255
 - . diffusion à toutes les stations du réseau IP local (à l’émetteur)
 - **Diffusion distante** : netid-A.255.255.255, netid-B.255.255, netid-C.255
 - . diffusion à toutes les stations sur réseau IP distant identifié par “netid”



Diffusion locale



Diffusion distante

Note : dû à une implémentation historiquement erronée, les adresses comportant l’hostid à 0 peuvent être interprétées comme requérant la diffusion.

2.2.2 Les groupes particuliers

- Chaque adresse identifie un ensemble de stations (routeurs) rendant un service particulier
- Optimisation : supprime les interruptions intempestives
- Exemples d'adresses de groupes particuliers :
 - 224.0.0.1 : toutes les machines (stations + routeurs) appartenant à un groupe quelconque du réseau IP local
 - 224.0.0.2 : tous les routeurs du réseau IP local
 - 224.0.0.9 : tous les routeurs RIP-2 appartenant au réseau IP local

2.3. Résolution directe d'adresse

- adresse multicast IP ➡ adresse MAC de groupe

- adresse IP **multicast** :

1110 xxxx xabc defg hijk lmno pqrs tuvw

23 bits de poids faibles

- adresse IEEE 802 **de groupe** :

0000 0001 0000 0000 0101 1110 0abc defg hijk lmno pqrs tuvw

Note :

- . plusieurs (16) adresses IP multicasts sont associées avec la même adresse IEEE 802
par exemple : 224.1.2.3 ou 225.129.2.3 ➡ 01 00 5E 01 02 03



3. Le protocole IGMP

3.1. Introduction

“Internet group management protocol”

RFC 1112 : “Host extensions for IP multicasting”

Protocole utilisé par les stations pour **informer les routeurs multicasts des groupes actifs**.

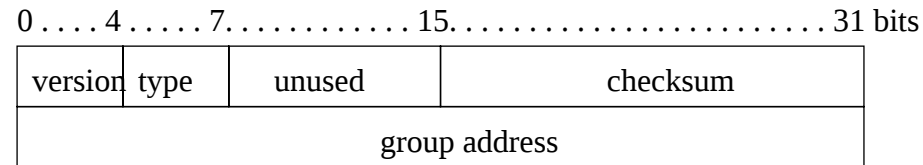
Définition : un groupe est actif localement à un réseau IP si au moins une station de ce réseau IP appartient au groupe.

Les messages IGMP sont encapsulés dans des datagrammes IP :

- champ Protocol du datagramme = 2

3.2. Format général du message IGMP

Message de taille fixe : 8 octets



Version :

- Version courante 1 (rfc 988 : ancienne version 0)

2 types de messages ICMP :

- “Host membership **query**” = 1
- “Host membership **report**” = 2

Calcul du checksum :

- somme de mots de 16 bits en complément à 1
- même procédé que TCP, UDP ou IP.

Group address :

- L’adresse IP multicast identifiant le groupe

3.3. Principe

Pour **adhérer à un groupe** :

- une station émet un message “IGMP report”
 - le champ “group address” contient l’adresse du groupe auquel la station veut adhérer.
- le message ICMP est encapsulé dans un datagramme dont le champ “Destination address” est l’adresse du groupe
 - 👉 les routeurs multicasts sont à l’écoute de tous les datagrammes multicasts
- cette première transmission est répétée après un délai aléatoire
 - 👉 lutte contre les pertes

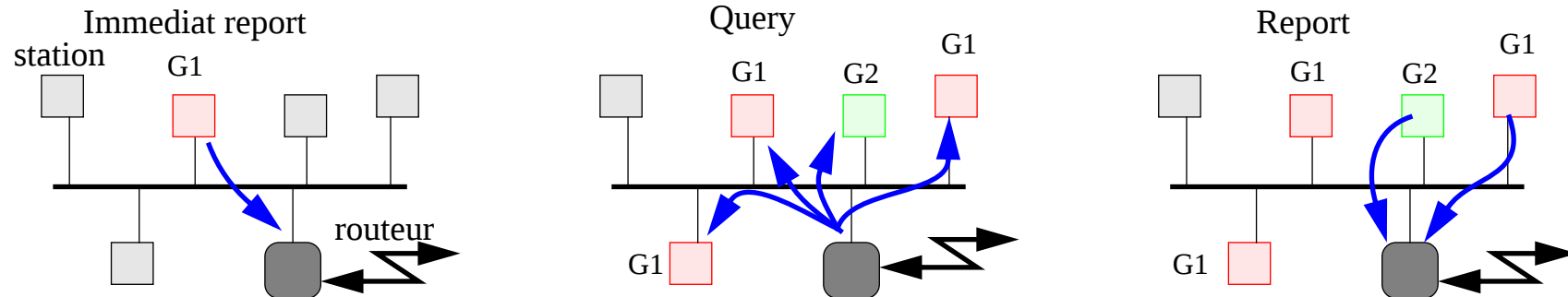
Les routeurs multicasts **surveillent les groupes actifs** :

- les routeurs envoient un message “IGMP query”
 - périodiquement (mais pas trop souvent pour limiter la surcharge) : $> 1mn$
 - permettent à un routeur de rafraîchir sa connaissance
 - 👉 panne de station
- le message ICMP est encapsulé dans un datagramme dont le champ “Destination address” = 224.0.0.1 et le champ TTL = 1.

Chaque station appartenant à un groupe **répond à l'interrogation** :

- en émettant un message “**IGMP report**” après un **délai aléatoire** [0 - 10 s] !
- le message ICMP est encapsulé dans un datagramme dont le champ “Destination address” est l'adresse du groupe
- si avant l'écoulement du délai une autre station appartenant au même groupe répond l'émission est annulée

☞ en général une seule réponse par groupe (minimisation du trafic)



Une station **quitte un groupe** :

- simplement en cessant tout échange
- si cette station est la dernière du groupe, lors de la prochaine interrogation le routeur ne recevra aucun rapport concernant ce groupe.

☞ le routage peut ne pas être optimal pendant qq instants

3.4. Gestion des erreurs

Les datagrammes dont le checksum est erroné ou dont l'adresse de destination du datagramme ne correspond pas au champ "Group address" sont détruits silencieusement.

Les pertes de messages IGMP sont traitées par répétition :

👉 transitoirement la distribution des messages multicasts peut être incorrecte

La panne d'un routeur multicast peut isoler les stations d'un groupe

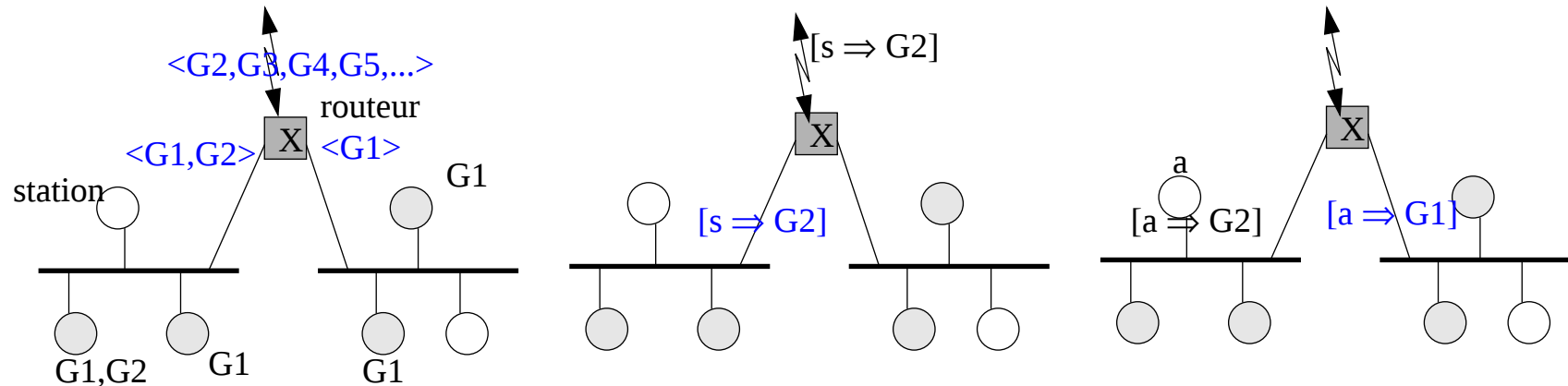


3.5. Routage local des datagrammes multicasts

Un routeur recevant un datagramme muni d'une adresse multicast identifiant un groupe le diffuse sur tous les interfaces où le groupe est actif (sauf celui d'où provient le datagramme).

Un routeur multicast n'a besoin de ne connaître que l'activité d'un groupe. Il ne mémorise que cette seule information.

☞ chaque routeur pour chacune de ses interfaces maintient la liste des groupes actifs.



Toute machine multicast est abonnée au groupe 224.0.0.1. Ce groupe est considéré comme toujours actif. Aucune action de surveillance n'est nécessaire pour ce groupe.

4. Le protocole DVMRP

4.1. Présentation

“Distance vector multicast routing protocol”: rfc 1075

- procédé simple de routage multipoint, implémenté par le processus “mrouted”
- utilisé par les routeurs du Mbone
- même principe que RIP-2 : “distance vector”
- utilise l’algorithme du RPF avec élagage

Les routeurs échangent avec leurs voisins des messages de mise-à-jour du vecteur de distance :

- liste de couples <destination + distance>
 - destination : l’adresse et le masque de sous-réseau de l’émetteur multicast !
 - distance : nombre de routeurs intermédiaires entre la destination et le routeur local

Chaque routeur gère ses tunnels :

- l’adresse du routeur de l’autre extrémité du tunnel
- le coût du tunnel (≥ 1)
- un seuil

Chaque routeur mémorise pour chacune de ses interfaces

- si cette interface héberge un autre routeur proposant un plus court chemin vers l'émetteur

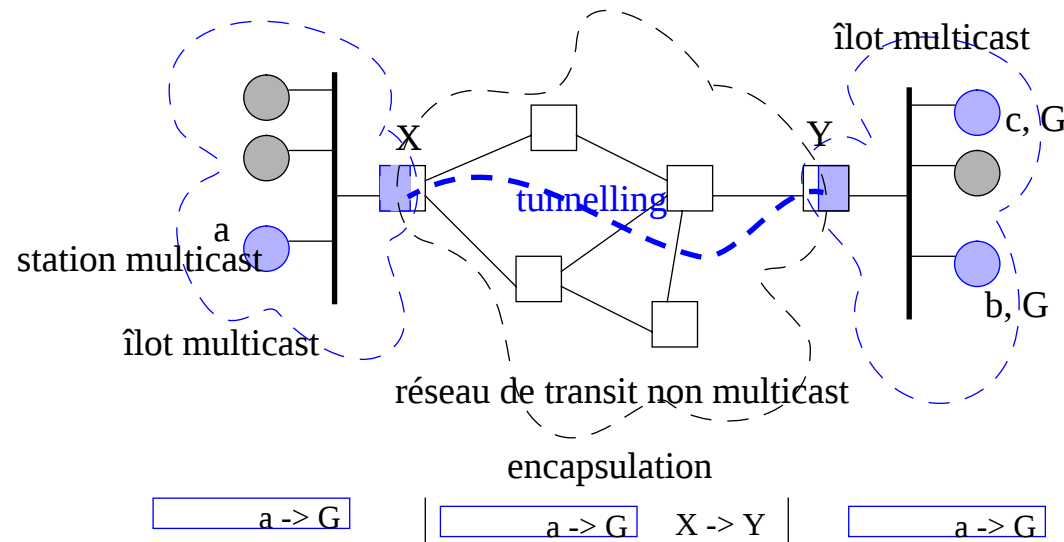
Les routeurs du Mbone relayent un paquet sur une interface

- seulement si son **TTL est plus élevé que son seuil**
 - limite la portée des paquets multicasts
 - valeur conventionnelle des seuils des routeurs en frontière :
 - . d'une organisation = 32
 - . d'une région du monde = 64
 - . d'un continent = 128
 - identique au champ spécifique "scope" de IPv6

4.2. L'infrastructure du Mbone

“**Multicast Backbone**” : Interconnexion de réseau IP multicasts

- chaque réseau multicast forme un îlot
- interconnectés par des “tunnels”
 - liaisons virtuelles
- utilise un réseau de transit non multicast : l'Internet

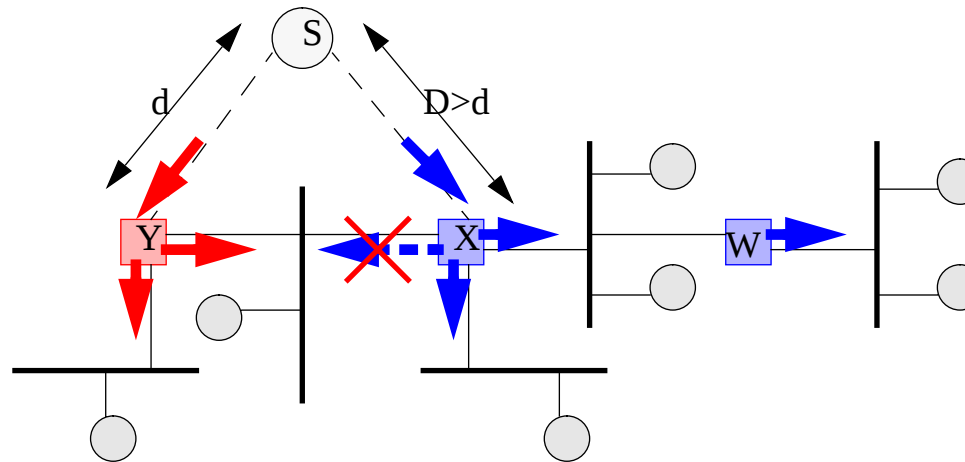


Encapsulation des datagrammes multicasts dans des datagrammes unicasts (IP in IP) :

- encapsulation au routeur d'interface réseau multicast/réseau de transit
- désencapsulation à l'autre routeur

4.3. L'algorithme RPF ("Reverse path forwarding")

- Quand un routeur multicast reçoit un paquet multipoint sur une interface il vérifie si cette interface est celle du (plus court) chemin **vers l'émetteur** :
 - si ce n'est pas le cas le paquet est détruit
 - sinon il relaye alors le paquet **sur toutes les interfaces multipoints** sauf celles qui hébergent un routeur proposant un plus court chemin vers l'émetteur



- Cet algorithme construit implicitement un **arbre recouvrant** dont la racine est l'émetteur
- Cet arbre a 2 propriétés :
 - chemin du plus court de la source à chaque destination
 - l'arbre dépend de la source

👉 pour un même groupe la charge des différentes sources est répartie

4.4. l'algorithme RPF avec élagage

Dans la version précédente tout paquet multicast est diffusé à tous les routeurs du réseau mondial si les seuils le permettent.

☞ risque de congestion

- les routeurs terminaux, s'ils constatent qu'il n'y a aucun abonné sur leurs réseaux locaux détruiront le paquet

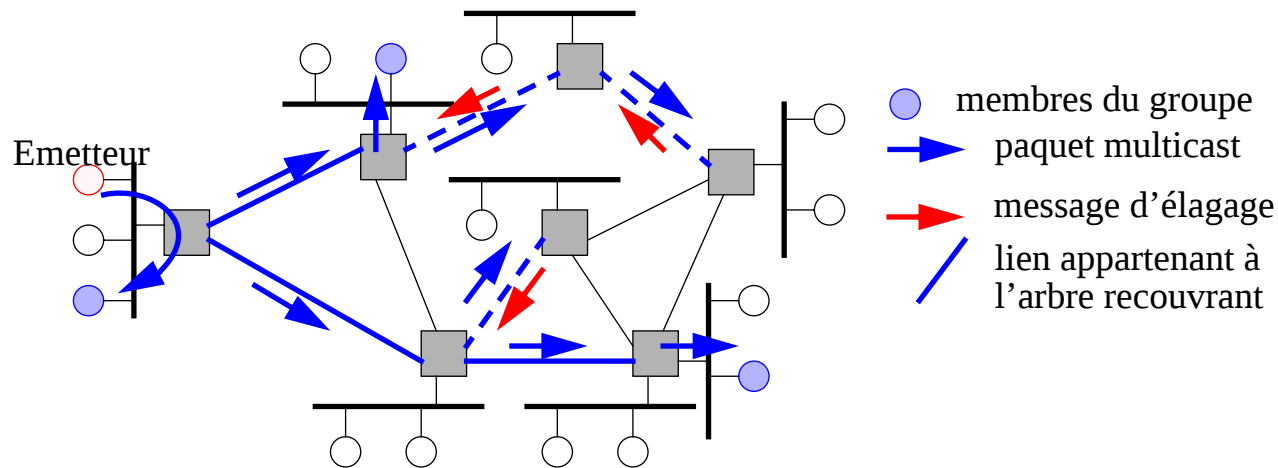
On se propose d'élaguer les branches inutiles :

- le routeur terminal (aval) lors de la réception d'un paquet multicast va envoyer
- un **message d'élagage** ("pruning") :
 - "n'envoyez plus de paquets multicasts ayant pour émetteur, pour le groupe et sur l'interface considérés"
- au routeur amont (qui mène à l'émetteur)

De manière récursive :

- si le routeur amont reçoit sur toutes ses interfaces aval un message d'élagage
- il envoie un message d'élagage à son propre routeur amont

- Exemple :



Après convergence, l'arbre ne contiendra que les seules branches qui mènent à un abonné.

Ces informations ne sont **mémorisées qu'un court instant** au sein des routeurs :

- périodiquement des paquets de données du groupe seront propagées
 - s'il n'y a toujours pas d'abonné un message d'élagage est reçu en retour
 - s'il y a de **nouveaux abonnés**, il reçoivent alors les paquets multicasts
- si **l'émetteur se tait**
 - l'état des routeurs sera nettoyé naturellement

5. Le protocole MOSPF

5.1. Présentation

La technique précédente (Mbone + DVMRP) est inadaptée si on envisage la généralisation des transmissions multicasts:

- La multiplicité des tunnels :
 - maillage complet des liaisons virtuelles
 - gestion spécifique et lourde
 - induit des répétitions multiples des mêmes paquets sur la même interface
- technique précédente peu efficace:
 - inondation + élagage ... puis inondation + élagage puis ...

MOSP : **multicast OSPF** (“Open short path first”)

- rfc 1584 : “Multicast extensions to OSPF”
- routage à l’intérieur d’un système autonome (AS)
- définit de nouveaux enregistrements :
 - permet de prendre connaissance des groupes actifs

5.2. Principe

MOSP, c'est :

- Extension du protocole OSPF : version 2 (rfc 1247)
- **Protocole à état des liaisons** (“link-state protocol”) :
 - chaque routeur diffuse à tous les routeurs l'état des liaisons
 - chaque routeur obtient donc une connaissance de la totalité réseau
 - chaque routeur peut calculer localement les meilleures routes
- Définit de **nouveaux enregistrements** :
 - chaque routeur prend connaissance des groupes actifs
- Calcul local des **arbres RPF avec élagage**

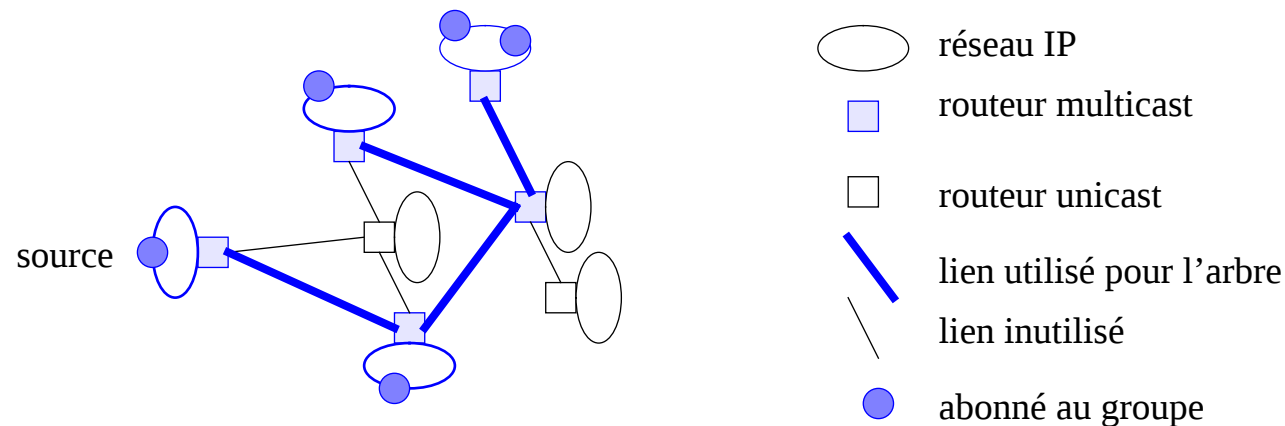
Lorsqu'un routeur MOSPF reçoit un paquet multicast (s --> G) :

- il utilise la BdD locale des liaisons
- pour calculer localement un **arbre recouvrant ayant pour racine s**
- il utilise la connaissance de la répartition des abonnés au groupe G pour **élaguer l'arbre**
- le paquet multicast est relayé vers les bons interfaces

5.3. Quelques raffinements

La zone de routage peut être hétérogène :

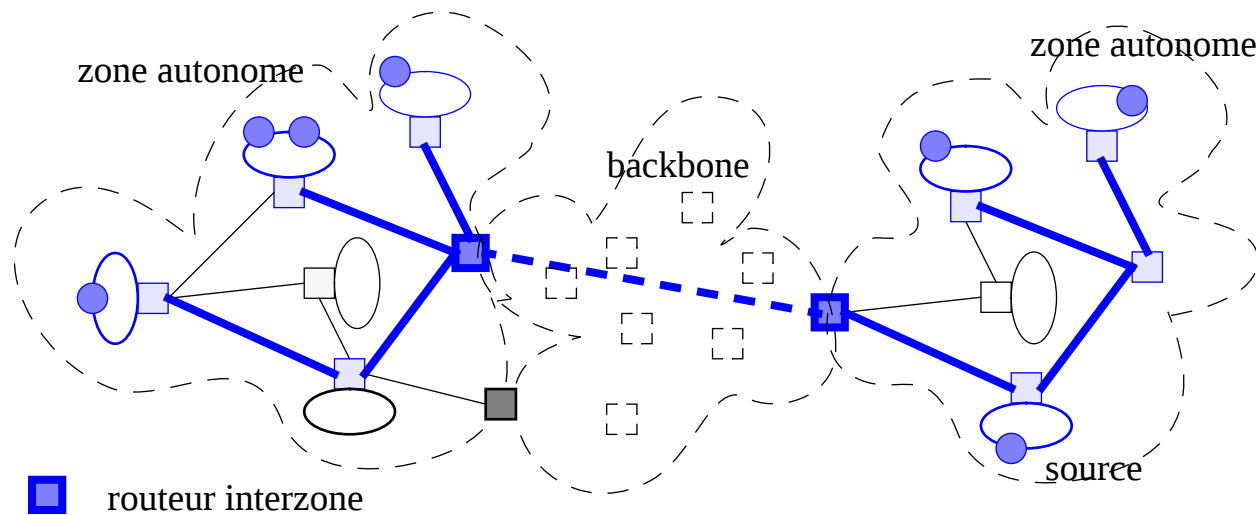
- les routeurs multicasts ou non seront désignés par un bit spécifique du champ *option* des messages OSPF
- les chemins multicasts seront construits de telle sorte qu'ils ne passeront que par des routeurs multicasts



Communication entre les zones de routage :

- Les routeurs interzones collectionnent tous les abonnés de leur zone
 - ils annoncent sur le “backbone” les groupes actifs de leur zone
- Il peut y avoir plusieurs routeurs interzones pour une même zone
 - la métrique permet de n'utiliser que le routeur situé sur le meilleur chemin

- un procédé parfaitement défini permet de séparer les ex-aequos
- Un routeur interzone est toujours **noeud de l'arbre local** :
 - il reçoit une copie de tous les paquets multicasts émis localement, qu'il propage sur le "backbone"
- Un routeur interzone est **racine de l'arbre local** lorsque l'émetteur est externe



6. Le protocole PIM

6.1. Présentation

Si on envisage un réseau très large (le “backbone”) **MOSPF n'est pas adapté**

- le nombre de calculs croît avec le carré de la taille de la zone
- trop de diffusion de messages
- même si on lance les calculs “à la demande” : quand arrive le 1^{er} paquet

 le protocole PIM (“Protocol independent multicast”)

2 protocoles PIM :

- PIM en mode dense (“dense mode”)
- PIM en mode clairsemé (“sparse mode”)
 - dépend de la probabilité de trouver un membre du groupe dans une zone

PIM en mode dense :

- implémente le routage RPF avec élagage
- ressemble à DVMRP mais n'utilise pas des tables de routage spécifiques :
 - on utilise les tables de routage du protocole de routage point-à-point

- on fait l'hypothèse que les chemins sont symétriques :
 - 👉 le chemin le plus court est le même à l'aller et au retour

6.2. Le protocole PIM en mode clairsemé

Protocole proposé pour le routage entre “systèmes autonomes de routage”

Pour les groupes clairsemés :

- le nombre de zones membres du groupe est très petit devant le nombre total de zones dans l’Internet
- cependant l’étendue du groupe est très vaste

Un routeur est considéré comme un routeur terminal d’un groupe

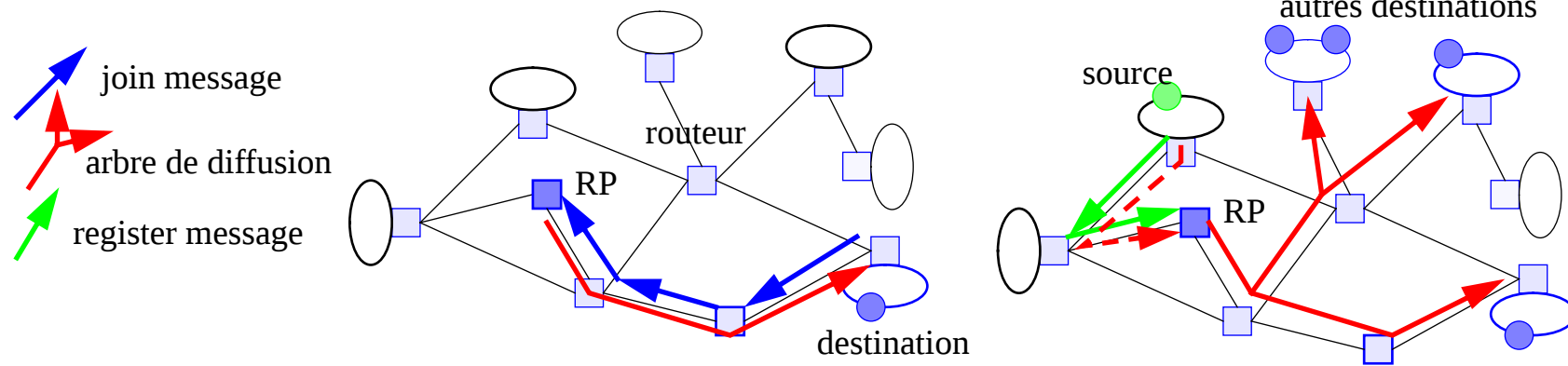
- s’il est connecté directement à un réseau qui héberge au moins un membre du groupe

6.3. Principe de fonctionnement: de PIM-SM

- Un routeur est choisi comme **le RP** (“rendez-vous point”) du groupe
 - les RP doivent être répartis uniformément
- Un routeur terminal d'un groupe envoie un message :
 - un message “d'**inscription au groupe**” (“join message”)
 - vers le RP du groupe, en point-à-point

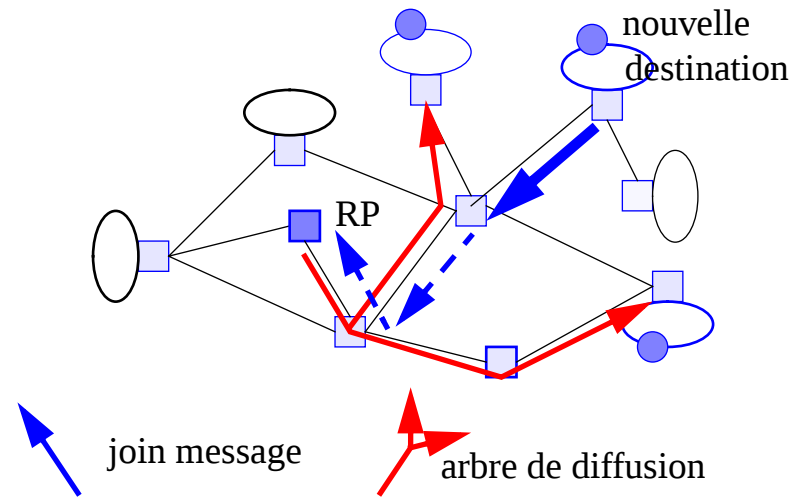
☞ le chemin suivi deviendra la branche de l'arbre de diffusion des paquets de données multicasts (utilisée en sens inverse !) vers les destinataires gérés par le routeur terminal

- Tout émetteur envoie un message :
 - un message d'**enregistrement** (“register message”)
 - vers le RP en utilisant le chemin unicast le plus court



6.4. Optimisations

Les messages adressés au RP peuvent être traités par un routeur situé sur le chemin et qui fait déjà parti de l'arborescence du groupe



Jusqu'à maintenant,

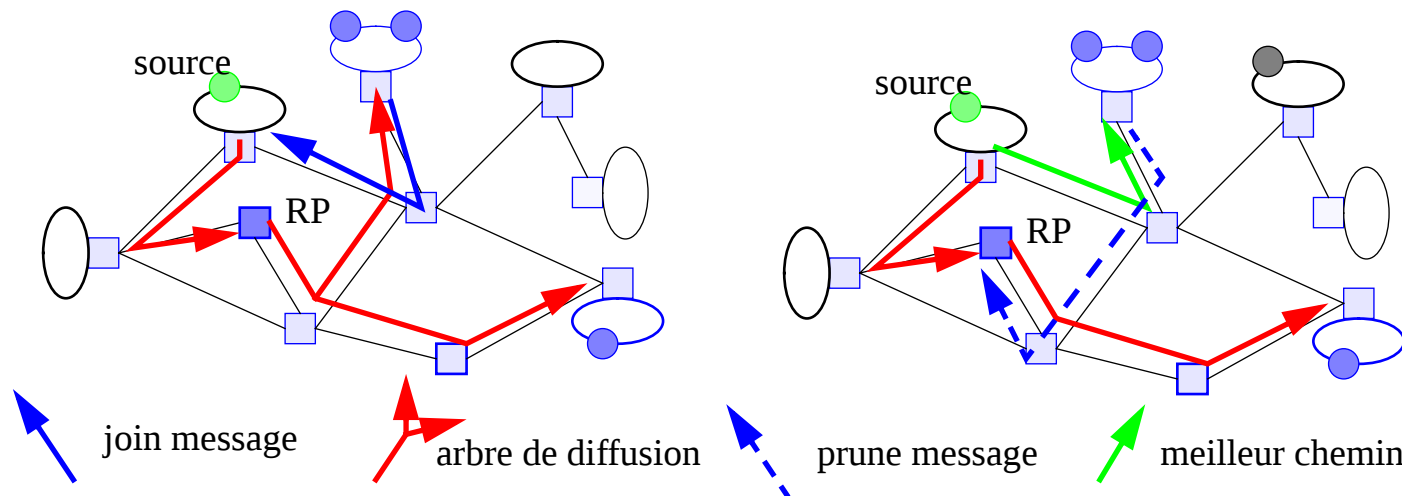
- Les paquets multicasts :
 - suivent **le plus court chemin** de l'émetteur vers le RP
 - utilisent un arbre de diffusion pour **minimiser le nombre de duplications**
- Les paquets de routage :
 - sont **en nombre limité** puisque seuls les routeurs terminaux en émettent

Il persiste un inconvénient :

- le chemin suivi par les données n'est pas le plus court
 - 👉 le routeur terminal peut décider d'utiliser **le meilleur chemin** pour recevoir les paquets multicasts d'un groupe

Construction d'un meilleur chemin :

- le routeur terminal prévient l'émetteur
 - envoie un “**join message**” directement à l'émetteur
- L'émetteur émet à la fois vers le RP et le routeur terminal
- le routeur terminal dès qu'il reçoit des données provenant du meilleur chemin :
 - envoie une **message d'élagage** vers le RP
 - ☞ les paquets multicasts utilisent alors uniquement le chemin le plus court entre l'émetteur et le routeur terminal



Note : toutes les sources continuent d'émettre leurs données vers le RP et donc les autres destinations continuent d'utiliser l'arbre de diffusion

7. Conclusion

7.1. Conclusion générale

Le service de multicast est un service utile.

Le procédé d'acheminement et l'adressage IP permet aisément à un paquet d'atteindre ses destinataires, pour peu que les tables de routage soient correctement configurées.

De nombreux protocoles sont chargés de cette mise à jour :

- local entre stations et routeurs terminaux : IGMP
- au sein d'une zone :
 - “distance vector” : DVMRP
 - “link state” : MOSPF
- entre inter-zone : PIM-SM

Actuellement le Mbone se développe doucement.

Autres protocoles :

- de gestion des groupes
- de transport de données fiables (TCP est inadapté au multicast)

7.2. L'adressage multicast dans IPv6

Les fonctionnalités de IGMP sont incluses dans ICMPv6

Ajout de la fonction “leave group”

Modification de l'adressage :

IPv6 address : 16 octets



- Champ *Flag* [4 bits] :
 - adresse affectée de manière permanente ou non
- Champ *Scope* [4 bits]
 - 1 : étendue restreinte au noeud
 - 2 : étendue restreinte à la liaison
 - 5 : étendue restreinte au site
 - 8 : étendue restreinte à l'organisme
 - E : étendue mondiale
- Champ *Group identifier* [14 octets]