

**Introduction à la
problématique des
Réseaux avec QoS
- Internet Multicast -**

Ingénierie des Réseaux d'Entreprise (Cycle C),
Compléments Réseaux de Transport et Application (Cycle B)

Janvier 2002

Eric Gressier-Soudan

Multicast vs Unicast

Applications :

- . visio/vidéo-conférence sur l'Internet
- . distribution audio
- . mode "push" sur le Web
- . gestion du routage
- . synchronisation d'horloge (avec NTP)
- . annonce de services/découverte de ressources sur un réseau
- . mise à jour de données (gestion de stock)

Effacité multicast p/r au broadcast : 1 message au lieu de n

- les routeurs ne traitent les messages du flot qu'une seule fois,
 - seuls les membres du groupe sont atteints par les messages multicast
- > économie de bande passante et de ressources

NB : Pour limiter la portée des datagrammes multicast, on joue sur le TTL, par exemple un TTL de 1 ne fait pas dépasser le premier routeur d'un réseau.

Multicast et Adresses Classe D

adresses unicast -> adresses un seul site (classes A,B,C)

adresses broadcast -> tous les sites d'un réseau (adresse unicast avec le champ station tout à 1)

adresses multicast -> un groupe de stations, adresses de classe D, transverse au réseau et à une localisation donnée

dans la classe D:

- . les adresses 224.0.0.x et 224.0.1.x (w=0 à 255) sont réservées pour des protocoles de routage ou autres protocoles de service :
 - 224.0.0.1 : les routeurs multicast du sous-réseau (IGMP)
 - 224.0.0.2 : tous les routeurs du sous-réseaux
 - 224.0.0.4 : tous les routeurs DVRMP
 - ...
- . les adresses 239.0.0 à 239.255.255.255 servent pour des adresses multicast privées, elles sont non routables à l'extérieur des entreprises, en particulier 239.253.0.0/16 est limité à un site

Adresses Multicast LAN et Multicast IP

IPmc multicast avec adresses MAC des LANs 802.x: on utilise les adresses multicast des LANs.

On extrait 23 bits de l'adresse IP classe D (sur les 32 - 4, soit 28 disponibles) qui correspondent à la fin de l'adresse

Les deux 1^{ère} lignes donnent le rang de la colonne du bit (dizaines)

```
0000..0000.0..011 1111 1111 2222 2222 2233
0123 4567 8 901 2345 6789 0123 4567 8901
```

```
1110 [^^^^ ^][^^^ ^^^^ ^^^^ ^^^^ ^^^^ ^^^^]
      < 5b > < 23b >
```

On ajoute ces 23 bits à la fin d'une adresse MAC préfixée par 0x01-00-5E, l'adresse (OUI pour Organization Unit Identifier) attribuée par l'IEEE avec le bit multicast à 1, le bit adresse universelle à 0, suivi d'un bit à 0 :

```
[01-00-5E] 0[^^^ ^^^^ ^^^^ ^^^^ ^^^^ ^^^^]
< 48b >
```

Cette approche est adoptée pour Ethernet et FDDI, alors que le Token-Ring utilise l'adresse c0:00:40:00:00:00 ou ff:ff:ff:ff:ff:ff:ff:ff (source CISCO SYSTEMS, Networkers99, Introduction to IP Multicast)

Adresses Ethernet et leur représentation- Rappels

représentation dans le standard 802 suivant l'ordre des bits tels qu'ils sont émis sur le support:

ML[OUI 22b][24b alloués par constructeur]

M = 1 adresse de diffusion

L = 0 adresse universelle (OUI est attribuée par l'IEEE)

OUI = partie telle que 2 constructeurs ont des id différents

reste = identification d'un coupleur telle que 2 coupleurs d'un même constructeur n'ont pas le même identificateur

représentation dans les documents de l'Internet :

| | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---------|---|---|---|---|---|---|---|---|
| octet 1 | | | 0 | | | | L | M |
| octet 2 | | | | U | | | | |
| octet 3 | | | | | | I | | |
| octet 4 | | | | | | | | |
| octet 5 | | | | | | | | |
| octet 6 | | | | | | | | |

Entête des trames Multicast IETF (OUI 0x01-00-5E)

| 01-00-5E | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|----------|---|---|---|---|---|---|---|---|
| octet 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| octet 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| octet 3 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| octet 4 | 0 | | | | | | | |
| octet 5 | | | | | | | | |
| octet 6 | | | | | | | | |

Exemple

@IP : 224.128.64.32, 0xE0.80.40.20

Le deuxième octet : 0x80 s'écrit **1000 0000** en binaire, on ne prend que les 7 bits les plus à gauche de cet octet pour l'adresse multicast Ethernet

ce qui donne :

#Eth : 01:00:5E:00:40:20

Attention, plusieurs adresses IP peuvent donner la même adresse MAC (5bits de l'adresse IP ne sont pas utilisés). @IP : **225**.128.64.32, soit 0xE1.80.40.20 donne aussi 01:00:5E:00:40:20

Le passage d'une adresse IP multicast à une adresse MAC multicast n'est pas bijectif.

IGMP- Internet Group Management Protocol

plutôt que "Management", on aurait envie de "Membership"

IGMPv1 : RFC1112, protocole de gestion de groupe basé sur deux types de requête : Interrogation-Rapport

IGMPv2 : RFC2236 plus efficace que v1 pour la gestion d'appartenance à un groupe, incorpore un protocole de routage (DVRMP), remplace de fait IGMPv1 (mais compatibilité conservée)

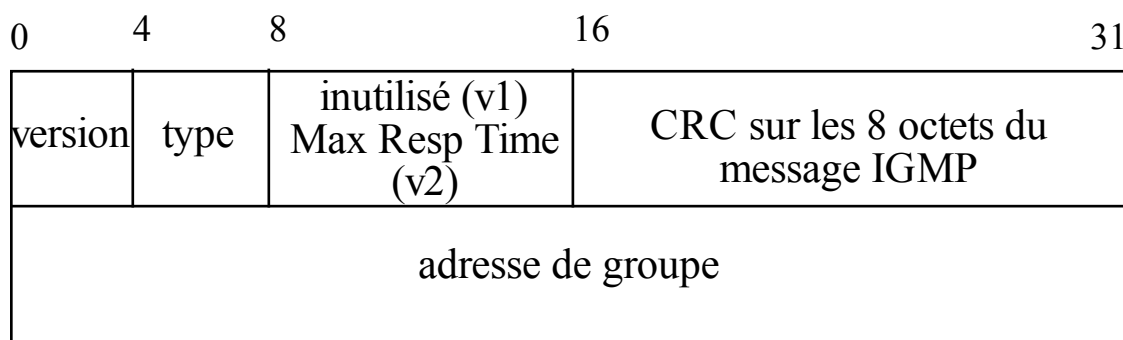
IGMPv3 : sélection de la source d'une diffusion par un récepteur (non stable), définit une API socket à ce propos.

Les opérations IGMP ne font intervenir aucun routage multicast.

Protocole entièrement repris et intégré dans IPv6.

Eléments de protocole IGMPv2 (1)

Un message IGMP est encapsulé dans un datagramme IP avec pour champ protocole de niveau supérieur valant 2



Types:

- Interrogation sur l'appartenance au groupe/**Host Membership Query** (valeur 1), message à tous les hôtes pour tous les groupes. En IGMPv2, les interrogations peuvent adresser un groupe multicast (Group-Specific Query).
- Appartenance au groupe/**Host Membership Report** (valeur 2 pour v1 ou 6 pour v2), un hôte fait une réponse en multicast avec chaque groupe auquel il appartient sur chacun des groupes
- Quitte le groupe/Leave Group propagé sur l'adresse 224.0.0.2 (extension IGMP V2), permet l'élagage d'une branche de l'arbre de routage multicast plus rapidement

Les Messages IGMP ont un TTL de 1, ils ne dépassent pas le réseau/sous réseau et donc leur routeur de rattachement.

Eléments de protocole IGMPv2 (2)

- . **Election du gestionnaire de groupes local** : Routeur multicast qui a la plus petite adresse IP d'un LAN, Election par écoute et auto-élimination, l' élu maintient la liste des groupes actifs sur le réseau (ou sous-reseau)
- . **L'entrée dans un groupe** est signalée par un multicast de type *Report* sur l'adresse de groupe. La latence d'entrée représente le temps entre le premier *Report* effectué concernant un groupe et le temps d'attachement du sous-réseau à l'arbre de routage associé au groupe défini par l'adresse de classe D. Dès que le routeur du sous-réseau a enregistré l'arrivée, le demandeur reçoit le trafic associé.
- . **Validation abonnements** : Périodiquement (60s), le gestionnaire de groupe local interroge les sites (ou plutôt le réseau/sous-réseau classe D 224.0.0.1 avec TTL de 1) par un multicast de type *Query*, les stations répondent par *Report*, une seule station du groupe suffit à maintenir l'abonnement (multicast de la réponse est donc entendu par les autres membres du LAN), si pas de réponse, le routeur ne propage plus les informations sur ce groupe.
Chaque hôte écoute les messages *Report*, si un des autres membres se signale dans son groupe, il ne génère pas de *Report*. Des *Reports* d'appartenance sont envoyés périodiquement pour chaque groupe auquel un hôte appartient.

Ajouts IGMPv3

Extension d'IGMP v2

Ajout d'un mécanisme de filtrage pour sélectionner les sources des multicast qu'un hôte veut recevoir.

Deux formats :

Tout excepté {liste d'adresses IP unicast}, mode EXCLUDE

Seulement {liste d'adresses IP unicast}, mode INCLUDE

Exemples :

- Join en IGMPv2 est équivalent à IPMulticastListen (socket, interface, multicast-address, EXCLUDE, {...})
- Leave en IGMPv2 est équivalent à IPMulticastListen (socket, interface, multicast-address, INCLUDE, {...})

Pour des informations précises voir :

<http://search.ietf.org/internet-drafts/draft-ietf-idmr-igmp-v3-09.txt>
url valide en Janvier 2002.

Concepts Généraux pour le routage multicast

Routage Multicast: (Multicast Routing) gestion du routage de datagrammes d'une source vers des destinataires

Arbre de Distribution: (Distribution Tree) Ensemble des routeurs et sous-réseaux qui permettent aux membres d'un groupe de recevoir les informations d'une source. Il y a un arbre de distribution par source.

Chemin de Montée: (Upstream) Meilleur chemin d'un routeur vers la source d'un arbre de distribution

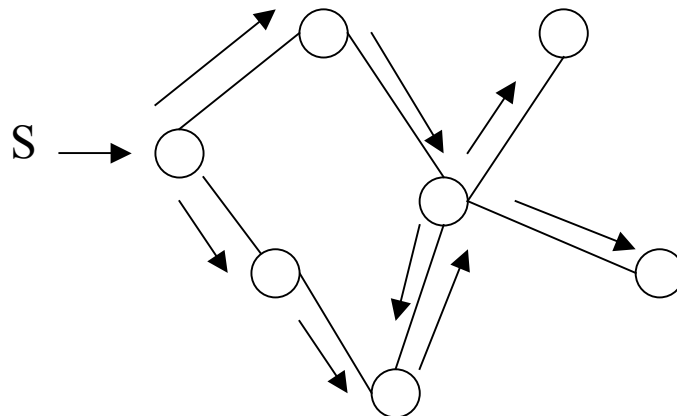
Chemin de Descente: (Downstream) Tout chemin qui n'est pas un chemin de montée, et qui mène aux destinataires accessibles depuis le routeur considéré

Propagation de Multicast: (Multicast Forwarding) Action de recopier un datagramme diffusé d'un chemin de montée (incoming interface) vers un chemin de descente (outgoing interface)

Identification d'un flot multicast :
(@IPsource,@IPgroupe)

Construction de l'arbre couvrant - Protocoles "Broadcast-and-Prune" (1)

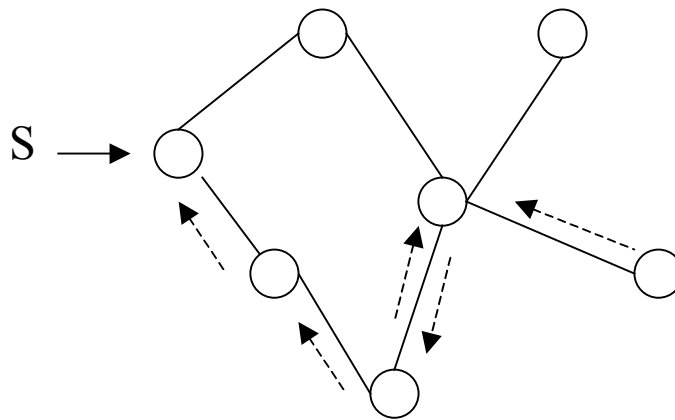
Le trafic multicast est la base de la construction de l'arbre. C'est une approche orientée donnée.



Inondation du réseau par le premier datagramme multicast

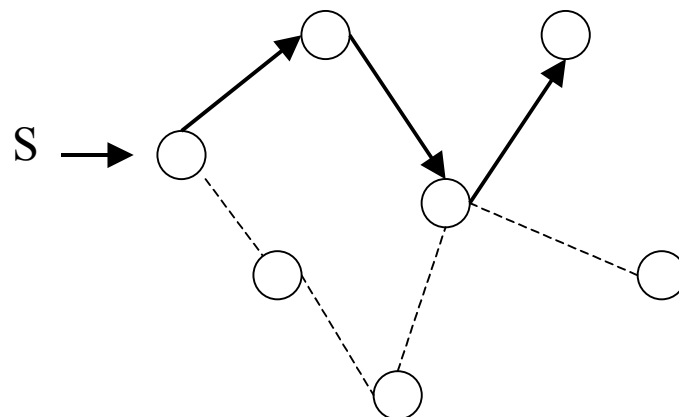
Ce broadcast a pour effet de créer une entrée ($@IPS, @IPG$) sur chaque routeur. Cette information est conservée même si le routeur n'est plus dans l'arbre de diffusion.

Construction de l'arbre couvrant - Protocoles "Broadcast-and-Prune" (2)



←----- Elagage(prune), trafic non souhaité

Aux feuilles, les bases de données IGMP sur les routeurs gestionnaires de groupes donnent les groupes actifs et inactifs. Quand un groupe est devenu inactif, le routeur envoie un message sur le chemin en montée qui demande de le faire disparaître de l'arbre de distribution.



On dit que les protocoles de type inondation-et-élagage sont des protocoles de type appartenance au groupe implicite, ils supposent que les participants d'un groupe veulent recevoir le trafic à priori.

Une branche élaguée peut-être à nouveau greffée (Grafted)
!

Construction de l'arbre couvrant - Protocoles "Broadcast-and-Prune" (3)

Il y a un arbre par trafic multicast.

Il faut périodiquement rafraîchir l'arbre.

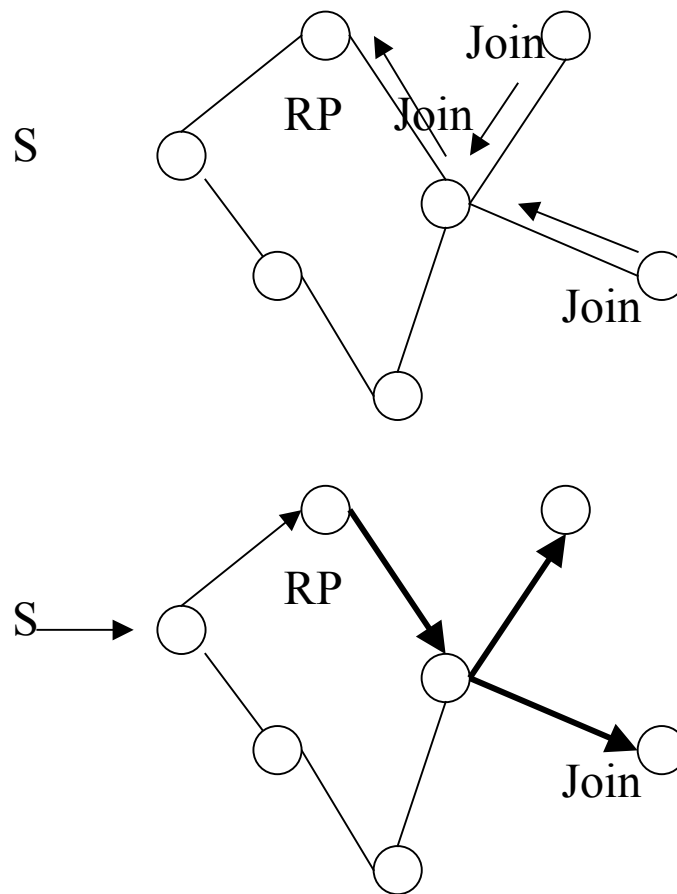
Technique qui consomme de la bande passante puisqu'il faut transmettre pour gérer l'arbre de distribution.

Elle est adaptée aux petits réseaux, LAN en particulier, voire MAN.

Autre désavantage, les membres du groupe doivent être regroupés de façon relativement compacte pour éviter des élagage de longues branches.

Construction de l'arbre couvrant - Protocoles Arbres partagés (1)

Un centre/point de RdV (RP, Rendez-Vous Point), qui n'est pas la source du flot de données, distribue le trafic vers les destinataires. L'arbre est construit par ajout de membres explicite.



Un centre gère plusieurs groupes, et les données de tous ces groupes sont distribuées suivant le même arbre indépendamment des sources. La source ne fait pas nécessairement partie du groupe.

Construction de l'arbre couvrant - Protocoles Arbres partagés (2)

Les protocoles à arbre partagé sous-entendent une appartenance au groupe explicite, ils supposent que les participants d'un groupe doivent se déclarer pour recevoir le trafic.

Cette technique nécessite qu'un destinataire indique explicitement qu'il rejoint un groupe.

Technique plus extensible que les arbres de distribution classiques depuis une source. Par contre, le centre reste un goulot d'étranglement.

Techniques de routage multicast

Techniques de Routage :

- . Core-Based Trees (CBT), visiblement ces travaux n'ont pas abouti comme tels par rapport aux protocoles concurrents, il semble que PIM-SM en ait bénéficié.
- . Distance-Vector Multicast Routing Protocol (DVRMP) utilisé pour le M-Bone
- . Multicast extension to OSPF (M-OSPF)
- . Protocol Independent Multicast - Dense Mode (PIM-DM)
- . Protocol Independent Multicast - Sparse Mode (PIM-SM)

Stratégies de routage vis-à-vis du mode d'abonnement à un groupe:

| Routage | Arbre basé sur la source | Arbre partagé |
|----------------------|--------------------------|---------------|
| Abonnement implicite | DVRMP PIM-DM | |
| Abonnement explicite | MOSPF PIM-SM | CBT PIM-SM |

Propagation des datagrammes Multicast – Technique du Reverse Path Forwarding

La recherche des chemins de sorties fonctionnent à l'envers par rapport au routage unicast.

Unicast : approche "où va le paquet", recherche du réseau vers le destinataire, on trouve dans la table l'interface pour joindre le prochain routeur

Multicast : approche "d'où vient le paquet" recherche de toutes les interfaces différentes de l'interface par où le datagramme multicast est arrivé, on éloigne le datagramme de la source !

On cherche dans la table l'entrée correspondant au chemin vers la source. Si on ne la trouve pas, le datagramme est éliminé. Si on la trouve, le datagramme est émis sur toutes les autres interfaces du routeur associées au routage multicast. Technique dite du relais sur le chemin inverse.

Différentes approches pour le RPF (1)

Inondation (Flooding) : protocole de base simple mais très coûteux.

On distingue : "parent link" un plus court chemin qui mène vers la source, et "child link" un lien downstream.

Reverse-Path Broadcasting (1978) : Si un message arrive sur un lien suppose le chemin le plus court vers la sources, alors le message est recopié vers toutes les interfaces sortantes. Ceci suppose l'existence d'un protocole de routage unicast.

Enhanced RPB : Ajout d'une technique dite "poison-reverse" qui consiste à éliminer de la liste des routeurs candidats à la propagation d'un message, les routeurs qui ont déjà un lien "parent" vers la source. En effet un routeur qui a déjà un lien parent élimine les messages qu'il a déjà propagé. D'autres critères de choix peuvent se cumuler.

Différentes approches pour le RPF (2)

Truncated RPB : Enhanced RPB+information d'appartenance à un groupe délivrée par IGMP (une branche est éliminée de la propagation s'il n'y a pas de membre du groupe sur cette partie du réseau. Cette limitation touche la propagation d'un message mais pas la construction de l'arbre de diffusion.

Reverse-Path Multicasting : Le premier message est envoyé suivant la technique TRPB vers tous les routeurs du réseau, les routeurs feuille sont assuré d'avoir le premier multicast. Un routeur feuille envoie à tous les sous-réseaux pour lesquels il existe au moins un membre associé à l'adresse multicast. S'il n'y a pas de membre, le routeur fait un élagage, le message d'élagage (prune) va vers le nœud père (one hop). Si un membre d'active, un message greffe est alors envoyé de la même façon. DVMRP et PIM-DM sont de ce type.

De façon générale, distinguer la construction de l'arbre couvrant pour la diffusion de la propagation des messages multicast.

Démonstration du fonctionnement des algorithmes de routage à partir du support de cours de CISCO téléchargeable sur le site, dont la version originale :

<ftp://ftpeng.cisco.com/ipmulticast/networkers00/index.html>

Présentation d'un sous-ensemble du Support 2214

Multicast Internet – Domaine R&D public

l'Internet Multicast est mis en oeuvre à travers le M-Bone

Il y a 3 types d'adresses :

- . Les adresses permanentes,
- . Les adresses privées,
- . Il y a des adresses pour des groupes créés de façon transitoire qui sont donc allouées dynamiquement en utilisant le protocole SAP (Session Advertising Protocol) du M-Bone.

Cartes et informations diverses :

en Europe : <http://www.dante.net/mbone/>

en France : <http://www.urec.cnrs.fr/fmbone/>

voir exposé :

<http://www.urec.fr/fmbone/Jres99/Mbone.pdf>