



# Multicast

## protocoles de routage

cours@urec.cnrs.fr



## Multicast

- **Création**
  - 1996 Bernard RAPACCHI, Bernard TUY

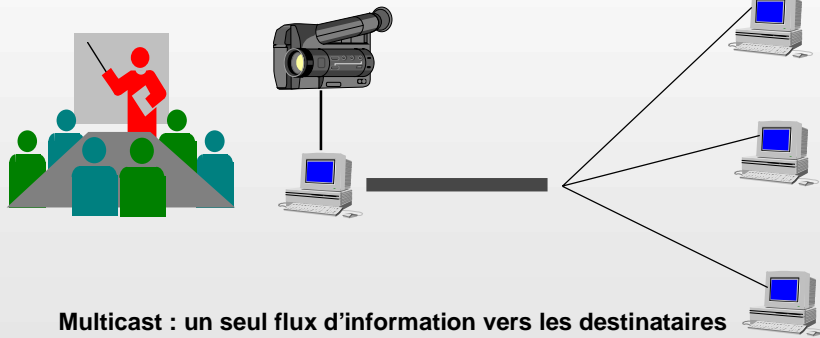
## Plan

- Définitions
- Exemples d'applications
- Notions générales
- Le MBONE
- Les Protocoles
  - IGMP
  - DVMRP
  - PIM
- Organisation du routage sur un site
  - principes et recommandations
  - exemple

## Définitions

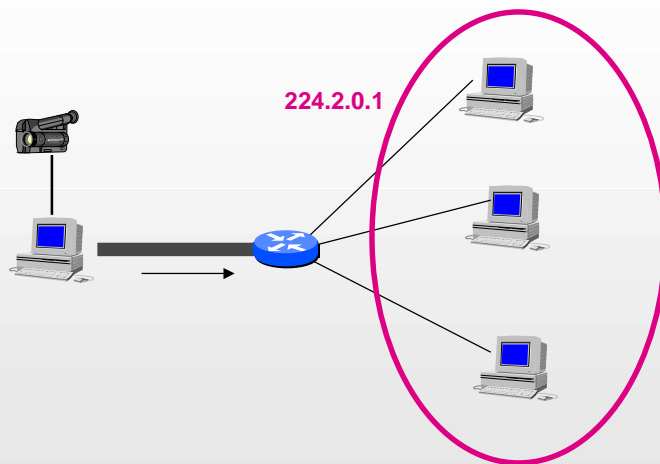
- Multicast = Multipoint
- Applications "habituelles" entre une source et **un** destinataire
  - applications **Unicast** (ou point à point)
- "Nouveau" type d'applications entre une source et **plusieurs** destinataires
  - => applications **Multicast**
- Les participants à une application multicast constituent un **groupe multicast**
- Le groupe multicast a une **adresse multicast** de classe D
  - 224.0.0.0 -> 239.255.255.255
- le groupe multicast n'est pas limité au réseau local
  - routage vers les adresses des destinataires abonnées au groupe multicast : **routage multicast**

## Exemple : téléseminaire



Multicast : un seul flux d'information vers les destinataires

## Téléseminaire (2)



— Groupe Multicast d'adresse 224.2.0.1

## Types d'applications Multicast

- **Diffusion de 1 vers plusieurs (1 vers n)**
  - pas d'interaction des destinataires*
  - Téléséminaire
  - Diffusion des tables de routage RIPv2
  - Bootstrap sur le réseau
  - découverte automatique de ressources sur le réseau
  - ...
- **Diffusion de plusieurs vers plusieurs (n vers m)**
  - chaque membre du groupe peut être source du flux multicast*
  - Téléconférence
  - Jeux interactifs (Star war ...)

## Multicast : généralités

- **Unicast** : vers un seul destinataire
- **Broadcast** : vers tous les équipements du **LAN**
- **Multicast** : vers tous ceux qui appartiennent au même groupe multicast
  - qui se sont "abonnés" à ce groupe



u r e c

## Multicast : généralités

- **meilleure utilisation de la bande passante**
  - les mêmes données ne circulent qu'une seule fois sur le même lien
- **les émetteurs (sources) et les receveurs (membres) sont distincts**
- **les hôtes disent aux routeurs de quels groupes ils sont membres**
  - et ne reçoivent que les datagrammes de ces groupes
  - ils ne disent rien sur les groupes multicast auxquels ils envoient des informations
- **les routeurs doivent écouter toutes les adresses multicast**
  - pour être capables de transmettre les datagrammes multicast
- **les routeurs utilisent des protocoles de routage multicast pour administrer les groupes multicast**



u r e c

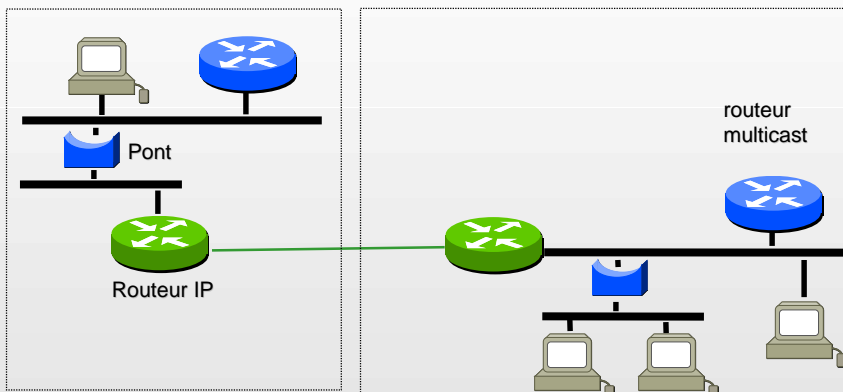
## le MBone

- **MBone = Multicast backbone**
  - Virtual Internet backbone for Multicast IP
- **Réseau virtuel interconnectant des équipements multicast sur Internet**
  - les sites multicast sont reliés entre eux par des "tunnels"
  - au-dessus de la topologie unicast (*overlay network*)
- **Ensemble d'outils :**
  - **SDR :**
    - pour annoncer la diffusion de sessions multimédia
    - pour permettre aux utilisateurs de rejoindre les groupes multicast
  - pour suivre ou diffuser ces programmes :
    - *vat, rat* : audio
    - *vic* : vidéo

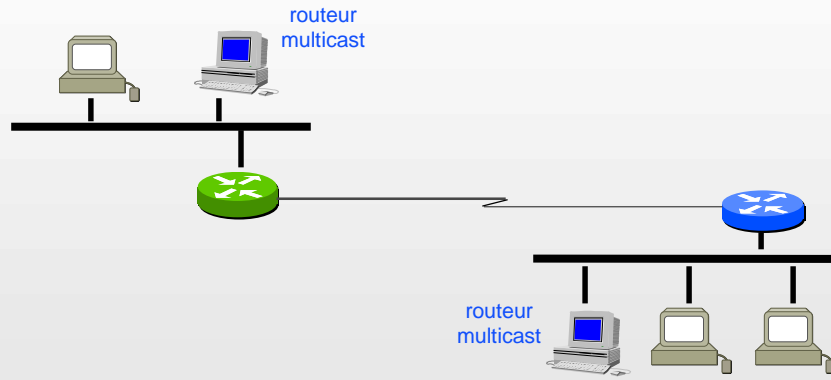
## le Mbone (2)

- **Un projet de coopération *expérimental***
  - 1992 : première diffusion réseau d'une réunion de l'IETF (son)
  - 1993 : première diffusion vidéo
- **Topologie :**
  - **interconnexion** des ilots multicast
    - par des « mrouteurs »
    - à travers des tunnels
  - en étoile au niveau du réseau de site

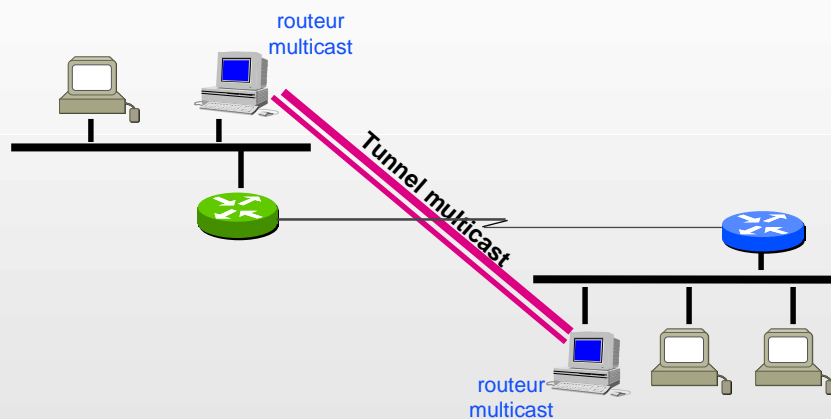
## La topologie du Mbone



## La topologie du MBone (2)



## La topologie du MBone (3)





## Tunnels IP

- Structure logique faisant abstraction de la topologie physique sous-jacente du réseau
- Dans RFC 1075, on utilise “*Loose Source Routing*”
- En réalité, on encapsule les paquets multicast dans des paquets unicast (protocole = 4, tunnel DVMRP)
- Association d'un **métrique** et d'un **seuil** (*threshold*) à chaque tunnel (en plus d'une adresse IP à chaque extrémité ...)
  - **métrique** permet de privilégier une interface parmi *n*
  - le **seuil** définit la portée du datagramme multicast
    - < 16 *site local*
    - < 32 *regional*
    - < 48 *national*
    - < 64 *transcontinental*



## Algorithme de transmission d'un datagramme multicast

- Si **TTL**(de l'entête IP) > **Seuil** (fixé sur l'interface)
- alors :
  - $TTL = TTL - 1$
  - on transmet le paquet
- On détruit le paquet si :
  - Si  $TTL < 2$
  - pas de route
  - Paquet reçu sur une interface non-RPF
  - Destination = 224.0.0.{1,2}





## Notions d'adressage multicast

- Adresses de classe A, B, C ...
  - et D : **224.0.0.0** à **239.255.255.255**
  
  - **réserveation de plages d'adresses spécifiques :**  
*cf le RFC 1700 ou <http://www.iana.org/iana/assignments.html>*
- Exemples :**
- les adresses **224.0.0.0 - 224.0.0.255**
  - **239.\*.\*** : "*administratively scoped addresses*"
    - adresses reservees pour des usages prives
- 
- **toutes les autres adresses de 224.0.1.0 à 239.255.255.255**
    - . attribuees de facon permanente a differentes applications
    - . reservees pour une allocation dynamique via SDR ou autres methodes



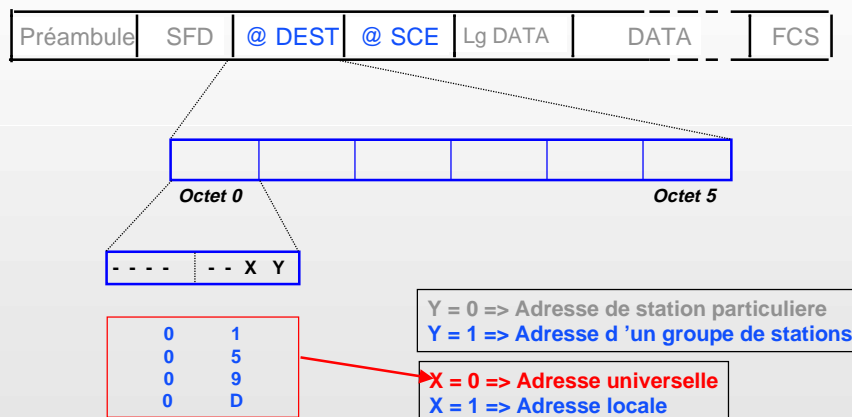
## Adresses multicast réservées : exemple

- **224.0.0.1** : tous les hosts Xcasts du LAN
- **224.0.0.2** : tous les routeurs Xcasts du LAN
- **224.0.0.4** : tous les routeurs DVMRP du LAN
- **224.0.0.9** : tous les routeurs RIPv2
- **224.0.0.13** : tous les routeurs PIM du LAN

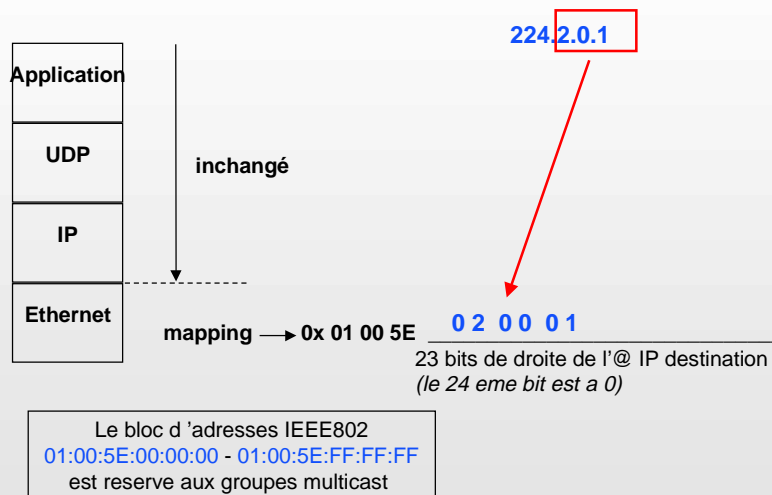
## Envoi de paquets

- Une **adresse multicast** ne peut être **que destinataire**
  - c'est l'adresse d'un groupe de machines abonnées à une session multicast
- les **sources (emetteurs)** ont toujours une **adresse unicast**
- le niveau **Liaison de données** n'utilise pas **ARP** :
  - **mécanisme de correspondance** (pour les @ IEEE-802)
    - @IP Xcast -> @Ethernet Xcast
- **Etre membre d'un groupe est indépendant d'envoyer à ce groupe**
  - une source n'est pas obligatoirement membre du groupe auquel elles envoient un flux multicast

## Les adresses IEEE 802.3 et le multicast



## Les adresses IEEE 802.3 et le multicast (2)



## Réception de paquets

- **Par défaut, le coupleur Ethernet d'une station écoute**
  - son adresse Ethernet (fixée en PROM)
  - et l'adresse de broadcast (FF...FF)
- **Les autres adresses Ethernet doivent être explicitement programmées dans le driver du coupleur**
- **Pour le multicast, il faut écouter au minimum :**
  - équivalent Ethernet de 224.0.0.1 (tous les hôtes Xcast du LAN)
  - équivalent Ethernet du répertoire des sessions Mbone  
*annonçant la liste des groupes multicast actifs*

## Internet Group Management Protocol ( IGMP )

RFC 1112 (IGMP version 1)

RFC 2236 (IGMP version 2)

### IGMP : généralités

- **Protocole d'interaction entre**
  - le(s) routeur(s) Xcast du LAN
  - et les hôtes Xcast du LAN
  
- **Permet à un hôte de s'abonner (désabonner) à un groupe**
- **et dire au routeur :**
  - “envoyez-moi une copie des paquets recus par cette adresse de groupe”
  
- **deux versions existent, IGMPv1 et v2**
- **IGMP version 3 en cours d'élaboration (IETF/ IDMR)**

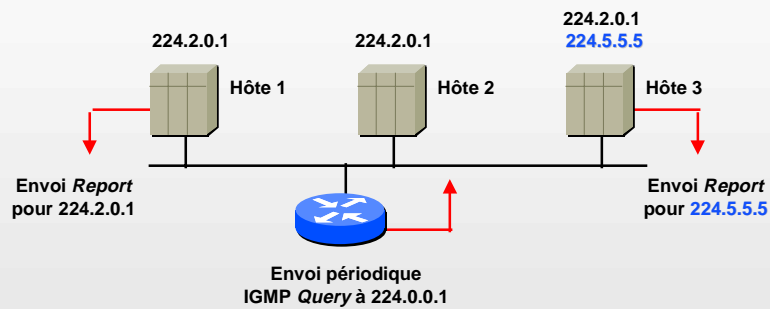
## IGMP: un seul routeur sur le LAN

- **le routeur envoie** toutes les 60 (120) secondes
  - une sollicitation aveugle à l'@ 224.0.0.1 (*query*) avec un TTL = 1
  - "à quel(s) groupe(s) voulez vous vous abonner ?"
  - et attend les réponses
- **le(s) hôte(s) renvoie(nt) un "IGMP report"**
  - qui indique l'adresse du ou des groupes qui l'(es) intéressent
- **si le routeur ne reçoit aucune réponse pour un groupe donné**
  - il **arrête la réémission** des paquets multicast de ce groupe
  - le groupe est réputé sans abonné local

## IGMP: un seul routeur (2)

- **quand l' hôte reçoit l'invitation query**
  - il fixe un délai aléatoire avant de répondre  
*pour éviter que toutes les réponses arrivent au même moment*
  - quand un hôte a répondu, les autres n'ont plus besoin de répondre  
*=> une réponse par groupe multicast et par LAN*
- **le routeur arme une temporisation sur les abonnements aux groupes multicast avant de solliciter à nouveau tous les hôtes**
  - *sollicitation periodique*

## IGMP : s'abonner à un groupe



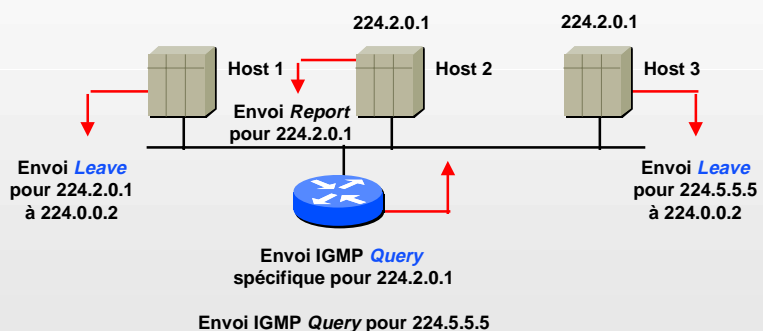
## IGMP : plusieurs routeurs sur le LAN

- **Un routeur est élu entre tous les routeurs**
  - c'est le **Dominant Router (DR)** ou **Designated Router**
  - il est seul à émettre les **IGMP Queries**
  - en v1, le mécanisme d'élection est fonction du routage Xcast et n'appartient pas à IGMP
  - en version 2, le **DR** est le routeur dont l'**@IP est la plus petite**
- **le DR n'est pas forcément le routeur qui transmet les paquets Xcast**

## IGMP : version 2

- Election du *DR*
  - @IP la plus petite
- *timers* programmables
- nouveaux type de paquets envoyés par l' hôte :
  - de désabonnement : *leave*
  - au reçu d'un *leave*, le routeur envoie un *query* **directionnel** aux hôtes qui ont été abonnés à ce groupe  
=> réduction du temps de latence pour arrêter la diffusion d'un groupe qui n'a plus d'abonné
- IGMP v2 doit obligatoirement supporter la version 1

## IGMP : quitter un groupe



## Les Protocoles de routage Multicast

- **IGMP** permet la distribution des datagrammes Xcast sur le LAN
- **Protocole de routage** pour acheminer ces paquets hors du LAN
- **Protocoles de routage multicast -distincts des protocoles de routage unicast :**
  - tous les routeurs ne savent pas traiter les datagrammes multicast
  - la topologie multicast est -en general- distincte de la topologie Unicast (basee sur les tunnels inter-routeurs multicast)
  - ils construisent des arbres de diffusion du trafic multicast
    - où l'émetteur est la racine de l'arbre de diffusion*
  - l'arbre minimal de diffusion ...
    - toutes les branches sont utiles (id. ont au moins un abonne)*
  - ... est en constante evolution
    - ajout (suppression) d'une feuille/branche*

## Les Protocoles de routage Multicast (2)

- **On distingue deux familles de protocoles en fonction du mode de diffusion des paquets multicast utilisé :**
  - Mode dense : **inondation et elagage**
    - **DVMRP, PIM-DM et MOSPF**
    - *suppose que les abonnés aux groupes Xcast sont nombreux*
  - Mode épars (clairseme) : **greffe et elagage**
    - **PIM-SM et CBT**
    - *faible population abonnée*



## Distance Vector Multicast Routing Protocol ( DVMRP )

RFC 1075

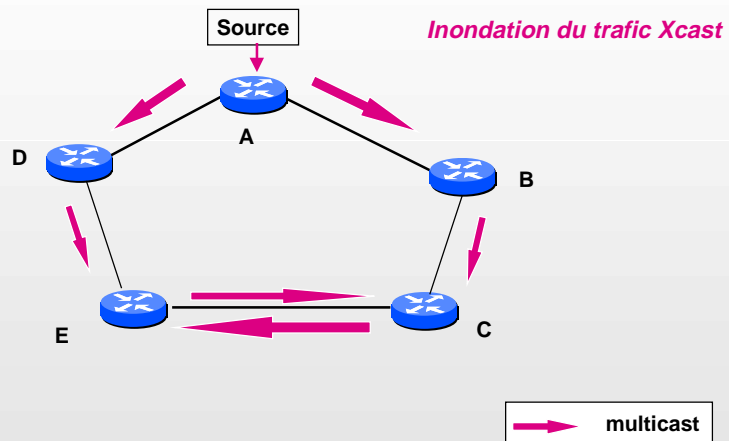
### DVMRP : généralités

- **Version 3 du protocole**
  - draft-ietf-idmr-dvmrp-v3-04.txt
- **“mrouterd” sous Unix**
  - pas natif dans tous les OS
  - **s ’assurer que la version du mrouterd est au moins de 3.8**
- Agit en **mode dense** : **flooding** + **pruning**
  - on **inonde** (*flooding*) tout l’arbre multicast
  - ceux qui ne sont pas intéressés le disent
  - ils sont **élagués** de l’arbre (*pruning*)
- **Pour éviter les boucles => algorithme RPF**
  - Reverse Path Forwarding

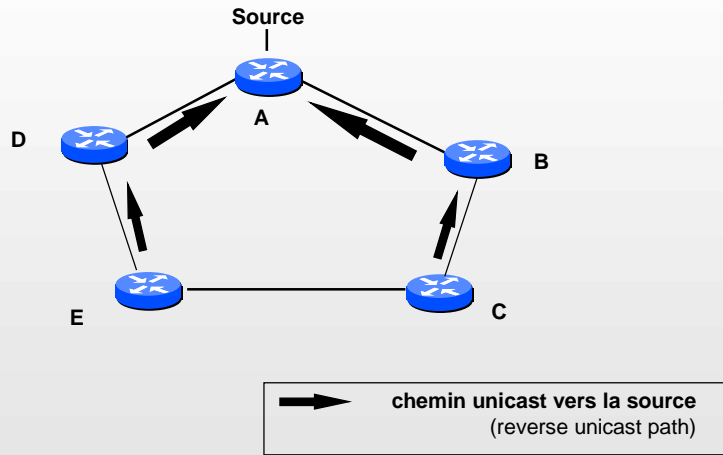
## Reverse Path Forwarding (RPF)

- un routeur transmet un paquet multicast
  - si le datagramme est reçu sur l'interface utilisée pour envoyer un paquet unicast vers la source (*reverse*)
- Test RPF :
  - Oui : paquet retransmis, on inonde
  - Non : paquet est mis à la poubelle
- un paquet est retransmis vers toutes les interfaces du routeur SAUF l'interface RPF d'entrée qui sont dans la liste des interfaces Xcast
  - Toutes les interfaces non élaguées (not pruned)

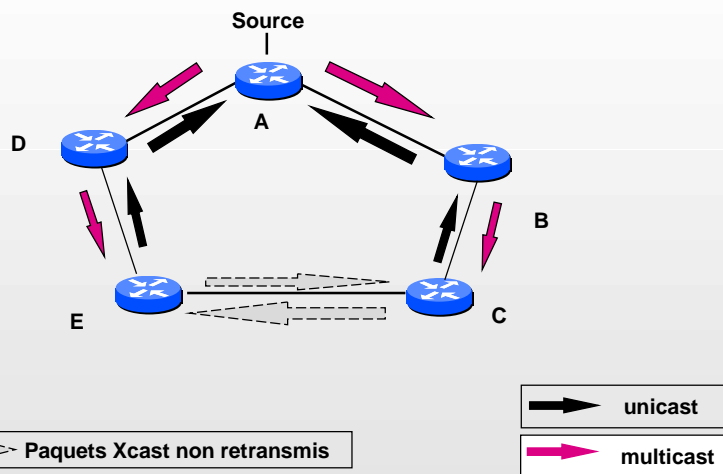
## Reverse Path Forwarding (2)



### Reverse Path Forwarding (3)



### Reverse Path Forwarding (4)



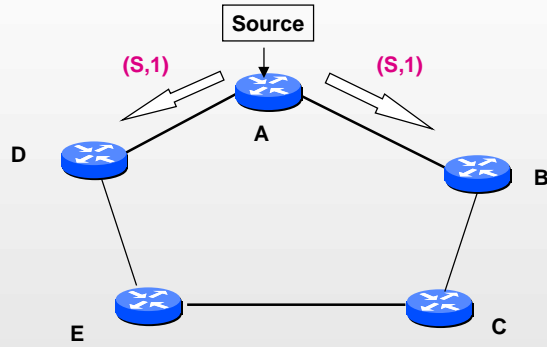
## Routage DVMRP

- **DVMRP utilise son propre routage unicast**
  - (variante de RIP)
  - pour déterminer son critère RPF et
  - sa décision de transmettre un datagramme Xcast
- Le routage Unicast est nécessaire pour localiser les Sources Xcast
- **les paramètres du protocole**
  - le **nombre de sauts (hops)**, les **métriques** et les **seuils (Threshold)**  
le seuil indique si un datagramme multicast peut être réémis.
- **obligation d'utiliser des tunnels**

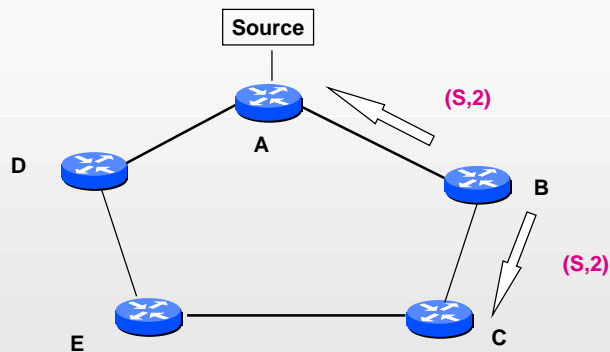
## Routage DVMRP

- **échange de tables de routage entre routeurs DVMRP**
  - Destination / Masque / Métrique
- Les **destinations** sont les **@ des sources Xcast** (émetteurs des datagrammes multicast)
- La **métrique** (distance) est le nombre de **routeurs multicast** à franchir pour atteindre la source
- L'optique est de toujours construire un **arbre minimal** à partir de la **source**
- Le même protocole de routage est utilisé dans tout l'arbre de diffusion (*tous les routeurs ont la même vue de la topologie multicast*)

### Echange des tables de routage (théorie)



### Echange des tables de routage (théorie)



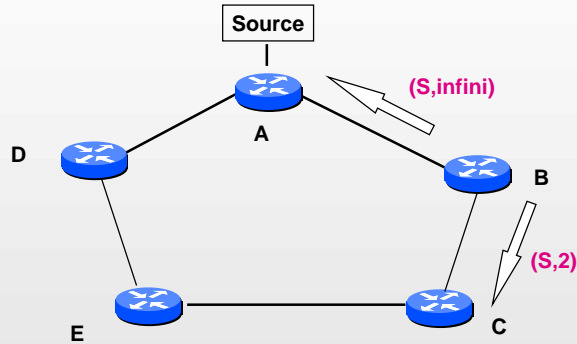
## Poison Reverse

- **Le routeur B va décider**
  - que le routeur **A** voisin est en “amont” vers la source **S**
  - il envoie à **A** une information de routage vers **S** dont la métrique est dite *empoisonnée*
  
- **Conséquence :**
  - **B** attend le flux multicast de **A** pour la source **S**
  - **A** ne doit pas compter sur **B** pour ce même flux

## Métrique empoisonnée

- **Le RFC 1112 prévoit d'envoyer :**
  - @IP Source, m = infini (16), + un flag à 1
  
- **Dans mrouterd :**
  - Source, m = vraie métrique vers S + infini (32)

## Poison Reverse



## DVMRP: échange des routes

- **Les échanges d'informations de routage utilise IGMP type 3**
  - **Envoi à 224.0.0.4** (tous les routeurs DVMRP)
  - d'un *route report* toutes les 60 secondes
  - d'un *probe message* toutes les 10 secondes pour découvrir les voisins dvmrp
- **Utilisation de sous-types :**
  - *Response* : envoie les routes vers les destinations (Sources)
  - *Request* : demande les routes vers les destinations
  - *Prune* : rapport d'aucun membre
  - *Graft* : greffe d'une nouvelle branche sur l'arbre multicast

## Elagage et greffe sur l'arbre de diffusion

- **Elagage :**
  - le routeur multicast n'a plus de récepteurs locaux ni de routeurs multicast en aval
  - il envoie un *prune packet* à la (les) sources émettrices pour ce groupe multicast via les interfaces RPF
  - il arme un timeout = 2h
- **Greffe :**
  - A la réception d'une nouvelle demande d'un récepteur local ou d'un routeur en aval pour un nouveau groupe (ou un groupe précédemment élagué)
  - il envoie un *graft packet* vers le routeur RPF pour éviter d'attendre l'expiration du timeout d'élagage

## Protocol Independent Multicast

"Internet Draft"  
<http://netweb.usc.edu/pim/>



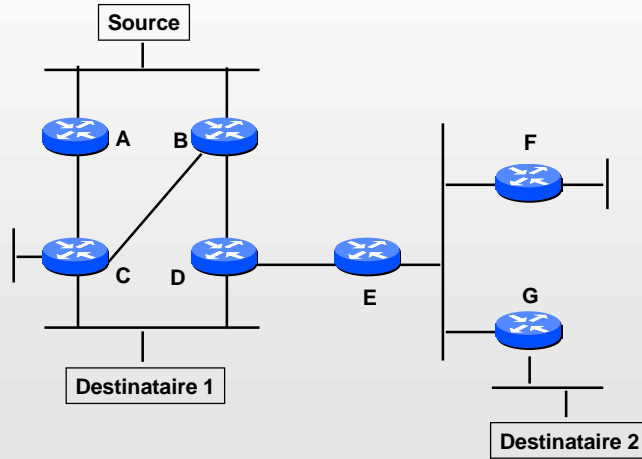
## PIM : généralités

- **Indépendant du protocole de routage unicast**
- **DVMRP**
  - prend les décisions de RPF
  - a son propre protocole de routage
- **PIM repose sur le protocole de routage unicast sous-jacent**
  - pour les décisions RPF
  - et les *poison reverse routes*
- **PIM peut fonctionner selon deux modes :**
  - *dense mode*
  - *sparse mode*
- **PIM utilise des arbres partagés pour réduire la gestion des groupes à faible trafic**
- **et des arbres basés sur la source pour réduire les délais de diffusion dans les groupes à forte population**

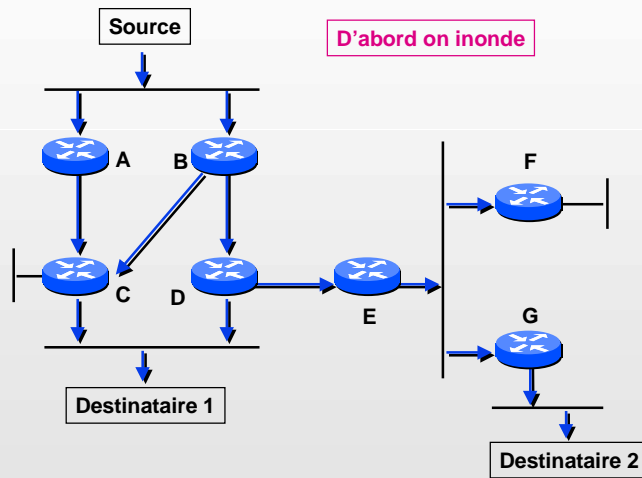
## PIM : Dense Mode

- **Ressemble à DVMRP**
  - sauf pour le routage
- **mécanismes de flooding et pruning et de greffe,**
  - Pruning vers les voisins non RPF
- **Arbres construits par rapport aux sources émettrices en utilisant l'algorithme RPF**
- **Utilisation de déclaration (*assert*) pour élire un transmetteur sur un LAN à plusieurs routeurs**
- **Faible *overhead* pour les groupes denses**

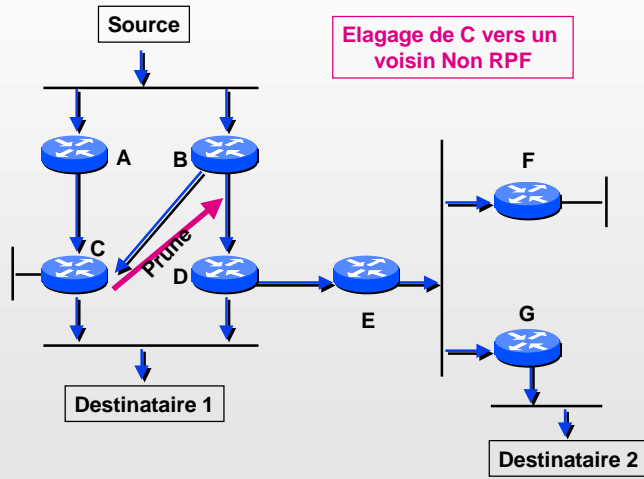
### PIM, Dense-Mode : exemple



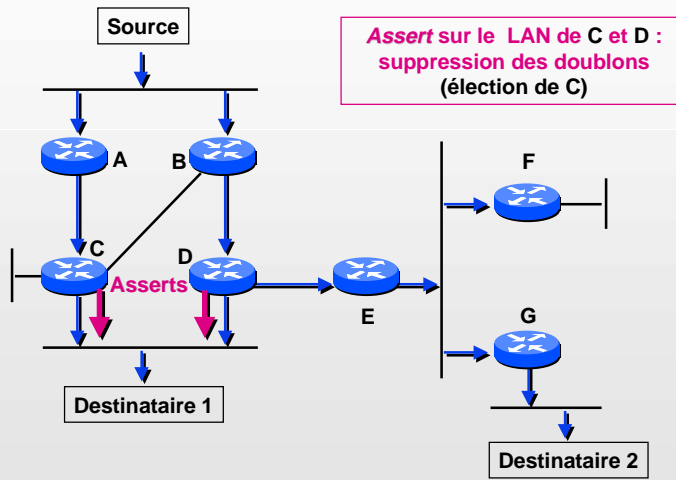
### PIM, Dense-Mode : exemple



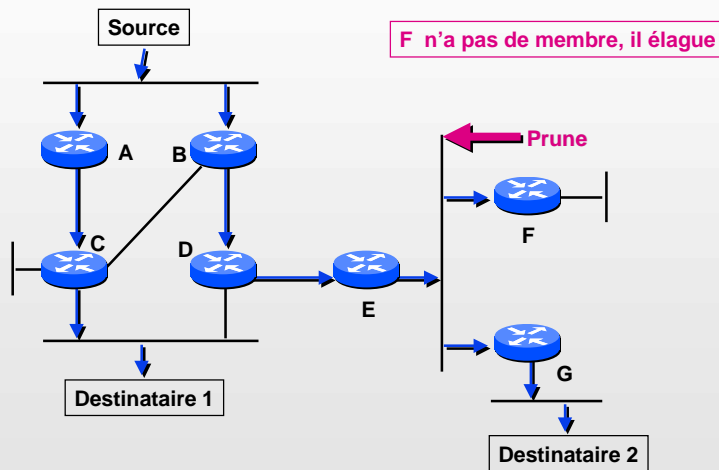
### PIM, Dense-Mode : exemple



### PIM, Dense-Mode : exemple



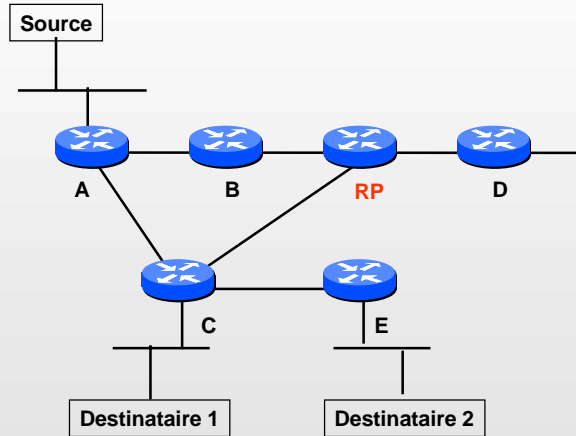
## PIM, Dense-Mode : exemple



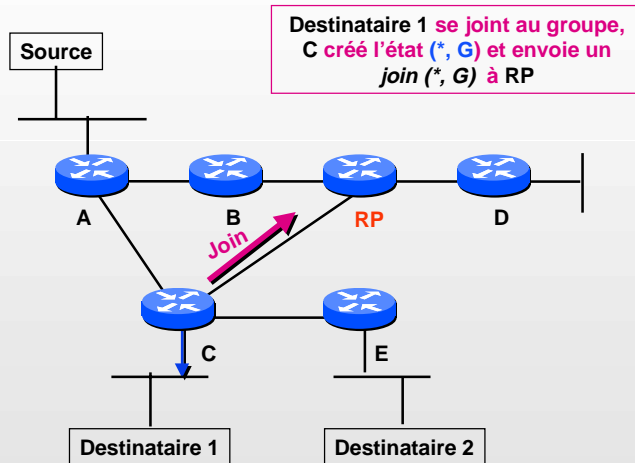
## PIM : Sparse mode

- **Mode d'abonnement explicite (Join) :**
  - La **source** s'enregistre auprès d'un **Point de Rendez-vous (RP)**
  - Le **RP** est la racine de l'arbre de diffusion multicast
  - c'est une adresse connue de tous
  - Pour s'abonner le destinataire envoie un **Join** au **RP**
  - Il peut y avoir **plusieurs RP** pour différents groupes
  - **Pas d'inondation**
- **Le flux multicast parcourt un arbre partagé**
  - les routeurs feuilles peuvent se joindre à l'arbre
  - **les paquets ne vont que là où c'est utile**
  - on utilise le **RP** pour tester les interfaces RPF de l'arbre partagé
    - **état (\*,G)**
  - on utilise la **Source** pour tester les interfaces RPF de l'arbre basé sur la Source
    - **état (S,G)**
  - on préfère les états **(S,G)** aux états **(\*,G)**

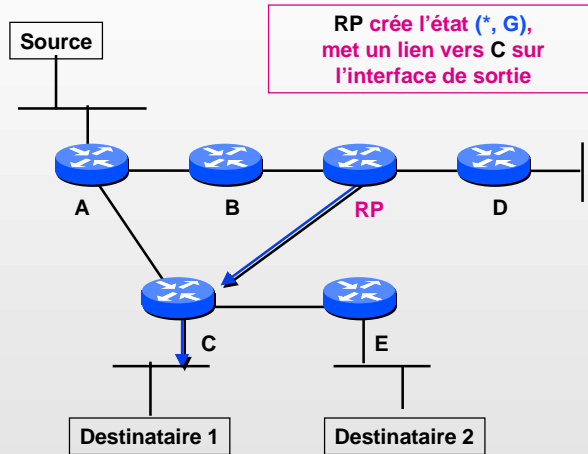
### PIM, Sparse-Mode : exemple



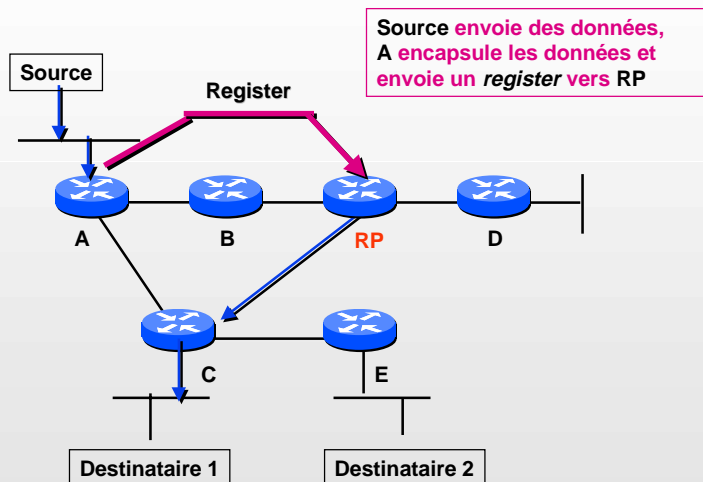
### PIM, Sparse-Mode : exemple



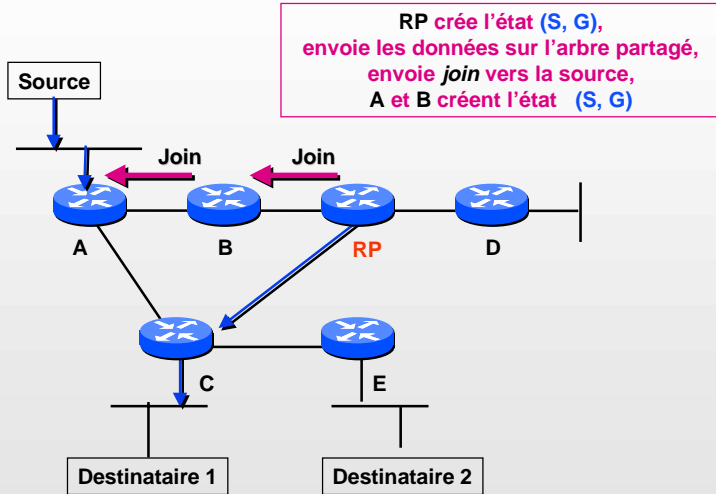
## PIM, Sparse-Mode : exemple



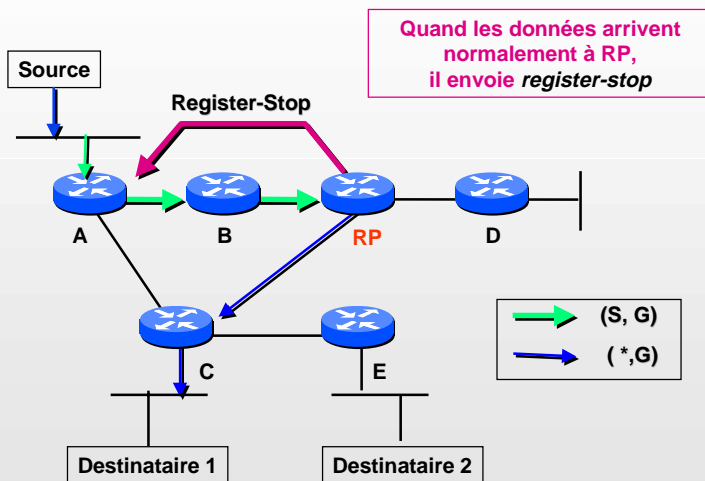
## PIM, Sparse-Mode : exemple



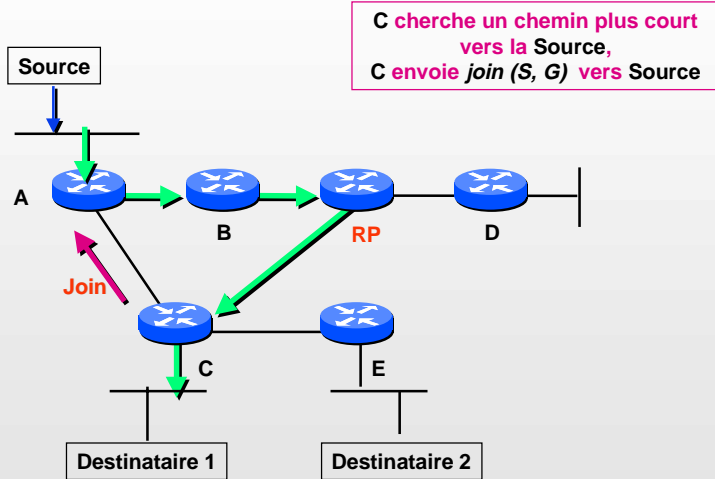
## PIM, Sparse-Mode : exemple



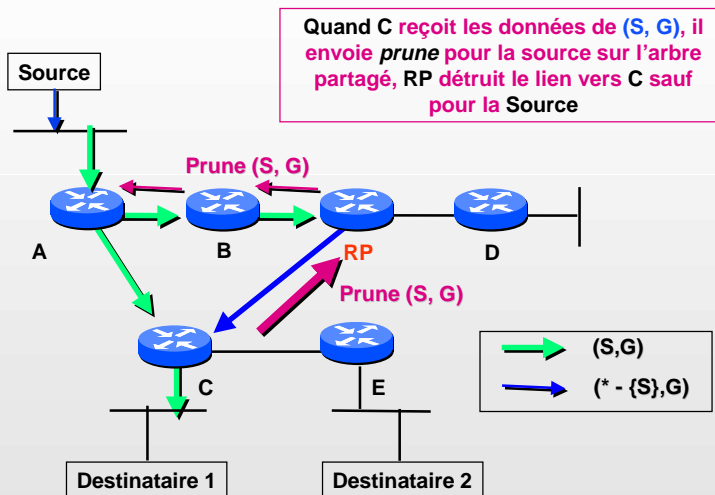
## PIM, Sparse-Mode : exemple



## PIM, Sparse-Mode : exemple

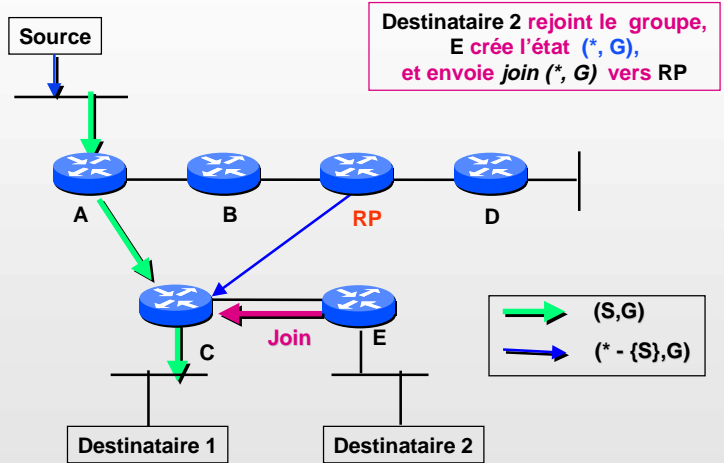


## PIM, Sparse-Mode : exemple

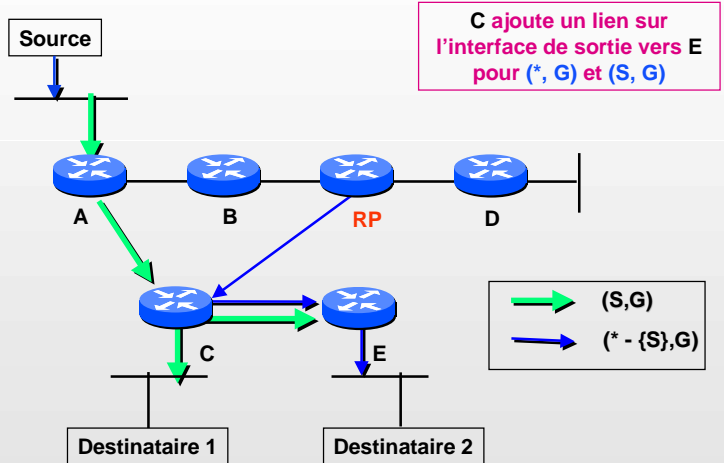




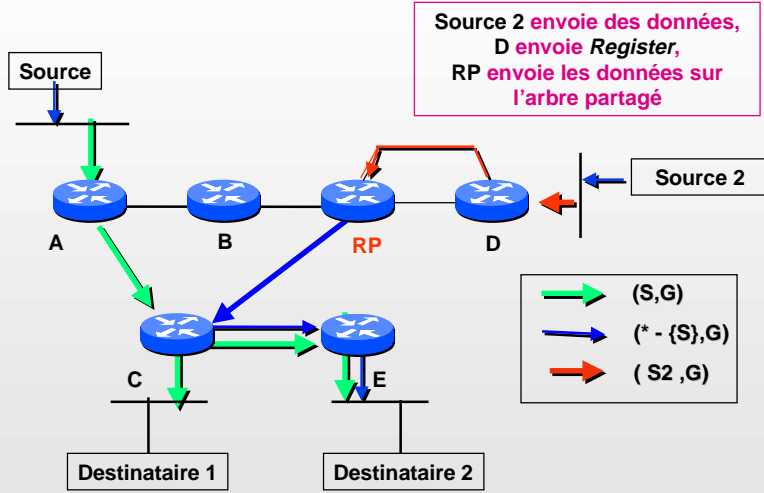
### PIM, Sparse-Mode : exemple



### PIM, Sparse-Mode : exemple



## PIM, Sparse-Mode : exemple



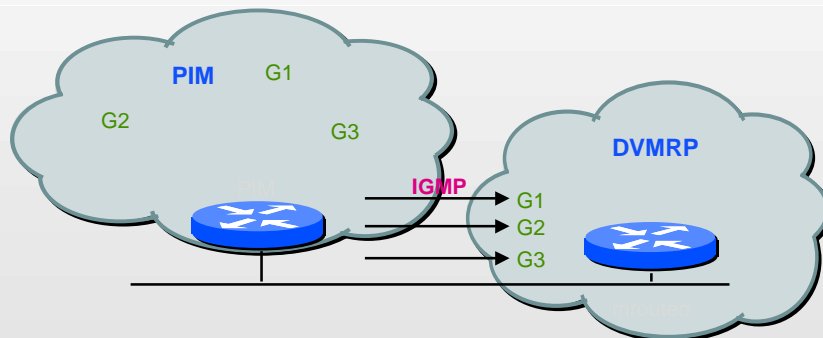
## Interactions PIM - DVMRP

## Principes

- Les sources du nuage PIM doivent être annoncées dans DVMRP
- Les membres dans le nuage PIM doivent être connus dans DVMRP
- Le routeur de frontière PIM-DVMRP
  - doit savoir parler le protocole de routage unicast DVMRP
  - doit convertir les métriques propres à chaque protocole
  - il a une vision particulière pour chacun des protocoles
  - utilise IGMP pour le rapport des groupes

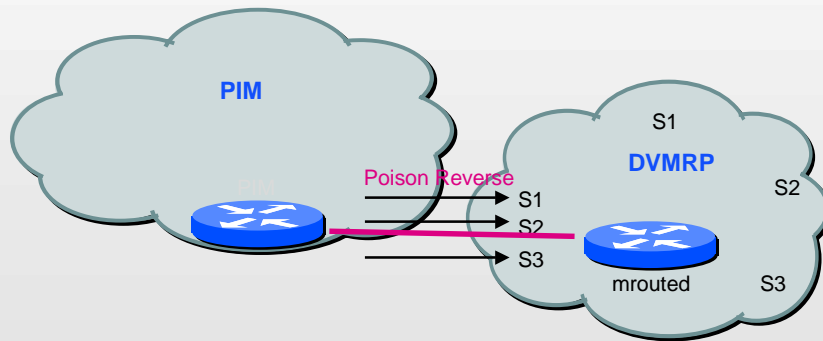
## Groupes : bordure native LAN

routeur PIM envoie *IGMP reports* pour tous les groupes qui ont des membres dans le nuage PIM



## Groupes : bordure par un tunnel

routeur PIM *Poison Reverse* toutes les sources connues du DVMRP puis élague les groupes sans intérêt



## Recevoir les rapports DVMRP

- **Sur un tunnel**
  - “*cache*” les routes apprises
  - *Poison Reverse* vers l'émetteur
- **Interface native**
  - si pas besoin de transmettre à d'autres DVMRP les routes apprises peuvent être abandonnées mais les routes unicast vers les sources doivent être même interface
  - sinon on *cache* les routes apprises

## Ce qu'on envoie à DVMRP

- Les interfaces directement connectées avec PIM
- Les routes connues d'autres nuages DVMRP
- Les routes des sources dans le nuage PIM
- si nécessaire *Poison-Reverse* les routes des sources dont DVMRP est en amont

## Annonces aux voisins DVMRP

- **Calcul des métriques pour :**
  - les routes directement connectées = 1
  - Apprises de DVMRP = métrique reçue + 1
  - Apprises d'un routeur PIM qui fait du routage unicast  
DVMRP = métrique reçue + 1
  - Apprise d'une table de routage unicast existante = 1
  - Poison-Reverse = métrique connue + infini (32)



## Modification des Métriques (option)

- Mettre manuellement les valeurs
- Ajouter manuellement un *offset*
- Appliquer aux routes de toutes origines
- Appliquer des *access-lists* (par défaut , tout est filtré)
- Possibilité de déclarer des *route maps*



## Configuration d'un routeur Cisco : commandes globales

- **[no] ip multicast-routing**
  - pour activer le routage Xcast
- **[no] ip dvmrp route-limit 8000**
  - limiter la taille de la table de routage Xcast
- **[no] ip dvmrp routehog-notification <route-count>**
- **[no] ip mroute <src> <mask> ... <@RPF> | @ interface**
  - pour déclarer une route Xcast statique



## Configuration d'un routeur : commandes d'interface (IGMP)

- **ip igmp access-group <access-list>**
  - configure les groupes autorisés sur une interface donnée
  
- **ip igmp join-group <group-address>**
  - déboguer le trafic qui arrive pour ce groupe
  
- **ip igmp query-interval ...**
  - fixe l'intervalle des requêtes envoyées aux hôtes du LAN

---
  
- **[no] ip sdr listen**
  - connaître les sessions multicast qui sont annoncées  
on les visualise avec **show ip sdr**



## Configuration d'un routeur : commandes d'interface (PIM)

- **[no] ip pim [dense-mode | sparse-mode ]**
  - *dense mode* est le défaut
  
- **ip pim query interval < time in seconds >**
  
- **[no] ip pim rp-address <ip-address>**
  - pour définir le point de Rendez-Vous
  
- **ip dvmrp unicast-routing**
  - (!) utiliser le routage unicast DVMRP sur une interface PIM
  
- **ip multicast ttl-threshold <valeur>**
  - fixer un seuil sur une interface

## Configuration d'un routeur : commandes d'interface (DVMRP)

- **[no] ip dvmrp metric-offset in | out <offset>**
  - modifier la metrique de 1 par défaut
- **[no] ip dvmrp metric <nombre> [list <access-list>]**
  - *cette commande pose problème dans certaines versions d 'IOS*
- **[no] ip dvmrp default-information only**
  - ne transmettre que la route par défaut
- **ip multicast rate-limit in | out <valeur en kbps>**
- **tunnel mode dvmrp**
  - établir un tunnel entre un routeur et une station par exemple
  - les tunnels entre routeurs (Cisco) sont en mode GRE

## Organisation du routage multicast

- **Principes :**
  - sur un campus, l'équipe réseau doit**
    - organiser la redistribution des flux multicast !
      - *Mieux vaut éviter les tunnels multiples ...*
    - Participer au Fmbone
      - *se connecter à la distribution régionale du Fmbone*
    - préférer **PIM** aux tunnels quand cela est possible
    - informer / **former** les utilisateurs potentiels
      - *configuration réseau et des applications (réglages des débits)*



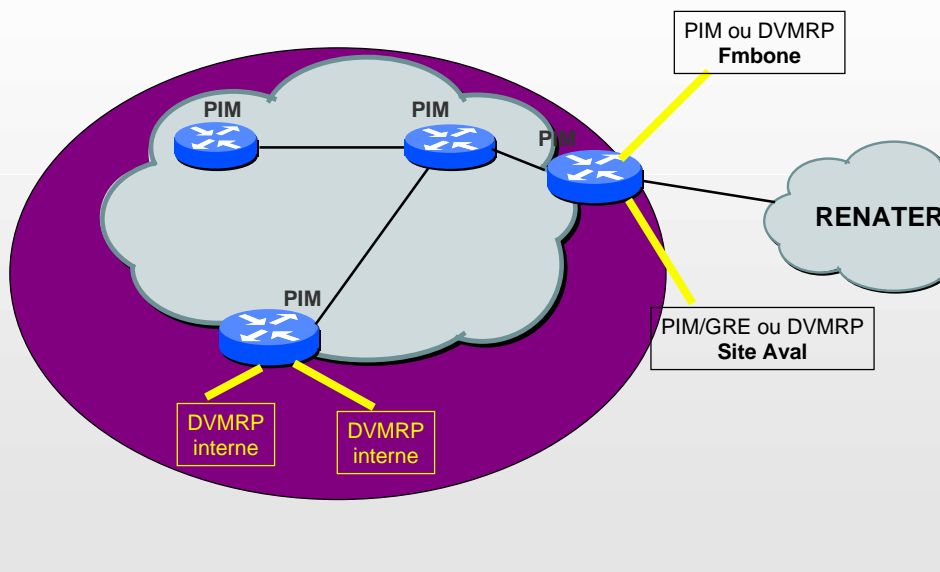
## Organisation du routage multicast (2)

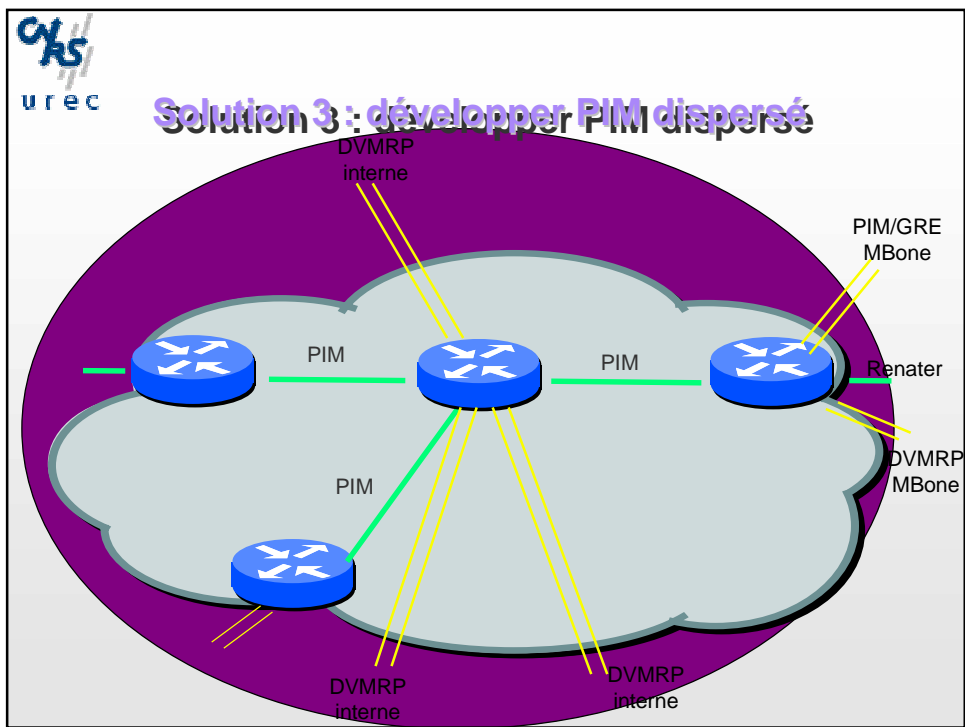
○ **Principes :**


**dans un laboratoire :**

- mettre en place **un seul routeur multicast**
- **PIM** si possible (type du routeur, niveau d'IOS, ...)
- quand le besoin existe !

## Exemple d'organisation d'un campus






**Bibliographie**

- **C. Huitema**
  - Le Routage dans l'Internet, Eyrolles, 1995
- **draft-ietf-mboned-intro-multicast-03.txt**
- **draft-ietf-mboned-mdh-00.txt**
- **<http://www.urec.cnrs.fr/Xcast/>**
- **<http://www.services.cnrs.fr/multicast>**