



Le protocole de routage

Multicast PIM

Le protocole PIM

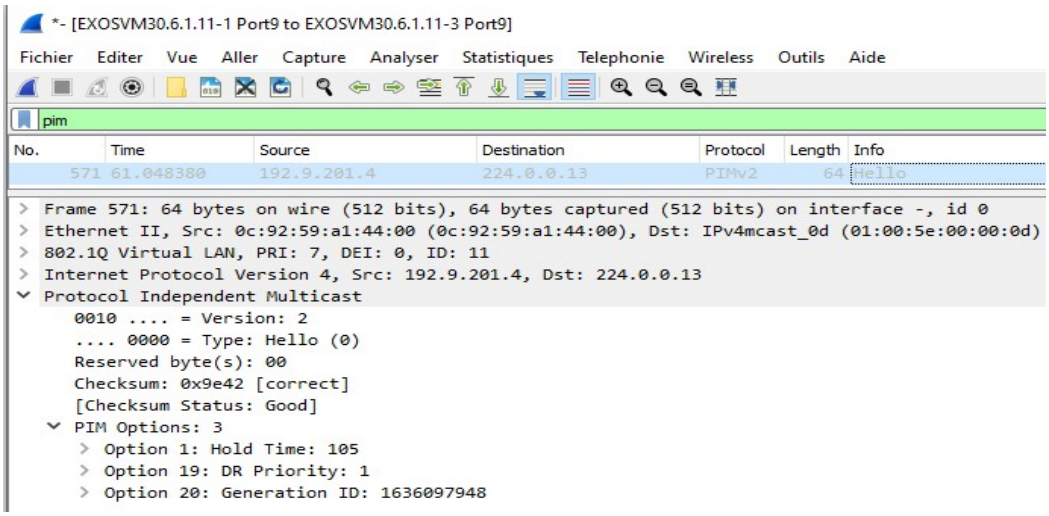
PIM est un protocole de routage multicast pouvant fonctionner selon deux modes:

- ✓ Dense Mode (PIM-DM)
- ✓ Sparse Mode (PIM-SM)

1) **PIM Dense Mode** (**PIM**-DM)

PIM Dense Mode (**PIM**-DM) utilise le routage multicast **dense**. Il diffuse par défaut le paquet multicast à travers les routeurs jusqu'à ce que des messages de pruning indiquent qu'aucun abonné du groupe n'est intéressé sur les routeurs connectés les routeurs en aval.

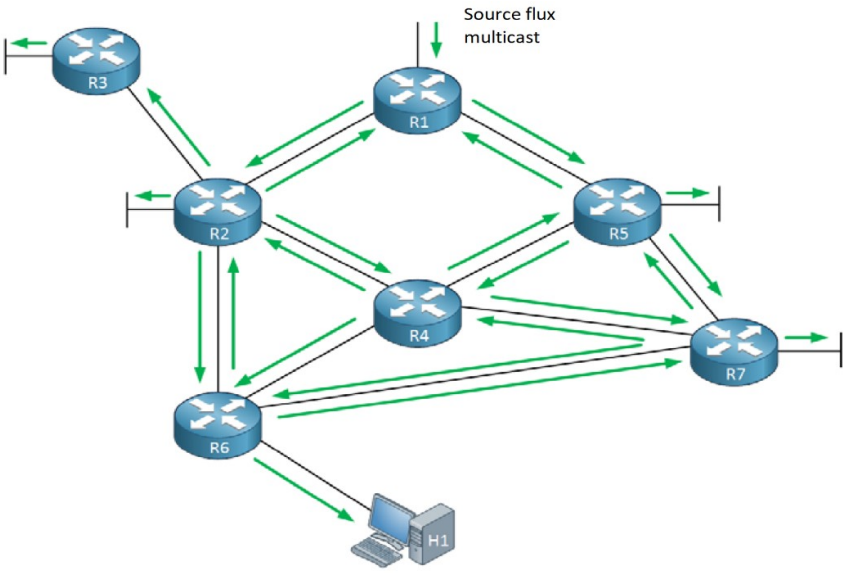
Quand le mode dense est actif sur le commutateur , La trame PIMhello est envoyé sur toutes les interfaces et le flux multicast inonde tout le monde. Le commutateur lorsqu'il recoit un PIM hello connait alors son plus proche voisin en PIM et lorsqu'il transmet il s'identifie PIM auprès de ses plus proches voisins.
Le 192.9.201.4 s'identifie auprès de ses plus proches voisins



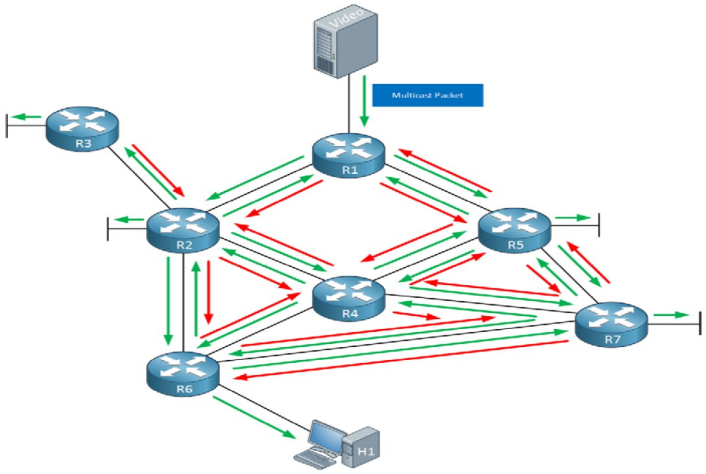

```
EXOS-VM.3 # show pim cache
Index  Dest Group      Source             InVlan   Origin
[0001] 239.0.0.1       192.9.200.6 (S)    VLAN1    Dense   Not Pruned

Number of multicast cache = 1
```

Quand le routeur reçoit un paquet Multicast. Le routeur envoie le paquet à toutes les interfaces en sortie sauf à l'interface d'origine de l'envoi.



Cette inondation peut causer un problème on risque de créer des boucles de routage multicast (exemple R1 envoie à R2,R4,R5 et retourne vers lui). Les routeurs vont envoyer un message prune pour se retirer du groupe multicast



Plusieurs raisons amènent le routeur à envoyer un message prune

- ✓ Si le routeur découvre qu'aucune interface n'a de destinataire pour ce groupe Multicast, il envoie un message « prune » en retour à l'envoyeur lui déclarant qu'il n'est pas nécessaire d'envoyer des messages à cette interface/sous réseau/lien/hôte.
- ✓ Que les routeurs proche de lui n'ont aucun destinataire pour ce groupe Multicast.
- ✓ Qu'il recoit du traffic sur une interface non RPF.

Le **reverse path forwarding** (RPF) est une technique utilisée dans le multicast pour limiter l'inondation des paquets dans le réseau.

Il consiste à ignorer un paquet s'il ne provient pas de l'interface que le routeur aurait utilisé pour router la source du paquet.

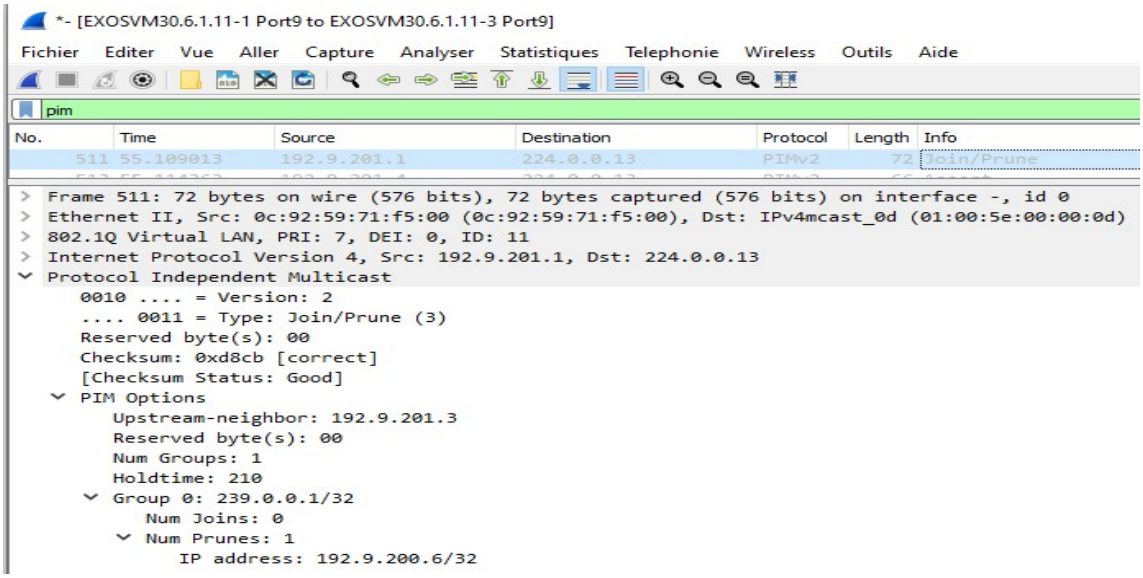
C'est un contrôle additionnel afin d'éviter les boucles multicast

Quand un routeur recoit un paquet multicast il contrôle l'@IP source de ce paquet dans sa table de routage unicast. Il contrôle aussi le prochain saut et l'interface sortante qu'il utilisera pour atteindre la source.

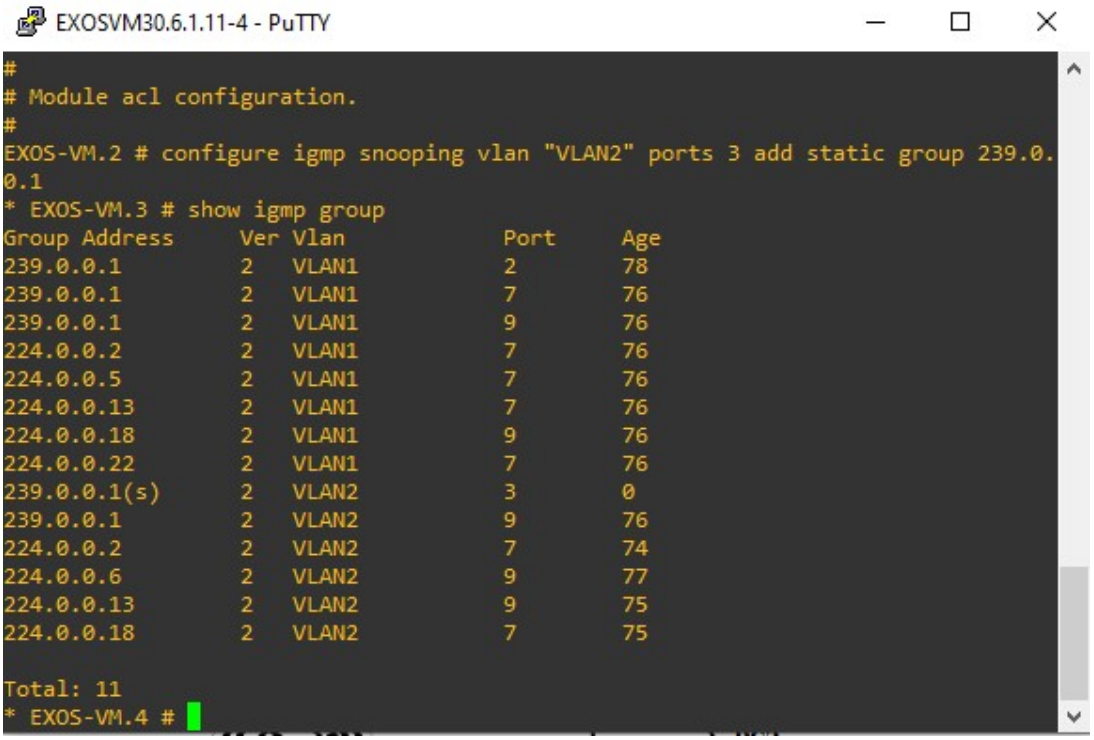
Quand le paquet multicast a été recu sur l'interface que nous utilisons pour atteindre la source le RPF est positif.

Quand le paquet multicast a été recu sur une autre interface le RPF est négatif la trame est rejeté et un message prune est envoyé.

Le 192.9.201.1 indique qu'il quitte le groupe multicast (prune) car il n'a pas de destinataire pour cette adresse multicast



Pour connaitre les ports abonnés aux flux multicast



Pour voir ses plus proches voisins sur un vlan

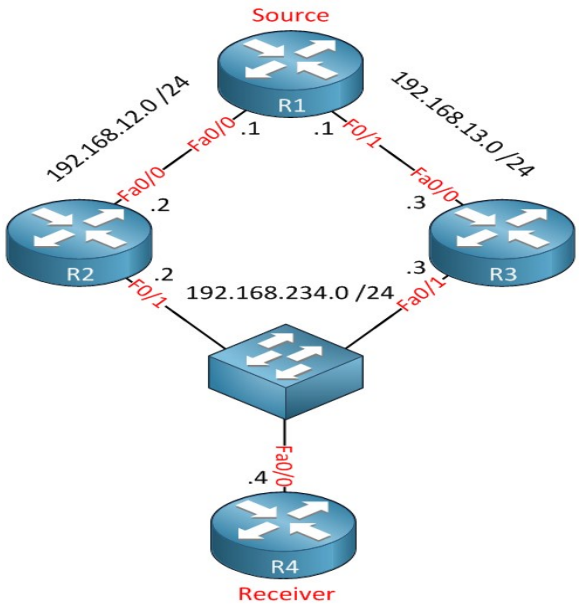
```
* EXOS-VM.6 # show igmp snooping vlan2
Router Timeout                :    260 sec
Host Timeout                  :    260 sec
Igmp Snooping Fast Leave Time :   1000 ms

VLAN VLAN2           (11) Snooping=Enabled
    Group           Sender          Age
    224.0.0.5       192.9.201.1     15
    224.0.0.5       192.9.201.2     18
    224.0.0.5       192.9.201.3     6
    224.0.0.13      192.9.201.1     21
    224.0.0.13      192.9.201.2     5
    224.0.0.13      192.9.201.3     19
    224.0.0.18      192.9.201.3     292
    239.0.0.1       192.9.200.6     0
    8 Incoming multicast streams

    Port   Host            Subscribed        Age    Group-Limit
    3      192.9.201.4     239.0.0.1(s)      0      0
    7      192.9.201.3     All Groups        3      0
    9      192.9.201.2     239.0.0.1         119    0
    9      192.9.201.2     All Groups        2      0


    s = static igmp member
```

Les messages Multicast PIM Assert



Ci-dessus, vous voyez 4 routeurs configurés pour utiliser la multidiffusion. R1 est la source et R4 est notre récepteur. Comme vous pouvez le voir, R2, R3 et R4 sont connectés au même commutateur.

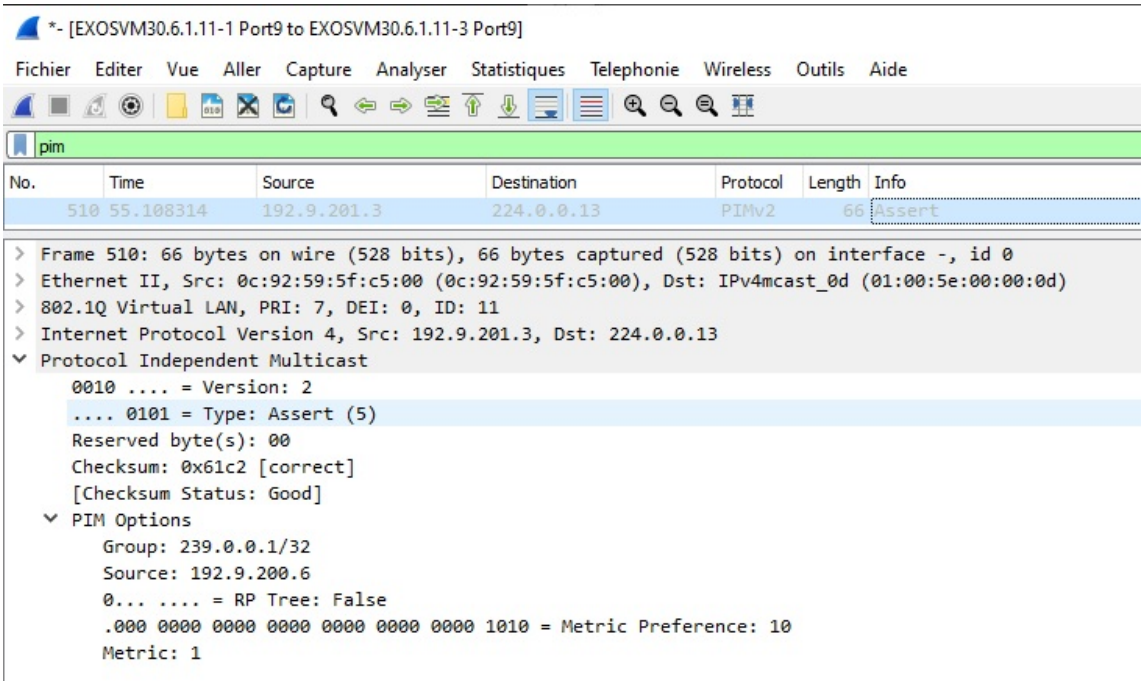
Maintenant, lorsque R1 commence à diffuser du trafic de multidiffusion vers R2 et R3, ils transfèrent tous deux les paquets de multidiffusion vers R4, ce qui entraîne un trafic en double. Pour arrêter ce **PIM, un retour d'acheminement PIM sera choisi** pour ce segment. PIM n'a pas d'informations de routage lui-même mais s'appuie sur d'autres protocoles de routage configurés, il utilisera ces informations pour sélectionner le meilleur chemin de transfert avec le mécanisme d'affirmation PIM.

Ne confondez pas le routeur d'acheminement PIM avec le PIM DR (Designated Router). ce sont deux choses différentes!

Lorsque R2 et R3 transfèrent tous deux des paquets de multidiffusion vers le segment 192.168.234.0 /24, ils voient mutuellement le trafic de multidiffusion, cela déclenche le mécanisme d'affirmation PIM. Nous élirons un chemin PIM sur la base des règles suivantes :

1. Le routeur avec la distance administrative la **plus faible** par rapport à la source du flux de multidiffusion sera le retour d'acheminement PIM élu. Cela ne se produit que si vous utilisez deux protocoles de routage ou si vous avez utilisé un itinéraire statique pointant vers la source.
2. Si l'AD est égal, nous comparerons la **métrique de routage unicast** vers la source.
3. Si l'AD et la métrique sont identiques, nous choisirons le PIM transféré en fonction de **l'adresse IP** la plus élevée.

Le routeur de transit PIM élu continuera à transmettre le trafic au récepteur tandis que le perdant élaguera son interface.



Le Dense Mode est utilisable pour les réseaux de petite taille, car il fonctionne par inondation.
Lorsqu'on souhaite faire du Multicast de niveau 3 (couche 3) avec IGMP sur un réseau de grande taille, on ne peut utiliser le dense-mode car son fonctionnement par inondation et la non-optimisation des chemins pose d'importants problèmes de trafic. De plus, si le réseau n'est pas bouclé, le dense-mode risque de ne pas fonctionner sur l'ensemble du réseau.

2) Le mode PIM sparse (PIM-SM)

Sparse mode Dans ce mode, il est supposé qu'il n'y ait aucun hôte qui veuille des paquets Multicast à moins qu'un hôte ne le demande spécifiquement.
Donc contrairement à l'approche « brutale » du dense mode, un routeur devient le point de rendez-vous RP. Ce RP log tous les hôtes qui veulent recevoir du Multicast. De plus les paquets Multicast sont envoyés uniquement à ces hôtes. Le Sparse Mode fonctionne par abonnement et référencement auprès d'un point de Rendez-vous comme nous le verrons dans la section suivante.
Le seul inconvénient du Sparse-Mode avec RP fixe est que si le RP disparait du réseau, il faut reconfigurer tous les routeurs pour utiliser un nouveau RP, ce qui suspend les possibilités de connexion Multicast durant ce temps.
Par conséquent, il existe une seconde méthode, qui est toujours du Sparse-Mode mais cette fois utilisant un RP dynamique automatiquement sélectionné et automatiquement remplacé en cas de disparition

2.1) Le point de rendez-vous (RP)

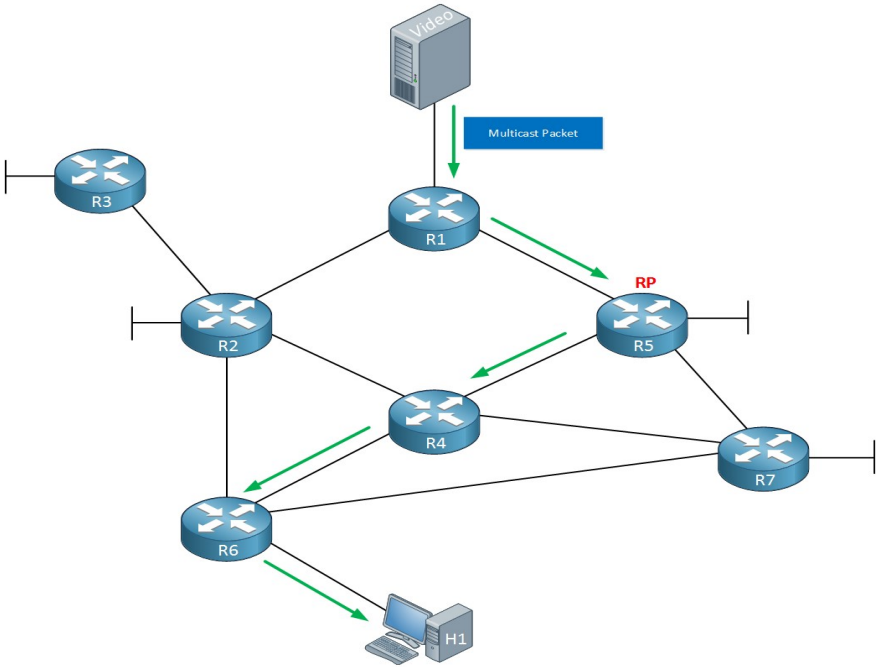
PIM-SM est un protocole de routage des flux de données multicast qui se base sur les notions de Rendez-vous Point (RP) et d'arbre partagé (Shared Tree).
Le RP est le point de rencontre initial entre une source diffusant un groupe multicast et les abonnés membre de ce groupe. Le RP constitue la racine d'un arbre partagé servant à la diffusion d'un flux multicast.
Pour comprendre comment les chemins sont établis à travers un réseau PIM, des sources aux abonnés, il est utile d' examiner le processus des deux côtés.
Page suivante, nous avons un serveur vidéo qui diffuse du trafic de multidiffusion vers 239.1.1.1. En bas, il y a R6 qui FAIT PARTIE du groupe d' abonné multicast.

Comment R6 est-il censé savoir que le serveur vidéo à 192.168.1.1 envoie du trafic de multidiffusion pour ce groupe? Avec le mode PIM dense, c'est simple. Le trafic de multidiffusion a été inondé partout, ce qui a permis à R6 d'apprendre l'adresse source.

Pour résoudre ce problème, le mode éparse PIM utilise un **RP (Rendez-vous Point)** dans le réseau. Voici comment cela fonctionne :

1. Chaque routeur qui reçoit du trafic de multidiffusion d'une source le transmettra au RP.
2. Chaque routeur qui souhaite recevoir du trafic de multidiffusion ira au RP.

Le RP est comme un « point de rencontre » pour le trafic de multidiffusion. Voici un exemple :



Ci-dessus, vous pouvez voir que R5 est devenu le RP de notre réseau. R1 reçoit le trafic de multidiffusion du serveur vidéo et le transmet au RP. En bas, nous avons R6 qui reçoit ce trafic de multidiffusion du RP et il est transmis à l'hôte.

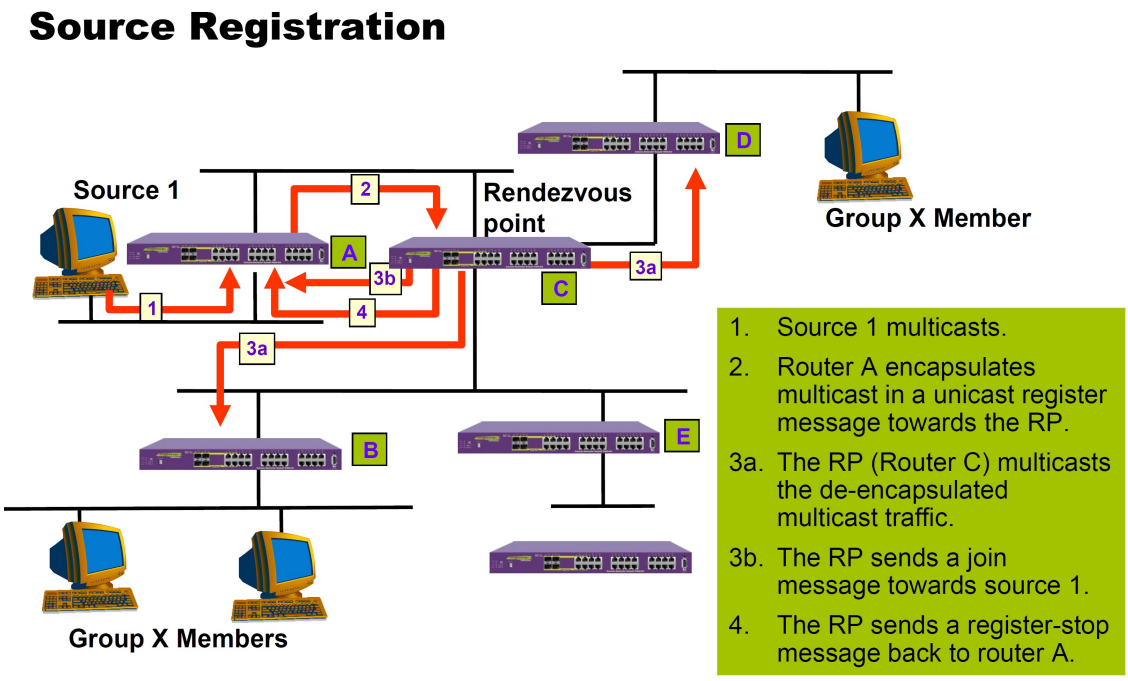
Cela introduit cependant un autre problème... Comment R6 peut-il savoir où trouver le RP ? Il existe plusieurs options différentes pour cela:

1. Nous pouvons configurer manuellement l'adresse IP du RP sur chaque routeur.

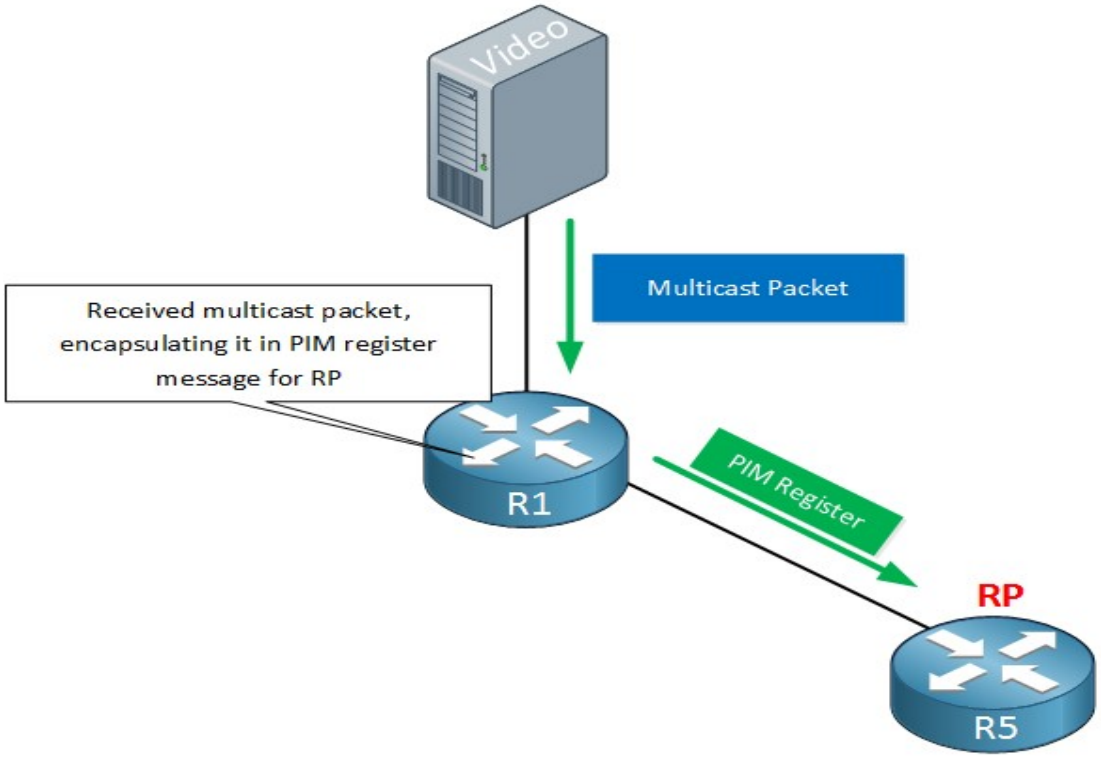
2. Il existe certains protocoles qui peuvent découvrir automatiquement le RP dans notre réseau. Nous ne garderons que cette solution par la suite car la première impose de reconfigurer les routeurs si le point de rendez-vous ne fonctionne plus.

Lorsque PIM-SM est activé sur un commutateur, il envoie un message PIM-SM Hello sur tous ses PIM-SM interfaces et écoute les messages Hello de ses voisins PIM-SM. Lorsqu'un commutateur reçoit un message Hello, il enregistre l'interface, l'adresse IP, la priorité pour devenir un DR et le délai d'expiration pour les informations du voisin. Le commutateur envoie régulièrement des messages Hello au Hello Time intervalle.

2.2) Enregistrement auprès du RP de la source



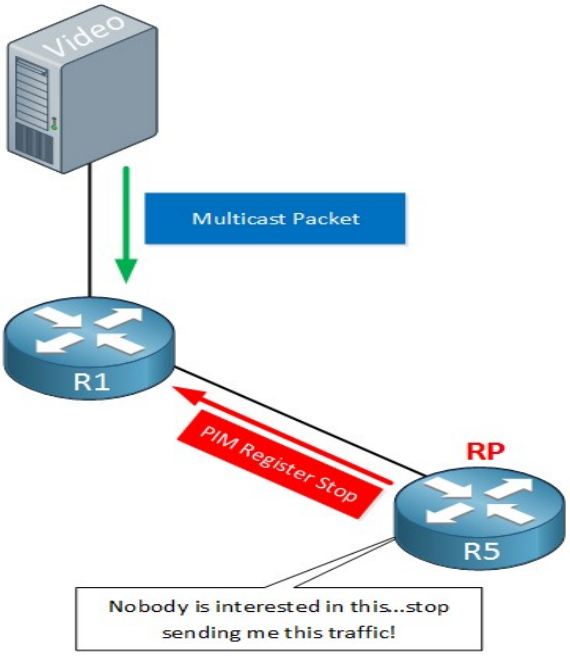
Imaginons que nous utilisions la topologie ci-dessus mais *pour le moment, personne ne s'intéresse au trafic de multidiffusion.* Voici ce qui se passe :



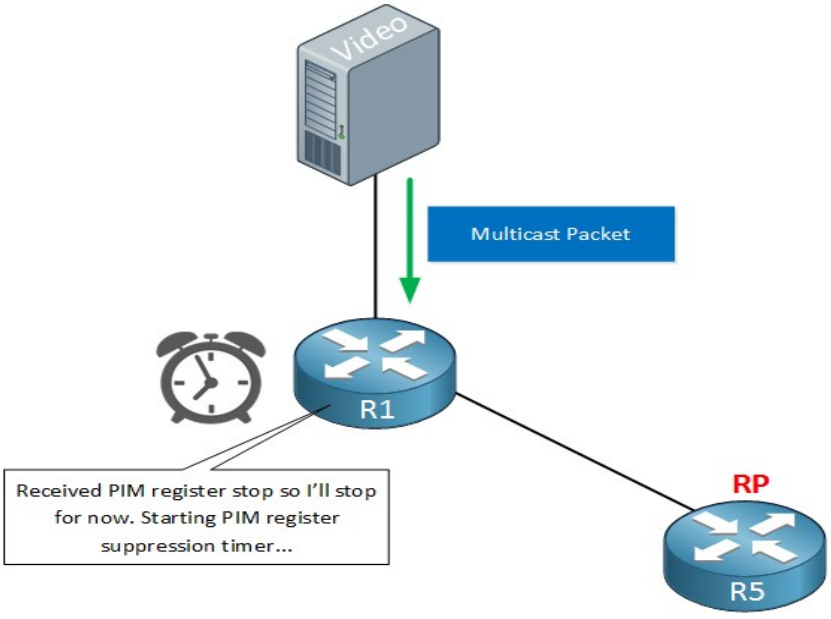
Notre serveur vidéo transfère le trafic multicast sur son interface qui est reçu par R1. Étant donné que nous utilisons le mode PIM clairsemé, ce routeur devra transférer le trafic de multidiffusion vers le RP. Au lieu de tout transférer, R1 n'enverra que le **premier paquet de multidiffusion**. Ce paquet est encapsulé dans un message **de registre PIM** et transféré au RP. Une fois que le RP reçoit le message du registre PIM, deux options s'offrent à vous :

1. Lorsque personne n'est intéressé par le trafic de multidiffusion, le RP **rejettera** le message du registre PIM.
2. Lorsqu'il y a au moins un récepteur, le RP **accepte** le message de registre RP.

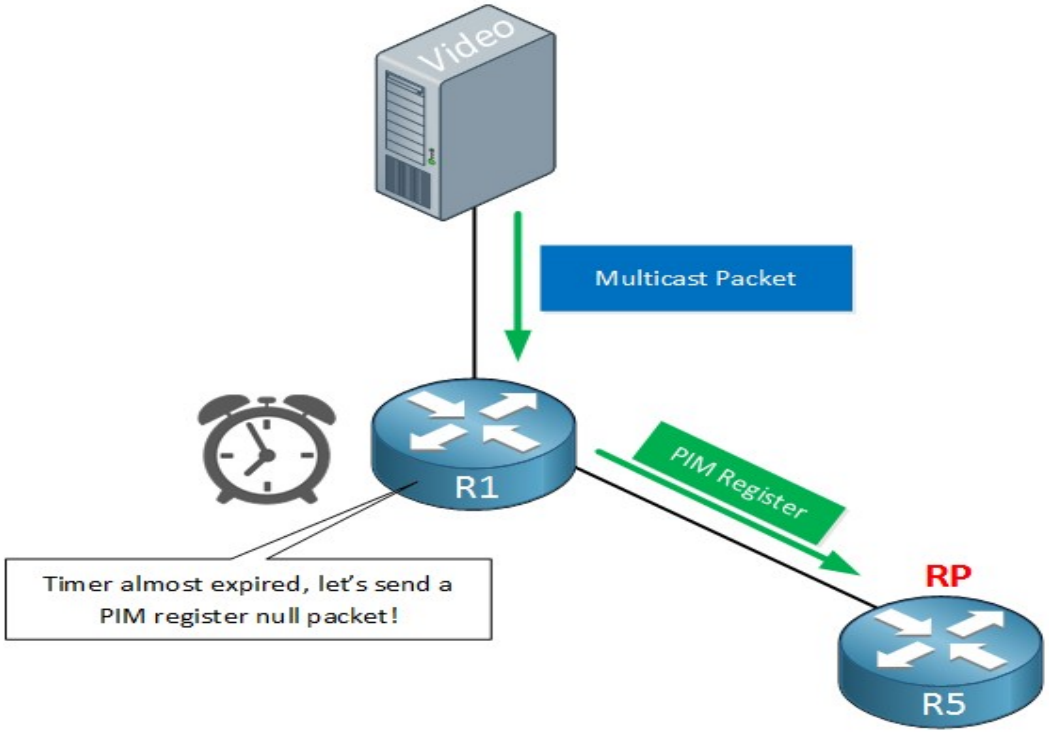
Pour l'instant, disons que nous n'avons pas de récepteurs:



Notre RP répondra à R1 avec un message **d'arrêt du registre** PIM. Cela informera R1 d'arrêter de transférer tout trafic de multidiffusion pour le moment. R1 restera silencieux pour l'instant mais pas pour longtemps :

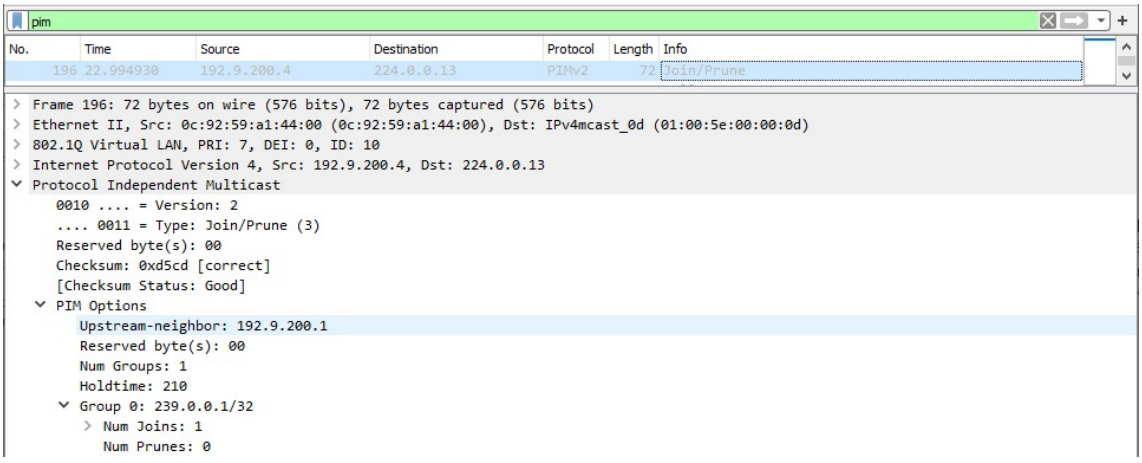


Lorsque nous recevons le paquet d'arrêt du registre PIM, R1 démarre une temporisation. Par défaut, cette temporisation est de **60 secondes** et lorsque la temporisation est presque expirée, R1 enverra un autre paquet:



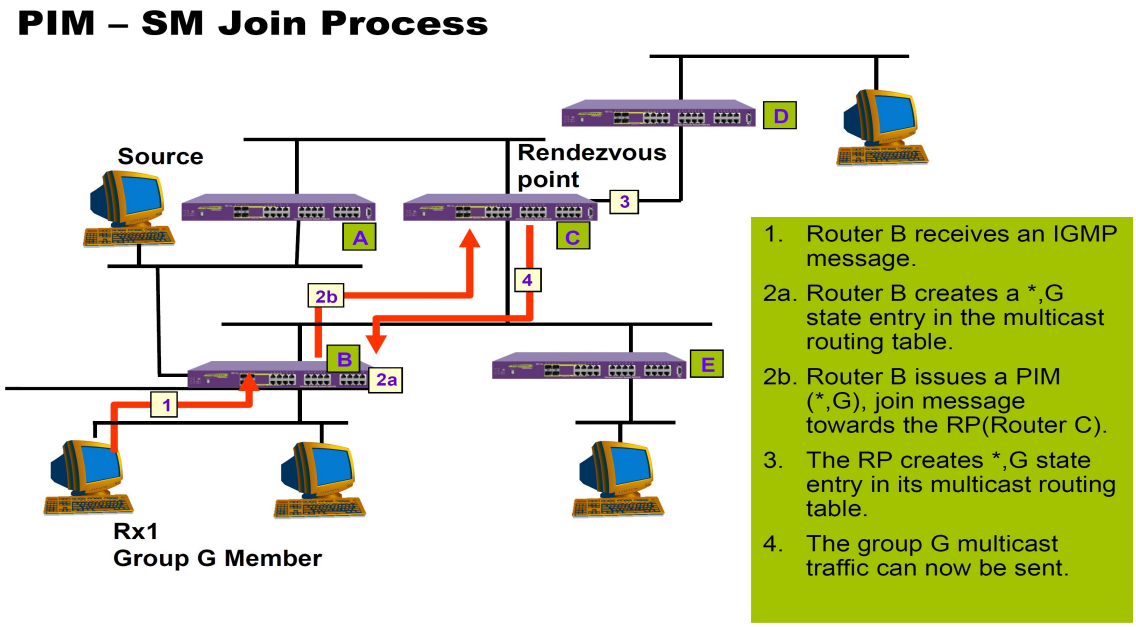
R1 enverra un autre message de registre PIM, mais celui-ci ne porte pas le paquet de multidiffusion encapsulé. C'est une simple demande de demander au RP s'il est intéressé maintenant. Ce paquet est appelé paquet null du **registre PIM.** Lorsque nous n'avons toujours pas de récepteurs, le RP enverra un autre message d'arrêt du registre PIM. Lorsque nous avons des récepteurs, le RP n'enverra pas de message d'arrêt du registre PIM et R1 commencera à transférer le trafic de multidiffusion.

Si on a un abonné le RP renvoie un PIM join vers la source exemple 200.4 vers 200.1

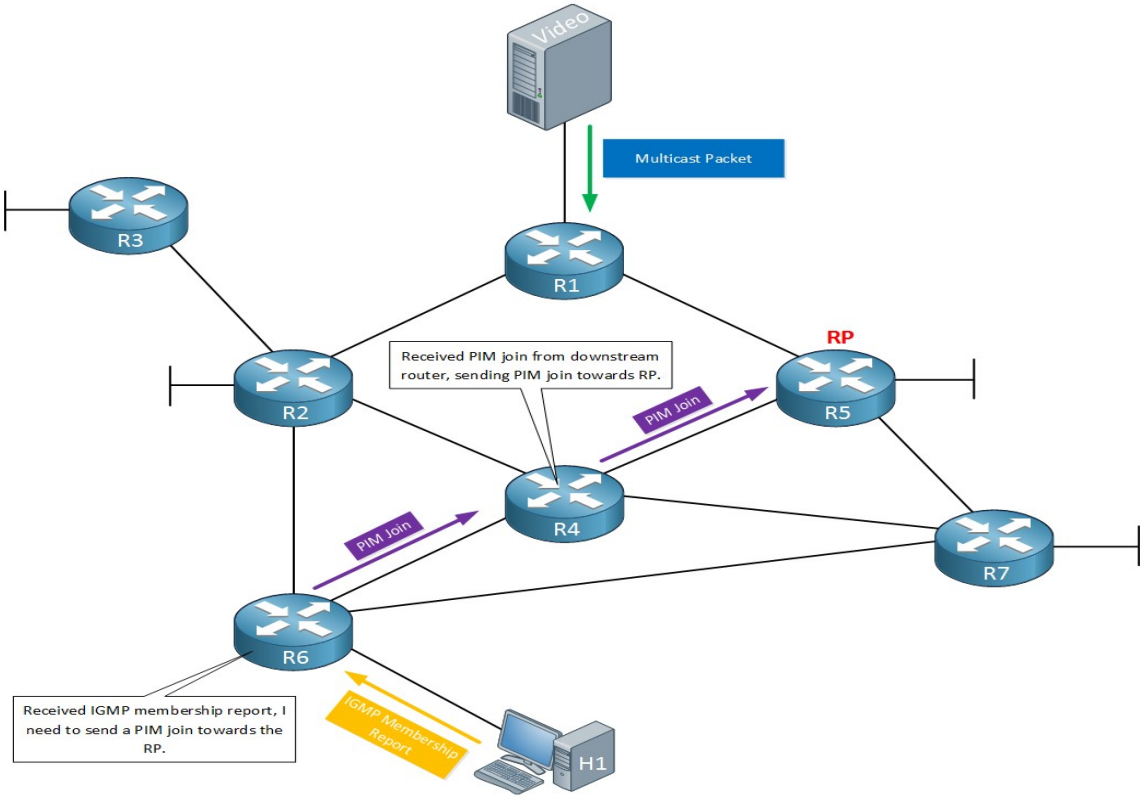


Vous savez maintenant ce qui se passe lorsqu'une source commence à envoyer du trafic de multidiffusion sans aucun récepteur.

2.3) L'Arbre partagé (Shared Tree) : Enregistrement auprès du RP de l'abonné



Voyons ce qui se passe lorsque nous avons des récepteurs, n'est-ce pas? Jetons un coup d'œil à l'image ci-dessous:



L'hôte connecté à R6 souhaite recevoir le trafic de multidiffusion afin d'envoyer un rapport d'appartenance IGMP pour le groupe de multidiffusion souhaité. R6 doit maintenant comprendre comment accéder au RP et lui demander de commencer à transférer le trafic de multidiffusion. Il vérifiera sa table de routage unicast pour l'adresse IP du RP et enverra un message de mise en relation PIM sur l'interface utilisée pour atteindre le RP. Dans ce cas, cela signifie que la de mise en relation PIM est redirigée vers R4.

Lorsque R4 reçoit la de mise en relation PIM, il doit demander au RP de commencer à transférer le trafic de multidiffusion afin d'envoyer également une demande de mise en relation PIM. Il vérifiera sa table de routage unicast, trouvera l'interface utilisée pour atteindre le RP et enverra un message de de mise en relation PIM vers le RP.

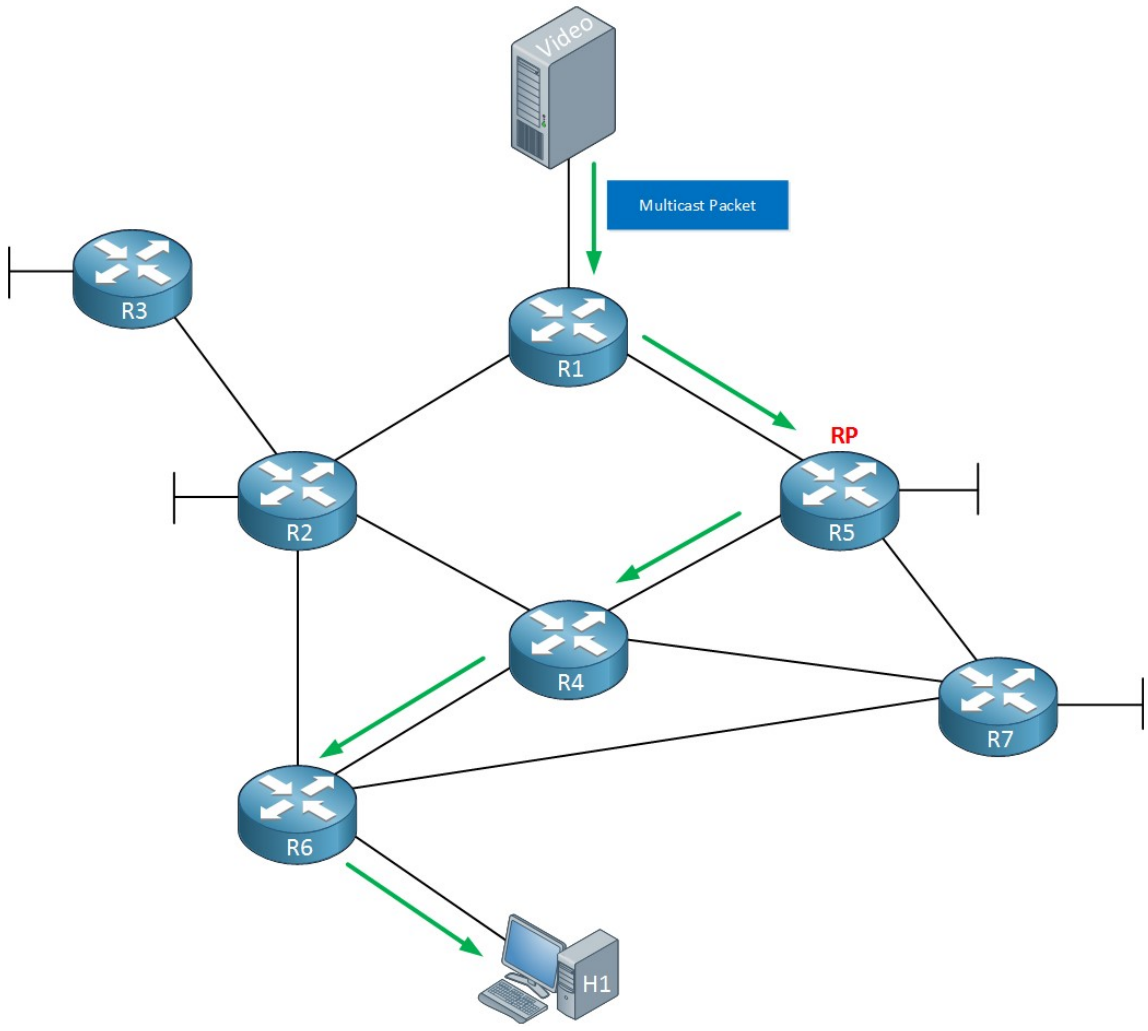
Pour résumer cela, les routeurs clairsemés PIM enverront un message de mise en relation PIM lorsque :

1. Le routeur a reçu un rapport d'adhésion IGMP d'un hôte sur une interface directement connectée.
2. Le routeur a reçu une de mise en relation PIM d'un routeur en aval.

Lorsque le RP reçoit la de mise en relation PIM, il commence à transférer le trafic de multidiffusion.

Ce concept de jonction du RP est appelé **RPT (Root Path Tree)** ou arbre de **distribution partagée.** Le RP est la racine de notre arbre qui décide où transférer le trafic de multidiffusion. Chaque groupe de multidiffusion peut avoir des sources et des récepteurs différents, de sorte que nous pouvons avoir des RPG différents dans notre réseau.

Le résultat final ressemblera à ceci :



Le trafic de multidiffusion circule maintenant de R1 vers le RP, jusqu'à R4, R6 et vers notre récepteur (H1).

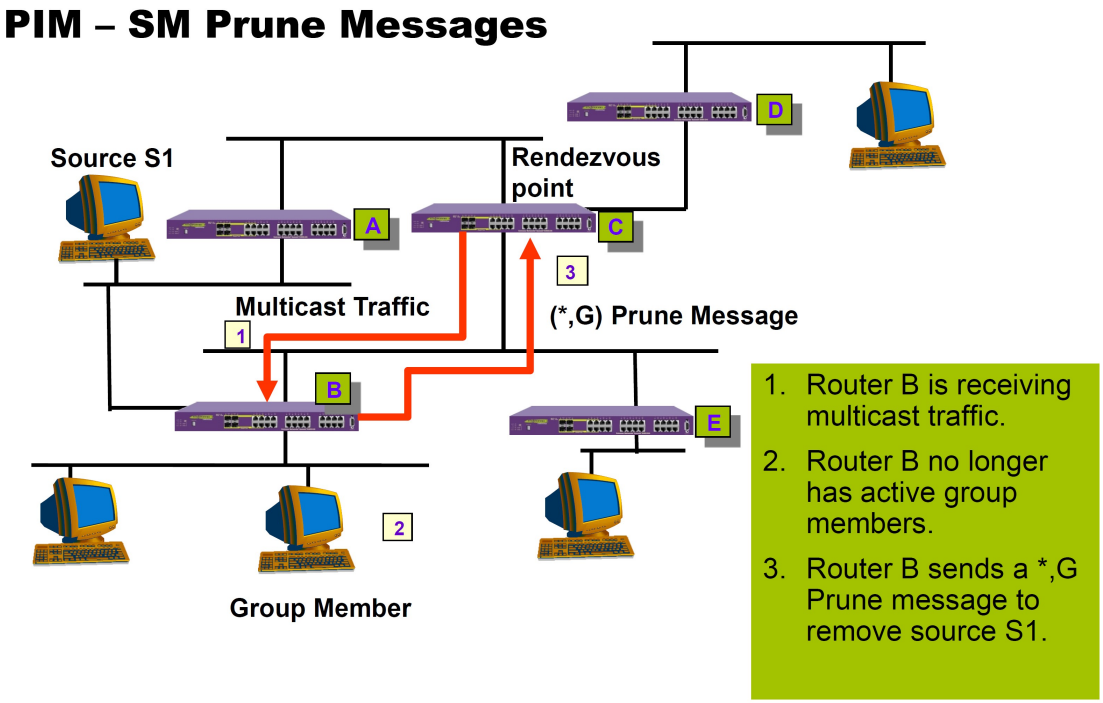
Avons-nous terminé maintenant? Presque...

Si vous regardez attentivement l'image ci-dessus, vous avez peut-être remarqué que R6 a plusieurs chemins vers la source. À l'heure actuelle, notre trafic de multidiffusion se déroule comme ceci:

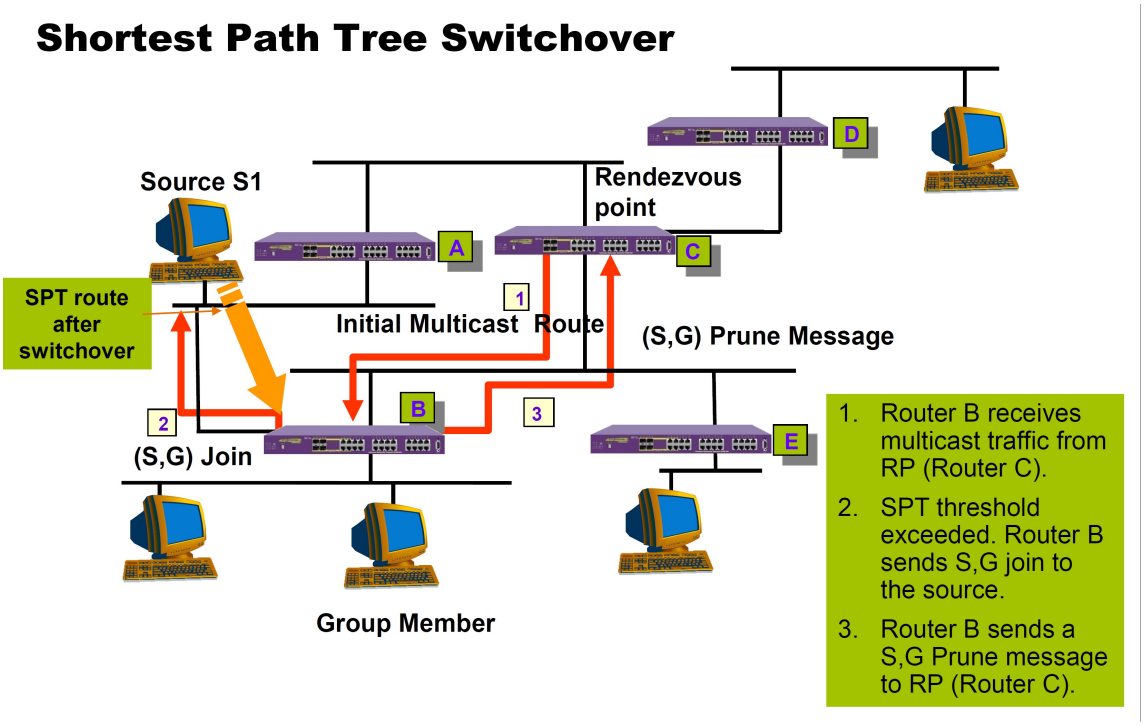
1. R1 > R5 > R4 > R6

Il est cependant possible que ce ne soit pas le chemin le plus optimal. Le chemin de R1 > R2 > R6 a un routeur de moins que le chemin actuel, donc si toutes les interfaces sont égales, ce chemin est probablement meilleur.

Si la demande parvient à un routeur qui reçoit déjà le flux en question, avant d'arriver au RP, ce routeur commencera à envoyer le flux dans la direction à partir de laquelle la mise en relation est arrivé et ne transmettra pas la demande d'abonnement plus en amont.
Si la demande finit par aller jusqu'au RP, le RP commencera à transférer le flux vers le bas dans la direction d'où la demande d'abonnement est arrivée. Si le RP n'a jamais réellement reçu le flux demandé, alors il ne fera tout simplement rien - il n'y a pas de mécanisme dans PIM pour un routeur de dire « désolé, ce flux n'existe pas ».
Une fois que le routeur demandeur reçoit le flux, il verra l'adresse IP source du flux. À ce moment-là,
il peut en profiter pour envoyer une demande directement à la source, plutôt que d'avoir à recevoir le flux via le RP. La RFC ne spécifie pas combien de temps après avoir commencé à recevoir le flux un routeur doit commencer à envoyer des demandes de mise en relation directement vers la source.
Dès qu'il reçoit un paquet dans le flux, il apprend l'adresse IP source du flux, puis commence à envoyer des demandes directement vers la source.



2.4) Arbre centré sur la Source (Source Tree)



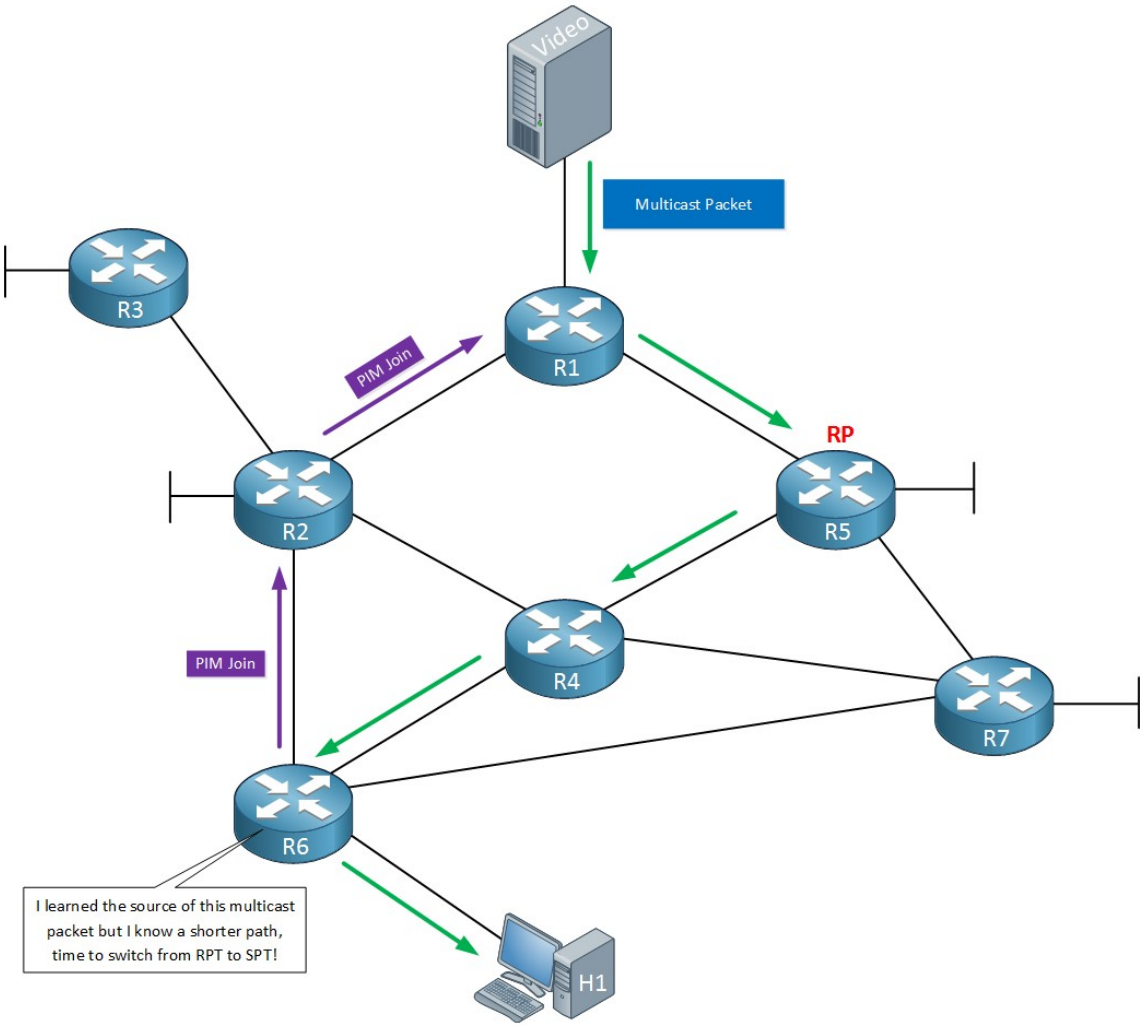
Une fois que nous recevons du trafic de multidiffusion via le RP, il est possible de passer au **SPT (Source Path Tree).**

Une fois l'arbre partagé construit et le flux acheminé vers le routeur de la source, un arbre de plus court chemin vers la source (Source Tree) est créé et la diffusion stoppée sur l'arbre partagé.

Dans ce cas, le DR émet un message « PIM (S,G) JOIN » vers l'émetteur. Cela crée des états spécifiques (S, G) dans les routeurs rencontrés sur le chemin vers l'émetteur. Le message atteint le DR de la source ou un autre routeur ayant déjà l'état (S, G).

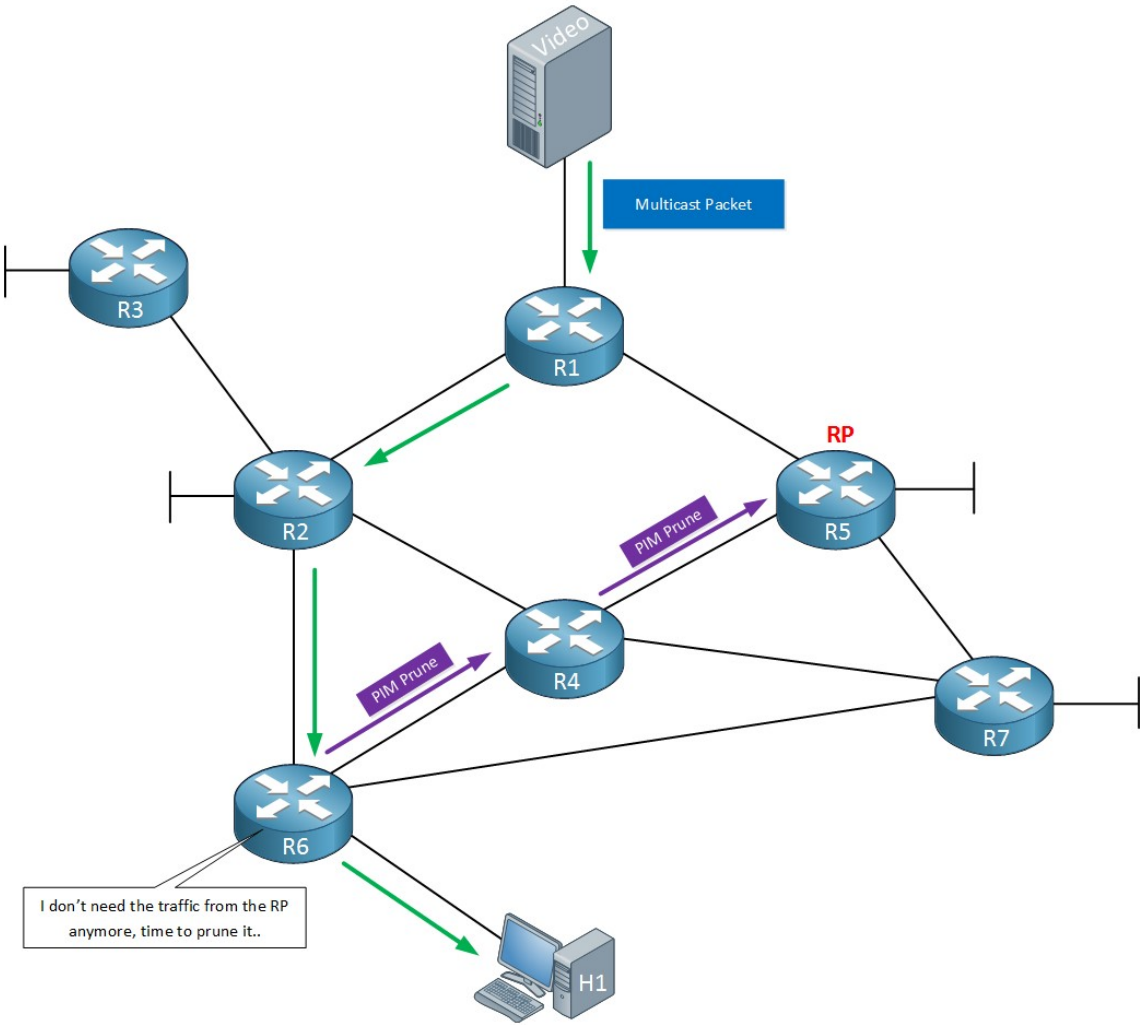
Les paquets multicast dorénavant émis par l'émetteur S suivront les états (S, G).
À partir de cet instant le DR du destinataire peut recevoir deux copies de chaque paquet multicast : une provenant du RP et ayant suivi l'arbre partagé associé, l'autre provenant directement de l'émetteur en ayant emprunté l'arbre centré sur la source.

Dès que le premier paquet multicast est reçu en provenance de l'arbre centré sur la source, le DR du destinataire détruit les paquets qui arrivent via le RP. Il envoie un message d'élagage « PIM (S,G) PRUNE » qui est propagé de routeur en routeur vers le RP. Dans chaque routeur rencontré ce message place un état indiquant que le trafic multicast (S, G) ne doit plus être propagé sur l'arbre partagé.
De manière plus imagée voici ce qui se passe ensuite :



Lorsque R6 a reçu le trafic de multidiffusion via le RP, il a également appris l'adresse source de ce trafic. R6 peut maintenant décider qu'il veut utiliser le SPT au lieu du RPT pour recevoir ce trafic. Quand il le souhaite, il enverra des messages de de mise en relation PIM vers la source.

R1 et R2 commenceront à transférer le trafic de multidiffusion vers R6, en utilisant le meilleur chemin:



Étant donné que R6 reçoit maintenant du trafic de multidiffusion via R2 et R1, il n' a plus besoin du RP. Il enverra des messages d'élagage PIM (message prune) vers le RP, informant tous les routeurs entre les deux qu'ils peuvent arrêter de transférer le trafic de multidiffusion pour ce groupe.

Vous devriez maintenant avoir une idée générale du fonctionnement du mode clairsemé PIM.

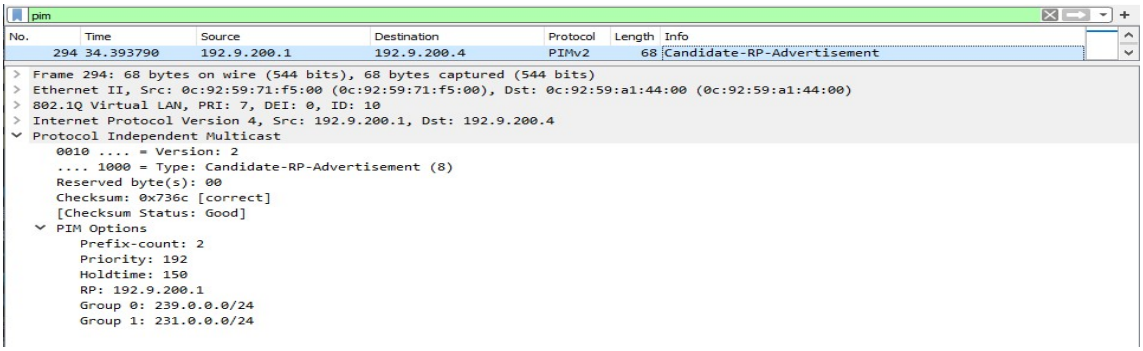
2.5) Protocole de de mise en place de l'Adresse dynamique IP du RP

Il existe certains protocoles qui peuvent découvrir automatiquement le RP dans notre réseau.
Le protocole d'attribution d'une adresse IP dynamique du point de rendez-vous est assuré grâce au routeur bootstrap (BSR) .

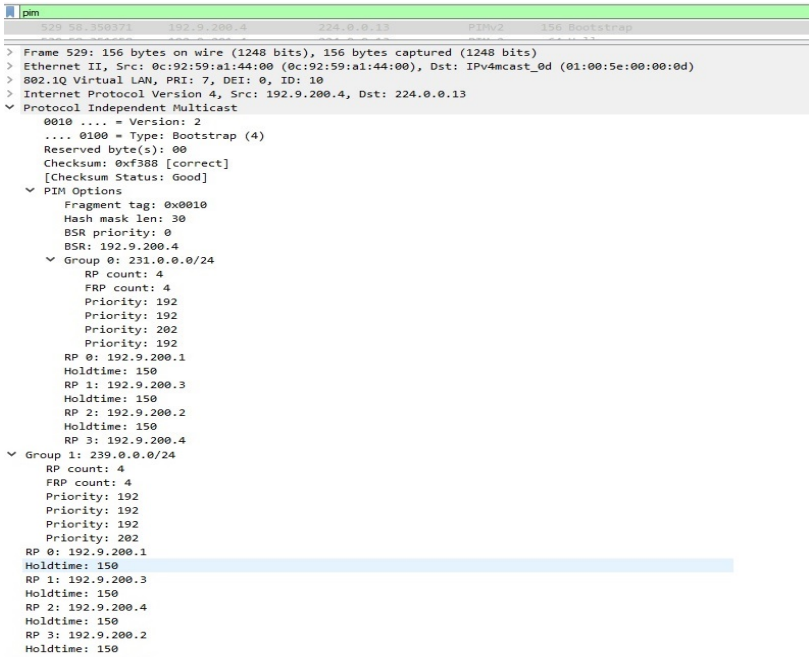
Bootstrap Router (BSR)

Encore une fois, le routeur Bootstrap pour un réseau est choisi par élection. Un certain nombre de routeurs dans le réseau peut être configuré pour être candidat au rôle BSR (CBSR). Chaque réseau PIM-SM doit avoir au moins un Candidat Au routeur d'amorçage (BSR), sauf si tous les commutateurs du domaine sont configurés statiquement avec des informations sur tous les PDR du domaine. Chaque commutateur qui est un candidat BSR envoie périodiquement un Message d'annonce de candidat Bootstrap pour annoncer qu'il est disponible en tant que candidat BSR.
Les candidats BSR du réseau élisent le commutateur avec la valeur la plus élevée pour être le BSR.
Le BSR élu écoute les messages du candidat PIM RP (CRP) spécifiant les candidats RP pour les groupes de multidiffusion. Il maintient une liste de candidats RP et envoie un message bootstrap à chaque BSR :

Exemple d'émission d'un candidat message CRP du 192.9.200.1



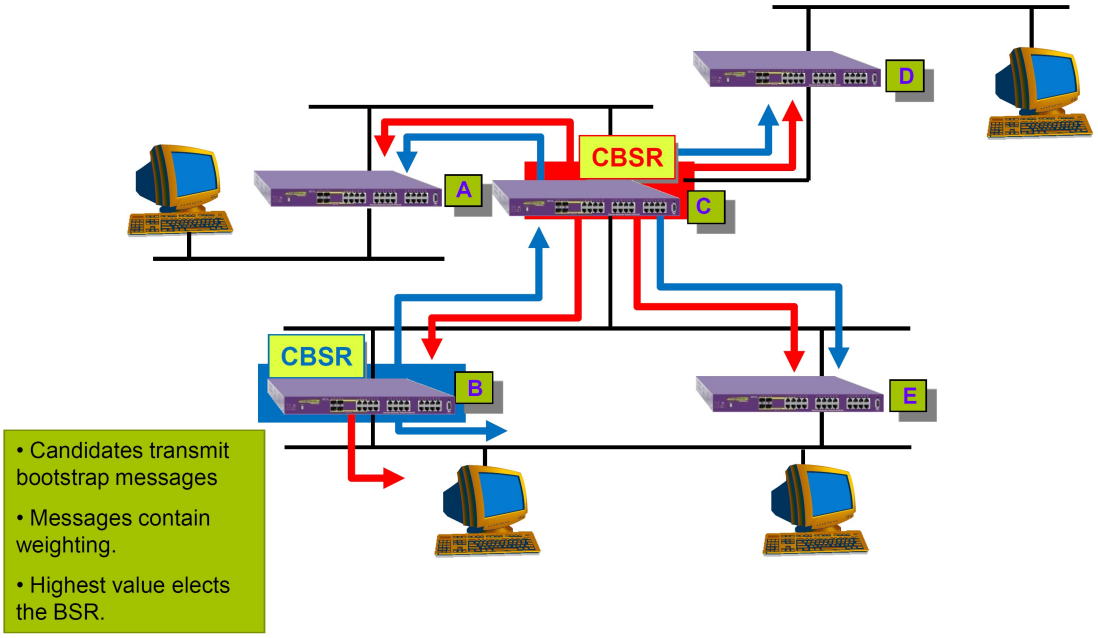
Exemple d'émission d'un message bootstrap du 192.9.200.4



En spécifiant tous les groupes de multidiffusion du réseau PIM et leurs candidats RP. Chaque routeur utilise ces informations et un algorithme de sélection standardisé pour déterminer le RP pour chaque groupe.
En résumé :

- ✓ Chaque groupe de multidiffusion doit avoir au moins un candidat RP
- ✓ Chaque domaine PIM-SM doit avoir au moins un candidat BSR, sauf si tous les routeurs du domaine sont configuré statiquement avec des informations sur tous les PDR du domaine
- ✓ Chaque sous-réseau doit avoir au moins un candidat DR.

BSR Election



Le BSR génère des messages d'amorçage qui permettent de:

- ✓ Distribuer les informations du point de rendez-vous.
- ✓ De les propager dans tout le domaine.
- ✓ D'Organiser des élections BSR si nécessaire.

Les routeurs du domaine sont configurés en tant que BSR candidats (CBSR).

Le processus électoral sélectionne un seul routeur comme BSR.
Tout d'abord, regardons comment est établi la demande de l'abonné remis au RP.

L'auditeur doit s'appuyer sur un RP pour lui fournir le flux. En recherchant son mappage Group-to-RP (qu'il a appris du BSR) le routeur connaît l'adresse du RP qui devrait être en mesure de fournir le flux demandé.

Le routeur envoie une demande de mise en relation vers le RP. La demande n'est pas unicast vers le RP, elle est envoyée comme une multidiffusion vers le voisin qui est le prochain saut sur le chemin vers le RP. Ce voisin, à son tour, transmettra la demande à son voisin du saut suivant sur le chemin vers le RP. La demande continuera d'être transmise de voisin à voisin de cette manière jusqu'à ce que l'un ou l'autre arrive à un routeur qui reçoit déjà le flux demandé, ou il atteint le RP.

Annexe

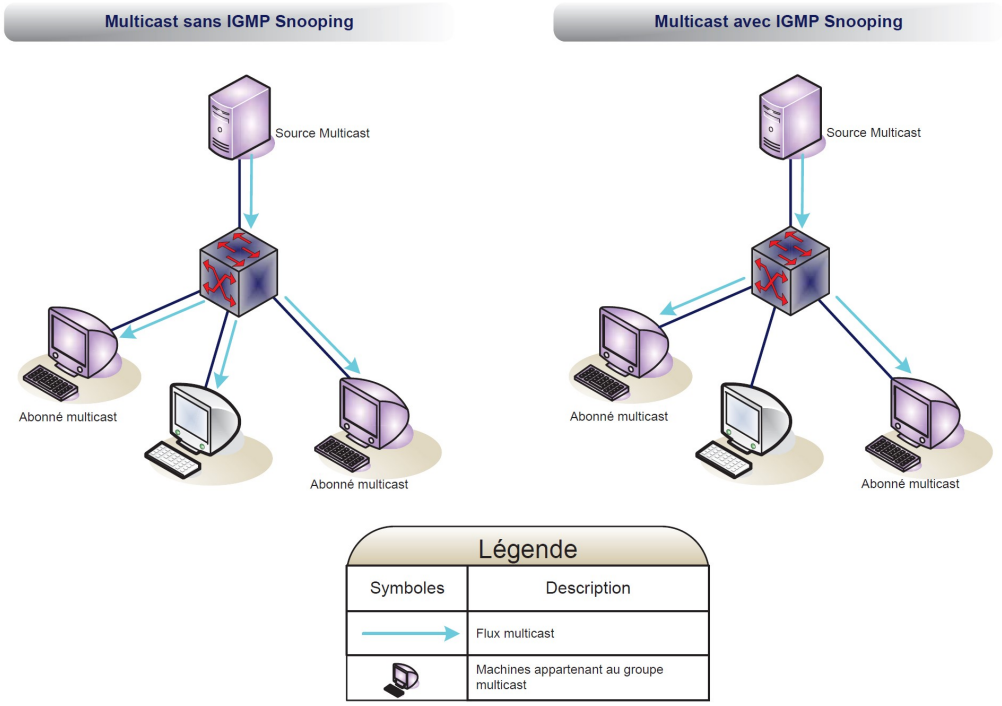


Figure 11 - IGMP Snooping

Designated Router

Lorsque deux routeurs PIM-SM se trouvent sur le même LAN, il y a élection d'un Designated Router (DR). L'élection est faite via des messages « PIM HELLO », le DR est élu selon les critères suivants : la priorité du DR ou l'adresse IP la plus élevée.
Le DR a pour rôle de prendre en charge le routage des flux multicast.

