|  |
| --- |
| Une brève présentation Multicast  Par Ph. Jounin |

# Définition et intérêt :

Un flux multicast est un flux unidirectionnel qui possède une source unique et un nombre de récepteurs quelconque. Des flux de type télévision/radio sont de bons candidats à une transmission en multicast.

Pour transmettre un flux multicast, les équipements de communication routeurs et commutateurs doivent répliquer la trame reçue pour la transmettre vers l’ensemble des récepteurs.

Dans un fonctionnement pur multicast (sans overlay), sur un lien donné, un flux multicast de N b/s utilise soit N b/s (il y a au moins un récepteur derrière le lien) soit 0 b/s (il n’y a pas de récepteurs).



# Format de trame

Un flux multicast IPv4 utilise la plage d’adresse 224.0.0.0/4 pour l’adresse de destination, on parle alors d’adresse de groupe. L’adresse source est l’adresse de la machine source.

En général, le flux est encapsulé dans une trame UDP pour le démultiplexage sur les récepteurs. Certains flux fonctionnent toutefois en IP, c’est le cas pour les protocoles de routage par exemple.

Au niveau Ethernet, l’adresse source est l’adresse de l’émetteur sur le LAN, l’adresse de destination est calculée par un mapping de l’adresse IP (attention il y a recouvrement, plusieurs adresses IP correspondent à la même adresse MAC multicast).

# Principe de routage :

Les récepteurs doivent se faire connaître auprès des équipements réseau pour recevoir le flux multicast à destination d’une adresse donnée (en /32). [Protocole IGMP]. Les commutateurs peuvent intercepter ces échanges pour créer une table de commutation optimisée [IGMP snooping].

Les routeurs remontent de proche en proche vers la -ou les- source en mémorisant l’interface par laquelle ils ont reçus la demande d’abonnement à l’adresse de groupe. [Protocole PIM]

Quand ils reçoivent le flux multicast, les routeurs vérifient que le flux est reçu par l’interface attendue (en général l’interface par laquelle l’adresse source est routée). [Check RPF]. Si ce n’est pas le cas, le flux reçu est droppé.

Ainsi, chaque routeur mémorise un état sous la forme :

*Adresse Source, Adresse de Groupe, Interface d’entrée 🡪 Liste des interfaces de destinations*

Si l’adresse source est connue l’état est noté (S,G), si elle n’est pas connue, l’état est noté (\*,G).

# Différentes versions de PIM

PIM existe en 3 versions incompatibles entre elles :

* ASM
* SSM
* Bidir

## PIM ASM (Any Source Multicast)

Les demandes d’abonnement à un flux n’utilise que l’adresse de groupe. Le récepteur signale un abonnement pour le flux multicast a.b.c.d/32. Cette demande est correctement acceptée par le routeur IGMP querier, mais celui-ci n’a aucun moyen de remonter directement vers la source.

Toutefois il possède l’adresse (unicast) d’un point de rendez-vous auprès de qui transférer les demandes d’abonnement. Après avoir mémorisé la demande d’abonnement sous la forme d’un arbre (\*, G), il translate la requête IGMP en requête PIM adressée vers le routeur le plus proche (au sens de sa table de routage) du point de rendez-vous.

Le point de rendez-vous joue alors 2 rôles :

* Mémorisation de tous les états multicast
* Réception des flux initiaux

### Initialisation d’un flux multicast (spécifique ASM)

|  |  |  |
| --- | --- | --- |
| **Phase** | **Type** | **Description** |
| 1 | Data plane | La source commence à émettre |
| 2 | Mixte | Le routeur encapsule le flux en unicast et l’envoie au RP |
| 3 | Control plane | Le RP crée un état (S, G) avec une liste de destination vide, un message JOIN est envoyé vers le routeur le plus près de la source (ici vers I) |
| 4 | Control plane | Le routeur I crée un état (S,G) avec interface d’entrée vers S et interface de sortie vers RP, il envoie un message JOIN vers S |
| 5a | Control Plane | Le routeur S crée un état S,G avec interface d’entrée le LAN, interface de sortie vers I |
| 5b | Data Plane | Le routeur S utilise l’état (S,G) pour forwarder le flux vers I |
| 6 | Data Plane | Le routeur I forwarde le flux vers RP, RP le droppe (pas de destination pour son état S,G) |
| 7 | Control Plane | RP reçoit deux copies du même flux. Il envoie un message register-stop vers S pour lui signaler qu’il recoit bien le flux nativement. S cesse alors d’encapsuler le flux en unicast |



### Routage d’un flux multicast (spécifique ASM)

|  |  |  |
| --- | --- | --- |
| **Phase** | **Type** | **Description** |
| 0 | Data plane | La source émet vers le RP (condition initiale) |
| 1 | Control plane | Le poste récepteur s’abonne au flux |
| 2 | Control Plane + Data Plane | Le switch intercepte la trame IGMP et place l’interface vers le poste récepteur comme interface de destination pour l’adresse Ethernet correspondante au groupe multicast. |
| 3 | Control plane | Le routeur crée un état (\*, G) avec son LAN comme liste de destination, il envoie un message JOIN vers le RP |
| 4 | Control plane | Le RP ajoute l’interface vers D à la liste de destination des états (\*,G) et (S,G) |
| 5 | Data Plane | L’état (S,G) possède une interface de destination et le flux est acheminé vers D |
| 6 | Data Plane + Control Plane | D reçoit le flux (S,G) en provenance de RP. Il crée l’état (S,G) mais l’interface d’entrée est incorrecte (il s’attend à recevoir le flux par S). Au niveau data plane, le flux est droppé. Il envoie un message JOIN (S,G) vers S et un message PRUNE (S,G) vers le RP. |
| 7 | Data Plane + Control Plane | Le RP supprime l’interface vers D de son état S,G et ne forwarde plus le flux. |
| 8 | Data Plane + Control Plane | S ajoute l’interface vers D à son état (S,G) et forwarde le flux vers I et D (duplication des trames) |
| 9 | Data Plane | D accepte le flux en provenance de S et le forwarde vers le switch qui le commute sur la seule interface active |



## PIM SSM (Source Specific Multicast)

En SSM, les récepteurs connaissent l’adresse source (unicast) des flux auxquels ils s’abonnent. Il n’y a pas d’états (\*,G) et les routeurs ne manipulent que des états (S,G). Le routage est plus simple, il n’y pas de RP et pas d’optimisation du chemin de flux (sauf en cas de changement de topologie).

## PIM Bidir

Bien qu’il ait été publié en RFC, ce protocole n’est implémenté que par Cisco. Cette fois, les routeurs ne gère que les états (\*,G) ce qui diminue fortement le nombre d’états si le nombre de source est important (par exemple si les récepteurs sont également source). En revanche, il n’y a plus d’optimisation de flux.

Un point de rendez-vous fictif est configuré sur les routeurs. Les flux sont systématiquement envoyés vers ce point et les demandes d’abonnement également. En d’autres mots, un état (\*,G) possède au moins l’interface vers le RP dans sa liste de destinations (sauf peut-être pour les routeurs directement connectés au RP).

|  |  |  |
| --- | --- | --- |
| **Phase** | **Type** | **Description** |
| 0 | Data plane | La source émet vers S. S crée l’état (\*,G) et remplit la liste de destination avec I1. I1 utilise son LAN comme interface de destination.  Le switch qui forme le LAN du RP broadcaste le flux sur tous les ports (car il n’a pas reçu de message IGMP). I2 reçoit le flux. |
| 1 | Control plane | Le poste récepteur s’abonne au flux |
| 2 | Control Plane + Data Plane | Le switch intercepte la trame IGMP et place l’interface vers le poste récepteur comme interface de destination pour l’adresse Ethernet correspondante au groupe multicast.  Le routeur crée un état (\*, G) avec son LAN comme liste de destination, il envoie un message JOIN vers I2 (next-hop pour atteindre le RP). |
| 3 | Data plane | R2 ajoute l’interface vers D dans la liste de destination de (\*, G) et commence à forwarder le flux vers le récepteur. |
| 4 | Data plane | D utilise son état (\*,G) pour forwarder le flux sur le LAN. |
| 5 | Control Plane | Un second récepteur s’abonne auprès de S |
| 6 | Control Plane +  Data Plane | S ajoute une destination à son état (\*,G) et commence à diffuser le flux vers le nouveau récepteur |



# Comparaison

|  |  |  |  |
| --- | --- | --- | --- |
|  | **ASM** | **SSM** | **Bidir** |
| Nombre d’états | (\*,G) et (S,G) 🡪 Bcp | (S,G) 🡪 Bcp | (\*,G) 🡪 Peu |
| Chemin de Flux | Un peu de gaspillage si le RP est loin des sources | Aussi optimisé qu’en unicast | Pas de convergence |
| Stabilité | Le RP doit être secouru | - | Le RP doit toujours être routé |
| Ressources | Encapsulation entre le routeur source et le RP | - | - |
| Complexité | complexe | Simple | Simple |
| Troubleshooting | complexe | Simple | Simple |
| Load balancing | Oui, (par S,G) | Oui, (par S,G) | Sans |
|  |  |  |  |
|  |  |  |  |

# Quelques Astuces

## Secours du RP en ASM

Possibilité de déclarer le RP en dynamique (protocole BSR). Chaque RP s’annonce avec une priorité différente ou un range d’adresse de destination différent (avec overlapping).

Déclaration en Anycast : les RP possède la même adresse et c’est le routage unicast qui décide si un routeur s’adresse à RP1 ou RP2. Les RP s’échangent les sources actives entre eux par le protocole MSDP (IPv4 seulement).

## Non congruence

Il est possible de dissocier le routages des adresses sources unicast et multicast (par exemple pour utiliser un tunnel pour que les flux multicast puisse traverser une zone unicast). La vérification RPF utilise alors les tables multicast BGP (safi 2), le routage static spécifique au multicast puis le routage unicast classique.

## Des erreurs classiques

* L’adresse source pour l’encapsulation vers le RP n’est pas précisée (et non connue du RP) 🡪 le routeur ne reçoit jamais de register-stop
* Utilisation de routes multicast statiques : elles sont utilisées dans l’ordre de la configuration (pas selon la finesse du masque)