Comprendre le mécanisme d'affirmation PIM

Contenu

Introduction

Ce document décrit le mécanisme d'affirmation PIM (Protocol Independent Multicast), se concentre sur les critères de vainqueur de PIM et plonge plus en profondeur dans certains cas d'angle.

Conditions préalables

Conditions requises

Cisco recommande que vous connaissiez le mécanisme d'assertion PIM.

Components Used

Les informations de ce document sont basées sur Cisco CSR1000V version 16.4.1

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. Si votre réseau est en ligne, assurez-vous de bien comprendre l’incidence possible des commandes.

Qu'est-ce que le mécanisme d'affirmation PIM ?

Lorsque plusieurs routeurs PIM sont activés sur un segment partagé, il est possible que ces routeurs rencontrent un trafic de multidiffusion en double. Cela peut être le cas parce que deux routeurs ou plus sur le même segment partagé peuvent avoir une entrée valide (S, G) ou (*, G) qui remplit l'interface sortante vers le segment partagé pour le même groupe IP/destination source.

Le mécanisme d'assertion PIM est utilisé pour détecter et éliminer la duplication du trafic de multidiffusion sur un segment partagé. Il est important de noter que ce mécanisme n'empêche pas la duplication de se produire, au lieu de cela il utilise la duplication du trafic de multidiffusion comme déclencheur afin d'activer ce mécanisme qui sélectionne un seul redirecteur pour ce flux.

Lorsque vous avez une duplication du trafic de multidiffusion sur un segment partagé, vous pouvez supposer que plusieurs routeurs envoient la même (S, G) ou (*, G) sur un segment partagé. Si vous choisissez un routeur pour transférer efficacement ce flux, il élimine la duplication.

PIM utilise les messages d'assertion PIM qui sont déclenchés lorsque vous recevez un paquet de multidiffusion sur la liste OIL (Outgoing Interface List). Ces messages d'assertion contiennent des métriques qui sont ensuite utilisées pour calculer qui deviendra gagnant d'assertion. Les routeurs en aval du LAN reçoivent également des messages d'assertion PIM. Ces messages sont ensuite utilisés par les périphériques en aval pour envoyer les messages Join/Prune appropriés au routeur en amont qui a remporté l'élection d'assertion.

Scénario 1. Justification LHR

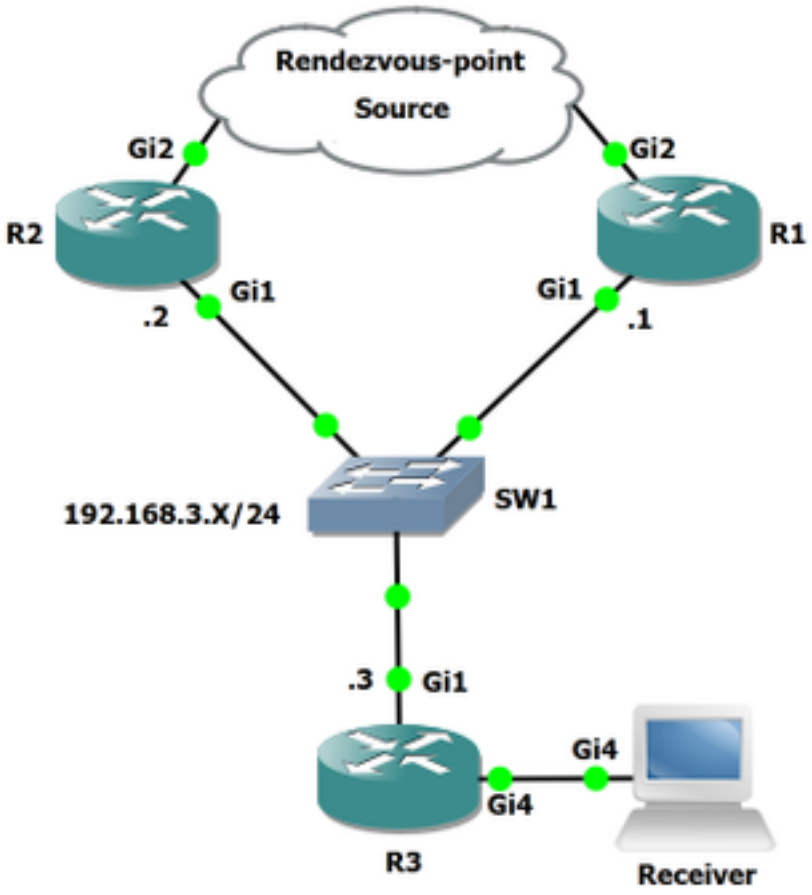


Figure 1.

Dans le schéma de réseau, R3 est le routeur de dernier saut (LHR), R3 se connecte à R2 et R1 via un segment partagé.

Lorsque vous recevez un rapport IGMP (Internet Group Management Protocol) du récepteur, R3 vérifie qui est le voisin RPF vers le RP. Dans la topologie, R1 est le voisin RPF vers le RP, d'où l'envoi par R3 d'une jointure (*, G) vers R1. Une fois que R1 descend le flux (supposons que le groupe est actif), R3 envoie une jointure (S, G) vers la source et descend l'arborescence source. R2 est le voisin RPF vers l'arborescence source, ce qui signifie que R3 enverra la jointure (S, G) vers R2. R3 a la même interface RPF vers le RP et la source. Vous pouvez voir ici la table mroute de R3 pour le groupe 239.1.1.1.

```
R3#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode
(*, 239.1.1.1), 00:00:55/stopped, RP 192.168.0.100, flags: SJC
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet4, Forward/Sparse, 00:00:55/00:02:04

(10.0.0.2, 239.1.1.1), 00:00:52/00:02:07, flags: JT
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.2, Mroute
  Outgoing interface list:
    GigabitEthernet4, Forward/Sparse, 00:00:52/00:02:07

(*, 224.0.1.40), 00:01:22/00:02:09, RP 192.168.0.100, flags: SJPCL
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
```

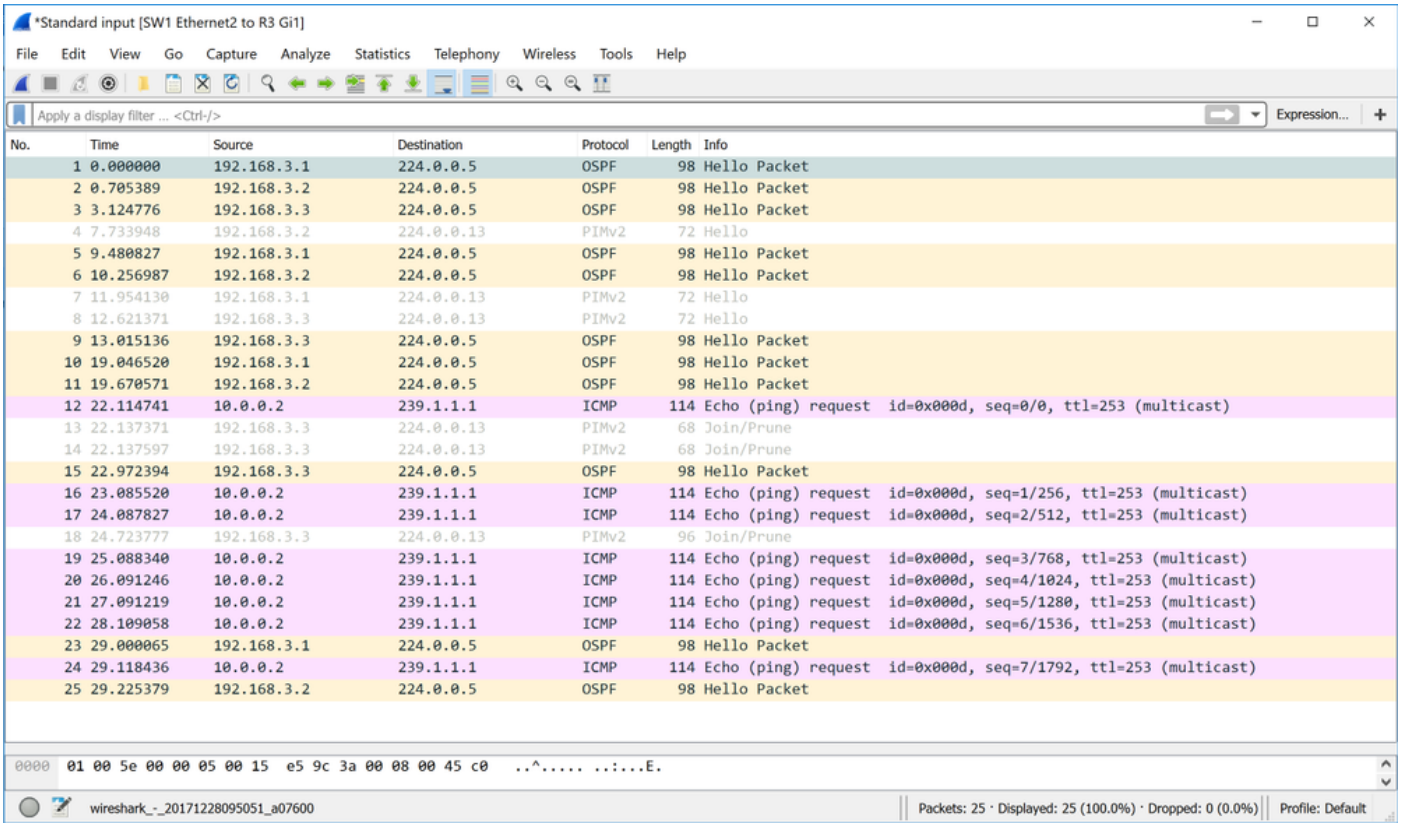
Comme vous pouvez le voir sur R3, le voisin RPF (*, G) est 192.168.3.1 et le voisin RPF vers (S, G) est 192.168.3.2. Maintenant, **cela devrait faire en sorte que R1 et R2 aient un OIL valide vers R3**. Examinons ces entrées :

```
R1#show ip mroute
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:15:02/00:02:33, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:15:02/00:02:33

(10.0.0.2, 239.1.1.1), 00:13:24/00:02:33, flags: PR
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list: Null

(*, 224.0.1.40), 00:29:17/00:02:51, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:16:06/00:02:51
  Outgoing interface list: Null
```

```
R2#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:08:00/stopped, RP 192.168.0.100, flags: SP
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null

(10.0.0.2, 239.1.1.1), 00:00:03/00:02:56, flags: T
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:00:03/00:03:26

(*, 224.0.1.40), 01:37:30/00:02:22, RP 192.168.0.100, flags: SJPL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
```

Comme mentionné précédemment, un assertion peut être déclenché lorsqu'il existe deux routeurs en amont dont l'OIL valide est renseigné sur un segment partagé. Puisque R1 et R2 ont tous deux une OIL valide, vérifiez s'il existe un mécanisme d'assertion dans la capture de paquets.

Cette capture de paquets a été capturée sur l'interface Gi1 de R3 vers SW1.



Dans cette capture de paquets, aucun paquet d'assertion ne s'affiche même si toutes les conditions préalables à la création de duplication sur le segment partagé entre R1, R2 et R3 sont réunies. Pourquoi ne voyez-vous aucun paquet d'assertion PIM lorsque le flux (S, G) a été activé ?

Il semble que le document RFC 7761 puisse répondre à ces questions.

Résumé de la RFC 7761 Section 4.2.2.

```
4.2.2.  Setting and Clearing the (S,G) SPTbit

   Basically, Update_SPTbit(S,G,iif) will set the SPTbit if we have the
   appropriate (S,G) join state, and if the packet arrived on the
   correct upstream interface for S, and if one or more of the following
   conditions apply:

   1.  The source is directly connected, in which case the switch to the
       SPT is a no-op.

   2.  The RPF interface to S is different from the RPF interface to the
       RP.  The packet arrived on RPF_interface(S), and so the SPT must
       have been completed.

   3.  No one wants the packet on the RP tree.
```

4. RPF'(S,G) == RPF'(*,G). In this case, the router will never be able to tell if the SPT has been completed, so it should just switch immediately. The RPF'(S,G) != NULL check ensures that the SPTbit is set only if the RPF neighbor towards S is valid.

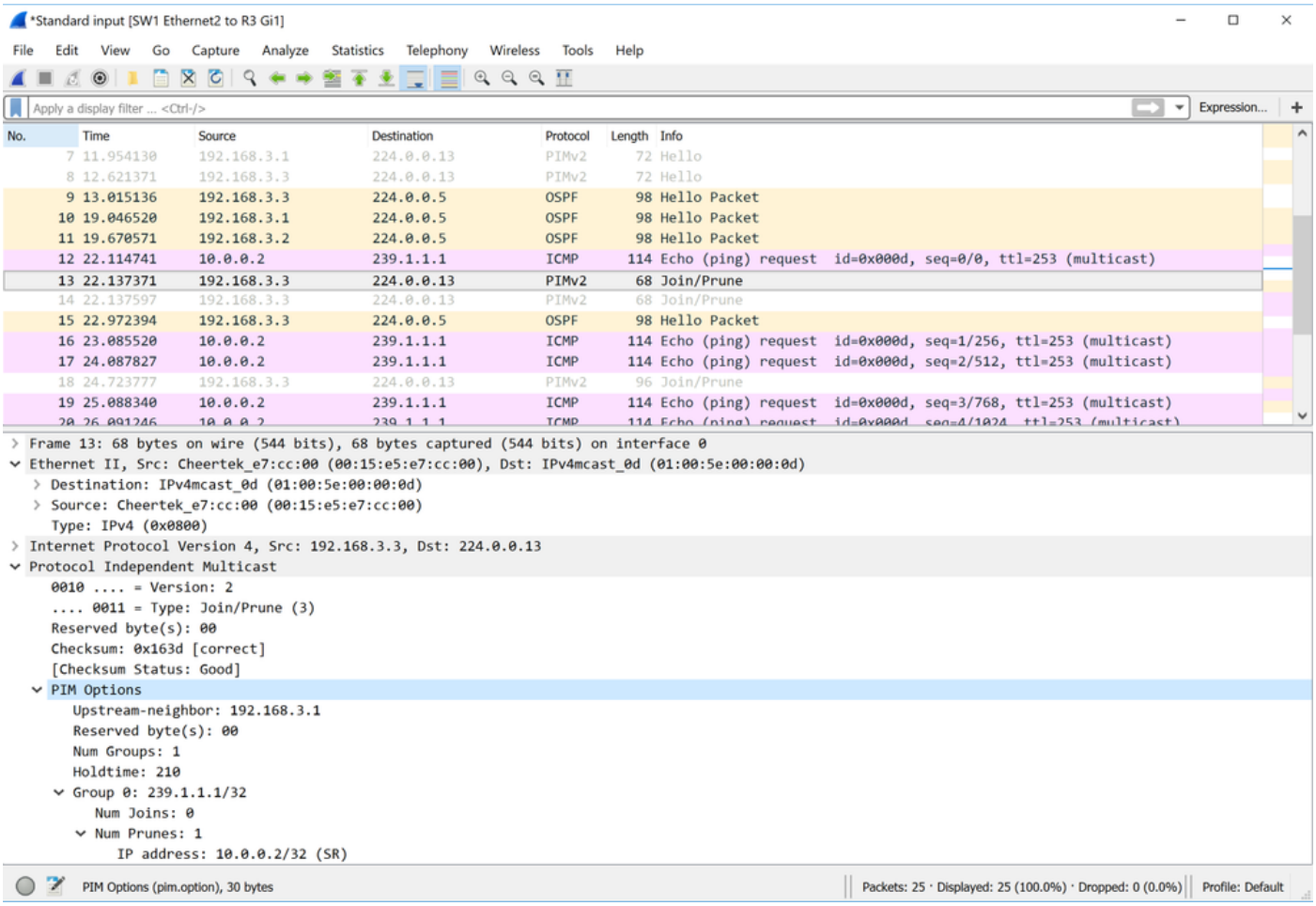
In the case where the RPF interface is the same for the RP and for S, but RPF'(S,G) and RPF'(*,G) differ, we wait for an Assert(S,G), which indicates that the upstream router with (S,G) state believes the SPT has been completed.

Le SPTbit (S, G) est utilisé pour déterminer si le transfert doit être effectué à l'état (*, G) ou à l'état (S, G). Lorsque vous passez de l'arborescence RP à l'arborescence source, il y a une période de transition lorsque les données arrivent en raison de l'état en amont (*, G) alors que l'état en amont (S, G) est établi, à ce moment-là, le routeur doit continuer à transmettre uniquement sur l'état (*, G). Cela empêche les trous noirs temporaires qui seraient provoqués par l'envoi d'une ou plusieurs pruneaux(S, G, rpt) avant que l'état en amont (S, G) ne soit établi.

Bien qu'il semble que le scénario puisse être corrélé avec le dernier point mentionné ci-dessus. Dans le cas où l'interface RPF est identique pour le RP et pour S,
mais RPF'(S, G) et RPF'(*, G) diffèrent, nous attendons un Assert(S, G), qui indique que le routeur en amont avec l'état (S, G) croit que le SPT est terminé.

Pour déclencher l'assertion, le routeur doit recevoir un paquet dupliqué sur son OIL déjà rempli pour le même groupe IP/destination source sur le segment. R3 est également un LHR, ce qui signifie qu'il est désigné pour passer de (*, G) à SPT (S, G) lorsqu'un paquet est reçu de (*, G).

Dans la capture de paquets, nous observons qu'aucune assertion n'est déclenchée. Bien que nous voyions une élingue envoyée immédiatement après la réception du premier écho ICMP.

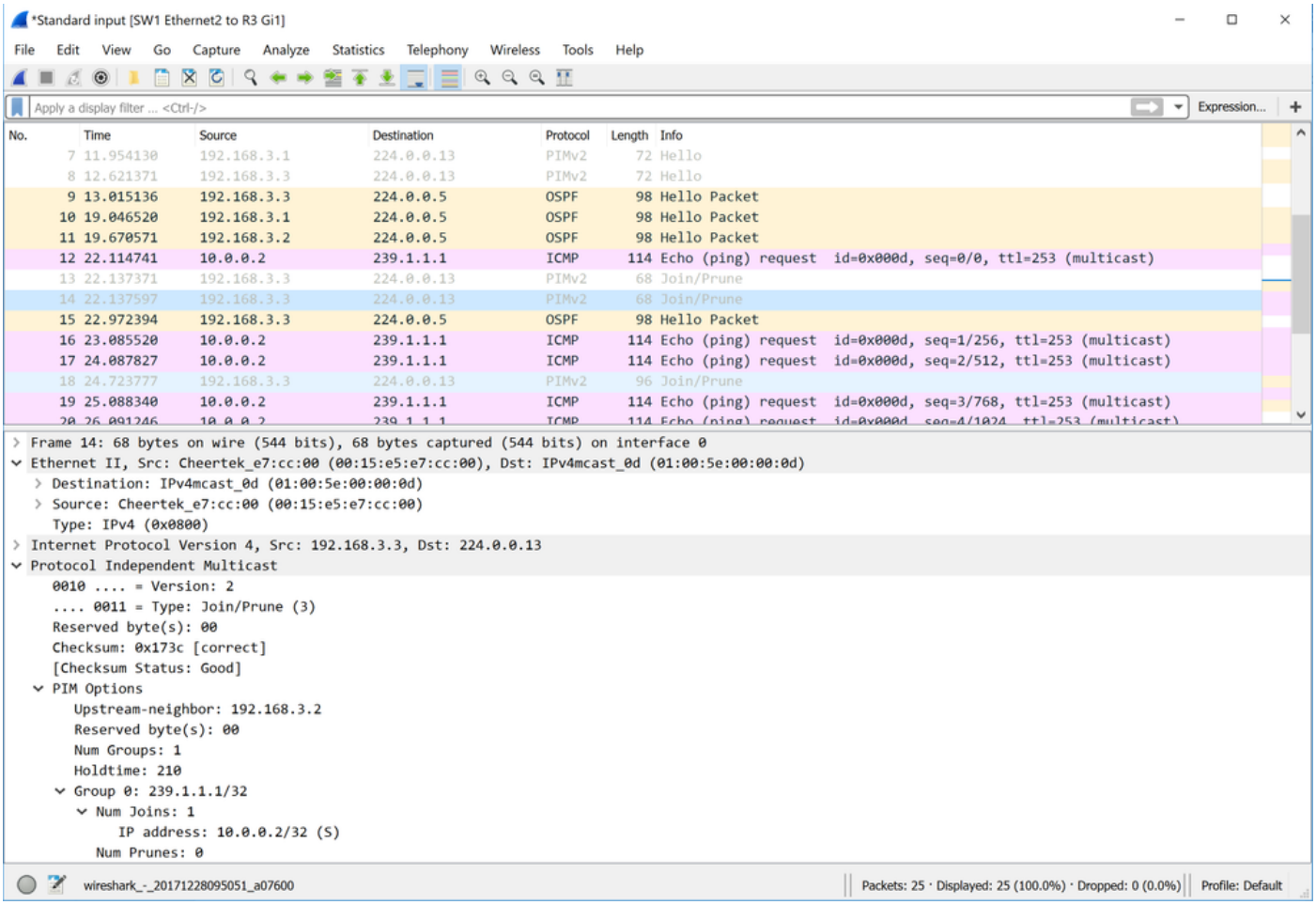


Comme vous pouvez le voir, une fois que le premier paquet de requête ICMP (Internet Control Message Protocol) est reçu sur l'interface G1 de R3, une valeur SR-bit (*, G) est envoyée vers le voisin en amont 192.168.3.1. Cette valeur est (*, G) pour la source spécifique définie.

Vous pouvez également voir ces indicateurs définis : (SR) :

```
The S flag: indicates that the multicast group is a sparse mode group.
The R flag: The R flag is the RP-bit flag and indicates that the information in the (S, G) entry
is applicable to the shared tree.
```

Dans le deuxième paquet PIM no 14, vous pouvez voir que R3 tente de rejoindre l'arborescence (S, G).



Il est observé qu'une fois le premier plan de données reçu, le paquet R3 exécute le (*, G) et crée le (S, G). C'est la raison pour laquelle vous ne voyez pas de paquets d'assertion PIM. Ce scénario est en vigueur lorsque vous avez un LHR qui a la même interface RPF pour (S, G) et (*, G). Bien que ce comportement puisse différer légèrement de celui de la RFC 7761, il ne doit pas causer de problèmes.

Maintenant, poursuivons avec le scénario 2. Le schéma de ce scénario peut être vu ici :

Scénario 2. Sélection du chemin d'assertion

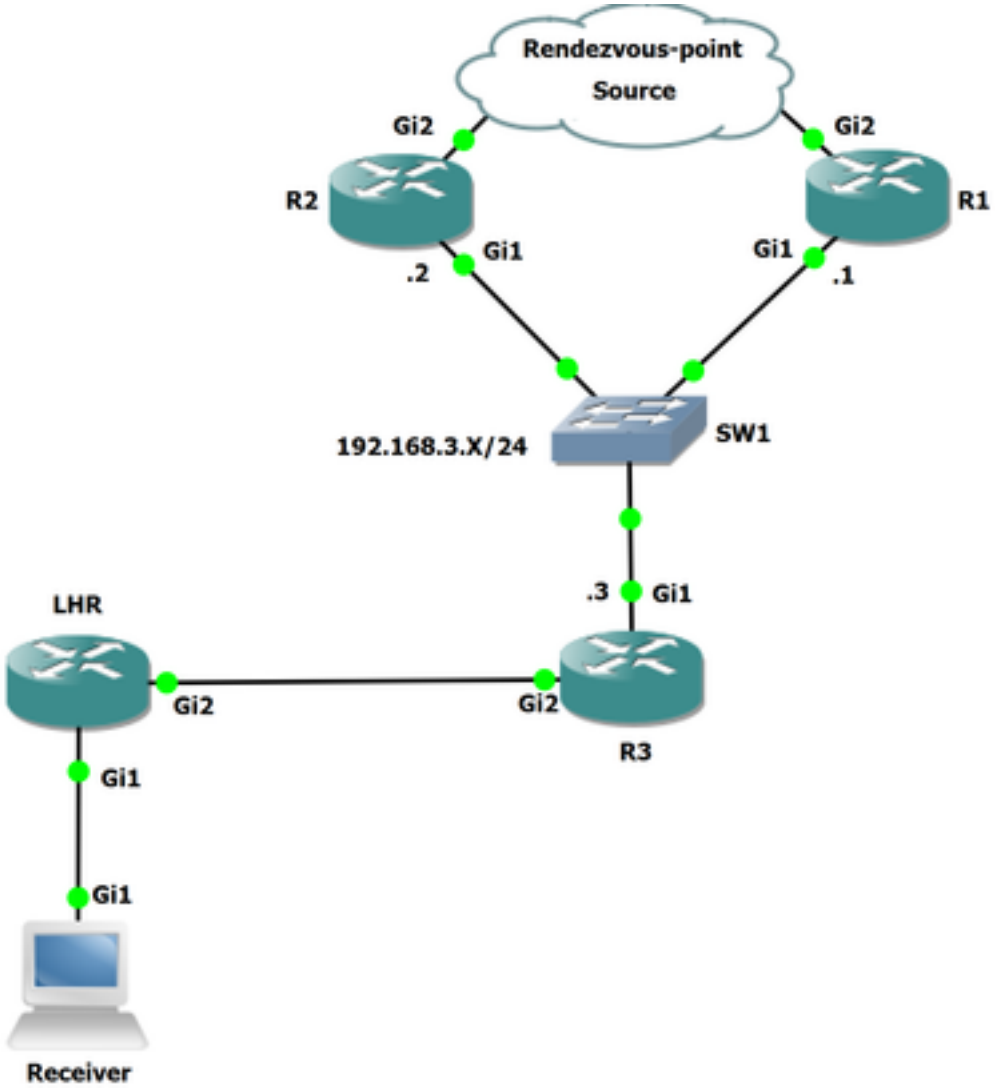


Figure 2.

Dans cette topologie, un autre routeur connecté à R3 est le LHR. Le LHR se connecte directement au récepteur. La source et le RP sont tous deux situés au-dessus de R2 et R1. Le voisin RPF sur R3 en direction du RP est R1 et le voisin RPF en direction de la source est R2.

Vérifions le voisin RPF pour la source et le RP.

Ici, vous voyez le voisin RPF vers le RP : 192.168.0.100 est 192.168.3.1.

```
R3#show ip rpf 192.168.0.100
RPF information for ? (192.168.0.100)
  RPF interface: GigabitEthernet1
  RPF neighbor: ? (192.168.3.1)
  RPF route/mask: 192.168.0.100/32
  RPF type: unicast (ospf 1)
  Doing distance-preferred lookups across tables
  RPF topology: ipv4 multicast base, originated from ipv4 unicast base
```

Ici, vous voyez le voisin RPF vers la source : 10.0.0.2 est 192.168.3.2.

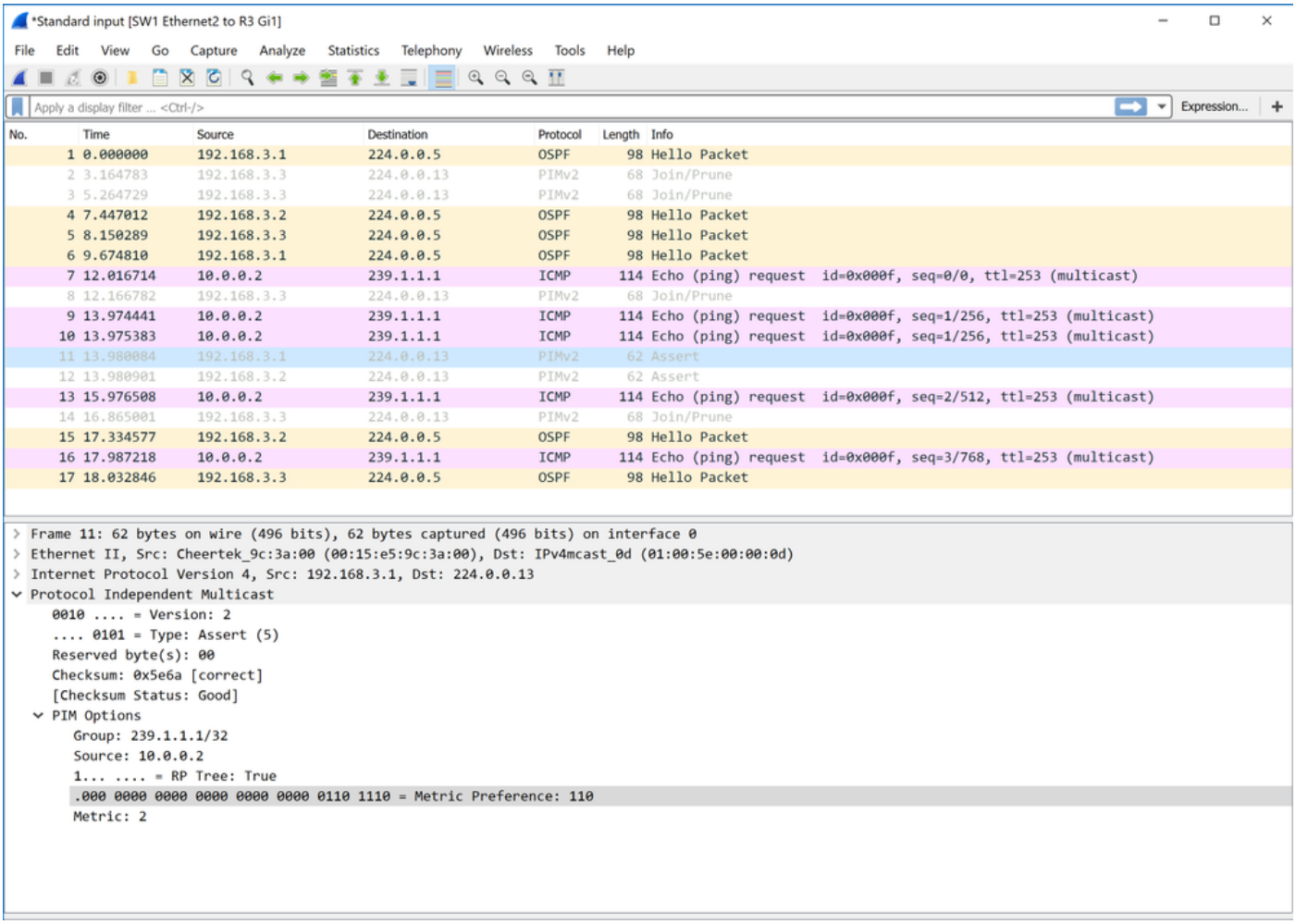
```
R3#show ip rpf 10.0.0.2
RPF information for ? (10.0.0.2)
  RPF interface: GigabitEthernet1
```

```
  RPF neighbor: ? (192.168.3.2)
  RPF route/mask: 10.0.0.0/24
  RPF type: unicast (ospf 1)
  Doing distance-preferred lookups across tables
  RPF topology: ipv4 multicast base, originated from ipv4 unicast base
```

Avant d'activer la source, jetons un coup d'oeil à la table mroute sur R3, car vous pouvez voir qu'il existe déjà (*, G) pour le groupe 239.1.1.1. En effet, le récepteur connecté à LHR a déjà demandé pour le groupe spécifié.

```
R3#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:00:57/00:02:32, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet2, Forward/Sparse, 00:00:57/00:02:32

(*, 224.0.1.40), 00:11:24/00:02:41, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet2, Forward/Sparse, 00:02:02/00:02:41
```

Maintenant, activez la source et capturez les paquets sur l'interface Gi1 de R3.



Comme vous pouvez le voir dans cette capture de paquets, PIM affirme que les paquets sont déjà présents.

Trame 11 :

```
> Frame 11: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
> Ethernet II, Src: Cheertek_9c:3a:00 (00:15:e5:9c:3a:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
> Internet Protocol Version 4, Src: 192.168.3.1, Dst: 224.0.0.13
v Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0101 = Type: Assert (5)
    Reserved byte(s): 00
    Checksum: 0x5e6a [correct]
    [Checksum Status: Good]
  v PIM Options
      Group: 239.1.1.1/32
      Source: 10.0.0.2
      1... .... = RP Tree: True
      .000 0000 0000 0000 0000 0000 0110 1110 = Metric Preference: 110
      Metric: 2
```

Trame 12 :

```
> Frame 12: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
> Ethernet II, Src: Cheertek_8b:3e:00 (00:15:e5:8b:3e:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
> Internet Protocol Version 4, Src: 192.168.3.2, Dst: 224.0.0.13
v Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0101 = Type: Assert (5)
    Reserved byte(s): 00
    Checksum: 0xde6a [correct]
    [Checksum Status: Good]
  v PIM Options
      Group: 239.1.1.1/32
      Source: 10.0.0.2
      0... .... = RP Tree: False
      .000 0000 0000 0000 0000 0000 0110 1110 = Metric Preference: 110
      Metric: 2
```

Lorsque vous examinez ces paquets, vous devriez être en mesure de déterminer qui est le gagnant de l'assertion. Maintenant, jetons un coup d'oeil à la sélection du redirecteur d'assertion PIM.

La préférence de métrique est la distance administrative (AD). Il s’agit de la distance administrative du protocole de routage qui installe la route dans la table de routage, utilisée pour rechercher l’adresse IP source et la métrique est le coût de la route.

D'autres attributs sont également utilisés pour déterminer qui est le gagnant de l'affirmation. Vous pouvez voir ces détails dans le document RFC 7761.

Résumé de la section 4.6.3 de la RFC 7761.

```
4.6.3.  Assert Metrics

   Assert metrics are defined as:

     struct assert_metric {
```

```
       rpt_bit_flag;
       metric_preference;
       route_metric;
       ip_address;
     };
   When comparing assert_metrics, the rpt_bit_flag, metric_preference,
   and route_metric fields are compared in order, where the first lower
   value wins.  If all fields are equal, the primary IP address of the
   router that sourced the Assert message is used as a tie-breaker, with
   the highest IP address winning.
```

Avec l'utilisation de ces champs définis et la sélection du chemin, vous pouvez déterminer qui sera le gagnant de l'assertion dans ce scénario. Si vous examinez à nouveau les paquets assertion, vous pouvez voir que la préférence de métrique n'est pas comparée puisque la décision est prise sur le tout premier critère de sélection qui est rpt_bit_flag.

Dans ce scénario, la comparaison entre R1 et R2 est effectuée. Les deux routeurs envoient des messages d'assertion qui ont été vus précédemment et une fois que les deux périphériques voient les messages d'assertion les uns des autres, ils peuvent comparer les métriques entre eux afin de déterminer qui est le gagnant.

Puisque R2 envoie un message d'assertion avec l'arborescence RP : False qui a la valeur 0, elle est en effet inférieure à ce que R1 a envoyé avec une arborescence RP : True dont la valeur est 1. Le bit de l'arborescence RP est défini sur 0 ou 1.

Le bit d'arborescence RP lorsque défini sur 1 signifie que vous êtes actuellement sur l'arborescence partagée ; le bit RPT effacé indique que l'expéditeur de l'assertion avait l'état de transmission (S, G) sur une interface.

Comme (S, G) affirme avoir la priorité sur (*, G) affirme, R2 doit être le gagnant de l'affirmation. Transition vers l'état « Je suis un vainqueur assertif ». Comme indiqué dans l'instruction précédente de la RFC 7761, la valeur la plus faible est plus privilégiée.

Regardons à la fois R1 et R2 afin de voir qui est le gagnant de l'affirmation.

```
R2#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:42:52/stopped, RP 192.168.0.100, flags: SP
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null

(10.0.0.2, 239.1.1.1), 00:42:52/00:01:40, flags: T
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:42:52/00:03:07, A

(*, 224.0.1.40), 00:43:23/00:02:25, RP 192.168.0.100, flags: SJPL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null
```

Dans cette sortie, vous pouvez voir que l'indicateur A (S, G) de R2 est défini sur l'OIL, ce qui indique qu'il s'agit du gagnant de l'assertion. Ici sur R1, vous n'avez pas d'OIL sur le (S, G) et

l'indicateur P est défini, ce qui signifie que le particulier (S, G) a été élagué dans ce cas : ce n'est pas le vainqueur de l'affirmation.

> **Note**: Lorsque le paramètre assert est présent sur un segment partagé, les voisins en aval envoient des messages périodiques Join(*, G) et Join(S, G) au voisin RPF approprié, c'est-à-dire au voisin RPF tel que modifié par le processus assert. Ils ne sont pas toujours envoyés au voisin RPF comme indiqué par la MRIB.

```
R1#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:44:32/00:03:09, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:44:32/00:03:09, A

(10.0.0.2, 239.1.1.1), 00:44:19/00:03:09, flags: PR
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list: Null

(*, 224.0.1.40), 00:44:50/00:02:53, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:43:56/00:02:53
```

Si R1 et R2 ont tous deux un bit d'arborescence RP défini sur 1. vous pouvez ensuite considérer le routeur avec la distance administrative la plus faible ; si égale, examinez la métrique. Si le bit d'arborescence RP est vrai sur les deux routeurs, la métrique est comparée à l'adresse IP RP. Si le bit de l'arborescence RP est 0, la métrique est comparée à la source du flux de multidiffusion.

Si toutes ces valeurs sont identiques, le message d'assertion d'origine d'adresses IP le plus élevé est le gagnant.

Résumé

Dans le scénario 1, vous n'avez pas observé de paquets assertion, cependant, par RFC, ils auraient dû être déclenchés. Comme nous l'avons mentionné, c'est parce que R3 était en élagage (*, G) avant la construction du plan de contrôle pour (S, G).

Dans le deuxième scénario, vous voyez des paquets assertion. Lorsque le premier paquet a été reçu sur LHR, il envoie une (S, G) jointure/élingue vers R3 pour extraire la source/le groupe. R3 envoie alors un paquet de jointure/élingue vers R2 pour la même source/le même groupe. Ainsi, R1 et R2 disposeraient d'OIL valides. Maintenant, R3 effectue uniquement des pruneaux (S, G) avec le bit RP défini lorsque l'indicateur T est rempli à l'état R3 (S, G). Pour ce faire, vous devez recevoir un autre paquet de plan de données du segment partagé. Comme le plan de contrôle a déjà été construit pour (S, G), cela entraîne une duplication sur le segment partagé déclenchant des messages d'assertion.