Groupe de travail Réseau
Request for Comments : 3446
Catégorie : Information
Traduction Claude Brière de L'Isle

D. Kim, Verio
D. Meyer, H. Kilmer & D. Farinacci
Procket Networks
janvier 2003

Mécanisme de point de rendez-vous (RP) en envoi à la cantonade utilisant la diffusion groupée indépendante du protocole (PIM) et le protocole de découverte de source de diffusion groupée (MSDP)

Résumé
Le présent document décrit un mécanisme pour permettre un nombre arbitraire de points de rendez-vous (RP, *Rendevous Point*) par groupe dans un seul domaine en mode épars de diffusion groupée indépendante du protocole (PIM-SM, *Protocol Independent Multicast-Sparse Mode*) à arborescence partagée.

1. Introduction

PIM-SM, comme défini dans la RFC 2362, permet un seul RP actif par groupe, et à ce titre, la décision du placement optimal du RP peut devenir problématique pour un réseau multi régional qui déploie PIM-SM.

Le RP à la cantonade supprime une importante contrainte de PIM-SM, à savoir qu'il ne peut y avoir qu'une seule transposition de groupe en RP active à un moment donné. La propriété d'une seule transposition a plusieurs implications, incluant de concentration du trafic, de manque d'adaptabilité de désencapsulation d'enregistrement (lorsque on utilise l'arborescence partagée) une lente convergence lorsque un RP actif a une défaillance, une éventuelle transmission sous optimale des paquets en diffusion groupée, et les dépendances aux RP distants. Ces propriétés de PIM-SM ont été démontrées dans les déploiements natifs de diffusion groupée à l'échelle continentale ou intercontinentale. Par suite, il est clair que les cœurs de réseau des fournisseurs d'accès Internet (FAI) exigent un mécanisme qui permette la définition de plusieurs RP actifs par groupe dans un seul domaine PIM-SM. De plus, un tel mécanisme devrait aussi traiter les problèmes évoqués ci-dessus.

Le mécanisme décrit ici est destiné à régler le besoin d'une meilleure récupération (temps de convergence) et d'un meilleur partage de la charge de désencapsulation d'enregistrement (là encore, lorsque on utilisé l'arborescence partagée) entre les RP dans un domaine. Il est principalement destiné aux applications dans les réseaux qui utilisent le protocole de routeur frontière multi protocoles (MBGP, *Multiprotocol Border Gateway Protocol*) le protocole de découverte de source de diffusion groupée (MSDP, *Multicast Source Discovery Protocol*) [RFC3618] et les protocoles PIM-SM, pour le déploiement natif de diffusion groupée, bien qu'il ne soit pas limité à ces protocoles. En particulier, le RP à la cantonade est applicable dans tout réseau PIM-SM qui prend aussi en charge MSDP (MSDP est nécessaire pour que les divers RP dans le domaine conservent une vue cohérente des sources qui sont actives). Noter cependant qu'un domaine qui déploie des RP à la cantonade n'est pas obligé de fonctionner avec MBGP. Finalement, une exigence générale du schéma de RP à la cantonade est que l'adresse d'envoi à la cantonade NE DOIT PAS être utilisée comme adresse de RP dans les messages SA du RP.

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

2. Définition du problème

La solution du RP à la cantonade traite les deux problèmes de la récupération rapide et de l'équilibrage de charge de l'arborescence partagée entre tout nombre de RP actifs dans un domaine.

2.1 Concentration du trafic et distribution de la charge de désencapsulation entre les RP

Alors que PIM-SM permet que plusieurs RP soient définis pour un certain groupe, une seule transposition de groupe en RP peut être active à un moment donné. Un mécanisme traditionnel de déploiement pour équilibrer la charge de désencapsulation d'enregistrement entre plusieurs RP couvrant l'espace du groupe de diffusion groupée est de partager l'espace 224.0.0.0/4 entre plusieurs RP définis. C'est une solution acceptable tant que le trafic de diffusion groupée reste faible, mais pose des problèmes lorsque le trafic en diffusion groupée augmente, en particulier parce que l'opérateur de réseau qui définit l'espace de groupe partagé entre les RP n'a pas toujours connaissance a priori de la distribution du trafic entre les groupes. Cela peut être surmonté par des reconfigurations périodiques, mais des considérations de fonctionnement font que ce type de solution est peu adaptable.

2.2 Transmission sous optimale des paquets de diffusion groupée

Lorsque un seul RP dessert un certain groupe de diffusion groupée, toutes les jonctions à ce groupe seront envoyées à ce RP sans considération de la distance topologique entre le RP et les sources et receveurs. Les données initiales seront envoyées aussi au RP jusqu'à ce que soit atteint le seuil de commutation de l'arborescence de plus court chemin configuré, ou les données seront toujours envoyées au RP si le réseau est configuré à toujours utiliser l'arborescence partagée dont la racine est le RP. Cela reste vrai même si toutes les sources et receveurs sont dans une seule région, et si le RP est topologiquement distant des sources et des receveurs. C'est un artifice de la nature dynamique des membres d'un groupe de diffusion groupée, du fait que les opérateurs ne peuvent pas toujours avoir connaissance a priori du placement topologique des membres du groupe.

Pris ensemble, ces effets peuvent signifier (par exemple) que bien que toutes les sources et receveurs d'un certain groupe soient en Europe, ils se joignent au RP aux USA et les données vont traverser deux fois un tuyau relativement coûteux, une fois pour aller au RP, et de retour à nouveau à l'arborescence enracinée au RP, créant un usage inefficace de ressources coûteuses.

2.3 Dépendance au RP distant

Comme mentionné ci-dessus, un seul RP actif par groupe peut être cause que les sources et receveurs locaux deviennent dépendants d'un RP topologiquement distant. De plus, lorsque plusieurs RP sont configurés, il peut y avoir un délai de convergence considérable impliqué par la commutation au RP de secours. Ce délai peut exister indépendamment de la localisation topologique des RP principaux et de secours.

3. Solution

Compte tenu du problème présenté ci-dessus, une bonne solution serait de permettre à un opérateur de configurer plusieurs RP par groupe, et de répartir ces RP d'une manière significative du point de vue topologique entre les sources et receveurs.

3.1 Mécanismes

Tous les RP qui desservent un certain groupe ou ensemble de groupes sont configurés avec une adresse identique d'envoi à la cantonade, en utilisant une interface numérotée des RP (fréquemment une interface logique comme de rebouclage). Les RP annoncent alors la transposition de groupe en RP en utilisant cette adresse d'interface. Cela va causer la jonction (enregistrement) des membres du groupe (envoyeurs) au RP le plus proche topologiquement. Les RP homologues MSDP les uns avec les autres utilisent une adresse unique à chaque RP. Comme l'adresse d'envoi à la cantonade n'est pas une adresse unique (par définition) un routeur NE DOIT PAS choisir l'adresse d'envoi individuel à la cantonade comme identifiant de routeur, car cela peut empêcher les relations d'homologue à homologue et/ou l'établissement des adjacences.

En résumé, les étapes suivantes sont alors nécessaires :

3.1.1 Créer l'ensemble des transpositions de groupe à adresse de RP à la cantonade

La première étape est de créer l'ensemble de transpositions de groupe à adresse de RP en envoi à la cantonade à utiliser

dans le domaine. Chaque RP qui participe à un ensemble de RP en envoi à la cantonade doit être configuré avec un ensemble cohérent de transpositions de groupe en adresse de RP. Cette transposition sera utilisée par les routeurs non RP dans le domaine.

3.1.2 Configurer chaque RP pour la gamme de groupe avec l'adresse de RP à la cantonade

L'étape suivante est de configurer chaque RP pour la gamme de groupes avec l'adresse de RP en envoi à la cantonade. Si un mécanisme dynamique, comme un mécanisme d'auto-RP ou le mécanisme bootstrap de PIMv2, est utilisé pour annoncer les transpositions de groupe à RP, l'adresse IP d'envoi à la cantonade devrait être utilisée pour l'adresse de RP.

3.1.3 Configurer les relations d'homologue MSDP entre chaque RP en envoi à la cantonade de l'ensemble

À la différence des annonces de transposition de groupe en RP, les relations d'homologues MSDP doivent utiliser une adresse IP qui soit unique pour les points d'extrémité ; c'est-à-dire, les points d'extrémité de relations d'homologue à homologue MSDP DOIVENT utiliser une adresse d'envoi individuel plutôt qu'une adresse d'envoi à la cantonade. Une ligne directrice générale serait de suivre l'adressage des relations d'homologue à homologue de BGP, par exemple, les rebouclages pour les relations d'homologue à homologue iBGP, les adresses d'interface physique pour les relations d'homologue à homologue pour eBGP. Noter que l'adresse d'envoi à la cantonade NE DOIT PAS être utilisée comme l'adresse de RP dans les messages SA (car cela ferait échouer la vérification de la transmission sur chemin inverse (RPF) de l'homologue).

3.1.4 Configurer les non RP avec les transpositions de groupe à adresse de RP en envoi à la cantonade

Finalement, chaque routeur non RP doit prendre connaissance de l'ensemble de transpositions de groupe en RP. Cela peut se faire via une configuration statique, auto-RP, ou par le mécanisme bootstrap de PIMv2.

3.1.5 S'assurer que l'adresse IP d'envoi à la cantonade est accessible par tous les routeurs dans le domaine

Ceci est normalement réalisé en faisant que chaque RP injecte le /32 dans l'IGP du domaines .

3.2 Interaction avec la vérification du RPF d'homologue MSDP

Chaque homologue MSDP reçoit et transmet le message à partir de l'adresse de RP dans un "arrosage en transmission sur le chemin inverse à l'homologue". La notion d'arrosage en transmission sur le chemin inverse à l'homologue est relative à la transmission des messages SA [RFC3618]. Les tableaux d'acheminement de BGP sont examinés pour déterminer quel homologue est le prochain bond vers le RP générateur du message SA. Un tel homologue est appelé un "homologue RPF". Voir les détails dans la [RFC3618] sur la vérification de la transmission sur le chemin inverse à l'homologue.

3.3 Implications sur l'état

On notera qu'utiliser MSDP de cette façon force la création de l'état (S,G) le long du chemin du receveur à la source. Cet état peut n'être pas présent si un seul RP a été utilisé et si les receveurs ont été forcés de rester sur l'arborescence partagée.

4. Considérations sur la sécurité

Comme la solution décrite ici fait une grosse utilisation de l'adressage à la cantonade, il faut veiller à éviter les usurpations d'identité. En particulier l'acheminement en envoi individuel et les RP PIM doivent être protégés.

4.1 Acheminement en envoi individuel

L'acheminement en envoi individuel interne aussi bien qu'externe peut être protégé de façon faible par MD5 chiffré [RFC1828], comme mis en œuvre dans un protocole interne comme OSPF [RFC2328] ou dans BGP [RFC2385]. Plus généralement, IPsec [RFC2401] pourrait être utilisé pour assurer l'intégrité du protocole pour le système d'acheminement en envoi individuel.

4.1.1 Effets de l'instabilité de l'acheminement en envoi individuel

Bien que ce ne soit pas un problème de sécurité, on notera que si l'acheminement en envoi individuel n'est pas stable, le RP réel que la source ou le receveur utilise sera soumis à la même instabilité.

4.2 Intégrité du protocole de diffusion groupée

Les mécanismes décrits dans la [RFC2362] devraient être utilisés pour assurer la protection de l'intégrité des messages du protocole et l'authentification de l'origine de message au niveau du groupe.

4.3 Intégrité de l'homologue MSDP

Comme c'est le cas pour BGP, les homologues MSDP peuvent être protégés en utilisant MD5 chiffré [RFC1828].

5. Remerciements

John Meylor, Bill Fenner, Dave Thaler et Tom Pusateri ont fourni des commentaires pénétrants sur des versions antérieures de cette idée.

Le présent mémoire a été produit par le groupe de travail de déploiement MBONE (MBONED) dans la zone Opérations et gestion de l'équipe d'ingénierie de l'Internet (IETF, *Internet Engineering Task Force*). Envoyer vos commentaires à <mboned@ns.uoregon.edu> ou aux auteurs.

6. Références

[RFC1828] P. Metzger et W. Simpson, "Authentification IP avec du MD5 à clés", août 1995. *(Historique)*

[RFC2119] S. Bradner, "Mots clés à utiliser dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997.

[RFC2328] J. Moy, "OSPF version 2", STD 54, avril 1998. *(MàJ par la* RFC6549)

[RFC2362] D. Estrin et autres, "Mode épars de diffusion groupée indépendante du protocole (PIM-SM) : Spécification du protocole", juin 1998. *(Obsolète, voir* RFC4601, RFC5059)

[RFC2385] A. Heffernan, "Protection des sessions de BGP via l'option de signature MD5 de TCP", août 1998. *(P.S.* (*MàJ par la* RFC6691) *(Remplacée par* RFC5925)

[RFC2401] S. Kent et R. Atkinson, "Architecture de sécurité pour le protocole Internet", novembre 1998. (*Obsolète, voir* RFC4301)

[RFC2403] C. Madson, R. Glenn, "Utilisation de HMAC-MD5-96 au sein d'ESP et d'AH", novembre 1998. *(P.S.)*

[RFC3618] B. Fenner et D. Meyer, éd., "Protocole de découverte de source de diffusion groupée (MSDP)", octobre 2003. *(Exp.)*

7. Adresses des auteurs

Dorian Kim
Verio, Inc.
mél : dorian@blackrose.org

Hank Kilmer
mél : hank@rem.com

Dino Farinacci
Procket Networks
mél : dino@procket.com

David Meyer
mél : dmm@maoz.com

8. Déclaration complète de droits de reproduction

Remerciement
Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.