Groupe de travail Réseau
Request for Comments : 1027
Traduction Claude Brière de L'Isle

Smoot Carl-Mitchell, Texas Internet Consulting
John S. Quarterman, Texas Internet Consulting
octobre 1987

Utilisation d'ARP pour mettre en œuvre des passerelles de sous réseau transparentes

Statut de ce mémoire
La présente RFC décrit l'utilisation du protocole de résolution d'adresse (ARP, *Address Resolution Protocol*) Ethernet par les passerelles de sous réseau pour permettre aux hôtes sur les sous réseaux connectés de communiquer sans connaître l'existence des sous réseaux, en utilisant la technique du "mandataire ARP" [RFC1009]. Elle se fonde sur les [RFC0950], [RFC0922], et [RFC0826] et est un sous ensemble restreint du mécanisme de la [RFC0925]. La distribution du présent mémoire n'est soumis à aucune restriction.

Remerciements
Le travail décrit dans le présent mémoire a été effectué alors que les auteurs étaient employés par le département d'informatique de l'Université du Texas à Austin.

Introduction
L'objet de ce mémoire est de décrire en détails la mise en œuvre de passerelles de sous réseau ARP transparentes en utilisant la technique du mandataire ARP. L'intention est de documenter cette technique largement utilisée.

1. Motivation

L'Ethernet de l'Université du Texas à Austin est une grande installation connectant une dizaine de bâtiments. Il y a actuellement plus de cent hôtes qui lui sont connectés [Nameservers]. La taille de l'Ethernet et la quantité de trafic qu'il traite interdit de les lier ensemble en utilisant des répéteurs. L'utilisation de sous réseaux fournit une solution de remplacement intéressante pour diviser le réseau en plus petites unités distinctes.

Ceci est exactement la situation pour laquelle sont destinés les sous réseaux Internet décrits dans la RFC0950. Malheureusement, de nombreux fabricants n'ont pas encore mis en œuvre de sous réseaux, et il n'était pas pratique de modifier la demie douzaine de systèmes d'exploitation différents qui fonctionnent sur les hôtes des réseaux locaux.

Donc, une méthode pour cacher l'existence de sous réseaux aux hôtes était très souhaitable. Comme tous les réseaux de zone locale prenaient en charge ARP, une méthode fondée sur ARP (couramment appelée "mandataire ARP" ou le "hack ARP") a été choisie. Dans le présent mémoire, chaque fois que le terme "sous réseau" est utilisé on sous entend la "méthode de sous réseau de la RFC0950".

2. Concept

2.1 Méthode de base

Sur un réseau qui prend en charge ARP, quand l'hôte A (la source) diffuse une demande ARP pour l'adresse réseau correspondant à l'adresse IP de l'hôte B (la cible) l'hôte B va reconnaître l'adresse IP comme la sienne et va envoyer une réponse ARP en point à point. L'hôte A garde la transposition de IP en adresse de réseau trouvée dans la réponse dans une antémémoire locale et l'utilise pour les communications ultérieures avec l'hôte B.

Si les hôtes A et B sont sur des réseaux physiques différents, l'hôte B ne va pas recevoir la demande de diffusion ARP provenant de l'hôte A et ne peut pas y répondre. Cependant, si le réseau physique de l'hôte A est connecté par une passerelle au réseau physique de l'hôte B, la passerelle va voir la demande ARP provenant de l'hôte A. En supposant que les numéros de sous réseau sont faits pour correspondre aux réseaux physiques, la passerelle peut aussi dire que la demande est pour un hôte qui est sur un réseau physique différent de celui de l'hôte demandeur. La passerelle peut alors répondre pour l'hôte B, en disant que l'adresse réseau pour l'hôte B est celle de la passerelle elle-même. L'hôte A va voir cette réponse, la mettre en antémémoire, et envoyer les futurs paquets IP pour l'hôte B à la passerelle. La passerelle va transmettre ces

paquets à l'hôte B par les mécanismes d'acheminement IP usuels. La passerelle agit comme un agent pour l'hôte B, et c'est pourquoi cette technique est appelée "mandataire ARP" ; on va se référer à cela comme à une passerelle transparente de sous réseau ou passerelle de sous réseau ARP.

Quand hôte B réplique au trafic provenant de l'hôte A, le même algorithme s'applique en sens inverse : la passerelle connectée au réseau de l'hôte B répond à la demande pour l'adresse réseau de l'hôte A, et l'hôte B envoie alors des paquets IP pour l'hôte A à la passerelle. Les réseaux physique des hôtes A et B n'ont pas besoin d'être connectés à la même passerelle. Tout ce qui est nécessaire est que les réseaux soient accessibles à partir de la passerelle.

Avec cette approche, tout le traitement de sous réseau ARP est fait dans les passerelles de sous réseau ARP. Aucun changement au protocole ou acheminement ARP normal n'a besoin d'être fait aux hôtes source et cible. Du point du vue de l'hôte, il n'y a pas de sous réseaux, et leurs réseaux physiques sont simplement un gros réseau IP. Si un hôte a une mise en œuvre de sous réseaux, ses gabarits de réseau doivent être réglés à couvrir seulement le numéro de réseau IP, excluant les bits de sous réseau, pour que le système fonctionne correctement.

2.2 Acheminement

Au titre de la mise en œuvre de sous réseaux, il est prévu que les éléments des tableaux d'acheminement vont inclure des numéros de réseau comportant à la fois le numéro de réseau IP et les bits de sous réseau, comme spécifié par le gabarit de sous réseau, quand c'est approprié. Quand une demande ARP est vue, la passerelle de sous réseau ARP peut déterminer si elle connaît un chemin pour l'hôte cible en cherchant dans le tableau d'acheminement ordinaire. Si les tentatives d'accès aux réseaux IP étrangers sont éliminées précocement (voir les vérifications de bonne santé ci-dessous) seule une demande pour une adresse sur le réseau IP local va atteindre ce point. On va supposer que le même gabarit de réseau s'applique à tout sous réseau du même réseau IP. Le gabarit de réseau de l'interface réseau sur laquelle la demande ARP est arrivée peut alors être appliquée à l'adresse IP cible pour produire la partie réseau à rechercher dans le tableau d'acheminement.

Dans 4.3BSD (et probablement dans d'autres systèmes d'exploitation) un chemin par défaut est possible. Ce chemin par défaut spécifie une adresse où transmettre un paquet quand aucun autre chemin n'est trouvé. Le chemin par défaut ne doit pas être utilisé quand on vérifie un chemin pour l'hôte cible d'une demande ARP. Si le chemin par défaut était utilisé, la vérification réussirait toujours. Mais l'hôte spécifié par le chemin par défaut a peu de chances de connaître un acheminement de sous réseau (car c'est généralement une passerelle Internet) et donc les paquets qui lui sont envoyés vont probablement être perdus. Ce cas particulier de la méthode de recherche d'acheminement est le seul changement de mise en œuvre nécessaire au mécanisme d'acheminement.

Si les interfaces réseau sur lesquelles la demande a été reçue et à travers lesquelles le chemin pour la cible passe sont les mêmes, la passerelle ne doit pas répondre. Dans ce cas, soit l'hôte cible est sur le même réseau physique que la passerelle (et donc l'hôte devrait répondre pour lui même) soit cette passerelle n'est pas sur le chemin le plus direct au réseau désiré, c'est-à-dire, il y a une autre passerelle sur le même réseau physique qui est sur un chemin plus direct et l'autre passerelle devrait répondre.

La [RFC0925] décrit un mécanisme général pour l'acheminement dynamique de sous réseau en utilisant le mandataire ARP et les antémémoires d'acheminement dans les passerelles. Notre technique est restreinte aux sous réseaux de la RFC0925, dans lesquels on utilise des chemins de sous réseau statiques qui sont déterminés administrativement. Par suite, nos passerelles de sous réseau transparentes n'exigent pas de nouvelles entrées de tableau d'acheminement réseau ni d'entrée d'antémémoire ARP ; les seuls tableaux qui sont affectés sont les antémémoires ARP dans l'hôte.

Dans notre mise en œuvre, les boucles d'acheminement sont empêchées par l'administration appropriée des tableaux d'acheminement de sous réseau dans les passerelles.

2.3 Passerelles multiples

La plus simple organisation de sous réseau à administrer est une structure d'arborescence, qui ne peut pas avoir de boucles. Cependant, il peut être souhaitable pour la fiabilité ou le traitement du trafic d'avoir plus d'une passerelle (ou chemin) entre deux réseaux physiques. Les passerelles de sous réseau ARP peuvent être utilisées dans une telle situation : un hôte demandeur va utiliser la première réponse ARP qu'il reçoit, même si plus d'une passerelle en fournit une. Cela peut même fournir un service rudimentaire d'équilibrage de charge, car si deux passerelles sont par ailleurs similaires, celui qui est le plus légèrement chargé est probablement celui qui va répondre le premier.

Des mécanismes plus complexes pourraient être construits sous la forme de protocoles de passerelle à passerelle, et ils vont

sans doute devenir nécessaires dans les réseaux avec un grand nombre de sous réseaux et passerelles, de la même façon que les protocoles de passerelle à passerelle sont généralement nécessaires parmi les passerelles IP.

2.4 Vérification de bonne santé

Les administrateurs de réseau et de passerelle doivent veiller à garder les gabarits de réseau les mêmes sur toutes les machines de passerelle de sous réseau. L'erreur la plus commune est de régler le gabarit de réseau sur un hôte sans une mise en œuvre de sous réseau pour inclure le numéro de sous réseau. Ceci cause l'échec de l'hôte quand il tente d'envoyer des paquets aux hôtes qui ne sont pas sur son sous réseau local. L'ajustement de ses tableaux d'acheminement ne va pas aider, car il ne va pas savoir comment acheminer aux sous réseaux.

Si les réseaux IP des hôtes source et cible d'une demande ARP sont différents, une mise en œuvre de passerelle de sous réseau ARP ne devrait pas répondre. C'est pour empêcher la passerelle de sous réseau ARP d'être utilisée pour atteindre des réseaux IP étrangers et donc éventuellement de court-circuiter les vérifications de sécurité fournies par les passerelles IP.

Une mise en œuvre de passerelle de sous réseau ARP ne doit pas répondre si les réseaux physiques de source et cible d'une demande ARP sont les mêmes. Dans ce cas, soit l'hôte cible est probablement sur le même réseau physique que l'hôte source et peut répondre pour lui-même, soit l'hôte cible est dans la même direction de la passerelle que l'hôte source, et une réponse ARP en provenant causerait une boucle.

Une demande ARP pour une adresse de diffusion doit ne pas répondre, sans considération de l'adresse de source ou des réseaux physiques impliqués. Si la passerelle devait répondre avec une réponse ARP dans cette situation, cela inviterait la source originale à envoyer du trafic réel à une adresse de diffusion. Il pourrait en résulter un "effet Tchernobyl" par lequel chaque hôte sur le réseau réplique à un tel trafic, causant la "fusion" du réseau.

2.5 Plusieurs sous réseaux logiques par réseau physique

La façon la plus directe d'allouer des numéros de sous réseau est un à un avec les réseaux physiques. Il y a, cependant des circonstances dans lesquelles plusieurs sous réseaux logiques par réseau physique sont assez utiles. Une des plus courantes est quand il est prévu qu'un groupe de stations de travail vont être mises sur leur propre réseau physique mais que la passerelle pour le nouveau réseau physique doit d'abord être vérifiée. (Un répéteur pourrait être utilisé quand la passerelle n'est pas utilisable). Si une règle d'un sous réseau par réseau physique est appliquée, les adresses des stations de travail doivent être changées chaque fois que la passerelle est testée. Si on peut allouer des adresses en utilisant un nouveau numéro de sous ensemble alors qu'ils sont encore sur l'ancien réseau physique, aucun autre changement d'adresse n'est nécessaire.

Pour permettre plusieurs sous réseaux par réseau physique, une passerelle de sous réseau ARP doit utiliser l'interface de réseau physique, et non le numéro de sous réseau, pour déterminer quand répondre à une demande ARP. C'est-à-dire, elle devrait envoyer une réponse de mandataire ARP seulement quand l'interface réseau de source diffère de l'interface réseau cible. De plus, les entrées de tableau d'acheminement appropriées pour ces sous réseaux "fantômes" doivent être ajoutées aux tableaux d'acheminement de passerelle de sous réseau.

2.6 Adresses de diffusion

Il y a deux sortes d'adresses de diffusion IP : de diffusion de réseau dirigé IP principal, principale et de diffusion de sous réseau. Une adresse de diffusion de réseau IP consiste en le numéro de réseau plus une valeur bien connue dans le reste (partie locale) de l'adresse. Une diffusion de sous réseau IP est similaire, sauf que les bits de numéro de réseau IP et de numéro de sous réseau sont inclus. La RFC0922 a normalisé l'utilisation de tous les bits à un dans la partie locale, mais il y avait deux conventions utilisées avant cela : tout de uns et tout de zéros. Par exemple, 4.2BSD utilisait tout à zéro, et 4.3BSD utilise tout de uns. Donc il y a quatre sortes d'adresses de diffusion IP toujours actuellement utilisées sur de nombreux réseaux.

Avec le sous réseautage transparent, une passerelle de sous réseau ne doit pas produire une diffusion IP en utilisant l'adresse de diffusion de sous réseau, par exemple, 128.83.138.255. Les hôtes sur les réseaux physiques qui reçoivent la diffusion ne vont pas comprendre une telle adresse comme adresse de diffusion, car ils ne vont pas avoir de sous réseaux activés (ou ne vont pas avoir de mise en œuvre de sous réseau). En fait, les hôtes 4.2BSD (avec ou sans mise en œuvre de sous réseau) vont plutôt traiter une adresse avec tous les bits à un dans la partie locale comme une adresse spécifique d'hôte et essayer de transmettre le paquet. Comme il n'y a pas de tel hôte cible, il ne va pas y avoir d'entrée dans les tableaux ARP de l'hôte transmetteur et il va générer une demande ARP pour l'hôte cible. Cela présente le scénario (actuellement observé)

d'une passerelle 4.3BSD fonctionnant avec le programme rwho, qui diffuse un paquet une fois par minute, causant la génération par chaque hôte 4.2BSD sur le réseau physique local d'une demande ARP au même moment. Le même problème se produit avec toute adresse de diffusion de sous réseau, que la partie locale soit toute de zéros ou de uns.

Donc une passerelle de sous réseau dans un réseau avec des hôtes qui ne comprennent pas les sous réseaux doit veiller à ne pas utiliser des adresses de diffusion de sous réseau : elle doit à la place utiliser l'adresse de diffusion dirigée sur le réseau IP.

Finalement comme de nombreux hôtes fonctionnant avec un logiciel dépassé vont encore utiliser (et attendre) des adresses de diffusion de réseau IP de vieux style toutes de zéros, la passerelle doit envoyer ses adresses de diffusion sous cette forme, par exemple, 128.83.0.0. Il peut être sûr d'envoyer aussi un paquet dupliqué avec la partie locale toute de uns, par exemple, 128.83.255.255. Il n'est pas clair que l'adresse de diffusion de réseau local toute de uns, 255.255.255.255, ait des effets néfastes, mais il est très probable qu'elle ne va pas être reconnue par de nombreux hôtes qui fonctionnent avec l'ancien logiciel.

3. Mise en œuvre dans 4.3BSD

Les passerelles de sous réseau qui utilisent ARP ont été mises en œuvre par un certain nombre de gens différents. La méthode particulière décrite dans le présent mémoire a d'abord été mise en œuvre dans 4.2BSD par dessus un code de sous réseau 4.3BSD beta-test retro compatible, et a depuis été remis en œuvre comme ajout aux sources 4.3BSD réparties. On décrit ici la dernière mise en œuvre.

La plus grande partie du nouveau code de noyau pour la fonction de passerelle ARP de sous réseau est dans le module générique d'interface Ethernet, netinet/if_ether.c. Elle consiste en huit lignes dans in_arpinput qui effectuent un couple de vérifications rapides (pour s'assurer que la facilité est activée sur l'interface de source et que les adresses de source et de cible sont sur des sous réseaux différents), qui invoquent un nouveau sous programme, if_subarp, pour des vérifications ultérieures, et construit ensuite la réponse ARP si toutes les vérifications ont réussi. Ce code n'est atteint que quand une demande ARP est reçue, et ne fait rien si la facilité n'est pas activée sur l'interface de source. Donc les performances de la passerelle devraient être très peu dégradées par cet ajout. (Les performances de l'hôte demandeur devraient aussi être similaires au dernier cas, car la seule différence est entre l'efficacité de l'antémémoire ARP et celle des tableaux d'acheminement).

Le sous programme if_subarp (environ soixante lignes) s'assure que les adresses de source et de cible sont sur le même réseau IP et que l'adresse de cible n'est d'aucune des quatre sortes d'adresses de diffusion dirigée. Il tente ensuite de trouver un chemin vers la cible soit en trouvant une interface réseau avec le sous réseau désiré, soit en cherchant dans les tableaux d'acheminement. Même si une interface réseau est trouvée qui conduit à la cible, pour qu'une réponse soit envoyée, la passerelle ARP doit être activée sur cette interface et les interfaces cible et source doivent être différentes.

Le fichier netinet/route.c a une définition de structure d'entrée d'acheminement statique qui est ajoutée, et des modifications d'environ huit lignes sont faites au principal sous programme de recherche de tableau d'acheminement, rtalloc, pour reconnaître un pointeur sur cette structure (quand on passe par if_subarp) comme directive de ne pas utiliser le chemin par défaut dans cette vérification d'acheminement. Le niveau de priorité de processeur (protection critique de section) autour de la vérification interne de recherche d'acheminement est changé en une valeur supérieure, car le sous programme peut maintenant être invoqué à partir des interruptions d'interface réseau ainsi qu'à partir des interruptions de logiciel internes qui pilotent le traitement de IP et des autres protocoles de haut niveau. On peut concevoir que cette priorité de processeur relevée pourrait un peu ralentir tout le noyau si il y a de nombreuses vérifications d'acheminement, mais comme la section critique est rapide, l'effet devrait être faible.

Une modification clé du noyau est l'ajout d'environ quinze lignes au sous programme ip_output dans netinet/ip_output.c. Cela change les adresses de diffusion de sous réseau dans les paquets qui ont leur origine à la passerelle d'adresses de diffusion de réseau IP afin que les hôtes sans code de sous réseau (ou avec leur gabarit de réseau réglé à ignorer les sous réseaux) les reconnaissent comme adresses de diffusion. Cette section de code n'est utilisée que si la passerelle ARP est activée pour l'interface sortante, et n'affecte que les adresses de diffusion de sous réseau.

Un nouveau sous programme, in_mainnetof, d'environ quinze lignes, est ajouté à netinet/in.c pour retourner le numéro de réseau IP (sans numéro de sous réseau) à partir d'une adresse IP. Il est invoqué à partir de if_subarp et ip_output.

Deux fichiers de paramètre du noyau ont chacun une ligne ajoutée : net/if.h a une définition d'un bit dans la structure d'interface réseau pour indiquer si les passerelles ARP de sous réseau sont activées, et netinet/in.h se réfère à in_mainnetof.

En plus de ces environ 110 lignes d'ajout au noyau source, il y a une modification de niveau utilisateur. La source de la commande ifconfig, qui est utilisée pour régler les adresses et les gabarits de réseau des interfaces réseau, a quatre lignes ajoutées pour lui permettre d'activer ou désactiver la facilité de passerelle ARP de sous réseau, pour chaque interface. Ceci est documenté dans onze nouvelles lignes dans l'entrée manuelle pour cette commande.

4. Disponibilité

La mise en œuvre 4.3BSD est actuellement disponible par FTP anonyme (login anonymous, mot de passe guest) à partir de sally.utexas.edu comme pub/subarp, qui est un listing 4.3BSD "diff -c" à partir des sources 4.3BSD qui ont été distribuées en septembre 1986.

Cette mise en œuvre n'a pas été incluse dans la distribution 4.3BSD propre parce que le CSRG de l'U.C. Berkeley pensait que cela réduirait l'incitation des fabricants à mettre en œuvre des sous réseaux conformément à la RFC0950. Les auteurs y ont concouru. Néanmoins, il y a des circonstances dans lesquelles l'utilisation de passerelles ARP de sous réseau transparentes est indispensable.

Références

[Nameservers] Carl-Mitchell, S., et J. S. Quarterman, "Nameservers in a Campus Domain", SIGCUE Outlook, Vol.19, No.1/2, pp.78-88, ACM SIG Computer Uses in Education, P.O. Box 64145, Baltimore, MD 21264, Spring/Summer 1986.

[RFC0950] J. Mogul et J. Postel, "Procédure standard de sous-réseautage Internet", (STD 5) août 1985.

[RFC0922] J. Mogul, "Diffusion des datagrammes Internet en présence de sous-réseaux", octobre 1984.

[RFC0826] D. Plummer, "Protocole de résolution d'adresses Ethernet : conversion des adresses de protocole réseau en adresses Ethernet à 48 bits pour transmission sur un matériel Ethernet", STD 37, novembre 1982.

[RFC0925] J. Postel, "Résolution d'adresse dans les multi-LAN", octobre 1984.

[RFC1009] R. Braden et J. Postel, "Exigences pour les routeurs de l'Internet", juin 1987. *(Obsolète, voir* RFC1812*) (Historique)*