Groupe de travail Réseau
Request for Comments : 4274
Catégorie : Information
Traduction Claude Brière de L'Isle

D. Meyer, Cisco Systems
K. Patel, Cisco Systems

janvier 2006

Analyse du protocole BGP-4

Statut de ce mémoire
Le présent document fournit des informations pour la communauté de l'Internet. Il ne spécifie aucune sorte de norme de l'Internet. La distribution du présent mémoire n'est soumise à aucune restriction.

Résumé
L'objet de ce rapport est de documenter comment les exigences de publication d'un protocole d'acheminement comme projet de norme de l'Internet ont été satisfaites par la version 4 du protocole de routeur frontière (BGP-4, *Border Gateway Protocol version 4*).

Le présent rapport satisfait les exigences de "second rapport", comme décrit au paragraphe 6.0 de la RFC 1264. Afin de satisfaire cette exigence, le présent rapport complète la RFC 1774 et résume les caractéristique clés de BGP-4, ainsi qu'analyse le protocole sous les aspects de l'adaptabilité et des performances.

Tables des Matières

1. Introduction

BGP-4 est un protocole d'acheminement inter systèmes autonomes conçu pour les internets TCP/IP. La version 1 de BGP a été publiée dans la [RFC1105]. Depuis lors, les versions 2, 3, et 4 de BGP ont été développées. La version 2 était documentée dans la [RFC1163]. La version 3 est documentée dans la [RFC1267]. La version 4 est documentée dans la [RFC4271] (on se réfère ici à la la version 4 de BGP comme BGP). Les changements entre les versions sont expliqués dans l'Appendice A de la [RFC4271]. Les applications possibles de BGP dans l'Internet sont documentées dans la [RFC1772].

BGP introduit la prise en charge de l'acheminement inter domaines sans classe (CIDR, *Classless Inter-Domain Routing*) [RFC1519]. Comme les versions antérieures de BGP n'avaient pas la prise en charge de CIDR, elles sont considérées comme obsolètes et inutilisables dans l'Internet d'aujourd'hui.

L'objet du présent rapport est de documenter comment les exigences de publication d'un protocole d'acheminement comme projet de norme de l'Internet ont été satisfaites par le protocole de routeur frontière version 4 (BGP-4, *Border Gateway Protocol version 4*).

Le présent rapport satisfait aux exigences d'un "second rapport", comme décrit au paragraphe 6.0 de la [RFC1264]. Afin de satisfaire aux exigences, ce rapport complète la [RFC1774] et résume les caractéristiques clés de BGP-4, ainsi qu'analyse le protocole sous les aspects de l'adaptabilité et des performances.

2. Caractéristiques et algorithmes clés de BGP

Cette section résume les caractéristiques clés et les algorithmes de BGP. BGP est un protocole d'acheminement inter systèmes autonomes ; il est conçu pour être utilisé entre plusieurs systèmes autonomes. BGP suppose que l'acheminement au sein d'un système autonome est fait par un protocole d'acheminement intra système autonome. BGP suppose aussi que les paquets de données sont acheminés de la source vers la destination indépendamment de la source. BGP ne fait aucune hypothèse sur les protocole d'acheminement intra système autonome déployés au sein de divers systèmes autonomes. Précisément, BGP n'exige pas que tous les systèmes autonomes fonctionnent avec le même protocole d'acheminement intra système autonome (c'est-à-dire, le protocole de routeur intérieur ou IGP).

On notera finalement que BGP est un protocole réel d'acheminement inter système autonome et qu'à ce titre, il n'impose pas de contraintes à la topologie d'interconnexion sous-jacente des systèmes autonomes. Les informations échangées via BGP sont suffisantes pour construire un graphe de la connexité des systèmes autonomes à partir duquel les boucles d'acheminement peuvent être élaguées, et de nombreuses décisions de politique d'acheminement au niveau du système autonome peuvent être mises en application.

2.1 Caractéristiques clés

Les caractéristiques clés du protocole sont la notion d'attributs de chemin et d'agrégation des informations d'accessibilité de couche réseau (NLRI, *Network Layer Reachability Information*).

Les attributs de chemin donnent à BGP la souplesse et l'extensibilité. Les attributs de chemin sont soit bien connus, soit facultatifs. La disposition d'attributs facultatifs permet l'expérimentation qui peut impliquer un groupe de routeurs BGP sans affecter le reste de l'Internet. De nouveaux attributs facultatifs peuvent être ajoutés au protocole de la même façon que de nouvelles options sont ajoutées, par exemple, au protocole Telnet [RFC0854].

Un des plus importants attributs de chemin est le chemin de système autonome, ou AS_PATH. Lorsque les informations d'accessibilité traversent l'Internet, ces informations (AS_PATH) sont augmentées de la liste des systèmes autonomes qui ont été traversés jusqu'alors, formant le AS_PATH. Le AS_PATH permet une suppression directe des informations d'acheminement en boucle. De plus, AS_PATH sert de mécanisme puissant et polyvalent pour l'acheminement fondé sur la politique.

BGP améliore l'attribut AS_PATH pour y inclure des ensembles de systèmes autonomes ainsi que des listes via l'attribut AS_SET. Ce format étendu permet de générer des routes agrégées pour porter les informations de chemin depuis les routes les plus spécifiques utilisées pour générer l'agrégat. On notera cependant qu'au moment de la rédaction de ce mémoire, les attributs AS_SET sont rarement utilisés dans l'Internet [ROUTEVIEWS].

2.2 Algorithmes de BGP

BGP utilise un algorithme qui n'est ni un pur algorithme de vecteur de distance ni un pur algorithme d'état de liaison. À la place, il utilise un algorithme de vecteur de distance modifié, appelé un algorithme de "vecteur de chemin". Cet algorithme utilise les informations de chemin pour éviter les problèmes traditionnels de vecteur de distance. Chaque route au sein de BGP apparie les informations sur la destination avec les informations de chemin pour cette destination. Les informations de chemin (aussi appelées informations AS_PATH) sont mémorisées au sein de l'attribut AS_PATH dans BGP. Les informations de chemin aident BGP à détecter les boucles d'AS, permettant ainsi aux locuteurs BGP de choisir des routes sans boucle.

BGP utilise une stratégie de mise à jour par incrément pour préserver la bande passante et la puissance de traitement. C'est-à-dire qu'après l'échange initial des informations complètes d'acheminement, une paire de routeurs BGP n'échangent plus que les changements à ces informations. Une telle conception de mise à jour incrémentaire exige un transport fiable entre une paire de routeurs BGP afin de fonctionner correctement. BGP résout ce problème en utilisant TCP comme transport fiable.

En plus des mises à jour incrémentaires, BGP a ajouté le concept d'agrégation de route afin que les informations sur les groupes de destinations qui utilisent une allocation hiérarchique d'adresse (par exemple, CIDR) puissent être agrégées et envoyées comme une seule NLRI.

Finalement, on notera que BGP est un protocole autonome. C'est-à-dire que BGP spécifie comment les informations d'acheminement sont échangées, à la fois entre les locuteurs BGP dans des systèmes autonomes différents, et entre des locuteurs BGP au sein d'un seul système autonome.

2.3 Automate à états finis de BGP

L'automate à états finis (FSM, *Finite State Machine*) BGP est un ensemble de règles qui sont appliquées à un ensemble d'homologues locuteurs BGP configurés pour le fonctionnement de BGP. Une mise en œuvre de BGP exige qu'un locuteur BGP se connecte et écoute sur l'accès TCP 179 pour accepter toute nouvelle connexion BGP avec ses homologues. L'automate à états finis de BGP doit être initié et maintenu pour chaque nouvelle connexion d'homologue entrante et sortante. Cependant, en fonctionnement en régime permanent, il y aura seulement un FSM BGP par connexion par homologue.

Il peut y avoir une courte période durant laquelle un homologue BGP peut avoir des connexions entrantes et sortantes séparées résultant en la création de deux (au lieu d'un seul) FSM BGP différents se rapportant à un homologue. Cela peut être résolu en suivant les règles de collision de connexions de BGP définies dans la [RFC4271].

Le FSM BGP a les états suivants associés à chacun de ses homologues :

IDLE *(repos)* : état dans lequel l'homologue BGP refuse toute connexion entrante.

CONNECT *(connecter)* : état dans lequel l'homologue BGP attend que sa connexion TCP soit réalisée.

ACTIVE *(actif)* : état dans lequel l'homologue BGP essaye d'acquérir un homologue en écoutant et acceptant une connexion TCP.

OPENSENT *(ouvert-envoyé)* : l'homologue BGP attend le message OPEN de son homologue.

OPENCONFIRM *(ouvert-confirmé)* : l'homologue BGP attend le message KEEPALIVE *(garder en vie)* ou NOTIFICATION de son homologue.

ESTABLISHED *(établi)* : la connexion BGP avec l'homologue est établie et échange des messages UPDATE, NOTIFICATION, et KEEPALIVE avec son homologue.

Il y a un certain nombre d'événements BGP qui opèrent sur les états susmentionnés du FSM BGP pour les homologues BGP. La prise en charge de ces événements BGP est obligatoire ou facultative. Ces événements sont déclenchés par la logique du protocole au titre de BGP ou en utilisant l'intervention de l'opérateur via une interface de configuration au protocole BGP.

Ces événements BGP sont des types suivants : événement facultatifs reliés aux attributs de session facultatifs, événements administratifs, événements de temporisateur, événements fondés sur la connexion TCP, et événements BGP fondés sur un message. Le FSM et les événements BGP sont expliqués en détail dans la [RFC4271].

3. Capacités de BGP

Le mécanisme de capacité BGP [RFC3392] fournit un moyen souple et aisé d'introduire de nouvelles caractéristiques au sein du protocole. En particulier, le mécanisme de capacité BGP permet à un locuteur BGP d'annoncer au démarrage à ses homologues diverses caractéristiques facultatives prises en charge par le locuteur (et de recevoir des informations similaires des homologues). Cela permet au BGP de base de ne contenir que les fonctionnalités essentielles, tout en fournissant un mécanisme souple pour signaler les extensions au protocole.

4. Oscillations persistantes d'homologue BGP

Chaque fois qu'un locuteur BGP détecte une erreur dans une connexion d'homologue, il ferme la connexion et change l'état de son FSM en REPOS. Un locuteur BGP a besoin d'un événement Démarrage pour réinitialiser une connexion d'homologue au repos. Si l'erreur persiste et si le locuteur BGP génère automatiquement un événement Démarrage, il peut alors en résulter une oscillation persistante de l'homologue. Bien que les oscillations d'homologue soient largement répandues dans les mises en œuvre de BGP, les méthodes pour empêcher les oscillations persistantes d'homologue sortent du domaine d'application de la spécification BGP de base.

5. Lignes directrices de mise en œuvre

Une mise en œuvre robuste de BGP est "conservatrice du travail". Cela signifie que si le nombre de préfixes est limité, des niveaux élevés arbitraires de changement de route peuvent être tolérés. Des niveaux élevés peuvent être tolérés avec un impact limité sur la convergence de route pour des changements occasionnels dans des routes généralement stables.

Une mise en œuvre robuste de BGP devrait avoir les caractéristiques suivantes :

1. Elle est capable de fonctionner dans des niveaux élevés presque arbitraires d'oscillation de route sans perdre la relation d'homologue à homologue (sauf à envoyer des "garder en vie") ou perdre d'autres adjacences de protocole par suite de la charge BGP.

2. L'instabilité d'un sous ensemble de routes ne devrait pas affecter les annonces de route et les transmissions associées à l'ensemble de routes stables.

3. L'instabilité ne devrait pas être causée par des homologues avec de hauts niveaux d'instabilité ou avec des vitesses ou charges de CPU différentes qui résultent en un traitement plus rapide ou plus lent des routes. Ces homologues instables devraient avoir un impact limité sur le temps de convergence pour les routes généralement stables.

Il existe de nombreuses mises en œuvre robustes de BGP. Produire une mise en œuvre robuste n'est pas un affaire triviale, mais est clairement réalisable.

6. Caractéristiques et adaptabilité des performances de BGP

Dans cette section, on fournit des réponses "d'ordre de grandeur" aux questions sur la quantité de bande passante de liaison, de mémoire de routeur et de cycles de CPU de routeur que BGP va consommer dans des conditions normales. En particulier, on va traiter de l'adaptabilité de BGP et de ses limitations.

6.1 Bande passante de liaison et utilisation de CPU

Immédiatement après l'établissement initial de la connexion BGP, les homologues BGP échangent les ensembles complets d'informations d'acheminement. Si on note N le nombre total de routes dans l'Internet, A le total des attributs de chemin (pour toutes les routes N) reçus d'un homologue, et si on suppose que les réseaux sont uniformément distribués parmi les systèmes autonomes, alors le pire cas de consommation de bande passante durant l'échange initial entre une paire de locuteurs BGP (P) est BW = O((N + A) * P)

BGP-4 a été créé spécifiquement pour réduire la taille de l'ensemble d'entrées de NLRI, qui doivent être portées et échangées par les routeurs de frontière. Le schéma d'agrégation, défini dans la [RFC1519], décrit le schéma d'agrégation fondée sur le fournisseur en usage dans l'Internet d'aujourd'hui.

Du fait des avantages d'annoncer quelques grands blocs agrégés (au lieu de nombreux plus petits réseaux individuels fondés sur la classe) il est difficile d'estimer la réduction réelle de bande passante et de traitement que BGP-4 a fourni par rapport à BGP-3. On énumère simplement tous les blocs agrégés dans leur réseau individuel, fondé sur la classe, on ne prendra pas en compte l'espace "mort" qui avait été réservé pour de futures expansions. La meilleure métrique pour déterminer le succès de l'agrégation de BGP est d'échantillonner le nombre d'entrées de NLRI dans l'Internet mondialement connecté d'aujourd'hui, et de le comparer aux taux de croissance qui ont été projetés avant le déploiement de BGP.

Au moment de cette rédaction, l'ensemble complet des routes extérieures portées par BGP est approximativement de 134 000 entrées de réseaux [ROUTEVIEWS].

6.1.1 Utilisation de CPU

Une caractéristique importante et fondamentale de BGP est que l'utilisation de CPU de BGP ne dépend que de la stabilité de son réseau qui se rapporte à BGP en termes d'annonces de messages UPDATE de BGP. Si le réseau BGP est stable, tous les routeurs BGP au sein de son réseau sont dans l'état "régime permanent". Donc, la seule bande passante de liaison et les seuls cycles de CPU de routeur consommés par BGP sont dus à l'échange des messages KEEPALIVE de BGP. Les messages KEEPALIVE ne sont échangés qu'entre les homologues. La fréquence suggérée de l'échange est toutes les 30 secondes. Les messages KEEPALIVE sont assez courts (19 octets) et n'exigent virtuellement pas de traitement. Par suite, la bande passante consommée par les messages KEEPALIVE est d'environ 5 bits/s. L'expérience du fonctionnement confirme que la redondance (en termes de bande passante et de CPU) associée aux messages KEEPALIVE devrait être vue comme négligeable.

Durant les périodes d'instabilité du réseau, les routeurs BGP au sein du réseau génèrent des mises à jour d'acheminement qui sont échangées en utilisant les messages BGP UPDATE. La plus grande redondance par message UPDATE se produit quand chaque message UPDATE contient un seul réseau. On souligne que, en pratique, les changements d'acheminement présentent un fort caractère local par rapport aux attributs de route. C'est-à-dire que les routes qui changent vont probablement avoir des attributs de route communs. Dans ce cas, plusieurs réseaux peuvent être groupés en un seul message UPDATE, réduisant donc significativement la quantité de bande passante requise (voir aussi l'Appendice F.1 de la RFC4271]).

6.1.2 Exigences de mémoire

Pour quantifier les exigences de mémoire dans le pire des cas pour BGP, on note N le nombre total de réseaux dans l'Internet, M la distance moyenne d'AS dans l'Internet (distance au niveau d'un système autonome, exprimée en termes de nombre de systèmes autonomes) et comme A le nombre total de chemins d'AS uniques. Les exigences de mémoire dans le pire des cas (MR) peuvent être exprimées par MR = O(N + (M * A))

Parce qu'une distance moyenne d'AS M est une fonction à mouvement lent de l'inter connexité ("treillage") de l'Internet, pour tous les aspects pratiques des pires cas d'exigences de mémoire des routeur sont de l'ordre du nombre total de réseaux dans l'Internet multiplié par le nombre d'homologues que le système local apparie. On s'attend à ce que le nombre total de réseaux dans l'Internet croisse beaucoup plus vite que le nombre moyen d'homologues par routeur. Il en résulte que les propriétés d'adaptation de mémoire de BGP sont en relation linéaire avec le nombre total de réseaux dans l'Internet.

Le tableau qui suit illustre les exigences typiques de mémoire d'un routeur fonctionnant avec BGP. On note par N le nombre moyen de routes annoncé par chaque homologue, par A le nombre total de chemins d'AS uniques, comme M la distance moyenne d'AS de l'Internet (distance au niveau d'un système autonome, exprimée en termes de nombre de systèmes autonomes) comme R le nombre d'octets requis pour mémoriser un réseau, et comme P le nombre d'octets requis pour mémoriser un AS dans un chemin d'AS. On suppose que chaque réseau est codé sur quatre octets, chaque AS est codé sur deux octets, et chaque réseau est accessible via une fraction de tous les homologues (nombre d'homologues BGP par réseau). Pour les besoins de notre estimation, on calculera MR = (((N * R) + (M * A) * P) * S).

Nombre réseaux (N)	Distance moyenne d'AS (M)	Nombre d'AS (A)	Nombre d'homo BGP / réseau (P)	Exigence mémoire (MR)
100 000	20	3 000	20	10 400 000
100 000	20	15 000	20	20 000 000
120 000	10	15 000	100	78 000 000
140 000	15	20 000	100	116 000 000

Pour analyser les exigences de mémoire de BGP, on se concentre sur la taille du tableau de RIB de BGP (en ignorant les détails de mise en œuvre). En particulier, on déduit les limites supérieures pour la taille du tableau de RIB de BGP. Par exemple, au moment de cette rédaction, les tableaux de RIB BGP d'un routeur normal de cœur de réseau portent de l'ordre de 120 000 entrées. Connaissant ce nombre, on peut se demander si il serait possible d'avoir un routeur fonctionnel avec un tableau contenant 1 000 000 d'entrées. Il est clair que la réponse à cette question réside dans la façon dont BGP est mis en œuvre. Une mise en œuvre robuste de BGP avec une CPU et une mémoire raisonnable ne devrait pas avoir de problème à s'adapter à de telles limites.

7. Expression de la politique de BGP et implications

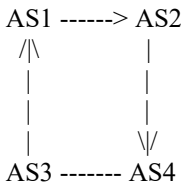
BGP est unique parmi les protocoles d'acheminement IP déployés en ce que l'acheminement est déterminé en utilisant des politiques d'acheminement sémantiquement riches. Bien que les politiques d'acheminement soient la première question de BGP qui vienne à l'esprit d'un opérateur de réseau, il est important de noter que les langages et techniques pour spécifier les politiques d'acheminement de BGP ne font pas partie de la spécification du protocole (voir la [RFC2622] pour un exemple d'un tel langage de politique). De plus, la spécification de BGP contient peu de restrictions, explicites ou implicites, sur les langages de politique d'acheminement. Ces langages ont normalement été développés par les fabricants et ont évolué à travers les interactions avec les ingénieurs réseau dans un environnement manquant de normes indépendantes des fabricants.

La complexité des configurations typiques de BGP, au moins dans les réseaux de fournisseurs d'accès, a été en croissance constante. Les fabricants de routeurs fournissent normalement des centaines de commandes particulières à utiliser dans la configuration de BGP, et cet ensemble de commandes s'étend continuellement. Par exemple, les communautés BGP [RFC1997] permettent aux auteurs de politiques de joindre de façon sélective des étiquettes à des routes et de les utiliser pour signaler les informations de politique aux autres routeurs locuteurs BGP. De nombreux fournisseurs permettent à leurs clients, et parfois à leurs homologues, d'envoyer des communautés qui déterminent la portée et les préférences de leurs routes. À cause de ces développements, la tâche d'écrire des configurations BGP a de plus en plus d'aspects associés à la programmation ouverte. Cela permet aux opérateurs de réseau de coder des politiques complexes afin de traiter de nombreuses situations imprévues, et a ouvert la porte à une grande créativité et à des expérimentations de politiques d'acheminement. Cette souplesse de politique est une des principales raisons pour lesquelles BGP convient si bien à l'environnement commercial de l'Internet actuel.

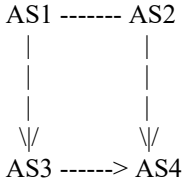
Cependant, cette riche expression de politique s'est faite à un prix qui n'est pas toujours reconnu. En particulier, il est possible de construire des politiques d'acheminement définies localement qui peuvent conduire à une divergence du protocole et à des anomalies d'acheminement global inattendues telles que du non déterminisme (non intentionnel). Si les politiques inter agissantes qui causent de telles anomalies sont définies dans différents systèmes autonomes, ces problèmes peuvent alors être très difficiles à déboguer et corriger. Dans les paragraphes qui suivent, on décrit deux de ces cas qui se rapportent à l'existence (ou au manque) d'acheminement stable.

7.1 Existence d'acheminements uniques stables

On peut facilement construire des ensembles de politiques pour lesquels BGP ne peut pas garantir que des acheminements stables sont uniques. Ceci est illustré par le simple exemple suivant. Considérons quatre systèmes autonomes, AS1, AS2, AS3, et AS4. AS1 et AS2 sont des homologues, et ils assurent le transit pour respectivement AS3 et AS4. Supposons que AS3 fournit le transit pour AS4 (dans ce cas, AS3 est un client de AS1, et AS4 est un client multi rattachements de AS3 et AS2). AS4 peut vouloir utiliser la liaison pour AS3 comme une liaison de "sauvegarde", et envoie à AS3 une valeur de communauté que AS3 a configurée pour diminuer la préférence des routes de AS4 à un niveau en dessous de celui de son fournisseur en amont, AS1. L'acheminement "de sauvegarde" prévu pour AS4 est illustré ci-dessous :



C'est-à-dire que la liaison AS3-AS4 est destinée à être utilisée seulement quand la liaison AS2-AS4 est empêchée (pour le trafic sortant, AS4 donne simplement aux routes provenant de AS2 une préférence locale supérieure). C'est un scénario courant dans l'Internet d'aujourd'hui. Mais on notera que cette configuration a une autre solution stable :



Dans ce cas, AS3 ne traduit pas la valeur de communauté "depref my route" reçue de AS4 dans une communauté "depref my route" pour AS1. Donc, si AS1 écoute la route pour AS4 qui transite par AS3, il va préférer cette route (parce que AS3 est un client). Cet état pourrait être atteint, par exemple, en commençant par l'acheminement de sauvegarde "correct"

montré d'abord, terminant la session BGP AS2-AS4, et l'activant à nouveau. En général, BGP n'a pas de moyen de préférer la solution "imprévue" à une solution anormale. La solution retenue va dépendre de l'ordre imprévisible des messages BGP.

Bien que cet exemple soit relativement simple, de nombreux opérateurs peuvent manquer à reconnaître que la vraie source du problème est que les politiques de BGP des AS peuvent interagir de façons inattendues, et que ces interactions peuvent résulter en plusieurs acheminements stables. On peut imaginer que les interactions pourraient être beaucoup plus complexes dans l'Internet réel. On soupçonne que de telles anomalies ne deviendront plus courantes que lorsque BGP aura continué d'évoluer vers une expressivité plus riche des politiques. Par exemple, des communautés étendues fournissent un moyen encore plus souple d'informations de signalisation au sein et entre les systèmes autonomes qu'il n'est possible avec les communautés de la [RFC1997]. En même temps, les applications des communautés par les opérateurs réseau évoluent pour traiter les problèmes complexes de l'ingénierie de trafic inter domaines.

7.2 Existence d'acheminements stables

On peut aussi construire un ensemble de politiques pour lesquelles BGP ne peut pas garantir qu'il existe un acheminement stable (ou pire, qu'on ne trouvera jamais un acheminement stable). Par exemple, la [RFC3345] présente plusieurs scénarios qui conduisent à des oscillations de route associées à l'utilisation de l'attribut "Discriminant multi sorties" (MED, *Multi-Exit Discriminator*). Les oscillations de route vont se produire dans BGP quand un ensemble de politiques n'a pas de solution. C'est-à-dire, quand il n'y a pas d'acheminement stable qui satisfasse aux contraintes imposées par la politique, BGP n'a pas d'autre choix que de continuer d'essayer. De plus, même si les configurations de BGP peuvent avoir un acheminement stable, le protocole peut n'être pas capable de le trouver ; BGP peut se "trouver piégé" dans une impasse qui n'a pas de solution.

La divergence de protocole n'est pas, cependant, un problème associé seulement à l'utilisation de l'attribut MED. Ce potentiel existe dans BGP même sans l'utilisation de l'attribut MED. Donc, comme l'indétermination involontaire décrite au paragraphe précédent, ce type de divergence de protocole est une conséquence involontaire de la nature non contrainte des langages de politique de BGP.

8. Applicabilité

Dans cette section on identifie les environnements pour lesquels BGP convient bien, et les environnements pour lesquels il ne convient pas. Cette question est partiellement traitée dans la Section 2 de BGP [RFC4271], qui déclare :

"Pour caractériser l'ensemble de décisions de politique qui peut être mis en application en utilisant BGP, on doit se concentrer sur la règle qu'un AS annonce à ses AS voisins seulement les routes qu'il utilise lui-même. Cette règle reflète le paradigme de l'acheminement "bond par bond" généralement utilisé partout dans l'Internet actuel. Noter que certaines politiques ne peuvent pas être prises en charge par le paradigme d'acheminement "bond par bond" et donc exigent des techniques telles que l'acheminement de source. Par exemple, BGP ne permet pas à un AS d'envoyer du trafic à un AS voisin dans l'intention que ce trafic prenne une route différente de celle prise par le trafic généré dans l'AS voisin. Par ailleurs, BGP peut prendre en charge toute politique se conformant au paradigme de l'acheminement "bond par bond". Comme l'Internet actuel utilise seulement le paradigme de l'acheminement "bond par bond" et comme BGP peut prendre en charge toute politique qui se conforme à ce paradigme, BGP est tout à fait applicable à un protocole d'acheminement inter AS pour l'Internet actuel".

Un des points importants est ici que BGP ne contient que les fonctions essentielles, tandis qu'en même temps, il fournit un mécanisme souple au sein du protocole qui nous permet d'étendre ses fonctions. Par exemple, les capacités de BGP donnent un moyen souple et facile d'introduire de nouvelles caractéristiques au sein du protocole. Finalement, comme BGP a été conçu pour être souple et extensible, de nouvelles exigences et/ou des évolutions des exigences peuvent être satisfaites par les mécanismes existants.

Pour résumer, BGP convient bien comme protocole d'acheminement inter systèmes autonomes pour tout internet qui se fonde sur IP [RFC791] comme le protocole Internet et le paradigme d'acheminement "bond par bond".

9. Remerciements

Nous tenons à remercier Paul Traina comme auteur des précédentes versions de ce document. Elwyn Davies, Tim Griffin, Randy Presuhn, Curtis Villamizar et Atanu Ghosh ont aussi fourni de nombreux commentaires pertinents sur les versions antérieures de ce document.

10. Considérations sur la sécurité

BGP fournit des mécanismes souples avec des niveaux de complexité variables pour les besoins de la sécurité. Les sessions BGP sont authentifiées en utilisant des adresses de session BGP et les numéros alloués aux AS. Comme les sessions BGP utilisent TCP (et IP) pour un transport fiable, les sessions BGP sont de plus authentifiées et sécurisées par tous les mécanismes d'authentification et de sécurité utilisés par TCP et IP.

BGP utilise l'option MD5 de TCP pour valider les données et les protéger contre l'usurpation des segments TCP échangés entre ses sessions. L'usage de l'option TCP MD5 pour BGP est décrit en détails dans la [RFC2385]. La gestion de clé TCP MD5 est exposée dans la [RFC3562]. Le chiffrement des données de BGP est fourni en utilisant le mécanisme IPsec, qui chiffre les données de la charge utile IP (incluant les données TCP et BGP). Le mécanisme IPsec peut être utilisé dans les deux modes transport et tunnel. Il est décrit dans la [RFC2406]. L'option TCP MD5 et le mécanisme IPsec ne sont tous deux pas des mécanismes de sécurité largement déployés pour BGP dans l'Internet d'aujourd'hui. Donc, il est difficile de juger leur réel impact de performances avec BGP. Cependant, comme les deux mécanismes sont fondés sur TCP et IP, la bande passante de liaison, l'utilisation de CPU et la mémoire de routeur consommées par BGP seraient les mêmes que pour tout autre protocole fondé sur TCP et IP.

BGP utilise la valeur de TTL de IP pour protéger ses sessions BGP externes (EBGP) contre toute attaque de consommation intensive de CPU fondée sur TCP ou IP. C'est un mécanisme simple qui suggère l'utilisation du filtrage des segments BGP (TCP) en utilisant la valeur de TTL IP portée dans l'en-tête IP des segments BGP (TCP) qui sont échangés entre les sessions BGP. Le mécanisme du TTL BGP est décrit dans la [RFC3682]. L'usage de la [RFC3682] impacte les performances d'une façon similaire à celle de l'utilisation de toute politique de liste de contrôle d'accès pour BGP.

De tels mécanismes de sécurité souples fondés sur TCP et IP permettent à BGP d'empêcher l'insertion, la suppression, ou la modification des données de BGP, tout espionnage des données et vol de session, etc. Cependant, BGP est vulnérable aux mêmes attaques que celles qui sont présentes dans TCP. La [RFC4272] explique en détails la vulnérabilité de la sécurité de BGP. Au moment de la rédaction de ce mémoire, plusieurs efforts sont en cours pour créer et définir une infrastructure de sécurité appropriée au sein du protocole BGP pour assurer l'authentification et la sécurité de ses informations d'acheminement ; ces efforts incluent [SBGP] et [SOBGP].

11. Références

11.1 Références normatives

[RFC0791] J. Postel, éd., "Protocole Internet - Spécification du protocole du programme Internet", STD 5, septembre 1981.

[RFC1519] V. Fuller, T. Li, J. Yu et K. Varadhan, "Acheminement inter domaine sans classe (CIDR) : stratégie d'allocation et d'agrégation d'adresses", septembre 1993. *(D.S., rendue obsolète par la* RFC4632)

[RFC1997] R. Chandra, P. Traina, T. Li, "Attribut Community de BGP", août 1996. *(P.S.)*

[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)

[RFC3345] D. McPherson et autres, "Condition d'oscillation de chemin persistante du protocole de routeur frontière (BGP)", août 2002. *(Information)*

[RFC3392] R. Chandra et J. Scudder, "Annonces de capacités avec BGP-4", novembre 2002. *(Obsolète, voir* RFC5492)

[RFC3562] M. Leech, "Considérations sur la gestion de clés pour l'option de signature MD5 dans TCP", juillet 2003. *(Information)*

[RFC3682] V. Gill, J. Heasley, D. Meyer, "Mécanisme TTL de sécurité généralisé (GTSM)", février 2004. (*Obsolète, voir* RFC5082) *(Expérimentale)*

[RFC4271] Y. Rekhter, T. Li et S. Hares, "Protocole de routeur frontière version 4 (BGP-4)", janvier 2006. *(D.S.) (MàJ par* RFC6608, RFC8212)

[RFC4272] S. Murphy, "Analyse des faiblesses de la sécurité de BGP", janvier 2006. *(Information)*

[SBGP] Kent, S., Lynn, C. et Seo, K., "Secure Border Gateway Protocol (Secure-BGP)", IEEE Journal on Selected Areas in Communications, Vol. 18, n° 4, avril 2000, pp. 582-592.

11.2 Références pour information

[RFC0854] J. Postel et J. Reynolds," Spécification du protocole TELNET", STD 8, mai 1983.

[RFC1105] K. Lougheed et Y Rekhter, "Protocole de routeur frontière BGP)", juin 1989. *(obsolète, voir 1163 et 1267, Historiques)*

[RFC1163] K. Lougheed, Y. Rekhter, "Protocole de routeur frontière", juin 1990. *(Obsolète, voir* RFC1267, *Historique*)

[RFC1264] R. Hinden, "Critères de normalisation du protocole d'acheminement Internet de l'IETF", octobre 1991. *(Historique, remplacée par la* RFC4794*)*

[RFC1267] K. Lougheed et Y Rekhter, "Protocole de routeur frontière 3 (BGP 3)", octobre 1991. *(Historique)*

[RFC1772] Y. Rekhter, P. Gross, "Application du protocole de routeur frontière dans l'Internet", mars 1995. *(D.S.)*

[RFC1774] P. Traina, éd., "Analyse du protocole BGP-4", mars 1995] *(Information)*

[RFC2406] S. Kent et R. Atkinson, "Encapsulation de charge utile de sécurité IP (ESP)", novembre 1998. *(Ob., voir* RFC4303)

[RFC2622] C. Alaettinoglu et autres, "Langage de spécification de politique d'acheminement (RPSL)", juin 1999. (*MàJ par* RFC4012, RFC7909) *(P.S.)*

[ROUTEVIEWS] Meyer, D., "The Route Views Project", http://www.routeviews.org.

[SOBGP] White, R., "Architecture and Deployment Considerations for Secure Origin BGP (soBGP)", non publiée, mai 2005.

Adresse des auteurs

David Meyer
mél : dmm@1-4-5.net

Keyur Patel
Cisco Systems
mél : keyupate@cisco.com

Déclaration complète de droits de reproduction

prétend avoir accompli aucun effort pour identifier de tels droits. Les informations sur les procédures de l'ISOC au sujet des droits dans les documents de l'ISOC figurent dans les BCP 78 et BCP 79.
Des copies des dépôts d'IPR faites au secrétariat de l'IETF et toutes assurances de disponibilité de licences, ou le résultat de tentatives faites pour obtenir une licence ou permission générale d'utilisation de tels droits de propriété par ceux qui mettent en œuvre ou utilisent la présente spécification peuvent être obtenues sur répertoire en ligne des IPR de l'IETF à http://www.ietf.org/ipr .
L'IETF invite toute partie intéressée à porter son attention sur tous copyrights, licences ou applications de licence, ou autres droits de propriété qui pourraient couvrir les technologies qui peuvent être nécessaires pour mettre en œuvre la présente norme. Prière d'adresser les informations à l'IETF à ietf-ipr@ietf.org .

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