RFC 7454 : BGP operations and security

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 19 février 2015

Date de publication du RFC : Février 2015

https://www.bortzmeyer.org/7454.html

Tout l'Internet repose sur le protocole BGP, qui permet l'échange de routes entre opérateurs Internet. (BGP est normalisé dans le RFC 4271 [1].) La sécurité de BGP est donc cruciale pour l'Internet, et elle a fait l'objet de nombreux travaux. Ce nouveau RFC résume l'état actuel des bonnes pratiques en matière de sécurité BGP. Du classique, aucune révélation, juste la compilation de l'état de l'art. Ce RFC porte aussi bien sur la protection du routeur, que sur le filtrage de l'information (les routes) reçues et transmises.

Ce genre de compilation aurait plutôt due être faite dans le cadre du project BCOP <http://bcop.nanog.org/> mais celui-ci semble mort.

La section 2 de ce RFC rappelle qu'il n'a pas de caractère obligatoire : il expose des pratiques de sécurité générales, et il est toujours permis de faire des exceptions, en fonction des caractéristiques spécifiques du réseau qu'on gère.

Donc, au début (sections 4 et 5 du RFC), la protection de la discussion entre deux routeurs, deux pairs BGP qui communiquent (sécurité du canal). Ensuite (sections 6 à 11), la protection des informations de routage échangées, le contrôle de ce qui est distribué (sécurité des données).

Commençons par sécuriser le routeur (section 4). Il devrait idéalement être protégé par des ACL qui interdisent les connexions vers le port 179, celui de BGP, sauf depuis les pairs connus. Les protections de TCP ne suffisent pas forcément, la mise en œuvre de TCP dans les routeurs est parfois faite de telle façon qu'on peut planter le routeur juste en envoyant plein de demandes de connexion. Il faut donc les jeter avant même que TCP ne les voit. De telles ACL sont parfois mise en place automatiquement par le logiciel du routeur, mais dans d'autres cas elles doivent être installées manuellement.

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc4271.txt

Rappelez-vous qu'un router a à la fois des fonctions de contrôle ("*control plane*", ce qui inclut BGP) et de transmission ("*data plane*"). Idéalement, les ACL pour protéger le contrôle devraient être spécifiques à cette fonction et ne pas affecter la transmission des paquets (mais le matériel et le logiciel ne permettent pas toujours cette séparation). Certains routeurs permettent également de mettre en place un limiteur de trafic, pour éviter du trafic excessif, même en provenance de pairs connus. Le RFC 6192 décrit avec plus de détails la protection des fonctions de contrôle d'un routeur.

Ensuite (section 5 du RFC), la protection des sessions BGP avec les pairs légitimes (cf. RFC 6952). Si les deux routeurs ne prennent aucune précaution, un attaquant pourrait, par exemple, couper leur session BGP en envoyant de faux paquets TCP de type `RST` (cf. RFC 5961). Pire, il pourrait, avec des techniques comme l'usurpation ARP, injecter de faux paquets dans une session BGP existante. Pour se protéger contre les attaques TCP, il faut utiliser une authentification TCP, comme la traditionnelle (et bien dépassée) TCP-MD5 du RFC 2385. Beaucoup d'opérateurs exigent une telle authentification lorsqu'on fait du BGP avec eux (particulièrement sur un point d'échange, où des inconnus peuvent facilement fabriquer de faux paquets). Mais on ne peut pas dire que 100 % des sessions BGP dans le monde sont protégées, en raison du surcoût d'administration qui en résulte (choisir les mots de passe, les distribuer, les changer, etc). En outre, MD5 étant désormais bien affaibli (RFC 6151), il faudrait désormais utiliser le mécanisme AO du RFC 5925. Le RFC note que, malgré le caractère antédiluvien de TCP-MD5, c'est toujours la solution la plus utilisée par les opérateurs. Mes lecteurs qui configurent tous les jours des appairages BGP connaissent-ils des gens qui utilisent AO ?

Une autre solution serait de se servir d'IPsec entre les routeurs mais personne ne le fait.

Autre précaution, filtrer les paquets IP en entrée du réseau de l'opérateur pour interdire les paquets prétendant avoir une adresse IP source située dans le réseau de l'opérateur. Sans cette précaution, même les sessions iBGP pourraient être attaquées.

Dernière protection des sessions BGP, GTSM (RFC 5082) qui consiste à tester que le TTL des paquets entrants est à la valeur maximale (255), et que le paquet vient donc bien du réseau local (s'il était passé par, ne serait-ce qu'un seul routeur, le TTL aurait été décrémenté).

Après avoir protégé les routeurs, et la session BGP sur TCP qui les relie, voyons les données. Sécuriser la session ne sert à rien si le pair légitime et authentifié nous envoie des informations fausses. La section 6 de notre RFC se consacre donc au filtrage des préfixes annoncés. D'abord, les préfixes non routables (ceux marqués "*Global : false*" dans le registre des adresses spéciales IPv4 <`https://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registry.xml`> ou son équivalent IPv6 <`https://www.iana.org/assignments/iana-ipv6-special-registry/iana-ipv6-spe`
`xml`>) devraient évidemment être rejetés. Mais il est également recommandé de ne pas accepter les préfixes non alloués (par le système d'allocation d'adresses IP IANA-¿RIR-¿LIR). Comme la liste de ces préfixes change tout le temps, les filtres **doivent** être mis à jour automatiquement, à partir de la liste à l'IANA <`https://www.iana.org/assignments/ipv6-unicast-address-assignments/ipv6-unicast-`
`xml`>. Comme il y a un délai entre l'allocation d'un préfixe à un RIR et son utilisation réelle sur ce terrain, il n'est pas nécessaire de tester tous les jours (le RFC recommande de tester au moins une fois par mois). Si, pour une raison ou pour une autre, on ne peut pas vérifier la liste en ligne, il vaut mieux ne pas filtrer les préfixes, plutôt que de le faire sur la base d'une liste dépassée : une des plaies de l'Internet est la nécessité de "*dé-bogoniser*" (faire retirer des listes de "*bogons*" ces listes d'adresses IP non allouées) tout nouveau préfixe, processus qui peut être lent et nécessite pas mal de tests sur les "*looking glasses*". En IPv4, il ne reste plus de préfixes non alloués <`https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml`> et ce test régulier n'est donc plus nécessaire.

Tester auprès de l'IANA ne permet que des filtres grossiers, éliminant les annonces de préfixes non alloués à un RIR, ce qui ne sert que pour IPv6, ne vérifie pas les préfixes plus spécifiques que ce que

l'IANA alloue, et n'empêche pas un malveillant ou un maladroit d'annoncer les préfixes d'un autre AS. Il peut donc être intéressant de filtrer de manière plus précise, en regardant les IRR. Un IRR est une base de données publiquement accessible, stockant les préfixes et l'AS autorisé à les annoncer (en langage RPSL, cf. RFC 4012). Ces IRR sont gérés par certains opérateurs, ou par les RIR. Par exemple, la base de données du RIPE-NCC contient cette information :

```
route:          217.70.176.0/20
descr:          GANDI is an ICANN accredited registrar
descr:          for more information:
descr:          Web:   http://www.gandi.net
origin:         AS29169
mnt-by:         GANDI-NOC
```

On voit ici que seul l'AS 29169 (Gandi) est autorisé à annoncer le préfixe `217.70.176.0/20`. En opérant récursivement (car un AS peut être un fournisseur d'un autre AS situé derrière lui et il faudra donc suivre les informations sur les AS relayés), on peut établir une liste de tous les préfixes qu'un pair peut annoncer, et ne pas accepter les autres. Des outils existent pour produire automatiquement des filtres sur le routeur à partir du RPSL (comme IRRToolSet <http://irrtoolset.isc.org>). Malheureusement, aucun IRR n'est parfait (et certains sont vraiment imparfaits) : préfixes absents (surtout les plus spécifiques, en cas de dés-agrégation des annonces), information dépassée, etc. Les IRR des RIR sont proches des opérateurs et donc a priori ont une information fiable mais ce n'est que théorique (les préfixes IP « du marais <https://www.bortzmeyer.org/nettoyage-marais.html> », alloués avant l'existence des RIR, sont une source particulièrement importante de problèmes). En outre, l'IRR d'un RIR ne couvre que la région de ce RIR, et on peut donc avoir besoin d'en consulter plusieurs (on a un pair états-unien, on regarde la base de l'ARIN, mais ce pair a des clients sud-américains et on doit donc aussi regarder la base de LACNIC...), ce qui justifie les IRR privés, comme RADB <http://www.radb.net>, qui essaient de consolider l'information des RIR.

Si vous trouvez que cette imperfection des IRR est bien ennuyeuse, le RFC recommande que vous agissiez de votre côté : vérifiez que **vos** préfixes sont correctement publiés dans les IRR.

Actuellement, la sécurité des données BGP repose essentiellement sur ce filtrage à partir des IRR et sur la réactivité des administrateurs réseau <https://www.bortzmeyer.org/securite-bgp-et-reaction-rapid
html>. Dans le futur, il est possible qu'un système plus fiable soit adopté et déployé, le couple RPKI+ROA <https://www.bortzmeyer.org/securite-routage-bgp-rpki-roa.html>, alias SIDR pour "*Secure Inter-Domain Routing*". SIDR repose sur une infrastructure de certification, la RPKI (RFC 6480), et sur des objets signés, les ROA (RFC 6482), annonçant quel AS peut annoncer tel préfixe. SIDR fournit deux services, dont seul le premier est un peu déployé aujourd'hui :

— La validation de l'origine de l'annonce (le premier AS sur le chemin). Décrite dans le RFC 6811, elle est aujourd'hui disponible dans la plupart des routeurs, et des ROA sont effectivement publiés.
— La validation du chemin d'AS complet. Surnommé « BGPsec » (RFC 7353 et RFC 7132), la normalisation technique de ce service est loin d'être complétée et il n'existe donc pas de mise en œuvre disponible.

Ces mécanismes SIDR devraient, une fois largement déployés, résoudre la plupart des problèmes décrits dans cette section 6 de notre RFC 7454. Mais cela prendra de nombreuses années et il est donc nécessaire de ne pas abandonner les méthodes actuelles comme les systèmes d'alarme <https://www.bortzmeyer.
org/alarmes-as.html>.

Pour la validation de l'origine de l'annonce, notre RFC recommande que la politique de filtrage des annonces (qui est une décision locale de chaque routeur) suive les règles du RFC 7115. Pour les résumer, lorsqu'une annonce BGP est comparée au contenu de la RPKI :

— S'il existe un ROA et que l'annonce est valide selon ce ROA, on accepte l'annonce,

— S'il existe un ROA mais que l'annonce n'est pas valide, on rejette l'annonce (attention, rappelez-vous qu'au début de toute nouvelle technique de sécurité, il y a pas mal de faux positifs),
— S'il n'existe pas de ROA, on accepte l'annonce, avec une préférence plus faible.

Le système RPKI+ROA pose de nouveaux et intéressants problèmes et il est donc recommandé de lire « "*On the Risk of Misbehaving RPKI Authorities*" <http://www.cs.bu.edu/~goldbe/papers/hotRPKI.pdf> » d'abord.

D'autres filtrages sont possibles en entrée. Par exemple, les opérateurs filtrent les annonces trop spécifiques, afin notamment d'éviter la croissance indéfinie de leurs bases de données et tables de routage. Chacun choisit les valeurs quantitatives précises et il n'y a pas de consensus documenté sur ce point (on peut consulter les documents RIPE-399 <https://www.ripe.net/ripe/docs/ripe-399> et RIPE-532 <https://www.ripe.net/ripe/docs/ripe-532>) mais on peut observer qu'un préfixe plus long que /24 en IPv4 et /48 en IPv6 a très peu de chances d'être accepté dans l'Internet. Voici un exemple de filtrage IPv4 sur JunOS :

```
policy-statement no-small-and-big-prefixes {
        from {
            route-filter 0.0.0.0/0 prefix-length-range /25-/32 reject;
            route-filter 0.0.0.0/0 prefix-length-range /0-/7 reject;
        }
}
protocols {
            bgp {
               ...
               import no-small-and-big-prefixes;
```

Typiquement, on filtre aussi en entrée les annonces portant sur les préfixes internes. Normalement, ce n'est pas à nos voisins d'annoncer nos routes !

Autres préfixes souvent filtrés, les routes par défaut, 0.0.0.0/0 en IPv4 et ::0/0 en IPv6.

Naturellement, les recommandations de filtrage dépendent du type d'appairage BGP : on ne filtre pas pareil selon qu'on parle à un pair, à un client ou à un transitaire (voir la section 6 du RFC pour tous les détails). Ainsi, pour reprendre le paragraphe précédent, sur la route par défaut, certains clients d'un opérateur demandent à recevoir une telle route et c'est tout à fait acceptable.

La section 7 est consacrée à une pratique très utilisée et très discutée, l'amortissement ("*damping*"). Lorsqu'une route vers un préfixe IP donné passe son temps à être annoncée et retirée, on finit par l'ignorer, pour éviter que le routeur ne passe son temps à recalculer ses bases de données. Pour réaliser cela, à chaque changement d'une route, on lui inflige une pénalité, et au bout d'un certain nombre de pénalités, on retire la route. Malheureusement, cette technique mène parfois à supprimer des routes et à couper un accès (voir RIPE-378 <https://www.ripe.net/ripe/docs/ripe-378>). Avec de meilleurs paramètres numériques, comme recommandé par le RFC 7196 et RIPE-580 <https://www.ripe.net/ripe/docs/ripe-580>, l'amortissement redevient utilisable et recommandable.

Autre technique de filtrage des erreurs, décrite en section 8, l'imposition d'un nombre maximum de préfixes annoncés par un pair BGP. S'il en annonce davantage, on coupe la session BGP. Un dépassement du nombre maximal est en effet en général le résultat d'une fuite, où, suite à une erreur de configuration, le routeur ré-annonce des routes reçues d'un autre. Parfois, c'est toute la DFZ qui est ainsi annoncée par accident aux pairs ! Notre RFC demande donc instamment qu'on limite le nombre de préfixes accepté pour une session BGP. Pour un pair sur un point d'échange, le seuil devrait être inférieur au nombre de routes de la DFZ (dans les 530 000 en IPv4 aujourd'hui, et 21 000 en IPv6), pour détecter les annonces accidentelles de toute la DFZ. On peut aussi avoir des seuils par pair, fondés sur le nombre de routes qu'ils sont censés annoncer. Pour un transitaire, par contre, le seuil doit être plus élevé que le nombre de routes dans la DFZ, puisqu'on s'attend à tout recevoir d'eux (mais une valeur maximale est quand même utile en cas de désagrégation intensive). Comme l'Internet change tout le temps, il faut réviser ces limites, et suivre les alertes (sur certains routeurs, on peut configurer deux seuils, un pour déclencher une alerte, et un autre, supérieur, pour réellement couper la session). Voici un exemple sur JunOS :

```
group Route-Server-LINX-V4 {
    family inet {
        unicast {
            prefix-limit {
               maximum 100000;
            }
}
}
```

Après le filtrage par préfixe, il peut aussi y avoir du filtrage par chemin d'AS (section 9 de notre RFC) et par routeur suivant ("*next hop*", section 10). Voyons d'abord le filtrage par chemin d'AS. Par exemple, un client d'un opérateur ne devrait pas annoncer des routes avec n'importe quels AS mais seulement avec un chemin comportant uniquement son propre AS (et, si le client a lui-même des clients, avec l'AS de ces clients secondaires). Si l'opérateur n'arrive pas à avoir une liste complète des AS qui peuvent se retrouver dans les chemins de ses clients, au moins peut-il limiter la longueur de ces chemins, pour éviter la ré-annonce accidentelle de routes. D'autre part, on n'accepte pas, dans le cas normal, de routes où un AS privé (64512 à 65534 et 4200000000 à 4294967294, voir RFC 6996) apparait dans le chemin. Conséquence logique, on n'annonce pas à ses voisins de routes avec des chemins d'AS qui incluent un AS privé, sauf arrangement spécifique. Et le chemin d'AS dans une annonce BGP doit toujours commencer par l'AS du voisin (sauf si on parle à un serveur de routes). Enfin, un routeur BGP n'acceptera pas d'annonces où il voit son propre AS dans le chemin, et ce comportement par défaut ne devrait pas être débrayé.

Quant au filtrage sur le routeur suivant (section 10 du RFC), il consiste à refuser une route si l'attribut BGP `NEXT_HOP` (RFC 4271, section 5.1.3) n'est pas l'adresse du voisin. Attention, ce test doit être débrayé si on parle à un serveur de routes, celui-ci ne souhaitant pas traiter les paquets IP. Idem (débrayage du test) si on fait du RTBH (RFC 6666).

Pour finir, je recommande trois lectures,

— Le Rapport sur la résilience de l'Internet en France <`https://www.ssi.gouv.fr/agence/rayonnement-scientifique/lobservatoire-de-la-resilience-de-linternet-francais/`> (ANSSI/AFNIC), qui contient plein d'études intéressantes sur les chiffres BGP, par exemple sur le nombre d'annonces BGP invalides (possibles détournements) observées,
— Le guide des bonnes pratiques BGP <`https://www.ssi.gouv.fr/administration/guide/le-guide-des-bonnes-pratiques-de-configuration-de-bgp/`> de la même ANSSI,
— La conf' BGP de Sarah Nataf <`http://www.iletaitunefoisinternet.fr/bgp-sarah-nataf/`>.