RFC 6888 : Common requirements for Carrier Grade NATs (CGNs)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 1 mai 2013

Date de publication du RFC : Avril 2013

https://www.bortzmeyer.org/6888.html

Qu'on les approuve (ce qui n'est pas mon cas) ou pas, les CGN, ces gros routeurs NAT installés dans le réseau du FAI et gérant des centaines ou des milliers de clients, sont d'ores et déjà une réalité douloureuse dans de nombreux pays d'Asie, pour les pauvres utilisateurs de la 3G, et peut-être demain sur d'autres continents <http://www.pcpro.co.uk/news/broadband/381646/customers-fume-as-bt-introduc pour les accès fixes. Pour limiter les dégâts qu'ils causent, ce RFC 6888 [1] pose un certain nombre de règles que **devraient** respecter ces CGN.

D'abord, qu'est-ce qu'un CGN ? Fonctionnellement, il ressemble au petit routeur NAT (RFC 2663) que vous avez chez vous. Comme lui, il fait du NAPT ("*Network Address and Port Translation*"), c'est-à-dire qu'il traduit dynamiquement des adresses IP du réseau interne vers celles utilisées à l'extérieur (et vice-versa pour les paquets dans l'autre sens). La grosse différence avec le petit routeur ou "*box*" chez vous est qu'il travaille pour de nombreux clients du FAI. Au lieu que l'adresse IP externe soit partagée uniquement entre les membres de votre cercle de famille, elle sera partagée avec des centaines d'inconnus.

Avec le CGN, il y a souvent « double NAT », une fois dans le CPE et une fois dans le CGN. Mais ce n'est pas obligatoire. Lorsqu'il y a double NAT, on parle souvent de NAT444 <https://www.bortzmeyer.org/nats.html> (deux traductions, d'IPv4 en IPv4).

Sur cette image, on voit du double NAT. Un FAI gère un CGN. Les adresses IP publiques du FAI, allouées par son RIR sont ici 198.51.100.0/25. Ce sont celles qui seront vues, par exemple, par les sites Web où se connectent les clients du FAI. Le FAI n'a pas assez d'adresses IP publiques pour son réseau interne et il utilise donc le 10.0.0.0/8 du RFC 1918. Prenons maintenant le client 1 : son réseau

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6888.txt

local a des adresses dans la plage 192.168.3.0/24. Les paquets seront donc émis avec ces adresses, NATés une première fois par la machine CPE 1, vers l'adresse 10.0.5.22. Ils seront ensuite NATés une seconde fois par le CGN.

Le CGN a des conséquences, par exemple dans le domaine légal (en théorie, vous pourrez voir les agents de la HADOPI débarquer chez vous parce qu'un autre utilisateur du même CGN a téléchargé illégalement, cf. RFC 6269). Et cela limite sérieusement les activités que vous pourrez avoir sur l'Internet. Par exemple, un moyen courant d'accepter les connexions entrantes, normalement interdites par le NAPT, est de configurer sa "*box*" pour rediriger tel numéro de port vers tel service sur le réseau local (par exemple pour faire fonctionner des applications pair-à-pair <https://www.bortzmeyer.org/emule-ports-linux.html>). Le routeur CGN étant partagé entre de nombreuses personnes qui ne se connaissent pas, cela n'est plus possible. Contrairement au routeur NAT à la maison, le CGN n'est **pas** géré par les abonnés, qui ne peuvent pas modifier sa configuration.

D'autre part, la simple taille du CGN en fait un engin compliqué et fragile, et sa panne affecte bien plus de clients que le redémarrage d'une "*box*". Enfin, ayant moins de ports sources à sa disposition par client, par rapport au routeur de la maison ou du SOHO, il risque plus souvent de tomber à cours, si les clients font beaucoup de connexions sortantes.

Mais, alors, pourquoi met-on des CGN ? Juste pour embêter les clients ? En fait, les FAI n'ont pas forcément beaucoup le choix. La pénurie d'adresses IPv4 <https://www.bortzmeyer.org/epuisement-adress
html>, bien que niée par beaucoup de négationnistes, est réelle depuis de nombreuses années. Par exemple, cela fait longtemps qu'on ne peut plus avoir une adresse IP par machine à la maison. Mais, jusqu'à récemment, on pouvait encore avoir une adresse IP publique par foyer. Désormais, l'espace d'adressage IPv4 disponible se resserrant chaque jour, ce n'est même plus le cas. S'il n'y a pas d'adresse IPv4 publique disponible pour chaque client, le CGN est la seule solution pour pouvoir continuer à faire de l'IPv4 (passer à IPv6 ferait moins de travail mais les acteurs du marché sont frileux).

Cela, c'était la raison avouable pour mettre des CGN. Il y en a une moins avouable, c'est que le CGN, grosse machine centrale et point de passage obligatoire pour tous les abonnés, est bien en phase avec la mentalité des opérateurs telco traditionnels. C'est ainsi que, dans l'Internet mobile où les libertés sont bien plus limitées, et l'utilisateur bien plus restreint dans ce qu'il a le droit de faire, tous les opérateurs 3G ont déployé cette solution depuis longtemps, refusant de donner une adresse IPv4 publique à chaque abonné, même avant l'épuisement complet des réserves. Comme le note prudemment le RFC, « "*there are driving forces other than the shortage of IPv4 addresses*" ».

À noter qu'un RFC décrit le déploiement de CGN comme une des techniques pouvant aider à la transition vers IPv6 : RFC 6264. Le CGN est un des composants de la solution DS-Lite, décrite dans le RFC 6333 et ce composant doit donc suivre les règles données plus loin.

Maintenant qu'on a présenté les CGN, que contient ce RFC ? Il liste des conseils qui, s'ils étaient appliqués, rendraient les CGN un peu moins insupportables. Comme le note la section 1, la publication de ce RFC ne signifie **pas** que l'IETF approuve ou même accepte les CGN, simplement que, puisqu'ils existent, autant essayer de limiter leurs effets négatifs. Comme un CGN est un routeur NAT, les précédents documents du groupe de travail Behave <http://tools.ietf.org/wg/behave>, concernant tous les types de NAT, s'appliquent aussi au CGN : RFC 4787, RFC 5382 et RFC 5508 notamment.

Les section 3 à 5 représentent donc le cœur de ce RFC. Elles contiennent les recommandations spécifiques pour les CGN. Chacune est nommée REQ-n où n est un numéro d'ordre. Ainsi, REQ-1 dit

que, pour chaque protocole de transport, le CGN doit suivre les recommandations spécifiques à ce protocole (RFC 4787 pour UDP, RFC 5382 pour TCP, RFC 5508 pour ICMP, et RFC 5597 pour DCCP). Notons que les autres protocoles de transport (comme SCTP) ont donc de sérieuses chances de ne pas fonctionner au travers du CGN.

Exigence suivante, REQ-2, qui dit que le comportement de l'"*IP address pooling*" doit être "*paired*". Cela signifie quoi ? Qu'une machine donnée doit toujours obtenir la même adresse IP externe (un CGN, contrairement au petit routeur NAT de la maison, a en général plusieurs adresses IP externes à sa disposition). Cela limite les risques de casser les applications. Le RFC 4787, section 4.1, donne les détails sur ce concept de "*paired*".

À propos du nombre d'adresses externes disponible pour le CGN : REQ-3 exige que ces adresses puissent être en nombre quelconque et dans des plages non-contigües (avec l'épuisement des adresses IPv4, il sera de plus en plus dur de trouver des plages contigües).

REQ-4 traite un problème spécifique aux CGN, que n'avaient pas les petits routeurs NAT de la maison : il faut pouvoir limiter le nombre de ports utilisables par un utilisateur donné. Dans le NAPT, il y a moins d'adresses externes que d'adresses internes. On se sert donc des ports pour désambigüiser. Un seul utilisateur gourmand pourrait lancer plein de sessions et épuiser à lui seul la plage de ports disponibles (cela peut même être volontaire, pour faire un déni de service, cf. section 8). C'est pour se protéger contre cette attaque que REQ-4 exige une limitation, configurable, par client. Le RFC recommande aussi qu'on puisse limiter le rythme de création de sessions par utilisateur. C'est nécessaire pour protéger les autres utilisateurs, mais c'est un bon exemple des inconvénients du CGN, qui casse sérieusement la transparence du réseau et le principe de bout en bout. Les routeurs NAT cassaient ces principes mais uniquement entre membres d'un même foyer ou d'un même SOHO. Avec le CGN, on dépend de gens dont on ne connait rien.

Même chose avec REQ-5, consacrée à d'autres ressources partagées dans le CGN, comme sa mémoire. Un CGN a des problèmes d'équité que n'ont pas les routeurs NAT classiques, vue son utilisation entre personnes qui n'ont rien en commun a priori.

Parfois, les cibles que vise l'utilisateur sont situées dans le réseau du FAI, et donc atteignables sans traduction d'adresse. Pour ce cas, REQ-6 demande que le CGN puisse être configuré pour ne pas traduire si la destination est dans une liste d'adresses configurées par l'administrateur réseaux.

Lorsqu'un routeur NAT accepte des paquets entrants du moment qu'il a, dans sa table des traductions, une entrée pour ce couple {adresse,port} de destination, indépendemment de l'adresse IP source d'émission, on dit qu'il fait du "*Endpoint-Independent Filtering*" (RFC 4787, section 5). REQ-7 souhaite que ce soit le comportement du CGN, car il casse moins d'applications (notamment jeux en ligne et pair-à-pair) que le "*Address-Dependent Filtering*" (où les paquets sont refusés si l'adresse source n'est pas celle dans l'entrée de la table).

REQ-8 légifère sur la réutilisation d'un port externe lorsqu'il n'est plus utilisé. Par défaut, il devrait rester réservé pendant deux minutes, le "*Maximum Segment Lifetime*" de TCP (RFC 793, section 3.3), après lequel on est sûr que le paquet ne traîne plus dans le réseau. Cela permet d'éviter une collision entre une ancienne correspondance et une nouvelle, qui utiliserait le même port externe et recevrait par erreur de vieux paquets. Des exceptions sont prévues, notamment :

— Si le CGN met en œuvre la machine à états de TCP, il peut savoir quand une connexion est terminée et réutiliser alors tout de suite le port,
— Si certains ports sont statiquement affectés, ils restent utilisables en permanence.

REQ-9 est plus ambitieux car il exige quelque chose qui n'existe pas dans les CGN actuels : un mécanisme permettant à l'utilisateur de changer les affectations de port, et que ce mécanisme soit de préférence PCP ("*Port Control Protocol*", RFC 6887). Aujourd'hui, sur les petits routeurs NAT de la maison ou du SOHO, ce mécanisme d'affectation manuelle des correspondances {adresse externe, port externe} -¿ {adresse interne, port interne} est typiquement fait par l'interface Web d'administration du routeur, ou bien par UPnP. La disparition d'un tel mécanisme dans les CGN actuels est un des plus gros problèmes que posent les CGN. Il était donc nécessaire de le rétablir, pour que toutes les applications puissent bien fonctionner. Mais PCP est encore très récent et peu déployé.

REQ-10 concerne plutôt les besoins de l'opérateur plus que ceux des simples utilisateurs : il demande que les routeurs CGN soient gérables (MIB du RFC 4008, etc).

Une autre exigence, REQ-11 concerne le traitement des erreurs. Si un routeur CGN ne peut pas créer une corrrespondance entre adresse+port externe et interne, il doit jeter le paquet et devrait renvoyer un message ICMP "*Destination unreachable*" / code 1 ("*Host unreachable*"). et déclencher une "*trap*" SNMP. Le « devrait » (et pas « doit ») est parce que ces deux types de messages peuvent légitimement avoir un débit limité.

Pourquoi "*Host unreachable*" et pas "*Communication administratively prohibited*" comme c'était le cas dans le RFC 5508, section 6 ? Parce que le code 1 ("*Host unreachable*") est une erreur « douce » (cf. RFC 1122) et ne coupera pas les connexions TCP en cours entre les mêmes machines (RFC 5461).

En outre, le RFC interdit de supprimer des entrées existantes dans la table de correspondances, ce qui permettrait pourtant de faire de la place pour de nouvelles connexions. C'est parce que la plupart des applications gèrent beaucoup mieux une impossibilté à établir une connexion qu'une perturbation ou une interruption d'une connexion en cours (RFC 5508, section 6).

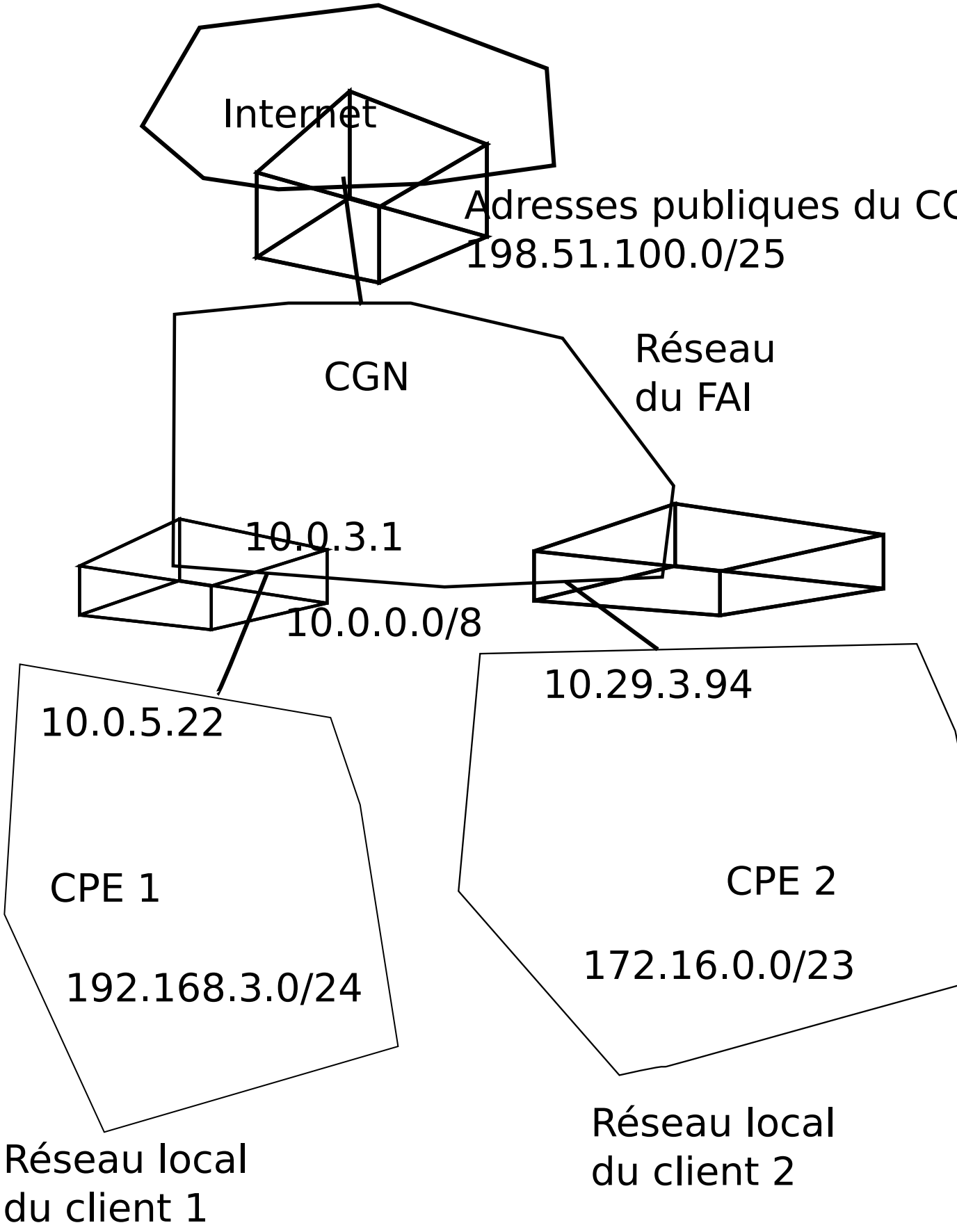
Un bon routeur CGN devrait pouvoir journaliser son activité. La section 4 encadre cette tâche. La principale question qui se pose est l'identification des clients, si une adresse IP externe est repérée pour son comportement abusif (envoi de spam, écriture d'un commentaire de blog défavorable aux autorités, partage de fichiers musicaux, et autres crimes). Le site distant va noter que « 192.0.2.42 a fait quelque chose de mal » et les enquêteurs vont chercher la personne qui est derrière cette adresse. En raison du CGN, des dizaines ou des centaines de personnes peuvent avoir utilisé cette adresse. Pour que l'enquêteur ait la moindre chance de retrouver le coupable, il faut que le site distant journalise, non seulement l'adresse IP source (comme c'est typiquement le cas aujourd'hui), mais aussi le port source <https://www.bortzmeyer.org/loguer-adresse-et-port.html> (cf. RFC 6302). **Et** il faut que le routeur CGN ait journalisé ses correspondances, protocole de transport, adresse+port internes, adresse+port externes et heure exacte. (C'est une obligation légale dans certains pays, par exemple en France avec la LCEN.)

Non seulement c'est très Big Brother comme comportement mais, en plus, cela peut rapidement produire une énorme quantité de données, ce qui ne va pas faciliter la tâche de l'opérateur. C'est pour cela que REQ-12 demande que ce ne soit **pas** activé par défaut, à la fois pour préserver la vie privée des utilisateurs et pour éviter de remplir les disques durs. (Notez aussi que le RFC ne conseille pas de loguer l'adresse IP de **destination**, ce qui serait bien pire.)

Et, au fait, comment allouer aux utilisateurs cette ressource limitée, les numéros de ports externes ? La section 5 formule trois exigences. Le problème est que les trois sont contradictoires. REQ-13 dit qu'il faudrait chercher à maximiser l'utilisation des ports (ils sont en nombre limités, 65 536 par adresse IP), REQ-14 (qui a suscité beaucoup de discussions dans le groupe de travail) qu'il faudrait minimiser le nombre d'entrées produites dans le journal, par exemple en allouant toute une plage de ports par adresse

IP, et REQ-15 que, pour des raisons de sécurité, il faudrait rendre difficile pour un attaquant la prédiction des ports sélectionnés (RFC 6056). Pas de miracle, ces trois exigences ne peuvent pas être satisfaites simultanément et un routeur CGN doit donc choisir lequel est le plus important pour lui. J'emprunte à Simon Perreault une description de la situation en mai 2013 : « Le CGN de Cisco allouait les ports 100 % dynamiquement, comme un NAT traditionnel, et génèrait donc beaucoup de logs. Mais ceux d'autres vendeurs, dont Alcatel, Juniper et Huawei, [ainsi que Cisco désormais, note de Nicolas Février] allouent les ports en bloc, ce qui génère beaucoup moins de logs au prix d'une légère perte de taux d'utilisation des adresses IPv4 externes. Il y a aussi le "*deterministic NAT*" qui alloue les ports statiquement tout en conservant une plage d'allocation dynamique pour les débordements. »

Merci à André Sintzoff pour la relecture technique et la détection d'une grosse faute technique. Merci à Nicolas Février pour ses utiles commentaires, même s'il n'est pas d'accord avec moi sur les défauts du CGN.


Internet
Adresses publiques du CG
198.51.100.0/25
CGN
Réseau
du FAI
10.0.3.1
10.0.0.0/8
10.29.3.94
10.0.5.22
CPE 1
192.168.3.0/24
CPE 2
172.16.0.0/23
Réseau local
du client 1
Réseau local
du client 2