RFC 7215 : LISP Network Element Deployment Considerations

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 23 avril 2014

Date de publication du RFC : Avril 2014

https://www.bortzmeyer.org/7215.html

Le protocole réseau LISP a été normalisé il y a plus d'un an et plusieurs déploiements de taille significative ont déjà eu lieu. La norme LISP n'impose pas un modèle unique de déploiement et laisse bien des choix à la discrétion de l'administrateur réseaux sur le terrain, notamment pour le placement des différents éléments qui composent un réseau LISP. Ce nouveau RFC fait le point sur les différents modèles de déploiement possibles. Il est donc orienté vers les opérationnels, ceux (et celles) qui vont déployer LISP en production.

Petit rappel, LISP vise à résoudre le problème de la croissance de la table de routage globale de l'Internet (problème décrit dans le RFC 4984 [1]), par le moyen d'une séparation entre **identificateurs**, les EID ("*Endpoint IDentifiers*") et les **localisateurs**, les RLOC ("*Routing LOCators*"). LISP est normalisé dans le RFC 6830. À l'heure actuelle, les RFC sur LISP ont le statut « expérimental » (y compris ce RFC 7215), reflétant le caractère assez disruptif de ce protocole. Il n'y a notamment aucune solution aux différents problèmes de sécurité étudiés dans la section 15 du RFC 6830.

Un principe essentiel de LISP est la distinction faite entre les réseaux de bordure, qui ne travaillent qu'avec des EID, et le cœur de l'Internet qui ne travaille qu'avec des RLOC. Mais où placer la frontière, frontière d'autant plus importante que c'est là où vont devoir se situer les routeurs LISP (les routeurs de bordure et du cœur sont, eux, des routeurs IP non modifiés) ? L'idée initiale était que la frontière était aux limites des AS « feuille », ceux qui n'ont pas de trafic de transit. Mais, en fait, LISP permet plusieurs choix. Un « site LISP » peut être un AS feuille mais peut aussi être un simple réseau local.

LISP tunnelise les paquets d'un site LISP à l'autre, à travers le cœur. Le routeur LISP, connecté à la fois à la bordure et au cœur se nomme un ITR ("*Ingress Tunnel Router*") à l'entrée (encapsulation des

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc4984.txt

paquets) et un ETR ("*Egress Tunnel Router*") à la sortie (décapsulation des paquets). Quand on veut parler des deux types en même temps, on dit juste « un xTR ».

Passons maintenant aux scénarios, section 2.1. Premier scénario possible de déploiement, le « "*Customer Edge*" ». Le routeur LISP est dans les locaux du client, contrôlé par lui, et sur la frontière entre réseau du client et réseau de l'opérateur. Ce sera a priori le cas le plus fréquent chez les LISPiens et c'est la solution recommandée par notre RFC. A priori, si on déploie LISP, c'est qu'on a un réseau de grande taille, "*multi-homé*", et qu'on souhaite faire de la gestion de trafic (faire entrer le trafic Internet préferentiellement par un des opérateurs, par exemple). Dans le scénario "*Customer Edge*", le client a le complet contrôle des xTR et peut déployer les politiques qu'il veut sans rien demander à personne. L'information de joignabilité des ETR, si importante en LISP, peut être maintenue correctement, tous les routeurs LISP étant sous le contrôle de la même organisation. Les "*Locator Status Bits*" mis par l'ITR sont également toujours conformes à la réalité.

Autre avantage, le réseau interne, son plan d'adressage, son protocole de routage et ses régles de filtrage ne changent pas.

Comme toutes les techniques utilisant des tunnels, LISP est très sensible aux problèmes de MTU (RFC 4459), le tunnel consommant quelques octets qui vont diminuer la MTU du lien. Si la connexion entre le client et son opérateur a une MTU supérieure aux traditionnels 1 500 octets d'Ethernet, il n'y a pas de problème. Sinon, ce scénario peut entraîner des problèmes de MTU.

Second scénario possible, le « "*Provider Edge*" ». Cette fois, on met les xTR chez l'opérateur, sous son contrôle. Le client n'a pas à mettre à jour ses routeurs, ou à les configurer. Et, avec un seul xTR, l'opérateur peut servir plusieurs clients LISP.

Par contre, cela fait perdre à LISP une de ses principales qualités, la possibilité de contrôler la répartition du trafic entrant ("*ingress traffic engineering*"). En effet, les xTR ne sont plus sous le contrôle du client. Et si le client est "*multi-homé*", c'est encore pire, les xTR étant chez des opérateurs concurrents.

Les xTR qui servent un site LISP n'étant plus coordonnés, ils ne vont pas forcément avoir de l'information correcte et à jour sur la joignabilité, et ne pourront donc pas servir cette information.

Par contre, ce scénario peut limiter les risques d'un problème de MTU, les xTR étant directement dans le réseau de l'opérateur, où la MTU disponible est souvent plus grande.

Mais on peut commettre des perversions bien pires avec LISP. Par exemple, on peut mettre les xTR derrière des routeurs NAT, par exemple si on n'a pas assez d'adresses IPv4. Dans ce troisième scénario, l'ITR encapsule les paquets qui sont ensuite NATés. Attention, les paquets LISP "*Map Requests*" n'auront alors pas le même en-tête IP que les "*Map Reply*" correspondants, qui seront alors jetés par le routeur NAT. Il faudra une configuration explicite (par exemple diriger tous les paquets UDP de port source 4342 vers l'ITR) pour éviter cela.

Ce RFC ne s'arrête pas à ces trois scénarios possibles. Il décrit aussi comment les fonctions LISP peuvent être réparties dans divers équipements. Par exemple, les fonctions d'ITR et d'ETR ne sont pas forcément présentes dans le même boîtier. On a le droit de les placer dans deux routeurs différents, par exemple pour mettre les ITR très près des machines terminales, afin d'encapsuler le plus tôt possible.

Comme toutes les solutions de séparation de l'identificateur et du localisateur, LISP dépend énormément des mécanismes de correspondance (RFC 9301), permettant de trouver un localisateur (RLOC) lorsqu'on

connait l'identificateur (EID). Pour résoudre le problème de "*bootstrap*", les serveurs de correspondance doivent être désignés par des RLOC uniquement. Le système de correspondance a deux composants, les "*Map Servers*" et les "*Map Resolvers*". Voyons d'abord les premiers (section 3.1).

Les "*Map Servers*" apprennent la correspondance EID-¿RLOC de leurs ETR (message "*Map Register*") et publient ensuite cette information, par exemple via le protocole ALT (RFC 6836) ou via le protocole DDT (RFC 8111). ALT s'inspire plutôt de BGP, DDT plutôt du DNS. Les gérants des "*Map Servers*" se nomment les MSP ("*Mapping Service Provider*"). Ils peuvent être opérateurs réseaux ou bien des organisations spécialisées dans le service de correspondance, se faisant payer pour publier un préfixe EID (section 5.2). A priori, les bonnes pratiques existantes pour la gestion de BGP s'appliquent à ALT, et celles pour le DNS à DDT. Mais notre RFC ne va pas plus loin : il est encore trop tôt pour graver dans le marbre de strictes politiques pour les gérants de "*Map Servers*".

DDT (RFC 8111) est arborescent et repose donc sur une racine `<http://ddt-root.org/>`. Cette racine est actuellement gérée par plusieurs organisations volontaires. Comme pour la racine du DNS, elle est servie par plusieurs serveurs, chacun géré par une organisation différente, et désignés par une lettre de l'alphabet grec. Au moins un de ces serveurs est "*anycasté*". La racine doit normalement vérifier que les organisations qui enregistrent un EID sont autorisés à le faire, via un RIR qui leur a attribué le préfixe en question.

Et les "*Map Resolvers*" (section 3.2) ? Leur travail (RFC 9301) est de recevoir des requêtes "*Map Request*", typiquement envoyées par un ITR, de trouver une correspondance EID-¿RLOC dans la base de données distribuée, et de la renvoyer au demandeur. (Les habitués du DNS peuvent se dire qu'un "*Map Server*" est un « serveur faisant autorité » et un "*Map Resolver*" un « résolveur » ou « serveur récursif ».) Vu leurs « clients », les "*Map Resolvers*" ont tout intérêt à être situés près des ITR qu'ils servent.

Les ITR vont devoir être configurés avec les adresses de leurs "*Map Resolvers*". Un préfixe "*anycast*" (RFC 4786) commun faciliterait cette tâche, l'ITR trouvant ainsi automatiquement le résolveur le plus proche et donc en général le plus rapide.

Comme toute technologie nouvelle sur le réseau, LISP doit faire face au problème de la coexistence avec les anciens systèmes. Aujourd'hui, il y a peu de sites LISP. Ceux-ci doivent donc se poser la question de la coexistence avec les sites non-LISP. Plusieurs techniques sont envisagées pour cela, comme les P-ITR, "*Proxy ITR*" du RFC 6832. Un site LISP qui veut envoyer des paquets à un site non-LISP peut le faire simplement en n'utilisant pas l'encapsulation LISP. Par contre, pour en recevoir, il **doit** déployer une technique comme le P-ITR.

Et puisqu'on parle de coexistence avec les sites non-LISP, il faut aussi envisager le processus de migration provisoire depuis l'état actuel vers LISP (section 5 et annexe A). Le RFC est ambitieux, partant de l'état initial (peu de sites LISP) en allant vers un état intermédiaire où il y aurait une majorité de sites LISP, pour terminer par un Internet complètement LISPisé.

Au début, un site LISP n'a pas le choix. Sauf à ne communiquer qu'avec les autres sites LISP, il a intérêt à annoncer ses préfixes dans le "*Map system*" LISP mais aussi en BGP, pour que les sites non-LISP sachent où le trouver. On notera donc que, dans cette situation, LISP ne contribue guère à la réduction de la table de routage globale, puisque tous les réseaux doivent toujours être publiés dans BGP. Heureusement, au fur et à mesure que le nombre de sites LISP augmente, il y aura de moins en moins besoin d'utiliser les techniques de "*traffic engineering*" pour contrôler les flux de données. Comme ces techniques (par exemple la désagrégation des préfixes) sont largement responsables de la croissance de la table de routage globale, LISP aura donc déjà un intérêt concret à ce stade (encore lointain).

On l'a vu, l'inconvénient de cette méthode (annoncer les préfixes en BGP) est qu'on ne diminue pas la taille de la DFZ. Pire, si on est un nouveau réseau sans infrastructure BGP existante, on augmente cette taille. Pour ces réseaux neufs, notre RFC recommande donc plutôt de ne pas déployer BGP du tout et d'utiliser les P-ITR. C'est alors le titulaire d'un préfixe englobant qui annonce le préfixe et le route vers les ETR du client. Ainsi, il n'y a pas de désagrégation des préfixes. Par contre, l'opérateur qui fournit ce service doit router tout le trafic non-LISP du client, ce qui n'est pas forcément raisonnable si c'est un gros client (comparez cela aux tunnels IPv6 que fournissent gratuitement des opérateurs comme Hurricane Electric : cela n'est réaliste que si le trafic est faible). Le problème disparaitra petit à petit si LISP se développe, lorsqu'il n'y aura plus que des reliquats non-LISP (on en est très loin).

L'annexe A du RFC décrit un plan de migration concret pour les responsables opérationnels, sous forme d'une liste d'étapes, avec des points à vérifier à chaque étape :

— Faire un état des lieux quantitatif du réseau, pour avoir combien de paquets par seconde et de bits par seconde il faudra acheminer.
— Vérifier les capacités LISP des routeurs existants : peuvent-ils être utilisés ou bien va t-il falloir en acheter d'autres ?
— Faire bien attention aux questions de MTU, LISP utilisant des tunnels. Si tous les liens externes peuvent accepter une MTU de 1 556 octets, parfait. Dans tous les cas, testez que cela passe et qu'il n'y a pas un pare-feu trop zélé qui bloque les messages ICMP indispensables à la découverte de la MTU du chemin.
— Vérifiez que vos préfixes IP sont utilisables pour LISP. Si c'est du PI, pas de problème.
— Configurez les routeurs LISP.
— Testez la joignabilité des ETR (par exemple avec ping) et l'enregistrement des préfixes avec lig (RFC 6835).
— etc (la liste est longue...)