Groupe de travail Réseau
Request for Comments : 2671
Catégorie : En cours de normalisation

P. Vixie, ISC
août 1999
Traduction Claude Brière de L'Isle

Mécanismes d'extension pour le DNS (EDNS0)

Statut du présent mémoire
Le présent document spécifie un protocole de l'Internet en cours de normalisation pour la communauté de l'Internet, et appelle à des discussions et suggestions pour son amélioration. Prière de se référer à l'édition en cours des "Normes officielles des protocoles de l'Internet" (STD 1) pour connaître l'état de la normalisation et le statut de ce protocole. La distribution du présent mémoire n'est soumise à aucune restriction.

Notice de Copyright
Copyright (C) The Internet Society (1999). Tous droits réservés.

Résumé
Le protocole de câblage du système des noms de domaines comporte un certain nombre de champs fixés dont la gamme est épuisée ou sur le point de l'être ce qui ne permet pas aux clients d'annoncer leurs capacités aux serveurs. Le présent document décrit les mécanismes de rétrocompatibilé pour permettre la croissance du protocole.

1. Motifs et portée

1.1 Le DNS (voir la [RFC1035]) spécifie un format de message et au sein de tels messages il y a des formats standard pour les options de codage, les erreurs, et la compression de nom. La taille maximum admissible d'un message DNS est fixée. Beaucoup des limites du protocole du DNS sont trop petites pour les utilisations courantes ou dont on désire qu'elles deviennent courantes. Les mises en œuvre n'ont aucun moyen d'annoncer leurs capacités.

1.2 Les clients existants ne vont pas savoir comment interpréter les extensions de protocole exposées ici. En pratique, ces clients seront mis à niveau lorsque ils auront besoin d'une nouvelle fonctionnalité, et seules les nouvelles fonctionnalités vont faire usage des extensions. On doit cependant prendre en compte le comportement du client en face de nouveaux champs, et concevoir un schéma de repli pour l'interopérabilité avec ces clients.

2. Éléments de protocole affectés

2.1 Le second mot complet de 16 bits de l'en-tête du message DNS (voir au paragraphe 4.1.1 de la [RFC1035]) est divisé en un OPCODE de 4 bits, un RCODE de 4 bits, et un certain nombre de fanions de 1 bit. Les bits Z réservés à l'origine ont été alloués à diverses fins, et la plupart des valeurs de RCODE sont maintenant utilisées. On a besoin de plus de fanions et de plus de possibilités de RCODE.

2.2 Les deux premiers bits d'une étiquette de domaine en format réseau sont utilisées pour noter le type de l'étiquette. Le paragraphe 4.1.4 de la [RFC1035] alloue deux des quatre types possibles et réserve les deux autres. Les propositions d'utilisation des deux types restants dépassent de beaucoup ce qui est disponible. Il y a besoin de plus de types d'étiquettes.

2.3 Les messages DNS sont limités à 512 octets lorsque ils sont envoyés sur UDP. Bien que la taille minimum de la mémoire tampon de réassemblage maximum permette toujours une limite de 512 octets de charge utile UDP, la plupart des hôtes maintenant connectés à l'Internet sont capables de réassembler de plus gros datagrammes. Certains mécanismes doivent être créés pour permettre aux demandeurs d'annoncer de plus grosses tailles de mémoire tampon à ceux qui font les réponses.

3. Types d'étiquette étendus

3.1 Le type d'étiquette "0 1" indiquera maintenant un type d'étiquette étendu, dont la valeur est codée dans les six bits de moindre poids du premier octet d'une étiquette. Tous les types d'étiquette développés ultérieurement devraient être codés en utilisant un type d'étiquette étendu.

3.2 Le type d'étiquette étendu "1 1 1 1 1 1" sera réservé pour les extensions futures de l'espace de code de type d'étiquette étendu.

4. Pseudo-RR OPT

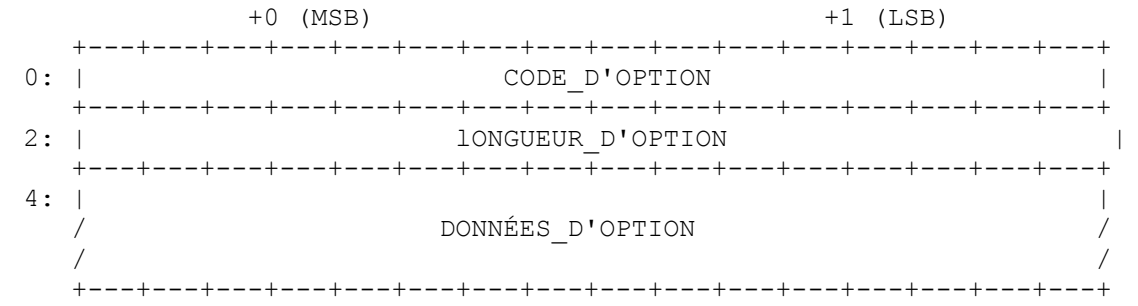
4.1 Un pseudo-RR OPT peut être ajouté à la section Données supplémentaires d'une demande ou d'une réponse. Un OPT est appelé pseudo-RR parce qu'il relève d'un message de niveau transport particulier et d'aucune données réelles du DNS. Les RR OPT ne doivent jamais être placés en antémémoire, transmis, ni mémorisés dans des fichiers maîtres ou chargés à partir d'eux. La quantité de pseudo-RR OPT par message devra être zéro ou un, mais pas plus.

4.2 Un RR OPT a une partie fixe et un ensemble variable d'options exprimées par des paires {attribut, valeur}. La partie fixe contient des métadonnées du DNS et aussi une petite collection des nouveaux éléments de protocole dont nous espérons qu'ils seront assez populaires pour que ce ne soit pas un gaspillage de l'espace réseau de las coder comme paires {attribut, valeur}.

4.3 La partie fixe d'un RR OPT est structurée comme suit :

Nom du champ	Type de champ	Description
NAME	nom de domaine	vide (domaine racine)
TYPE	u_int16_t	OPT
CLASS	u_int16_t	taille de charge utile UDP de l'envoyeur
TTL	u_int32_t	RCODE étendu et fanions
RDLEN	u_int16_t	décrit les RDATA
RDATA	flux d'octets	paires {attribut, valeur}

4.4 La partie variable d'un RR OPT est codée dans son RDATA et est structurée comme zéro ou plus de ce qui suit :



CODE_D'OPTION (Alloué par l'IANA.)

LONGUEUR_D'OPTION Taille (en octets) de DONNÉES_D'OPTION.

DONNÉES_D'OPTION Varie selon le CODE_D'OPTION.

4.5 La taille de charge utile UDP de l'envoyeur (que OPT mémorise dans le champ CLASSE du RR) est le nombre d'octets de la plus grosse charge utile UDP qui peut être réassemblée et livrée dans la pile réseau de l'envoyeur. Noter que la MTU de chemin, avec ou sans fragmentation, peut être plus petite que cela.

4.5.1 Noter qu'une charge utile UDP de 512 octets exige une mémoire tampon de réassemblage IP de 576 octets. Choisir 1280 sur un demandeur connecté à un Ethernet serait raisonnable. La conséquence du choix d'une trop grande valeur peut être un message ICMP provenant d'une passerelle intermédiaire, ou même un abandon silencieux du message de réponse.

4.5.2 Les demandeurs et ceux qui répondent devront tous deux tenir compte de la MTU de découverte de chemin (si elle est déjà connue) lors du calcul des tailles de message.

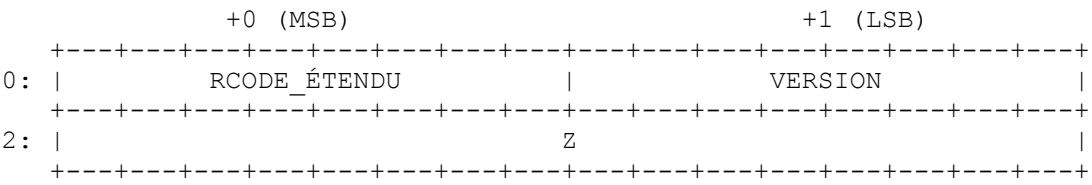
4.5.3 La taille maximum de charge utile du demandeur peut changer avec le temps, et ne devrait donc pas être mise en antémémoire pour être utilisée au delà de la transaction dans laquelle elle est annoncée.

4.5.4 La taille maximum de charge utile de celui qui répond peut changer avec le temps, mais on peut raisonnablement s'attendre à ce qu'elle reste constante entre deux transactions consécutives ; par exemple, une INTERROGATION

sans signification pour découvrir la taille maximum de charge utile UDP d'un répondant, suivie immédiatement par une MISE A JOUR qui tire parti de cette taille. (Ceci est estimé préférable à l'utilisation de TCP pour les demandes surdimensionnées, si il y a une raison quelconque de suspecter que celui qui répond met en œuvre EDNS, et si une demande ne tient pas dans la limite de taille de charge utile par défaut de 512.)

4.5.5 Du fait de la redondance de transaction, il n'est pas sage d'annoncer une limite architecturale comme taille maximum de charge utile UDP. Ce n'est pas parce que votre pile de protocole peut réassembler 64 kbits de datagrammes que vous pouvez supposer que vous voulez dépenser plus d'environ 4 kbits de mémoire d'état pour la transaction en cours.

4.6 Le RCODE étendu et les fanions (que OPT mémorise dans le champ TTL du RR) sont structurés comme suit :



RCODE_ÉTENDU Forme les 8 bits de poids fort des 12 bits du RCODE étendu.
Noter que la valeur "0" du RCODE_ÉTENDU indique qu'un RCODE non étendu est utilisé (les valeurs "0" à "15").

VERSION Indique le niveau de mise en œuvre de celui qui le règle. La pleine conformité à la présente spécification est indiquée par version "0". Les demandeurs sont invités à régler cela au plus faible niveau mis en œuvre qui est capable d'exprimer une transaction, pour minimiser la charge de ceux qui répondent et du réseau pour découvrir le plus fort niveau de mise en œuvre commun entre demandeur et répondant. Une stratégie de numérotage de version d'un demandeur devrait idéalement être une option de configuration au démarrage.

Si un répondant ne met pas en œuvre le niveau de VERSION de la demande, il répond alors par RCODE=BADVERS. Toutes les réponses seront limitées en format au niveau de VERSION de la demande, mais la VERSION de chaque réponse sera le plus haut niveau de mise en œuvre de celui qui répond. De cette façon, le demandeur va apprendre le niveau de mise en œuvre d'un répondant comme effet collatéral de chaque réponse, y compris les réponses d'erreur, incluant RCODE=BADVERS.

Z Réglé à zéro par les envoyeurs et ignoré par les receveurs, sauf modification dans une spécification ultérieure.

5 Considérations pour le transport

5.1 La présence d'un pseudo-RR OPT dans une demande devrait être prise comme une indication que le demandeur met pleinement en œuvre la version mentionnée de EDNS, et qu'il peut correctement comprendre toute réponse qui se conforme à la spécification de cette disposition.

5.2 L'absence d'utilisation de ces dispositions dans une demande doit être prise comme une indication que le demandeur ne met pas en œuvre la présente spécification et que celui qui répond ne peut pas faire usage des extensions de protocole décrites ici dans sa réponse.

5.3 Les répondants qui ne comprennent pas ces extensions de protocole sont supposés envoyer une réponse avec RCODE NOTIMPL, FORMERR, ou SERVFAIL. L'utilisation des extensions devrait donc être "sondée" de telle sorte qu'un répondant dont on ne sait pas si il les prend en charge soit autorisé à recommencer sans extension si il répond avec un tel RCODE. Si le niveau de capacités d'un répondant est mis en antémémoire par un demandeur, un nouveau sondage devrait être fait périodiquement pour vérifier les changements de capacités de ce répondant.

6 Considérations pour la sécurité

La spécification côté demandeur de la taille maximum de mémoire tampon peut ouvrir une nouvelle attaque de déni de service DNS si on peut faire que ceux qui répondent envoient des messages trop grands pour que les passerelles intermédiaires les transmettent, ce qui conduirait éventuellement à des tempêtes ICMP entre les passerelles et les envoyeurs de réponses.

7 Considérations relatives à l'IANA

L'IANA a alloué le code de type de RR 41 pour OPT.

La recommandation du présent document et de son groupe de travail que l'IANA crée un registre pour les types d'étiquette étendus EDNS, pour les codes d'option EDNS, et pour les numéros de version EDNS.

Le présent document alloue le type d'étiquette 0b01xxxxxx comme "type d'étiquette étendue EDNS." Il est demandé à l'IANA d'enregistrer cette allocation.

Le présent document alloue le type d'étiquette étendue 0bxx111111 comme "Réservé pour des types futurs d'étiquette étendue". Il est demandé à l'IANA d'enregistrer cette allocation.

Le présent document alloue le code d'option 65535 à "Réservé pour une expansion future".

Le présent document étend l'espace RCODE de 4 bits à 12 bits. Cela va permettre à l'IANA d'allouer plus que les 16 valeurs de RCODE distinctes permises par la [RFC1035].

Le présent document alloue le RCODE étendu EDNS de "16" à "BADVERS".

L'approbation de l'IESG devrait être exigée pour créer de nouvelles entrées dans les registres de type d'étiquette EDNS étendue ou de numéro de version EDNS, et l'allocation de tout nouveau code d'option EDNS devrait s'appuyer sur la publication d'une RFC (d'information, expérimentale, ou BCP).

8 Remerciements

Paul Mockapetris, Mark Andrews, Robert Elz, Don Lewis, Bob Halley, Donald Eastlake, Rob Austein, Matt Crawford, Randy Bush, et Thomas Narten ont tous participé activement à la création et à l'affinage de cette spécification.

9 Références

[RFC1035] P. Mockapetris, "Noms de domaines – Mise en œuvre et spécification", STD 13, novembre 1987.

10 Adresse de l'auteur

Paul Vixie
Internet Software Consortium
950 Charter Street
Redwood City, CA 94063
USA
téléphone : +1 650 779 7001
mél : vixie@isc.org

11 Déclaration complète de droits de reproduction

Copyright (C) The Internet Society (2000). Tous droits réservés.

Ce document et les traductions de celui-ci peuvent être copiés et diffusés, et les travaux dérivés qui commentent ou expliquent autrement ou aident à sa mise en œuvre peuvent être préparés, copiés, publiés et distribués, partiellement ou en totalité, sans restriction d'aucune sorte, à condition que l'avis de droits de reproduction ci-dessus et ce paragraphe soient inclus sur toutes ces copies et œuvres dérivées. Toutefois, ce document lui-même ne peut être modifié en aucune façon, par exemple en supprimant le droit d'auteur ou les références à l'Internet Society ou d'autres organisations Internet, sauf si c'est nécessaire à l'élaboration des normes Internet, auquel cas les procédures pour les droits de reproduction définies dans les processus des normes de l'Internet doivent être suivies, ou si nécessaire pour le traduire dans des langues autres que l'anglais.

Les permissions limitées accordées ci-dessus sont perpétuelles et ne seront pas révoquées par la Société Internet ou ses successeurs ou ayants droit.

Ce document et les renseignements qu'il contient sont fournis "TELS QUELS" et l'INTERNET SOCIETY et l'INTERNET ENGINEERING TASK FORCE déclinent toute garantie, expresse ou implicite, y compris mais sans s'y limiter, toute garantie que l'utilisation de l'information ici présente n'enfreindra aucun droit ou aucune garantie implicite de commercialisation ou d'adaptation à un objet particulier.

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