

RFC 2671 : Extension Mechanisms for DNS (EDNS0)

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 13 Mars 2007. Dernière mise à jour le 7 Février 2008

Date de publication du RFC : Août 1999

<http://www.bortzmeyer.org/2671.html>

Le protocole DNS a presque vingt ans d'âge dans sa forme actuelle. Il souffre de nombreuses limites et notre RFC vise à permettre d'en surmonter certaines, notamment celle de la taille maximale des réponses.

Le DNS, dans sa forme originale, spécifiée dans le RFC 1034¹, ne permettait pas de négocier des options, d'indiquer au serveur ce que sait faire le client, en plus des capacités minimales qu'impose le protocole. Par exemple, la norme originale (RFC 1035, section 2.3.4) imposait une limite de 512 octets aux messages DNS envoyés sur UDP. Une telle limite est bien trop basse aujourd'hui, à la fois compte-tenu des nouvelles demandes (IPv6, IDN, DNSSEC, tous demandent des données DNS plus grandes) et des capacités des réseaux et des machines modernes. Notre RFC a donc été écrit pour traiter ce problème.

EDNS0 est un mécanisme d'extension du DNS et une première extension, pour indiquer une taille supérieure aux 512 octets. L'extension se fait en squattant des champs inutilisés du paquet (DNS est un format binaire rigide, il ne permet donc pas facilement d'ajouter de nouvelles possibilités) et en créant un pseudo-type d'enregistrement, le type OPT.

La nouvelle extension pour indiquer la taille permet au client de spécifier la quantité d'octets qu'il est capable de recevoir. Avec le client DNS dig, cela se fait avec l'option `bufsize`.

Prenons par exemple le TLD de Hong Kong, `.hk` car c'est un des plus gros en nombre de serveurs de noms. Si je demande cette liste :

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc1034.txt>

```

% dig NS hk.

; <<>> DiG 9.3.4 <<>> NS hk
;; global options:  printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 23662
;; flags: qr rd ra; QUERY: 1, ANSWER: 15, AUTHORITY: 0, ADDITIONAL: 6

;; QUESTION SECTION:
;hk.                IN      NS

;; ANSWER SECTION:
hk.                 604781 IN      NS      NS-HK.RIPE.NET.
hk.                 604781 IN      NS      B.DNS.TW.
hk.                 604781 IN      NS      NS1.HKIRC.NET.hk.
hk.                 604781 IN      NS      NS2.CUHK.EDU.hk.
hk.                 604781 IN      NS      NS2.HKIRC.NET.hk.
hk.                 604781 IN      NS      NS3.CUHK.EDU.hk.
hk.                 604781 IN      NS      SEC3.APNIC.NET.
hk.                 604781 IN      NS      TLD1.ULTRADNS.NET.
hk.                 604781 IN      NS      TLD2.ULTRADNS.NET.
hk.                 604781 IN      NS      TLD3.ULTRADNS.ORG.
hk.                 604781 IN      NS      TLD4.ULTRADNS.ORG.
hk.                 604781 IN      NS      TLD5.ULTRADNS.INFO.
hk.                 604781 IN      NS      TLD6.ULTRADNS.CO.UK.
hk.                 604781 IN      NS      ADNS1.BERKELEY.EDU.
hk.                 604781 IN      NS      ADNS2.BERKELEY.EDU.

;; ADDITIONAL SECTION:
NS1.HKIRC.NET.hk.   26468  IN      A       203.119.2.18
NS2.CUHK.EDU.hk.   133160 IN      A       137.189.6.21
NS2.HKIRC.NET.hk.   26468  IN      A       203.119.2.19
NS3.CUHK.EDU.hk.   133160 IN      A       202.45.188.19
SEC3.APNIC.NET.     43917  IN      A       202.12.28.140
SEC3.APNIC.NET.     43917  IN      AAAA    2001:dc0:1:0:4777::140

;; Query time: 1 msec
;; SERVER: 192.134.4.162#53(192.134.4.162)
;; WHEN: Tue Mar 13 10:22:41 2007
;; MSG SIZE rcvd: 508

```

On voit que la réponse était proche des 512 octets et que, pour qu'elle tienne dans cette limite, le serveur a dû sérieusement réduire la taille de la section additionnelles ("*additional section*"). Si le serveur avait dû réduire encore plus, jusqu'à retirer des enregistrements de la section réponse ("*answer section*"), il aurait dû mettre le bit TC (troncation) à VRAI, imposant ainsi au client de reessayer en TCP.

Mais EDNS0 permet d'avoir la totalité de la section additionnelle (notez la pseudo-section lié à l'enregistrement OPT) :

```

% dig +bufsize=4096 NS hk

; <<>> DiG 9.3.4 <<>> +bufsize=4096 NS hk
;; global options:  printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 36374
;; flags: qr rd ra; QUERY: 1, ANSWER: 15, AUTHORITY: 0, ADDITIONAL: 12

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:

```

```

;hk.                IN      NS

;; ANSWER SECTION:
hk.                604800  IN      NS      B.DNS.TW.
hk.                604800  IN      NS      NS1.HKIRC.NET.hk.
hk.                604800  IN      NS      NS2.CUHK.EDU.hk.
hk.                604800  IN      NS      NS2.HKIRC.NET.hk.
hk.                604800  IN      NS      NS3.CUHK.EDU.hk.
hk.                604800  IN      NS      SEC3.APNIC.NET.
hk.                604800  IN      NS      TLD1.ULTRADNS.NET.
hk.                604800  IN      NS      TLD2.ULTRADNS.NET.
hk.                604800  IN      NS      TLD3.ULTRADNS.ORG.
hk.                604800  IN      NS      TLD4.ULTRADNS.ORG.
hk.                604800  IN      NS      TLD5.ULTRADNS.INFO.
hk.                604800  IN      NS      TLD6.ULTRADNS.CO.UK.
hk.                604800  IN      NS      ADNS1.BERKELEY.EDU.
hk.                604800  IN      NS      ADNS2.BERKELEY.EDU.
hk.                604800  IN      NS      NS-HK.RIPE.NET.

;; ADDITIONAL SECTION:
NS1.HKIRC.NET.hk.  26487  IN      A        203.119.2.18
NS2.CUHK.EDU.hk.  133179 IN      A        137.189.6.21
NS2.HKIRC.NET.hk.  26487  IN      A        203.119.2.19
NS3.CUHK.EDU.hk.  133179 IN      A        202.45.188.19
SEC3.APNIC.NET.    43936  IN      A        202.12.28.140
SEC3.APNIC.NET.    43936  IN      AAAA     2001:dc0:1:0:4777::140
TLD2.ULTRADNS.NET. 105673 IN      A        204.74.113.1
TLD3.ULTRADNS.ORG. 19273  IN      A        199.7.66.1
TLD5.ULTRADNS.INFO. 3385   IN      A        192.100.59.11
ADNS1.BERKELEY.EDU. 2047   IN      A        128.32.136.3
ADNS2.BERKELEY.EDU. 65275  IN      A        128.32.136.14

;; Query time: 22 msec
;; SERVER: 192.134.4.162#53(192.134.4.162)
;; WHEN: Tue Mar 13 10:22:22 2007
;; MSG SIZE rcvd: 599

```

Et voilà, tout le monde est désormais content.

EDNS0 a été normalisé il y a sept ans et est mis en œuvre dans BIND depuis la version 9. Malheureusement, bien des serveurs et des résolveurs ne le déploient pas encore ou, s'ils le font, sont bloqués par un coupe-feu mal programmé ou mal configuré <<http://www.bortzmeyer.org/dns-size.html>>. Ainsi, l'ICANN, avant d'autoriser l'ajout d'adresses IPv6 aux serveurs de noms de la racine, a dû se lancer dans un long processus de test <<http://www.icann.org/committees/security/sac017.htm>>, prenant notamment en compte les obstacles au déploiement de EDNS0. (Ces adresses ont finalement été ajoutées le 4 février 2008 <<http://www.icann.org/announcements/announcement-04feb08.htm>>.) Le même problème se reposera sans doute pour la signature de la racine en 2010 <<http://www.bortzmeyer.org/signature-racine.html>>.

Pour les amateurs de programmation, du code C d'analyse d'un paquet DNS contenant de l'EDNS est dans mon article « Décoder les paquets DNS capturés avec pcap <<http://www.bortzmeyer.org/pcap-decodage-dns.html>> ».