RFC 5798 : Virtual Router Redundancy Protocol Version 3 for IPv4 and IPv6

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 mars 2010

Date de publication du RFC : Mars 2010

https://www.bortzmeyer.org/5798.html

Lorsqu'on configure une machine connectée à l'Internet, on indique le routeur par défaut qu'elle doit utiliser. Que faire s'il tombe en panne ? Utiliser VRRP, le protocole que normalise notre RFC, qui permet à plusieurs routeurs de se surveiller mutuellement, et de se remplacer en cas de défaillance.

Prenons par exemple une machine Unix. Sa table de routage va être :

```
% netstat -r -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags   MSS Window  irtt Iface
10.0.0.0        0.0.0.0         255.255.255.0   U         0 0          0 wlan0
0.0.0.0         10.0.0.1        0.0.0.0         UG        0 0          0 wlan0
```

Ici, `10.0.0.1` est le **routeur par défaut** (on utilise parfois l'ancien terme, incorrect, de passerelle par défaut). Que se passe t-il s'il tombe en panne ? La machine n'a plus accès qu'à une petite portion de l'Internet, son réseau local (ici `10.0.0.0/24`) et ceux pour lesquels il existe une route via un autre routeur. En IPv6, des mécanismes comme la découverte de voisin (RFC 4861 [1]) peuvent aider à trouver un autre routeur, s'il existe, mais les délais sont souvent trop élevés.

C'est évidemment inacceptable lorsqu'on veut pouvoir compter sur son accès Internet. D'où le protocole VRRP, normalisé à l'origine dans le RFC 2338, puis dans le RFC 3768. Cette ancienne version était spécifique à IPv4 et notre RFC 5798 est la première version de VRRP à être commune à IPv4 et IPv6.

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc4861.txt`

Elle porte le numéro 3. (Bien à tort, je trouve, le RFC parle de « IPvX » lorsqu'il veut désigner les deux familles, au lieu de simplement dire « IP ».)

La section 1 du RFC commence par introduire le problème et par expliquer quelles sont les solutions possibles, avant VRRP. Ainsi (section 1.2), en IPv4, les méthodes classiques étaient de faire tourner un protocole de routage comme RIP (RFC 2453) en mode « écoute seule », où la machine reçoit les mises à jour de route, pour en déduire les routeurs possibles, mais n'en envoie pas elle même. C'est ainsi que, à une époque lointaine, toutes les machines Unix en réseau faisaient tourner le programme de routage `routed`. L'utilisation d'un protocole de routage par des machines non-routeuses a entraîné tellement de mauvaise surprises que cette méthode n'est plus recommandée.

Il reste donc la découverte de routeur par le biais de messages ICMP (RFC 1256), jamais réellement déployée, ou bien les routes statiques, mises à la main ou via DHCP. Cette solution a l'inconvénient de l'absence de résistance aux pannes. Si le routeur par défaut s'arrête, il n'y a pas de mécanisme de bascule automatique.

En IPv6 (section 1.3), il y a une possibilité supplémentaire, le protocole de découverte de voisin du RFC 4861 qui permet, via la fonction "*Neighbor Unreachability Detection*" (section 7.3 du RFC 4861) de s'apercevoir qu'un routeur est en panne et d'en chercher un autre. Avec les paramètres par défaut de la découverte de voisin, un tel changement prend environ 40 secondes, ce qui est trop pour la plupart des applications.

La section 1.6 décrit le vocabulaire de VRRP. Notons qu'un « routeur VRRP » est un routeur qui parle le protocole VRRP mais qu'un « routeur virtuel » est l'ensemble des routeurs (un maître et plusieurs remplaçants) qui contribuent au fonctionnement continu d'une adresse IP vers laquelle pointent les routes.

La section 2 est un résumé du cahier des charges de VRRP : service continu pour une adresse IP, avec minimisation du temps de bascule vers un autre routeur physique, possibilité d'exprimer une préférence entre les différents routeurs physiques d'un même routeur virtuel, minimiser les bascules inutiles (par exemple lorsqu'un ancien maître redémarre), fonctionnement sur un réseau local coupé par des ponts qui doivent apprendre l'adresse MAC (section 2.4, qui détaille ce problème), etc. VRRP fonctionne sur tout type de réseau local mais, en pratique, est surtout utilisé sur Ethernet (l'annexe A décrit les spécificités des autres protocole comme le "*Token Ring*").

La section 3 donne une vision générale du protocole VRRP. C'est donc un protocole d'élection. Les routeurs physiques communiquent par la diffusion restreinte sur le réseau local. Pour chaque routeur virtuel, identifié par un nombre nomé VRID ("*virtual router identifier*"), un maître est élu, et il sera le seul à router. Les routes des machines non-routeuses pointeront vers le routeur virtuel (cf. section 4.1 pour un schéma). Si on veut faire de la répartition de charge, il faut plusieurs routeurs virtuels, avec des VRID différents, cf. section 4.2 pour un bon exemple. Chaque routeur virtuel a une adresse MAC (il n'y a donc pas de problème avec les caches ARP).

Le maître diffuse périodiquement des messages "*VRRP advertisement*". Si le maître n'en envoie plus, un remplaçant le... remplace, avec la même adresse IP (celle du routeur virtuel) et la même adresse MAC.

Le protocole est normalisé en section 5. Le format des paquets est en section 5.1 et 5.2. À noter :
— L'adresse IPv4 de destination des paquets de diffusion est `224.0.0.18` et la IPv6 est `FF02:0:0:0:0:0:0:12`.
— Le TTL des paquets doit être à 255 (cf . RFC 5082).
— VRRP n'utilise pas un des protocoles de transport classiques comme UDP, il a son propre protocole, numéroté 112.

— Le VRID est sur huit bits et il n'existe pas de mécanisme pour son allocation sur un réseau local, c'est à l'administrateur réseau de s'assurer de son unicité.

La section 6 est consacréee à la machine à états du protocole. Dans l'état d'attente ("*Backup*", section 6.4.2), le routeur VRRP écoute passivement les annonces du maître et ne répond **pas** aux requêtes ARP ou ND pour l'adresse IP du routeur virtuel, et ignore les paquets envoyés à cette adresse. Si le maître n'envoie plus d'annonces (par défaut, c'est après trois annonces non reçues), le routeur passe dans l'état Maître.

Inversement, le routeur dans l'état Maître (section 6.4.3), répond aux requêtes ARP et ND pour l'adresse IP du routeur virtuel, traite les paquets destinés à cette adresse, route les paquets qu'il reçoit et envoie des annonces périodiques pour manifester qu'il est toujours en service.

La section 7 décrit de manière très détaillée ce que doit faire un routeur VRRP lorsqu'il reçoit ou émet les paquets VRRP. C'est là qu'est spécifié le fait qu'un routeur maître doit utiliser l'adresse MAC du routeur virtuel (et non pas celle du routeur physique) lorsqu'il envoie les annonces de bon fonctionnement. C'est pour permettre aux ponts et commutateurs de trouver le routeur physique. Cette adresse MAC est calculée (sections 7.3 et 12) et vaut `00-00-5E-00-01-{VRID}` en IPv4 et `00-00-5E-00-02-{VRID}` en IPv6.

Comme souvent sur un réseau, le diable est dans les détails pratiques. Ils font l'objet de la section 8. Ainsi, la section 8.1.2 rappelle que, puisque le maître utilise toujours comme adresse MAC celle du routeur virtuel présentée au paragraphe précédent, le client ne peut pas découvrir qu'un routeur physique en a remplacé un autre.

Parmi ces problèmes opérationnels, celui de la sécurité a droit à une section entière, la 9. En gros, VRRP n'a aucune sécurité. Les versions précédentes avaient tenté de mettre en place quelques mécanismes mais ils n'ont eu aucun succès et notre version 3 de VRRP les supprime. Notons que le problème n'est pas créé par VRRP : sur le réseau local, un méchant a énormément de moyens de perturber bien des protocoles indispensables, à commencer par DHCP et ARP. Par exemple, le méchant peut toujours répondre aux requêtes ARP pour l'adresse IP du routeur et lui voler ainsi le trafic. VRRP, où le méchant peut se faire désigner comme maître, n'aggrave donc pas tellement la situation.

VRRP a toujours souffert d'une polémique récurrente sur un brevet que détient Cisco et qui est apparemment effectivement appliqué (des développeurs VRRP ou bien de protocoles similaires ont, semble t-il, reçu des menaces des avocats de Cisco). L'existence de ce brevet n'est pas en soi contraire à la politique de l'IETF. En effet, celle-ci accepte des protocoles brevetés (il est difficile de faire autrement, compte-tenu du membre, et du caractère ridiculement futile, de la grande majorité des brevets logiciels) et demande juste que les prétentions soient publiques, ce qui est le cas ici <`https://datatracker.ietf.org/ipr/19/`>. D'innombrables messages, souvent courroucés, ont été échangé sur les listes IETF au sujet de ce brevet. Cette situation a mené les développeurs d'OpenBSD à développer un protocole concurrent, CARP. On peut lire un point de vue (très outrancier) sur cette polémique dans l'interview de Ryan McBride <`http://kerneltrap.org/node/2873`>. La réalité est plus complexe : les développeurs d'OpenBSD ont adopté une attitude d'affrontement immédiatement (« comme souvent », disent ceux qui connaissent les gens d'OpenBSD) et la bureaucratie IETF n'a pas fait preuve de bonne volonté et a réagi par un blocage complet. Aujourd'hui, les positions semblent hélas figées, au point que CARP, n'ayant pas eu de numéro de protocole officiel (en raison de leur refus de se plier aux procédures IANA, procédures d'ailleurs incorrectement décrites par McBride), a tout simplement pris celui de VRRP. Sur un réseau local, si on voit des paquets du protocole 112, il peut donc s'agir de CARP ou bien de VRRP.

En raison du brevet Cisco sur HSRP (le précurseur de VRRP), brevet dont la licence n'est apparemment disponible que selon les conditions RAND (bien insuffisantes pour du logiciel libre), il n'est pas évident de faire une mise en œuvre libre de VRRP. Il existe toutefois vrrpd <`http://www.imagestream.`

com/VRRP.html> et keepalived <http://www.keepalived.org/> (aucun des deux ne semble intégrer IPv6).

La configuration de VRRP sur un routeur Juniper est documentée en *"Configuring VRRP and VRRP for IPv6"* <http://www.juniper.net/techpubs/software/junos/junos92/swconfig-network-interface configuring-vrrp-and-vrrp-for-ipv6.html> et *"Configure Basic VRRP Support"* <http://www.juniper.net/techpubs/software/junos/junos56/swconfig56-interfaces/html/interfaces-html>. Cela donne, par exemple :

```
address 192.0.2.0/25 {
                    vrrp-group 12 {
                        virtual-address 192.0.2.65;
                        priority 80;
                    }
```

Notez que `vrrp-group` est un terme purement Juniper, c'est ce que le RFC appelle VRID (ici 12). Quant au concept de priorité (ici 80, donc moins que la priorité par défaut), il est décrit en section 5.2.4. Avec `vrrpd`, la même configuration serait :

```
ip address 192.0.2.2 255.255.255.128
vrrp 12 ip 192.0.2.1
vrrp 12 priority 80
```