

PMTU

Jean-Yves BRUCKER

Path Maximum Transmission Unit (**PMTU**)

La description complète de ce protocole est décrite dans le **RFC1191**.

Le besoin

Lorsque plusieurs réseaux doivent communiquer au travers de routeurs, les MTU (1500 pour Ethernet, 4096 pour TokenRing...) peuvent être différentes.

Afin d'éviter la fragmentation et l'engorgement des routeurs, la valeur du paramètre "**tcp_mssdflt**" (par défaut 512) doit être modifiée.

Pour un même volume de données transférées, le nombre de paquets va augmenter ainsi que la charge globale du réseau.

Pour éviter ces inconvénients, le PMTU a été défini et mis en œuvre sur l'AIX 4.3.3.

Comment ça marche

A l'établissement de la connexion, par exemple d'un réseau TokenRing vers un réseau Ethernet, le système envoie vers la machine distante un paquet d'ICMP (*ping*) de 4096 avec l'information "DF" (ne pas fragmenter).

Mais si le routeur qui reçoit ce paquet ne peut pas le transférer sans fragmentation, le paquet sera perdu. Ce paquet perdu ne revenant pas à l'émetteur, un nouveau paquet, plus petit, sera envoyé vers le routeur, toujours avec l'information "DF", jusqu'à l'obtention d'un paquet de retour ICMP.

L'information de PMTU sera alors stockée dans la table de routage (**netstat -rn**).

Dans notre cas, il y aura amélioration du trafic en envoyant des paquets de MTU 1460, au lieu de 512, sans provoquer de fragmentation.

Cette information est valide pendant 30 minutes (**pmtu_rediscover_interval**) et sera conservée pendant 10 minutes en cas d'inactivité (**pmtu_default_age**).

Performances

Problèmes

Le PMTU est activé par défaut sur l'AIX 4.3.3. et l'AIX 5L.

Quelques routeurs ne savent pas manipuler ces informations, ce qui peut entraîner des problèmes. Dans certains cas, la ML 08/09 peut provoquer des blocages (*hangs*). Une PTF corrigeant ce problème est disponible auprès du Point Service.

L'utilisation du PMTU avec plusieurs centaines de connexions entraînera régulièrement des "tempêtes de pings" qui pourront :

- dans certains cas, provoquer un blocage des réseaux pendant plusieurs secondes (par exemple, sur des terminaux distants sur des lignes à 19200)
- et, dans tous les cas, augmenter la charge du réseau.

Solutions

● Conserver le PMTU

Conserver le PMTU en cas de réseaux complexes utilisant plusieurs valeurs de MTU. Dans ce cas, afin de réduire la trafic ICMP, augmenter les valeurs "**pmtu_rediscover_interval**" et "**pmtu_discover_age**" (attention toutefois au routage dynamique qui pourrait prendre une route différente au retour).

● Supprimer le PMTU

Solution fortement recommandée !

Dans ce cas, il faut positionner le paramètre "**tcp_mssdflt=1452**". Cette valeur convient à la majorité des cas.

● Cas particuliers

Si un réseau "giga-Ethernet" utilisant des *jumbos frames* communique avec un autre réseau "giga-Ethernet" (*jumbo* sur toute la liaison), on peut utiliser le paramètre "**tcp_mssdflt**" avec uniquement l'interface de réseau correspondante (par exemple : `ifconfig en0 tcp_mssdflt=9000`).

Remarque

Toutes ces modifications sont dynamiques et doivent être incluses à la fin de "**rc.net**" :

```
/etc/rc.net
/usr/sbin/no -o udp_pmtu_discover=0
.....
/etc/ifconfig en0 tcp_mssdflt=9000
.....
```

