

## Dégradations de débit pour les messages d'un seul paquet

La transmission de parole numérisée sur l'ARPANET représente une nouvelle dimension de l'utilisation des systèmes à commutation de paquet. Les exigences de débit et de délai pour ce domaine d'application émergent sont assez différentes des exigences de débit et de délai des utilisations interactives ou des transferts de fichiers. En particulier, il faut réaliser un haut débit pour de petits messages car les messages longs résultent en longs délais de source pour remplir mes grandes mémoires tampon. Nous étudions donc actuellement les limites de débit pour les messages d'un seul paquet. Nous avons réalisé que jusqu'à maintenant peu de tentatives ont été faites pour optimiser le débit pour le trafic à faible délai. Il a cependant été surprenant de constater que le débit observé pour les messages à un seul paquet est dans de nombreux cas de seulement un quart de ce qu'on attendait. Dans ce qui suit, nous allons expliquer pourquoi cela arrive et ce que l'on peut faire pour corriger cette situation.

Le 1<sup>er</sup> avril 1974, nous avons envoyé, à l'aide du générateur de message de l'IMP, des messages d'un seul paquet au plus fort débit possible ("piloté par RFNM") à partir du MOFFET-IMP, au SRI-IMP. Il y a deux chemins de trois bonds du MOFFET au SRI, l'un impliquant deux circuits à 230,4 kbit/s. Comme il n'y avait presque pas d'interférence de trafic, nous attendions un délai moyen d'aller-retour de plus de 100 ms. En supposant qu'il y ait, en moyenne, 3 messages en transmission entre MOFFET et SRI et en supposant une longueur de message d'environ 1000 bits, il aurait dû en résulter un débit de plus de 30 kbit/s. Le débit observé était cependant de moins de 8 kbit/s. Une répétition de l'expérience a montré le même résultat. Une analyse plus détaillée des données collectées a révélé qu'un nombre moyen de 3,5 messages étaient simultanément en transmission entre le MOFFET et le SRI. La dégradation du débit ne pouvait donc pas être imputée au trafic interférant entre les deux sites. L'utilisation du canal pour tous les canaux qui étaient impliqués dans la transmission était aussi de moins de 40 pour cent. Les temps moyens d'aller-retour observés entre le MOFFET et le SRI étaient, cependant, d'environ 500 ms. Comme ces gros temps d'aller-retour n'étaient visiblement pas dus à des limitations physiques, nous avons étudié le mécanisme de contrôle de flux pour les messages d'un seul paquet et avons été capables d'aboutir à une explication de ce comportement indésirable.

Lorsque un message d'un seul paquet arrive à l'IMP de destination en désordre (c'est-à-dire que le message logiquement précédant n'y est pas encore arrivé) il n'est pas accepté par l'IMP de destination. Il est plutôt traité comme une demande d'allocation d'une mémoire tampon de réassemblage. Le message ALLOCATE correspondant n'est alors renvoyé à l'IMP de source qu'après que le RFNM pour le message précédent a été traité. On peut donc avoir la séquence d'événements suivante :

- 1 MSG(i) envoyé de SOURCE-IMP (le message i est envoyé de l'IMP de source à l'IMP de destination).
- 2 MSG(i+1) envoyé de SOURCE-IMP.
- 3 MSG(i+1) arrive à DEST-IMP (du fait d'un chemin de remplacement ou d'une erreur de ligne, le message (i+1) arrive à l'IMP de destination décalé il est traité comme une demande d'allocation de mémoire tampon de réassemblage puis éliminé).
- 4 MSG(i) arrive à DEST-IMP (le message i arrive à l'IMP de destination; il est mis sur la file d'attente de sortie de l'hôte approprié).
- 5 RFNM(i) envoyé de DEST-IMP (après que le message i a été accepté par l'hôte de destination, le RFNM est envoyé à l'IMP de source).
- 6 ALL(i+1) envoyé de DEST-IMP (le ALLOCATE pour le message i + 1 ne peut être envoyé qu'après que le RFNM pour le message i a été traité).
- 7 RFNM(i) arrive à SOURCE-IMP.
- 8 ALL(i+1) arrive à SOURCE-IMP.
- 9 MSG(i+1) arrive à DEST-IMP (maintenant le message i+1 est mis sur la file d'attente de sortie de l'hôte approprié.)
- 10 RFNM(i+1) envoyé par DEST-IMP.

11 RFNM(i+1) arrive à SOURCE-IMP.

Noter que le délai d'aller-retour pour le message i+1 est l'intervalle de temps entre l'événement 2 et l'événement 11. Donc le délai d'aller-retour pour le message i+1 est plus de deux fois celui qu'il aurait eu si il était arrivé dans l'ordre, les autres conditions restant inchangées. Donc dans de nombreux cas, une erreur de ligne retarde non seulement le message erroné mais aussi le message d'un seul paquet suivant si ce message sui le précédant dans le 125 ms, l'intervalle de temporisation d'erreur de retransmission. De plus, un chemin de remplacement plus rapide pour l'IMP de destination peut en fait ralentir la transmission car il provoque l'arrivée des messages dans le désordre.

Cette situation devient encore pire lorsque on considère le trafic de message d'un seul paquet conduite par RFNM. Le Tableau 1 montre une séquence possible d'événements. Nous supposons ici encore que le message i+1 atteint l'IMP de destination avant le message i.

IMP de SOURCE	IMP de DESTINATION
MSG(i) envoyé	
MSG(i+1) envoyé	
MSG(i+2) envoyé	MSG(i+1) arrive
MSG(i+3) envoyé	MSG(i) arrive
	RFNM(i) envoyé
	ALL(i+1) envoyé
	MSG(i+2) arrive
	MSG(i+3) arrive
RFNM(i) arrive	
MSG(i+4) envoyé	
ALL(i+1) arrive	MSG(i+4) arrive
MSG(i+1) envoyé	MSG(i+1) arrive
	RFNM(i+1) envoyé
RFNM(i+1) arrive	ALL(i+2) envoyé
MSG(i+5) envoyé	
ALL(i+2) arrive	MSG(i+5) arrive
MSG(i+2) envoyé	MSG(i+2) arrive
	RFNM(i+2) envoyé
RFNM(i+2) arrive	ALL(i+3) envoyé
MSG(i+6) envoyé	
ALL(i+3) arrive	MSG(i+6) arrive
MSG(i+3) envoyé	MSG(i+3) arrive
	RFNM(i+3) envoyé
RFNM(i+3) arrive	ALL(i+4) envoyé
MSG(i+7) envoyé	
ALL(i+4) arrive	MSG(i+7) arrive
MSG(i+4) envoyé	MSG(i+4) arrive
	RFNM(i+4) envoyé
RFNM(i+4) arrive	ALL(i+5)
MSG(i+8) envoyé	
ALL(i+5) arrive	
MSG(i+5) envoyé	

**Tableau 1. Schéma de retransmission pour le contrôle de flux en cours**

(Comme le trafic est piloté par RFNM, l'arrivée de RFNM i, i+1, ... est suivie par l'envoi des messages i+4, i+5, ...)

Le fait le plus intéressant de cette séquence d'événements est que l'arrivée du message i+1 avant le message i à l'IMP de destination IMP cause non seulement la retransmission des messages i+1 mais aussi de tous les messages futurs – bien qu'on suppose qu'aucun des messages futurs n'arrive décalé. Le tableau montre aussi que le délai d'aller-retour pour le message i+4 et tous les messages futurs est quatre fois plus long qu'il n'aurait été sans les retransmissions indésirables. Il vaut aussi de noter que, une fois que ce schéma de retransmission s'est établi lui-même, il n'y a pratiquement aucun moyen de récupérer de cette condition autrement qu'en interrompant le flux d'entrée à l'IMP de source. Une seule arrivée décalée du dernier utilisateur ou des messages de contrôle, par exemple, va changer ce schéma de retransmission. Le flux normal de messages d'un seul paquet ne sera rétabli de lui-même que si, par exemple, les messages i+4, i+5, et i+6 sont simultanément retardés de plusieurs centaines de millisecondes de telles sorte que les messages i+1, i+2, et i+3 puissent être retransmis dans l'intervalle. La probabilité d'occurrence d'un tel événement est, cependant, extrêmement faible. Donc on peut considérer que le système s'est enfermé dans une condition de retransmission indésirable. D'un autre côté, le flux "normal"

de messages ne représente que le comportement transitoire du système dans la mesure où il y a toujours une probabilité finie que deux messages arrivent décalés du fait d'erreurs de transmission.

Comme mentionné ci-dessus, le système ne peut récupérer de cette dégradation du débit (et de délai) que si le flux d'entrée de messages d'un seul paquet est interrompu. Dans le cas de transmission de parole, cependant, cela peut ne pas survenir pendant un long moment. Donc les systèmes de transmission de la parole vont dans de nombreux cas avoir à travailler avec un quart de la bande passante espérée pour les messages à un seul paquet. Comme ceci est évidemment une condition inacceptable, on va chercher à modifier le schéma actuel de contrôle de flux pour corriger cette situation. Dans ce qui suit, nous décrivons deux méthodes qui pourraient être utilisées pour éviter la retransmission indésirable des messages.

Rappelons qu'un message d'un seul paquet est rejeté à l'IMP de destination et retransmis plus tard si le RFNM pour le message précédent n'a pas déjà été envoyé à l'IMP de source. Cela est fait principalement pour empêcher l'occurrence de conditions de blocage de réassemblage [1]. Donc le problème ne peut pas être résolu en acceptant simplement tous les messages d'un seul paquet sans mesures supplémentaires pour empêcher les impasses. Cela pourrait conduire à un blocage du réassemblage si un grand nombre de messages d'un seul paquet provenant de plusieurs IMP de source arrivent décalés à leur IMP commun de destination. Dans ce cas, l'IMP de destination pourrait n'être pas capable d'accepter les messages qui sont en ordre à cause de l'insuffisance des mémoires tampon de réassemblage. Il en résulte que le système est dans une impasse. Toute solution du problème de la dégradation du débit doit garantir que tous les messages qui arrivent en ordre peuvent être acceptés par l'IMP de destination.

Une façon d'arriver à cet objectif est de ne rejeter les messages d'un seul paquet qui arrivent en désordre que si les exigences de mémoire tampon du ou des messages précédents ne sont pas connues. Dans les exemples précédents, nous avons vu que l'IMP de destination rejetait en continu les messages bien qu'il connaisse les exigences de mémoire tampon pour les messages qui devaient être livrés en premier. Lorsque les exigences de mémoire tampon commencent à être connues, la quantité nécessaire de mémoire tampon peut être mise de côté et les messages futurs d'un seul paquet peuvent être acceptés sans danger de blocage. Donc le schéma de retransmission indésirable ne peut plus s'établir tout seul. Le Tableau 2 montre la séquence des événements de cette politique si le message  $i+1$  arrive avant le message  $i$  à l'IMP de destination.

IMP de SOURCE	IMP de DESTINATION
MSG (i) envoyé	
MSG(i+1) envoyé	
MSG(i+2) envoyé	MSG(i+1) arrivé. (rejeté)
MSG(i+3) envoyé	MSG(i) arrivé (sortie vers l'hôte)
	RFNM(i) envoyé
	ALL (i+1) envoyé
	MSG(i+2) arrivé (mémemorisé)
	MSG(i+3) arrivé (mémemorisé)
RFNM(i) arrivé	
MSG(i+4) envoyé	
ALL(i+1) arrivé	MSG(i+4) arrivé (mémemorisé)
MSG(i+1) envoyé	MSG(i+1) arrivé (sortie vers l'hôte)
	RFNM(i+1) envoyé
RFNM(i+1) arrivé	RFNM(i+2) envoyé
MSG(i+5) envoyé	RFNM(i+3) envoyé
	RFNM(i+4) envoyé

**Tableau 2 : Séquence des événements pour les schéma modifié de contrôle de flux**

Noter que dans ce schéma modifié un seul message, le message  $i+1$ , est retransmis. En considérant le fait que les IMP ont beaucoup d'espace de mémoire tampon de réassemblage, il est cependant souhaitable d'éviter aussi cette seule retransmission. Ceci est particulièrement important pour les transmissions de parole qui dépendent d'un flux de données solide et sera perturbé par un grand retard soudain. Nous suggérons donc une seconde méthode pour résoudre le problème de la dégradation du débit qui, dans la plupart des cas, n'exige aucune retransmission.

Supposons que tous les messages d'un seul paquet soient acceptés initialement (ou mémorisés). Actuellement les messages d'un seul paquet qui arrivent en désordre sont rejetés à cause de la possibilité d'un blocage. Mais regardons de plus près la situation où tous les messages d'un seul paquet sont acceptés (ou mémorisés) de telle sorte qu'il n'y ait pas de mémoire tampon de réassemblage disponible pour les messages qui doivent être livrés à leur prochain hôte. Ceci n'est pas réellement une condition de blocage parce que les IMP de source conservent une copie des tous les messages d'un seul paquet pour lesquels un RFNM n'a pas encore été reçu. Donc, tout message d'un seul paquet qui est arrivé décalé mais a été cependant accepté par l'IMP de destination peut être supprimé ultérieurement sans que le message ne soit perdu. L'IMP de destination a seulement à envoyer un ALLOCATE pour chaque message d'un seul paquet supprimé à l'IMP de source correspondant

lorsque de l'espace de mémoire tampon de réassemblage devient disponible. Cela peut aussi être considéré comme un rejet différé. Mais maintenant, une retransmission n'est nécessaire que si l'IMP de destination est réellement à court de mémoires tampon de réassemblage. Dans ce cas, les limitations physiques du système sont atteintes et on ne peut espérer obtenir des augmentations du débit qu'au moyen de changements de protocole.

Notre intention est de poursuivre l'étude de ce problème avec le groupe de développement de l'IMP au BBN à l'avenir. Il y a accord sur le fait que des deux solutions suggérées amélioreraient la situation. Cependant, il y a peut-être des solutions de remplacement.

Je remercie Bill Naylor et Joe Katz de leur aide pour effectuer les expériences qui ont mené à la découverte de la dégradation du débit.

**Référence :**

[1] L.Kleinrock et H. Opderbeck, "Sur une possible condition de blocage dans le sous-réseau IMP à cause de la séquence de message", RFC 626, mars 1974.

[ La présente RFC a été mise en forme pour son entrée dans les archives en ligne des RFC par Alex McKenzie avec le soutien de GTE, anciennement BBN Corp. en 11/99 ]