

Groupe de travail Réseau
Request for Comments : 626
NIC n° 22161
Traduction Claude Brière de L'Isle

L. Kleinrock (UCLA-NMC)
H. Opderbeck (UCLA-NMC)
mars 1974

"Sur une possible condition de blocage dans le sous-réseau IMP à cause de la séquence de message"

Les conditions de blocage ou d'impasse sont les dysfonctionnement système les plus sérieux qui puissent survenir dans un système informatique ou un réseau. Les protocoles de communication doivent être conçus avec le plus grand soin pour éviter l'apparition de ces blocages. Leur caractéristique commune est qu'ils ne surviennent que dans des circonstances inhabituelles qui n'étaient pas prévues ou qui sont réputées trop invraisemblables pour qu'elles surviennent par les concepteurs du protocole. (Cependant, ces concepteurs ne sont souvent pas en position d'évaluer quantitativement une telle vraisemblance.)

Le blocage le plus connu qui s'est produit dans l'ARPANET est le blocage de réassemblage [1]. Le blocage du système de distribution différée, lui aussi décrit dans la référence 1, a été évité dans le nouveau IMPSYS par l'observation scrupuleuse de l'heuristique de Kahn [1]. Le dernier blocage que l'on connaisse dans le sous réseau est survenu le 21 décembre 1973 ("le blocage de Noël"). Ces conditions de blocage dormantes sont mises en lumière en collectant des messages de mesures instantanées de tous les sites simultanément. Le blocage de Noël est survenu quand les messages de situations instantanées sont arrivés à leur IMP d'UCLA à qui a été allouée la tâche de la mémorisation de leur réassemblage et où aucun bloc de réassemblage n'était libre. (Un bloc de réassemblage est un élément de mémorisation utilisé dans le processus réel de réassemblage des paquets en messages) [2].

Des conditions d'impasse ont été observées non seulement dans le sous-réseau mais aussi dans des protocoles de niveau supérieur. La conception originale de l'ICP, par exemple, avait une faute qui n'a été découverte qu'après plusieurs mois d'utilisation [3,4]. Plus récemment BBN a rapporté un problème d'impasse impliquant l'échange d'informations d'état d'hôtes par les programmes du serveur RSEXEC (RSSER) [5].

Tant qu'il ne sera pas possible de concevoir des protocoles de communication pratiques qui garantissent un fonctionnement sans blocage, il est vital de vérifier continuellement les protocoles actuellement utilisés à la recherche de telles défaillances – même si elle paraissent "très invraisemblables". Dans la présente RFC, on commente une condition possible de blocage dans le sous-réseau IMP, qui à notre connaissance, ne s'est pas encore produite, ni n'a encore été identifiée. Bien que l'on n'ait jamais vu ce problème se produire dans la réalité, il peut survenir à l'avenir si aucune précaution n'est prise. Cette possible condition de blocage est due au séquençement des messages dans le sous-réseau.

Pour éviter l'occurrence d'un blocage de réassemblage, le mécanisme de contrôle de flux dans le sous-réseau a été modifié de plusieurs façons significatives. Il y a actuellement une limite de quatre messages qui peuvent être simultanément en transmission entre toute paire d'IMP de source et de destination. Par suite de la suppression du traitement de liaison sur l'ancien IMPSYS, il est devenu nécessaire d'introduire un mécanisme de séquençement de message.

Les messages quittent l'IMP de destination dans le même ordre que celui dans lequel ils sont entrés dans l'IMP de source. (Noter que le séquençage est fait IMP par IMP, et non d'hôte à hôte. Cela peut introduire un "délai de séquençage" indésirable si deux hôtes rattachés au même IMP de destination reçoivent des messages du même IMP de source).

Le séquençage des messages a, en général, un potentiel d'introduction de conditions de blocage. La raison en est que tout message, disons le message (n+1), qui est décalé (et donc ne peut pas être livré à son hôte de destination) peut utiliser des ressources qui sont requises par le message (n) qui doit être livré ensuite. Donc, le message (n) ne peut pas atteindre son IMP de destination qui, à son tour, empêche les autres messages (n+1), etc.) qui sont décalés, d'être livrés à leur hôte de destination. Pour cette raison, on doit faire très attention à ne pas allouer trop de ressources (par exemple, de mémoire tampon) aux messages qui sont décalés.

Pour éviter les conditions de blocage l'IMPSYS actuel prend les deux précautions suivantes :

1. Les demandes d'allocation de mémoire tampon sont toujours servies dans l'ordre des numéros de message ; c'est à dire qu'aucun "ALLOCATE" n'est retourné pour le message (n+1) si le message (n) (ou une demande d'allocation pour le message (n)) n'a pas encore été reçue et servie.
2. Les messages d'un seul paquet (régulier et prioritaire) qui arrivent à l'IMP de destination décalés ne sont pas acceptés, sauf si ils ont été retransmis en réponse à une allocation précédente de mémoire tampon. Ces messages sont plutôt

traités comme une demande d'allocation d'une mémoire tampon (conformément au 1 ci-dessus) puis éliminés.

Avec ces deux précautions, il paraît impossible qu'apparaisse une condition de blocage. Cependant, il y a un second mécanisme d'allocation de mémoire tampon qui n'est pas lié au séquençage de message, à savoir le "ALLOCATE" qui est porté sur le RFNM pour un message de plusieurs paquets. Le ALLOCATE porté représente une allocation de mémoire tampon pour le prochain message à plusieurs paquets, et NON pour le prochain message en séquence. Donc, si le prochain message en séquence est un message d'un seul paquet, le ALLOCATE porté actif alloue les mémoires tampon à un message qui est déclassé.

Voyons comment cette situation peut conduire à une condition de blocage. Supposons qu'il y a un nombre maximum de 24 mémoires tampon de réassemblage dans chaque IMP.

Soient les IMP A, B, et C qui transmettent continuellement des messages de huit paquets au même IMP de destination D de telle sorte que toutes les 24 mémoires tampon de réassemblage de l'IMP D sont entièrement utilisées par cette transmission de messages à paquets multiples. Si maintenant, dans le flux de messages de huit paquets, l'IMP A envoie un message d'un seul paquet, il ne sera généralement pas accepté par l'IMP de destination D car il n'y a pas d'espace de mémoire tampon de réassemblage disponible. (Il peut y avoir une mémoire tampon de réassemblage libre si il se trouve que le message d'un seul paquet arrive au moment où un des trois messages de huit paquets est transmis à son hôte). Le message d'un seul paquet va donc être traité comme une demande d'allocation de mémoire tampon. Cette demande ne sera pas servie avant l'envoi du RFNM du message à paquets multiples précédent. À ce moment cependant, toutes les mémoires tampon de réassemblage libres sont déjà allouées au prochain message à paquets multiples via le mécanisme ALLOCATE porté. La seule chance pour que le message à un seul paquet obtienne la satisfaction de sa demande d'allocation est de saisir une mémoire tampon de réassemblage de l'un des deux autres messages de huit paquets. Cette tentative peut être sans succès parce qu'elle dépend de la succession dans le temps des événements dans l'IMP. Une condition de blocage peut survenir si l'IMP B et l'IMP C envoient aussi un message d'un seul paquet dans leur flux de messages de huit paquets qui ne peut pas être servi pour la même raison. Dans ce cas, les trois messages de huit paquets qui vont arriver ensuite à l'IMP D ne pourront pas être livrés à leurs hôtes de destination parce qu'ils sont décalés. Les trois messages d'un seul paquet qui devraient être livrés ensuite ne vont cependant jamais atteindre l'IMP de destination car il n'y a pas d'espace de réassemblage disponible.

Une séquence d'événements possible qui conduit à une condition de blocage peut être décrite comme suit :

Exemples pour la notation :

événement : A8 arrive -> tous les 8 paquets du message à huit paquets de l'IMP A sont arrivés à l'IMP D

événement : C1 arrive -> un message d'un seul paquet de l'IMP C est arrivé à l'IMP D (et est traité comme demande d'allocation de mémoire tampon)

événement : B8 s'achève -> le dernier paquet du message de huit paquets de l'IMP B a été reçu par son hôte de destination

événement : A8 RFNM/ALL -> un RFNM avec le ALLOCATE porté est envoyé à l'IMP A

	Nombre de mémoires tampon de réassemblage allouées	Nombre de mémoires tampon de réassemblage utilisées	Nombre de mémoires tampon de réassemblage libres
Initialement	24	0	0
1. A8 arrive	16	8	0
2. B8 arrive	8	16	0
3. C8 arrive	0	24	0
4. A1 arrive	0	24	0
5. B1 arrive	0	24	0
6. C1 arrive	0	24	0
7. A8 achevé	0	16	8
8. B8 achevé	0	8	16
9. C8 achevé	0	0	24
10. A8 RFNM/ALL	8	0	16
11. B8 RFNM/ALL	16	0	8
12. C8 RFNM/ALL	24	0	0
13. A8 arrive	16	8	0
14. B8 arrive	8	16	0
15. C8 arrive	0	24	0
16. - blocage -			

Noter qu'un ALLOCATE pour un des messages à un seul paquet A1, B1 et C1 ne peut être retourné, respectivement, à

l'IMP de source A, B, et C, qu'après que le RFNM (avec son ALLOCATE porté) pour le précédent message à huit paquets a été envoyé. Si ces RFNM sont envoyés en séquence, c'est-à-dire, sans un ALLOCATE pour un des messages à un seul paquet entre eux, la mémoire de réassemblage temporairement libre (événements (7) à (9)) est implicitement allouée au prochain message multi paquets (événements (10) à (12)) et non à un des messages à un seul paquet. Les prochains messages à huit paquets sont cependant décalés et ne peuvent être livrés à leur hôte de destination.

Dès à présent, il semble qu'un tel blocage ne peut survenir que si le nombre de mémoires tampon de réassemblage est un multiple de huit. Bien sûr, la probabilité d'un blocage est plus forte dans ce dernier cas. Cependant, des blocages peuvent aussi survenir si le nombre de mémoires tampon de réassemblage n'est pas un multiple de huit.

Supposons qu'il y a 26 mémoires tampon de réassemblage au lieu de 24. Supposons aussi que, du fait de chemins de remplacement, de défaillances de ligne, ou de retransmissions, le second paquet d'un message de deux paquets arrive à l'IMP D avant un message d'un seul paquet provenant du même IMP de source A. Le message d'un seul paquet a un plus petit numéro de séquence et doit donc être livré à son hôte de destination avant le message de deux paquets. Lorsque le second paquet du message de deux paquets arrive à l'IMP D, l'IMP réalise que seulement deux des huit mémoires tampon allouées seront nécessaires. Donc six mémoires tampon sont retournées au pôle des mémoires tampon de réassemblage libres. Si il y avait avant $26 - 3 \times 8 = 2$ mémoires tampon dans le pôle, la taille du pôle est augmentée de 6 à 8 mémoires tampon. Ces huit mémoires sont cependant juste assez pour envoyer un autre ALLOCATE porté. Le message d'un seul paquet va donc trouver vide le pôle des mémoires tampon de réassemblage libres, bien que le nombre total de mémoires de réassemblage ne soit pas un multiple de huit. Le message de deux paquets ne peut pas être livré à son hôte de destination parce qu'il est décalé. Donc, la condition de blocage que nous décrivions précédemment peut survenir encore.

On est d'accord que la séquence d'événements mentionnée ci-dessus a une faible probabilité de survenir (autrement quelqu'un l'aurait déjà observée). Ceci est particulièrement vrai car le nombre maximum actuel de mémoires tampon de réassemblage (58) est bien plus élevé que 24. En jugeant à partir de l'expérience du passé des systèmes et réseaux informatiques, on sait cependant que même les événements très peu probables ont une tendance à survenir à long terme. Et aussi, la probabilité de cette condition de blocage augmente proportionnellement au trafic sur le réseau. Donc, il vaut certainement la peine de modifier le IMPSYS d'une façon telle que ce blocage ne puisse pas survenir. Il se trouve qu'une modification mineure a déjà réalisé l'effet désiré. Souvenez vous que le blocage décrit ne peut survenir que parce que des messages d'un seul paquet et des messages à paquets multiples utilisent le même pôle de mémoires tampon de réassemblage. Si on met à part une seule mémoire tampon de réassemblage (ou une pour chaque hôte de destination) qui ne peut être utilisée que par les messages d'un seul paquet, cette condition de blocage due au séquençage de message ne peut pas survenir.

Un autre solution à ce problème est, bien sûr, de passer le séquençage de message du niveau IMP à celui de l'hôte. Cela ne signifie pas qu'un problème similaire de blocage ne puisse pas survenir au niveau de l'hôte. Et aussi, il est actuellement beaucoup plus facile de résoudre ce problème par une légère modification de l'IMPSYS plutôt que d'avoir à coordonner tous les divers hôtes. Une autre solution de remplacement est de cesser d'utiliser des messages à plusieurs paquets. Cependant, cela implique aussi des changements beaucoup plus radicaux qui nécessitent une attention approfondie.

Références

1. Kahn, R. E. and W. R. Crowther, "Flow Control in a Resource-Sharing Computer Network," IEEE Transactions on Communication, Volume COM-20,3 juin 1972.
2. BBN Report No. 2717, "Interface Message Processors for the ARPA Computer Network," Quarterly Technical Report No. 4, janvier 1974.
3. W. Naylor, J. Wong, C. Kline et J. Postel, "À propos de l'offre officielle de ICP," RFC 143, NIC 6728, avril 1971.
4. Postel, J. B. "A Graph Model Analysis of Computer Communications Protocols," Thèse de doctorat, University of California, Los Angeles.
5. BBN Report No. 2670. "Natural Communication with Computers IV," Quarterly Progress Report No. 12, octobre 1973.