Groupe de travail Réseau
Request for Comments : 2914
BCP : 41
Catégorie : Bonnes pratiques actuelles

S. Floyd, ACIRI
septembre 2000

Traduction Claude Brière de L'Isle

Principes du contrôle d'encombrement

Résumé
Le but du présent document est d'expliquer le besoin du contrôle d'encombrement dans l'Internet, et de discuter ce qui constitue un contrôle d'encombrement correct. Un but spécifique est d'illustrer les dangers de négliger d'appliquer un contrôle d'encombrement approprié. Un second but est de discuter le rôle de l'IETF dans la normalisation de nouveaux protocoles de contrôle d'encombrement.

1. Introduction

Le présent document emprunte beaucoup à des RFC antérieures, et dans certains cas reproduit des paragraphes entiers du texte de documents antérieurs [RFC2309], [RFC2357]. On a aussi beaucoup fait d'emprunts à des publications antérieures qui traitent du contrôle d'encombrement de bout en bout [FF99].

2. Normes actuelles sur le contrôle d'encombrement

Les normes de l'IETF qui concernent le contrôle d'encombrement de bout en bout se concentrent soit sur des protocoles spécifiques (par exemple, TCP [RFC2581], des protocoles de diffusion groupée fiable [RFC2357]) soit sur la syntaxe et la sémantique des communications entre les nœuds d'extrémité et les routeurs sur les informations d'encombrement (par exemple, la notification explicite d'encombrement [RFC2481]) ou la qualité de service désirée (diff-serv)). Le rôle du contrôle d'encombrement de bout en bout est aussi discuté dans une RFC pour information des "Recommandations sur la gestion de file d'attente et l'évitement d'encombrement dans l'Internet" [RFC2309]. La RFC2309 recommande le déploiement de mécanismes de gestion active de file d'attente dans les routeurs, et la poursuite des efforts de conception sur les mécanismes dans les routeurs pour traiter les flux qui ne répondent pas aux notifications d'encombrement. On emprunte librement à la RFC2309 une partie de sa discussion générale du contrôle d'encombrement de bout en bout.

À la différence des RFC mentionnées ci-dessus, le présent document est une discussion plus générale des principes du contrôle d'encombrement. Une des clés du succès de l'Internet a été le mécanismes d'évitement d'encombrement de TCP. Alors que TCP est toujours le protocole de transport dominant dans l'Internet, il n'est plus omniprésent, et il y a un nombre croissant d'applications qui, pour une raison ou une autre, choisissent de ne pas utiliser TCP. Un tel trafic inclut non seulement le trafic en diffusion groupée, mais aussi du trafic en envoi individuel, comme le multimédia en direct qui n'exige pas de fiabilité; et du trafic comme le DNS ou l'acheminement de messages qui consistent en courts transferts réputés critiques pour le fonctionnement du réseau. Beaucoup de ce trafic n'utilise aucune forme de réservation de bande passante ou de contrôle d'encombrement de bout en bout. L'utilisation continue du contrôle d'encombrement de bout en bout par du trafic au mieux est critique pour maintenir la stabilité de l'Internet.

Le présent document discute aussi le rôle général de l'IETF dans la normalisation de nouveaux protocoles de contrôle d'encombrement.

La discussion des principes du contrôle d'encombrement pour les services différenciés ou les services intégrés n'est pas traitée dans le présent document. Certaines catégories de services intégrés ou différenciés incluent une garantie, par le réseau, de bande passante de bout en bout, et à ce titre n'exigent pas de mécanisme de contrôle d'encombrement de bout en bout.

3. Développement du contrôle d'encombrement de bout en bout

3.1 Prévention de l'écroulement par encombrement

L'architecture du protocole Internet se fonde sur un service de paquet de bout en bout sans connexion qui utilise le protocole IP. Les avantages de sa conception sans connexion, souplesse et robustesse, ont été amplement démontrés. Cependant, ces avantages ne sont pas sans coût : une conception soigneuse est requise pour fournir un bon service face à une lourde charge. En fait, le manque d'attention à la dynamique de la transmission de paquet peut résulter en une sévère dégradation de service ou "débâcle de l'Internet". Ce phénomène a été observé pour la première fois durant la phase précoce de croissance de l'Internet du milieu des années 1980 [RFC0896], et est appelée techniquement "écroulement par encombrement".

La spécification d'origine de TCP [RFC0793] incluait un contrôle de flux fondé sur la fenêtre comme moyen pour le receveur de diriger la quantité de données envoyées par l'expéditeur. Ce contrôle de flux était utilisé pour prévenir le débordement de l'espace de mémoire tampon de données du receveur disponible pour cette connexion. La [RFC0793] disait que les segments pouvaient être perdus soit à cause d'erreurs, soit à cause d'encombrement du réseau, mais n'incluait pas d'ajustement dynamique de la fenêtre de contrôle de flux en réponse à l'encombrement.

Le remède original à la débâcle de l'Internet a été fourni par Van Jacobson. En commençant en 1986, Jacobson a développé les mécanismes d'évitement d'encombrement qui sont maintenant exigés dans les mises en œuvre de TCP [Jacobson88], [RFC2581]. Ces mécanismes fonctionnent dans les hôtes pour forcer les connexions TCP à un "repli" durant l'encombrement. On dit que les flux TCP sont "réactifs" (c'est-à-dire, ils éliminent des paquets) aux signaux d'encombrement provenant du réseau. Ce sont ces algorithmes d'évitement d'encombrement de TCP qui empêchent l'écroulement par encombrement dans l'Internet d'aujourd'hui.

Cependant, ceci n'est pas la fin de l'histoire. Des recherches considérables ont été effectuées sur la dynamique de l'Internet depuis 1988, et l'Internet a grandi. Il est devenu clair que les mécanismes TCP d'évitement d'encombrement [RFC2581], bien que nécessaires et puissants, ne sont pas suffisants pour fournir un bon service en toutes circonstances. En plus du développement de nouveaux mécanismes de contrôle d'encombrement [RFC2357], des mécanismes fondés sur le routeur sont en cours de développement et complètent les mécanismes d'évitement d'encombrement de point d'extrémité.
Une question majeure qui reste encore à traiter est celle du potentiel de futurs écroulements par encombrement de l'Internet dus aux flux qui n'utilisent pas de contrôles d'encombrement de bout en bout réactifs. La [RFC0896] suggérait en 1984 que les passerelles devraient détecter et "éteindre" les hôtes qui se conduisent mal : "Le manquement à réagir à un message ICMP Extinction de source, devrait cependant être considéré comme justifiant une action de déconnexion d'un hôte par une passerelle. Détecter de tels manquements n'est pas trivial mais c'est un domaine qui vaut la peine qu'on y effectue des recherches plus approfondies". Les articles actuels proposent toujours que les routeurs détectent et pénalisent les flux qui n'utilisent pas de contrôle d'encombrement de bout en bout acceptable [FF99].

3.2 Équité

En plus du souci de l'écroulement par encombrement, il y a celui de "l'équité" à l'égard du trafic au mieux. Comme TCP "se replie" durant l'encombrement, un grand nombre de connexions TCP peuvent partager une seule liaison encombrée d'une façon telle que la bande passante soit partagée d'une façon raisonnablement équitable entre les flux de situation similaire. Le partage équitable de bande passante entre les flux dépend du fait que tous les flux utilisent des algorithmes compatibles de contrôle d'encombrement. Pour TCP, cela signifie que les algorithmes de contrôle d'encombrement se conforment à la spécification TCP actuelle [RFC0793], [RFC1122], [RFC2581].

La question de l'équité entre les flux en compétition devient de plus en plus importante pour plusieurs raisons. D'abord, en utilisant l'adaptation de fenêtre [RFC1323], les TCP individuels peuvent utiliser une forte bande passante même sur des chemins à fort délai de propagation. Ensuite, avec la croissance de la Toile, les utilisateurs de l'Internet veulent de plus en plus de grosses bandes passantes et des communications à faibles délais, plutôt que le transfert tranquille d'un long fichier en arrière plan. La croissance du trafic au mieux qui n'utilise pas TCP souligne ce problème de l'équité entre des trafics au mieux en concurrence dans les périodes d'encombrement.

La popularité de l'Internet a causé une prolifération du nombre de mises en œuvre de TCP. Certaines d'entre elles peuvent échouer à mettre en œuvre correctement les mécanismes TCP d'évitement d'encombrement à cause d'une mauvaise mise en œuvre [RFC2525]. D'autres peuvent être délibérément mises en œuvre avec des algorithmes d'évitement d'encombrement qui sont plus agressifs dans leur utilisation de la bande passante que d'autres mises en œuvre TCP ; cela permettrait à un fabricant de prétendre avoir un "TCP plus rapide". La conséquence logique de telles mises en œuvre serait une spirale de mises en œuvre de TCP de plus en plus agressives, ou de protocoles de transport de plus en plus agressifs, ramenant au point où il n'y a effectivement plus d'évitement d'encombrement et où l'encombrement de l'Internet est chronique.

Il y a un moyen bien connu d'obtenir des performance plus agressives sans même changer le protocole de transport, en changeant le niveau de granularité : ouvrir plusieurs connexions avec le même lieu, comme cela a été fait dans le passé par certains navigateurs de la Toile. Et donc, au lieu d'une spirale de protocoles de transport de plus en plus agressifs, on aurait à la place une spirale de navigateurs de la Toile de plus en plus agressifs, ou d'applications de plus en plus agressives.

Cela soulève le problème de la granularité appropriée d'un "flux", où on définit un "flux" comme le niveau de granularité approprié pour l'application de l'équité et du contrôle d'encombrement. D'après la RFC2309 : "Il y a quelques réponses "naturelles" : 1) une connexion TCP ou UDP (adresse/accès de source, adresse/accès de destination) ; 2) une paire d'hôtes de source/destination ; 3) un certain hôte de source ou de destination. On pourrait deviner que la paire d'hôtes de source/destination donne la granularité la plus appropriée dans de nombreuses circonstances. La granularité des flux pour la gestion de l'encombrement est, au moins en partie, une question de politique qui doit être réglée dans la plus large communauté de l'IETF".

Empruntant encore à la RFC2309, on utilise le terme "compatible TCP" pour un flux qui se comporte en présence d'encombrement comme un flux produit par un TCP conforme. Un flux TCP compatible répond aux notifications d'encombrement, et en état normal n'utilise pas plus de bande passante qu'un TCP conforme fonctionnant dans des conditions comparables (taux d'abandon, RTT, MTU, etc.).

Il est pratique de diviser les flux en trois classes : (1) flux compatibles TCP, (2) flux non réactifs, c'est-à-dire, flux qui ne ralentissent pas lorsque survient l'encombrement, et (3) flux qui réagissent mais ne sont pas compatibles TCP. Les deux dernières classes contiennent des flux plus agressifs qui font peser des menaces significatives sur les performances de l'Internet, comme on l'expose ci-dessous.

En plus de l'équité en état normal, l'équité du démarrage lent initial est aussi un problème. Il y a l'effet transitoire sur les autres flux d'un flux qui a une procédure de démarrage lent sur-agressive. Les performances de démarrage lent sont particulièrement importantes pour les nombreux flux qui ont une durée de vie brève, et n'ont qu'une petite quantité de données à transférer.

3.3 Optimisation des performances à l'égard du débit, du délai, et des pertes

En plus d'empêcher l'écroulement par encombrement et les problèmes d'équité, une troisième raison pour qu'un flux utilise le contrôle d'encombrement de bout en bout peut être d'optimiser ses propres performances en ce qui concerne le débit, le délai, et les pertes. Dans certaines circonstances, par exemple, dans des environnements de fort multiplexage statistique, le délai et le taux de perte rencontrés par un flux sont largement indépendants de son propre taux d'envoi. Cependant, dans des environnements avec de plus faibles niveaux de multiplexage statistique ou avec une programmation par flux, le délai et le taux de perte rencontrés par un flux sont en partie fonction du propre taux d'envoi du flux. Donc, un flux peut utiliser le contrôle d'encombrement de bout en bout pour limiter le délai ou la perte subis par ses propres paquets. On notera cependant que dans un environnement comme l'Internet au mieux actuel, les soucis concernant l'écroulement par encombrement et l'équité de traitement des flux concurrents limitent la gamme des comportements de contrôle d'encombrement disponibles pour un flux.

4. Rôle du processus de normalisation

La normalisation d'un protocole de transport inclut non seulement la normalisation des aspects du protocole qui pourraient affecter l'interopérabilité (par exemple, les informations échangées par les nœuds d'extrémité) mais aussi la normalisation des mécanismes réputés critiques pour les performances (par exemple, dans TCP, la réduction de la fenêtre d'encombrement en réponse à un abandon de paquet). En même temps, les détails spécifiques de la mise en œuvre et autres aspects du protocole de transport qui n'affectent pas l'interopérabilité et n'interfèrent pas significativement avec les performances n'exigent pas de normalisation. Les zones de TCP qui n'exigent pas de normalisation incluent les détails de la procédure de récupération rapide de TCP après une retransmission rapide [RFC2582]. La Section 9 utilise des exemples de TCP pour exposer plus en détails le rôle du processus de normalisation dans le développement du contrôle d'encombrement.

4.1 Développement de nouveaux protocoles de transport

En plus de s'intéresser au danger de l'écroulement par encombrement, le processus de normalisation pour de nouveaux protocoles de transport prend soin d'éviter une "course aux armements" du contrôle d'encombrement entre les protocoles en compétition. Par exemple, dans la [RFC2357] les directeurs de zone TSV et leurs conseils de direction soulignent les critères pour la publication comme RFC de projets Internet sur les protocoles fiables de transport de diffusion groupée. D'après la [RFC2357] : "Un souci particulier de l'IETF est l'impact du trafic fiable en diffusion groupée sur les autres trafic de l'Internet dans les périodes d'encombrement, en particulier l'effet du trafic fiable en diffusion groupée sur le trafic TCP en compétition avec lui.... Le défi pour l'IETF est d'encourager la recherche et les mises en œuvre de diffusion groupée fiable, et de permettre que le besoin de diffusion groupée fiable des applications soit satisfait aussi rapidement que possible, tout en protégeant en même temps l'Internet de l'écroulement ou désastre par encombrement qui pourrait résulter de la large utilisation d'applications ayant des mécanismes inappropriés de diffusion groupée fiable."

La liste des critères techniques qui doivent être pris en compte par les RFC sur les nouveaux protocoles de transport de diffusion groupée fiable inclut les suivants : "Y a-t-il un mécanisme de contrôle d'encombrement ? Quelles sont ses performances ? Quand est-il mis en échec ? Noter que les mécanismes de contrôle d'encombrement qui fonctionnent plus agressivement sur le réseau que TCP vont devoir prouver réellement qu'ils ne menacent pas la stabilité du réseau."

Il est raisonnable de s'attendre à ce que ces soucis quant à l'effet de nouveaux protocoles de transport sur le trafic concurrent vont s'appliquer non seulement aux protocoles de diffusion groupée fiable, mais aussi à l'envoi individuel non fiable, à l'envoi individuel fiable, et à la diffusion groupée non fiable.

4.2 Questions de niveau application qui affectent le contrôle d'encombrement

La question spécifique d'un navigateur qui ouvre plusieurs connexions avec la même destination a été réglée par la [RFC2616], qui déclare au paragraphe 8.1.4 que "les clients qui utilisent des connexions persistantes DEVRAIENT limiter le nombre de connexions simultanées qu'elles entretiennent avec un serveur donné. Un client d'un seul utilisateur NE DEVRAIT PAS entretenir plus de deux connexions avec un serveur ou mandataire quelconque."

4.3 Nouveaux développements dans le processus de normalisation

Les développements les plus évidents qui dans l'IETF pourraient affecter l'évolution du contrôle d'encombrement sont ceux des services intégrés et différenciés [RFC2212], [RFC2475] et de la notification explicite d'encombrement (ECN, *Explicit Congestion Notification*) [RFC2481]. Cependant, d'autres développements moins radicaux vont vraisemblablement affecter aussi le contrôle d'encombrement.

Un de ces efforts est de construire le gestionnaire d'encombrement de point d'extrémité [RFC3124], pour permettre à plusieurs flux concurrents d'un envoyeur au même receveur de partager l'état de contrôle d'encombrement. En permettant à plusieurs connexions avec la même destination d'agir comme un flux en termes de contrôle d'encombrement de bout en bout, un gestionnaire d'encombrement pourrait permettre à des connexions individuelles en démarrage lent de tirer parti des informations précédentes sur l'état d'encombrement du chemin de bout en bout. De plus, l'utilisation d'un gestionnaire d'encombrement pourrait supprimer les dangers du contrôle d'encombrement de flux multiples ouverts entre la même paire de source/destination, et pourrait peut-être être utilisée pour permettre à un navigateur d'ouvrir des connexions simultanées avec la même destination.

5. Description de l'écroulement par encombrement

Cette section expose l'écroulement par encombrement provenant de paquets non livrés avec un certain niveau de détail, et montre comment les flux non réactifs pourraient contribuer à un écroulement par encombrement dans l'Internet. Cette section s'appuie fortement sur les matériaux de [FF99].

De façon informelle, l'écroulement par encombrement survient lorsque une augmentation de la charge du réseau résulte en une diminution du travail utile fait par le réseau. Comme exposé à la Section 3, l'écroulement par encombrement a été rapporté pour la première dois au milieu des années 1980 [RFC0896], et été largement dû à des connexions TCP retransmettant inutilement des paquets qui étaient soit en transit, soit avaient déjà été reçus à destination. On appelle l'écroulement par encombrement qui résulte de la retransmission inutile de paquets l'écroulement par encombrement classique. L'écroulement par encombrement classique est une condition stable qui peut résulter en un débit qui est une petite fraction du débit normal [RFC0896]. Les problèmes de l'écroulement par encombrement classique ont généralement

été corrigés par les améliorations des temporisations et les mécanismes de contrôle d'encombrement dans les mises en œuvre modernes de TCP [Jacobson88].

Une seconde forme potentielle d'écroulement par encombrement survient à cause des paquets non livrés. L'écroulement par encombrement venant de paquets non livrés survient lorsque la bande passante est gaspillée par la livraison de paquets à travers le réseau qui sont abandonnés avant d'atteindre leur destination ultime. C'est probablement le plus grand danger non résolu par rapport à l'écroulement par encombrement dans l'Internet d'aujourd'hui. Différents scénarios peuvent résulter en différents degrés d'écroulement par encombrement, en termes de fraction de la bande passante de la liaison encombrée utilisée pour un travail productif. Le danger d'écroulement par encombrement provenant de paquets non livrés est principalement dû au déploiement croissant d'applications en boucle ouverte qui n'utilisent pas le contrôle d'encombrement de bout en bout. Encore plus destructives seraient des applications au mieux qui *augmentent* leur taux d'envoi en réponse à une augmentation du taux d'abandon de paquets (par exemple, en utilisant automatiquement un niveau de FEC supérieur).

Le Tableau 1 donne les résultats d'un scénario avec écroulement par encombrement provenant de paquets non livrés, où une bande passante faible est gaspillée par des paquets qui n'atteignent jamais leur destination. La simulation utilise un scénario avec trois flux TCP et un flux UDP qui sont en compétition sur une liaison encombrée à 1,5 Mbit/s. La liaison d'accès pour tous les nœuds est à 10 Mbit/s, sauf la liaison d'accès au receveur du flux UDP qui est à 128 kbit/s, seulement 9 % de la bande passante de la liaison partagée. Lorsque le taux de la source UDP excède 128 kbit/s, la plupart des paquets UDP seront éliminés à l'accès de sortie de cette liaison finale.

Taux d'arrivée UDP	Débit UDP	Débit TCP	Débit total
0,7	0,7	98,5	99,2
1,8	1,7	97,3	99,1
2,6	2,6	96,0	98,6
5,3	5,2	92,7	97,9
8,8	8,4	87,1	95,5
10,5	8,4	84,8	93,2
13,1	8,4	81,4	89,8
17,5	8,4	77,3	85,7
26,3	8,4	64,5	72,8
52,6	8,4	38,1	46,4
58,4	8,4	32,8	41,2
65,7	8,4	28,5	36,8
75,1	8,4	19,7	28,1
87,6	8,4	11,3	19,7
105,2	8,4	3,4	11,8
131,5	8,4	2,4	10,7

Tableau 1 : Simulation avec trois flux TCP et un flux UDP.

Le Tableau 1 montre le taux d'arrivée UDP depuis l'envoyeur, le débit utile UDP (défini comme la bande passante délivrée au receveur) le débit utile TCP (comme délivré aux receveurs TCP) et le débit utile agrégé sur la liaison encombrée à 1,5 Mbit/s. Chaque taux est donné comme fraction de la bande passante de la liaison encombrée. Lorsque le taux de la source UDP augmente, le débit utile TCP diminue en gros de façon linéaire, et le débit utile UDP est presque constant. Donc, comme le flux UDP augmente sa charge offerte, son seul effet est de comprimer le débit utile TCP et agrégé. Sur la liaison encombrée, le flux UDP "gaspille" finalement la bande passante qui aurait pu être utilisée par le flux TCP, et réduit le débit utile global dans le réseau à une petite fraction de la bande passante de la liaison encombrée.

Les simulations du Tableau 1 illustrent à la fois l'inéquité et l'écroulement par encombrement. Comme l'expose [FF99], le contrôle d'encombrement compatible n'est pas le seul moyen d'assurer l'équité ; la programmation par flux chez les routeurs encombrés est un autre mécanisme qui garantit l'équité au niveau des routeurs. Cependant, comme l'explique [FF99], on ne peut pas s'appuyer sur la programmation par flux pour prévenir l'écroulement par encombrement.

Il n'y a que deux solutions pour éliminer le danger d'écroulement par encombrement provenant des paquets non livrés. La première est l'utilisation effective du contrôle d'encombrement de bout en bout par les nœuds d'extrémité. Plus précisément, l'exigence serait qu'un flux évite un schéma de pertes significatives sur les liaisons en aval de la première liaison encombrée sur le chemin. (Ici, on va considérer toute liaison comme "liaison encombrée" si un flux quelconque utilise de la bande passante qui aurait autrement été utilisée par d'autre trafic sur la liaison.) Étant donné qu'un nœud d'extrémité est généralement incapable de distinguer entre un chemin avec une liaison encombrée et un chemin avec plusieurs liaisons encombrées, la façon la plus fiable pour qu'un flux évite un schéma de pertes significatives à une liaison encombrée en aval est que le flux utilise le contrôle d'encombrement de bout en bout, et réduise son taux d'envoi en présence de pertes.

Une seconde solution de remplacement pour prévenir l'écroulement par encombrement provenant de paquets non livrés serait la garantie par le réseau que les paquets acceptés à une liaison encombrée du réseau seront livrés de toute façon au receveur [RFC2212], [RFC2475]. On note que le choix entre la première solution de contrôle d'encombrement de bout en bout et la seconde alternative de garanties de bande passante de bout en bout n'a pas à être une décision exclusive ; l'écroulement par encombrement peut être empêché par l'utilisation d'un contrôle d'encombrement effectif de bout en bout par une partie du trafic, et l'utilisation des garanties de bande passante de bout en bout de la part du réseau pour le reste du trafic.

6. Formes de contrôle d'encombrement de bout en bout

Le présent document a exposé les problèmes de l'écroulement par encombrement et de l'équité pour TCP pour de nouvelles formes de contrôle d'encombrement. Cela ne signifie pas cependant que les problèmes concernant l'écroulement par encombrement et l'équité pour TCP nécessitent que tout le trafic au mieux déploie le contrôle d'encombrement sur la base de l'algorithme d'augmentation additive, diminution multiplicative (AIMD, *Additive-Increase Multiplicative-Decrease*) de TCP de réduction de moitié du taux d'envoi en réponse à chaque abandon de paquet. Cette section expose séparément les implications de ces deux problèmes d'écroulement par encombrement et d'équité avec TCP.

6.1 Contrôle d'encombrement de bout en bout pour éviter un écroulement par encombrement

L'évitement de l'écroulement par encombrement provenant de la non livraison de paquets exige que les flux évitent un scénario de forts taux d'envoi, plusieurs liaisons encombrées, et un fort taux persistant d'abandon de paquet sur la liaison aval. Comme l'écroulement par encombrement provenant de la non livraison de paquets consiste en paquets qui gaspillent une bande passante précieuse pour être ensuite éliminés en aval, cette forme d'écroulement par encombrement n'est pas possible dans un environnement où chaque flux traverse seulement une liaison encombrée, ou lorsque seulement un petit nombre de paquets sont éliminés sur des liaisons en aval de la première liaison encombrée. Donc, toute forme de contrôle d'encombrement qui réussit à éviter un fort taux d'envoi en présence d'un fort taux de perte de paquets devrait être suffisante pour éviter l'écroulement par encombrement provenant de la non livraison de paquets.

On notera que l'ajout de la notification explicite d'encombrement (ECN, *Explicit Congestion Notification*) à l'architecture IP ne va pas, par sa seule vertu, supprimer le danger d'écroulement par encombrement pour le trafic au mieux. ECN permet aux routeurs d'établir un bit dans l'en-tête des paquet pour indiquer l'encombrement aux nœuds d'extrémité, plutôt que d'être forcés de s'appuyer sur les abandons de paquets pour indiquer l'encombrement. Cependant, avec ECN, le marquage des paquets ne va remplacer l'abandon de paquet que dans les périodes d'encombrement modéré. En particulier, lorsque l'encombrement est sévère et que les mémoires tampon d'un routeur débordent, le routeur n'a d'autre choix que d'éliminer les paquets qui arrivent.

6.2 Contrôle d'encombrement de bout en bout pour l'équité envers TCP

Le souci exprimé dans la [RFC2357] au sujet de l'équité avec TCP fait peser une contrainte significative mais pas paralysante sur la gamme des mécanismes viables de contrôle d'encombrement de bout en bout pour le trafic au mieux. Un environnement avec programmation par flux à toutes les liaisons encombrées isolerait les flux les uns des autres, et éliminerait le besoin que les mécanismes de contrôle d'encombrement soient compatibles TCP. Un environnement de services différenciés, où les flux marqués comme appartenant à une certain classe diff-serv seraient programmés à l'abri du trafic au mieux, pourrait permettre l'émergence d'une classe de trafic diff-serv entière où le contrôle d'encombrement n'aurait pas besoin d'être compatible TCP. De même, dans un environnement à tarif contrôlé, ou avec une classe diff-serv qui a ses propres règles de tarification, le souci de l'équité vis à vis de TCP pourrait être oublié. Cependant, pour l'environnement Internet actuel, où les autres trafics au mieux peuvent être en compétition, dans une file d'attente au premier entré, premier sorti, avec du trafic TCP, l'absence d'équité vis à vis de TCP pourrait conduire à ce qu'un flux "affame" un autre flux dans un temps de fort encombrement, comme cela a été illustré dans le Tableau 1 ci-dessus.

Cependant, la liste des procédures de contrôle d'encombrement compatible TCP n'est pas limitée à AIMD ave les mêmes paramètres d'augmentation/diminution que TCP. D'autres procédures de contrôle d'encombrement compatible TCP incluent des variantes d'AIMD fondées sur le taux ; AIMD avec différents ensembles de paramètres d'augmentation/diminution qui donnent le même comportement en état de marche ; le contrôle d'encombrement fondé sur des équations où l'envoyeur ajuste son taux d'envoi en réponse aux informations sur le taux d'abandon de paquets à long terme ; la diffusion groupée en couches où les receveurs s'abonnent et se désabonnent à des groupes de diffusion groupée rangés en couches ; et éventuellement d'autres formes qu'on a pas encore commencé d'envisager.

7. Remerciements

Beaucoup du présent document s'inspire directement des précédentes RFC sur le contrôle d'encombrement de bout en bout. Il essaye d'être un résumé des idées qui ont été discutées depuis de nombreuses années, et par de nombreuses personnes. En particulier, il faut mentionner les membres du groupe de recherche de bout en bout, le groupe de recherche sur la diffusion groupée fiable, et le conseil de direction de la zone Transport. Le présent document a aussi bénéficié des discussions et des réactions du groupe de travail de la zone Transport. Des remerciements particuliers sont dus à Mark Allman pour ses réactions sur une version antérieure du présent document.

8. Références

[FF99] Floyd, S. et K. Fall, "Promoting the Use of End-to-End Congestion Control in the Internet", IEEE/ACM Transactions on Networking, August 1999. URL http://www.aciri.org/floyd/end2end-paper.html

[Jacobson88] V. Jacobson, Congestion Avoidance et Control, ACM SIGCOMM '88, août 1988.

[RFC0793] J. Postel (éd.), "Protocole de commande de transmission – Spécification du protocole du programme Internet DARPA", STD 7, septembre 1981.

[RFC0896] J. Nagle, "Contrôle de l'encombrement dans l'inter-réseau IP/TCP", janvier 1984.

[RFC1122] R. Braden, "Exigences pour les hôtes Internet – couches de communication", STD 3, octobre 1989. *(MàJ par la* RFC6633)

[RFC1323] V. Jacobson, R. Braden et D. Borman, "Extensions TCP pour de bonnes performances", mai 1992.

[RFC2119] S. Bradner, "Mots clés à utiliser dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997.

[RFC2212] S. Shenker, C. Partridge, R. Guerin, "Spécification de la qualité de service garantie", septembre 1997. *(P.S.)*

[RFC2309] B. Braden et autres, "Recommandations sur la gestion de file d'attente et l'évitement d'encombrement dans l'Internet", avril 1998.

[RFC2357] A. Mankin, A. Romanov, S. Bradner et V. Paxson, "Critères de l'IETF pour l'évaluation des protocoles de transport et d'application de diffusion groupée fiable", juin 1998. *(Information)*

[RFC2414] M. Allman, S. Floyd, C. Partridge, "Accroissement de la fenêtre initiale de TCP", septembre 1998. *(Expérimentale)* (*Obsolète, voir* RFC3390)

[RFC2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang et W. Weiss, "Architecture pour services différenciés", décembre 1998. *(MàJ par* RFC3260*)*

[RFC2481] K. Ramakrishnan S. Floyd, "Proposition d'ajout de la notification d'encombrement explicite (ECN) à IP", janvier 1999.

[RFC2525] V. Paxson et autres, "Problèmes connus de mise en œuvre de TCP", mars 1999. *(Information)*

[RFC2581] M. Alman, V. Paxson et W. Stevens, "Contrôle d'encombrement avec TCP", avril 1999. *(Obsolète, voir* RFC5681)

[RFC2582] S. Floyd, T. Henderson, "Modification NewReno à l'algorithme de récupération rapide de TCP", avril 1999. (*Obsolète, voir* RFC3782) *(Expérimentale)*

[RFC2616] R. Fielding et autres, "Protocole de transfert hypertexte -- HTTP/1.1", juin 1999. *(D.S., MàJ par* 2817, 6585*)*

[RFC2861] M. Handley, J. Padhye, S. Floyd, "Validation de fenêtre d'encombrement TCP", juin 2000. *(Expérimentale)*

[RFC3124] H. Balakrishnan et S Seshan, "Le gestionnaire d'encombrement", juin 2001. *(P.S)*

[RFC3150] S. Dawkins et autres, "Implications des liaisons lentes sur les performances de bout en bout", juillet 2001. (BCP0048)

[SCWA99] S. Savage, N. Cardwell, D. Wetherall, et T. Anderson, "TCP Congestion Control with a Misbehaving Receiver", ACM Computer Communications Review, octobre 1999.

[TCPB98] Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan, Mark Stemm, et Randy H. Katz, "TCP Behavior of a Busy Internet Server: Analysis et Improvements", IEEE Infocom, mars 1998. Disponible à http://www.cs.berkeley.edu/~hari/papers/infocom98.ps.gz

[TCPF98] Dong Lin et H.T. Kung, "TCP Fast Recovery Strategies: Analysis et Improvements", IEEE Infocom, mars 1998. Disponible à http://www.eecs.harvard.edu/networking/papers/infocom-tcp-final-198.pdf

9. Problèmes spécifiques de TCP

Dans cette section, on expose certaines particularités du contrôle d'encombrement TCP, pour illustrer une réalisation des principes du contrôle d'encombrement, incluant certains des détails qui ressortent lorsque on les incorpore dans une production de protocole de transport.

9.1 Démarrage lent

L'envoyeur TCP ne peut pas ouvrir une nouvelle connexion en envoyant une grosse salve de données (par exemple, la fenêtre annoncée du receveur) tout d'un coup. L'envoyeur TCP est limité à une petite valeur initiale par la fenêtre d'encombrement. Durant le démarrage lent, l'envoyeur TCP peut augmenter le taux d'envoi d'au plus un facteur de deux par délai d'aller-retour. Le démarrage lent se termine lorsque de l'encombrement est détecté, ou lorsque la fenêtre d'encombrement de l'envoyeur est supérieure au seuil de démarrage lent ssthresh.

Une question qui peut affecter le contrôle d'encombrement global, et a donc été explicitement visée par le processus de normalisation, inclut l'augmentation de la valeur de la fenêtre initiale [RFC2414], [RFC2581].

Les questions qui n'ont pas été visées par le processus de normalisation, et sont généralement considérées comme n'exigeant pas de normalisation, incluent des questions comme l'utilisation (ou non utilisation) de ralentisseurs fondés sur le taux d'envoi, et des mécanismes pour terminer plus tôt le démarrage lent, avant que la fenêtre d'encombrement n'atteigne ssthresh. De tels mécanismes résultent en un comportement de démarrage lent qui est aussi prudent ou plus prudent que celui du TCP standard.

9.2 Augmentation additive, diminution multiplicative

En l'absence d'encombrement, l'envoyeur TCP augmente sa fenêtre d'encombrement d'au moins un paquet par délai d'aller-retour. En réponse à une indication d'encombrement, l'envoyeur TCP divise sa fenêtre d'encombrement par deux. (Plus précisément, la nouvelle fenêtre d'encombrement est la moitié du minimum de la fenêtre d'encombrement et de la fenêtre annoncée du receveur.)

Une question qui peut affecter le contrôle d'encombrement global, et qui pourrait donc vraisemblablement être explicitement traitée par le processus de normalisation, inclurait une proposition d'ajout des "purs ACK" au contrôle d'encombrement du flux de retour.

Une question qui n'a pas été visée par le processus de normalisation, et n'est généralement pas considérée comme exigeant de normalisation, serait de changer la fenêtre d'encombrement pour appliquer une limite supérieure au nombre d'octets présumés être dans le tuyau, au lieu d'appliquer une fenêtre glissante commençant à partir de l'accusé de réception cumulatif. (En clair, la fenêtre annoncée du receveur s'applique comme une fenêtre glissante commençant au champ Accusé de réception cumulatif, parce que les paquets reçus au dessus du champ Accusé de réception cumulatif sont détenus dans la mémoire tampon du receveur TCP, et n'ont pas été livrés à l'application. Cependant, la fenêtre d'encombrement s'applique au nombre de paquets en cours dans le réseau, et n'a pas nécessairement à inclure les paquets qui ont été reçus déclassés par le receveur TCP.)

9.3 Temporisateurs de retransmission

L'envoyeur TCP établit un temporisateur de retransmission pour déduire qu'un paquet a été éliminé dans le réseau. Lorsque le temporisateur de retransmission arrive à expiration, l'envoyeur en déduit qu'un paquet a été perdu, il règle ssthresh à la moitié de la fenêtre actuelle, et passe en démarrage lent, en retransmettant le paquet perdu. Si le temporisateur de retransmission expire parce que aucun accusé de réception n'a été reçu pour une paquet retransmis, le temporisateur de retransmission est aussi "réduit", en doublant la valeur du prochain intervalle de temporisation de retransmission.

Une question qui peut affecter le contrôle d'encombrement global, et donc sera vraisemblablement visée explicitement par le processus de normalisation, pourrait inclure une modification du mécanisme d'établissement du temporisateur de retransmission qui pourrait significativement augmenter le nombre de temporisations de retransmission qui expirent prématurément, lorsque l'accusé de réception n'est pas encore arrivé à l'envoyeur, mais qu'en fait aucun paquet n'a été abandonné. Cela pourrait poser un problème au processus de normalisation de l'Internet parce que les temporisateurs de retransmission qui expirent prématurément pourraient conduire à un accroissement du nombre de paquets transmis inutilement sur une liaison encombrée.

9.4 Retransmission rapide et récupération rapide

Après avoir vu trois accusés de réception dupliqués, l'envoyeur TCP déduit une perte de paquet. L'envoyeur TCP règle ssthresh à la moitié de la fenêtre actuelle, réduit la fenêtre d'encombrement à au plus la moitié de la fenêtre précédente, et retransmet le paquet perdu.

Une question qui peut affecter le contrôle d'encombrement global, et donc sera vraisemblablement visée explicitement par le processus de normalisation, pourrait inclure une proposition (il faudrait que quelqu'un la fasse) pour déduire la perte d'un paquet après seulement un ou deux accusés de réception dupliqués. Si elle est mal conçue, une telle proposition pourrait conduire à augmenter le nombre de paquets inutilement transmis sur une liaison encombrée.

Une question qui n'a pas encore été visée par le processus de normalisation, et dont on ne s'attend pas à ce qu'elle exige de normalisation, serait une proposition d'envoyer un "nouveau" paquet ou un paquet présumé perdu en réponse à un accusé de réception dupliqué ou partiel, si c'est permis par la fenêtre d'encombrement. Un exemple en serait l'envoi d'un nouveau paquet en réponse à un seul accusé de réception dupliqué, pour garder "l'horloge d'accusé de réception" en marche pour le cas où aucun autre accusé de réception n'arriverait. Un telle proposition est un exemple de changement avantageux qui n'implique pas d'interopérabilité et n'affecte pas le contrôle d'encombrement global, et qui pourrait donc être mis en œuvre par les fabricants sans exiger l'intervention du processus de normalisation de l'IETF. (Cette question a en fait été traitée dans la [RFC3150], qui suggère que des "chercheurs peuvent souhaiter faire des expériences en injectant du nouveau trafic dans le réseau lorsque des accusés de réception dupliqués sont reçus, comme décrit dans [TCPB98] et [TCPF98]."

9.5 Autres aspects du contrôle d'encombrement TCP

D'autres aspects du contrôle d'encombrement TCP qui n'ont été discutés dans aucun des paragraphes ci-dessus incluent la récupération de TCP d'une période d'inactivité ou d'activité limitée par l'application [RFC2861].

10. Considérations pour la sécurité"

Le présent document traite des risques associés au contrôle d'encombrement, ou à l'absence of contrôle d'encombrement. Le paragraphe 3.2 traite de l'inéquité potentielle si des flux concurrents n'utilisent pas de mécanismes de contrôle d'encombrement compatibles, et la Section 5 considère les dangers de l'écroulement par encombrement si les flux n'utilisent pas le contrôle d'encombrement de bout en bout.

Comme le présent document ne propose aucun mécanisme de contrôle d'encombrement spécifique, il n'est pas nécessaire de présenter de mesure de sécurité spécifique associée au contrôle d'encombrement. Cependant, on notera qu'il y a une gamme de considérations pour la sécurité associée au contrôle d'encombrement qui devrait être prise en considération dans les documents de l'IETF.

Par exemple, des mécanismes de contrôle d'encombrement individuels devraient être aussi robustes que possible aux tentatives de nœuds d'extrémité individuels de subvertir le contrôle d'encombrement de bout en bout [SCWA99]. C'est un souci particulier pour le contrôle d'encombrement en diffusion groupée, à cause de la distribution à grande distance du trafic et de plus grandes opportunités pour les receveurs individuels de manquer à rapporter l'encombrement.

La RFC2309 discutait aussi des dangers potentiels pour l'Internet de flux non réactifs, c'est-à-dire, de flux qui ne réduisent pas leur taux d'envoi en présence d'encombrement, et décrit le besoin de mécanismes dans le réseau pour traiter les flux qui ne réagissent pas aux notifications d'encombrement. On notera qu'il y a encore des efforts de recherche, d'ingénierie, de mesures, et de déploiement à faire dans ces domaines.

Comme l'Internet agrège un très grand nombre de flux, le risque pour l'infrastructure globale de la subversion du contrôle d'encombrement par quelques flux individuels est limité. Le risque pour l'infrastructure viendrait plutôt du large déploiement de nombreux nœuds d'extrémité qui ne respecteraient pas le contrôle d'encombrement de bout en bout.

Adresse de l'auteur

Sally Floyd
AT&T Center for Internet Research at ICSI (ACIRI)
téléphone : +1 (510) 642-4274 x189
mél : floyd@aciri.org
URL : http://www.aciri.org/floyd/

Déclaration de droits de reproduction

Remerciement

Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.