Groupe de travail Réseau
Request for Comments : 4341
Catégorie : Sur la voie de la normalisation
Traduction Claude Brière de L'Isle

S. Floyd, ICIR
E. Kohler, UCLA
mars 2006

Profil d'identifiant 2 de protocole de contrôle d'encombrement de datagrammes (DCCP) : Contrôle d'encombrement de style TCP

Statut du présent mémoire
Le présent document spécifie un protocole de l'Internet en cours de normalisation pour la communauté de l'Internet, et appelle à des discussions et suggestions pour son amélioration. Prière de se référer à l'édition en cours des "Protocoles officiels de l'Internet" (STD 1) pour voir l'état de normalisation et le statut de ce protocole. La distribution du présent mémoire n'est soumise à aucune restriction.

Résumé
Le présent document contient le profil pour l'identifiant 2 de contrôle d'encombrement (CCID 2, *Congestion Control Identifier 2*), le contrôle d'encombrement de style TCP, dans le protocole de contrôle d'encombrement de datagrammes (DCCP, *Datagram Congestion Control Protocol*). CCID 2 devrait être utilisé par les envoyeurs qui aimeraient tirer parti de la bande passante disponible dans un environnement de conditions changeant rapidement, et qui sont capables de s'adapter à des changements brusques de la fenêtre d'encombrement typiques du contrôle d'encombrement à augmentation additive, diminution multiplicative (AIMD, *Additive Increase Multiplicative Decrease*) de TCP.

Table des matières

1. Introduction....2
2. Conventions et notation....2
3. Usage....2
3.1 Relations avec TCP....2
3.2 Exemple de demie connexion....3
4. Établissement de connexion....3
5. Contrôle d'encombrement sur des paquets de données....4
5.1 Réponse aux périodes de repos et d'application limitée....5
5.2 Réponse aux données éliminées et au receveur lent....5
5.3 Taille de paquet....5
6. Accusé de réception....6
6.1 Contrôle d'encombrement sur les accusés de réception....6
6.2 Accusé de réception des accusés de réception....7
7. Notification explicite d'encombrement....7
8. Options et caractéristiques....8
9. Considérations sur la sécurité....8
10. Considérations relatives à l'IANA....8
10.1 Codes de réinitialisation....8
10.2 Types d'option....8
10.3 Numéros de caractéristiques....8
11. Remerciements....8
Appendice A. Déduction de diminution du ratio d'accusé de réception....9
Appendice B. Coût des erreurs d'inférence de pertes sur le ratio d'accusés de réception....9
Références normatives....10
Références pour information....11
Adresse des auteurs....11
Déclaration complète de droits de reproduction....11

1. Introduction

Le présent document contient le profil pour l'identifiant 2 de contrôle d'encombrement (CCID 2, *Congestion Control Identifier 2*), le contrôle d'encombrement de style TCP, dans le protocole de contrôle d'encombrement de datagrammes (DCCP, *Datagram Congestion Control Protocol*) [RFC4340]. DCCP utilise des identifiants de contrôle d'encombrement (CCID, *Congestion Control Identifier*) pour spécifier le mécanisme de contrôle d'encombrement utilisé sur une demie connexion.

Le CCID de contrôle d'encombrement de style TCP envoie des données en utilisant une proche variante des mécanismes de contrôle d'encombrement de TCP, incorporant une variante de l'accusé de réception sélectif (SACK, *Selective Acknowledgement*) [RFC2018], [RFC3517]. CCID 2 convient pour des envoyeurs qui peuvent s'adapter à des changements brusques de la fenêtre d'encombrement typiques du contrôle d'encombrement à augmentation additive, diminution multiplicative (AIMD, *Additive Increase Multiplicative Decrease*) de TCP, et est particulièrement utile pour des envoyeurs qui aimeraient tirer parti de la bande passante disponible dans un environnement de conditions à changement rapide. Voir à la Section 3 les exigences d'application.

2. Conventions et notation

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

Une demie connexion DCCP consiste en l'envoi des données d'application par un point d'extrémité et en les accusés de réception correspondants envoyés par l'autre point d'extrémité. Les termes "envoyeur HC" et "receveur HC" notent les points d'extrémité qui envoient respectivement les données d'application et les accusés de réception. Comme les CCID s'appliquent au niveau des demies connexions, on abrège envoyeur HC en "envoyeur" et receveur HC en "receveur" dans ce document. Voir plus d'explications dans la [RFC4340].

Pour simplifier, on dit que les envoyeurs envoient des paquets de données DCCP et que les receveurs envoient des paquets DCCP-Ack. Ces deux catégories sont destinées à inclure les paquets DCCP-DataAck.

Les phrases "marqué ECN" et "marqué" se réfèrent aux paquets marqués "encombrement ECN rencontré" sauf mention contraire.

3. Usage

Le contrôle d'encombrement CCID 2 de style TCP est approprié pour les flux DCCP qui voudraient recevoir autant de bande passante que possible sur le long terme, en cohérence avec l'utilisation du contrôle d'encombrement de bout en bout. Les flux CCID 2 doivent aussi tolérer de grosses variation des caractéristiques de taux d'envoi de contrôle d'encombrement AIMD, incluant de diviser par deux la fenêtre d'encombrement en réponse à un événement d'encombrement.

Les applications qui ont simplement besoin de transférer autant de données que possible dans un délai aussi court que possible devraient utiliser CCID 2. Cela diffère de CCID 3, contrôle en douceur de débit TCP (TFRC, *TCP-Friendly Rate Control*) [RFC4342], qui est approprié pour les flux qui préfèrent minimiser les changements brusques du taux d'envoi. Par exemple, CCID 2 est recommandé plutôt que CCID 3 pour les applications de supports en direct qui mettent en mémoire tampon une quantité considérable de données chez l'application receveuse avant le moment de l'exécution, isolant en quelque sorte l'application des changements abrupts du taux d'envoi. De telles applications pourraient facilement choisir le CCID 2 de DCCP plutôt que TCP lui-même, ajoutant éventuellement une forme de fiabilité sélective à la couche d'application. CCID 2 est aussi recommandé plutôt que CCID 3 pour les applications où la division par deux du taux d'envoi en réponse à l'encombrement ne va probablement pas interférer avec les performances de niveau application.

Un avantage supplémentaire de CCID 2 est que ses mécanismes de contrôle d'encombrement de style TCP sont raisonnablement bien compris, avec des dynamiques de trafic assez similaires à celles de TCP. Bien que la communauté de la recherche sur les réseaux en apprenne toujours sur la dynamique de TCP après 15 années de protocole de transport dominant dans l'Internet, certaines applications peuvent préférer la dynamique mieux connue du contrôle d'encombrement de style TCP à celle de mécanismes de contrôle d'encombrement plus récents, qui n'ont pas encore subi l'épreuve d'un large

déploiement dans l'Internet.

3.1 Relations avec TCP

Les mécanismes de contrôle d'encombrement décrits ici suivent étroitement les mécanismes normalisés par l'IETF pour l'utilisation de TCP fondés sur SACK, et on s'appuie partiellement sur la documentation TCP existante, comme les [RFC0793], [RFC2581], [RFC3465], et [RFC3517]. Le contrôle d'encombrement TCP continue d'évoluer, mais les mises en œuvre de CCID 2 DEVRAIENT attendre des mises à jour explicites de CCID 2 plutôt que de suivre directement l'évolution de TCP.

Les différences entre CCID 2 et le contrôle d'encombrement TCP direct incluent :
- o CCID 2 applique le contrôle d'encombrement aux accusés de réception, un mécanisme non actuellement normalisé pour l'usage de TCP.
- o DCCP est un protocole de datagrammes, de sorte que plusieurs paramètres dont les unités sont spécifiées en octets dans TCP, comme la fenêtre d'encombrement cwnd, ont des unités de paquets dans DCCP.
- o Comme protocole non fiable, DCCP ne retransmet jamais un paquet, de sorte que les mécanismes de contrôle d'encombrement qui distinguent les retransmissions des nouveaux paquets ont été redessinés pour le contexte de DCCP.

3.2 Exemple de demie connexion

Cet exemple montre la progression normale d'une demie connexion utilisant le contrôle d'encombrement de style TCP de CCID 2, n'incluant pas l'initiation et la terminaison de connexion. Cet exemple pour information n'est pas normatif.

1. L'envoyeur envoie des paquets DCCP-Data, où le nombre de paquets envoyés est gouverné par une fenêtre d'encombrement, cwnd, comme dans TCP. Chaque paquet DCCP-Data utilise un numéro de séquence. L'envoyeur envoie aussi une option de caractéristique Ack Ratio *(ratio d'accusés de réception)* qui spécifie le nombre de paquets de données couverts par un paquet Ack du receveur ; le ratio de Ack est deux par défaut. Le champ CCVal de l'en-tête DCCP est réglé à zéro.
   En supposant que la demie connexion est capable de notification explicite d'encombrement (ECN, *Explicit Congestion Notification*) (la caractéristique Incapable de ECN est zéro, par défaut) chaque paquet DCCP-Data est envoyé comme à capacité ECN avec le codet ECT(0) ou ECT(1) établi, comme décrit dans la [RFC3540].

2. Le receveur envoie un paquet DCCP-Ack qui accuse réception des paquets de données pour chaque paquet de données Ack Ratio transmis par l'envoyeur. Chaque paquet DCCP-Ack utilise un numéro de séquence et contient un Vecteur Ack. Le numéro de séquence acquitté dans un paquet DCCP-Ack est celui du paquet reçu avec le plus fort numéro de séquence ; il n'est pas un accusé de réception cumulatif de style TCP.
   Le receveur retourne la somme des noms occasionnels ECN reçus via l'option Vecteur Ack, ce qui permet à l'envoyeur de vérifier de façon probabiliste que le receveur se comporte bien. Les paquets DCCP-Ack provenant du receveur sont aussi envoyés comme capables de ECN, car l'envoyeur va contrôler le taux d'accusés de réception d'une façon en gros favorable à TCP en utilisant la caractéristique Ack Ratio. Il y a peu d'intérêt pour le receveur à vérifier les noms occasionnels de ses paquets DCCP-Ack, car l'envoyeur ne peut pas tirer un parti significatif d'un faux rapport du taux de marquage des Ack.

3. L'envoyeur continue d'envoyer des paquets DCCP-Data selon le contrôle de la fenêtre d'encombrement. À réception des paquets DCCP-Ack, l'envoyeur examine leur Vecteur Ack pour s'informer sur les paquets de données marqués ou éliminés et ajuste sa fenêtre d'encombrement en conséquence. Parce que ce transfert est non fiable, l'envoyeur ne retransmet pas les paquets éliminés.

4. Parce que les paquets DCCP-Ack utilisent des numéros de séquence, l'envoyeur a des informations sur les paquets DCCP-Ack perdus ou marqués. L'envoyeur répond aux paquets DCCP-Ack perdus ou marqués en modifiant le ratio d'Ack envoyé au receveur.

5. L'envoyeur accuse réception des accusés de réception du receveur au moins une fois par fenêtre d'encombrement. Si les deux demies connexions sont actives, l'accusé de réception par l'envoyeur des accusés de réception du receveur est inclus dans l'accusé de réception de l'envoyeur des paquets de données du receveur. Si le chemin inverse de la demie connexion est au repos, l'envoyeur envoie au moins un paquet DCCP-DataAck par fenêtre d'encombrement.

6. L'envoyeur estime les délais d'aller-retour, soit en gardant trace des temps d'aller-retour de l'accusé de réception comme le fait TCP, soit par une option Horodatage explicite, et calcule une valeur de fin de temporisation (TO, *TimeOut*) assez

semblable à la façon dont la fin de temporisation de retransmission (RTO, *Retransmit Timeout*) est calculée dans TCP. Le TO détermine quand un nouveau paquet DCCP-Data peut être transmis quand l'envoyeur a été limité par la fenêtre d'encombrement et qu'aucun retour n'a été reçu du receveur.

4. Établissement de connexion

L'utilisation de Vecteur Ack est OBLIGATOIRE sur les demies connexions CCID 2, de sorte que l'envoyeur DOIT envoyer une option "Change R(Send Vecteur Ack, 1)" au receveur au titre de l'établissement de connexion. L'envoyeur NE DEVRAIT PAS envoyer de données tant qu'il n'a pas reçu le "Confirm L(Send Vecteur Ack, 1)" correspondant du receveur, sauf qu'il PEUT envoyer des données sur les paquets Demande DCCP.

5. Contrôle d'encombrement sur des paquets de données

Les mécanismes de contrôle d'encombrement de CCID 2 se fondent sur ceux de TCP fondé sur SACK [RFC3517], car le Vecteur Ack donne toutes les informations qui peuvent être transmises dans les options SACK.

Un envoyeur de données CCID 2 tient trois paramètres d'entiers mesurés dans les paquets.

1. La fenêtre d'encombrement "cwnd", qui est égale au nombre maximum de paquets de données permis dans le réseau à tout moment. ("paquet de données" signifie tout paquet DCCP qui contient des données d'utilisateur : DCCP-Data, DCCP-DataAck, et occasionnellement DCCP-Request et DCCP-Response.)

2. Le seuil de démarrage lent "ssthresh", qui contrôle les réglages de cwnd.

3. La valeur du tuyau "pipe", qui est l'estimation de l'envoyeur du nombre de paquets de données en instance dans le réseau.

Ces paramètres sont manipulés, et leurs valeurs initiales déterminées, en accord avec le comportement de TCP fondé sur SACK, sauf qu'ils sont mesurés en paquets, et non en octets. Le reste de la section donne des indications plus spécifiques.

L'envoyeur PEUT envoyer un paquet de données quand "pipe" < cwnd mais NE DOIT PAS envoyer un paquet de données quand "pipe" ≥ cwnd. Chaque paquet de données envoyé augmente "pipe" de 1.

L'envoyeur réduit "pipe" lorsque il déduit que les paquets de données ont quitté le réseau, soit en les ayant reçus, soit qu'ils ont été éliminés. En particulier :

1. Paquets de données acquittés. L'envoyeur réduit "pipe" de 1 pour chaque paquet de données nouvellement acquitté lorsque il a reçu (Vecteur Ack State 0 ou State 1) d'un DCCP-Ack.

2. Paquets de données éliminés. L'envoyeur réduit "pipe" de 1 pour chaque paquet de données qu'il peut déduire comme perdu à cause de l'équivalent DCCP de "accusé de réception dupliqué" de TCP. Cela dépend du paramètre NUMDUPACK, le nombre d'accusés de réception dupliqués nécessaires pour déduire une perte. Le paramètre NUMDUPACK est réglé à trois, comme c'est couramment le cas dans TCP. Un paquet P est déduit comme perdu plutôt que retardé, quand au moins NUMDUPACK paquets transmis après P ont été acquittés comme reçus (Vecteur Ack State 0 ou 1) par le receveur. Noter que les paquets acquittés après le trou peuvent être des DCCP-Ack ou d'autres paquets non de données.

3. Fins de temporisation de transmission. Finalement, l'envoyeur a besoin de transmettre les fins de temporisation, traitées comme de fins de temporisation de retransmission de TCP, dans le cas où une fenêtre de paquets entière est perdue. L'envoyeur estime le délai d'aller retour au plus une fois par fenêtre de données et utilise les algorithmes de TCP pour maintenir le délai d'aller-retour moyen, l'écart moyen, et la valeur de fin de temporisation [RFC2988]. (Si plus d'une mesure par temps d'aller-retour a été utilisée pour ces calculs, alors les pondérations des éléments participant à la moyenne devraient être ajustées pour s'assurer que le délai d'aller retour moyen est effectivement déduit de mesures sur plusieurs délais d'aller-retour.) Parce que DCCP ne retransmet pas les données, DCCP n'exige pas la temporisation minimum recommandée d'une seconde de TCP. Le recul exponentiel du temporisateur est exactement comme dans TCP. Quand une fin de temporisation de transmission se produit, l'envoyeur règle "pipe" à zéro. Les ajustements à cwnd et ssthresh sont décrits ci-dessous.

L'envoyeur NE DOIT PAS décrémenter pipe plus d'une fois par paquet de données. Les vrais accusés de réception dupliqués, par exemple, NE DOIVENT PAS affecter pipe. L'envoyeur NE DOIT aussi PAS décrémenter pipe à nouveau à réception de l'accusé de réception d'un paquet précédemment déduit comme perdu. De plus, l'envoyeur NE DOIT PAS décrémenter pipe pour des paquets non de données, comme les DCCP-Ack, même si le Vecteur Ack contient des informations sur eux.

Les événements d'encombrement causent la réduction par CCID 2 de sa fenêtre d'encombrement. Un événement d'encombrement contient au moins un paquet perdu ou marqué. Comme dans TCP, deux pertes ou marques sont considérées faire partie d'un seul événement d'encombrement quand le second paquet a été envoyé avant la détection de la perte ou marque du premier paquet. Pour une approximation, un envoyeur peut considérer que deux pertes ou marques font partie d'un seul événement d'encombrement quand les paquets ont été envoyés dans une estimation de RTT d'un autre, en utilisant une estimation de RTT en cours au moment de l'envoi des paquets. Pour chaque événement d'encombrement, indiqué explicitement comme un accusé de réception Vecteur Ack State 1 (marqué ECN) ou déduit via des "accusés de réception dupliqués", cwnd est diminué de moitié, puis ssthresh est réglé au nouveau cwnd. Cwnd n'est jamais réduit en dessous d'un paquet. Après une fin de temporisation, le seuil de démarrage lent est réglé à cwnd/2, puis cwnd est réglé à un paquet. Quand ils sont diminués de moitié, les valeurs de cwnd et ssthresh sont arrondies, sauf que cwnd n'est jamais moins de un et ssthresh n'est jamais moins de deux.

Quand cwnd < ssthresh, ce qui signifie que l'envoyeur est en démarrage lent, la fenêtre d'encombrement est augmentée d'un paquet tous les deux paquets de données nouvellement acquittés avec Vecteur Ack State 0 (non marqué ECN) jusqu'à un maximum de Ack Ratio/2 paquets par accusé de réception. C'est une forme modifiée du compte d'octets approprié [RFC3465] qui est cohérent avec la norme TCP actuelle (qui n'inclut pas de compte d'octets) mais permet à CCID 2 d'augmenter aussi agressivement que TCP quand le Ack Ratio de CCID 2 est supérieur à la valeur par défaut de deux. Quand cwnd ≥ ssthresh, la fenêtre d'encombrement est augmentée d'un paquet pour chaque fenêtre de données acquittée sans paquet perdu ou marqué. Le paramètre cwnd est initialisé à au plus quatre paquets pour les nouvelles connexions, suivant les règles de la [RFC3390] ; le paramètre ssthresh est initialisé à une valeur arbitrairement haute.

Les envoyeurs PEUVENT utiliser une forme de rythme fondée sur le taux quand ils envoient de multiples paquets de données libérés par un seul paquet Ack plutôt que d'envoyer tous les paquets de données libérés dans une seule salve.

5.1 Réponse aux périodes de repos et d'application limitée

CCID 2 est conçu pour suivre les mécanismes de contrôle d'encombrement de TCP dans toute la mesure du possible, mais TCP n'a pas une normalisation complète de sa réponse de contrôle d'encombrement aux périodes d'inactivité (quand aucun paquet de données n'est envoyé) ou aux périodes d'application limitée (quand le taux d'envoi est inférieur à ce qui est permis par cwnd). Ce paragraphe est un bref guide des normes de TCP dans ce domaine.

Pour les périodes inactives, la [RFC2581] recommande que l'envoyeur TCP DEVRAIT faire un démarrage lent après une période d'inactivité, où une période d'inactivité est définie comme une période excédant l'intervalle de temporisation. La [RFC2861], actuellement expérimentale, suggère un mécanisme légèrement plus modéré où la fenêtre d'encombrement est diminuée de moitié pour chaque délai d'aller-retour où l'envoyeur est resté inactif.

Il n'y a actuellement pas de norme gouvernant l'usage par TCP de la fenêtre d'encombrement durant une période d'application limitée. En particulier, il est possible que la fenêtre d'encombrement de TCP devienne assez grosse durant une longue période sans encombrement quand l'envoyeur est d'application limitée, envoyant à un faible taux. La [RFC2861] suggère essentiellement que la fenêtre d'encombrement de TCP ne soit pas augmentée durant les périodes d'application limitée quand la fenêtre d'encombrement n'est pas pleinement utilisée.

5.2 Réponse aux données éliminées et au receveur lent

L'option Données éliminées de DCCP permet à un receveur de déclarer qu'un paquet a été éliminé chez l'hôte d'extrémité avant sa livraison à l'application -- par exemple, à cause de corruption ou d'un débordement de la mémoire tampon de réception. L'option Receveur lent de DCCP permet à un receveur de déclarer qu'il a des problèmes pour faire face aux paquets de l'envoyeur, bien que rien n'ait encore été éliminé. Les envoyeurs CCID 2 répondent à ces options comme décrit dans la [RFC4340], avec les précisions suivantes.

- Code d'abandon 2 ("élimination de mémoire tampon de réception"). La fenêtre d'encombrement "cwnd" est réduite de un à chaque paquet nouvellement acquitté comme Code d'abandon 2, sauf qu'elle n'est jamais réduite en dessous de un.

o Sortie de démarrage lent. L'envoyeur DOIT sortir du démarrage lent chaque fois qu'il reçoit une option Données éliminées ou Receveur lent pertinente.

5.3 Taille de paquet

CCID 2 est optimisé pour les applications qui utilisent généralement une taille fixe de paquet et font varier leur taux d'envoi en paquets par seconde en réponse à l'encombrement. CCID 2 n'est pas approprié pour les applications qui requièrent un intervalle de temps fixe entre les paquets et font varier leur taille de paquet au lieu de leur taux de paquets en réponse à l'encombrement.

CCID 2 tient une fenêtre d'encombrement en paquets et n'augmente pas la fenêtre d'encombrement en réponse à une diminution de la taille de paquet. Cependant, une certaine attention pourrait être requise pour les applications utilisant CCID 2 qui font varier leur taille de paquet non en réponse à l'encombrement, mais en réponse à d'autres exigences de niveau application.

Les mises en œuvre de CCID 2 PEUVENT vérifier que les applications ne manipulent pas la taille de paquet de façon inappropriée. Par exemple, une application pourrait envoyer de petits paquets pendant un moment, croissant à un taux rapide, puis passer à de grands paquets qui tirent parti du taux rapide. (Des simulations préliminaires indiquent que des applications peuvent n'être pas capables d'augmenter leur taux de transfert global de cette façon, donc il n'est pas clair que cette manipulation puisse se produire en pratique [V03].)

6. Accusé de réception

Les accusés de réception CCID 2 sont généralement régulés par les paquets de données de l'envoyeur . Chaque accusé de réception requis DOIT contenir les options Vecteur Ack qui déclarent exactement quels paquets sont arrivés et si ces paquets étaient marqués ECN. Les données d'accusé de réception dans les options Vecteur Ack DEVRAIENT généralement couvrir la fenêtre d'accusés de réception entière du receveur ; voir le paragraphe 11.4.2 de la [RFC4340]. Toutes les options Données éliminées DEVRAIENT de même couvrir la fenêtre d'accusés de réception entière du receveur.

Les envoyeurs CCID 2 utilisent la caractéristique Ack Ratio de DCCP pour influencer le taux de génération des paquets DCCP-Ack par les receveurs, contrôlant donc l'encombrement du chemin inverse. Cela diffère de TCP, qui n'a présentement pas de contrôle d'encombrement pour le pur trafic d'accusés de réception. Le contrôle d'encombrement du chemin inverse de CCID 2 n'essaye pas d'être favorable à TCP ; il essaye juste d'éviter le collapsus d'encombrement, et d'être un peu meilleur que TCP en présence d'une forte perte de paquets ou d'un fort taux de marquage sur le chemin inverse. Le Ack Ratio par défaut est de deux, et CCID 2 avec cet Ack Ratio se comporte comme TCP avec des accusés de réception retardés. Le paragraphe 11.3 de la [RFC4340] décrit le Ack Ratio plus en détails, incluant sa relation à la régulation des paquets d'accusé de réception et de DCCP-DataAck. La Section 6 du présent document décrit comment un envoyeur CCID 2 détecte les accusés de réception perdus ou marqués, et le paragraphe 6.1.2 décrit comment il change le Ack Ratio.

6.1 Contrôle d'encombrement sur les accusés de réception

Quand Ack Ratio est R, le receveur envoie un paquet DCCP-Ack par, plus ou moins, R paquets de données. Comme l'envoyeur envoie cwnd paquets de données par temps d'aller-retour, le taux d'accusés de réception est égal à cwnd/R DCCP-Ack par temps d'aller-retour. L'envoyeur garde le taux d'accusés de réception à peu près favorable à TCP en surveillant le flux d'accusés de réception pour les paquets DCCP-Ack perdus et marqués et modifie R en conséquence. Pour chaque RTT contenant un événement d'encombrement DCCP-Ack (c'est-à-dire, un DCCP-Ack perdu ou marqué) l'envoyeur diminue de moitié le taux d'accusés de réception en doublant Ack Ratio ; pour chaque RTT ne contenant pas d'événement d'encombrement DCCP-Ack, il augmente le taux d'accusés de réception par une diminution graduelle de Ack Ratio.

6.1.1 Détection des accusés de réception perdus et marqués

Tous les paquets provenant du receveur contiennent des numéros de séquence, de sorte que l'envoyeur peut détecter les pertes et les marques sur les paquets du receveur. L'envoyeur déduit les pertes de paquet du receveur de la même façon qu'il déduit les pertes de ses paquets de données : un paquet provenant du receveur est considéré perdu après qu'au moins

NUMDUPACK paquets avec des numéros de séquence supérieurs ont été reçus.

Les paquets DCCP-Ack sont généralement petits, de sorte qu'ils imposent moins de charge sur les liaisons de réseau encombrées que les paquets DCCP-Data et DCCP-DataAck.
Pour cette raison, Ack Ratio dépend des pertes et marques sur les paquets non de données du receveur, et non des pertes et marques agrégées sur tous les paquets du receveur. La catégorie de paquets non de données consiste en les types de paquets qui ne peuvent pas porter des données d'application : DCCP-Ack, DCCP-Close, DCCP-CloseReq, DCCP-Reset, DCCP-Sync, et DCCP-SyncAck. L'envoyeur peut facilement distinguer les marques non de données des autres marques. C'est cependant plus difficile pour les pertes, car l'envoyeur ne peut pas toujours savoir si un paquet perdu portait des données. Sauf si il a de meilleures informations, l'envoyeur DEVRAIT supposer, pour les besoins du calcul de Ack Ratio, que chaque paquet perdu était un paquet non de données. De meilleures informations sont disponibles via l'option DCCP de compte NDP, si nécessaire. (L'Appendice B discute des coûts de prendre à tort une perte de paquet de données pour une perte de paquet de non de données.)

Un receveur qui met en œuvre son propre contrôle d'encombrement d'accusés de réception indépendant de Ack Ratio NE DEVRAIT PAS réduire son taux d'accusés de réception de DCCP-Ack à cause de pertes ou marques sur ses paquets de données.

6.1.2 Changement du ratio d'accusés de réception

Le ratio d'accusés de réception satisfait toujours à trois contraintes : (1) Ack Ratio est un entier ; (2) Ack Ratio n'excède pas cwnd/2, arrondi, sauf que Ack Ratio 2 est toujours acceptable ; (3) Ack Ratio est deux ou plus pour une fenêtre d'encombrement de quatre paquets ou plus.

L'envoyeur change Ack Ratio dans les limites de ces contraintes comme suit. Pour chaque fenêtre d'encombrement de données avec des paquets DCCP-Ack perdus ou marqués, Ack Ratio est doublé ; et pour chaque cwnd/(R^2 - R) fenêtres d'encombrement consécutives de données sans paquet DCCP-Ack perdu ou marqué, Ack Ratio est diminué de 1. (Voir le calcul à l'Appendice A.) Les changements de Ack Ratio sont signalés par la négociation des caractéristiques ; voir le paragraphe 11.3 de la [RFC4340].

Pour une fenêtre d'encombrement constante, cela donne un taux d'envoi de Ack qui est à peu près favorable à TCP. Bien sûr, cwnd varie généralement dans le temps ; la dynamique est assez complexe, mais en gros favorable à TCP. On recommande que l'envoyeur utilise la plus récente valeur de cwnd quand il détermine si il diminue Ack Ratio de 1.

L'envoyeur n'a pas besoin de garder le Ack Ratio complètement à jour. Par exemple, il PEUT limiter les renégociations de taux de Ack Ratio à une fois tous les quatre ou cinq temps d'aller-retour, ou à une fois par une ou deux secondes. L'envoyeur NE DEVRAIT PAS tenter de renégocier le Ack Ratio plus d'une fois par temps d'aller-retour. De plus, il PEUT appliquer un minimum d'Ack Ratio de deux, ou il PEUT régler Ack Ratio à un pour les demies connexions qui ont une fenêtre d'encombrement persistantes de 1 ou 2 paquets.

L'un dans l'autre, le receveur envoie toujours au moins un accusé de réception par fenêtre de données quand cwnd = 1, et au moins deux accusés de réception par fenêtre de données autrement. Donc, le receveur pourrait envoyer deux paquets d'accusé de réception par fenêtre de données même en présence d'un très lourd encombrement sur le chemin inverse. On notera cependant que si l'encombrement est suffisamment lourd, tous les paquets d'accusé de réception sont éliminés, et alors l'envoyeur revient aux temporisations à croissance exponentielle du retard, comme dans TCP. Donc, si l'encombrement est suffisamment lourd sur le chemin inverse, l'envoyeur réduit alors son taux d'envoi sur le chemin de transmission, ce qui réduit aussi le taux sur le chemin inverse.

6.2 Accusé de réception des accusés de réception

Un envoyeur DCCP actif A DOIT occasionnellement accuser réception des accusés de réception de son homologue DCCP B afin que le DCCP B puisse libérer l'état du Vecteur Ack. Quand les deux demies connexions sont actives, les accusés de réception de A des accusés de réception de B sont automatiquement contenues dans les accusés de réception de A des données de B. Cependant, si la demie-connexion de B à A est au repos, DCCP A doit occasionnellement envoyer des accusés de réception de façon proactive, comme d'envoyer un paquet DCCP-DataAck qui inclut un numéro d'accusé de réception dans l'en-tête.

Un envoyeur actif DEVRAIT accuser réception des accusés de réception du receveur au moins une fois par fenêtre d'encombrement. Bien sûr, l'application de l'envoyeur peut rester silencieuse. Cela ne pose pas de problème ; quand aucun

des côtés n'envoie de données, un envoyeur peut attendre une durée d'une longueur arbitraire avant d'envoyer un ack.

6.2.1 Détermination de repos

Ce paragraphe décrit comment un receveur CCID 2 détermine que l'envoyeur correspondant n'envoie aucune donnée et est donc passé au repos. Voir au paragraphe 11.1 de la [RFC4340] les informations générales sur le repos.

Soit T égal au plus grand de 0,2 seconde et deux temps d'aller-retour. (Le receveur peut connaître le temps d'aller-retour dans son rôle d'envoyeur pour l'autre demie-connexion. Si il ne le connaît pas, il devrait utiliser un RTT par défaut de 0,2 s, comme décrit au paragraphe 3.4 de la [RFC4340].) Une fois que l'envoyeur a accusé réception des Vecteurs Ack du receveur et que l'envoyeur n'a pas envoyé de données supplémentaires pendant au moins T secondes, le receveur peut déduire que l'envoyeur est au repos. Plus précisément, le receveur déduit que l'envoyeur est passé au repos quand au moins T secondes se sont écoulées sans qu'il reçoive de données de l'envoyeur, et quand l'envoyeur a accusé réception des Vecteurs Ack du receveur qui couvrent tous les paquets de données reçus chez le receveur.

7. Notification explicite d'encombrement

CCID 2 prend en charge la notification explicite d'encombrement (ECN, *Explicit Congestion Notification*) [RFC3168]. L'envoyeur va utiliser le nom occasionnel ECN pour les paquets de données, et le receveur va faire écho à ces noms occasionnels dans ses vecteurs Ack, comme spécifié au paragraphe 12.2 de la [RFC4340]. Les informations sur les paquets marqués sont aussi retournées dans le Vecteur Ack. Parce que les informations dans le Vecteur Ack sont transférées de façon fiable, DCCP n'a pas besoin des fanions TCP Écho ECN et Fenêtre d'encombrement réduite.

Pour les paquets de données non marqués, le receveur calcule l'écho de nom occasionnel ECN comme dans la [RFC3540] et le retourne au titre de ses options de Vecteur Ack. L'envoyeur DEVRAIT vérifier ces échos de nom occasionnel ECN par rapport aux valeurs attendues, se protégeant donc contre la dissimulation accidentelle ou malveillante de paquets marqués.

Parce que les accusés de réception CCID 2 sont à encombrement contrôlé, ECN peut aussi être utilisé pour ses accusés de réception. Dans ce cas, on n'utilise pas le nom occasionnel ECN, parce qu'il ne serait pas aisé d'assurer la protection contre la dissimulation des paquets d'accusé de réception marqués par l'envoyeur, et parce que l'envoyeur n'aurait pas de motif de mentir sur le taux de marquage sur les accusés de réception.

8. Options et caractéristiques

L'option DCCP Vecteur Ack, et ses caractéristiques Capacité ECN, Ratio d'Ack, et Vecteur d'Ack envoyé, sont pertinentes pour CCID 2.

9. Considérations sur la sécurité

Les considérations sur la sécurité pour DCCP ont été discutées dans la [RFC4340], et les considérations de sécurité pour TCP ont été discutées dans la [RFC2581].

La [RFC2581] discute les façons dont un attaquant pourrait dégrader les performance d'une connexion TCP en éliminant des paquets, ou en fabricant des accusés de réception supplémentaires dupliqués ou des accusés de réception pour de nouvelles données. On n'a pas connaissance de nouvelles considérations de sécurité créées par le présent document dans son utilisation du contrôle d'encombrement de style TCP.

10. Considérations relatives à l'IANA

La présente spécification définit la valeur 2 dans l'espace de noms DCCP CCID géré par l'IANA. Cette allocation est aussi mentionnée dans la [RFC4340]. CCID 2 introduit aussi trois ensembles de nombres dont les valeurs devraient être allouées par l'IANA, à savoir les codes de réinitialisation spécifiques de CCID 2, les types d'option, et les numéros de caractéristiques. Ces gammes vont empêcher toute future allocation spécifique de CCID 2 de polluer les espaces de noms

globaux correspondants de DCCP ; voir au paragraphe 10.3 de la [RFC4340]. Cependant, le présent document ne fait pas d'allocation particulière d'une gamme, sauf pour l'usage expérimental et d'essais de la [RFC3692]. On se réfère à la politique d'action de normalisation mentionnée dans la [RFC2434].

10.1 Codes de réinitialisation

Chaque entrée dans le registre de code de réinitialisation DCCP CCID 2 contient un code de réinitialisation spécifique de CCID 2, qui est un nombre dans la gamme de 128 à 255 ; une brève description du code de réinitialisation, et une référence à la RFC qui définit le code de réinitialisation. Les codes de réinitialisation de 184 à 190 et de 248 à 254 sontréservés de façon permanente pour un usage expérimental et d'essai . Les codes de réinitialisation restants – 128 à 183, 191 à 247, et 255 -- sont actuellement réservés et devraient être alloués selon la politique d'action de normalisation, qui exige la revue et l'approbation par l'IESG et la publication d'une RFC sur la voie de la normalisation de l'IETF.

10.2 Types d'option

Chaque entrée dans le registre DCCP CCID 2 de type d'option contient un type d'option spécifique de CCID 2, qui est un nombre dans la gamme 128 à 255, le nom de l'option, et une référence à la RFC qui définit le type d'option. Les types d'option 184 à 190 et 248 à 254 sont réservés de façon permanente pour un usage expérimental et d'essai. Les types d'option restants – 128 à 183, 191 à 247, et 255 -- sont actuellement réservés et devraient être alloués selon la politique d'action de normalisation, qui exige la revue et l'approbation par l'IESG et la publication d'une RFC sur la voie de la normalisation de l'IETF.

10.3 Numéros de caractéristiques

Chaque entrée dans le registre DCCP CCID 2 des numéros de caractéristique contient un numéro de caractéristique spécifique de CCID 2, qui est un nombre dans la gamme 128 à 255, le nom de la caractéristique et une référence à la RFC qui définit le numéro de caractéristique. Les numéros de caractéristique 184 à 190 et 248 à 254 sont réservés de façon permanente pour un usage expérimental et d'essai. Les numéros de caractéristique restants -- 128 à 183, 191 à 247, et 255 -- sont actuellement réservés et devraient être alloués selon la politique d'action de normalisation, qui exige la revue et l'approbation par l'IESG et la publication d'une RFC sur la voie de la normalisation de l'IETF.

11. Remerciements

Merci à Mark Handley et Jitendra Padhye de leur aide pour définir CCID 2. Merci aussi à Mark Allman, Aaron Falk, Nils-Erik Mattsson, Greg Minshall, Arun Venkataramani, Magnus Westerlund, et aux membres du groupe de travail DCCP pour leurs réactions sur ce document.

Appendice A. Déduction de diminution du ratio d'accusé de réception

Cette Section justifie l'algorithme pour augmenter et diminuer le ratio d'accusés de réception donné au paragraphe 6.1.2.

La phase d'évitement d'encombrement de TCP divise par deux le cwnd pour chaque fenêtre avec encombrement. De même, CCID 2 double le ratio d'accusé de réception pour chaque fenêtre avec encombrement sur le chemin de retour, divisant en gros par deux le taux d'envoi de DCCP-Ack.

La phase d'évitement d'encombrement de TCP augmente cwnd de une taille de segment maximum pour chaque fenêtre libre d'encombrement. Quand ce comportement d'évitement d'encombrement est appliqué au trafic d'accusés de réception, cela va correspondre à augmenter le nombre de paquets DCCP-Ack par fenêtre de un après chaque fenêtre de paquets DCCP-Ack libre d'encombrement. On ne peut pas réaliser cela exactement en utilisant Ack Ratio, car c'est un entier. On doit plutôt diminuer Ack Ratio de un après que K fenêtres ont été envoyées sans un événement d'encombrement sur le chemin inverse, où K est choisi de façon à ce que le nombre de paquets DCCP-Ack par fenêtre d'encombrement à long terme soit en gros favorable à TCP, suivant le contrôle d'encombrement AIMD.

Dans CCID 2, un trafic d'accusés de réception en gros favorable à TCP peut être réalisé en réglant K à cwnd/(R^2 - R) où R est le ratio d'accusés de réception courant.

Ce résultat a été calculé comme suit :

R = Ack Ratio = nombre de paquets de données par paquets d'accusé de réception,
W = Fenêtre d'encombrement = nombre de paquets de données par fenêtre,
W/R = nombre de paquets d'accusé de réception par fenêtre.

Exigence : augmenter W/R de 1 par fenêtre libre d'encombrement. Comme on peut seulement réduire R par incréments de un, on trouve K de sorte que, après K fenêtres libres d'encombrement, W/R + K va être égal à W/(R-1).

(W/R) + K = W/(R-1), donc K = W/(R-1) - W/R = W/(R^2 - R).

Appendice B. Coût des erreurs d'inférence de pertes sur le ratio d'accusés de réception

Comme exposé au paragraphe 6.1.1, souvent l'envoyeur ne peut pas déterminer si les paquets perdus portaient des données. Cela entrave sa capacité à séparer les événements de perte non de données des autres événements de perte. En l'absence de meilleures informations, l'envoyeur suppose, pour les besoins du calcul du ratio d'accusés de réception, que tous les paquets perdus étaient des paquets non de données. Cela peut surestimer le taux d'événements de perte de non données, ce qui peut conduire à un trop fort ratio de Ack, et donc un taux d'accusé de réception trop faible. Toutes les informations d'accusé de réception vont quand même arriver – les accusés de réception DCCP sont fiables -- mais les informations d'accusé de réception vont arriver de façon plus saccadée. En l'absence de toute forme de régulation fondée sur le taux, cela pourrait conduire à augmenter la sporadicité du trafic de données pour l'envoyeur.

Il y a plusieurs cas de problème de ratio d'accusé de réception trop élevé où l'augmentation de sporadicité résultante du trafic de données ne va pas se produire. En particulier, si on appelle le receveur DCCP B et l'envoyeur DCCP A :

- Le problème ne se posera pas à moins que DCCP B envoie lui-même une quantité significative de données. Quand la demie connexion de B à A est au repos ou à taux faible, la plupart des paquets envoyés par DCCP B vont, en fait, être de purs accusés de réception, et l'estimation par DCCP A du taux de perte de DCCP-Ack va être raisonnablement précise.

- Le problème ne se posera pas si DCCP B fait habituellement porter ses informations d'accusé de réception dans ses paquets de données. Les accusés de réception portés ne sont pas limités par le ratio d'Ack, de sorte qu'ils peuvent arriver assez fréquemment pour empêcher la sporadicité.

- Le problème ne se posera pas si le taux d'envoi de DCCP A est faible, car la sporadicité n'est pas un problème aux faibles débits.

- Le problème ne se posera pas si le taux d'envoi de DCCP B est élevé par rapport à celui de DCCP A, car le taux de perte de B vers A doit être faible pour supporter le taux d'envoi de DCCP. Cela limite le ratio d'accusé de réception à des valeurs raisonnables même quand DCCP A étiquette chaque perte comme une perte de DCCP-Ack.

- Le problème ne se posera pas si DCCP B envoie des options Compte NDP quand c'est approprié (quand la caractéristique Compte NDP envoyé/B est vraie). L'envoyeur peut alors utiliser les options Compte NTP du receveur pour détecter, dans la plupart des cas, si les paquets perdus étaient des paquets de données ou des DCCP-Ack.

- Finalement, le problème ne va pas se poser si DCCP A régule le taux de ses paquets de données.

Cela laisse le cas où DCCP B envoie en gros la même quantité de paquets de données et de paquets non de données, sans options Compte NDP, et avec toutes les informations d'accusé de réception dans les paquets DCCP-Ack. On quantifie le coût potentiel, en termes de ratio d'accusés de réception trop grand, dû au mauvais classement des pertes de paquet de données par l'envoyeur comme des pertes de DCCP-Ack. Pour simplifier, on suppose un environnement de multiplexage statistique à grande échelle où le taux d'élimination de paquets est indépendant du taux d'envoi de toute connexion individuelle.

En supposant que quand DCCP A compte correctement les pertes non de données, le ratio d'Ack est réglé de telle façon que le trafic de données de B à A et d'accusés de réception aient tous deux un taux d'envoi de D paquets par seconde. Ensuite, quand DCCP A compte incorrectement les pertes de données comme pertes non de données, le taux d'envoi pour le trafic de données de B à A est toujours D paquets par seconde, mais le taux d'envoi réduit pour le trafic d'accusés de réception de

B à A est f*D paquets par seconde, avec f < 1. Soit p le taux de perte de paquet. L'envoyeur estime incorrectement le taux de perte de non données à (pD + pfD)/fD, ou, en simplifiant, p(1 + 1/f). Comme le mécanisme de contrôle d'encombrement pour le trafic d'accusé de réception est en gros favorable à TCP, et donc que les taux d'envoi non de données et de données croissent tous deux comme 1/√x) pour le taux d'élimination de paquet x, on a :

fD/D = √(p)/√(p(1 + 1/f)),

donc

f^2 = 1/(1 + 1/f).

La résolution nous donne f = 0,62. Si l'envoyeur compte incorrectement les paquets de données perdus comme non de données dans ce scénario, le taux d'accusés de réception est diminué d'un facteur 0,62. Il en résulterait une augmentation modérée de la sporadicité du trafic de données de A à B, qui pourrait être atténuée par l'envoi d'options Compte NDP ou d'accusés de réception portés, ou en régulant le taux d'envoi des données.

Références normatives

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

[RFC2018] M. Mathis et autres, "Options d'accusé de réception sélectif sur TCP", octobre 1996. (*Remplace* RFC1072) *(P.S.)*

[RFC2119] S. Bradner, "Mots clés à utiliser dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. *(MàJ par* RFC8174)

[RFC2434] T. Narten et H. Alvestrand, "Lignes directrices pour la rédaction d'une section Considérations relatives à l'IANA dans les RFC", BCP 26, octobre 1998. *(Rendue obsolète par la* RFC5226)

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

[RFC2988] V. Paxson, M. Allman, "Calcul du temporisateur de retransmission de TCP", novembre 2000. *(P.S.)(Obs., voir* RFC6298)

[RFC3168] K. Ramakrishnan et autres, "Ajout de la notification explicite d'encombrement (ECN) à IP", septembre 2001. *(P.S. ; MàJ par* RFC8311)

[RFC3390] M. Allman, S. Floyd, C. Partridge, "Augmentation de la fenêtre initiale de TCP", octobre 2002. *(P.S.)*

[RFC3517] E. Blanton et autres, "Algorithme de récupération de perte fondé sur l'accusé de réception sélectif prudent (SACK) pour TCP", avril 2003. (*Remplacée par* RFC6675) *(P.S.)*

[RFC3692] T. Narten, "L'allocation de numéros expérimentaux et d'essai est considérée comme utile", janvier 2004. (BCP0082)

[RFC4340] E. Kohler et autres, "Protocole de contrôle d'encombrement de datagrammes (DCCP)", mars 2006. *(P.S.) (MàJ par* 6773)

Références pour information

[RFC2861] M. Handley, J. Padhye, S. Floyd, "Validation de fenêtre d'encombrement TCP", juin 2000. *(Historique, Remplacée par* RFC7661)

[RFC3465] M. Allman, "Contrôle d'encombrement sur TCP avec compte d'octets approprié (ABC)", février 2003. *(Expérimentale)*

[RFC3540] N. Spring, D. Wetherall, D. Ely, "Signalisation de notification robuste d'encombrement explicite (ECN) avec des noms occasionnels", juin 2003. *(Expérimentale)*

[RFC4342] S. Floyd et autres, "Profil d'identifiant 3 de protocole de contrôle d'encombrement de datagrammes (DCCP) : Contrôle en douceur de débit TCP (TFRC)", mars 2006. *(P.S. ; MàJ par* RFC5348, RFC8311)

[V03] Arun Venkataramani, août 2003. Citation pour mémoire.

Adresse des auteurs

Sally Floyd
ICSI Center for Internet Research
1947 Center Street, Suite 600
Berkeley, CA 94704
USA
mél : floyd@icir.org

Eddie Kohler
4531C Boelter Hall
UCLA Computer Science Department
Los Angeles, CA 90095
USA
mél : kohler@cs.ucla.edu

Déclaration complète de droits de reproduction

Remerciement
Le financement de la fonction d'édition des RFC est fourni par l'activité de soutien administratif (IASA) de l'IETF.